

# UCSF

## UC San Francisco Previously Published Works

### Title

Mutational Tail Loss Is an Evolutionary Mechanism for Liberating Marapsins and Other Type I Serine Proteases from Transmembrane Anchors\*

### Permalink

<https://escholarship.org/uc/item/4jp2n51n>

### Journal

Journal of Biological Chemistry, 288(15)

### ISSN

0021-9258

### Authors

Raman, Kavita

Trivedi, Neil N

Raymond, Wilfred W

et al.

### Publication Date

2013-04-01

### DOI

10.1074/jbc.m112.449033

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Mutational Tail Loss Is an Evolutionary Mechanism for Liberating Marapsins and Other Type I Serine Proteases from Transmembrane Anchors<sup>\*[5]</sup>

Received for publication, January 7, 2013, and in revised form, February 21, 2013. Published, JBC Papers in Press, February 27, 2013, DOI 10.1074/jbc.M112.449033

Kavita Raman<sup>‡</sup>, Neil N. Trivedi<sup>§</sup>, Wilfred W. Raymond<sup>‡</sup>, Rajkumar Ganesan<sup>¶</sup>, Daniel Kirchhofer<sup>¶</sup>, George M. Verghese<sup>||</sup>, Charles S. Craik<sup>\*\*</sup>, Eric L. Schneider<sup>\*\*</sup>, Shilpa Nimishakavi<sup>‡</sup>, and George H. Caughey<sup>‡§##§§1</sup>

From the <sup>‡</sup>Cardiovascular Research Institute and Departments of <sup>§</sup>Medicine and <sup>\*\*</sup>Pharmaceutical Chemistry, University of California, San Francisco, California 94143, <sup>¶</sup>Department of Early Discovery Biochemistry, Genentech, Inc., South San Francisco, California 94080, <sup>||</sup>Department of Medicine, University of Virginia, Charlottesville, Virginia 22908, and <sup>##</sup>Northern California Institute for Research and Education and <sup>§§</sup>Veterans Affairs Medical Center, San Francisco, California 94121

**Background:** Vertebrate marapsins can be either type I transmembrane proteases or unanchored.

**Results:** Point mutations liberated marapsins from transmembrane peptides independently in human-related primates and other mammalian clades. Soluble marapsins are active and inhibitor-resistant.

**Conclusion:** Mutational tail loss transformed transmembrane marapsins and related proteins into soluble proteases.

**Significance:** These findings suggest a general evolutionary mechanism for evolving proteases with new properties and functions.

Human and mouse marapsins (Prss27) are serine proteases preferentially expressed by stratified squamous epithelia. However, mouse marapsin contains a transmembrane anchor absent from the human enzyme. To gain insights into physical forms, activities, inhibition, and roles in epithelial differentiation, we traced tail loss in human marapsin to a nonsense mutation in an ancestral ape, compared substrate preferences of mouse and human marapsins with those of the epithelial peptidase prostatic, designed a selective substrate and inhibitor, and generated *Prss27*-null mice. Phylogenetic analysis predicts that most marapsins are transmembrane proteins. However, nonsense mutations caused membrane anchor loss in three clades: human/bonobo/chimpanzee, guinea pig/degu/tuco-tuco/mole rat, and cattle/yak. Most marapsin-related proteases, including prostasins, are type I transmembrane proteins, but the closest relatives (prosemins) are not. Soluble mouse and human marapsins are tryptic with subsite preferences distinct from those of prostatic, lack general proteinase activity, and unlike prostasins resist antiproteases, including leupeptin, aprotinin, serpins, and  $\alpha_2$ -macroglobulin, suggesting the presence of non-canonical active sites. *Prss27*-null mice develop normally in barrier conditions and are fertile without overt epithelial defects, indicating that marapsin does not play critical, non-redundant roles in development, reproduction, or epithelial differentiation. In conclusion, marapsins are conserved, inhibitor-resistant, tryptic peptidases. Although marapsins are type I transmembrane proteins in their typical form, they mutated

independently into anchorless forms in several mammalian clades, including one involving humans. Similar pathways appear to have been traversed by prosemins and tryptases, suggesting that mutational tail loss is an important means of evolving new functions of tryptic serine proteases from transmembrane ancestors.

The human degradome is estimated to include 176 serine proteases and serine protease homologs (1). Of these, the largest subset belongs to the S1 family of trypsin-like proteins, including familiar proteases associated with digestion of food; with clotting and fibrinolytic and complement cascades; and with neutrophil, lymphocyte, and mast cell granules and host defense against infection. S1 serine proteases serve wide ranging functions and take many forms. The majority is secreted and soluble. A minority is membrane-associated via peptide or glycosylphosphatidylinositol (GPI)<sup>2</sup> anchors. The membrane group contains proteins with a transmembrane anchor that is C- or N-terminal to the catalytic domain (types I and II, respectively) (2, 3). Although genes encoding type I transmembrane proteins comprise but a small fraction of the serine protease degradome, phylogenetic analysis suggests that at least one group of extant soluble serine proteases (namely, mast cell  $\alpha/\beta/\delta$ -tryptases) evolved from type I proteins related to  $\gamma$ -tryptases and prostasins (4–9). However, the putative conversion lacks a mechanism and has not been observed within a specific clade of mammalian proteases. Seeking such evidence, we focused on marapsin (also known as pancreasin and PRSS27), which belongs to a group of serine proteases featuring members that can be either soluble or type I transmembrane proteins but are otherwise closely related (e.g. human *versus* mouse (10)) with

\* This work was supported, in whole or in part, by National Institutes of Health Grant P01 HL024136 (to K. R., W. W. R., S. N., and G. H. C.). This work was also supported by the Northern California Institute for Research and Education (to N. N. T. and G. H. C.) and United States Department of Defense Grant PC111318 (to C. S. C.).

[5] This article contains supplemental Figs. S1–S4 and Tables S1 and S2.

<sup>1</sup> To whom correspondence should be addressed: Veterans Affairs Medical Center, 111D, 4150 Clement St., San Francisco, CA 94121. Tel.: 415-221-4810; Fax: 415-379-5538; E-mail: george.caughey@ucsf.edu.

<sup>2</sup> The abbreviations used are: GPI, glycosylphosphatidylinositol; 4NA, 4-nitroaniline; bis-Tris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol.

highly conserved catalytic domains in mammals and some other vertebrates (7, 11). Marapsins, therefore, provide an opportunity to test the hypothesis that some soluble serine proteases evolved from membrane-anchored proteases and to explore genetic mechanisms of the conversion.

Characterization of marapsins has been limited. Insights into enzymatic activity derive from studies of recombinant human marapsin/pancreasin, which is tryptic, preferring P1 Arg as assessed with a small panel of peptidyl 4-nitroanilides (10). No general protease activity has been identified for the human enzyme, which has the characteristic (unusual for a tryptic serine protease) of resisting inactivation by broad spectrum inhibitors such as aprotinin and soybean trypsin inhibitor. This suggests that the active site is restricted compared with classical trypsin-like serine proteases. The observed conservation of marapsin catalytic domains led us to hypothesize that substrate selectivity and inhibitor resistance are general characteristics of marapsins. Both human and mouse marapsins are modified by *N*-glycosylation (10, 12). Additionally, the mouse enzyme contains a 34-residue, C-terminal hydrophobic extension (absent in the human enzyme) predicted to span cell membranes (10). Mouse marapsin expressed in mammalian cells is membrane-associated but shed by phosphatidylinositol-specific phospholipase C (3), suggesting that the transmembrane peptide is swapped for a GPI anchor (12). In agreement with hydrophobicity analysis as well as with predictions that the human enzyme lacks a C-terminal anchor, native human marapsin is secreted from an esophageal cell line (13). The marapsin gene is expressed by a variety of human, mouse, and opossum tissues (9, 10, 12, 13). In humans and mice, it is especially abundant in stratified squamous epithelia of cornea, larynx, esophagus, and cervix (13, 14) and is induced in wounded and psoriatic skin (13, 14). These patterns of expression in combination with evidence of gene conservation in mammals encouraged speculation that marapsin serves critical functions, such as epithelial differentiation. To test hypotheses concerning origination of soluble marapsins from ancestral transmembrane proteases, clade-specific enzymatic properties of marapsins, and importance to epithelial development, we analyzed vertebrate marapsin and related protein sequences, compared mouse and human marapsin extended substrate preferences with those of prosta-sin, and characterized a marapsin-deficient mouse.

## EXPERIMENTAL PROCEDURES

**Data Mining**—Full amino acid sequences of marapsins not previously published or annotated were obtained via BLAST searches of high throughput genome sequence and whole genome shotgun databases at the National Center for Biotechnology Information. Human and mouse marapsin genes and cDNAs (10) were used as query sequences. Previously unreported amino acid sequences of marapsins and marapsin-like proteases were predicted from genomic DNA using existing human and mouse gene structures as a template following standard rules for placement of intron-exon boundaries as detailed in prior publications involving phylogenetic studies of related serine proteases (5, 7, 8, 10). Resulting protein sequences were aligned using Geneious software (Biomatters, Auckland, New Zealand). Candidate marapsins were consid-

ered to be intact if there were no catalytic domain insertions or deletions relative to published, cDNA-validated human and mouse sequence (3, 10, 13) and if tryptic serine protease catalytic and specificity triad residues were conserved.

**Sequencing of Marapsin Genes**—Gorilla genomic DNA from cultured fibroblasts was obtained from the Primate Cell Repositories of the Coriell Institute for Medical Research (Camden, NJ). DNA encoding the marapsin gene sixth exon was amplified by PCR using the primer pair 5'-GGGTTCTTGATGAGGAA-GTCCGTTGAG and 5'-AGCTGGCACACAGGCTGGGTT-TTTATT. Amplimers were cloned into pCR2.1-TOPO (Invitrogen) and sequenced.

