



Cleavage Specificity Analysis of Six Type II Transmembrane Serine Proteases (TTSPs) Using PICS with Proteome-Derived Peptide Libraries

Olivier Barré^{1,9}, Antoine Dufour^{1,9}, Ulrich Eckhard^{1,9}, Reinhild Kappelhoff^{1,9}, François Béliveau², Richard Leduc², Christopher M. Overall^{1,3*}

1 Centre for Blood Research, Department of Oral Biological & Medical Sciences, University of British Columbia, Vancouver, BC, Canada, **2** Department of Pharmacology, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada, **3** Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada

Abstract

Background: Type II transmembrane serine proteases (TTSPs) are a family of cell membrane tethered serine proteases with unclear roles as their cleavage site specificities and substrate degradomes have not been fully elucidated. Indeed just 52 cleavage sites are annotated in MEROPS, the database of proteases, their substrates and inhibitors.

Methodology/Principal Finding: To profile the active site specificities of the TTSPs, we applied Proteomic Identification of protease Cleavage Sites (PICS). Human proteome-derived database searchable peptide libraries were assayed with six human TTSPs (matriptase, matriptase-2, matriptase-3, HAT, DESC and hepsin) to simultaneously determine sequence preferences on the N-terminal non-prime (P) and C-terminal prime (P') sides of the scissile bond. Prime-side cleavage products were isolated following biotinylation and identified by tandem mass spectrometry. The corresponding non-prime side sequences were derived from human proteome databases using bioinformatics. Sequencing of 2,405 individual cleaved peptides allowed for the development of the family consensus protease cleavage site specificity revealing a strong specificity for arginine in the P1 position and surprisingly a lysine in P1' position. TTSP cleavage between R↓K was confirmed using synthetic peptides. By parsing through known substrates and known structures of TTSP catalytic domains, and by modeling the remainder, structural explanations for this strong specificity were derived.

Conclusions: Degradomics analysis of 2,405 cleavage sites revealed a similar and characteristic TTSP family specificity at the P1 and P1' positions for arginine and lysine in unfolded peptides. The prime side is important for cleavage specificity, thus making these proteases unusual within the tryptic-enzyme class that generally has overriding non-prime side specificity.

Citation: Barré O, Dufour A, Eckhard U, Kappelhoff R, Béliveau F, et al. (2014) Cleavage Specificity Analysis of Six Type II Transmembrane Serine Proteases (TTSPs) Using PICS with Proteome-Derived Peptide Libraries. PLoS ONE 9(9): e105984. doi:10.1371/journal.pone.0105984

Editor: Alexander Wlodawer, NCI-Frederick, United States of America

Received: May 11, 2011; **Accepted:** July 31, 2014; **Published:** September 11, 2014

Copyright: © 2014 Barré et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was funded by Proteolytic Signatures in Breast Cancer C.I.H.R. Grant M.O.P. 111058. O.B. was supported by the Swiss National Science Foundation (SNF) and a post-doctoral fellowship by the Canadian Institutes of Health Research (CIHR). A.D. and U.E. were supported by post-doctoral fellowships from CIHR and the Michael Smith Foundation for Health Research (MSFHR), respectively. A Canada Research Chair in Metalloproteinase Proteomics and Systems Biology supports C.M.O. This work was supported by grants to C.M.O. and R.L. from the CIHR and the Canadian Cancer Society. The University of British Columbia, Centre for Blood Research Mass Spectrometry Suite is supported by the Canada Foundation for Innovation (CFI) and MSFHR. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: chris.overall@ubc.ca

⁹ These authors contributed equally to this work.

Introduction

Pericellular proteolysis is involved in many important cellular processes such as the transduction of signals across the cell membrane, the release of bioactive growth factors, cytokines and peptide hormones, as well as the interactions with other cells, basement membrane and extracellular matrix proteins [1–3]. The increasing relevance of these processes at the cell surface has focused attention on membrane-associated proteolytic systems, including the family of type II transmembrane serine proteases (TTSPs) [4,5]. The TTSP family is composed of more than 20 members that share a number of structural features including an N-terminal cytoplasmic domain, a transmembrane domain, a

central region containing various domains potentially involved in protein-protein interaction, and a C-terminal extracellular serine protease domain (Figure 1A). TTSPs are synthesized as inactive single chain proenzyme zymogens and enzyme activation proceeds by cleavage after a basic amino acid residue in a conserved activation motif N-terminal the catalytic domain. However, the catalytic domain remains membrane-associated after activation because of a disulfide bond that links the prodomain and catalytic domain [6]. Although a few of the TTSPs are expressed across several tissue and cell types, in general these enzymes demonstrate relatively restricted expression patterns, indicating tissue specific regulation or functions [1,4,5,7–11].

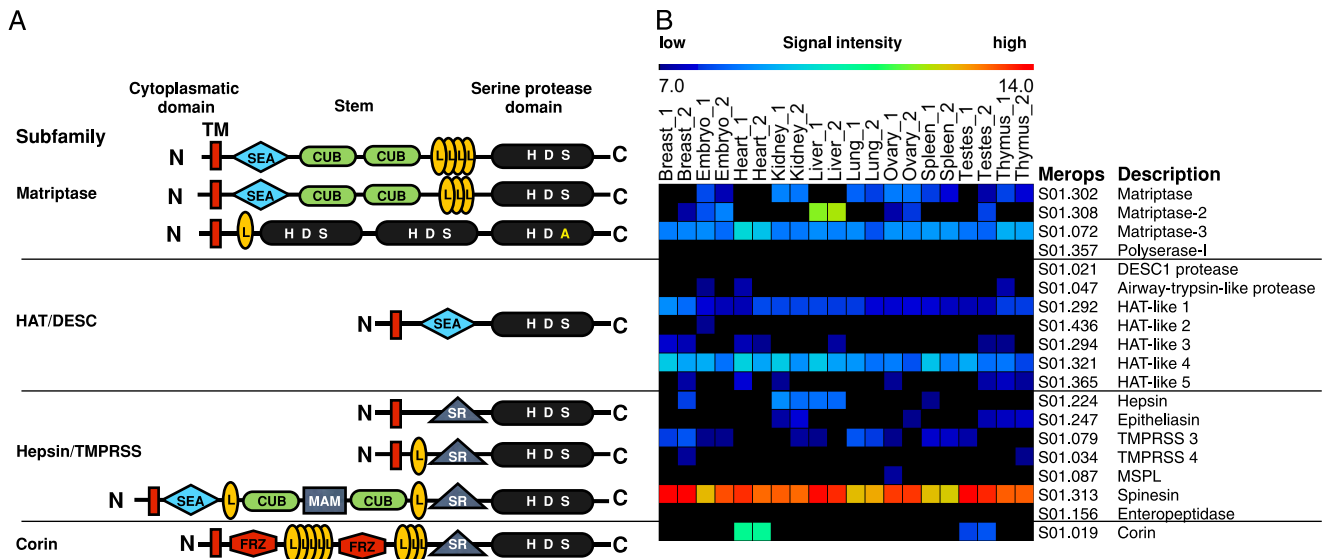


Figure 1. A. Schematic representation of the protein structures and arrangements of the four TTSP subfamilies: HAT/DESC, hepsin/TMPRSS, matriptase and corin. All TTSPs contain a N-terminal transmembrane signal anchor domain (TM) and a C-terminal serine protease domain (H D S). In case of polyserase-1 one of the 3 catalytic domains is inactive (H D A). The stem region of the TTSPs contains 1–6 different domains: the sea urchin sperm protein/enteropeptidase/agrin domain (SEA), group A scavenger receptor domain (SR), low-density lipoprotein receptor class A domain (L), Cls/Clr, urchin embryonic growth factor, bone morphogenetic protein-1 domain (CUB) and Corin contains two frizzled (FRZ) domain. **B. Murine CLIP-CHIP RNA expression profile and its distribution of 19 members of the TTSP family in 10 murine tissues in duplicates according to their average signal intensity.**
doi:10.1371/journal.pone.0105984.g001

Based on phylogenetic analysis of the serine protease domains and the domain structure of the extracellular stem region, the TTSPs have been divided into four subfamilies (Figure 1A). The largest is the HAT/DESC subfamily, with currently 7 proteases: human/murine airway trypsin-like (HAT/MAT), differentially expressed in squamous cells carcinoma (DESC)1, and HAT-like 1–5 protease. HAT-like 2 and 3 are only expressed in rodents [12]. HAT is predominantly expressed in the trachea [13,14] whereas DESC1 is restricted to the epithelia of the skin and oral cavity [15,16]. TTSPs of the hepsin/TMPRSS/enteropeptidase subfamily include hepsin, mosaic serine protease large form (MSPL), type II transmembrane serine protease (TMPRSS) 2, 3, 4, 5 and enteropeptidase. These TTSPs are expressed predominantly in fetal liver and kidney [1], prostate [17] and on the brush-border of the duodenum, respectively [18]. The matriptase subfamily contains three highly homologous proteases; matriptase, matriptase-2, and matriptase-3, as well as a protein with an atypical mosaic structure, polyserase-1. Matriptase was originally identified from a human breast cancer cell line and shows the most ubiquitous pattern of expression of the TTSP family [9–11]. Matriptase-2 plays an important role in iron homeostasis [19–21] and is expressed primarily in the liver, where it suppresses hepcidin expression [22]. Matriptase-3 is expressed in brain, ovary, testis and salivary gland [23], whereas polyserase-1 is mainly restricted to skeletal muscle, liver, placenta and heart [24]. The fourth subfamily consists of a single TTSP, corin, which is mainly expressed in cardiomyocytes [25] and is involved in the activation of proatrial natriuretic peptide, a cardiac hormone that regulates blood pressure and cardiac function by promoting natriuresis, diuresis, and vasodilatation [26].

