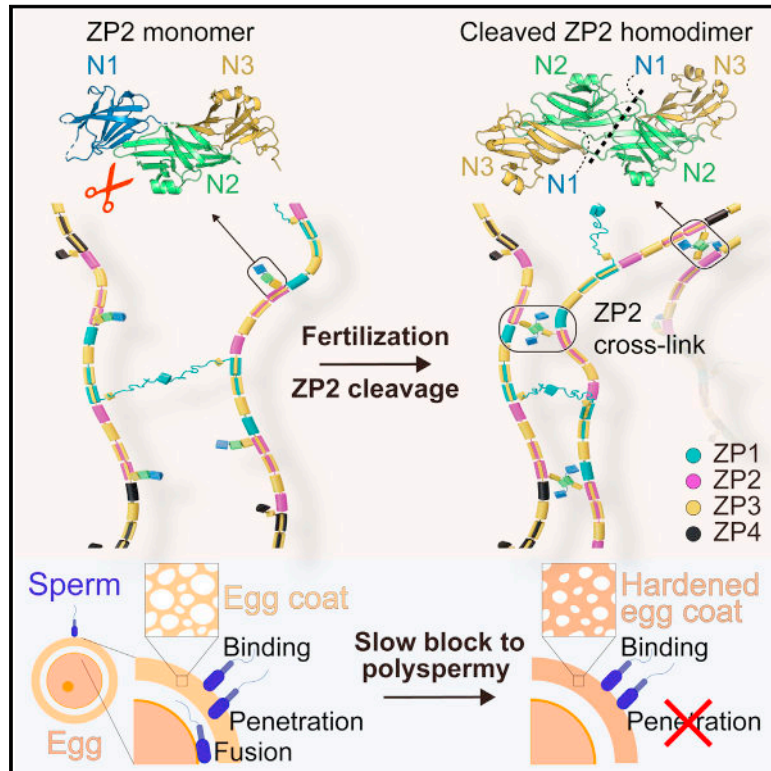


# ZP2 cleavage blocks polyspermy by modulating the architecture of the egg coat

## Graphical abstract



## Authors

Shunsuke Nishio, Chihiro Emori,  
Benjamin Wiseman, ...,  
Shigeki Yasumasu, Masahito Ikawa,  
Luca Jovine

## Correspondence

luca.jovine@ki.se

## In brief

Post-fertilization cleavage of the egg zona pellucida (ZP) subunit ZP2 triggers its oligomerization, which in turn cross-links the filaments that make up the ZP. By tightening its mesh, this hardens the egg coat matrix and hinders the penetration of further sperm, thereby ensuring a permanent block to polyspermy.

## Highlights

- Post-fertilization cleavage of zona pellucida (ZP) subunit ZP2 makes it oligomerize
- Premature ZP2 oligomerization causes infertility by impairing sperm penetration of ZP
- A native egg coat fragment structure sheds light on heteromeric ZP filament assembly
- Cleaved ZP2 cross-links ZP filaments to harden the egg coat and prevent polyspermy



Article

# ZP2 cleavage blocks polyspermy by modulating the architecture of the egg coat

Shunsuke Nishio,<sup>1,11</sup> Chihiro Emori,<sup>2,3</sup> Benjamin Wiseman,<sup>1,12</sup> Dirk Fahrenkamp,<sup>1,13</sup> Elisa Dioguardi,<sup>1,14</sup> Sara Zamora-Caballero,<sup>1,15</sup> Marcel Bokhove,<sup>1,16</sup> Ling Han,<sup>1</sup> Alena Stsiapanava,<sup>1</sup> Blanca Algarra,<sup>1</sup> Yonggang Lu,<sup>2,3</sup> Mayo Kodani,<sup>2,4</sup> Rachel E. Bainbridge,<sup>5,17</sup> Kayla M. Komondor,<sup>5</sup> Anne E. Carlson,<sup>5</sup> Michael Landreh,<sup>6,7</sup> Daniele de Sanctis,<sup>8</sup> Shigeki Yasumasu,<sup>9</sup> Masahito Ikawa,<sup>2,3,4,10</sup> and Luca Jovine<sup>1,18,\*</sup>

<sup>1</sup>Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden

<sup>2</sup>Department of Experimental Genome Research, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, Japan

<sup>3</sup>Immunology Frontier Research Center, Osaka University, Suita, Osaka, Japan

<sup>4</sup>Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka, Japan

<sup>5</sup>Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA, USA

<sup>6</sup>Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

<sup>7</sup>Department of Cell and Molecular Biology, Uppsala University, 75124 Uppsala, Sweden

<sup>8</sup>ESRF - The European Synchrotron, Grenoble, France

<sup>9</sup>Department of Materials and Life Sciences, Faculty of Science and Technology, Sophia University, Tokyo, Japan