**Identification of Truncation Mutations**—Transcript alterations leading to tail loss in translated sequences of vertebrate marapsins were identified by alignment of nucleotide and predicted amino acid sequence bracketing the open reading frame of marapsin gene exon 6. Molecular time estimates of truncation mutations were calculated using TimeTree.

**Prediction of Post-translational Modifications**—Sites containing sequence Asn-Xaa-Ser/Thr (where Xaa is any amino acid) were considered potential sites of *N*-glycosylation. Candidate C-terminal transmembrane segments were identified by the Dense Alignment Surface algorithm (15) as applied to marapsins with open reading frame extensions relative to the human catalytic domain. Possible sites of GPI anchor attachment in marapsins with predicted C-terminal transmembrane peptide segments were identified using BigPI (16).

**Phylogenetic Analysis**—To probe cladistic relationships between catalytic domains, the analysis was restricted to sequence starting from Met<sup>1</sup> of the mature catalytic domain to avoid potential errors or biases arising from miscalls of prepro sequence that lie on small exons (10)). On the C-terminal end, aligned sequences were cut off at conserved His<sup>239</sup> of marapsin (or equivalent residue in related proteases) to avoid biasing effects of length variations in protease C termini, which do not universally include a membrane-spanning domain, which can extend protease sequence by 30 or more residues. A rooted tree was generated by unweighted pair group with arithmetic mean analysis using methods described previously (7).

**Recombinant Expression of Soluble Proteases**—Mouse marapsin was expressed in CHO cells as a soluble protein comprising residues Met<sup>1</sup>–Gln<sup>292</sup> with the native hydrophobic C-terminal transmembrane peptide replaced by a His<sub>8</sub> epitope. This activated enzyme, which contained the native, disulfide-linked propeptide, was purified from supernatants of conditioned medium by nickel affinity chromatography (13). Recombinant full-length human marapsin (Met<sup>1</sup>–Lys<sup>290</sup>–Ala–Ala–Ala–His<sub>8</sub>) was expressed and purified similarly in its activated form (13). Combinatorial substrate profiling studies used mouse myeloma cell-expressed recombinant mouse marapsin (R&D Systems, Minneapolis, MN) comprising residues Ala<sup>23</sup>–Thr<sup>290</sup> with the native hydrophobic C-terminal transmembrane peptide replaced by a His<sub>6</sub> epitope. Met<sup>1</sup>–Gln<sup>292</sup>–His<sub>8</sub> mouse marapsin was subjected to PAGE (4–12% NuPAGE bis-Tris gel with MES-SDS buffer, Invitrogen) under reducing (50 mM dithiothreitol) and non-reducing conditions followed by Coomassie Blue staining. Recombinant soluble mouse prosta-sin with C-terminal transmembrane sequence deleted at Ser<sup>313</sup>

## Evolution of Serine Proteases by Mutational Tail Loss

was produced in *Escherichia coli* and activated as described (17).

**Mapping of Protease Subsites by Profiling with Combinatorial Fluorogenic Tetrapeptide Library**—A positional scanning synthetic combinatorial library approach (18) as applied to other serine proteases (19, 20) was used to identify residues preferred by recombinant soluble mouse and human marapsins in each of four positions (P1–P4) among fluorogenic tetrapeptides. A similar approach was used for mouse prostaticin (170 nM). Because of limited recombinant material and prior work with mouse and human prostaticins showing tryptic activity and preference for P1 Arg, this work tested mouse prostaticin using libraries with Arg fixed in the P1 position. Recombinant marapsins were studied at 300 nM in 50 mM Tris, 250 mM NaCl, 0.05% Brij 35 (pH 8.0) at 25 °C.

**Design, Generation, and Validation of Custom Substrate and Inhibitor**—A colorimetric 4-nitroanilide (4NA) substrate based on residues preferred by mouse marapsin at each subsite as determined using the positional scanning combinatorial library was synthesized by Anaspec (San Jose, CA) along with a potentially selective, peptidic, covalent (chloromethyl ketone) inhibitor. Potency and selectivity of the custom synthesized substrate and inhibitor were assessed in comparisons of recombinant mouse and human marapsins, recombinant mouse prostaticin (R&D Systems) and matriptase (R&D Systems), and cattle trypsin (Sigma-Aldrich). Mouse marapsin also was tested against potential natural peptide and protein substrates containing favored sequences based on results of combinatorial screening of tetrapeptide fluorogenic substrates. Proteinaceous and small molecular weight inhibitors selected for the ability to inactivate other tryptic serine proteases were screened using the general tryptic protease substrate butyloxycarbonyl-QAR-4NA (0.5 mM; Bachem Americas, Torrance, CA). These inhibitors included 4-(2-aminoethyl)benzenesulfonyl fluoride, benzamidine, aprotinin, soybean trypsin inhibitor (all from Sigma-Aldrich), leupeptin (Enzo Life Sciences, Plymouth Meeting, PA), and nafamostat mesylate (Santa Cruz Biotechnology, Santa Cruz, CA).

**Gel Filtration**—Recombinant mouse marapsin (5.1  $\mu$ g) was added to 0.3 ml of pooled mouse sera (Sigma-Aldrich). After incubation for 15 min at 25 °C, 0.1 ml of the mixture was injected onto a size exclusion column (Superose 6 GL 10/300, GE Healthcare) equilibrated with PBS (pH 7.4) and eluted using an ÄKTA Purifier System (GE Healthcare). Eluate was monitored for absorbance at 280 nm and collected in 0.3-ml fractions, which were assayed for QAR-4NA-hydrolyzing activity in buffer containing 50  $\mu$ M aprotinin and leupeptin in 96-well assay plates (Costar 3370, Corning Life Sciences, Lowell, MA) sealed with TempPlate RT Optical Film (USA Scientific, Ocala, FL) to minimize evaporation. Initial change in absorbance at 410 nm at 25 °C was recorded on a temperature-controlled Synergy 2 plate reader (BioTek, Winooski, VT). As a negative control, an aliquot of unspiked pooled mouse sera was loaded, eluted, and assayed similarly. The column was precalibrated with globular proteins of known size: thyroglobulin (669 kDa), apoferritin (460 kDa),  $\gamma$ -globulin (158 kDa), BSA (66 kDa), soybean trypsin inhibitor (30 kDa), and aprotinin (6 kDa).

**Generation of *Prss27*<sup>-/-</sup> Mice**—Marapsin (*Prss27*)-deficient mice were generated in collaboration with Lexicon Pharmaceuticals (The Woodlands, TX) using gene trapping-based insertional mutagenesis (21) and retroviral vectors with splice acceptor sites to target expressed genes in mice. Disrupted genes were identified by 3' rapid amplification of cDNA ends, and a library of embryonic stem cell clones was created (21). Clone OST34330 contained a disrupted *Prss27* gene (deleting exons 2 through 6) as established by DNA blotting of EcoRI-digested genomic DNA using probes corresponding to genomic sequence internal and external to the targeting vector sequence. The internal probe yielded 27.0- and 5.7-kb bands for wild type and targeted genomes, respectively, whereas the external probe yielded 27.0- and 16.7-kb bands, respectively. Targeted ES cell (129/SvEvBrd) clones were microinjected into C57Bl/6 (albino) blastocysts to generate chimeric mice that were crossed with C57Bl/6 mice to generate 50% 129/SvEvBrd, 50% C57Bl/6 progeny. Resulting heterozygous offspring were interbred to produce *Prss27*<sup>-/-</sup> mice, which were backcrossed via single nucleotide polymorphism-assisted genome scanning (The Jackson Laboratory, Bar Harbor, ME) into a C57Bl/6J background until congenic. Single nucleotide polymorphism analysis suggested that the *Prss27*<sup>-/-</sup> mice are >99% identical to C57Bl/6J+/+ mice. Genotyping was performed on tail lysates via a PCR-based assay using primer pairs specific for wild type (forward, 5'-CAGGTAGGACTTAAGTGTCC and reverse, 5'-CCAGCAGGACCTGGTATATG, generating a 278-bp band) and mutant (forward, 5'-CCTGCCCTGCATCCTTGTATGG and reverse, 5'-GCAGCGCATCGCCTTC-TATC, generating a 266-bp band) chromosomes. The PCR was performed for 35 cycles of melting, annealing, and extension at 94, 60, and 72 °C, respectively. A marapsin genomic clone in pKOS and targeted ES cell genomic DNA served as wild type (+/+) and heterozygote (+/-) positive controls, respectively.

**Immunohistochemical Characterization of Mouse Tissues**—Formalin-fixed, paraffin-embedded sections were deparaffinized in xylene and hydrated. Sections were blocked with 10% goat serum for 30 min at room temperature and then stained for 1 h with 0.9  $\mu$ g/ml monoclonal hamster anti-mouse marapsin antibody 3B2 (13) or with hamster IgG control antibody (BioLegend, San Diego, CA; 0.9  $\mu$ g/ml). After two washes in PBS, sections were incubated for 30 min with biotinylated goat anti-hamster IgG (Santa Cruz Biotechnology; 1.3  $\mu$ g/ml). After two further washes, sections were stained by incubation for 5 min with Vectastain ABC Reagents (Vector Laboratories, Burlingame, CA) and metal-enhanced diaminobenzidine (Thermo Fisher Scientific, Waltham, MA). Sections then were counterstained with hematoxylin, dehydrated in a graded series of alcohols, cleared in xylene, and mounted.

