

Running title: Ovastacin – Substrate profiling

Substrate profiling of the metalloproteinase ovastacin – Implications for its physiological function in mammalian fertilization

Matthias Felten^a, Ute Distler^b, Nele v. Wiegen^c, Mateusz Łacki^b, Christian Behl^c, Stefan Tenzer^b, Walter Stöcker^a & Hagen Körschgen^{c*}

^aInstitute of Molecular Physiology, Cell and Matrix Biology, Johannes Gutenberg-University Mainz; Johann-Joachim-Becher-Weg 7; D-55128 Mainz (Germany).

^bInstitute for Immunology, University Medical Center of the Johannes Gutenberg-University Mainz, Langenbeckstr. 1; D-55131 Mainz (Germany).

^cThe Autophagy Lab, Institute for Pathobiochemistry, University Medical Center of the Johannes Gutenberg-University Mainz, Duesbergweg 6; D-55099 Mainz (Germany).

* Corresponding author

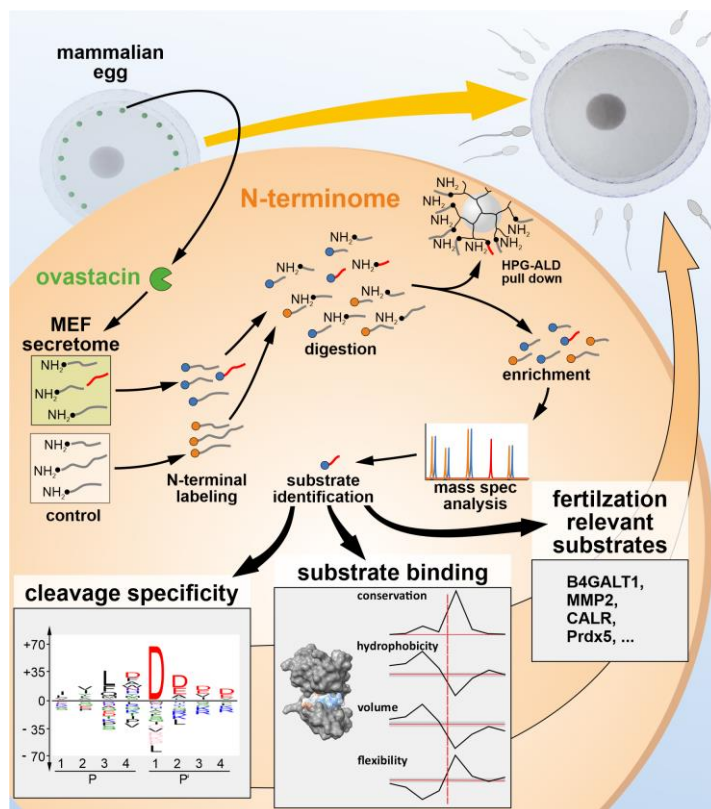
e-mail: hagen.koerschgen@uni-mainz.de

phone: (+49) 6131 3924481

This PDF file includes: Main Text, Figures 1 to 3 and Tables 1 & 2

Keywords: ovastacin, metalloproteinases, astacins, N-TAILS, cleavage specificity, fertilization

Graphical Abstract



Abstract

The metalloproteinase ovastacin is released by the mammalian egg upon fertilization and cleaves a distinct peptide bond in zona pellucida protein 2, a component of the enveloping extracellular matrix. This limited proteolysis causes zona pellucida hardening, abolishes sperm binding and thereby regulates fertility. Accordingly, this process is tightly controlled by the plasma protein fetuin-B, an endogenous competitive inhibitor. At present, little is known about how the cleavage characteristics of ovastacin differ from closely related proteases. Physiological implications of ovastacin beyond ZP2 cleavage are still obscure. In this study, we employed N-terminal amine isotopic labeling of substrates (N-TAILS) contained in the secretome of mouse embryonic fibroblasts to elucidate the substrate specificity and the precise cleavage site specificity. Furthermore, we were able to unravel the physicochemical properties governing enzyme-substrate interactions. Eventually, we identified several potential physiological substrates with significance for mammalian fertilization. These data suggest that ovastacin might regulate sperm-oocyte interaction and fertility beyond zona pellucida hardening.

Main Text

Introduction

Several germ cell specific proteolytic enzymes have been reported to exert essential functions during egg-sperm interaction in mammalian fertilization. Among others, these include ovastacin, acrosin, testisin (TESP5, PRSS21), ADAM1b (i.e. fertilin α) and plasminogen activators (1–7), revealing a complex, coordinated and antagonist-controlled proteolytic system reminiscent of those controlling blood coagulation and fibrinolysis. However, only for two of these proteinases, namely acrosin and ovastacin, fertilization relevant cleavage sites in distinct substrates have been identified so far (1, 8–10). Considering 50 million infertile couples estimated worldwide and a rate of about 30% idiopathic infertility, 15 million couples are affected by involuntary childlessness (11, 12). Hence, there is considerable demand for further elucidation of the molecular mechanisms underlying fertilization.

Ovastacin (encoded by the gene *ASTL*), a member of the astacin family of metalloproteinases, is one of the enzymes that regulate fertilization at the level of the zona pellucida (ZP), the egg-enveloping extracellular matrix. As a typical astacin-like proteinase, ovastacin comprises a catalytic domain made up of 198 amino acid residues containing the extended zinc binding motif (H¹⁸²ExxHxxGxxH¹⁹²) and the strictly conserved 1,4- β -turn, called the Met-turn (M²³⁶), typical for metzincin proteinases (13). Additionally, it contains a unique cortical granule localization motif (D⁵²KDIPAIN⁵⁹) within the N-terminal propeptide, upstream of the conserved zinc-binding aspartate ensuring latency, and a C-terminal domain of 150 residues of unknown function (13, 14).

Upon plasmogamy and oocyte activation, ovastacin is released into the perivitelline space during the cortical reaction and cleaves a distinct peptide bond (SRLA↓D¹⁶⁸ENQ) within zona pellucida protein 2 (ZP2), thereby converting it into ZP2f (1). Thus, sperm reception is disrupted and the ZP "hardens", which prevents sperm binding to the ZP and definitely abrogates its penetration for further sperm (1, 15). Even though ZP2 has so far been the only proven physiological substrate, cleavage of additional substrates in the complex secretome of fertilization is likely.

Phenotypically, ovastacin-deficient mice (*ASTL*^{-/-}) are inconspicuous except for a significant reduction in fecundity (16). Here, the absence of ZP hardening results in a soft ZP, which fails to mechanically protect the embryo until implantation. During *in vitro* fertilization (IVF).

ASTL^{-/-}-oocytes and embryos are covered by a vast number of bound sperm and many sperm in the perivitelline space (16, 17). However, this does not increase the rate of polyspermy. This further supports the concept that the decisive block against polyspermy is located at the oolemma level (17–19).

The absence of an endogenous ovastacin inhibitor, the plasma protein fetuin-B, leads to female infertility (20). Here small quantities of prematurely released ovastacin are sufficient to prevent fertilization. Hence, strict regulation of this proteinase is essential to maintain fertility. Conversely, supplementation of fetuin-B during IVF significantly improves the success of fertilization by effectively blocking prematurely released intracellularly activated ovastacin (16, 20, 21).

Recently, we were able to elucidate the precise mechanism of this inhibition (22–24). In mammals, only three members of the astacin family of proteinases, ovastacin and the closely related meprins (α & β) are inhibited by fetuin-B (22, 25). The underlying molecular mechanism differs greatly from the classical cystatin/papain interaction (26). This inhibition by the so called ‘raised elephant trunk’ mechanism is not due to the interaction of one or both cystatin domains of fetuin-B with the active site of ovastacin but is rather based on the warhead containing linker (C¹⁵⁴PDC¹⁵⁷) between the two cystatin-like domains and a hairpin loop in the second cystatin-like domain.

In this study, we utilized N-terminal amine isotopic labelling of substrates (N-TAILS) (27) on mouse embryonic fibroblast (MEF) secretome to analyze the cleavage specificity of ovastacin. We found unique characteristics of ovastacin, differing from other astacins. Furthermore, we identified a variety of potential physiological substrates including cleavage sites, which might extend the function of ovastacin during fertilization.

