Substrate profiling of the metalloproteinase ovastacin uncovers specific enzyme–substrate interactions and discloses fertilization-relevant substrates

Matthias Felten[1], Ute Distler[2], Nele von Wiegen[3], Mateusz Łącki[2], Christian Behl[3], Stefan Tenzer[2], Walter Stöcker[1] and Hagen Körschgen[3] (iD)

1 Institute of Molecular Physiology, Cell and Matrix Biology, Johannes Gutenberg-University Mainz, Germany
2 Institute for Immunology, University Medical Center of the Johannes Gutenberg-University Mainz, Germany
3 Institute of Pathobiochemistry, The Autophagy Lab, University Medical Center of the Johannes Gutenberg-University Mainz, Germany

The metalloproteinase ovastacin is released by the mammalian egg upon fertilization and cleaves a distinct peptide bond in zona pellucida protein 2 (ZP2), 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 ovastacin–substrate interactions as well as the individual characteristics that distinguish ovastacin from similar proteases, such as meprins and tolloid. Eventually, we identified several substrates whose cleavage could affect mammalian fertilization. Consequently, these substrates indicate newly identified functions of ovastacin in mammalian fertilization beyond zona pellucida hardening.

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 in general.

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 (H^{182}ExxHxxGxxH^{192}) and the strictly conserved 1,4-β-turn, called the Met-turn (M^{236}), typical for metzincin proteinases [13]. Additionally, it contains a unique cortical granule localization motif (D^{52}KDIPAIN^{59}) within the N-terminal propeptide, upstream of the conserved zinc-binding aspartate ensuring latency, and a C-terminal region (CTR) of 150 residues in mice with unknown function [14,15].

Upon plasmogamy and oocyte activation, ovastacin is released into the perivitelline space during the cortical reaction and cleaves a distinct peptide bond (SRLA↓D^{168}ENQ) 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 [1,16]. Even though ZP2 has so far been the only proven physiological substrate for ovastacin, cleavage of additional substrates in the complex secretome of fertilization appears likely.

Phenotypically, ovastacin-deficient mice ($ASTL^{-/-}$) are inconspicuous except for a significant reduction in fecundity [17]. 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 [17,18]. However, this does not increase the rate of polyspermy, which further supports the concept that the decisive block against polyspermy is located at the oolemma level [18–20]. Similarly, loss-of-function mutations in ovastacin also appear to affect human reproduction [21].

Absence of the plasma protein fetuin-B, an endogenous ovastacin inhibitor, leads to female infertility [22]. 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 [17,22,23].

Recently, we were able to elucidate the precise mechanism of this inhibition [24–26]. In mammals, only three members of the astacin family of proteinases, ovastacin, and the closely related meprins (α & β)

are inhibited by fetuin-B [24,27]. The underlying molecular mechanism differs greatly from the classical cystatin/papain interaction [28]. 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 (C^{154}PDC^{157}) between the two cystatin-like domains and a hairpin loop in the second cystatin-like domain.

In the present study, we analyzed the cleavage specificity of ovastacin by applying N-terminal amine isotopic labeling of substrates (N-TAILS) [29] to the secretome of mouse embryonic fibroblasts (MEFs). We found unique characteristics of ovastacin, differing from other astacins. Furthermore, we identified a variety of cleavage sites in novel potential physiological substrates. These cleavages might extend the functions of ovastacin during fertilization.

Results

N-TAILS analysis

We used the secretome proteins of mouse embryonic fibroblasts (MEFs) as a versatile resource for mass spectrometric N-TAILS analysis of the cleavage specificity of heterologously expressed murine ovastacin. We have favored this model, as ovastacin and its regulation by fetuin-B have been best studied for the mouse ortholog [1,17,22,24,30–33]. Since the physiological function, ZP2 cleavage, *per se*, and its regulation by fetuin-B are highly conserved [24,25,30], we assume this model is valid for most theria and thus also for humans as well as domestic animals. MEFs were chosen rather than germ cells to overcome the particularly low quantity of germ cell secretome and to avoid very high animal consumption. Based on this complex mixture of natively folded mouse proteins, we unraveled ovastacin's specific cleavage preference. On top, we detected a series of novel, potentially physiologically relevant ovastacin substrates. Altogether, 855 unique cleavage sites were identified in 489 proteins. After restriction to the physiological context, that is, the extracellular space, we allocated a dataset of 353 unique cleavage sites to 163 proteins (Table S1).

Ovastacin adds individual characteristics to the typical acidic prime side specificity of most astacin metalloproteinases

Results are presented according to the concept by Schechter and Berger [34], indicating amino acid residues located N- and C-terminally of the cleavage site

as "non-prime" (P) and "prime" (P′), respectively. The corresponding subsites of the protease, harboring the substrate's side chains, are termed S and S′, accordingly. The general cleavage profile of the 855 unique cleavage sites shown in Fig. 1 is highlighted by the graphic iceLogo representation (Fig. 1D). The cleavage site in mouse ZP2, the so far only known physiological ovastacin substrate, is SRLA↓D[168]ENQ [1]. Figure 1E shows a frequency plot of the corresponding ovastacin cleavage sites in mammalian ZP2-proteins (PGMA↓–DENA, for alignment see Fig. S1), which slightly deviates from the mouse sequence. The key preference of ovastacin is aspartate in position P1′. At this crucial position, aspartate even outnumbers glutamate at a striking ratio of 15 : 1. In general, aspartate is clearly favored over glutamate in all positions of the prime side (P1′ to P5′). Ovastacin significantly prefers negatively charged residues in positions P1′ to P6′, especially aspartate in positions P1′ (+70.1%, rel. to natural occurrence) and P2′ (+16.6%) (Fig. 1C,D). There is considerable variability in position P1, where 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–P4, hydrophobic residues are preferred, such as leucine, (P2, +19.8%), phenylalanine (P2, +8.2%), valine (P3, +5.0%), and isoleucine (P4, +5.3%). Condensed, on the non-prime side, bulky aromatic side chains are observed at higher frequency in positions from P3 to P1, whereas negatively charged residues, especially aspartate, predominate the prime side. This is highlighted by the graphic iceLogo representation (Fig. 1D). As reflected from our N-TAILS analysis, the most crucial side chains are P1′ and P2′. Adjacent side chains in P4, P3, P2 on the non-prime side and P3′, P4′ on the prime side are more variable, probably also reflected by the conservation of the cleavage site in the mammalian in ZP2-proteins (Fig. 1E). This probably contributes to species-specific cleavage.

## Physicochemical properties of the ovastacin substrate binding cleft compared to other astacins

The cleavage preference determined for ovastacin (Fig. 1D) 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′ [35]. To identify ovastacin-specific characteristics, we analyzed in detail the physicochemical properties of cleavage sites, that is, electrostatics, hydrophilicity/

hydrophobicity, volume, flexibility, and degree of conservation of distinct residues flanking the cleavage site (Fig. 2). All signatures of the individual astacin members (Fig. 2, Fig. S2) feature the highest degree of conservation in position P1′, and to lesser extent in the positions P2 to P2′ for meprin α only. The hydrophobicity of the residues is low in P1 and P1′, and high in P3 and P2. The side chains with the largest volumes tend to be in P3 to P2, those with the smallest in the range P1 to P2′. The flexibility of the residues is lowest between P3 and P2 and is highest in the positions between P1′ and P2′. Therefore, the typical substrate of the astacins may be characterized as follows: small, flexible, hydrophilic residues in P1′ and to a slightly lesser extent in the flanking positions P1 and P2′, and large, rigid, hydrophobic residues in positions P3 to P2. All investigated astacins exhibit a common pattern, despite significant differences in the individual substrate sequences (Fig. 2) and can be combined in a typical astacin signature of astacins ("astacin footprint") (Fig. 2A). How does the individual "ovastacin footprint" look like, in comparison? Ovastacin (Fig. 2B,C) shares the family-specific, strong preference for aspartate in P1′, but is preferring aspartate also in P2′–P4′, even over glutamate. The "ovastacin footprint" also includes pronounced hydrophobicity, rigidity, and bulkiness of residues in positions P4 to P2 (especially in P2) (Fig. 2C).