<sup>10</sup>Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Osaka, Japan

<sup>11</sup>Present address: Institute of Fermentation Sciences (IFeS), Faculty of Food and Agricultural Sciences, Fukushima University, Fukushima, Japan

<sup>12</sup>Present address: Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

<sup>13</sup>Present address: Crelux GmbH, Martinsried, Germany

<sup>14</sup>Present address: Chiesi Pharma AB, Stockholm, Sweden

<sup>15</sup>Present address: Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain

<sup>16</sup>Present address: Japan Synchrotron Radiation Research Institute (JASRI), Hyogo, Japan

<sup>17</sup>Present address: National Institute of Environmental Health Sciences, Durham, NC, USA

<sup>18</sup>Lead contact

\*Correspondence: [luca.jovine@ki.se](mailto:luca.jovine@ki.se)

<https://doi.org/10.1016/j.cell.2024.02.013>

## SUMMARY

Following the fertilization of an egg by a single sperm, the egg coat or zona pellucida (ZP) hardens and polyspermy is irreversibly blocked. These events are associated with the cleavage of the N-terminal region (NTR) of glycoprotein ZP2, a major subunit of ZP filaments. ZP2 processing is thought to inactivate sperm binding to the ZP, but its molecular consequences and connection with ZP hardening are unknown. Biochemical and structural studies show that cleavage of ZP2 triggers its oligomerization. Moreover, the structure of a native vertebrate egg coat filament, combined with AlphaFold predictions of human ZP polymers, reveals that two protofilaments consisting of type I (ZP3) and type II (ZP1/ZP2/ZP4) components interlock into a left-handed double helix from which the NTRs of type II subunits protrude. Together, these data suggest that oligomerization of cleaved ZP2 NTRs extensively cross-links ZP filaments, rigidifying the egg coat and making it physically impenetrable to sperm.