All TTSPs contain a highly conserved catalytic serine protease domain (Figure 1A) of the trypsin (S1) fold subfamily. The large variability of the central modular region together with the diverse expression patterns of TTSP family members suggest that these

enzymes may play diverse roles, but few of these functions have been identified. Therefore, substrate discovery is essential to establish *in vivo* functions [27]. As a first step, quenched fluorescence peptide assays combined with positional scanning synthetic combinatorial libraries and phage display approaches have identified a limited number of candidate peptide substrates with their cleavage sites [15,23,28–31]. However, the protease database MEROPS (<http://merops.sanger.ac.uk/>) shows that the number of known physiological cleavage sites for the TTSPs is only 52 (+92 synthetic peptide cleavages), and the peptide substrates used only extend from P4 to P4' [32]. Further, most of these studies do not provide any information of prime side specificity corresponding to the sequence of the carboxyl cleavage product.

By knowing the cleavage site and preferred amino acids in recognition sequences, more specific and selective inhibitors can be developed. To overcome these challenges, we used a proteomics approach to determine the cleavage specificity of six different TTSPs belonging to three subfamilies. PICS (Proteome-wide Identification of Cleavage site Specificity) is a high throughput mass spectrometric approach that allows for the simultaneous identification of the amino acid preference at the cleavage site on both the prime and non-prime sides [33,34]. This technique using proteome-wide peptide libraries has the possibility to precisely identify several hundred cleaved peptides by tandem mass spectrometry and searchable databases. Since 2008, thousands of new cleavage sites have been both discovered or confirmed using PICS [35–39]. In the present study, the amino acid preferences of six human TTSPs, from P6 to P6', were established from 2,405 cleavage sites and a rational interpretation of family-wide peptide substrate specificity was obtained by linking their specificities to 3D models of their catalytic domains.

Materials and Methods

Murine CLIP-CHIP microarray

The expression profile of a panel of 10 commercial murine RNAs (Life Technologies) was analyzed with the murine CLIP-CHIP microarray. mRNA sample preparation for the hybridization to the microarray chip and its analysis were conducted as previously described [40,41].

Proteases expression and purification

pRSET5a containing the cDNA of human matriptase (UniProt ID Q9Y5Y6) was a generous gift of Dr. U. Bodendorf, Novartis, Switzerland. Human matriptase-2 (Q8IU80) cDNA was a generous gift from Dr. C. Lopez-Otin (Universidad de Oviedo, Oviedo, Spain). Human hepsin (P05981) cDNA was cloned from a human liver cDNA library from Life Technologies. Human DESC1 (Q9UL52) cDNA was a generous gift from Dr. D. E. Schuller (Ohio State University, OH, USA). Human HAT (O60235) cDNA was purchased from GeneCopoeia (Rockville, MD, USA).

Proteases were cloned, expressed and purified as described [29]. In brief, cDNAs corresponding to amino acids 78–811 of matriptase-2, 45–417 of hepsin, 44–423 of DESC1 and 45–418 of HAT were amplified by PCR and ligated into the pMT/BiP/V5-His vector (Life Technologies, Burlington, Canada). These constructs each contained a C-terminal V5-His tag for affinity purification using immobilized metal-chelate affinity chromatography. *Drosophila* Schneider 2 (S2) cells (Life Technologies) were grown in six-well plates to a density of $2\text{--}4 \times 10^6$ cells/mL and transfected with 19 μg cDNAs in pMT/BiP/V5-His and 1 μg pCoBlast selection vector (Life Technologies) using calcium phosphate transfection kits (Life Technologies). The calcium phosphate solution was removed 16 h post-transfection and fresh medium was added. Cells were grown for an additional 2 days. Blasticidin (20 $\mu\text{g}/\text{mL}$) was added to the medium, and the cells were incubated for a 14-day selection period. Cells were then inoculated into 1 L of selection medium and grown to a density of 3×10^6 cells/mL. Recombinant TTSP expression was induced by adding 500 μM CuSO_4 . Cells were removed by centrifugation at $6,000 \times g$. Supernatants were loaded on HisTrap FF columns (GE, Mississauga, Canada) and bound TTSPs were eluted using 500 mM imidazole. Fractions were analyzed by immunoblotting and fractions containing TTSPs were pooled and for imidazole removal dialyzed for 16 h at 4°C against 50 mM Tris-HCl (pH 8.5), 10% glycerol and 250 mM NaCl. Dialyzed pooled fractions were then loaded on HiTrap Benzamide FF affinity columns (GE, Mississauga, Canada) and eluted with 10 mM HCl (pH 2.0), 0.5 M NaCl and then buffered to 50 mM HEPES (pH 9.0) with 20% glycerol. The four purified TTSPs were active-site titrated with the burst titrant 4-methylumbelliferyl-*p*-guanidino benzoate.

Spodoptera frugiperda Sf9 cells expressing mouse matriptase-3 (Q7RTY8) cDNA encoding the serine protease domain (amino acids 581–829) were a kind gift from Dr. T. Bugge (NIH, Bethesda, MD, USA). Matriptase-3 protease was purified as previously described [1,4,23] with slight modifications. Conditioned medium was collected from Blasticidin-resistant Sf9 cells, and sodium phosphate buffer (pH 7.5) with NaCl were added to a final concentration of 20 mM and 500 mM respectively. The solution was filtered through a 0.22 μm filter and applied to a HisTrap column using an AKTA explorer system. Chelated proteins were eluted with a linear gradient of 20–400 mM imidazole, and TTSP-positive fractions were identified by immunoblotting using the anti-V5 antibody (Life Technologies).

The positive fractions were dialysed overnight at 4°C against 200 volumes of 50 mM Tris-HCl (pH 8.0), 1 mM CaCl_2 and 0.1% Tween 20 using dialysis membranes of 10 kDa MWCO.

PICS analysis

To identify TTSP cleavage sites, we prepared two overlapping proteome-derived peptide libraries. Endoproteases GluC or chymotrypsin were used to digest HEK293 cell (Life Technologies) proteomes. Peptides were reduced, cysteines alkylated and primary amines protected by dimethylation [33]. The two peptide libraries were then separately incubated with the individual TTSPs after which the N-termini of prime-side cleavage products were tagged with biotin, affinity enriched and identified by liquid chromatography tandem mass spectrometry (LC-MS/MS), 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 UBC Centre for Blood Research Mass Spectrometry Suite. In brief, 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 data 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. MS and MS/MS data were acquired in centroid mode. QSTAR data were converted to the mzXML format with mzwiff (version 4.3.1). These sequences represent the prime side products of TTSP cleavage. The corresponding non-prime side sequences were derived from human proteome databases using bioinformatics searches. To increase the confidence of our mass spectrometry queries both search engines Mascot (version 2.2.04) [42] and X! Tandem (version 2007.07.01.2) [43] were used and the results were combined. Peptides were analyzed with PeptideProphet [44] and iProphet [45], both implemented within the Trans Proteomic Pipeline (Institute for Systems Biology, Seattle, WA, USA) [46]. Peptides were analyzed at a global false discovery rate of 5% at peptide level by using a decoy search strategy. This allowed for the selection of peptides having a minimum threshold probability corresponding to this false discovery rate percentage [46]. Search parameters included a mass tolerance of 0.4 Da for parental ions and 0.2 Da for fragment ions, allowing up to three missed cleavages, and for the following fixed peptide modifications: carbamidomethylation of cysteine residues (+57.02 Da), dimethylation of lysine amines (+28.03 Da), and variable methionine oxidation (+15.99 Da), variable thioacylation (+88.00 Da) and dimethylation (+28.03 Da) of peptide N-termini. N-terminally thioacylated peptides identified by MS/MS represent prime-side cleavage products of the assayed proteases. The complete cleavage sites were reconstructed by bioinformatic determination of the non-prime side sequences from the human database IPI v3.42 (http://ebi.edu.au/ftp/databases/IPI/last_release/msipi/old/HUMAN/) using WebPICS [39], available at <http://clipserve.clip.ubc.ca/pics/>. Subsite positions with ambiguous information coming from different protein isoforms were omitted and replaced by X, the one-letter notation for an unknown amino acid. Reconstructed cleavage sites were aligned and summarized as heat maps and iceLogos (University Ghent, Belgium) [7].

Peptide cleavage site confirmation

In order to confirm the PICS results for the cleavage specificity, mass spectrometry peptide cleavage assays were performed.