Materials and Methods

Cell culture and secretome preparation

We cultured mouse embryonic fibroblasts (MEF) in DMEM (Dulbecco's Modified Eagle Medium (Sigma-Aldrich) with 10% FBS and 100 U/ml Penicillin / Streptomycin (37°C, 5% CO₂) in adherent culture and passaged three times per week (detaching of adherent cells with 0.05% trypsin/EDTA solution).

For preparation of conditioned cell media, MEF cells were washed three times with PBS buffer (Dulbecco's Phosphate Buffered Saline - Sigma-Aldrich) and incubated for additional 24 h in DMEM/High Glucose (51449C, Sigma-Aldrich) with 4 mM Alanine-glutamine (G8541, Sigma-Aldrich) and 100 U/ml Penicillin / Streptomycin). Subsequently, we collected the supernatant, treated with protease inhibitors (cOmplete™ EDTA-free Protease Inhibitor Cocktail (Merck, 04693132001) or Pefabloc® SC (CarlRoth, A154.1)), and centrifuged for 3 min at 180 x g at 4°C. The supernatant was concentrated (Vivaspin 5 kDa MWCO) to 1 – 2 mg/ml and stored on ice until further use.

Activation of pro-ovastacin

Heterologously expressed murine ovastacin was purified and activated with human plasmin (Haematologic Technologies) as described (Karmilin *et al.*, 2019; Kuske *et al.*, 2021). Exemplarily, supplemental Figure 1 illustrates ovastacin purification and activation. Based on the short half-life of the ovastacin activity, the plasmin used for activation was not removed but inhibited using cOmplete™ EDTA-free Protease Inhibitor Cocktail (Merck, 04693132001) or Pefabloc® SC (CarlRoth, A154.1). The inhibition was checked via a fluorogenic assay (as exemplified in supplemental Figure 2).

Sample preparation for Terminal Isotopic Labeling of Substrates (TAILS) analysis

The secretome of MEF cells was incubated twice with active ovastacin at a mass ratio of 100:1 for 2 h, at 37 °C or with the equal volume of activation solution without ovastacin, respectively. Subsequently, samples were reduced in 5 mM dithiothreitol (DTT), at room temperature (RT). After incubation (70 °C, 60 min) the samples were alkylated with iodoacetamide (IAA) (final concentration 15 mM, 30 min, RT, light-protected). Thereafter, proteins were precipitated with trichloroacetic acid (TCA) (28) and resuspended (200 mM HEPES, pH 7.5) to a final concentration of approx. 2 mg/ml. N-termini were labeled with light

and heavy formaldehyde ($^{12}\text{CH}_2\text{O}$ in H_2O , and $^{13}\text{CD}_2\text{O}$ in D_2O (252549 and 596388, Sigma-Aldrich)), respectively. Sodium-cyanoborohydride was added to a final concentration of 40 mM, adjusted to pH 7 followed by overnight (ON) incubation at 37 °C(29). Excess formaldehyde was captured by adjusting the sample to 50 mM Tris/HCl pH 7.0 and incubation for 4 h (37 °C). Proteins were precipitated with 9x volume acetone and 1x volume methanol (each -20 °C) ON at -80 °C and centrifuged (4500 x g for 60 min at 4 °C). The pellet was washed four times with 1 ml methanol following centrifugation (4500 x g for 15 min at 4 °C) and dissolved in 200 mM HEPES buffer (pH 8). Proteolysis into peptides was accomplished using 0.5% (w/w) Trypsin (sequencing grade; Worthington Biochemical) (37 °C, overnight). After desalting via C18 Sep-Pak columns (Waters), we enriched the labeled protein N-termini by negative selection using a polymer for enrichment (hyperbranched polyglycerol-aldehyde, HPG-ALD (University of British Columbia, Vancouver)). Peptides were adjusted to 200 mM HEPES (pH 7) and incubated with a ratio of 1:5 (w/w) of polymer and 100 mM sodium cyanoborohydride (ON at 37 °C, 40 rpm). Free aldehyde groups were saturated using 100 mM glycine pH 7 for 60 min at RT. Subsequently, the N-terminal blocked peptides were isolated using centrifugal filtration (Vivaspin 500, 10 kDa MWCO) and two washing steps (1x 300 μl 100 mM ammonium bicarbonate, 1x 300 μl 2 M guanidine hydrochloride). The samples were desalted via C18 columns and fractionated via strong cation exchange chromatography (Stagetips)(30). Afterwards, the samples were desalted again. Peptide concentrations for mass spectrometric measurement were determined by a NanoDrop ND-1000 spectrophotometer (PeqLab). Protein or peptide concentration in the course preparation was checked by Bradford- or bicinchoninic acid-assays (31, 32).

TAILS analysis by mass spectrometry

Mass spectrometric analysis was performed using LC-MS (liquid chromatography/mass spectrometry) on a Synapt G2-S HDMS (Waters, MA, USA) in positive ion mode. Chromatographic separation of peptides was conducted on a nanoAcquity UPLC system (Waters). Peptides were separated using a C18 reversed-phase analytical separation column (HSS-T3, 75 μm x 250 mm, 1.8 μm particle size) (Waters). Samples were loaded directly onto the column (2.5 μL injection volume). Peptides were separated using a gradient of mobile phases A (water containing 0.1% (v/v) formic acid and 3% (v/v) DMSO) and B (acetonitrile containing 0.1% (v/v) formic acid and 3% (v/v) DMSO). The proportion of phase B increased from 5% (v/v) to 40% (v/v) over 90 min (300 nL/min). Afterwards, the column was re-

equilibrated with phase B (90% (v/v), 600 nL/min, 5 min). Total measurement time was 120 min. Data were acquired using HD-DDA (33). [Glu¹]-fibrinopeptide was used as lock mass at 100 fmol/μL and sampled every 30 s into the mass spectrometer via the reference sprayer of the NanoLockSpray source. Processing of raw data and database searching was performed using PEAKS (version 8.5; <https://www.bioinform.com/>) (34–36). We reviewed liquid chromatography plots individually to exclude false positive results. Single signals (FDR = 0.01) identified by PEAKS were checked for the presence of a double signal caused by 6 Da mass difference due to dimethylation, which was considered a duplicate and excluded from the data set. Peptides with naturally acetylated N-termini as well as peptides with unlabeled N-termini or double labeled with light and heavy formaldehyde were also excluded. All peptides identified by PEAKS that had N-terminal labeling by heavy formaldehyde and no analogue with N-terminal labeling by light formaldehyde were manually selected. Samples were prepared as two independent biological replicates and each analyzed in technical triplicates. The final data set only included peptides that could be assigned to a single UNIPROT accession number from *Mus musculus*. Amino acid residues N-terminally of the cleaved peptide bond (i.e. “nonprime side”) were determined by bioinformatic alignment of the sequences with the UNIPROT database. Cleavage sites less than 10 residues distant from N- or C-terminus were considered.

Comparing all identified cleavage sites with aspartate in position P1' (cleavage preference of astacin proteases against the full data set revealed a negative correlation with lysine and arginine in position P1 (tryptic cleavage specificity). Therefore, in order to exclude peptides generated e.g. by not entirely inactivated plasmin carried over from ovastacin activation, the data sets from both biological replicates were further subjected to cluster analysis (GibbsCluster 2.0) (37, 38). This is exemplified for one of the replicas in supplementary figures 3 and 4. The identified clusters from both biological replicates, with the typical acidic residues on the prime side (39), were combined for the further analyses. The dataset was limited to proteins within the physiological context of the extracellular space by analysis using the DAVID Bioinformatics database (40, 41).

Identification of the cleavage specificity and physicochemical properties of ovastacin

Determination of cleavage preferences derived from the substrate cleavage sites, their representation as sequence logos, and statistical variations in the abundances of residues was performed using iceLogo (42). In order to equally compare cleavage preferences of other

proteases based on the cleavage sites annotated in the MEROPS database (10), only cleavage sites with all positions from P4 to P4' identified were used. The "subLogo" function was used to check for cooperativity between residues in different positions. To calculate the degree of conservation of residues in each position the "Conservation Line" function with a BLOSUM62-substitution matrix (43) was used.