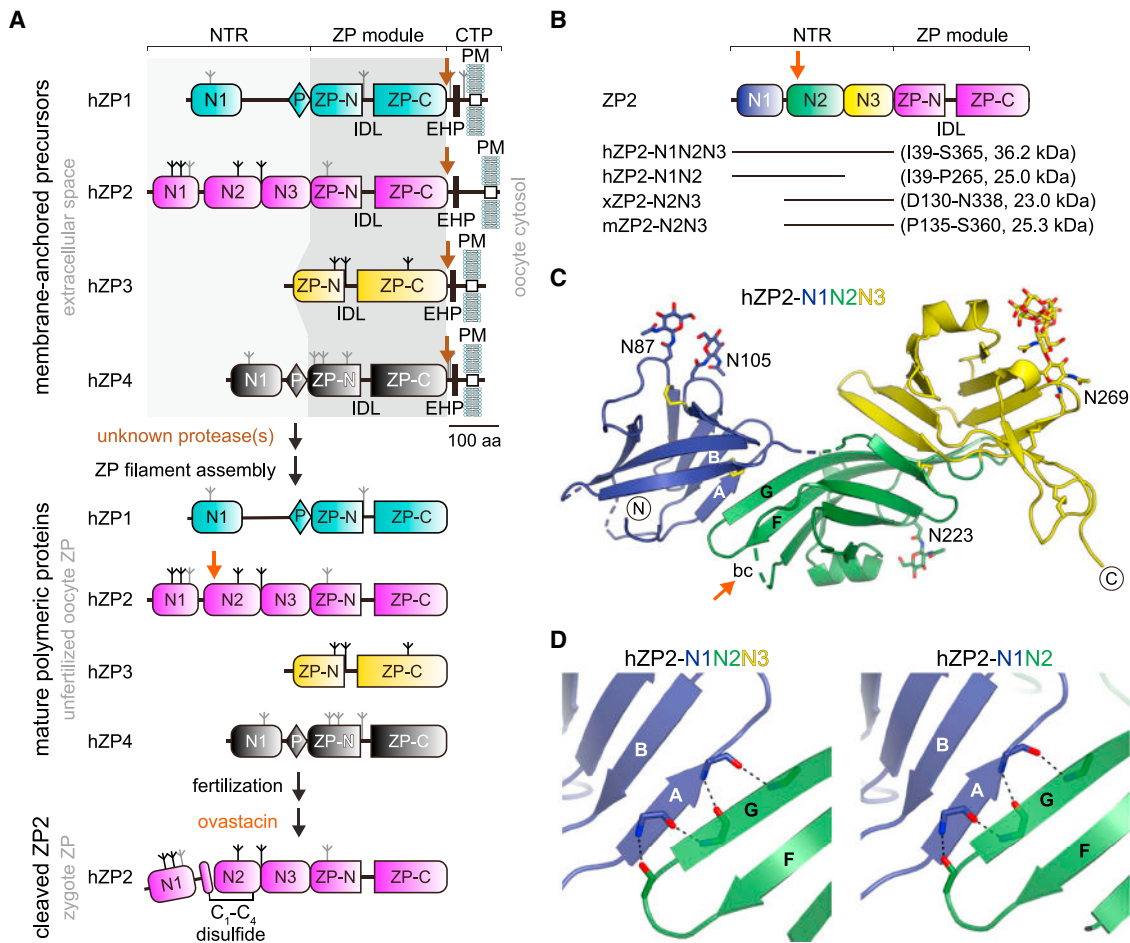
## INTRODUCTION

By interfering with the generation of a diploid zygote, fertilization of an egg by more than one sperm is lethal for most animals, including humans. Prevention of the fusion of multiple sperm with the egg plasma membrane is achieved using different mechanisms and, in marine-dwelling species, is often only transient. On the other hand, an evolutionarily conserved and permanent block to polyspermy is established by the irreversible modification of the egg coat, a specialized extracellular matrix called the zona pellucida (ZP) in mammals and vitelline envelope (VE) in

non-mammals.<sup>1,2</sup> In mammals and anuran amphibians, this process depends on site-specific cleavage of egg coat subunit ZP2 by a cortical granule (CG) activity released from the egg after fertilization,<sup>3,4</sup> which was identified as zinc metalloprotease ovastacin (Ovst) in the mouse.<sup>5</sup>

The human ZP (hZP) is composed of ZP proteins (hZP1–4; [Figure 1A](#)) that have homologs in other vertebrates and structurally similar counterparts in invertebrates. All of these molecules belong to the ZP module protein superfamily, which not only includes egg coat filament subunits from mollusk to mammal (such as abalone VE receptor for lysin [VERL] and mouse ZP [mZP]





**Figure 1. Crystal structures of the functional N-terminal region (NTR) of mature human ZP2 in its uncleaved, monomeric state**

(A) Multidomain architecture of hZP filament components, in relation to their proteolytic processing during egg coat assembly and after fertilization. P, trefoil domain; PM, plasma membrane. White squares, transmembrane domains; inverted black and gray tripods, experimentally supported and predicted N-glycans; brown arrows, CFCS; orange arrow, ZP2 post-fertilization cleavage site.

(B) Experimentally determined ZP2 structures from human (h), *Xenopus* (x), and mouse (m).

(C) Crystal structure of hZP2-N1N2N3. Disordered loops are represented by dashed lines, with the ZP2-N2 bc loop targeted by ovastacin indicated by an orange arrow.

(D) Conserved intramolecular ZP2-N1/ZP2-N2 domain interface in the crystal structures of hZP2-N1N2N3 and hZP2-N1N2. Dashed lines indicate hydrogen bonds.

See also Figures S1 and S2.

components mZP1–3) but also hundreds of other extracellular molecules from multicellular eukaryotes with highly different biological functions (for example, urinary glycoprotein uromodulin [UMOD]).<sup>6–8</sup> The functional diversity of ZP module proteins is imparted by the highly mosaic architecture of their N-terminal region (NTR), which generally contains different types and numbers of protein domains and is followed by a single C-terminal ZP module (previously referred to as ZP domain). This conserved element, consisting of two distinct immunoglobulin (Ig)-like domains (ZP-N and ZP-C) connected by an interdomain linker (IDL),<sup>9,10</sup> is both necessary and sufficient for protein polymerization into filaments.<sup>11</sup> Studies on mammalian ZP subunits and UMOD showed that incorporation of their mature forms into the respective nascent filaments depends on the dissociation of a C-terminal propeptide (CTP) that includes a polymeriza-

tion-blocking external hydrophobic patch (EHP).<sup>12,13</sup> This is triggered by cleavage of their membrane-bound precursors at a site located after ZP-C and called consensus furin cleavage site (CFCS) in ZP proteins (Figure 1A).<sup>14,15</sup>

While ZP3 essentially consists of a ZP module that includes a small ZP-C subdomain and has been implicated in sperm binding,<sup>9,16</sup> the different functions of other ZP subunits are determined by the respective NTRs, which are thought to protrude from the core of ZP filaments.<sup>10</sup> A single isolated ZP-N domain (ZP1-N1) in the NTR of ZP1, which also contains a signature trefoil domain immediately before the ZP module, cross-links ZP filaments by forming intermolecularly disulfide-bonded homodimers from chicken to human.<sup>17,18</sup> On the other hand, the corresponding ZP4-N1 domain of ZP4, a paralog that shares the same domain organization as ZP1 but is absent in the mouse