Figure 2. A. Multiple sequence alignment of the catalytic domains of six TTSPs was performed with Clustal Omega [59] and displayed using ESPrnt 2.2 (<http://esprnt.ibcp.fr/ESPrnt/cgi-bin/ESPrnt.cgi> [63]). Green stars indicate the residues of the catalytic triad; conserved residues are shaded in red. Blue boxes indicate amino acid with similar physico-chemical properties. Secondary structure elements are represented with flat arrows for β sheets and helices for α - and 3_{10} - (η) helices. Three conserved disulfide bridges are indicated in orange numerals below the alignment. Sequence numbering is according to the respective UniProt entries. **B. Topology diagram of a TTSP catalytic domain.** Residues of the catalytic triad are indicated as green stars, disulfide bonds (DSB) are shown in yellow. Loops (L-) are labeled corresponding to the thrombin nomenclature suggested by Bode et al. [64]. **C. Superposition of the catalytic domains of matriptase, matriptase-2, matriptase-3, hepsin, DESC1 and HAT,** including both known and modeled structures, in traditional serine protease standard orientation [65]. Catalytic triad and conserved disulfide bridges are shown in ball and stick representation. Important loops and exosites are indicated and labeled. doi:10.1371/journal.pone.0105984.g002

Scanning through the peptides identified, we selected one peptide sequence that was found for each of the TTSPs analyzed, namely AVIGRKFGDP. The peptide (theoretical mass of 1059.22 Da) was synthesized (GenScript) (>90% purity). For the dimethylated peptide, 30 mM formaldehyde, 30 mM of cyanoborohydride (ALD coupling solution, Sterogene, Carlsbad, CA) and 10 mM HEPES pH 7.4 were added to the peptide and incubated overnight (18 h) at room temperature to generate (dm)AVIGR(dm)KFGDP. To assay for cleavage, 200 μ g of AVIGR(dm)KFGDP or (dm)AVIGR(dm)KFGDP were added to 2 μ g of matriptase, matriptase-2, HAT and hepsin in 100 mM HEPES, pH 7.4, 150 mM NaCl and incubated at 37°C for 18 h. 20 μ L of the reaction mixtures were used and buffer exchanged using C18 OMIX 10 μ L tips (Agilent, Mississauga, Canada) using the manufacturer's instructions. Samples (0.8 μ L) were mixed with 0.5 μ L of 10 mg/mL of CHCA matrix solution (Sigma, Oakville, Canada) and spotted on a stainless steel MALDI plate. Spectra were acquired in reflectron positive mode and 2,175 random shots were performed and analyzed on a 4700 MALDI-TOF/TOF mass spectrometer (Applied Biosystems) with a mass range of 300–1,200 Da. Data analysis was conducted using protein data software provided by Applied Biosystems.

Structural analysis

For matriptase, hepsin, and DESC1, the crystal structures 3p8g [47], 1z8g [48] and 2oq5 [49] were used, respectively. For matriptase-2, -3 and HAT, structural models were generated using the interactive mode of the modeling server SWISS-MODEL [50,51], and crystal structures of closely related enzymes identified by the FFAS03 server [52,53]. We used DESC1 (PDB code 2oq5) [49] for matriptase-2 and HAT, and matriptase (PDB code 3p8g) [47] for matriptase-3. Sequence numbering was corrected and secondary structures assigned to all structures and models using the UCSF Chimera package [44]. The topology representation was created using TopDraw [54]. Peptide docking with the structural model of matriptase-2 was performed using the Rosetta FlexPepDock web server [55,56], allowing full flexibility to the peptide (AEAALR(P1).(P1')KLLEVA), and full side-chain flexibility to the protease. Initial models were based on available complex structures of matriptase with aprotinin (PDB code 1eaw) [47] and trypsin inhibitor 1 (PDB entry 3p8f) [57], and our PICS P6 to P6' sequence results. Molecular graphic figures were made using the molecular visualization system PyMOL (The PyMOL Molecular Graphics System, Version 1.5.03, Schrödinger, LLC) [58].

Results

Tissue distribution of TTSPs

The tissue distribution of TTSP family members was determined by transcript analysis of ten murine tissues using the dedicated murine CLIP-CHIP [1–3], with probes for all murine proteases, inactive homologues, and inhibitors (Figure 1B). The expression profile reveals that most of the TTSPs are very low or

non expressed (average signal intensity below 7.5) in the analyzed tissues with the exception of the high expression of matriptase-2 in liver and corin in heart, both of which have been previously described as predominantly expressed in these tissues [25,28]. Hepsin shows significant expression in liver and kidney, TMPRSS3 in breast and lung, matriptase in kidney and lung. Matriptase-3, Hat-like 1, Hat-like 4 and spinesin show expression in all 10 tissues. However, polyserase-1, DESC1 and enteropeptidase were not detected in these ten tested tissues. This suggests potential tissue specificity for some of the TTSPs whereas other TTSPs might be involved in more general processes due to their ubiquitous expression.

Sequence alignment and structural analysis of the catalytic domain of TTSPs

To better understand the cleavage specificity of the TTSP family, we first compared the sequences of their catalytic domains using Clustal Omega [59] and ESPrnt 3.0 [57] (Figure 2A). The catalytic domain of matriptase showed the highest sequence identity with matriptase-2 (45% identity) and matriptase-3 (31% identity). HAT showed the highest sequence identity with DESC1 (50% identity). From the three subfamilies, the conserved regions within the catalytic domain are clearly visible (indicated in red). The catalytic triad, indicated as green stars, consists of histidine 656, aspartate 711 and serine 805, following the numbering of matriptase (top row). Interestingly, strictly conserved regions are often observed in stretches that adopt a β -sheet structure (Figure 2A). A topology diagram of a typical TTSP catalytic domain is shown in Figure 2B. Conservative substitutions, for example such as leucine by valine and serine by threonine were identified in the β 4-strand close to the active site. This suggests that a specific chemical environment is necessary for the catalytic domain to perform catalysis.

Such conservation of key regions and three disulfide bridges led us to hypothesize that the catalytic domain of all the TTSPs had structural similarities. Crystal structures of matriptase [47], DESC1 [49] and hepsin [48,60] have been solved previously. Therefore, matriptase-2, matriptase-3 and HAT were modeled and analyzed by superimposition of all the catalytic domains (Figure 2C), and typical structural elements of serine proteases of the trypsin fold are depicted. The first striking aspect observed was that the structures of all the TTSP catalytic domains are similar and highly superimposable with few major differences. One of these differences occurs in matriptase-3 where an extended loop in its catalytic domain following the β 7 strand that is diverging from the other family members. This loop may interact with substrates or effectors that could induce a conformational change in the matriptase-3 catalytic domain and modulate substrate binding and presentation to the active site for catalysis. Another example is the extended 60-loop between the two 3_{10} -helices (η 1 and η 2) of matriptase, which potentially provides additional substrate specificity and possibly flip over the active site upon substrate binding, blocking further access until cleavage occurs. However, these differences in structure do not affect the active site geometry of the