For determination of parameters needed for evaluation of physicochemical properties the "Aa Parameter" function was used. Underlying parameters are listed in the AAindex1 database (<http://www.genome.jp/aaindex/>). The "Transfer Energy, organic solvent/water" template was used to determine hydrophobicity (44), the "Residue Volume" template to determine residue volume (45), and the "Average flexibility indices" template to determine residue flexibility (46).

In silico structure prediction

Models were visualized using Chimera 1.13.1 software. Homology models were calculated using Modeller (<https://salilab.org/modeller/>) (47, 48) via the Chimera interface. Alignments necessary for generating homology models, were also performed with Chimera (via BLOSUM-62 matrix (43), gap opening penalty 12 / gap extension penalty 1). For model creation with more than one template, the structure-specific secondary structure score was employed. The homology model with most favorable (negative) zDOPE (Discrete Optimized Protein Energy) value was selected.

The stereochemical quality of the models generated was evaluated using the PROCHECK software (www.servicesn.mbi.ucla.edu/PROCHECK) (49, 50). Interactions between ovastacin and identified substrates were simulated *via* the ClusPro software (<https://cluspro.org>) (51–53) using the generated models. From all calculated values, those 1000 with the most favorable (lowest) energy were selected and searched for clusters in the configurations.

Results

TAILS Analysis

We performed two N-TAILS analyses on native MEF secretome with heterologously expressed murine ovastacin to identify potential physiological relevant substrates of ovastacin, as well as its cleavage preference. Together, a total number of 855 unique cleavage sites were identified in 489 proteins. After restriction to the physiological context, the extracellular space, by database mining, we identified a dataset of 353 unique cleavage sites in 163 proteins (Supplemental Table S1 and Supplemental Table S2).

Ovastacin features the typical acidic prime side specificity of the astacin metalloproteinase family

Results are presented according to the concept by Schechter and Berger (54). Peptide residues N-terminal to the cleavage site are indicated as nonprime (P) and C-terminal as prime (P'). The summary of the cleavage profile of the 855 unique cleavage sites exhibits very high similarity to the only physiologically verified cleavage site SRLA↓D¹⁶⁸ENQ in ZP2 (1)(Figure 1). Ovastacin significantly prefers negatively charged residues in positions P1 to P6', especially for aspartate in positions P1' (+70.1%, rel. to natural occurrence) and P2' (+16.6%). No strong preference is evident for position P1. Nevertheless, aspartate (+7.6%), phenylalanine (+4.8%), alanine (+4.6%), tryptophan (+4.0%), asparagine (+3.7%) and histidine (+3.5%) are found predominantly. In position P2, leucine (+19.8%), phenylalanine (+8.2%) and glutamine (+5.4%) exhibit the highest abundance. Hydrophobic residues are preferred in the range from P4 to P1 with aromatic residues at higher frequency in positions from P3 to P1. On the prime side, on the other hand, acidic residues, especially aspartate, predominate in the range from P1 to P6'. This is highlighted by the graphic iceLogo representation (figure 1 D). Ovastacin cleavage sites in substrates reveal a strong cleavage preference for negatively charged residues on the prime side, especially for aspartate in position P1'. In general, aspartate is clearly favored over glutamate in all positions of the prime side. On the nonprime side, leucine, phenylalanine and methionine are particularly preferred in position P2, and tyrosine, valine, isoleucine and phenylalanine in position P3.

Physicochemical properties of ovastacin compared to other astacin proteinases

The cleavage preference determined for ovastacin (figure 1) shows high agreement with the cleavage characteristics of other astacins (BMP1, meprin α , meprin β , and LAST MAM) partly

due to the preference for the negatively charged aspartate residue in position P1' (39). To identify ovastacin-specific characteristics, we analyzed in detail the physicochemical properties of cleavage sites, i.e. electrostatics, hydrophilicity/hydrophobicity, volume, flexibility and degree of conservation of distinct residues flanking the cleavage site. All signatures of the individual astacin members (supplementary figure 3) feature the highest degree of conservation in position P1' of the cleavage site, and to lesser extent in the positions P2 to P2' for meprin α only. The hydrophobicity of the residues reaches its minimum in positions P1 and P1', respectively, and its maximum in positions P3 and P2. The side chains with the largest volumes tend to be in positions P3 to P2, those with the smallest in the range P1 to P2'. The flexibility of the residues is lowest between P3 to P2 and its highest in the positions P1' to P2. Therefore, the typical substrate of the astacins may be characterized as follows: non-hydrophobic, small and flexible residues in position P1' and to a slightly lesser extent in the flanking positions P1 and P2', and larger, non-flexible hydrophobic residues in positions P3 to P2. All investigated astacins exhibited a common pattern, despite significant differences in the individual substrate sequences, in the parameters studied. Therefore, these are combined in a single signature of astacins (figure 2 A). Compared directly with the other astacins, ovastacin (figure 2 B, C) exhibits a particularly high degree of conservation of residues in position P1' and, to a lesser extent, in position P2. The hydrophobicity in positions P4 to P1 (especially in P2) is even more pronounced, while position P1' is rather hydrophilic. Furthermore, the volume of side chains on the nonprime side, especially in position P2, is larger than for the other astacins. The flexibility of the residues in ovastacin substrates is comparatively lower in the region P4 to P1, but significantly higher in position P1'. Among astacins, the physicochemical cleavage signature of ovastacin shows the greatest similarity to that of meprin β (supplementary figure 3 C). However, in direct comparison, meprin β significantly favors glutamate in position P1'.

Substrate interaction

We first created homology models (supplementary figure 4) to analyze the substrate interaction with the catalytic domain of ovastacin based on the physicochemical properties of the surface. The model with the combined templates of ZHE-1 (PDB 3LQB(55)) and meprin β (PDBs 4GWN and 4GWM (56)) was used for this purpose. Based on the observation in standard orientation (57), positive charges are primarily located on the upper rim of the right half of the catalytic cleft (figure 3 A1). The bottom of the catalytic cleft displays non-charged hydrophilic properties, except for a hydrophobic region in the left periphery (figure 3 A2).

Hydrophobic regions stretch across the upper and lower rim of the active side cleft. Strongly hydrophobic regions are located at the left margin of the catalytic cleft as well as at its basis. In summary, there is a transition in the physicochemical properties from strongly hydrophobic to strongly hydrophilic with positive charges in the active side cleft, in standard orientation from left to right (i.e. from nonprime to prime side). Arginine, lysine and histidine are responsible for the formation of the hydrophilic, strongly positively charged regions in the S' region (figure 3 B). Also displayed are the large, rigid, and hydrophobic residues leucine, isoleucine, valine, phenylalanine, tryptophan, methionine and tyrosine (figure 3 D). Evaluation of the cleavage sites obtained from N-TAILS analyses revealed V-L-D↓D-D-D as the theoretically most preferred sequence spanning positions P3 to P3' in ovastacin cleavage sites (figure 1 D). This peptide was docked virtually into the catalytic cleft of ovastacin using Chimera (figure 3 C). It is apparent that the aliphatic substrate residues Val (P3) and Leu (P2) are in the region between Ile216, Ile219 and Phe216 on one side, Phe54 on the other side and Trp191 at the base of the active side cleft (figure 3 C). The aspartate residues of the prime side are, depending on the exact orientation of the side chains, proximate for interacting with the sidechain of Arg264. Arg174 appears too distant for a direct interaction at least for P1' and P2'. Likewise, Lys175 is too distant for a direct interaction with the aspartate residues in position P1'-P3'.

Physiological relevant substrates identified via TAILS analysis

Cleavage of ZP2, the only confirmed physiological substrate of ovastacin so far, is a key regulatory event in fertilization since it destroys the sperm receptor and mediates ZP-hardening. By screening the 489 substrates cleaved by ovastacin in our N-TAILS analysis for proteins involved in fertilization, we identified 12 different potentially fertilization-relevant proteins including 29 cleavage sites (Table 1 & 2). Several examples are to be mentioned explicitly. The β -1,4-galactosyltransferase (B4GALT1) is localized on the sperm membrane and binds sugar side chains of ZP3, thereby mediating additional attachment to the ZP (58). Calreticulin (CALR3) is localized on the surface of the acrosomal region of sperm (59) and also on the oolemma (60). There is evidence for its involvement in oocyte signal transduction (61) and the establishment of the polyspermy block after its exocytosis from the oocyte (62). Chaperonin containing TCP1 (Cct2) localizes also on the sperm surface and is directly involved in the binding of sperm to oocytes (63, 64). Matrix metalloprotease 2 (MMP2) is associated with the acrosome and potentially contributes to the penetration of the ZP (65). Taken together, detection of cleavages in these proteins strongly indicates a more complex

function of ovastacin within the proteolytic web of egg-sperm interaction beyond regulation of ZP hardening.