## RESULTS

**Marapsins Diverged from Related Serine Proteases in Pre-mammalian Vertebrates and Are Highly Conserved in Mammals**—As shown in the cladogram in Fig. 1, marapsins (as deduced from genomic or cDNA sequence) are present in a wide variety of placental and marsupial mammals and in reptiles. The breadth of representation in mammalian genomes is similar to that of prostaticin (7, 22), a related enzyme that is

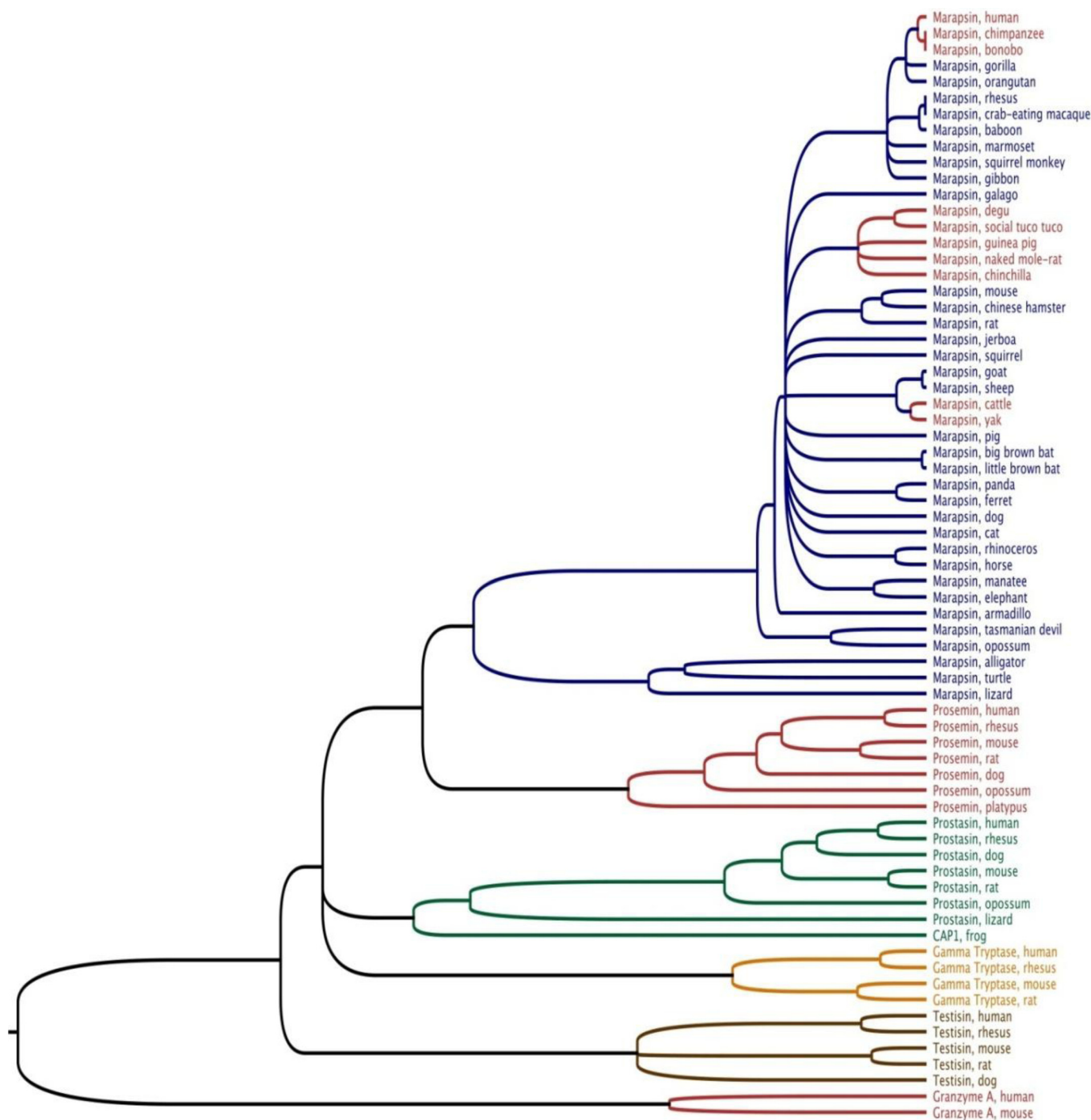
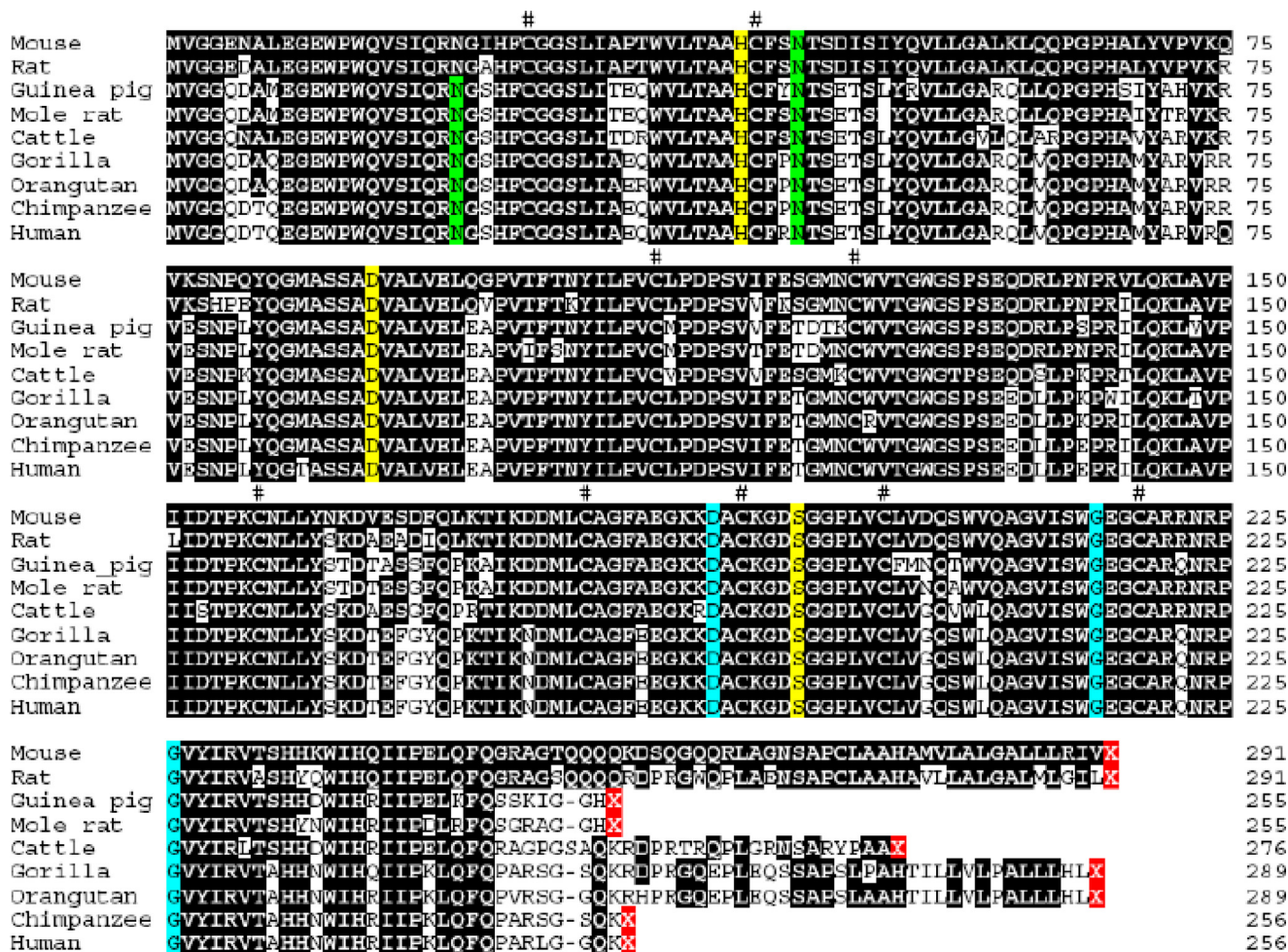


FIGURE 1. **Cladogram.** This rooted tree probing relationships between vertebrate marapsins and related serine proteases was generated by unweighted pair group with arithmetic mean analysis of aligned protein sequences. To allow comparison of proteins with signal peptides, propeptides, and C termini of varying length, the alignment was limited to mature catalytic domains with deletion of C-terminal extensions present in a subset of the proteases. Nodes were assigned if predicted by at least 550 of 1000 iterations of bootstrap resampling. Mouse granzyme A and human granzyme A, which are tryptic serine proteases not closely related to marapsins, prostates, or tryptases, together serve as an outgroup. Clades of proteases known or predicted to lack a membrane anchor are *red*. The other proteases are predicted to be type I transmembrane proteins based on the presence of a C-terminal, hydrophobic extension of membrane-spanning length. Note that channel-activating protease (CAP-1) appears to be a frog ortholog of mammalian prostates. Accession numbers of sequences used in tree construction are given in supplemental Table S1.

essential for embryonic development and postnatal survival in mice (23, 24). Marapsin genes are more widespread than those that encode  $\gamma$ -tryptase, which is absent from genomes of dogs and several other mammals and lacks an identified ortholog in non-mammalian vertebrates (7, 9). Marapsin catalytic domain sequence conservation is high in comparison with related

enzymes like testisin and  $\gamma$ -tryptase as reflected by cladogram line length, which is proportional to the number of sequence mismatches between a given pair of sequences. High level conservation also is evident from the alignment in Fig. 2, which shows little variation in sequence and even less in length except in the C-terminal tail, which is not part of the standard catalytic