Among astacins, the physicochemical cleavage signature of ovastacin shows the closest similarity to that of meprin β (Fig. 2E). However, in direct comparison, meprin β equally favors glutamate and aspartate in position P1′.

Substrate interaction

We first created homology models of murine ovastacin to analyze the substrate interaction with the catalytic domain of ovastacin based on the physicochemical properties of the active site cleft. We focused on the catalytic domain, since in the mouse model the catalytic domain is released as a consequence of fertilization without CTR [17]. As visualized by the model, positive charges are primarily located on the upper rim of the right half of the catalytic cleft in standard orientation [36] (Fig. 3A1). The bottom of the catalytic cleft displays non-charged hydrophilic properties, except for a hydrophobic region in the left periphery (Fig. 3A2). Hydrophobic regions stretch across the upper and lower rim of the active site cleft. Strongly hydrophobic regions are located at the left margin and the basis of the cleft. In summary, there is a gradient in the cleft's physicochemical properties from strongly hydrophobic to strongly hydrophilic with positive charges in the active
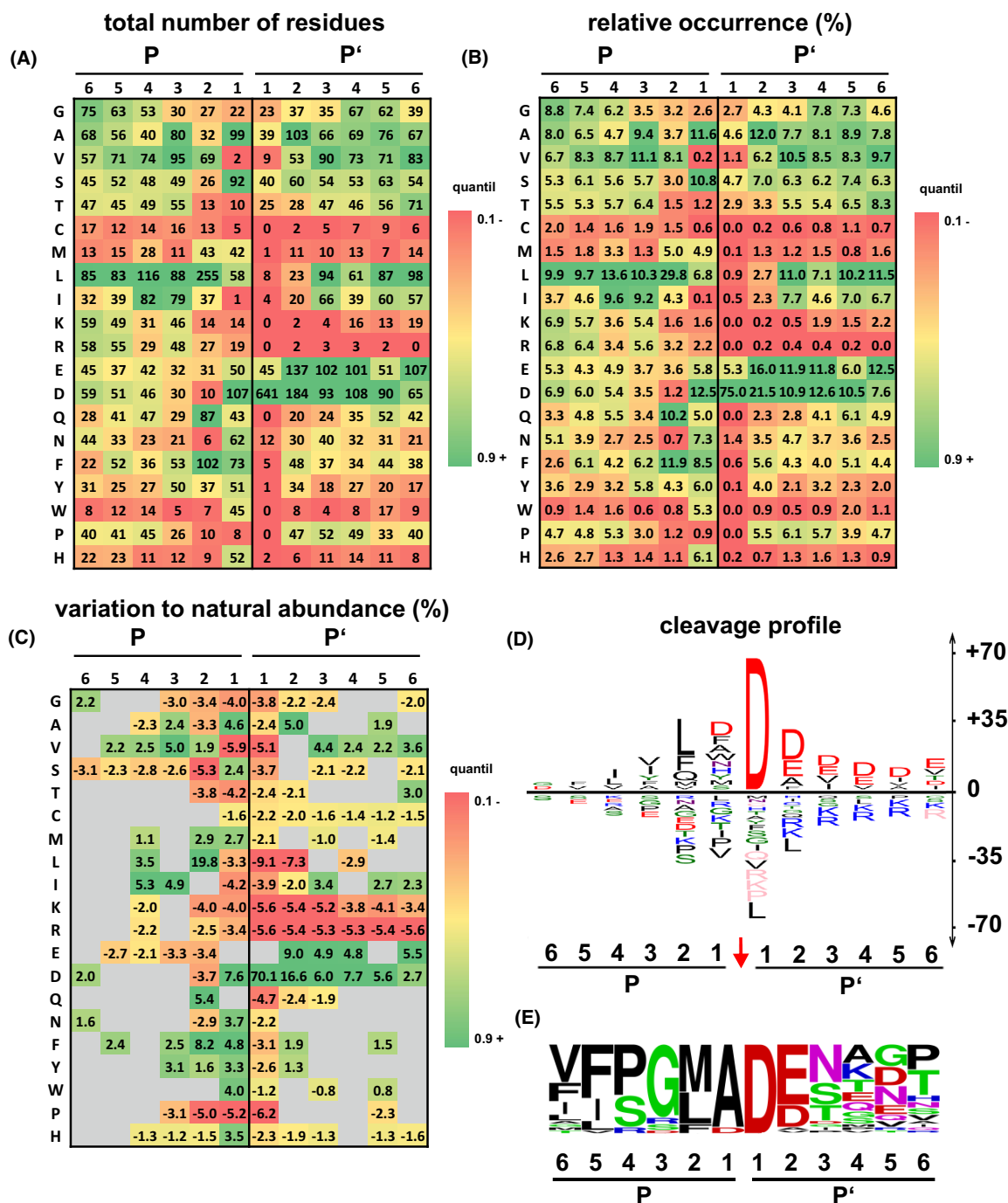
Fig. 1. Heatmaps including iceLogo of cleavage sites (*n* = 855) identified via N-TAILS and conservation of the ZP2 cleavage sites in mammals. Residues are normalized to their natural occurrence in the mouse proteome (*Mus musculus*). Distribution of residues in positions P6–P1 (non-prime side) and P1'–P6' (prime side). Displayed for each position are the total number of residues (A), their relative abundance in percent (B), and the relative abundance corrected by natural abundance in percentage points as heatmap (C) and as iceLogo (D). The deviations were determined with iceLogo [71]. 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. (E) Frequency plot of the primary ovastacin cleavage site in mammalian ZP2 (same species as in Fig. 4; alignment in Fig. S1).
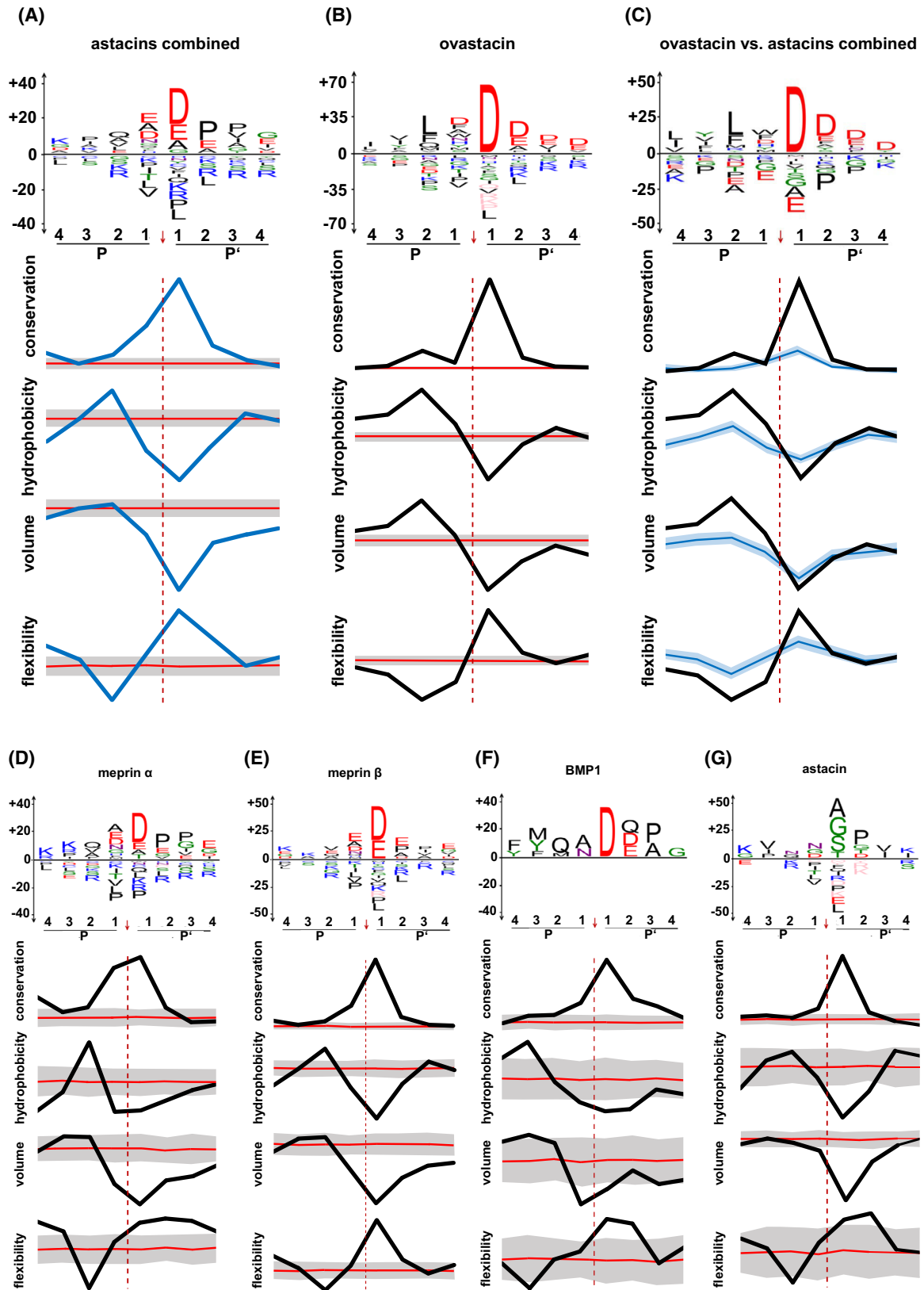
Fig. 2. Physicochemical properties of astacin proteases and ovastacin. Overview of the common cleavage preferences and physicochemical properties of all studied astacin proteases combined (A), ovastacin only (B) and comparatively of all astacin proteases combined versus ovastacin (C), meprin α only (D), meprin β only (E), BMP1 only (F), and astacin only (G). 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 [72] (upper graph), the hydrophobicity of the residues [73] (second graph from the top), the volume of the residues [74] (second graph from the bottom), and the flexibility of the residues [75] (bottom graph). The significance level is 95%. The error range displayed in gray/light blue. The diagrams allow a semi-quantitative evaluation, since iceLogo does not permit scale normalization.