Discussion

Our N-TAILS analysis identified hundreds of ovastacin substrates and revealed the substrate specificity of ovastacin. In fact, ovastacin shares a high degree of similarity with the substrate specificity of meprin β . Accordingly, some well-known meprin β substrates, such as the amyloid precursor protein (APP), were cleaved by ovastacin as well (Supplemental Table 1). However, due to the spatio-temporal restriction of ovastacin to oocytes or the functional sphere of the egg-sperm interaction, these are presumably of no physiological relevance. The restriction to substrates to the extracellular space or the context of fertilization reduced the number of physiologically relevant ovastacin substrates in the present study from 489 to 163 and twelve, respectively. Owing that ovastacin expression is extremely limited to oocytes, this rigorous restriction may be appropriate (66). Although our data suggest a physiological processing, further steps are required to provide the physiological evidence. At least, a confirmation of the data obtained in a cellular context, e.g. egg or sperm secretome, is required to verify the physiological aspects of a cleavage by ovastacin. The germ cell specific expression pattern might compromise the activity or substrate selection of ovastacin e.g. by potential pre-processing or interactions of substrates. This would also explain why we were not able to identify the only physiologically proven substrate. ZP2 is not expressed by MEF cells (67). Establishment of the N-TAILS approach on germ cell secretome would remedy these problems.

Most of the fertilization-relevant substrates listed in Table 2 have been found to positively influence sperm viability or sperm-oocyte binding (63–65, 68, 69). If cleavage of these substrates results in their loss of function, ovastacin could be the initiator of a multifactorial regulation of sperm interaction. For instance, β -1,4-galactosyltransferase (B4GALT1) mediates the binding of sperm head to oligosaccharides of ZP3 until exocytosis of the acrosome (70). Although this interaction is not essential, unlike sperm binding to ZP2, its function is poorly understood. Proteolysis of B4GALT1 at S²⁴⁸ or D²⁴⁹, respectively, in a loop region between two central β -sheets, is likely to alter the conformation and binding properties to ZP3.

Similarly, for instance, the D²⁰⁸, D²⁰⁹ and D²¹⁰ in MMP2 are in the β -sheet above the zinc-binding active site helix (Table 2). From there, the chain enters the first fibronectin-like domain, which is an important exosite for binding protein substrates. Hence, cleavage within this region would most likely disrupt (protein-) substrate binding to MMP2 and abrogate its function and thereby may impair the penetration capacity of sperm. Reduction of sperm viability and hindrance of sperm binding to the oocyte would extend the effect of the already known ZP hardening by ovastacin. Recently, the regulation of ovastacin activity has been shown to be clinically relevant in humans (21, 71, 72). Thus, these potential substrates are of major interest for this field.

The cleavage preferences identified in this study displayed a clear preference for aspartate in position P1' and, to a lesser extent, for aspartate and glutamate in the P2' to P4' region of the prime side. In addition, there is a preference for leucine in position P2 of the nonprime side. A similar preference for aspartate in position P1' has been also described for the astacin proteases BMP1, meprin β , and LAST MAM (MAM-domain containing astacin from the horseshoe crab *Limulus polyphemus*) (supplemental figure 3). The substrate binding was analyzed by modeling. This interaction might be mediated by interactions with Arg174 and Arg264 and by the properties of hydrophobic (aromatic as well as aliphatic) residues of the nonprime side. In standard orientation, Phe154 is located at the upper left edge of the catalytic cleft, Val157 and Trp191 are in the bottom section, and Ile216, Ile219, Phe214 and Phe243 are present at the lower left edge. Together, these residues wrap around the peptide chain of the substrate in the left half of the catalytic cleft in a cuff-like fashion.

Crayfish astacin and LAST, unlike ovastacin, do not display preference for negatively charged residues in position P1'. However, we found great similarities between them in terms of distribution of physicochemical properties across the different positions relative to the cleavage site. All analyzed astacins share similar physicochemical properties in their substrate binding clefts. Thus, based on these findings, the different astacin proteases can be designated as variants of a common prototype. Accordingly, this prototypical signature is independent of the substrate sequence preferred by the respective enzyme. Due to the limited data currently available, this analysis was performed for only seven astacins (astacin, BMP1, LAST, LAST MAM, meprin α , meprin β , and ovastacin (Supplemental figure 3)). However, this analysis of physicochemical properties allows for deeper understanding of cleavage preference and substrate selectivity of proteases in general. This could contribute

to higher precision when predicting potential protease substrates, as well as to improvement in the design of proteases with a specific substrate specificity.

The knowledge of the biochemical properties that distinguish the cleavage of substrates by ovastacin from the closely related meprins could be useful for the development of more specific and selective inhibitors. Such inhibitors would significantly reduce the amount of oocytes required for IVF (21, 73). Additionally, these could overcome the problems of proteinogenic inhibitors (e.g. half-life, application restrictions, production) such as fetuin-B.

Conclusion

In summary, using N-TAILS, we were able to identify new substrates of the metalloproteinase ovastacin with potential physiological relevance. This deepens insight into the complexity of mammalian fertilization. The cleavage preference of ovastacin towards these native substrate proteins was resolved and resulted in the recognition of a globally applicable cleavage site signature based on physicochemical properties. This signature might serve as a useful tool to elucidate protease-substrate interactions and provide broader explanations for the genesis of cleavage preference in proteases.

Data availability statement

The LC-MS/MS data have been deposited to the ProteomeXchange Consortium via the JPOST repository with the dataset identifier PXD038561 (ProteomeXchange) and JPST001944 (jPOST).

Acknowledgments

The MEF cell line was kindly provided by Oliver Schilling (University Medical Center, Freiburg, Germany). We thank Michael Plenikowski for preparation of the graphical abstract.

Author Contributions:

Substantial contributions to conception and design (MF, UD, WS, HK). Acquisition of data, or analysis and interpretation of data (MF, UD, MŁ, NvW, WS, HK). Drafting the article or revising it critically for important intellectual content (MF, NvW, ST, CB, WS, HK). Final approval of the version to be published (all authors).

Funding

This work was supported by a grant from Deutsche Forschungsgemeinschaft (DFG) to H.K. (KO 6071/2-1).

Competing Interest Statement

None of the authors has competing interests.

References

1. Burkart, A. D., Xiong, B., Baibakov, B., Jiménez-Movilla, M., and Dean, J. (2012) Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. *J. Cell Biol.* **197**, 37–44
2. Zaneveld, L. J. D., Polakoski, K. L., and Williams, W. L. (1972) Properties of a Proteolytic Enzyme From Rabbit Sperm Acrosomes. *Biol. Reprod.* **6**, 30–39
3. Honda, A., Yamagata, K., Sugiura, S., Watanabe, K., and Baba, T. (2002) A Mouse Serine Protease TESP5 Is Selectively Included into Lipid Rafts of Sperm Membrane Presumably as a Glycosylphosphatidylinositol-anchored Protein. *J. Biol. Chem.* **277**, 16976–16984
4. Coy, P., Jiménez-Movilla, M., García-Vázquez, F. a, Mondéjar, I., Grullón, L., and Romar, R. (2012) Oocytes use the plasminogen-plasmin system to remove supernumerary spermatozoa. *Hum. Reprod.* **27**, 1985–93
5. Huarte, J. (1987) Plasminogen activator and mouse spermatozoa: urokinase synthesis in the male genital tract and binding of the enzyme to the sperm cell surface. *J. Cell Biol.* **104**, 1281–1289
6. Kim, E., Nishimura, H., and Baba, T. (2003) Differential localization of ADAM1a and ADAM1b in the endoplasmic reticulum of testicular germ cells and on the surface of epididymal sperm. *Biochem. Biophys. Res. Commun.* **304**, 313–319
7. Hirose, M., Honda, A., Fulka, H., Tamura-Nakano, M., Matoba, S., Tomishima, T., Mochida, K., Hasegawa, A., Nagashima, K., Inoue, K., Ohtsuka, M., Baba, T., Yanagimachi, R., and Ogura, A. (2020) Acrosin is essential for sperm penetration through the zona pellucida in hamsters. *Proc. Natl. Acad. Sci.* **117**, 2513–2518
8. Kuske, M., Floehr, J., Yiallourous, I., Michna, T., Jahnen-Dechent, W., Tenzer, S., Stöcker, W., and Körschgen, H. (2021) Limited proteolysis by acrosin affects sperm-binding and mechanical resilience of the mouse zona pellucida. *Mol. Hum. Reprod.* **27**, 1–10
9. Rawlings, N. D., and Salvesen, G. (eds.) (2013) *Handbook of Proteolytic Enzymes 1, 2, 3*, 3rd Ed., Elsevier, London
10. Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014) MEROPS : the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503–D509
11. Mascarenhas, M. N., Flaxman, S. R., Boerma, T., Vanderpoel, S., and Stevens, G. A. (2012) National, Regional, and Global Trends in Infertility Prevalence Since 1990: A Systematic Analysis of 277 Health Surveys. *PLoS Med.* **9**, e1001356
12. Smith, S. (2003) Diagnosis and Management of Female Infertility. *JAMA.* **290**, 1767