and does not homodimerize in human,<sup>18,19</sup> contributes to sperm binding in bovine.<sup>20</sup> The NTR that is arguably the most important for the regulation of gamete interaction is, however, that of ZP2. This part of the protein, which is not involved in ZP2 incorporation into the ZP,<sup>11</sup> was predicted to contain three additional copies of the ZP-N domain arranged in tandem (ZP2-N1–3).<sup>21,22</sup> This hypothesis was corroborated by the crystal structure of mZP2-N1,<sup>23</sup> a domain that is present in mature mammalian ZP2 but proteolytically removed during maturation of anuran ZP2.<sup>4</sup> Despite this difference, in both amphibians and mammals the following ZP2-N2 domain harbors a conserved recognition site for the CG protease,<sup>4,24</sup> whose action generates two ZP2 protein fragments that remain covalently attached through an intramolecular disulfide bond (Figure 1A).<sup>3,24,25</sup> Interestingly, the NTR of mollusk VERL consists of a ZP-N repeat array that binds sperm,<sup>23</sup> and tandem ZP-N domains are also found in the NTRs of non-mammalian egg coat components such as the ZPAX subunits of the avian, amphibian, and fish VE (fVE).<sup>26</sup>

Post-fertilization cleavage of ZP2 was initially hypothesized to abolish sperm binding by either altering the conformation of ZP proteins or somehow modifying the overall architecture of the egg coat,<sup>24,25</sup> whereas in the current model of mammalian gamete recognition, it is postulated to act by inactivating a direct sperm-binding site on ZP2-N1.<sup>27,28</sup> At the same time, ZP2 processing has long been connected with ZP hardening, a global physicochemical change of the egg coat matrix associated with significant increases in stiffness, resistance to proteolytic digestion, and filament density.<sup>25,29</sup> Although gamete recognition and ZP hardening are intimately connected with the NTR and ZP module moieties of ZP2, there is no structural information on the key ZP2-N2 domain and the possible consequences of its cleavage. Similarly, except for the suggestion that ZP1 also incorporates into ZP filaments via its ZP module,<sup>10</sup> the current model of ZP structure has remained essentially the same as proposed almost 40 years ago, where filaments made up of alternating ZP2/ZP3 dimers are cross-linked by pairs of ZP1 molecules.<sup>8,30,31</sup> This is because, although recent cryoelectron microscopy (cryo-EM) studies showed that the homomeric filament of UMOD adopts a previously unobserved architecture resulting from major conformational changes of the ZP module during polymerization,<sup>32,33</sup> it is still unknown how egg coat subunits form heteromeric polymers. In this work, we combine studies of the ZP2 NTR and the ZP filament to bridge the aforementioned observations and reveal how ZP2 cleavage blocks polyspermy by altering the overall structure of the egg coat.

## RESULTS

### Structure of intact, monomeric ZP2 NTR

To investigate the structure of ZP2 NTR in its intact (pre-fertilization) and cleaved (post-fertilization) states, we first expressed the complete NTR of hZP2 (Figure 1B) in mammalian cells and determined its crystal structure at 2.7 Å resolution by experimental phasing (Table S1). This showed that, rather than being arranged like beads on a string, the three Ig-like domains of the NTR generate a V-shaped architecture where ZP2-N1 and ZP2-N3 extend the sheets of the ZP2-N2 β-sandwich (Figure 1C). This is facilitated by the second β sheet of ZP2-N2, whose F and