site cleft, from left to right, that is, from non-prime to prime side, respectively. Arginine, lysine, and histidine are responsible for the formation of the hydrophilic, strongly positively charged regions in the S′ region (Fig. 3B top). Also displayed are the bulky, rigid, and hydrophobic residues leucine, isoleucine, valine, phenylalanine, tryptophan, methionine, and tyrosine (Fig. 3B bottom). 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 (Fig. 1D). This peptide was docked virtually into the catalytic cleft of ovastacin using CHIMERA (Fig. 3C). Although a clustering of more than two aspartates in the P′ region is not frequently observed in the N-TAILS (Table S1), aside of Arg264 there are also Arg174 and Lys175, whose positively charged, long and flexible side chains can be attracted by substrates with multiple negative charges. It is apparent that the aliphatic substrate residues Val (P3) and Leu (P2) occupy the region between Ile216, Ile219, and Phe214 on one side, Phe154, Val157 on the other side, and Trp191 at the base of the active site cleft (Fig. 3C). The aspartate residues of the prime side are, depending on the exact orientation of the side chains, proximate for interacting with the side chain of Arg264. Lys175 and Arg174 appear too distant for a direct interaction at least for P1′ and P2′. However, they are likely involved in stabilizing negative charges in position P3′–P4′. Generally, in astacin-like metalloproteases, the side chain in P1′ points directly toward the bottom of the active site cleft [37]. This is the reason for its size limitation. Other residues are in contact with the upper and lower rim of the cleft. As seen in Fig. 3C,D, the potential to bind negatively charged side chains is located on the right-hand side of the cleft.

High sequence identity suggests functional conservation of ovastacin within the theria