13. Gomis-Rüth, F. X., Trillo-Muyo, S., and Stöcker, W. (2012) Functional and structural insights into astacin metallopeptidases. *Biol. Chem.* **393**, 1027–41
14. Xiong, B., Zhao, Y., Beall, S., Sadusky, A. B., and Dean, J. (2017) A Unique Egg Cortical Granule Localization Motif Is Required for Ovastacin Sequestration to Prevent Premature ZP2 Cleavage and Ensure Female Fertility in Mice. *PLOS Genet.* **13**, e1006580
15. Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., and Dean, J. (2010) Gamete Recognition in Mice Depends on the Cleavage Status of Egg's Zona Pellucida Protein. *Science* **329**, 216–219
16. Körschgen, H., Kuske, M., Karmilin, K., Yiallourous, I., Balbach, M., Floehr, J., Wachten, D., Jahnen-dechent, W., and Stöcker, W. (2017) Intracellular activation of ovastacin mediates pre-fertilization hardening of the zona pellucida. *Mol. Hum. Reprod.* **23**, 607–616
17. Nozawa, K., Satouh, Y., Fujimoto, T., Oji, A., and Ikawa, M. (2018) Sperm-borne phospholipase C zeta-1 ensures monospermic fertilization in mice. *Sci. Rep.* **8**, 1315
18. Bianchi, E., Doe, B., Goulding, D., and Wright, G. J. (2014) Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature.* **508**, 483–487
19. Bianchi, E., and Wright, G. J. (2014) Izumo meets Juno: Preventing polyspermy in fertilization. *Cell Cycle.* **13**, 2019–2020
20. Dietzel, E., Wessling, J., Floehr, J., Schäfer, C., Ensslen, S., Denecke, B., Rösing, B., Neulen, J., Veitinger, T., Spehr, M., Tropartz, T., Tolba, R., Renné, T., Egert, A., Schorle, H., Gottenbusch, Y., Hildebrand, A., Yiallourous, I., Stöcker, W., Weiskirchen, R., and Jahnen-Dechent, W. (2013) Fetuin-B, a liver-derived plasma protein is essential for fertilization. *Dev. Cell.* **25**, 106–12
21. Dietzel, E., Floehr, J., Van de Leur, E., Weiskirchen, R., and Jahnen-Dechent, W. (2016) Recombinant fetuin-B protein maintains high fertilization rate in cumulus cell-free mouse oocytes. *Mol. Hum. Reprod.* **23**, 25–33
22. Cuppari, A., Körschgen, H., Fahrenkamp, D., Schmitz, C., Guevara, T., Karmilin, K., Kuske, M., Olf, M., Dietzel, E., Yiallourous, I., de Sanctis, D., Goulas, T., Weiskirchen, R., Jahnen-Dechent, W., Floehr, J., Stoecker, W., Jovine, L., and Gomis-Rüth, F. X. (2019) Structure of mammalian plasma fetuin-B and its mechanism of selective metallopeptidase inhibition. *IUCrj* **6**, 317–330
23. Guevara, T., Körschgen, H., Cuppari, A., Schmitz, C., Kuske, M., Yiallourous, I., Floehr, J., Jahnen-Dechent, W., Stöcker, W., and Gomis-Rüth, F. X. (2019) The C-terminal region of human plasma fetuin-B is dispensable for the raised-elephant-trunk mechanism of inhibition of astacin metallopeptidases. *Sci. Rep.* **9**, 14683
24. Eckhard, U., Körschgen, H., von Wiegen, N., Stöcker, W., and Gomis-Rüth, F. X. (2021) The crystal structure of a 250-kDa heterotetrameric particle explains inhibition of sheddase

- mepriin β by endogenous fetuin-B. *Proc. Natl. Acad. Sci.* **118**, 1–6
25. Karmilin, K., Schmitz, C., Kuske, M., Körschgen, H., Olf, M., Meyer, K., Hildebrand, A., Felten, M., Fridrich, S., Yiallourous, I., Becker-Pauly, C., Weiskirchen, R., Jahnen-Dechent, W., Floehr, J., and Stöcker, W. (2019) Mammalian plasma fetuin-B is a selective inhibitor of ovastacin and mepriin metalloproteinases. *Sci. Rep.* **9**, 546
 26. Turk, V., and Bode, W. (1991) The cystatins: Protein inhibitors of cysteine proteinases. *FEBS Lett.* **285**, 213–219
 27. Kleifeld, O., Doucet, A., Auf Dem Keller, U., Prudova, A., Schilling, O., Kainthan, R. K., Starr, A. E., Foster, L. J., Kizhakkedathu, J. N., and Overall, C. M. (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol.* **28**, 281–288
 28. Sambrook, J., Fritsch, E. ., and Maniatis, T. (1989) *Molecular cloning : a laboratory manual*, Cold Spring Harbor Laboratory Press
 29. Kleifeld, O., Doucet, A., Prudova, A., auf dem Keller, U., Gioia, M., Kizhakkedathu, J. N., and Overall, C. M. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* **6**, 1578–1611
 30. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906
 31. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254
 32. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85
 33. Helm, D., Vissers, J. P. C., Hughes, C. J., Hahne, H., Ruprecht, B., Pachel, F., Grzyb, A., Richardson, K., Wildgoose, J., Maier, S. K., Marx, H., Wilhelm, M., Becher, I., Lemeer, S., Bantscheff, M., Langridge, J. I., and Kuster, B. (2014) Ion Mobility Tandem Mass Spectrometry Enhances Performance of Bottom-up Proteomics. *Mol. Cell. Proteomics.* **13**, 3709–3715
 34. Tran, N. H., Zhang, X., Xin, L., Shan, B., and Li, M. (2017) De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci.* **114**, 8247–8252
 35. Tran, N. H., Qiao, R., Xin, L., Chen, X., Liu, C., Zhang, X., Shan, B., Ghodsi, A., and Li, M. (2019) Deep learning enables de novo peptide sequencing from data-independent-acquisition mass

- spectrometry. *Nat. Methods*. **16**, 63–66
36. Tran, N. H., Rahman, M. Z., He, L., Xin, L., Shan, B., and Li, M. (2016) Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **6**, 31730
 37. Andreatta, M., Alvarez, B., and Nielsen, M. (2017) GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res.* **45**, W458–W463
 38. Andreatta, M., Lund, O., and Nielsen, M. (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* **29**, 8–14
 39. Becker-Pauly, C., Barré, O., Schilling, O., auf dem Keller, U., Ohler, A., Broder, C., Schütte, A., Kappelhoff, R., Stöcker, W., and Overall, C. M. (2011) Proteomic Analyses Reveal an Acidic Prime Side Specificity for the Astacin Metalloprotease Family Reflected by Physiological Substrates. *Mol. Cell. Proteomics*. **10**, M111.009233
 40. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13
 41. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57
 42. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods*. **6**, 786–787
 43. Henikoff, S., and Henikoff, J. G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **89**, 10915–10919
 44. Nozaki, Y., and Tanford, C. (1971) The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions. *J. Biol. Chem.* **246**, 2211–2217
 45. Bigelow, C. C. (1967) On the average hydrophobicity of proteins and the relation between it and protein structure. *J. Theor. Biol.* **16**, 187–211
 46. Bhaskaran, R., and Ponnuswamy, P. K. (1988) Positional flexibilities of amino acid residues in globular proteins. *Int. J. Pept. Protein Res.* **32**, 241–255
 47. Webb, B., and Sali, A. (2016) Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1–5.6.37
 48. Šali, A., and Blundell, T. L. (1993) Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **234**, 779–815
 49. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**,