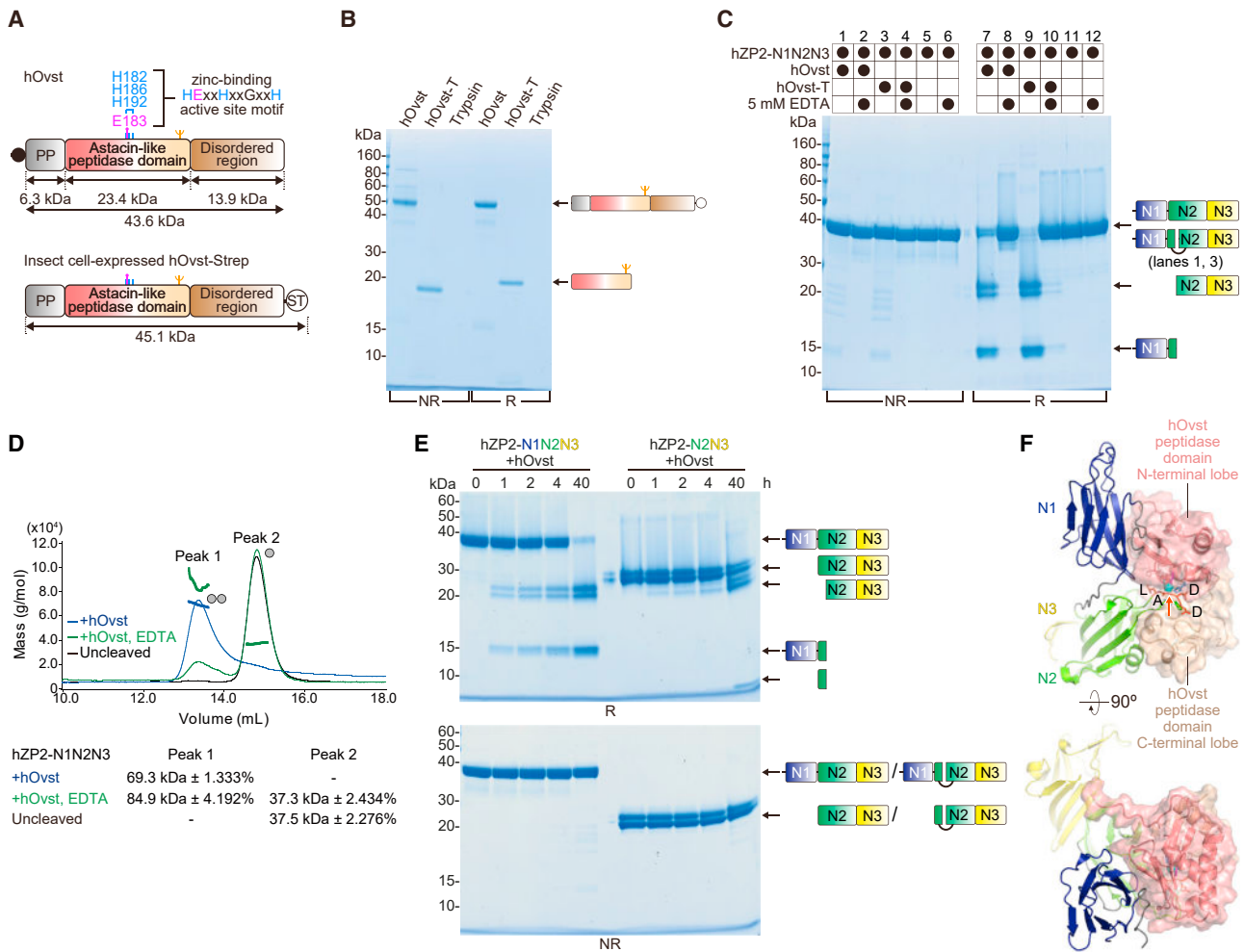
G strands are significantly longer than those of ZP2-N1/N3. Notably, although ZP2-N2 and ZP2-N3 share the same basic fold as ZP2-N1,<sup>23</sup> their cd loops are not engaged in the typical disulfide bond between conserved Cys 2 and 3 (C<sub>2</sub>–C<sub>3</sub>) found in other ZP-N domains,<sup>9,22,23,34</sup> but instead harbor an additional C' β strand (as previously observed in ZP1-N1<sup>18</sup>) and a helical region (Figure S1). Whereas ZP2-N2 and -N3 share an interface of ~770 Å<sup>2</sup>, the interaction between ZP2-N1 and -N2 is less extensive (~475 Å<sup>2</sup>) but also observed in a 3.2 Å resolution structure of hZP2-N1N2 that we subsequently solved by molecular replacement (MR) (Figures 1D and S2; Table S1). In both crystals, however, there is no density for the ZP2-N2 bc loop containing the CG protease cleavage site (171-LA|DD-174), suggesting that this is highly flexible in uncleaved ZP2.

### Cleavage triggers ZP2 NTR oligomerization

To investigate the molecular consequences of the post-fertilization processing of ZP2, we developed a fully recombinant system for *in vitro* cleavage of hZP2 NTR with hOvst (Figure 2). This demonstrated that insect cell-expressed hOvst (Figures 2A and 2B) directly cleaves hZP2 in a metal-dependent way (Figure 2C). Moreover, comparison of uncleaved and hOvst-cleaved hZP2 by size exclusion chromatography-multi-angle light scattering (SEC-MALS) revealed that processing of hZP2 induces its homodimerization (Figure 2D). Although we could not determine a structure of cleaved hZP2, we succeeded in obtaining structural information by taking advantage of the fact that the *Xenopus* homolog of ZP2 can be processed *in vitro* using collagenase. This enzyme closely mimics the action of the yet-to-be-identified anuran CG protease by specifically cleaving the bc loop of xZP2-N2 and triggering a block to sperm binding.<sup>4,35</sup> Based on this knowledge, we treated preparative amounts of recombinant xZP2-N2N3 with collagenase and discovered that cleavage induces almost complete tetramerization of the protein (Figures 3A–3C). The same phenomenon was observed in parallel small-scale experiments performed using a crude exudate from activated *Xenopus* eggs, which, however, also yielded a roughly equal amount of dimers (Figures 3D–3F). Crystallography and cryo-EM (Tables S1 and S2) reveal that cleavage of xZP2-N2N3, which is highly similar to the same region of hZP2 (average root-mean-square deviation [RMSD] 1.9 Å over 161 C $\alpha$ ), allows it to homodimerize through the reorientation and mutual interaction of the severed bc loops of two molecules (Figure 3G). The loops also insert conserved hydrophobic side chains into pockets of the opposite subunit (Figure 3H) and pack on top of paired βG strands (Figure 3I), generating an interface of ~1,600 Å<sup>2</sup> that—combined with a second type of interface mediated by intermolecular pairing between ZP2-N2 C' and E' strands (~460 Å<sup>2</sup>; Figure 3J)—results in the formation of a highly stable dimer of homodimers (Figures 3K, 3L, S3 and S4).