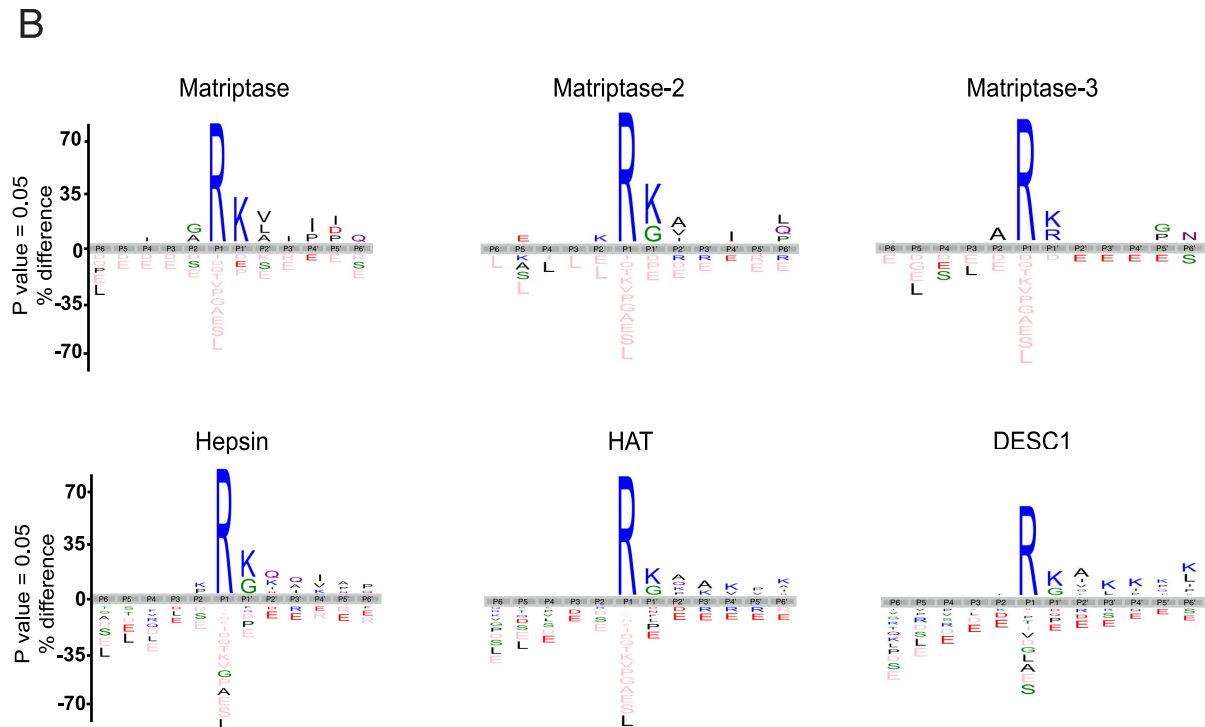
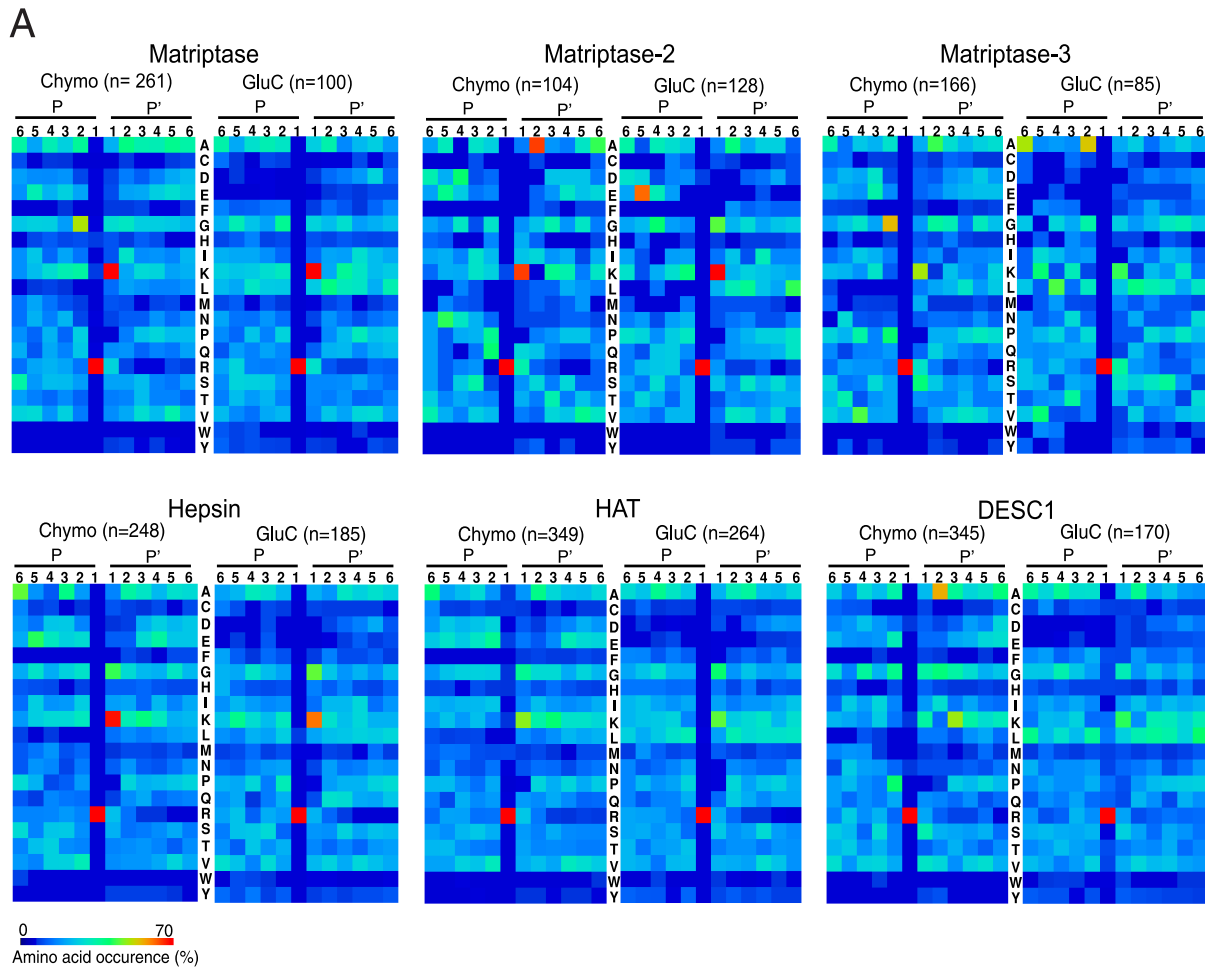


Figure 3. A. Matriptase, matriptase-2, matriptase-3, hepsin, DESC1, an HAT cleavage site specificities obtained by PICS using two different peptide libraries: GluC and chymotryptic. For each class of peptide library, the amino acid occurrences in P6–P6' are shown as heat maps. Above each heat map, the number of identified cleaved peptides is specified. **B.** Taking into account the relative natural abundances of amino acids, iceLogo representations of the different cleavage site specificities obtained from the GluC library are shown. Peptide sequences were aligned on the scissile bond between P1 and P1'. Statistically significant amino acid residue occurrences present ($P < 0.01$) were plotted and amino acid heights are indicative for the degree of conservation at the indicated position. Residues that were completely absent in the identified terminal peptides are shown below in pink. doi:10.1371/journal.pone.0105984.g003

catalytic triad in these TTSPs, and their specificity for peptidic substrates, but may explain differences in native substrate cleavage or in comparing peptide cleavage specificities with native protein cleavage specificities.

PICS analysis represented by heat maps and iceLogo

Using PICS, 2,405 cleavage sites were identified using both chymotryptic and GluC-generated proteome-derived peptide libraries, greatly expanding upon the known 144 cleavage sites annotated in MEROPS (Tables S1–S12 in File S1). These peptide