283–291

50. Laskowski, R., Rullmann, J. A., MacArthur, M., Kaptein, R., and Thornton, J. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR.* **8**, 477–486
51. Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., and Vajda, S. (2017) The ClusPro web server for protein–protein docking. *Nat. Protoc.* **12**, 255–278
52. Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., and Vajda, S. (2013) How good is automated protein docking? *Proteins Struct. Funct. Bioinforma.* **81**, 2159–2166
53. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., and Kozakov, D. (2017) New additions to the ClusPro server motivated by CAPRI. *Proteins Struct. Funct. Bioinforma.* **85**, 435–444
54. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157–162
55. Okada, A., Sano, K., Nagata, K., Yasumasu, S., Ohtsuka, J., Yamamura, A., Kubota, K., Iuchi, I., and Tanokura, M. (2010) Crystal Structure of Zebrafish Hatching Enzyme 1 from the Zebrafish *Danio rerio*. *J. Mol. Biol.* **402**, 865–878
56. Arolas, J. L., Broder, C., Jefferson, T., Guevara, T., Sterchi, E. E., Bode, W., Stöcker, W., Becker-Pauly, C., and Gomis-Rüth, F. X. (2012) Structural basis for the sheddase function of human meprin β metalloproteinase at the plasma membrane. *Proc. Natl. Acad. Sci.* **109**, 16131–16136
57. Gomis-Rüth, F. X., Botelho, T. O., and Bode, W. (2012) A standard orientation for metalloproteinases. *Biochim. Biophys. Acta - Proteins Proteomics.* **1824**, 157–163
58. Miller, D. J., Macek, M. B., and Shur, B. D. (1992) Complementarity between sperm surface beta-1,4-galactosyltransferase and egg-coat ZP3 mediates sperm-egg binding. *Nature.* **357**, 589–93
59. Nakamura, M., Moriya, M., Baba, T., Michikawa, Y., Yamanobe, T., Arai, K., Okinaga, S., and Kobayashi, T. (1993) An Endoplasmic Reticulum Protein, Calreticulin, Is Transported into the Acrosome of Rat Sperm. *Exp. Cell Res.* **205**, 101–110
60. Calvert, M. E., Digilio, L. C., Herr, J. C., and Coonrod, S. A. (2003) Oolemmal proteomics--identification of highly abundant heat shock proteins and molecular chaperones in the mature mouse egg and their localization on the plasma membrane. *Reprod. Biol. Endocrinol.* **1**, 27
61. Tutuncu, L., Stein, P., Ord, T. S., Jorgez, C. J., and Williams, C. J. (2004) Calreticulin on the

- mouse egg surface mediates transmembrane signaling linked to cell cycle resumption. *Dev. Biol.* **270**, 246–260
62. Saavedra, M. D., Mondéjar, I., Coy, P., Betancourt, M., González-Márquez, H., Jiménez-Movilla, M., Avilés, M., and Romar, R. (2014) Calreticulin from subolemmal vesicles affects membrane regulation of polyspermy. *Reproduction* **147**, 369–378
63. Redgrove, K. A., Anderson, A. L., Dun, M. D., McLaughlin, E. A., O’Bryan, M. K., Aitken, R. J., and Nixon, B. (2011) Involvement of multimeric protein complexes in mediating the capacitation-dependent binding of human spermatozoa to homologous zonae pellucidae. *Dev. Biol.* **356**, 460–474
64. Dun, M. D., Smith, N. D., Baker, M. A., Lin, M., Aitken, R. J., and Nixon, B. (2011) The Chaperonin Containing TCP1 Complex (CCT/TRiC) Is Involved in Mediating Sperm-Oocyte Interaction. *J. Biol. Chem.* **286**, 36875–36887
65. Ferrer, M., Rodriguez, H., Zara, L., Yu, Y., Xu, W., and Oko, R. (2012) MMP2 and acrosin are major proteinases associated with the inner acrosomal membrane and may cooperate in sperm penetration of the zona pellucida during fertilization. *Cell Tissue Res.* **349**, 881–895
66. Pires, E. S., Hlavin, C., Macnamara, E., Ishola-Gbenla, K., Doerwaldt, C., Chamberlain, C., Klotz, K., Herr, A. K., Khole, A., Chertihin, O., Curnow, E., Feldman, S. H., Mandal, A., Shetty, J., Flickinger, C., and Herr, J. C. (2013) SAS1B protein [ovastacin] shows temporal and spatial restriction to oocytes in several eutherian orders and initiates translation at the primary to secondary follicle transition. *Dev. Dyn.* **010425**, 1–22
67. Sarkar, P., Randall, S. M., Muddiman, D. C., and Rao, B. M. (2012) Targeted proteomics of the secretory pathway reveals the secretome of mouse embryonic fibroblasts and human embryonic stem cells. *Mol. Cell. Proteomics.* **11**, 1829–1839
68. van Gestel, R. A., Brewis, I. A., Ashton, P. R., Brouwers, J. F., and Gadella, B. M. (2007) Multiple proteins present in purified porcine sperm apical plasma membranes interact with the zona pellucida of the oocyte. *Mol. Hum. Reprod.* **13**, 445–54
69. Elliott, R. M. A., Lloyd, R. E., Fazeli, A., Sostaric, E., Georgiou, A. S., Satake, N., Watson, P. F., and Holt, W. V (2009) Effects of HSPA8, an evolutionarily conserved oviductal protein, on boar and bull spermatozoa. *Reproduction* **137**, 191–203
70. Miller, D. J., MacEk, M. B., and Shur, B. D. (1992) Complementarity between sperm surface β -1,4-galactosyl-transferase and egg-coat ZP3 mediates sperm-egg binding. *Nature* **357**, 589–593
71. Maddirevula, S., Coskun, S., Al, M., Omar, Q., and Saad, A. (2021) ASTL is mutated in female

infertility. *Hum. Genet.* 10.1007/s00439-021-02388-8

72. Floehr, J., Dietzel, E., Neulen, J., Rösing, B., Weissenborn, U., and Jahnen-Dechent, W. (2016) Association of high fetuin-B concentrations in serum with fertilization rate in IVF: A cross-sectional pilot study. *Hum. Reprod.* **31**, 630–637
73. Körschgen, H., Jäger, C., Tan, K., Buchholz, M., Stöcker, W., and Ramsbeck, D. (2020) A Primary Evaluation of Potential Small-Molecule Inhibitors of the Astacin Metalloproteinase Ovastacin, a Novel Drug Target in Female Infertility Treatment. *ChemMedChem* **15**, 1499–1504
74. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132