## Evolution of Serine Proteases by Mutational Tail Loss



**FIGURE 2. Alignment of marapsins.** Selected mammalian marapsin sequences beginning with Met<sup>1</sup> of the predicted mature catalytic domain after activation are aligned. The nine absolutely conserved Cys residues, one of which (Cys<sup>110</sup>) is predicted to link with propeptide Cys<sup>-9</sup> (not shown), are marked with a "#." Consensus conserved sites of *N*-glycosylation sites are green. The absolutely conserved catalytic triad residues (His<sup>41</sup>, Asp<sup>90</sup>, and Ser<sup>195</sup>) and specificity triad residues (Asp<sup>189</sup>, Gly<sup>216</sup>, and Gly<sup>226</sup>) essential for S1 serine protease function and tryptic specificity are yellow and cyan, respectively. Other aligned residues that are identical to those in the mouse marapsin sequence are black. Sequence-terminating stop codons are red, revealing marked length variations.

domain exemplified by trypsin and  $\beta$ -trypsinases (5, 25). Gene preservation and high level conservation of protein sequence can indicate biological importance and conserved function. In this regard, it is notable that all of the marapsins (a subset of which are shown in Fig. 2) are intact in that they lack major insertions, deletions, or truncations and possess the "catalytic triad" residues (His<sup>57</sup>, Asp<sup>102</sup>, and Ser<sup>195</sup> by standard chymotrypsinogen numbering) essential in serine proteases for attacking and hydrolyzing peptide bonds. Each of the sequences also contains the "specificity triad" residues (Asp<sup>189</sup>, Gly<sup>216</sup>, and Gly<sup>226</sup>) by chymotrypsinogen numbering) featured in serine proteases with trypsin-like primary specificity for basic amino acids at the site of cleavage (26–28). These triad residues line the pocket that accommodates the basic side chain of lysine or arginine substrates.

The large number of extant vertebrate marapsin and marapsin-related sequences contributing to the tree strongly reinforces prior suggestions that catalytic domain of marapsin is most closely related to that of the tryptic protease prosemmin/Prss22/trypsinase  $\epsilon$  (6, 9, 12). Identification of reptilian marapsins predicts that marapsins last shared ancestry with prosemmin before mammals evolved as a separate lineage from other vertebrates. One conserved and highly idiosyncratic feature of the

mature catalytic domain is methionine as the first residue, which is Ile or Val in the vast majority of other trypsin-like serine proteases, including the closest relatives of marapsin shown in Fig. 1. The number and position of cysteines are highly conserved in mammalian marapsins, including Cys<sup>110</sup> (Fig. 2), which is predicted to form a disulfide linkage with the propeptide (10), and are similar to those of prosemmin, prostasin, testisin, and  $\gamma$ -trypsin. The propeptide itself (not shown) includes highly conserved consensus sequence CGRPRMLNR (see supplemental Table S2), which begins with the Cys<sup>-9</sup> proposed to link with Cys<sup>110</sup> and ends at the proposed site of cleavage activation at universally conserved Arg<sup>-1</sup>. As shown by the alignment in Fig. 2, two consensus sites of predicted *N*-glycosylation (Asn<sup>21</sup> and Asn<sup>45</sup>) are highly conserved in mammalian marapsins. This conservation combined with direct evidence from electrophoretic responses to *N*-glycosidase treatment that sugars are attached to one or more of these sites in recombinant marapsins (3, 10) suggests that *N*-glycosylation may be required for stability or function.

*Human Marapsin and Selected Other Marapsins Evolved into Soluble Proteases via Nonsense Mutations in Ancestors with Transmembrane Segments*—As shown by the alignments in Fig. 2, the most conspicuous structural variation among mamma-

lian marapsins is the length of the C-terminal region. The majority of marapsins, including reptilian versions, have a C-terminal extension that is hydrophobic and predicted to span lipid bilayers (10). In the mouse, this sequence can be exchanged for a lipid (GPI) membrane anchor (3). Indeed, most marapsins with a hydrophobic C-terminal extension have a consensus GPI anchor attachment site (for example, Asn<sup>269</sup> of mouse and rat sequences and Ser<sup>268</sup> of gorilla, gibbon, and orangutan sequences; see Fig. 2). However, human marapsin (10) and several other mammalian marapsins identified in the present work lack this hydrophobic extension as well as consensus GPI attachment sites; indeed, native human marapsin is secreted as a soluble protein (13). Thus, some marapsins have a peptide or lipid transmembrane anchor (and are typical type I transmembrane proteins), whereas others described in ensuing paragraphs appear to be typical soluble proteins. Although marapsin-related prosemis lack a C-terminal hydrophobic sequence and are not predicted to be transmembrane proteins, the slightly more distant cousins of marapsin in Fig. 1, including prostatic,  $\gamma$ -tryptase, and testisin, all are known or predicted to be type I transmembrane tryptic proteases (2, 3, 5, 17, 29) and are basal to marapsins as well as to prosemis on the tree. This is evidence that soluble marapsins and prosemis evolved from type I proteases rather than the other way around.

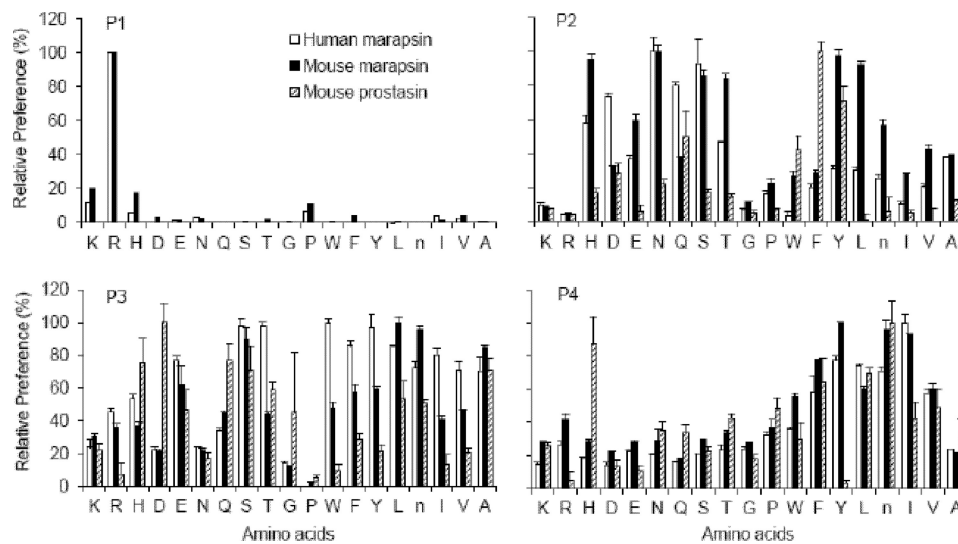
*A Nonsense Mutation in an Ancestral Great Ape Caused Loss of the C-terminal Transmembrane Anchor in Human, Chimpanzee, and Bonobo Marapsins*—As shown in Fig. 2, the predicted catalytic domains of human, chimpanzee, and bonobo end well before the corresponding sequences of other primates, including otherwise closely related great apes (gorilla and orangutan), a lesser ape (gibbon), and old and new world monkeys (sequences not shown; for species, see cladogram in Fig. 1) and a distantly related “primitive” primate (galago). The missing 33 residues include a predicted C-terminal membrane-spanning peptide and GPI anchor addition site. Therefore, human, chimpanzee, and bonobo marapsins are predicted to be soluble, *i.e.* neither peptide- nor lipid-anchored to cell membranes. As revealed by the nucleotide alignments in [supplemental Fig. S1](#), the genetic basis of the truncation is the same for humans and for the other two affected great apes: namely, a nonsense point mutation converting the Arg<sup>257</sup> codon into a stop codon. Identification of this mutation in humans and their two closest phylogenetic relatives among primates (as reflected by the clades in Fig. 1) suggests that this mutation occurred after humans last shared a common ancestor with gorillas ~8 million years ago (30) but before ancestors of chimpanzees, bonobos, and humans diverged into separate lineages ~6 million years ago (30). The high degree of homology in aligned exon 6 nucleotide sequences between human and ape sequences past the site of the nonsense mutation as revealed in [supplemental Fig. S1](#) further supports the comparatively recent occurrence of the mutation. Despite the introduction of the earlier stop codon, which truncates the translated sequence, the relic of the ancestral C-terminal hydrophobic tail persists as an untranslated open reading frame.

*A Separate Nonsense Mutation in an Ancestral Rodent Caused Transmembrane Anchor Loss in Marapsins of Guinea Pig-related Mammals Native to Two Continents*—As seen in

Fig. 2, early truncation of marapsin was also identified in members of rodent infraorder Hystricognathi, which are related only distantly to rats and mice. These animals (guinea pig, degu, social tuco-tuco, and naked mole rat) have a T→C nonsense mutation changing Gln<sup>266</sup> to a stop codon as shown in [supplemental Fig. S2](#). The resulting truncation occurs at a site very close to that seen in humans, chimpanzees, and bonobos and produces a catalytic domain of 255 residues, which is one residue shorter than human and 36 residues shorter than the rat and mouse catalytic domains, which contain a predicted transmembrane peptide and GPI attachment site. Thus, guinea pig and the related truncated marapsins like human marapsin provide evidence of a distinct but durable mutational tail loss mutation that occurred in an ancestral rodent after divergence of lineages leading to mice, rats, and squirrels and the ancestors of guinea pigs. Although the effect of this mutation is similar to the one identified in humans and closely related great apes, it is a distinct and otherwise unrelated event and occurred much longer ago (between 43 and 78 million years ago) given the much greater overall sequence divergence within this clade of affected rodents (as reflected by the *length of tines* in the Fig. 1 cladogram) as well as by the fact that the group includes mammals that are native to two long separated continents (Africa for mole rats and South America for guinea pig, degu, and social tuco-tuco).