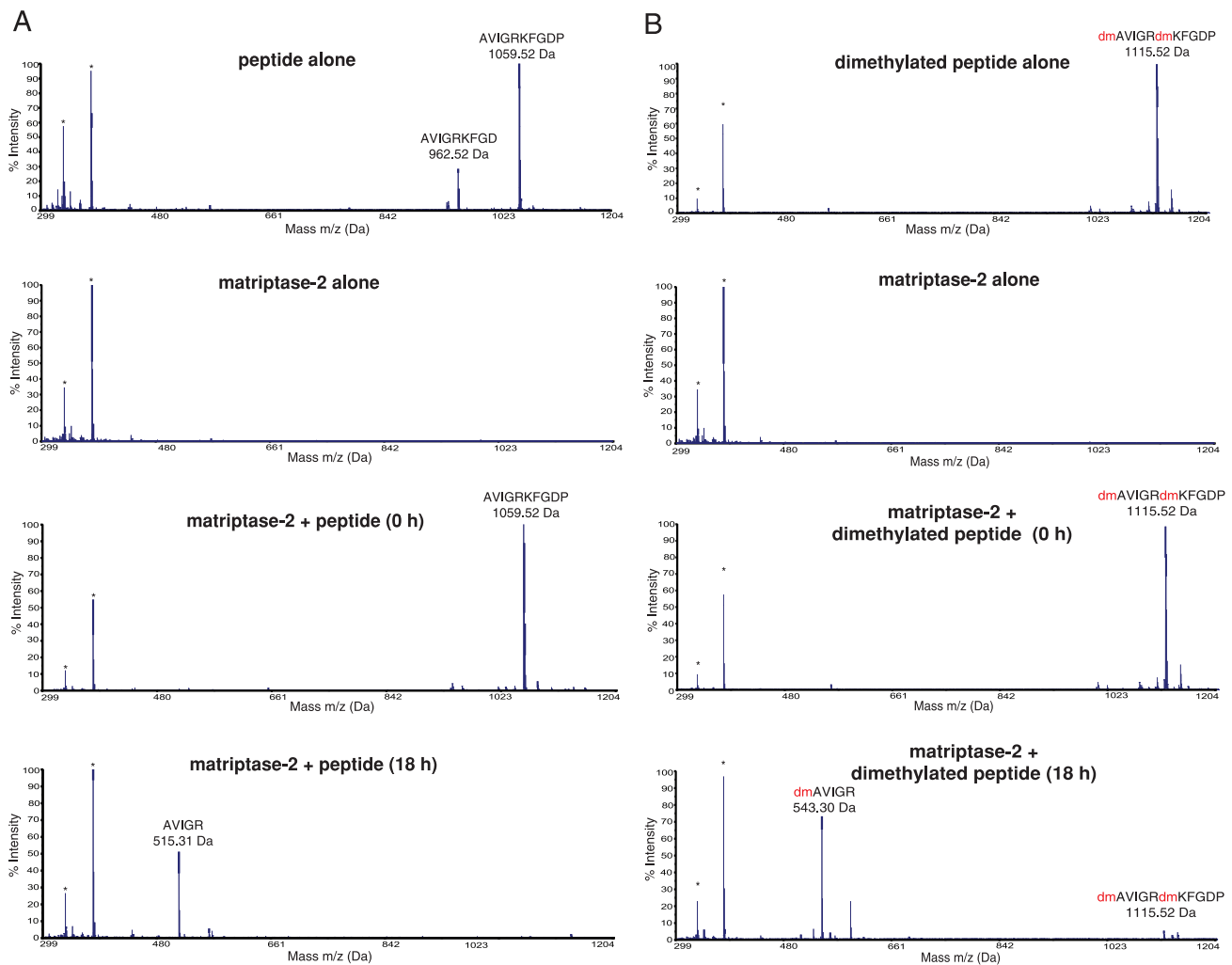


Figure 4. A. Upper panel. MALDI-TOF spectrum of the synthetic peptide AVIGRKFQDP. The sample contained a minor synthesis contaminant of AVIGRKFQD. The experimental determined $[M+H]^+$ is 1059.52 Da (predicted m/z is 1059.22 Da). The sequence is presented by the spectra. The two asterisk peaks represent the MALDI matrix peaks which are found in all the spectra in this m/z range. **Second panel.** MALDI-TOF spectrum of matriptase-2. **Third panel.** MALDI-TOF spectrum of the synthetic peptide added to matriptase-2 at 0 h. **Lower panel.** MALDI-TOF spectrum of the assay reaction products generated after incubation of the synthetic peptide with matriptase-2 for 18 h. The spectral peak at 1059.52 m/z disappeared and a new peak at 515.31 Da appeared corresponding to the cleavage product $[AVIGR+H]^+$ (predicted m/z is 514.96 Da). **B. Upper panel.** MALDI-TOF spectrum of the dimethylated (dm) synthetic peptide (dm)AVIGR(dm)KFQDP. The experimental determined $[M+H]^+$ is 1115.52 Da. The sequence is presented on the spectral peak. The two asterisk peaks represent the MALDI matrix peaks and they can be found in all the spectra. **Second panel.** MALDI-TOF spectrum of matriptase-2. **Third panel.** MALDI-TOF spectrum of the reaction products after incubation of the dimethylated synthetic peptide with matriptase-2 added at 0 h. **Lower panel.** MALDI-TOF spectrum of the reaction products after incubation of the dimethylated synthetic peptide with matriptase-2 for 18 h. The peak at 1115.52 Da disappeared and a new peak at 543.30 Da appeared, corresponding to $[dAVIGR+H]^+$. Red (dm) is dimethylation. doi:10.1371/journal.pone.0105984.g004

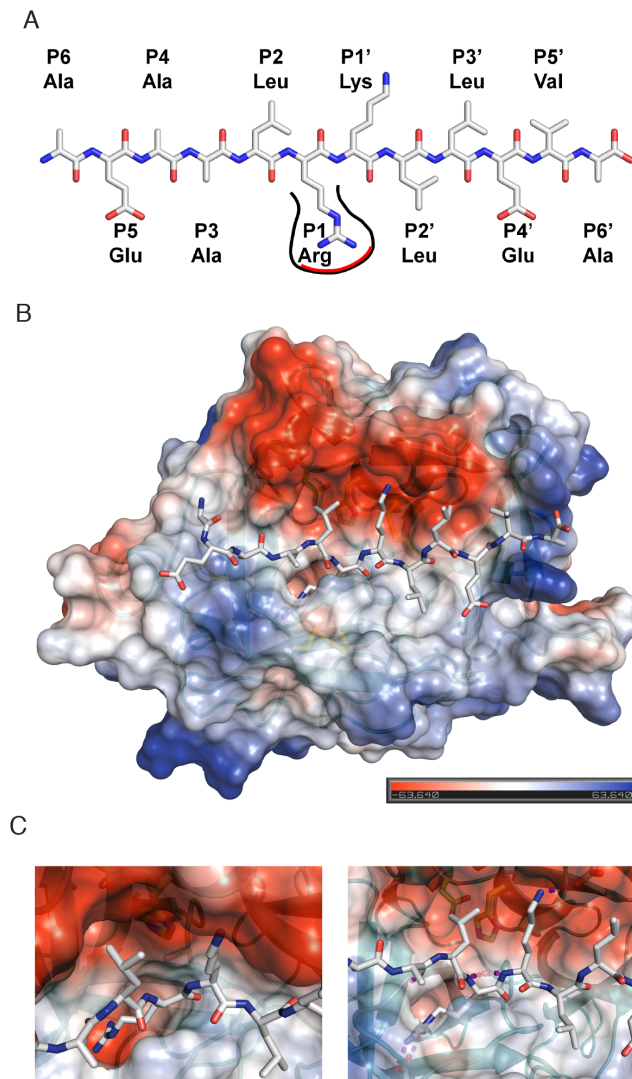


Figure 5. A Schematic representation of the dodecapeptide (AEAALR↓KLLEVA) used for active site docking. The deep TTSP S1 subsite, accommodating the P1-Arg, is shown. **B.** Structural model of matriptase-2 with the peptide substrate (AEAALRKLEVA) identified by PICS docked into the active site. Matriptase-2 was colored according to its surface charge distribution, and is shown in standard orientation with the modeled peptide occupying subsites S6 to S6' (from left to right). **C.** Close-up views of the active site illustrating the deep and negatively charged S1 pocket (left) and key stabilizing interaction between the protease and the modeled peptide (right). doi:10.1371/journal.pone.0105984.g005

libraries were selected because the TTSP family belongs to the trypsin-like family and so have a lysine or arginine specificity in P1. Thus, peptide libraries generated by trypsin for use with the PICS technique will contain peptides displaying an arginine or lysine at the C-terminus and hence will not contain internal basic residues. In such cases alternate peptide digestive proteases are easily used to generate appropriate peptide libraries. Here we used chymotryptic and GluC peptide libraries in which the C-terminal residues were aromatic/hydrophobic or negatively charged, respectively. These two libraries are overlapping and both libraries display internal lysine and arginine residues and complement each other by displaying acidic and hydrophobic residues within the peptide sequences. Using two or more libraries also allows for

experimental validation of results obtained from one library with the other.

From the heat maps generated for the six human TTSPs, a common feature was apparent—arginine is preferred in the P1 position, ~17-fold more than the natural occurrence of arginine in proteins, and lysine is preferred in P1' (Figure 3A). Note in the PICS workflow lysine is dimethylated. For several TTSPs, the P1' position has glycine as an alternative amino acid preference, which is strongly favored in the case of hepsin. Similar specificities also occur within subfamilies. For example, for matriptase and matriptase-3, glycine or alanine is important at P2, whereas matriptase-2 prefers glutamate in P5 and alanine in P2'. Another example is across the HAT/DESC1 subfamily, where alanine and lysine are preferred in the P2' and P3' positions, respectively, for DESC1, but these are less preferred for HAT. HAT requires an aspartate in P2, a feature that was shown by Wysocka et al [31] using positional scanning libraries.

The iceLogo representation (Figure 3B) also revealed the least preferred amino acid preferences at each position. For example, since arginine is the only preferred amino acid at P1, this means that all other residues are thereby under represented. Similar negative preferences result from the lysine or glycine in P1'. Least preferred residues also reinforces the differences among members of the same subfamily. For example, matriptase and matriptase-3 prefer a glycine in P2, but matriptase-2 does not. Conversely, matriptase-2 favors a glutamate in P5, but this is not the case for matriptase and matriptase-3. This also points out the utility of PICS as it is easy to profile 6 amino acid residues in both the non-prime and prime sides, which other techniques generally do not do cover. Another interesting finding was the preference for alanine in the P2' position for most of the TTSPs, especially important for matriptase-2 and DESC1, suggesting a specific feature for these two subfamilies.

To confirm the identified cleavage specificities of the six studied TTSPs we used synthetic peptides. We first examined the PICS mass spectrometry results in more detail and found that the AVIGRKFQDP peptide ($[M+H]^+ = 1059.52$) was cleaved by all TTSPs. This peptide was synthesized and incubated with matriptase-2 (Figure 4), matriptase, HAT and hepsin (Figures S1–S3), and the cleavage site was identified by MALDI-TOF MS. The N-terminal nonprime cleavage product AVIGR (515.30 Da) was identified after cleavage between Arg and Lys confirming the unusual P1' specificity.

In PICS, all primary amines within the peptide library are dimethylated resulting in a dimethylated lysine side chain with altered charge, raising the question about the involvement of lysine in P1'. We found that the dimethylated (dm) version of the peptide (dm)AVIGR(dm)KFGDP was also cleaved by matriptase-2 (Figure 4B) and all other tested TTSPs with similar kinetics, validating the identified PICS cleavage sites for these TTSPs. Overall, our data reveal that the prime side is also important for the cleavage specificity of the TTSPs where assayed using unstructured peptide libraries. This makes the TTSPs unique within the tryptic-enzyme class that are dominated by P1 interactions.

Interpreting cleavage specificity by structural modeling of matriptase-2

Based on the known crystal structures, a matriptase-2 3D homology model was generated (Figure 2C and Figure 5), and a PICS determined model peptide, AEAALR↓KLLEVA (Figure 5A) was incorporated in the active site by peptide docking. An electrostatic surface charge representation of matriptase-2 including the docked peptide is shown in Figure 5B, illustrating

Table 1. Known protein substrates obtained from MEROPS [32] and UniProt [66] showing the P4-P4' cleavage specificity in natural substrates of matriptase, matriptase-2, matriptase-3, HAT, hepsin and corin where cleaved as native proteins versus denatured peptides in PICS.