Figures and Tables

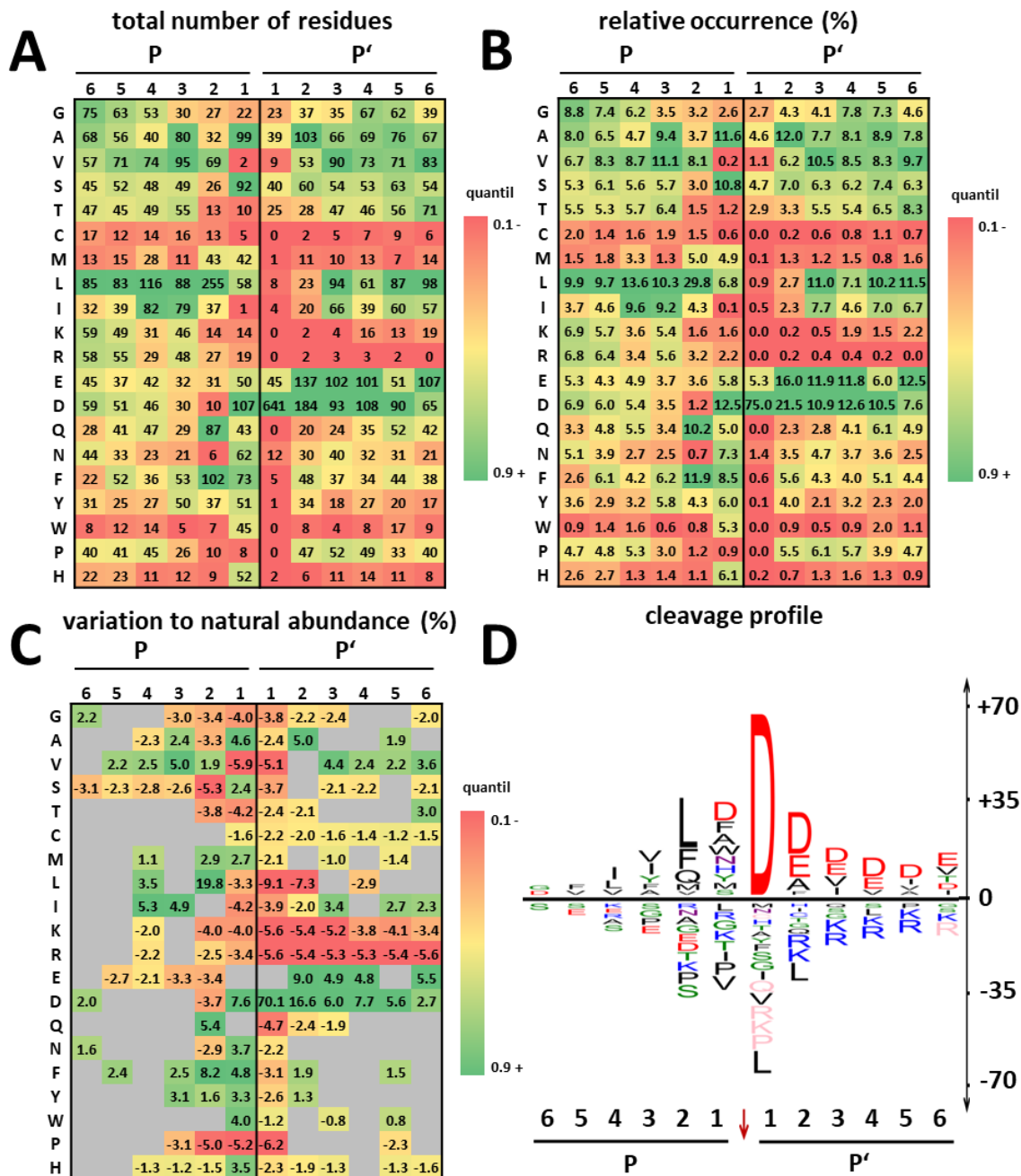


Figure 1. Heatmap and iceLogo of cleavage sites (n= 855) identified via N-TAILS. Residues are normalized to their natural occurrence in the mouse proteome (*Mus musculus*). Distribution of residues in positions P6-P1 (nonprime side) and P'1'-P6' (prime side). Displayed for each position are the total number of residues (A), their relative abundance in percent (B) and the difference to the natural abundance in percentage points as heatmap (C) and as iceLogo (D). The deviations were determined with iceLogo (42). Positions without significant changes (p>0.05) are depicted in gray for the heatmaps, the iceLogo displays only significant changes (p≤0.05). Amino acids abbreviated as single letter.

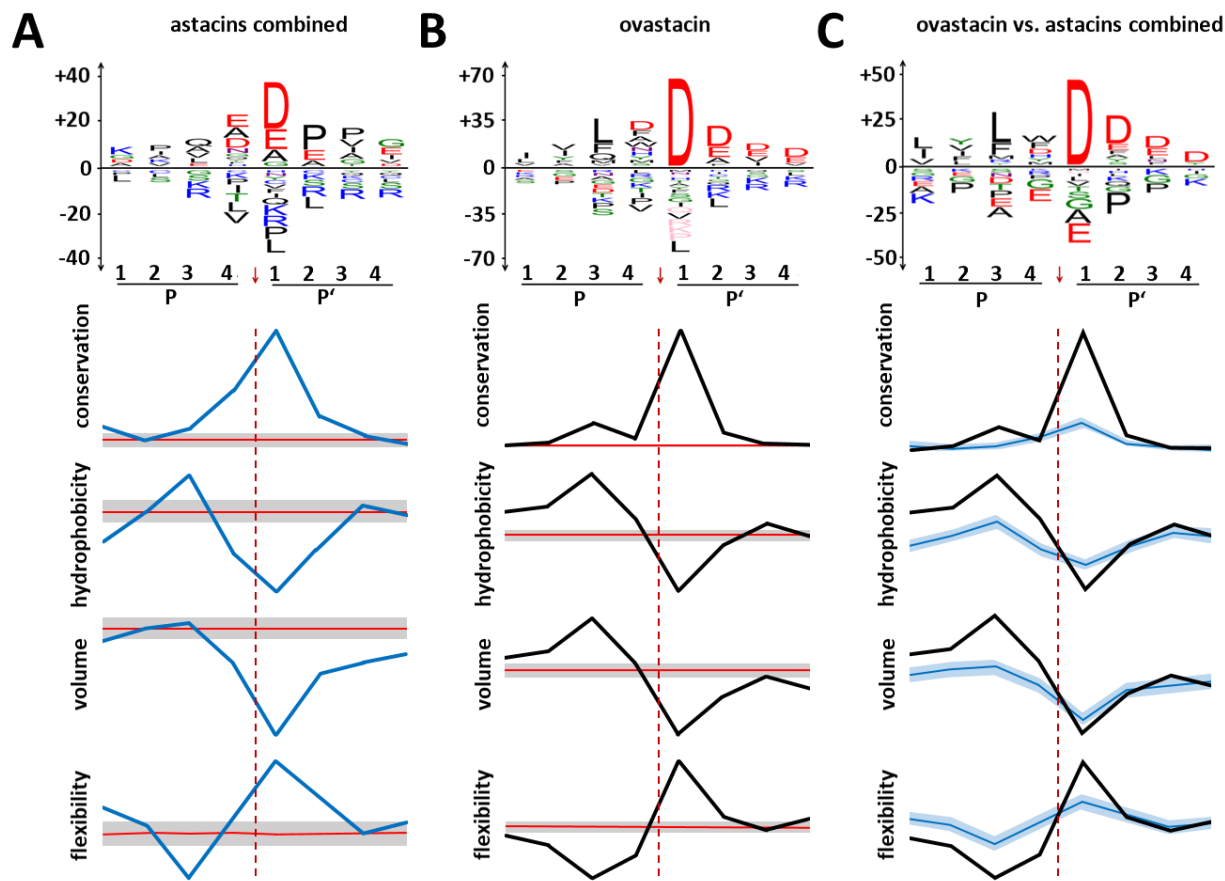


Figure 2. Physicochemical properties of astacin proteases and ovastacin. Overview of the common cleavage preferences and physicochemical properties of all studied astacin proteases combined (left), ovastacin only (middle) and comparatively of all astacin proteases versus ovastacin (right). The analysis was performed on the basis of the cleavage sites deposited in the MEROPS database (astacin (P07584), n=199; BMP1 (P13497), n=25; LAST_MAM (B4F320), n=415; LAST (B4F319), n=76; meprin α (Q16819), n=700; meprin β (Q16820), n=879) and ovastacin (Q6HA09), n=855 (this study). Depicted in each case is a sequence logo (upper panel) displaying the difference to the natural abundance of residues in the cleavage site positions P4-P4', the degree of conservation of positions according to the BLOSUM62 substitution matrix (43) (upper graph), the hydrophobicity of the residues (44) (second graph from the top), the volume of the residues (45) (second graph from the bottom) and the flexibility of the residues (46) (bottom graph). The significance level is 95%. The error range is displayed in gray/light blue. The diagrams allow a semi-quantitative evaluation, since iceLogo does not allow scale normalization.