In order to assess the potential translatability of the results generated here with murine ovastacin, we analyzed the conservation and phylogeny for all major mammalian taxa, especially primates, artiodactyls, and carnivores including relevant domestic animals. Here, we primarily focused on the catalytic domain, since, in the mouse model the catalytic domain is released as a consequence of fertilization without CTR [17]. Accordingly, we considered the high variability or even a partial loss of CTR would probably not affect the cleavage specificity, *vide supra*. Ovastacins of all species analyzed share the typical characteristics required for catalysis (Fig. S3). The sequence identity analysis confirmed very high conservation of ovastacins within all analyzed eutherian species (Fig. 4A). The identity is at least 74% to the mouse, and 81% to the human ortholog. Within primates, the identity is even > 95% (outlined in green); within the cetartiodactyls, it is ≥ 88% (red); and within the carnivores, it is ≥ 89% (blue). Even with the Metatheria, here represented by the koala, the mean sequence identity is only slightly reduced, to about 70%. Fish alveolin (from medaka, *Oryzias latipes*) and frog XHE1 (*Xenopus* hatching enzyme; also termed UVS.2 [38]) were chosen for comparison, because these enzymes belong to the hatching enzyme (HE) subfamily of astacin proteinases. They are functional homologs of ovastacins with respect to their cleavage of the egg envelope (for review see [39]). However, they are clearly distinct from the ovastacin subfamily and may serve as outgroup with sequence identities to ovastacin of 35–40%. Even protostome hatching enzymes, such as *Astacus* embryonic astacin (AEA), reveal this conservation level [40,41]. If the CTR is included to the analysis, it becomes evident that this region is only slightly conserved in comparison. Only within primates this region exhibits significant conservation. Phylogenetic analysis also reflects the strong conservation of the catalytic domain, with a high level of confidence between the different mammalian taxa (Fig. 4B). Solely within the segregation of certain laurasiatheria and, to a lesser extent, eurarchontoglires and afrotheria, the phylogeny was less well resolved. Finally, we modeled the structures of the
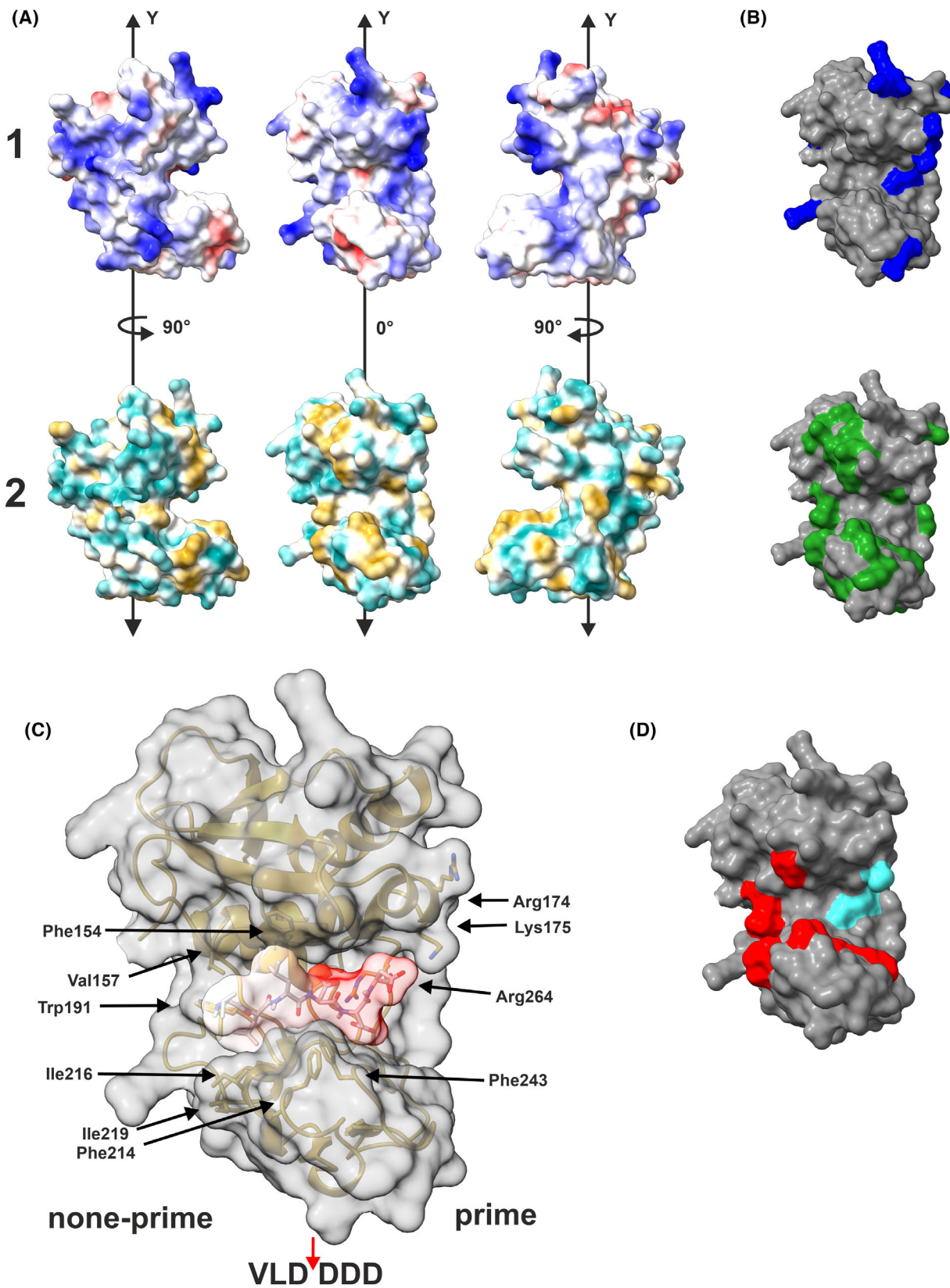
Fig. 3. Structural analysis of substrate interaction. Surface models of the catalytic domain of murine ovastacin (AAH64729.2; position 92–282) created with ALPHAFOLD2 using COLABFOLD [76]. (A1) Coulomb potential of the surface, highlighting areas with positive (blue, 10 kcal·(mol*e)$^{-1}$) and negative (red, −10 kcal·(mol*e)$^{-1}$) electrostatic potential. (A2) Surface areas of hydrophilic (cyan) and hydrophobic (yellow) residues according to the Kyte-Doolittle scale [86]. (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, Metl, and Tyr). (C) Synthetic peptide (Coulomb potential of the surface, *vide supra*) according to the most preferred residues in position P3 to P3′ (Fig. 1) and inserted into the catalytic cleft via CHIMERA; highlighted residues of ovastacin are discussed. (D) Hydrophobic (red) and hydrophilic (cyan) regions in the catalytic cleft.

catalytic domains of several ovastacins from different mammalian taxa, according to the phylogenetic tree (Fig. 4B) to allow conclusions regarding the conservation of function and substrate interaction among different species. Here, the ovastacins of all species reveal the almost identical astacin-like fold of the catalytic domain (Fig. 4C). Accordingly, XHE1 with only 40% identity reflects an almost identical fold. In order to obtain comparative conclusions about the substrate interaction, we have also considered the Coulomb potential of the surface (Fig. 3). Here, we observed a charge pattern very similar to mouse ovastacin in the other mammalian species. The bottom of the catalytic cleft is always uncharged (Fig. 4D). The positive charged residues are mainly located on the right half of the catalytic cleft (standard orientation). The residues Lys175 and Arg264, responsible for substrate recognition on the prime side, are conserved in ovastacins of all species, exept Arg174, which is only present in rodents. Apart from the preference for acidic residues in P1′, the anuran XHE1 exhibits a very different physicochemical surface pattern. These features suggest a cleavage specificity that is distinct from that of the mammalian ovastacins. The strong conservation of ovastacin including its surface properties in the catalytic cleft, as well as the completely conserved di-acidic cleavage site in ZP2 proteins (Fig. S1), indicates the overall conservation of its function within the eutheria.

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 29 cleavage sites (Tables 1 and 2) within 12 potentially fertilization-relevant proteins. Several examples are to be mentioned explicitly. β-1,4-galactosyltransferase (B4GALT1) is localized on the sperm membrane and has been shown to bind sugar side chains of ZP3,

thereby mediating additional attachment to the ZP [42]. Calreticulin (CALR3) is localized on the acrosomal surface of sperm [43] and also on the oolemma [44]. There is evidence for its involvement in oocyte signal transduction [45] and the establishment of the polyspermy block after its exocytosis from the oocyte [46]. Chaperonin containing TCP1 (Cct2) also localizes to the sperm surface and is directly involved in the binding of sperm to oocytes [47,48]. Matrix metalloproteinase 2 (MMP2) is associated with the acrosome and potentially contributes to the penetration of the ZP [49]. 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.
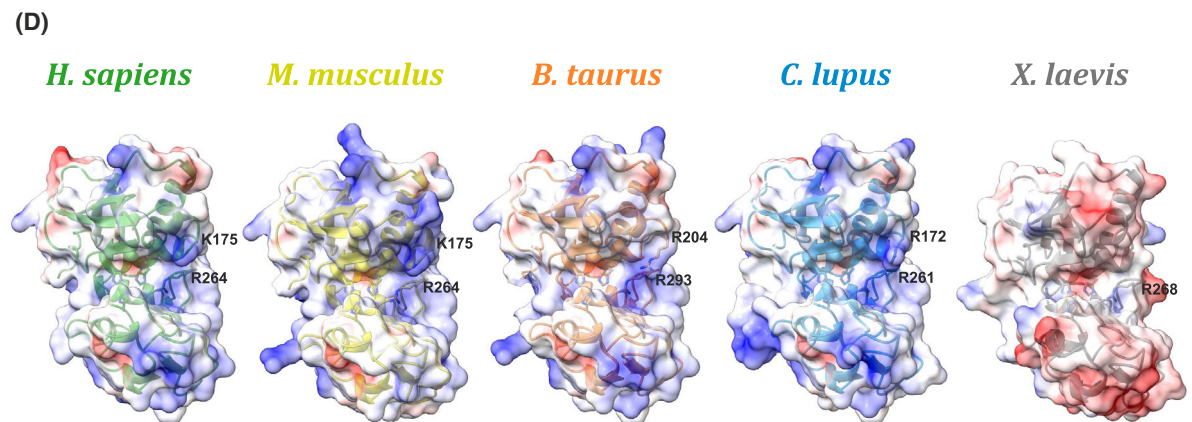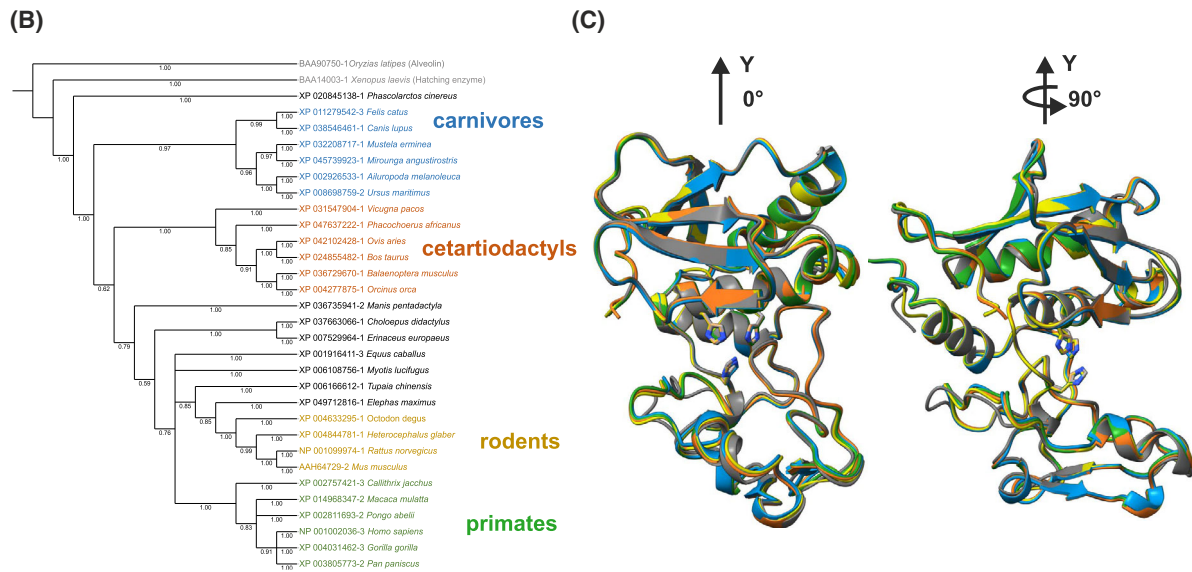
Discussion