*Recent Marapsin Gene Mutations Caused Truncation and Probable Transmembrane Anchor Loss in Cattle-like Mammals*—Another marapsin truncation was identified in a third small clade: cattle and yaks, which are a subset of ruminant mammals, which in turn are a subset of ungulates. The truncation was noted in genomes from at least two varieties of domestic cattle: *Bos primigenius taurus* and *Bos primigenius indicus*. As shown in Fig. 2, the mutation truncates marapsin at a site after that observed in the primate and rodent subgroups discussed above. Nonetheless, it eliminates the hydrophobic C-terminal peptide that is the requisite for forming a transmembrane anchor, which in turn is required for attaching GPI. Thus, cattle and yak marapsins are predicted to be unattached to membranes. As revealed by Fig. 2 and by the nucleotide sequence alignments in [supplemental Fig. S3](#), the truncation is not present in other bovids (sheep and goat) and ungulates (horse, rhinoceros, and pig) and therefore occurred between 4.7 and 30 million years ago. Two nucleotide changes are required to convert the corresponding codon in sheep to the stop codon in cattle and yaks. Without additional sequence information from other related species, it is not yet clear whether sheep and goats acquired an additional mutation after divergence from the ancestors of *Bos* species or whether the latter accumulated sequential mutations after divergence from ancestors of sheep. Based on overall sequence conservation and the cladogram in Fig. 1, the early termination event now present in cattle and yaks is more recent than the one observed in the rodent clade given that sheep appear to have shared a common ancestor with cattle more recently than guinea pigs shared an ancestor with naked mole rats.

## Evolution of Serine Proteases by Mutational Tail Loss



**FIGURE 3. Comparison of marapsin and prostaticin subsite preferences.** Mouse and human marapsins and mouse prostaticin were profiled using fluorogenic tetrapeptide substrates. For the marapsins, primary specificity was tested in a combinatorial library in which each of the designated 20 amino acids was held constant in turn as residues P2 through P4 were varied. Therefore, each assay condition tested a mixture of 8000 different peptide substrates for a given P1 residue. The amidolytic activity of marapsin liberates 7-amino-4-carbamoylmethylcoumarin. The initial readout in relative fluorescence units was normalized to the result for the most preferred amino acid in each subsite. The other panels show results of similar profiling at positions P2, P3, and P4, respectively, by fixing amino acids at the designated position and varying the residues at the remaining positions. Prostaticin was profiled at positions P2, P3, and P4 with P1 fixed at Arg. Abbreviations for amino acids are given in standard one-letter code; *n* is norleucine (substituting for Met). Error bars represent S.D.

*Human and Mouse Marapsin Subsite Preferences Differ from Those of Prostaticin*—As shown in Fig. 3 by screening of combinatorial tetrapeptidic substrates, recombinant soluble human, and mouse marapsin, preferences are similar with both peptidases preferring tryptic residues at position P1 at the site of hydrolysis as anticipated based on strict conservation of specificity triad residues Asp/Gly/Gly lining the primary specificity pocket of nearly every serine protease of tryptic specificity (20, 27). Although some tryptic serine proteases readily hydrolyze substrates with either Arg or Lys at P1 (18), the marapsins strongly prefer Arg. In the adjoining position at P2, polar residues (excepting Arg and Lys) are favored, and Gly and Pro are disfavored. P3 accommodates polar and hydrophobic amino acids with Gly and Pro disfavored as at P2. Hydrophobic residues are favored at P4. Substrate peptide preferences of recombinant soluble mouse prostaticin at P2, P3, and P4 (with P1 fixed at Arg) contrast strongly with those of the marapsins, thereby suggesting that the natural targets of these two classes of enzymes may not overlap.

*A Tetrapeptide Substrate Based on Marapsin Cleavage Preferences Is Selective for Marapsin versus Prostaticin*—To validate results of combinatorial peptide screening and to create a useful assay substrate for marapsin, a tetrapeptide substrate, succinyl-L-Tyr-Leu-Asn-Arg-4NA (YLNR-4NA), based on sequences preferred by marapsins and not by mouse prostaticin (see Fig. 3) was synthesized. As shown in Fig. 4A, this substrate is selective for mouse marapsin over mouse prostaticin, whereas commercially available substrate QAR-4NA (which is a substrate for human prostaticin and detects prostaticin and matriptase activity in cultured cells (31)) was preferred by mouse prostaticin.

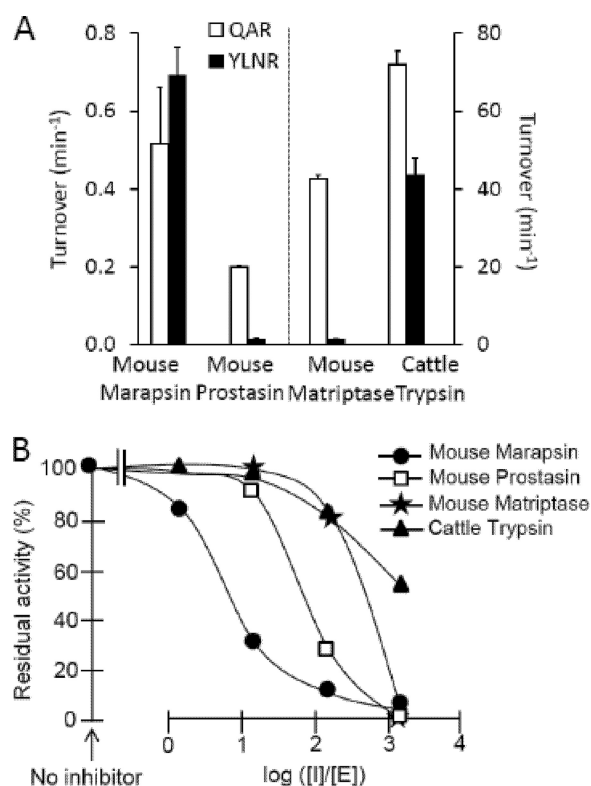
*A Tetrapeptide Covalent Inhibitor Is Selective for Mouse Marapsin*—To assess potential for developing a selective inhibitor, we tested peptidic compound biotinyl-L-Tyr-Leu-Asn-Arg-chloromethyl ketone (YLNR-chloromethyl ketone), which

was custom synthesized based on results of combinatorial screening. As shown in Fig. 4B, YLNR-chloromethyl ketone potently inactivates mouse marapsin and is selective based on weaker inhibition of other tryptic serine proteases, including mouse prostaticin, mouse matriptase and cattle trypsin.

*Marapsin Resists Inhibitors*—Although inactivated by YLNR-chloromethyl ketone, mouse marapsin is unusually resistant to other inhibitors that readily inhibit trypsin and other tryptic serine proteases as revealed by Fig. 5A. Mouse marapsin also resists inhibitors present in serum as indicated by the gel filtration chromatogram in Fig. 5B that shows preservation of marapsin activity upon addition to mouse serum, indicating that it resists inactivation by serum serpins, such as  $\alpha_1$ -antitrypsin. Most marapsin-like activity appears in an elution position of  $\sim 30$  kDa (relative to standard proteins), which is well below that of an  $\alpha_2$ -macroglobulin-bound complex ( $\sim 750$  kDa (32)), suggesting that marapsin also largely eludes capture by  $\alpha_2$ -macroglobulin. Smaller amounts of activity appear at  $\sim 160$  and  $\sim 700$  kDa (in addition to QAR-4NA-hydrolyzing activity present in unspiked serum), possibly representing marapsin homo- or hetero-oligomers and a small fraction of  $\alpha_2$ -macroglobulin-bound marapsin, respectively. Overall, 53% of marapsin activity added to serum was recovered. Furthermore, reducing and non-reducing SDS-PAGE of recombinant mouse marapsin (Fig. 5C) does not reveal formation of disulfide-linked oligomers as a potential explanation of inhibitor resistance.

*Marapsin Lacks General Peptidase and Proteinase Activity*—Notwithstanding its ability to hydrolyze tetrapeptides with fluorogenic or chromogenic leaving groups, marapsin lacks general peptidase and proteinase activity. Among potential natural peptide targets tested and not hydrolyzed by mouse marapsin despite prolonged incubations are LL-37 cathelicidin, substance P, calcitonin gene-related peptide, neurotensin,





**FIGURE 4. Selectivity of a tetrapeptide substrate and inhibitor synthesized based on mouse marapsin subsite preferences.** *A* shows the relative preference of mouse marapsin, prostatin, and matriptase and bovine trypsin for custom tetrapeptide substrate YLNR-4NA versus nonspecific tripeptide substrate QAR-4NA (both substrates 0.5 mM). YLNR-4NA synthesized based on results of combinatorial peptide substrate profiling of marapsin (as shown in Fig. 3) was highly selective for mouse marapsin over prostatin and matriptase but less so in relation to trypsin in which subsite preferences are less pronounced. QAR-4NA was preferred by all proteases except marapsin. Specific activity for both substrates was much lower for marapsin and prostatin than for matriptase and trypsin. Error bars represent S.E. *B* shows the effect on QAR-4NA-hydrolyzing activity of preincubating marapsin, prostatin, matriptase, and trypsin with custom tetrapeptide-based inhibitor YLNR-chloromethyl ketone in various inhibitor/enzyme ( $[I]/[E]$ ) molar ratios. Residual activity after incubation with inhibitor is shown as a percentage of activity without inhibitor.

vasoactive intestinal peptide, and peptide YY, some of which contain sequences potentially favored based on results of combinatorial peptide screening. Proteins not cleaved include mouse epidermal growth factor receptor, mouse hepatocyte growth factor, human NK4-like protein (33), *Staphylococcus aureus* protein A, human keratin 14, mouse interleukin 17, and mouse pro-matrix metalloproteinase-9 (data not shown). We detected no gelatinolytic activity of mouse and human marapsins by gelatin zymography and no direct caseinolytic activity of the mouse enzyme (data not shown).