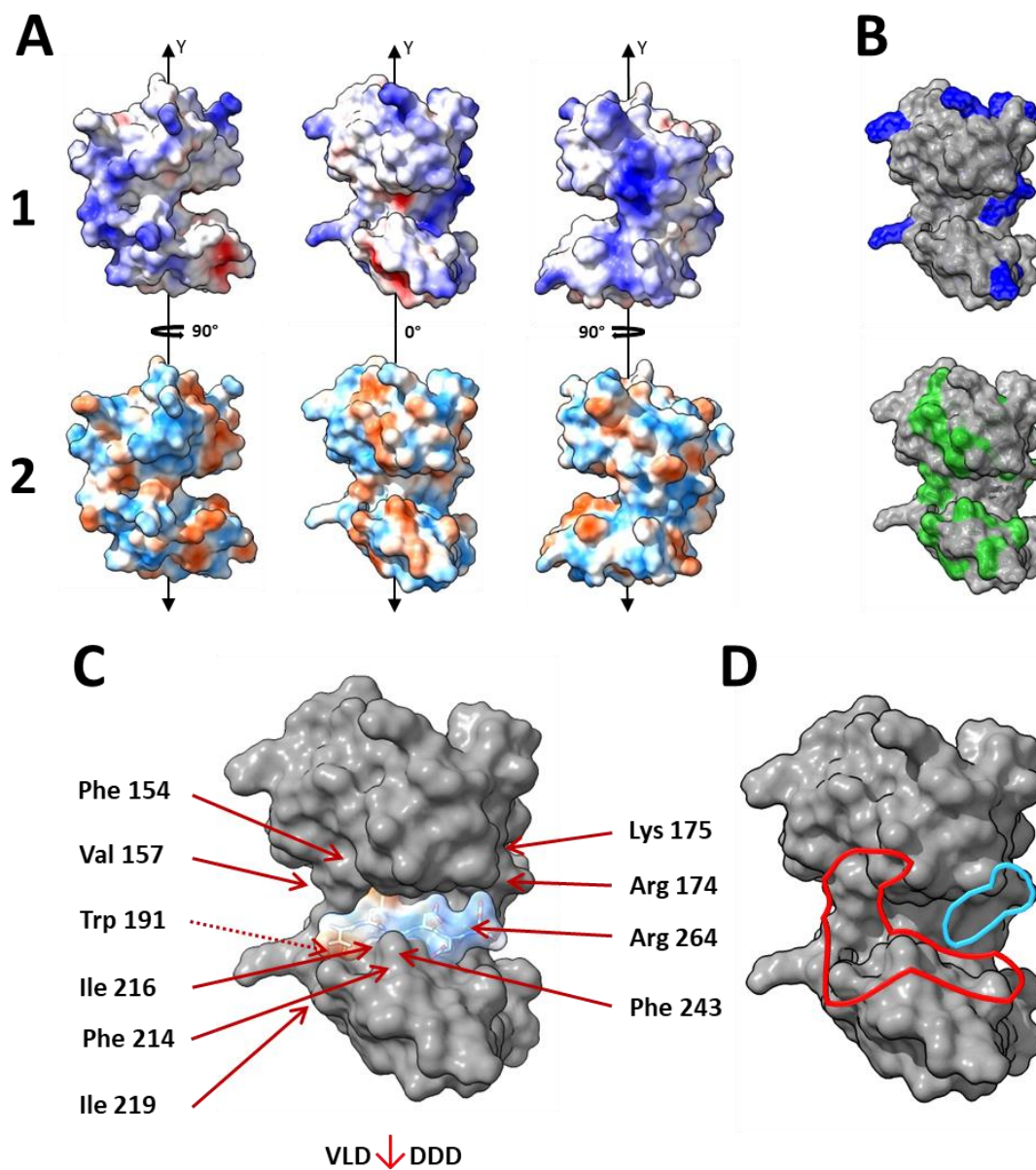


Figure 3. Structural analysis of substrate interaction. Surface models of the catalytic domain of murine ovastacin (Q6HA09; position 92-282) created with Modeller (47) using the templates of the catalytic domains of meprin β and ZHE-1 (PDBs: 4GWN, 4GWM, and 3LQB). (A 1) Coulomb potential of the surface highlighting areas with positive (blue, 10 kcal/(mol*e)) and negative (red, -10 kcal/(mol*e)) electrostatic potential. (A 2) Surface areas of hydrophilic (light blue) and hydrophobic (orange) residues according to the Kyte-Doolittle scale (74). (B) Upper model with all accessible positively charged residues (Arg, Lys) and lower model with all accessible hydrophobic residues (Leu, Ile, Val, Pro, Phe, Trp, Met and Tyr). (C) Synthetic peptide (surface areas of hydrophilic (light blue) and hydrophobic (orange) residues) according to the most preferred residues in position P3 to P3' (figure 1) and inserted into the catalytic cleft via Chimera; highlighted residues of ovastacin are discussed (D) Hydrophobic (outlined in red) and hydrophilic (outlined in blue) regions in the catalytic cleft.

Table 1. Search terms for identification of substrates associated with fertilization. Search terms in the DAVID bioinformatics database (40, 41) including their category, the p-value, the number of substrates in the dataset and their fraction of the total dataset.

search term	category	P-value	substrates n	proportion [%]
<i>zona pellucida</i> receptor complex	GOTERM_CC_DIRECT	3.5 E-3	4	0.8
binding of sperm to <i>zona pellucida</i>	GOTERM_BP_DIRECT	1.4 E-2	5	1.0
estrogen signaling pathway	KEGG_PATHWAY	7.4 E-2	8	1.6
<i>in vitro</i> fertilized eggs	UP_TISSUE	1.4 E-2	13	2.7

Table 2. Substrates of ovastacin with proposed function in fertilization. In TAILS analysis identified substrates whose annotations in the DAVID Bioinformatics database suggest a potential physiological impact on the fertilization. Listed are the substrates, the UniProt accession number, the search term, the total number of amino acids, the cleavage site position and the corresponding residues on the nonprime and prime side.

no.	Substrate	uniprot accession number	<i>in vitro</i> fertilized eggs <i>zona pellucida</i> binding complex binding of sperm to <i>zona pellucida</i> estrogen signaling pathway				n residues substrate	P1' no.	P6 - P1	P1' - P6'
1	calreticulin (Calr)	P14211	+	-	-	-	416	18	LGLAA	DPAIYF
								30	KEQFLD	GDAWTN
								165	KDIRCK	DDEFTH
								335	NFLITN	DEAYAE
2	T-complex protein 1 subunit beta (Cct2)	P80314	-	+	+	-	535	92	DMSRVQ	DDEVGD
								392	AERSLH	DALCVL
								452	LPTIIA	DNAGYD
								457	ADNAGY	DSADLV
508	QVLLSA	AEAAEV								
3	T-complex protein 1 subunit zeta (Cct6a)	P80317	-	+	+	-	531	442	GVQAFA	DALLII
4	eukaryotic translation initiation factor 4A1 (Eif4a1)	P60843	+	-	-	-	406	36	IVDSFD	DMNLSE
								182	IKMFVL	DEADEM
5	heat shock protein 8 (Hspa8)	P63017	-	-	-	+	646	79	LIGRRF	DDAVVQ
								109	VQVEYK	GETKSF
6	heat shock protein HSP 90-beta (Hsp90ab1)	P11499	-	-	-	+	724	364	RRVFIM	DSCDEL
7	heat shock protein HSP 90-alpha	P07901	-	-	-	+	733	373	RRVFIM	DNCEEL
669	LVILLY	ETALLS								
8	endoplasmic reticulum chaperone protein (Hsp90b1)	P08113	-	-	-	+	802	226	QHIWES	DSNEFS
9	heterogeneous nuclear ribonucleoprotein U (HnrnpU)	Q8VEK3	+	-	-	-	800	190	GPTSFL	AVTVAP
10	72 kDa type IV collagenase (MMP2)	P33434	-	-	-	+	662	208	GGDSHF	DDDELW
								209	GDSHFD	DDELWT
								210	DSHFDD	DELWTL
								264	RFSRIH	DGEADI
11	peroxiredoxin-5 (Prdx5)	P99029	+	-	-	-	210	123	QVVAQL	SVNDVF
								126	ACLSVN	DVFVIE
								162	ATDLLL	DDSLVS
								163	TDLLLD	DSLVSLL
12	β -1.4-galactosyltransferase 1 (B4galt1)	P15535	-	-	+	-	339	248	YNCVFV	SDVDLI
								249	NCFVFS	DVDLIP