	Cleavage site										References	
	P4	P3	P2	P1	P1'	P2'	P3'	P4'				
Matriptase substrates												
CDCP1	Lys	Gln	Ser	Arg	Lys	Phe	Val	Pro				[67]
fibronectin	Val	Thr	Gly	Arg	Gly	Asp	Ser	Pro				[68]
flaggrin	Arg	Lys	Arg	Arg	Gly	Ser	Arg	Val				[11]
flaggrin	Thr	Gly	His	Gly	Gln	Ala	Ser	Ser				[69]
hepatocyte growth factor	Lys	Gln	Leu	Arg	Val	Val	Asn	Gly				[11]
insulin-like growth factor-binding protein 1	Lys	Ala	Leu	His	Val	Thr	Asn	Ile				[11]
insulin-like growth factor-binding protein 7	Arg	Lys	Gly	Lys	Ala	Gly	Ala	Ala				[70]
laminin, alpha 1-like	Arg	Val	Val	Arg	Lys	Gly	Val	Ser				[68]
matriptase	Arg	Gln	Ala	Arg	Val	Val	Gly	Gly				[11]
prostasin	Ile	Gln	Pro	Arg	Ile	Thr	Gly	Gly				[71]
proteinase-activated receptor 2	Ser	Lys	Gly	Arg	Ser	Leu	Ile	Gly				[11]
stromelysin-1	Arg	Lys	Pro	Arg	Cys	Gly	Val	Pro				[72]
urokinase-type plasminogen activator	Pro	Arg	Phe	Lys	Ile	Ile	Gly	Gly				[11]
Matriptase-2 substrates												
matriptase-2 precursor	Pro	Ser	Ser	Arg	Ile	Val	Gly	Gly				[73]
matriptase-2	Pro	Gly	Val	Arg	Val	His	Tyr	Gly				[73]
hemajuvelin	Ile	Ile	Ile	Arg	Gln	Thr	Ala	Gly				[74]
Matriptase-3 substrates												
alpha-2-antiplasmin	Ala	Met	Ser	Arg	Met	Ser	Leu	Ser				[23]
antithrombin-III	Ile	Ala	Gly	Arg	Ser	Leu	Asn	Pro				[23]
heparin cofactor 2	Phe	Met	Pro	Leu	Ser	Thr	Gln	Val				[23]
plasma serine protease inhibitor	Phe	Thr	Phe	Arg	Ser	Ala	Arg	Leu				[23]
plasminogen activator inhibitor-1	Val	Ser	Ala	Arg	Met	Ala	Pro	Glu				[23]
Hepsin substrates												
coagulation factor IX	Asp	Phe	Thr	Arg	Val	Val	Gly	Gly				[60]
coagulation factor VII	Pro	Gln	Gly	Arg	Ile	Val	Gly	Gly				[75]
coagulation factor XII	Ser	Met	Thr	Arg	Val	Val	Gly	Gly				[60]
hepatocyte growth factor	Lys	Gln	Leu	Arg	Val	Val	Asn	Gly				[60]
hepsin precursor	Pro	Val	Asp	Arg	Ile	Val	Gly	Gly				[60]
laminin subunit beta-3	Ser	Gln	Leu	Arg	Leu	Gln	Gly	Ser				[61]
urokinase-type plasminogen activator	Pro	Arg	Phe	Lys	Ile	Ile	Gly	Gly				[76]
Corin substrates												
atrial natriuretic peptide-converting enzyme	Ser	Val	Val	Arg	Asn	Met	Glu	Met				[77]

Table 1. Cont.

	Cleavage site										References
	P4	P3	P2	P1	P1'	P2'	P3'	P4'			
Matriptase substrates											
atrial natriuretic peptide-converting enzyme	Gly	Asp	Gln	Arg	Cys	Leu	Tyr	Asn			[77]
pro-atrial natriuretic peptide	Thr	Ala	Pro	Arg	Ser	Leu	Arg	Arg			[18]
HAT substrates											
urokinase plasminogen activator surface receptor	Asn	Ser	Gly	Arg	Ala	Val	Thr	Tyr			[78]
urokinase plasminogen activator surface receptor	Thr	Tyr	Ser	Arg	Ser	Arg	Tyr	Leu			[78]

doi:10.1371/journal.pone.0105984.t001

the negative charge distribution of the upper half of the active site including subsite S2 and S1', and the slightly positive subsites S5 and S4', explaining their opposed preferences for charged residues. The deep S1 pocket with Asp756 forming the negative base and main interactor of the P1-Arg is depicted in Figure 5C. Both the groove and the electrostatic environment suggest that a positively charged amino acid composed of a long side chain can be readily accommodated and stabilized in S1. Indeed, this would explain why an arginine is highly favored at the P1 position. Noteworthy, both backbone nitrogens of the P1 arginine and the P1' lysine are held in position by the active site Ser762. Specificity for lysine in P1' is mainly provided by the negatively charged side-chain of Glu621 at the top of the S1 subsite (Figure 5C). We suggest that these features are the key players for the preferred cleavage between arginine in P1 and lysine in P1'.

Glycine 784 of the protease backbone connects to the P2 position of the peptide through hydrogen bonding and stabilizes this atom at the non-prime site of the peptide. The slight preference in this position for glycine might be guided by the high flexibility of its α carbon, which may compensate for less-preferred amino acids in other subsites. Subsite P3', which favors small hydrophobic residues for most of the investigated TTSPs, is stabilized by anti-parallel backbone-backbone interaction of the peptide to the protease.

Discussion

By utilizing diverse natural peptides of variable length and amino acid composition, PICS identifies the cleavage specificity of a protease and the exact sequence on both sides of the scissile bond by analysis of hundreds of cleaved peptides in each experiment [33,34]. The mass spectrometric determination of the active site specificity of six important members of the TTSP subfamilies using PICS revealed several important features. First, it is an unusual finding for tryptic proteases for their specificity to be determined by both the P1 and P1' positions. Second it is also unusual for tryptic proteases to have specificity exerted through extended features on the prime side. The specificity for arginine in P1 is in agreement with other recent reports using other techniques such as positional scanning libraries and/or fluorescent substrates [15,23,29,31,60,61]. Furthermore, almost all of the known protein substrates of the assayed TTSPs contain an arginine in P1 position (Table 1). However, we cannot rule out that native substrates with a lysine in P1 might be cleavable since all lysines in PICS are dimethylated to protect against NHS-biotin reactivity, which is used to selectively tag the neo-N terminal α -amine for enrichment after cleavage. Third, the specificity for lysine in P1' is shown for the first time using a peptide library technique in our study and is consistent with the natural substrate specificity of matriptase, which cleaves for example CDCP1 and laminin, alpha 1-like between arginine and lysine (Table 1). Consistent with our findings, Beliveau *et al* [29] demonstrated that negatively charged amino acids are the least preferred at P1' position as shown by having a very low k_{cat}/K_M .

PICS showed that if the nature of the peptide substrate was negatively charged, non-polar, or hydrophobic, then these characteristics affected the amino acid preferences accommodated in the TTSP active sites depending on which library was used. For example, using a chymotryptic library, a preference for aspartate and glutamate was identified at the P3, P4 and P5 positions for hepsin, matriptase, matriptase-3 and DESC1, whereas peptides generated by GluC had a leucine and/or lysine preferred at P3 and P4. Since aspartate and glutamate are displayed internally in the chymotryptic peptides these libraries better reflect the true

specificities for acidic residues than can be derived from GluC libraries, which are internally depleted for negatively charged residues. This reaffirms the need for using at least two libraries in PICS to accurately interpret cleavage specificity preferences for proteases [33,34].

It was notable that the prime side of a peptide is important for TTSP cleavage specificity. Thus alanine or other small hydrophobic residues are required in the P2' position, as well as nonpolar amino acids such as leucine and valine at the other prime positions. This feature can also be observed in natural substrates of matriptase (hepatocyte growth factor), matriptase-3 (plasminogen activator-inhibitor-1) and hepsin (coagulation factor IX and VII) (Table 1).

The application of PICS was shown to be highly successful by the identification of 2,405 cleavage sites for the six TTSP family members studied. These data greatly expand the 144 known cleavages sites for these proteases. The specific amino acid residue features found in our study of three subfamilies of the TTSPs provide new insight that might be used to design specific inhibitors and peptide substrates, such as AEAALR↓KLLEVA that we identified here and validated. It is important to keep in mind that the overall similarities in specificity for the TTSPs will likely render small molecule drug development challenging. The presence of several domains outside the catalytic domain suggests that these exosites could be critical for substrate binding and recognition [62] and undoubtedly influence native substrate binding and cleavage of intact proteins. Indeed many native protein substrates do not have Lys in P1'. Native protein structures often mask even highly preferred cleavage sites rendering these unavailable for protease access and cleavage. However, this can be levered to advantage in designing small molecule inhibitors. Thus future studies of exosite interactions, and possibly designing exosite inhibitors, will be insightful in revealing TTSP specificity for natural substrates *in vivo* and hence *in vivo* roles.