*Prss27<sup>-/-</sup> Mice Develop Normally and Are Fertile*—The deleted portion of marapsin gene affecting exons 2 through 6, including the entire catalytic domain, is shown schematically in Fig. 6A. Mice that were genetically modified to lack functional marapsin genes and raised in barrier facilities lacked overt phenotypic defects as *Prss27<sup>+/-</sup>* heterozygotes and *Prss27<sup>-/-</sup>* homozygotes. This was also true of mice backcrossed for 10 or more generations into a C57Bl/6 background. Fertility of *Prss27<sup>+/+</sup>*, *Prss27<sup>+/-</sup>*, and *Prss27<sup>-/-</sup>* mice was similar as was litter size, perinatal survival, and weight gain. Marapsin gene

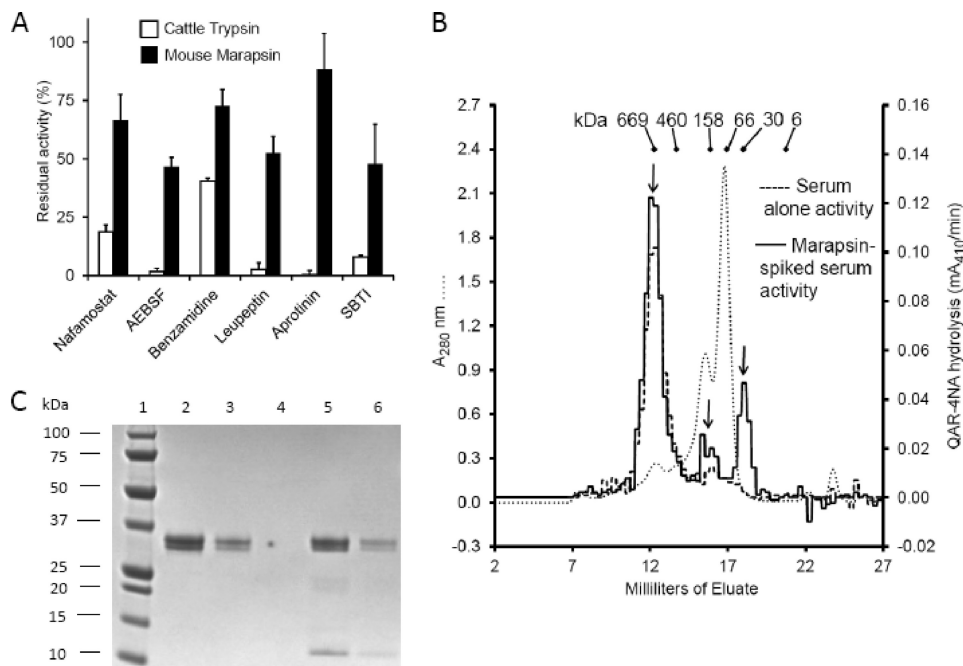
deletion in *Prss27<sup>-/-</sup>* mice was confirmed by PCR-based genotyping (Fig. 6) and by using antibodies raised against mouse marapsin to survey selected tissues for marapsin expression (Fig. 7).

*Epithelial Histology and Marapsin Expression in *Prss27<sup>+/+</sup>* and *Prss27<sup>-/-</sup>* Mice*—As shown in Fig. 7, the histological appearance of stratified squamous epithelium was similar in *Prss27<sup>+/+</sup>* and *Prss27<sup>-/-</sup>* esophagus, which was identified previously as a prominent site of marapsin expression (13), suggesting that marapsin is not essential for development, differentiation, or sloughing of esophageal epithelium. In tissue sections, incubation with anti-marapsin antibody detected immunoreactive material in stratified squamous epithelium derived from *Prss27<sup>+/+</sup>* but not *Prss27<sup>-/-</sup>* esophagus, confirming the specificity of this antibody for marapsin.

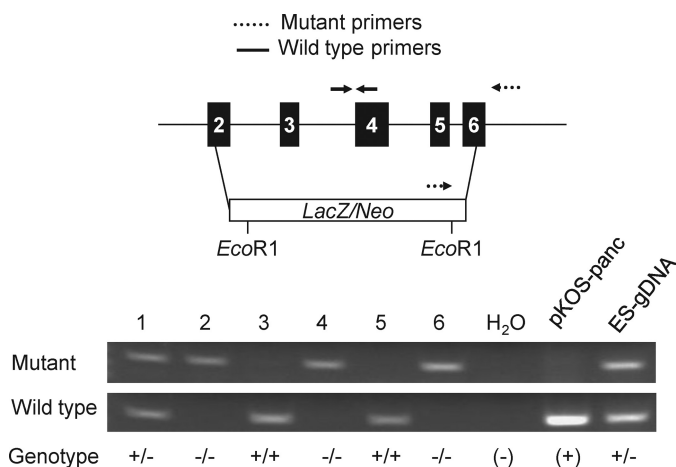
## DISCUSSION

The present work identifies mammals “caught in the act” of shedding the transmembrane domain of marapsin via nonsense mutation while maintaining high level conservation of the catalytic domain. Our observations that tail loss occurred independently in several clades and was preserved through millions of years of evolution suggest that tail loss is not deleterious and that positive selection may be at work. Theoretically, human marapsin and other marapsins lacking a C-terminal transmembrane hydrophobic peptide could have evolved from ancestors that were type I transmembrane proteins. Alternatively, transmembrane proteins could have evolved from soluble proteins by acquisition of hydrophobic tails. At least three lines of evidence in this study suggest that extant soluble marapsin evolved from transmembrane ancestors rather than the other way around. One piece of evidence is that most of the close relatives of marapsin are known or predicted to be type I transmembrane proteins. This includes prostatin, which is GPI-anchored in mice and humans and may have deeper origins in vertebrate evolution than marapsin, as revealed by supplemental Fig. S4 and by the analysis in Fig. 1. Second, tail loss in the three small clades depicted in Fig. 1 occurred comparatively recently in mammalian evolution and by distinct nonsense mutation events in each clade. These events are temporally and stochastically incompatible with an evolutionary sequence involving acquisition of a C-terminal anchor by tailless soluble proteins as a mechanism of generating the observed isoforms of marapsin. In specific reference to C-terminal anchor loss in a subset of primates, humans, chimpanzees, bonobos, and gorillas comprise subfamily Homininae within the family Hominidae (great apes), and humans, chimpanzees, and bonobos comprise tribe Hominini within subfamily Homininae. Thus, the appearance and preservation of the point mutation causing marapsin tail loss in Hominini but not in other great apes, monkeys, and other “lesser” primates follows the expected cladistic line of descent from early primates to humans (30). Finally, the evidence that human and other soluble marapsins were generated by recent mutational tail loss is supported by the presence of an open reading frame distal to the site of nonsense mutation that preserves the untranslated remnant of ancestral transmembrane sequence. In soluble mast cell tryptases, which are proposed to have evolved from type I transmembrane ancestral

## Evolution of Serine Proteases by Mutational Tail Loss

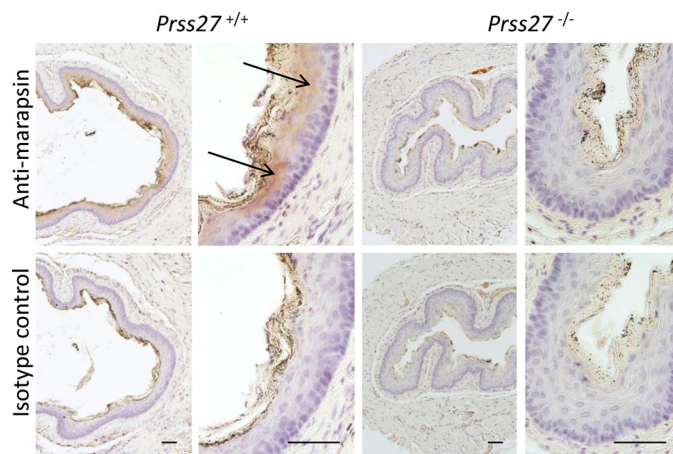


**FIGURE 5. Inhibitor susceptibility and resistance.** *A* compares the susceptibility of cattle trypsin (1 nM) and mouse marapsin (36 nM) to inhibitors of tryptic serine proteases. Enzymes were preincubated with nafamostat (300 nM), 4-(2-aminoethyl)benzenesulfonyl fluoride (AEBSF; 1 mM), benzamidine (1 mM), leupeptin (0.1 mM), aprotinin (0.1 mM), and soybean trypsin inhibitor (SBTI; 0.3 and 11  $\mu$ M for trypsin and marapsin, respectively). Results are expressed as the percentage of activity after preincubation with inhibitor relative to activity without inhibitor. Error bars represent S.E. *B* shows the activity and effective size of marapsin added to mouse serum and loaded on a gel filtration column calibrated with globular proteins, elution positions and molecular masses of which are shown in kDa. The *dotted line* shows absorbance at 280 nm ( $A_{280}$ ) of unspiked serum. QAR-4NA-hydrolyzing tryptic activity was measured in eluted fractions of native and marapsin-spiked serum (*dashed and solid lines*, respectively). *Arrows* indicate elution positions of potential marapsin activity, the majority of which appeared at  $\sim$ 30 kDa in the expected position of the monomer. *C* shows a Coomassie-stained gel after SDS-PAGE of marker proteins (*lane 1*) and recombinant mouse marapsin under native (*lanes 1–3*) and reducing (*lanes 5 and 6*) conditions. *Lane 4* contains buffer only. *Lanes 2 and 5* contain 3  $\mu$ g of marapsin; *lanes 3 and 6* contain 1  $\mu$ g. The  $<$ 10-kDa band appearing in reduced marapsin lanes is likely the disulfide-linked propeptide of activated marapsin.