Our N-TAILS analysis identified hundreds of ovastacin substrates and revealed the substrate specificity of ovastacin. In fact, ovastacin's cleavage specificity exhibits 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 (Table S1). However, due to the spatiotemporal restriction of ovastacin to oocytes and the functional sphere of the egg-sperm interaction, these are presumably of low physiological relevance. The restriction of substrates to the extracellular space or the context of fertilization reduces the number of physiologically relevant ovastacin substrates in the present study from 489 to 163 and 12, respectively. Owing that ovastacin expression is limited to oocytes, this rigorous restriction may be appropriate [50]. 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, for example, 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, for example, by potential pre-processing or interactions of

Fig. 4. Phylogenetic and interspecific analysis of ovastacin. (A) Identity matrix (% of identity) of 30 mammalian ovastacin (catalytic domain/ catalytic domain and C-terminal region as specified) orthologs representing all major therian taxa. In addition, astacin proteases from *Xenopus laevis* (hatching enzyme) and *Oryzias latipes* (alveolin) with a homologous cleavage of the egg envelope are displayed. Alignment was performed with CLUSTALO (Fig. S3). Data grayed out for species without CTR. (B) Consensus tree of the catalytic domain of ovastacins and homologous astacins presented in A. Computed using MR BAYES v3.2.7_0 via the NGPHYLOGENY tool [83] (100 000 generations, burn-in fraction 0.25) and displayed via ITOL [84] with specification of the credibility scores. (C, D) Models of the catalytic domain of ovastacins were created with ALPHAFOLD2 using COLABFOLD v1.5.2 [76]. (C) Superposition the catalytic domain of ovastacins from defined species with CHIMERA. (D) Individual representation of the models from C with Coulomb potential of the surface, highlighting areas with positive (blue, 10 kcal·(mol*e)$^{-1}$) and negative (red, -10 kcal·(mol*e)$^{-1}$) electrostatic potential.

substrates. This also explains why we were not able to identify the only physiologically proven substrate. ZP2 is not expressed by MEF cells [51]. Since the limited proteolysis results in only a few peptides with an ovastacin-specific neo-N-terminus (< 10% of the secretome), which therefore need to be enriched, the application of N-TAILS to germ cells is currently still elusive due to the limited quantity of available secretome. Optimization and 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 [47–49,52,53]. 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, B4GALT1 mediates the binding of sperm head to oligosaccharides of ZP3 until exocytosis of the acrosome [54]. Although this interaction is not strictly essential, unlike sperm binding to ZP2, its function is poorly understood. Proteolysis of B4GALT1 at S^{248} or D^{249}, 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 ovastacin cleavage sites in MMP2 (D^{208}, D^{209}, and D^{210}, Table 2) are located in a β-strand above the zinc-binding active site helix. From there, the chain enters the first fibronectin-like domain, which is an important exosite of MMP2 for binding protein substrates. Hence, cleavage within this region would most likely disrupt (protein-) substrate binding to MMP2 and abrogate its function, which might 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 [21,23,55], which underpins the importance of these potential substrates.

The cleavage preferences identified in this study displayed a high preference for aspartate in position P1′ and, to a lesser extent, for aspartate followed by glutamate in the P2′ to P4′ region. In addition, there is a preference for leucine in position P2 of the non-prime side. A similar preference for aspartate in position P1′ has been also described for the astacin proteases BMP1 (bone morphogenetic protein 1), meprin β, and LAST MAM (MAM-domain containing astacin from the horseshoe crab *Limulus polyphemus*) (Fig. 2E,F, Fig. S2). Substrate binding, analyzed by modeling, might be mediated by interactions with Arg264 and Lys175, and by the properties of hydrophobic (aromatic as well as aliphatic) residues of the non-prime side. In standard orientation, Phe154 is located at the upper left edge of the catalytic cleft, while Val157 and Trp191 form 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.

Both the very high conservation level of ovastacin, including its surface properties relevant for substrate interaction, and the presence of the typical di-acidic cleavage motif in all mammalian ZP2 suggest functional conservation of ZP2 cleavage and ZPH, respectively, in all mammalian species. This hypothesis is supported by strong conservation of the structural elements required for the inhibition of ovastacin in fetuin-B [24,30]. Perhaps for this reason, the highly variable residues on the prime side (starting with P3′) in ZP2 (Fig. 1E) could be of particular importance for substrate recognition and potentially for species-specific cleavage. The results presented here may thereby contribute to understanding a perplexing observation, published two decades ago by Dean and coworkers [56], who observed that mice carrying human ZP2 instead of endogenous mouse ZP2 were fertile *in vitro* and *in vivo*. However, only mouse, but not human sperm bound to these humanized mouse eggs and human ZP2 remained uncleaved by mouse ovastacin. As a possible reason for this observation, the authors considered humanized ZP on mouse oocytes would be glycosylated mouse specific and, thereby, triggered a binding mismatch for human,

Table 1. Search terms for identification of substrates associated with fertilization. Search terms in the DAVID bioinformatics database [69,70] 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 (%)
Zona pellucida receptor complex	GOTERM_CC_DIRECT	3.5E-3	4	0.8
Binding of sperm to *zona pellucida*	GOTERM_BP_DIRECT	1.4E-2	5	1.0
Estrogen signaling pathway	KEGG_PATHWAY	7.4E-2	8	1.6
In vitro fertilized eggs	UP_TISSUE	1.4E-2	13	2.7

Table 2. Substrates of ovastacin with proposed function in fertilization. In N-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 non-prime and prime side.