References

- Leytus SP, Loeb KR, Hagen FS, Kurachi K, Davie EW (1988) A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells. *Biochemistry* 27: 1067–1074.
- Rodríguez D, Morrison CJ, Overall CM (2010) Matrix metalloproteinases: What do they do? New substrates and biological roles identified by murine models and proteomics. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1803: 39–54. doi:10.1016/j.bbamcr.2009.09.015.
- Overall CM, Blobel CP (2007) In search of partners: linking extracellular proteases to substrates. *Nat Rev Mol Cell Biol* 8: 245–257. doi:10.1038/nrm2120.
- Netzel-Arnett S, Hooper JD, Szabo R, Madison EL, Quigley JP, et al. (2003) Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. *Cancer Metastasis Rev* 22: 237–258.
- Szabo R, Bugge TH (2008) Type II transmembrane serine proteases in development and disease. *Int J Biochem Cell Biol* 40: 1297–1316. doi:10.1016/j.biocel.2007.11.013.
- Hooper JD, Clements JA, Quigley JP, Antalis TM (2001) Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *J Biol Chem* 276: 857–860. doi:10.1074/jbc.R000020200.
- Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K (2009) Improved visualization of protein consensus sequences by iccLogo. *Nat Methods* 6: 786–787. doi:10.1038/nmeth1109-786.
- Wu F, Yan W, Pan J, Morser J, Wu Q (2002) Processing of pro-atrial natriuretic peptide by corin in cardiac myocytes. *J Biol Chem* 277: 16900–16905. doi:10.1074/jbc.M201503200.
- Bugge TH, List K, Szabo R (2007) Matriptase-dependent cell surface proteolysis in epithelial development and pathogenesis. *Front Biosci* 12: 5060–5070.
- List K, Haudenschild CC, Szabo R, Chen W, Wahl SM, et al. (2002) Matriptase/MT-SP1 is required for postnatal survival, epidermal barrier function, hair follicle development, and thymic homeostasis. *Oncogene* 21: 3765–3779. doi:10.1038/sj.onc.1205502.
- Uhland K (2006) Matriptase and its putative role in cancer. *Cell Mol Life Sci* 63: 2968–2978. doi:10.1007/s00018-006-6298-x.
- Sales KU, Hobson JP, Wagenaar-Miller R, Szabo R, Rasmussen AL, et al. (2011) Expression and Genetic Loss of Function Analysis of the HAT/DESC Cluster Proteases TMPRSS11A and HAT. *PLoS ONE* 6: e23261. doi:10.1371/journal.pone.0023261.
- Yamaoka K, Masuda K, Ogawa H, Takagi K, Umamoto N, et al. (1998) Cloning and characterization of the cDNA for human airway trypsin-like protease. *J Biol Chem* 273: 11895–11901.
- Yasuoka S, Ohnishi T, Kawano S, Tsuchihashi S, Ogawara M, et al. (1997) Purification, characterization, and localization of a novel trypsin-like protease found in the human airway. *Am J Respir Cell Mol Biol* 16: 300–308. doi:10.1165/ajrcmb.16.3.9070615.
- Hobson JP, Netzel-Arnett S, Szabo R, Réhault SM, Church FC, et al. (2004) Mouse DESC1 is located within a cluster of seven DESC1-like genes and encodes a type II transmembrane serine protease that forms serpin inhibitory complexes. *J Biol Chem* 279: 46981–46994. doi:10.1074/jbc.M403299200.
- Lang JC, Schuller DE (2001) Differential expression of a novel serine protease homologue in squamous cell carcinoma of the head and neck. *Br J Cancer* 84: 237–243. doi:10.1054/bjoc.2000.1586.
- Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, et al. (2001) Delineation of prognostic biomarkers in prostate cancer. *Nature* 412: 822–826. doi:10.1038/35090585.
- 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.
- Folgueras AR, de Lara FM, Pendás AM, Garabaya C, Rodríguez F, et al. (2008) Membrane-bound serine protease matriptase-2 (Tmprss6) is an essential regulator of iron homeostasis. *Blood* 112: 2539–2545. doi:10.1182/blood-2008-04-149773.
- Silvestri L, Pagani A, Nai A, De Domenico I, Kaplan J, et al. (2008) The serine protease matriptase-2 (TMPRSS6) inhibits hepcidin activation by cleaving membrane hemouvelin. *Cell Metab* 8: 502–511. doi:10.1016/j.cmet.2008.09.012.
- Du X, She E, Gelbart T, Truksa J, Lee P, et al. (2008) The serine protease TMPRSS6 is required to sense iron deficiency. *Science* 320: 1088–1092. doi:10.1126/science.1157121.
- Ramsay AJ, Hooper JD, Folgueras AR, Velasco G, López-Otín C (2009) Matriptase-2 (TMPRSS6): a proteolytic regulator of iron homeostasis. *Haematologica* 94: 840–849. doi:10.3324/haematol.2008.001867.

Supporting Information

Figure S1
(EPS)

Figure S2
(EPS)

Figure S3
(EPS)

File S1 Tables S1-S12.
(XLS)

Acknowledgments

O.B. was supported by the Swiss National Science Foundation (SNF) and a post-doctoral fellowship by the Canadian Institutes of Health Research (CIHR). A.D. and U.E. were supported by post-doctoral fellowships from CIHR and the Michael Smith Foundation for Health Research (MSFHR), respectively. A Canada Research Chair in Metalloproteinase Proteomics and Systems Biology supports C.M.O. This work was supported by grants to C.M.O. and R.L. from the CIHR and the Canadian Cancer Society. We thank Wei Chen, Suzanne Perry and Jason Rogalski for LC-MS/MS measurements. The University of British Columbia, Centre for Blood Research Mass Spectrometry Suite is supported by the Canada Foundation for Innovation (CFI) and MSFHR. We also thank Ursula Bodendorf, Thomas Bugge and Roman Szabo for their generous donation of plasmids and the cell line expressing Matriptase-3.

Author Contributions

Conceived and designed the experiments: OB FB RL CMO. Performed the experiments: OB AD UE RK. Analyzed the data: OB AD UE RK CMO. Contributed reagents/materials/analysis tools: OB AD UE RK FB RL CMO. Wrote the paper: OB AD RK UE CMO. Prepared all the figures: OB AD RK UE. Produced and purified the TTSPs: FB.