**FIGURE 6. Generation and genotyping of marapsin-deficient *Prss27*<sup>-/-</sup> mice.** The *top* portion of the figure is a schematic representation of exons 2 through 6 of the mouse *Prss27* gene showing locations of the *LacZ/Neo* cassette deleting all exons encoding the pancreasin catalytic domain. Locations of wild type and mutant-specific PCR primers in relation to the genomic sequence are depicted. The *bottom* portion of the figure is an agarose gel showing results of genotyping of six mice (*lanes 1–6*, including the three genotypes: *Prss27*<sup>+/+</sup>, *Prss27*<sup>+/-</sup>, and *Prss27*<sup>-/-</sup>) with plasmid containing wild type pancreasin/marapsin gene (*pKOS-panc*) as a positive/negative control and contains targeted embryonic stem cell genomic DNA (*ES-gDNA*) as a positive control for heterozygote detection.

proteases much earlier in mammalian evolution (5, 7, 34), this trace of the ancestral tail is no longer evident. This evolutionary mechanism as applied to marapsins, mast cell tryptases, and potentially to other proteases can be relevant for type I but not



**FIGURE 7. Esophageal histology and marapsin immunoreactivity in *Prss27*<sup>+/+</sup> and *Prss27*<sup>-/-</sup> mice.** The *top panels* are low and higher power photomicrographs of esophageal tissue sections incubated with anti-mouse marapsin antibody. *Arrows* indicate immunoreactive stratified squamous epithelial cells lining the esophageal lumen of *Prss27*<sup>+/+</sup> mice. The granular *brown* staining in the most superficial layer represents luminal debris and non-nucleated, keratinized cell-derived material. *Prss27*<sup>-/-</sup>-derived sections lacked specific staining. The *bottom panels* show the lack of immunoreactivity in serial sections of *Prss27*<sup>+/+</sup> and *Prss27*<sup>-/-</sup> esophageal squamous epithelium incubated with non-immune (isotype control) antibody. Sections were counterstained with hematoxylin. *Scale bars*, 50  $\mu$ m.

for type II proteases because a nonsense mutation in the transmembrane segment codon of a type II serine proteases would prevent the catalytic domain itself from being translated.

These observations reveal that although marapsins in at least three groups of mammals lack a C-terminal transmembrane

segment the transmembrane version is the typical mammalian form and was also the likely form of the enzyme in ancestral, premammalian vertebrates. Given the major variation in the length of marapsin C-terminal sequence, it is perhaps surprising that the catalytic domain of marapsin is strongly conserved in mammals and other vertebrates. The preservation of catalytic triad residues His<sup>57</sup>, Asp<sup>102</sup>, and Ser<sup>195</sup> (using standard chymotrypsinogen numbering) and trypsin-like specificity triad residues Asp<sup>189</sup>, Gly<sup>216</sup>, and Gly<sup>226</sup> indicates that proteolytic activity is likely to be important to marapsin function. The demonstration of similar peptidolytic activity in recombinant soluble versions of human and mouse marapsins further predicts that catalytic competence is a feature of the catalytic domain whether it originates as a transmembrane protein or as a protein secreted in a soluble form. Notwithstanding these considerations, mouse and human marapsins both are enzymatically weak in comparison with trypsin, matriptase, and mast cell tryptases and are similar to prostates in this regard (31) although with divergent peptide subsite preferences as revealed by the combinatorial peptide library screening results in Fig. 2. Although it is possible that marapsins cleave a yet-to-be-identified protein target with high efficiency, our studies to date identify little or no proteinase activity for mouse and human marapsins. Therefore, it is possible that marapsin serves a regulatory function that does not depend on peptidase activity. We also note that the last four residues of observed consensus propeptide sequence CGRPRMLNR are similar to the custom YLNR-4NA peptide synthesized based on results of combinatorial screening of mouse marapsin and determined to be cleaved preferentially by marapsin. This suggests the possibility of direct marapsin autoactivation, which is consistent with the observed spontaneous activation of recombinant human and mouse marapsins during expression and purification. Alternatively, promarapsin could be activated in a multi-step process by initial hydrolysis at Arg<sup>-5</sup> followed by dipeptidyl peptidase I- or cathepsin-mediated removal of the remnant peptide as may occur with some mast cell tryptases (35–37) and mastins (38).

The general reluctance of mouse and human marapsins to hydrolyze proteins and peptides at tryptic sites even at sites that would seem to be favored based on identified preferences in tetrapeptide substrates suggests that the active site may be blocked in a manner that hinders access to substrates (and inhibitors). The overall weak activity toward tetrapeptide 4NA substrates also is consistent with a partially blocked or “collapsed” active site potentially due to the absence of a yet-to-be-identified allosteric activator or cofactor (39) or to a requirement for active site conformation change induced by binding of a highly specific substrate as occurs for complement factor D (40). A potentially similar blockade was identified in structures derived from crystallized versions of prostates (41, 42), which as noted is a type I transmembrane tryptic serine peptidase related to marapsin but with differing substrate preferences and patterns of expression. The active site blockade may be more extreme in marapsin based on higher level resistance to inhibitors. Indeed, although the present work identifies low molecular weight inhibitors, marapsin appears to resist larger (*i.e.* proteinaceous) inhibitors, including broad spectrum tryptic

serine proteases inhibitors like aprotinin, serum serpins, and  $\alpha_2$ -macroglobulin. A partially blocked or collapsed active site can explain inhibitor resistance as well as weak general peptidase activity. Although some proteinases, notably mast cell tryptases (25) and mastins (38), achieve a measure of protection from larger inhibitors by self-associating into active site-shielding oligomers, the gel filtration behavior of mouse marapsin (Fig. 5) suggests that marapsin remains uninhibited as a monomer among many potential antipeptidases in serum. Furthermore, mouse and human marapsins do not seem to form disulfide-linked dimers or higher order oligomers. Therefore, we propose that the most likely explanation of inhibitor resistance is a prostates-like, partially blocked active site in the monomeric catalytic domain.

It should be stressed that our phylogenetic analysis shows that the closest relatives of marapsin are not prostates but prosemis, which appear to be soluble (*i.e.* not transmembrane) proteases in mammals. However, both marapsins and prosemis appear to have evolved from type I transmembrane proteases. On this basis, we propose that an early (possibly pre-mammalian) ancestral prostates underwent mutational loss of its C-terminal hydrophobic tail as some marapsins have done recently, leaving no extant membrane-anchored prosemis among identified mammalian proteases. Although marapsin and prostates share similarities as type I transmembrane tryptic P1 Arg-preferring peptidases with partially protected active sites, there are key contrasts, including as already noted functional differences in peptide substrate preferences, inhibitor susceptibility, and tissue patterns of expression. Perhaps the most compelling contrast in regard to function is the respective phenotypes of prostates-deficient (*Prss8*<sup>-/-</sup>) and marapsin-deficient (*Prss27*<sup>-/-</sup>) mice. *Prss8*<sup>-/-</sup> mice die during embryogenesis, and mice with skin-selective deficits of prostates die in the early postnatal period (23). *Prss27*<sup>-/-</sup> mice as shown in the present work are viable and fertile without overt defects when raised in specific pathogen-free barrier conditions, arguing against a non-redundant, critical role in fetal or early postnatal development. Nonetheless, strong conservation of the marapsin gene in mammals and other vertebrates suggests that roles await discovery. In particular, the availability of the marapsin-deficient mice reported here allows explorations of the importance of marapsin in epithelial responses to infection, injury, and other environmental challenges.

## REFERENCES

1. Puente, X. S., Sánchez, L. M., Overall, C. M., and López-Otín, C. (2003) Human and mouse proteases: a comparative genomic approach. *Nat. Rev. Genet.* **4**, 544–558
2. Netzel-Arnett, S., Hooper, J. D., Szabo, R., Madison, E. L., Quigley, J. P., Bugge, T. H., and Antalis, T. M. (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
3. Wong, G. W., and Stevens, R. L. (2005) Identification of a subgroup of glycosylphosphatidylinositol-anchored tryptases. *Biochem. Biophys. Res. Commun.* **336**, 579–584
4. Wong, G. W., Tang, Y., Feyfant, E., Sali, A., Li, L., Li, Y., Huang, C., Friend, D. S., Krilis, S. A., and Stevens, R. L. (1999) Identification of a new member of the tryptase family of mouse and human mast cell proteases which possesses a novel COOH-terminal hydrophobic extension. *J. Biol. Chem.* **274**, 30784–30793