No.	Substrate	Uniprot accession number	*In vitro* fertilized eggs	*Zona pellucida* binding complex	Binding of sperm to *zona pellucida*	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	Endoplasmin (Hsp90b1)	P08113	−	−	−	+	802	226	QHIWES	DSNEFS
9	Heterogeneous nuclear ribonucleoprotein U (Hnrnpu)	Q8VEK3	+	−	−	−	800	190	GPTSLF	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	QVVACL	SVNDVF
								126	ACLSVN	DVFVIE
								162	ATDLLL	DDSLVS
								163	TDLLLD	DSLVSL
12	β-1.4-galactosyltransferase 1 (B4galt1)	P15535	−	−	+	−	339	248	YNCFVF	SDVDLI
								249	NCFVFS	DVDLIP

but not for mouse sperm. This would explain the binding failure of human sperm, but not the inability of mouse ovastacin to cleave human ZP2. However, the latter might be due to species-specific differences of the cleavage sites in mouse and human ZP2. Human ZP2 contains two positively charged lysyl side chains in P4′

and P7′ (SGLA*DDSKGTK), which are absent in mouse ZP2 (SRLA↓DENQNVS). Mouse ovastacin, on the other hand, carries an additional arginyl residue (Arg174, Fig. 3C) in the far S′ side, where human ovastacin features an uncharged glutamine side chain. This additional positive charge would presumably cause repulsion of the positive charge in human ZP2 and thereby prevent cleavage.

Distantly related astacins, as crayfish astacin and LAST, unlike ovastacin, do not prefer negatively charged residues in position P1′ (Fig. 2G, Fig. S2). However, we found great similarities between them in terms of distribution of physicochemical properties of substrates across the different positions relative to the cleavage site. All analyzed astacins share similar physicochemical properties in their substrate binding clefts, independent of their differing cleavage specificity. 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. This analysis was performed for seven astacins (astacin, BMP1, LAST, LAST MAM, meprin α, meprin β, and ovastacin), which include four of the six mammalian astacins (Fig. 2). Therefore, analysis of physicochemical properties allows for deeper understanding of cleavage preference and substrate selectivity of this group of proteases in general. This could contribute to higher precision when predicting potential protease substrates. Knowledge of the biochemical properties that distinguish the cleavage of substrates by ovastacin from the closely related meprins appears to be of utmost importance for the development of more specific and selective inhibitors. Such inhibitors would significantly reduce the amount of oocytes required for IVF [23,33], and additionally would 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 resolve the cleavage preference of ovastacin towards native substrate proteins within a complex secretome. This resulted in a globally applicable cleavage site signature ("ovastacin footprint") based on physiochemical properties of ovastacin. This footprint might serve as a useful tool to elucidate protease-substrate interactions in general. Additionally, this study allowed us to identify new ovastacin substrates with potential physiological relevance and deepens insight into the biochemical complexity of mammalian fertilization.

Materials and methods

Cell culture and secretome preparation

We cultured mouse embryonic fibroblasts (MEF), which originated from [57] and were kindly provided by O. Schilling (University Medical Center, Freiburg, Germany), in DMEM (Dulbecco's Modified Eagle Medium; Sigma-Aldrich, Taufkirchen, Germany) with 10% FBS and 100 U·mL^{-1} Penicillin/Streptomycin (37 °C, 5% CO$_2$) in adherent culture and passaged three times per week (detaching of adherent cells with 0.05% trypsin/EDTA solution). Cells were regularly tested to be free of mycoplasma.

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 Alanyl-glutamine (G8541; Sigma-Aldrich) and 100 U·mL^{-1} Penicillin/Streptomycin. Subsequently, we collected the supernatant, treated with protease inhibitors (cOmplete™ EDTA-free Protease Inhibitor Cocktail (04693132001; Sigma-Aldrich) or Pefabloc® SC (A154.1; Carl Roth, Karlsruhe, Germany)), and centrifuged for 3 min at 180 *g* at 4 °C. The supernatant was concentrated (Vivaspin 5 kDa MWCO) to 1–2 mg·mL^{-1} and stored on ice until further use.

Activation of pro-ovastacin

Heterologously expressed murine ovastacin was purified and activated with human plasmin (Haematologic Technologies, Essex Junction, VT, USA) as described [8,27]. Exemplarily, Fig. S4 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 or Pefabloc® SC. The inhibition was checked via a fluorogenic assay (as exemplified in Fig. S5).

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, at room temperature (RT). After incubation (70 °C, 60 min) the samples were alkylated with iodoacetamide (final concentration 15 mM, 30 min, RT, light-protected). Thereafter, proteins were precipitated with trichloroacetic acid [58] and resuspended (200 mM HEPES, pH 7.5) to a final concentration of approx. 2 mg·mL^{-1}. N-termini were labeled with light and heavy formaldehyde ($^{12}CH_2O$ in H_2O, and $^{13}CD_2O$ 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 [59]. Excess formaldehyde was captured by adjusting the sample to 50 mM Tris/HCl pH 7 and incubation for 4 h (37 °C). Proteins were precipitated with 9× volume acetone and 1× volume methanol (each −20 °C) ON at −80 °C and centrifuged (4500 **g** for 60 min at 4 °C). The pellet was washed four times with 1 mL methanol following centrifugation (4500 **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, Lakewood, NJ, USA) (37 °C, overnight). After desalting via C18 Sep-Pak columns (Waters Corporation, Milford, MA, USA), we enriched the labeled peptide N-termini by negative selection using a polymer (hyperbranched polyglycerol-aldehyde, HPG-ALD (University of British Columbia, Vancouver, Canada)). 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 r.p.m.). 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 (1× 300 μL 100 mM ammonium bicarbonate, 1× 300 μL 2 M guanidine hydrochloride). The samples were desalted via C18 columns and fractioned via strong cation exchange chromatography (Stagetips) [60]. Afterward, the samples were desalted again. Peptide concentrations for mass spectrometric measurement were determined by a NanoDrop ND-1000 spectrophotometer (PeqLab, Erlangen, Germany). Protein or peptide concentration in the course preparation was checked by Bradford- or bicinchoninic acid-assays [61,62].

TAILS analysis by mass spectrometry

Mass spectrometric analysis was performed using LC–MS (liquid chromatography/mass spectrometry) on a Synapt G2-S HDMS (Waters Corporation) in positive ion mode. Chromatographic separation of peptides was conducted on a nanoAcquity UPLC system (Waters Corporation). Peptides were separated using a C18 reversed-phase analytical separation column (HSS-T3, 75 μm × 250 mm, 1.8 μm particle size) (Waters Corporation). 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^{-1}). Afterward, the column was re-equilibrated with phase B (90% (v/v), 600 nL·min^{-1}, 5 min). Total measurement time was 120 min. Data were acquired using HD-DDA [63]. [Glu^1]-fibrinopeptide was used as lock mass at 100 fmol·μL^{-1} 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.bioinfor.com/) [64–66]. 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 di-methylation, 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 analog 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., "non-prime side") were determined by bioinformatic alignment of the sequences with the UNIPROT database. Cleavage sites < 10 residues distant from N- or C-terminus were not 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, for example, 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) [67,68]. This is exemplified for one of the replicas in Figs S6 and S7. The identified clusters from both biological replicates, with the typical acidic residues on the prime side [35], 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 [69,70].

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 abundancies of residues was performed using iceLogo [71]. In order to 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 [72] 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 [73], the "Residue Volume" template to determine residue volume [74], and the "Average flexibility indices" template to determine residue flexibility [75].