### The ZP-N1 domain regulates the cleavage and dimerization of mammalian ZP2

Comparison of hZP2 *in vitro* cleavage experiments carried out using the complete hZP2 NTR or a fragment consisting of the ZP2-N2 and ZP2-N3 domains shows that the catalytic efficiency of hOvst is enhanced by the presence of the ZP2-N1 domain (Figure 2E), suggesting that the latter facilitates



**Figure 2. Ovastacin cleavage of human ZP2 NTR triggers its homodimerization and is facilitated by the ZP2-N1 domain**

(A) Schematic diagram of hOvst. Black circle, signal peptide; PP, propeptide; ST, Strep-tag II.

(B) Trypsinization of purified recombinant hOvst. hOvst-T, trypsinized hOvst. NR, non-reducing conditions; R, reducing conditions.

(C) Cleavage of hZP2-N1N2N3 with hOvst or hOvst-T. Note how EDTA inhibits the Zn<sup>2+</sup>-dependent proteolytic activity of hOvst.

(D) SEC-MALS profile of uncleaved and hOvst-cleaved hZP2-N1N2N3.

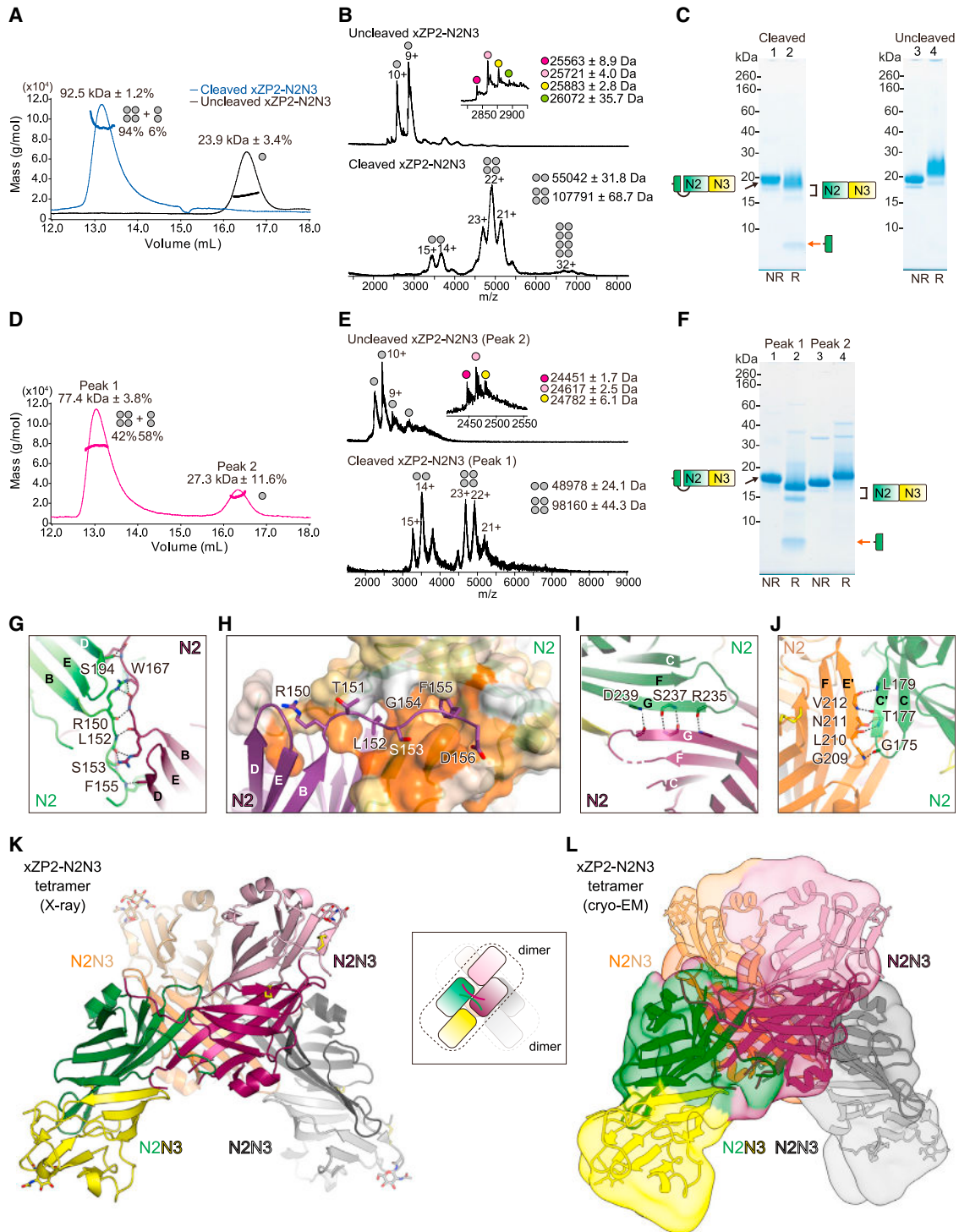
(E) Time course of hZP2-N1N2N3 or hZP2-N2N3 cleavage with hOvst.