23. Szabo R, Netzel-Arnett S, Hobson JP, Antalis TM, Bugge TH (2005) Matriptase-3 is a novel phylogenetically preserved membrane-anchored serine protease with broad serpin reactivity. *Biochem J* 390: 231–242. doi:10.1042/BJ20050299.
24. Cal S, Quesada V, Garabaya C, López-Otín C (2003) Polysaccharase-I, a human polyprotease with the ability to generate independent serine protease domains from a single translation product. *Proc Natl Acad Sci USA* 100: 9185–9190. doi:10.1073/pnas.1633392100.
25. Hooper JD, Scarman AL, Clarke BE, Normyle JF, Antalis TM (2000) Localization of the mosaic transmembrane serine protease corin to heart myocytes. *Eur J Biochem* 267: 6931–6937.
26. Chan JCY, Knudson O, Wu F, Morser J, Dole WP, et al. (2005) Hypertension in mice lacking the proatrial natriuretic peptide convertase corin. *Proc Natl Acad Sci USA* 102: 785–790. doi:10.1073/pnas.0407234102.
27. López-Otín C, Overall CM (2002) Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* 3: 509–519. doi:10.1038/nrm858.
28. Lee SL, Dickson RB, Lin CY (2000) Activation of hepatocyte growth factor and urokinase/plasminogen activator by matriptase, an epithelial membrane serine protease. *J Biol Chem* 275: 36720–36725. doi:10.1074/jbc.M007802200.
29. Béliveau F, Désilets A, Leduc R (2009) Probing the substrate specificities of matriptase, matriptase-2, hepsin and DESC1 with internally quenched fluorescent peptides. *FEBS Journal* 276: 2213–2226. doi:10.1111/j.1742-4658.2009.06950.x.
30. Velasco G, Cal S, Quesada V, Sánchez LM, López-Otín C (2002) Matriptase-2, a membrane-bound mosaic serine proteinase predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins. *J Biol Chem* 277: 37637–37646. doi:10.1074/jbc.M203007200.
31. Wysocka M, Spichalska B, Lesner A, Jaros M, Brzozowski K, et al. (2010) Substrate specificity and inhibitory study of human airway trypsin-like protease. *Bioorg Med Chem* 18: 5504–5509. doi:10.1016/j.bmc.2010.06.059.
32. Rawlings ND, Waller M, Barrett AJ, Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* 42: D503–D509. doi:10.1093/nar/gkt953.
33. Schilling O, Overall CM (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* 26: 685–694. doi:10.1038/nbt1408.
34. Schilling O, Huesgen PF, Barré O, auf dem Keller U, Overall CM (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. doi:10.1038/nprot.2010.178.
35. Starr AE, Bellac CL, Dufour A, Goebeler V, Overall CM (2012) Biochemical characterization and N-terminomics analysis of leukolysin, the membrane-type 6 matrix metalloprotease (MMP25): chemokine and vimentin cleavages enhance cell migration and macrophage phagocytic activities. *J Biol Chem* 287: 13382–13395. doi:10.1074/jbc.M111.314179.
36. Eckhard Ulrich, Huesgen PF, Brandstetter H, Overall CM (2014) Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen. *J Proteomics* 100: 102–114. doi:10.1016/j.jprot.2013.10.004.
37. Marino G, Huesgen PF, Eckhard Ulrich, Overall CM, Schröder WP, et al. (2014) Family-wide characterization of matrix metalloproteinases from *Arabidopsis thaliana* reveals their distinct proteolytic activity and cleavage site specificity. *Biochem J* 457: 335–346. doi:10.1042/BJ20130196.
38. Becker-Pauly C, Barré O, Schilling O, auf dem Keller U, Ohler A, et al. (2011) Proteomic analyses reveal an acidic prime site specificity for the astacin metalloprotease family reflected by physiological substrates. *Mol Cell Proteomics* 10: M111.009233. doi:10.1074/mcp.M111.009233.
39. Schilling O, auf dem Keller U, Overall CM (2011) Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. *Biol Chem* 392: 1031–1037. doi:10.1515/BC.2011.158.
40. Kappelhoff R, Overall C (2001) The CLIP-CHIP Oligonucleotide Microarray: Dedicated Array for Analysis of All Protease, Nonproteolytic Homolog, and Inhibitor Gene Transcripts in Human and Mouse. Hoboken, NJ, USA: John Wiley & Sons, Inc. doi:10.1002/0471140864.ps2119s56.
41. Kappelhoff R, auf dem Keller U, Overall CM (2010) Analysis of the degradome with the CLIP-CHIP microarray. *Methods Mol Biol* 622: 175–193. doi:10.1007/978-1-60327-299-5_10.
42. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20: 3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551:AID-ELPS3551>3.0.CO;2-2.
43. Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20: 1466–1467. doi:10.1093/bioinformatics/bth092.
44. Keller A, Nesvizhskii AI, Kolker E, 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.
45. Shteynberg D, Deutsch EW, Lam H, Eng JK, Sun Z, et al. (2011) iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. *Mol Cell Proteomics* 10: M111.007690. doi:10.1074/mcp.M111.007690.
46. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, et al. (2010) A guided tour of the Trans-Proteomic Pipeline. *Proteomics* 10: 1150–1159. doi:10.1002/pmic.200900375.
47. Friedrich R, Fuentes-Prior P, Ong E, Coombs G, Hunter M, et al. (2002) Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase. *J Biol Chem* 277: 2160–2168. doi:10.1074/jbc.M109830200.
48. Katz BA, Luong C, Ho JD, Somoza JR, Gjerstad E, et al. (2004) Dissecting and designing inhibitor selectivity determinants at the S1 site using an artificial Ala190 protease (Ala190 uPA). *J Mol Biol* 344: 527–547. doi:10.1016/j.jmb.2004.09.032.
49. Kyrieleis OJP, Huber R, Ong E, Oehler R, Hunter M, et al. (2007) Crystal structure of the catalytic domain of DESC1, a new member of the type II transmembrane serine proteinase family. *FEBS J* 274: 2148–2160. doi:10.1111/j.1742-4658.2007.05756.x.
50. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, et al. (2009) Protein structure modeling using SWISS-MODEL workspace. *Nat Protoc* 4: 1–13. doi:10.1038/nprot.2008.197.
51. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201. doi:10.1093/bioinformatics/bti770.
52. Xu D, Jaroszewski L, Li Z, Godzik A (2014) FFAS-3D: improving fold recognition by including optimized structural features and template re-ranking. *Bioinformatics* 30: 660–667. doi:10.1093/bioinformatics/btt578.
53. Jaroszewski L, Li Z, Cai X-H, Weber C, Godzik A (2011) FFAS server: novel features and applications. *Nucleic Acids Res* 39: W38–W44. doi:10.1093/nar/gkr441.
54. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19: 311–312.
55. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O (2011) Rosetta FlexPepDock web server—high resolution modeling of peptide-protein interactions. *Nucleic Acids Res* 39: W249–W253. doi:10.1093/nar/gkr431.
56. Raveh B, London N, Schueler-Furman O (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* 78: 2029–2040. doi:10.1002/prot.22716.
57. Yuan C, Chen L, Meehan EJ, Daly N, Craik DJ, et al. (2011) Structure of catalytic domain of Matriptase in complex with Sunflower trypsin inhibitor-1. *BMC Struct Biol* 11: 30. doi:10.1186/1472-6807-11-30.
58. DeLano WL (2005) The case for open-source software in drug discovery. *Drug Discov Today* 10: 213–217. doi:10.1016/S1359-6446(04)03363-X.
59. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7: 539. doi:10.1038/msb.2011.75.
60. Herter S, Piper DE, Aaron W, Gabriele T, Cutler G, et al. (2005) Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers. *Biochem J* 390: 125–136. doi:10.1042/BJ20041955.
61. Tripathi M, Nandana S, Yamashita H, Ganesan R, Kirchofer D, et al. (2008) Laminin-332 is a substrate for hepsin, a protease associated with prostate cancer progression. *J Biol Chem* 283: 30576–30584. doi:10.1074/jbc.M802312200.
62. Overall CM (2002) Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol Biotechnol* 22: 51–86. doi:10.1385/MB:22:1:051.
63. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31: 3320–3323.
64. Bode W, Mayr I, Baumann U, Huber R, Stone SR, et al. (1989) The refined 1.9 Å crystal structure of human alpha-thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J* 8: 3467–3475.
65. Huber R, Kukla D, Bode W, Schwager P, Bartels K, et al. (1974) Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. *J Mol Biol* 89: 73–101.
66. Magrane M, Consortium U (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011: bar009. doi:10.1093/database/bar009.
67. Bhatt AS, Erdjument-Bromage H, Tempst P, Craik CS, Moasser MM (2005) Adhesion signaling by a novel mitotic substrate of src kinases. *Oncogene* 24: 5333–5343. doi:10.1038/sj.onc.1208582.
68. Satomi S, Yamasaki Y, Tsuzuki S, Hitomi Y, Iwanaga T, et al. (2001) A Role for Membrane-Type Serine Protease (MT-SP1) in Intestinal Epithelial Turnover. *Biochemical and Biophysical Research Communications* 287: 995–1002. doi:10.1006/bbrc.2001.5686.
69. Darragh MR, Bhatt AS, Craik CS (2008) MT-SP1 proteolysis and regulation of cell-microenvironment interactions. *Front Biosci* 13: 528–539.
70. Ahmed S, Jin X, Yagi M, Yasuda C, Sato Y, et al. (2006) Identification of membrane-bound serine proteinase matriptase as processing enzyme of insulin-like growth factor binding protein-related protein-1 (IGFBP-rP1/angiomodulin/mac25). *FEBS Journal* 273: 615–627. doi:10.1111/j.1742-4658.2005.05094.x.
71. Netzel-Arnett S, Currie BM, Szabo R, Lin CY, Chen LM, et al. (2006) Evidence for a Matriptase-Prostasin Proteolytic Cascade Regulating Terminal Epidermal Differentiation. *Journal of Biological Chemistry* 281: 32941–32945. doi:10.1074/jbc.C600208200.
72. Jin X, Yagi M, Akiyama N, Hirotsaki T, Higashi S, et al. (2006) Matriptase activates stromelysin (MMP-3) and promotes tumor growth and angiogenesis. *Cancer Science* 97: 1327–1334. doi:10.1111/j.1349-7006.2006.00328.x.
73. Stürnberg M, Maurer E, Horstmeyer A, Kolp S, Frank S, et al. (2010) Proteolytic processing of the serine protease matriptase-2: identification of the cleavage sites

- required for its autocatalytic release from the cell surface. *Biochem J* 430: 87–95. doi:10.1042/BJ20091565.
74. Maxson JE, Chen J, Enns CA, Zhang AS (2010) Matriptase-2- and Proprotein Convertase-cleaved Forms of Hemojuvelin Have Different Roles in the Down-regulation of Hepsidin Expression. *Journal of Biological Chemistry* 285: 39021–39028. doi:10.1074/jbc.M110.183160.
75. Kazama Y, Hamamoto T, Foster DC, Kiesel W (1995) Hepsin, a putative membrane-associated serine protease, activates human factor VII and initiates a pathway of blood coagulation on the cell surface leading to thrombin formation. *J Biol Chem* 270: 66–72.
76. Moran P, Li W, Fan B, Vij R, Eigenbrot C, et al. (2006) Pro-urokinase-type plasminogen activator is a substrate for hepsin. *J Biol Chem* 281: 30439–30446. doi:10.1074/jbc.M605440200.
77. Jiang J, Wu S, Wang W, Chen S, Peng J, et al. (2011) Ectodomain shedding and autocleavage of the cardiac membrane protease corin. *Journal of Biological Chemistry* 286: 10066–10072. doi:10.1074/jbc.M110.185082.
78. Beaufort N, Leduc D, Eguchi H, Mengele K, Hellmann D, et al. (2007) The human airway trypsin-like protease modulates the urokinase receptor (uPAR, CD87) structure and functions. *AJP: Lung Cellular and Molecular Physiology* 292: L1263–L1272. doi:10.1152/ajplung.00191.2006.