In silico structure prediction

We created models of the catalytic domains of ovastacin using COLABFOLD v1.5.2 (https://colab.research.google.com) [76] based on ALPHAFOLD2 [77]. The best ranked model for each ovastacin was used for further analysis. Models were visualized using UCSF CHIMERA 1.13.1 or CHIMERAX software [78]. Interaction between ovastacin and the identified theoretical optimal substrate was simulated via the CLUSPRO software (https://cluspro.org) [79–81] using the generated model.

Phylogenetic analysis

Selected ovastacin or ZP2 orthologs deposited in the National Library of Medicine were aligned with CLUSTALO [82]. The sequence matrix was generated using BIOEDIT. From the alignment, a consensus tree was calculated using MR BAYES via the NGPHYLOGENY tool (https://ngphylogeny.fr/tools/tool/281/form) [83] (100 000 generations, burn-in fraction 0.25) and displayed via ITOL [84]. The frequency plot of ZP2 was generated using WEBLOGO [85] based on the alignment generated with CLUSTALO.

Acknowledgments

Conflict of interest

The authors declare no conflict of interest.

Author contributions

MF, UD, ST, WS, and HK were involved in conceptualization and design of the study. MF, UD, ST, WS, and HK involved in establishing methods. MF, UD, MŁ, NW, CB, ST, WS, and HK were involved in acquisition or analysis and interpretation of data. CB and HK were involved in funding acquisition. MF, NW, ST, CB, WS, and HK were involved in drafting the article or revising it critically for important intellectual content. Final approval of the version to be published and agreement to be accountable for all aspects of the work (all authors).

Peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/febs.16954.

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).

References

1 Burkart AD, Xiong B, Baibakov B, Jiménez-Movilla M & 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 LJD, Polakoski KL & Williams WL (1972) Properties of a proteolytic enzyme from rabbit sperm acrosomes. *Biol Reprod* **6**, 30–39.

3 Honda A, Yamagata K, Sugiura S, Watanabe K & 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 FA, Mondéjar I, Grullón L & Romar R (2012) Oocytes use the plasminogen-plasmin system to remove supernumerary spermatozoa. *Hum Reprod* **27**, 1985–1993.

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 & 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 *et al.* (2020) Acrosin is essential for sperm penetration through the zona pellucida in hamsters. *Proc Natl Acad Sci USA* **117**, 2513–2518.

8 Kuske M, Floehr J, Yiallouros I, Michna T, Jahnen-Dechent W, Tenzer S, Stöcker W & 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 ND & Salvesen G, eds (2013) Handbook of Proteolytic Enzymes 1, 2, 3. 3rd edn. Elsevier, London.

10 Rawlings ND, Waller M, Barrett AJ & Bateman A (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **42**, D503–D509.

11 Mascarenhas MN, Flaxman SR, Boerma T, Vanderpoel S & Stevens GA (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 FX & Stöcker W (2023) Structural and evolutionary insights into astacin metallopeptidases. *Front Mol Biosci* **9**, 1–20.

14 Gomis-Rüth FX, Trillo-Muyo S & Stöcker W (2012) Functional and structural insights into astacin metallopeptidases. *Biol Chem* **393**, 1027–1041.

15 Xiong B, Zhao Y, Beall S, Sadusky AB & 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.

16 Gahlay G, Gauthier L, Baibakov B, Epifano O & Dean J (2010) Gamete recognition in mice depends on the cleavage status of egg's zona pellucida protein. *Science* **329**, 216–219.

17 Körschgen H, Kuske M, Karmilin K, Yiallouros I, Balbach M, Floehr J, Wachten D, Jahnen-dechent W & Stöcker W (2017) Intracellular activation of ovastacin mediates pre-fertilization hardening of the zona pellucida. *Mol Hum Reprod* **23**, 607–616.

18 Nozawa K, Satouh Y, Fujimoto T, Oji A & Ikawa M (2018) Sperm-borne phospholipase C zeta-1 ensures monospermic fertilization in mice. *Sci Rep* **8**, 1315.

19 Bianchi E, Doe B, Goulding D & Wright GJ (2014) Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487.

20 Bianchi E & Wright GJ (2014) Izumo meets Juno: preventing polyspermy in fertilization. *Cell Cycle* **13**, 2019–2020.

21 Maddirevula S, Coskun S, Al-Qahtani M, Omar Q & Saad A (2021) ASTL is mutated in female infertility. *Hum Genet* **141**, 49–54.

22 Dietzel E, Wessling J, Floehr J, Schäfer C, Ensslen S, Denecke B, Rösing B, Neulen J, Veitinger T, Spehr M et al. (2013) Fetuin-B, a liver-derived plasma protein is essential for fertilization. *Dev Cell* **25**, 106–112.

23 Dietzel E, Floehr J, Van de Leur E, Weiskirchen R & Jahnen-Dechent W (2016) Recombinant fetuin-B protein maintains high fertilization rate in cumulus cell-free mouse oocytes. *Mol Hum Reprod* **23**, 25–33.

24 Cuppari A, Körschgen H, Fahrenkamp D, Schmitz C, Guevara T, Karmilin K, Kuske M, Olf M, Dietzel E, Yiallouros I et al. (2019) Structure of mammalian plasma fetuin-B and its mechanism of selective metallopeptidase inhibition. *IUCrJ* **6**, 317–330.

25 Guevara T, Körschgen H, Cuppari A, Schmitz C, Kuske M, Yiallouros I, Floehr J, Jahnen-Dechent W, Stöcker W & Gomis-Rüth FX (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.

26 Eckhard U, Körschgen H, von Wiegen N, Stöcker W & Gomis-rüth FX (2021) The crystal structure of a 250-kDa heterotetrameric particle explains inhibition of sheddase meprin β by endogenous fetuin-B. *Proc Natl Acad Sci USA* **118**, 1–6.

27 Karmilin K, Schmitz C, Kuske M, Körschgen H, Olf M, Meyer K, Hildebrand A, Felten M, Fridrich S, Yiallouros I et al. (2019) Mammalian plasma fetuin-B is a selective inhibitor of ovastacin and meprin metalloproteinases. *Sci Rep* **9**, 546.

28 Turk V & Bode W (1991) The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett* **285**, 213–219.

29 Kleifeld O, Doucet A, Auf Dem Keller U, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN & Overall CM (2010) Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* **28**, 281–288.

30 Schmitz C, Sadr SZ, Körschgen H, Kuske M, Schoen J, Stöcker W, Jahnen-Dechent W & Floehr J (2021) The E-modulus of the oocyte is a non-destructive measure of zona pellucida hardening. *Reproduction* **162**, 259–266.

31 Floehr J, Dietzel E, Schmitz C, Chappell A & Jahnen-Dechent W (2017) Down-regulation of the liver-derived plasma protein fetuin-B mediates reversible female infertility. *Mol Hum Reprod* **23**, 34–44.

32 Avella MA, Baibakov B & Dean J (2014) A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. *J Cell Biol* **205**, 801–809.

33 Körschgen H, Jäger C, Tan K, Buchholz M, Stöcker W & 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.

34 Schechter I & Berger A (1967) On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **27**, 157–162.

35 Becker-Pauly C, Barré O, Schilling O, Auf dem Keller U, Ohler A, Broder C, Schütte A, Kappelhoff R, Stöcker W & Overall CM (2011) Proteomic analyses

reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. *Mol Cell Proteomics* **10**, M111.009233.

36 Gomis-Rüth FX, Botelho TO & Bode W (2012) A standard orientation for metallopeptidases. *Biochim Biophys Acta* **1824**, 157–163.

37 Grams F, Dive V, Yiotakis A, Yiallouros I, Vassiliou S, Zwilling R, Bode W & Stöcker W (1996) Structure of astacin with a transition-state analogue inhibitor. *Nat Struct Mol Biol* **3**, 671–675.