(F) Top-ranked AlphaFold-Multimer model of the hZP2 NTR/hOvst complex. Orange sticks, hZP2-N2 residues flanking the ovastacin cleavage site (arrow). Marine blue/magenta sticks, hOvst HExxHxxGxxH motif residues; cyan sphere, computationally grafted Zn<sup>2+</sup>. Low-confidence hOvst C-terminal residues were omitted.

protease/substrate interaction. Accordingly, prediction of the structure of the hZP2 NTR/hOvst complex with AlphaFold-Multimer<sup>36</sup> consistently results in models where ZP2-N1 binds to the N-terminal lobe of hOvst and is displaced from ZP2-N2, which interacts with the C-terminal lobe of the protease and inserts the bc loop LADD sequence into the enzyme's active site (Figure 2F).

Notably, the ZP2-N2 βG strand that mediates the intermolecular pairing of cleaved xZP2 (Figure 3I) pairs intramolecularly with ZP2-N1 βA in the context of uncleaved hZP2 (Figure 1D). This raises the question of whether, in addition to its suggested role as a sperm receptor, the ZP2-N1 domain influences the post-fertilization cleavage of mammalian ZP2 in ways other than facilitating the action of Ovst. To address this point, we generated mice expressing ZP2 lacking ZP2-N1 instead of the wild-type (WT) protein (Figures S5A–S5C). These animals

are severely subfertile *in vivo* (Figure 4A) and sterile *in vitro* (Figure 4B). Unexpectedly, sperm still attach to the ZPs of animals homozygous for this ΔN1 mutation (Figure 4C); however, very few sperm manage to penetrate the mutant ZP (Figures 4D–4G), which—despite detectable lack of mZP2 cleavage by mOvst (Figures S5D–S5H)—is as resistant to collagenase digestion (used as an indicator for hardening) as that of fertilized WT eggs (Figure 4H). In agreement with the latter observation, crystallography of a mouse NTR ΔN1 construct (mZP2-N2N3) revealed that—at concentrations comparable with those in the ZP—this protein is able to establish a homodimeric interface similar to that of cleaved xZP2-N2N3, even though its N2 bc loop is intact and can thus only make limited interactions compared with its cleaved *Xenopus* counterpart (Figures 4I, 4J, and S5D–S5H; Table S1).



**Figure 3. Biochemical and structural analysis of *Xenopus* ZP2 NTR in its cleaved, homotetrameric state**

(A) SEC-MALS profile of uncleaved and collagenase-cleaved xZP2-N2N3. Gray dots indicate the oligomerization state of the protein. Protein tetramer and dimer peaks overlap, due to their very similar radii of gyration (deduced from the structure in [K] and [L]).

(B) Native mass spectrometry (MS) of the samples from (A).

(C) SDS-PAGE analysis of (A) samples. Orange arrow in this panel and (F), N-terminal fragment of cleaved xZP2 that separates from the rest of the protein upon reduction of the C<sub>1</sub>139–C<sub>4</sub>244 disulfide.

(D) SEC-MALS profile of *Xenopus* egg exudate-treated xZP2-N2N3.

(legend continued on next page)

Taken together, these results indicate that, unlike previously thought, mZP2-N1 is not required for sperm/ZP attachment; however, this domain is crucial for fertility by physically counteracting the association of uncleaved ZP2 molecules before fertilization, as well as facilitating their cleavage by Ovst after fertilization.