5. Caughey, G. H., Raymond, W. W., Blount, J. L., Hau, L. W., Pallaoro, M., Wolters, P. J., and Verghese, G. M. (2000) Characterization of human  $\gamma$ -tryptases, novel members of the chromosome 16p mast cell tryptase and prostatic gene families. *J. Immunol.* **164**, 6566–6575
6. Wong, G. W., Yasuda, S., Madhusudhan, M. S., Li, L., Yang, Y., Krilis, S. A., Sali, A., and Stevens, R. L. (2001) Human tryptase epsilon (PRSS22), a new member of the chromosome 16p13.3 family of human serine proteases expressed in airway epithelial cells. *J. Biol. Chem.* **276**, 49169–49182
7. Trivedi, N. N., Tong, Q., Raman, K., Bhagwandin, V. J., and Caughey, G. H. (2007) Mast cell  $\alpha$  and  $\beta$  tryptases changed rapidly during primate speciation and evolved from  $\gamma$ -like transmembrane peptidases in ancestral vertebrates. *J. Immunol.* **179**, 6072–6079
8. Trivedi, N. N., Raymond, W. W., and Caughey, G. H. (2008) Chimerism, point mutation, and truncation dramatically transformed mast cell  $\delta$ -tryptases during primate evolution. *J. Allergy Clin. Immunol.* **121**, 1262–1268
9. Reimer, J. M., Samollow, P. B., and Hellman, L. (2010) High degree of conservation of the multigene tryptase locus over the past 150–200 million years of mammalian evolution. *Immunogenetics* **62**, 369–382
10. Bhagwandin, V. J., Hau, L. W., Mallen-St Clair, J., Wolters, P. J., and Caughey, G. H. (2003) Structure and activity of human pancreasin, a novel tryptic serine peptidase expressed primarily by the pancreas. *J. Biol. Chem.* **278**, 3363–3371
11. Reimer, J. M., Enoksson, M., Samollow, P. B., and Hellman, L. (2008) Extended substrate specificity of opossum chymase—implications for the origin of mast cell chymases. *Mol. Immunol.* **45**, 2116–2125
12. Wong, G. W., Yasuda, S., Morokawa, N., Li, L., and Stevens, R. L. (2004) Mouse chromosome 17A3.3 contains 13 genes that encode functional tryptic-like serine proteases with distinct tissue and cell expression patterns. *J. Biol. Chem.* **279**, 2438–2452
13. Li, W., Danilenko, D. M., Bunting, S., Ganesan, R., Sa, S., Ferrando, R., Wu, T. D., Kolumam, G. A., Ouyang, W., and Kirchhofer, D. (2009) The serine protease marapsin is expressed in stratified squamous epithelia and is up-regulated in the hyperproliferative epidermis of psoriasis and regenerating wounds. *J. Biol. Chem.* **284**, 218–228
14. Adachi, W., Ulanovsky, H., Li, Y., Norman, B., Davis, J., and Piatigorsky, J. (2006) Serial analysis of gene expression (SAGE) in the rat limbal and central corneal epithelium. *Invest. Ophthalmol. Vis. Sci.* **47**, 3801–3810
15. Cserző, M., Eisenhaber, F., Eisenhaber, B., and Simon, I. (2002) On filtering false positive transmembrane protein predictions. *Protein Eng.* **15**, 745–752
16. Eisenhaber, B., Bork, P., and Eisenhaber, F. (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J. Mol. Biol.* **292**, 741–758
17. Verghese, G. M., Gutknecht, M. F., and Caughey, G. H. (2006) Prostaticin regulates epithelial monolayer function: cell-specific Gpld1-mediated secretion and functional role for the GPI anchor. *Am. J. Physiol. Cell Physiol.* **291**, C1258–C1270
18. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., and Craik, C. S. (2000) Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7754–7759
19. Raymond, W. W., Ruggles, S. W., Craik, C. S., and Caughey, G. H. (2003) Albumin is a substrate of human chymase: prediction by combinatorial peptide screening and development of a selective inhibitor based on the albumin cleavage site. *J. Biol. Chem.* **278**, 34517–34524
20. Raymond, W. W., Trivedi, N. N., Makarova, A., Ray, M., Craik, C. S., and Caughey, G. H. (2010) How immune peptidases change specificity: cathepsin G gained tryptic function but lost efficiency during primate evolution. *J. Immunol.* **185**, 5360–5368
21. Zambrowicz, B. P., Friedrich, G. A., Buxton, E. C., Lilleberg, S. L., Person, C., and Sands, A. T. (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611
22. Verghese, G. M., Tong, Z. Y., Bhagwandin, V., and Caughey, G. H. (2004) Mouse prostaticin gene structure, promoter analysis, and restricted expression in lung and kidney. *Am. J. Respir. Cell Mol. Biol.* **30**, 519–529
23. Leyvraz, C., Charles, R. P., Rubera, I., Guitard, M., Rotman, S., Breiden, B., Sandhoff, K., and Hummler, E. (2005) The epidermal barrier function is dependent on the serine protease CAP1/Prss8. *J. Cell Biol.* **170**, 487–496
24. Planès, C., Randrianarison, N. H., Charles, R. P., Frateschi, S., Cluzeaud, F., Vuagniaux, G., Soler, P., Clerici, C., Rossier, B. C., and Hummler, E. (2010) ENaC-mediated alveolar fluid clearance and lung fluid balance depend on the channel-activating protease 1. *EMBO Mol. Med.* **2**, 26–37
25. Pereira, P. J., Bergner, A., Macedo-Ribeiro, S., Huber, R., Matschiner, G., Fritz, H., Sommerhoff, C. P., and Bode, W. (1998) Human  $\beta$ -tryptase is a ring-like tetramer with active sites facing a central pore. *Nature* **392**, 306–311
26. Perona, J. J., and Craik, C. S. (1997) Evolutionary divergence of substrate specificity within the chymotrypsin-like serine protease fold. *J. Biol. Chem.* **272**, 29987–29990
27. Wouters, M. A., Liu, K., Riek, P., and Husain, A. (2003) A despecialization step underlying evolution of a family of serine proteases. *Mol. Cell* **12**, 343–354
28. Caughey, G. H., Beauchamp, J., Schlatter, D., Raymond, W. W., Trivedi, N. N., Banner, D., Mauser, H., and Fingerle, J. (2008) Guinea pig chymase is leucine-specific: a novel example of functional plasticity in the chymase/granzyme family of serine peptidases. *J. Biol. Chem.* **283**, 13943–13951
29. Chen, L. M., Skinner, M. L., Kauffman, S. W., Chao, J., Chao, L., Thaler, C. D., and Chai, K. X. (2001) Prostaticin is a glycosylphosphatidylinositol-anchored active serine protease. *J. Biol. Chem.* **276**, 21434–21442
30. Raaum, R. L., Sterner, K. N., Noviello, C. M., Stewart, C. B., and Disotell, T. R. (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J. Hum. Evol.* **48**, 237–257
31. Nimishakavi, S., Besprozvannaya, M., Raymond, W. W., Craik, C. S., Gruenert, D. C., and Caughey, G. H. (2012) Activity and inhibition of prostaticin and matriptase on apical and basolateral surfaces of human airway epithelial cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **303**, L97–L106
32. Raymond, W. W., Su, S., Makarova, A., Wilson, T. M., Carter, M. C., Metcalfe, D. D., and Caughey, G. H. (2009)  $\alpha$ 2-Macroglobulin capture allows detection of mast cell chymase in serum and creates a reservoir of angiotensin II-generating activity. *J. Immunol.* **182**, 5770–5777
33. Raymond, W. W., Cruz, A. C., and Caughey, G. H. (2006) Mast cell and neutrophil peptidases attack an inactivation segment in hepatocyte growth factor to generate NK4-like antagonists. *J. Biol. Chem.* **281**, 1489–1494
34. Caughey, G. H. (2007) Mast cell tryptases and chymases in inflammation and host defense. *Immunol. Rev.* **217**, 141–154
35. Wolters, P. J., Pham, C. T., Muilenburg, D. J., Ley, T. J., and Caughey, G. H. (2001) Dipeptidyl peptidase I is essential for activation of mast cell chymases, but not tryptases, in mice. *J. Biol. Chem.* **276**, 18551–18556
36. Le, Q. T., Gomez, G., Zhao, W., Hu, J., Xia, H. Z., Fukuoka, Y., Katunuma, N., and Schwartz, L. B. (2011) Processing of human protryptase in mast cells involves cathepsins L, B, and C. *J. Immunol.* **187**, 1912–1918
37. Le, Q. T., Min, H. K., Xia, H. Z., Fukuoka, Y., Katunuma, N., and Schwartz, L. B. (2011) Promiscuous processing of human  $\alpha/\beta$ -protryptases by cathepsins L, B, and C. *J. Immunol.* **186**, 7136–7143
38. Raymond, W. W., Sommerhoff, C. P., and Caughey, G. H. (2005) Mastin is a gelatinolytic mast cell peptidase resembling a mini-proteasome. *Arch. Biochem. Biophys.* **435**, 311–322
39. Gohara, D. W., and Di Cera, E. (2011) Allosterism in trypsin-like proteases suggests new therapeutic strategies. *Trends Biotechnol.* **29**, 577–585
40. Forneris, F., Ricklin, D., Wu, J., Tzekou, A., Wallace, R. S., Lambris, J. D., and Gros, P. (2010) Structures of C3b in complex with factors B and D give insight into complement convertase formation. *Science* **330**, 1816–1820
41. Rickert, K. W., Kelley, P., Byrne, N. J., Diehl, R. E., Hall, D. L., Montalvo, A. M., Reid, J. C., Shipman, J. M., Thomas, B. W., Munshi, S. K., Darke, P. L., and Su, H. P. (2008) Structure of human prostaticin, a target for the regulation of hypertension. *J. Biol. Chem.* **283**, 34864–34872
42. Spraggon, G., Hornsby, M., Shipway, A., Tully, D. C., Bursulaya, B., Danahey, H., Harris, J. L., and Lesley, S. A. (2009) Active site conformational changes of prostaticin provide a new mechanism of protease regulation by divalent cations. *Protein Sci.* **18**, 1081–1094