38 Sato SM & Sargent TD (1990) Molecular approach to dorsoanterior development in *Xenopus laevis*. *Dev Biol* **137**, 135–141.

39 Nagasawa T, Kawaguchi M, Nishi K & Yasumasu S (2022) Molecular evolution of hatching enzymes and their paralogous genes in vertebrates. *BMC Ecol Evol* **22**, 9.

40 Quesada V, Sánchez LM, Alvarez J & López-Otín C (2004) Identification and characterization of human and mouse ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. *J Biol Chem* **279**, 26627–26634.

41 Quesada V, Sánchez LM, Álvarez J & López-Otín C (2019) Withdrawal: identification and characterization of human and mouse ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. *J Biol Chem* **294**, 1432.

42 Miller DJ, Macek MB & Shur BD (1992) Complementarity between sperm surface beta-1,4-galactosyltransferase and egg-coat ZP3 mediates sperm-egg binding. *Nature* **357**, 589–593.

43 Nakamura M, Moriya M, Baba T, Michikawa Y, Yamanobe T, Arai K, Okinaga S & Kobayashi T (1993) An endoplasmic reticulum protein, calreticulin, is transported into the acrosome of rat sperm. *Exp Cell Res* **205**, 101–110.

44 Calvert ME, Digilio LC, Herr JC & Coonrod SA (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.

45 Tutuncu L, Stein P, Ord TS, Jorgez CJ & Williams CJ (2004) Calreticulin on the mouse egg surface mediates transmembrane signaling linked to cell cycle resumption. *Dev Biol* **270**, 246–260.

46 Saavedra MD, Mondéjar I, Coy P, Betancourt M, González-Márquez H, Jiménez-Movilla M, Avilés M & Romar R (2014) Calreticulin from suboolemmal vesicles affects membrane regulation of polyspermy. *Reproduction* **147**, 369–378.

47 Redgrove KA, Anderson AL, Dun MD, McLaughlin EA, O'Bryan MK, Aitken RJ & 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.

48 Dun MD, Smith ND, Baker MA, Lin M, Aitken RJ & Nixon B (2011) The chaperonin containing TCP1 complex (CCT/TRiC) is involved in mediating sperm-oocyte interaction. *J Biol Chem* **286**, 36875–36887.

49 Ferrer M, Rodriguez H, Zara L, Yu Y, Xu W & 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.

50 Pires ES, Hlavin C, Macnamara E, Ishola-Gbenla K, Doerwaldt C, Chamberlain C, Klotz K, Herr AK, Khole A, Chertihin O *et al.* (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* **242**, 1405–1426.

51 Sarkar P, Randall SM, Muddiman DC & Rao BM (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.

52 van Gestel RA, Brewis IA, Ashton PR, Brouwers JF & Gadella BM (2007) Multiple proteins present in purified porcine sperm apical plasma membranes interact with the zona pellucida of the oocyte. *Mol Hum Reprod* **13**, 445–454.

53 Elliott RMA, Lloyd RE, Fazeli A, Sostaric E, Georgiou AS, Satake N, Watson PF & Holt WV (2009) Effects of HSPA8, an evolutionarily conserved oviductal protein, on boar and bull spermatozoa. *Reproduction* **137**, 191–203.

54 Miller DJ, MacEk MB & Shur BD (1992) Complementarity between sperm surface β-l,4-galactosyl-transferase and egg-coat ZP3 mediates sperm-egg binding. *Nature* **357**, 589–593.

55 Floehr J, Dietzel E, Neulen J, Rösing B, Weissenborn U & 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.

56 Rankin TL, Coleman JS, Epifano O, Hoodbhoy T, Turner SG, Castle PE, Lee E, Gore-Langton R & Dean J (2003) Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Dev Cell* **5**, 33–43.

57 Reinheckel T, Hagemann S, Dollwet-Mack S, Martinez E, Lohmüller T, Zlatkovic G, Tobin DJ, Maas-Szabowski N & Peters C (2005) The lysosomal cysteine protease cathepsin L regulates keratinocyte proliferation by control of growth factor recycling. *J Cell Sci* **118**, 3387–3395.

58 Sambrook J, Fritsch E & Maniatis T (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

59 Kleifeld O, Doucet A, Prudova A, Auf dem Keller U, Gioia M, Kizhakkedathu JN & Overall CM (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **6**, 1578–1611.

60 Rappsilber J, Mann M & Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896–1906.

61 Bradford MM (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.

62 Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ & Klenk DC (1985) Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**, 76–85.

63 Helm D, Vissers JPC, Hughes CJ, Hahne H, Ruprecht B, Pachl F, Grzyb A, Richardson K, Wildgoose J, Maier SK *et al.* (2014) Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* **13**, 3709–3715.

64 Tran NH, Zhang X, Xin L, Shan B & Li M (2017) De novo peptide sequencing by deep learning. *Proc Natl Acad Sci USA* **114**, 8247–8252.

65 Tran NH, Qiao R, Xin L, Chen X, Liu C, Zhang X, Shan B, Ghodsi A & Li M (2019) Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat Methods* **16**, 63–66.

66 Tran NH, Rahman MZ, He L, Xin L, Shan B & Li M (2016) Complete de novo assembly of monoclonal antibody sequences. *Sci Rep* **6**, 31730.

67 Andreatta M, Alvarez B & Nielsen M (2017) GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res* **45**, W458–W463.

68 Andreatta M, Lund O & Nielsen M (2013) Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. *Bioinformatics* **29**, 8–14.

69 Huang DW, Sherman BT & Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1–13.

70 Huang DW, Sherman BT & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57.

71 Colaert N, Helsens K, Martens L, Vandekerckhove J & Gevaert K (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* **6**, 786–787.

72 Henikoff S & Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* **89**, 10915–10919.

73 Nozaki Y & Tanford C (1971) The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J Biol Chem* **246**, 2211–2217.

74 Bigelow CC (1967) On the average hydrophobicity of proteins and the relation between it and protein structure. *J Theor Biol* **16**, 187–211.

75 Bhaskaran R & Ponnuswamy PK (1988) Positional flexibilities of amino acid residues in globular proteins. *Int J Pept Protein Res* **32**, 241–255.

76 Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S & Steinegger M (2022) ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679–682.

77 Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.

78 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF chimera? A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.

79 Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D & Vajda S (2017) The ClusPro web server for protein–protein docking. *Nat Protoc* **12**, 255–278.

80 Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR & Vajda S (2013) How good is automated protein docking? *Proteins* **81**, 2159–2166.

81 Vajda S, Yueh C, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR & Kozakov D (2017) New additions to the ClusPro server motivated by CAPRI. *Proteins* **85**, 435–444.

82 Sievers F & Higgins DG (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135–145.

83 Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S & Gascuel O (2019) NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* **47**, W260–W265.

84 Letunic I & Bork P (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293–W296.

85 Crooks GE, Hon G, Chandonia J-M & Brenner SE (2004) WebLogo: a sequence logo generator: figure 1. *Genome Res* **14**, 1188–1190.

86 Kyte J & Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105–132.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Alignment of selected therian ZP2 orthologs.
Fig. S2. Physicochemical properties of astacin proteases.
Fig. S3. Alignment of therian ovastacin orthologs.
Fig. S4. Purification and activation of ovastacin.
Fig. S5. Evaluation of plasmin inhibition.
Fig. S6. Cluster analysis of cleavage sites identified via N-TAILS Part 1.
Fig. S7. Cluster analysis of cleavage sites identified via N-TAILS Part 2.
Table S1. Peptides identified via N-TAILS analysis.

Table S1. Peptides identified via N-TAILS analysis.