### Three-dimensional architecture of a vertebrate egg coat filament

Since ZP2 is a multidomain protein whose NTR and ZP module halves are involved in the regulation of gamete interaction and polymerization, respectively, the biological significance of our ZP2 NTR-related findings is intimately connected with the essential role of ZP2 as a building block of the egg coat.<sup>8,37</sup> This is because, depending on whether it involves the NTRs of ZP2 subunits that belong to the same or separate filaments, the oligomerization of cleaved ZP2 would differently affect the architecture of the ZP and, thus, its interaction with sperm.

As mentioned above, cryo-EM recently unveiled the homopolymeric architecture of urinary ZP module protein UMOD.<sup>32</sup> However, comparable studies of native heteromeric ZP filaments have been hindered by their paucity, covalent cross-linking, and extreme biochemical heterogeneity; moreover, there is currently no system to reconstitute these polymers *in vitro* from recombinant subunits. To overcome this impasse, we exploited the fact that relatively homogeneous native egg coat filament fragments can be obtained upon specific digestion of medaka fVE, which consists of ZP1 and ZP3 subunits (fZP1/3; Figure 5A, left and middle panels), with high and low choriolytic enzymes (HCE/LCE). The co-operative action of these proteases allows the mature fish embryo to hatch from the VE. This is achieved by first swelling the egg coat through digestion of the NTRs of fZP1 and fZP3, which, after fertilization, become cross-linked by  $\epsilon$ -( $\gamma$ -glutamyl)lysine bonds that harden the VE, and then dissolving the egg coat via site-specific fragmentation of its filaments.<sup>32,38</sup> Using the resulting material, we crystallized a ~74 kDa heteromeric complex consisting of an intact ZP module from fZP3 and single copies of the ZP-N and ZP-C domains of fZP1, whose IDL is site-specifically cleaved by LCE (Figure 5A, right panel). Although the ~4 Å resolution diffraction of the crystals hampered experimental phasing, we could solve the structure by MR using a partial AlphaFold<sup>39</sup> model of the fZP1/fZP3 complex, generated using a procedure developed to investigate UMOD polymerization *in silico*<sup>40</sup> (crystal form I, Figures S6A and S6B; Table S3). The resulting refined model was then used for MR phasing of data collected from a different crystal form of the same material, which diffracted highly anisotropically beyond 2.7 Å resolution (crystal form II, Figures S6D–S6H; Table S3).

The maps provide detailed information on the polymeric structure of egg coat proteins by revealing a common rod-shaped as-

sembly where fZP3 adopts a fully extended conformation, with its IDL embracing the ZP-C and ZP-N domains of preceding and subsequent fZP1 subunits, respectively (Figures 5B–5D). However, whereas crystal form I contains dimers of fZP1/fZP3 complexes held together by weak head-to-head contacts (Figures S6B and S6C), crystal form II consists of pseudo-infinite filaments where complexes are engaged in extensive head-to-tail interfaces (Figures 5C and S6H). Crucially, the latter position the hatching enzyme-cleaved termini of fZP1-N and fZP1-C moieties at a distance that is compatible with the covalent connectivity of intact fZP1 subunits (Figure 5C, red arrows). This strongly suggests that the crystal form II structure—which will be henceforth discussed—is biologically relevant. Its packing indicates a filament architecture where a protofilament composed of full-length fZP1 subunits and a protofilament of fZP3 subunits wrap around each other into a left-handed double helix (Figure 5D).

### Comparison with the assembly and overall structure of homopolymeric UMOD

Although the conformation of the ZP module of the precursor forms of ZP1/2/4 is unknown, such information is available for ZP3, whose precursors can be secreted as either obligatory or facultative homodimers.<sup>9,32,43</sup> Comparison of the structure of the homodimeric precursor of chicken ZP3 (cZP3) with that of polymeric fZP3 indicates that, although its precursor is very different from that of UMOD,<sup>9,34</sup> ZP3 also undergoes a major conformational rearrangement during polymerization. This is because, following dissociation of the EHP (which acted as  $\beta$  strand G of the precursor's ZP-C domain<sup>9</sup>), the ZP-N/ZP-C linker of both proteins becomes stretched for interaction with other subunits (Figure 5E). As a result, although the heteromeric egg coat polymer has significantly different helical parameters than homopolymeric UMOD<sup>32</sup> and lacks its characteristic zig-zag pattern (Figures 5F and 5G), the two types of filaments share the same overall architecture. Accordingly, the trefoil domain of fZP1 (Figure 5H)—which belongs to a distinct clade for which there was no structural information<sup>42</sup> and corresponds to ZP2 NTR by immediately preceding the ZP-N domain—protrudes from the core of the filament like previously observed for the epidermal growth factor (EGF) domain IV of UMOD<sup>32</sup> (Figure 5I).

### Protein-protein interactions in the heteropolymeric egg coat filament

Consistent with the suggested mechanism of UMOD assembly,<sup>32</sup> the egg coat heterofilament has a polarity dictated by the interaction of the ZP-N domain of an incoming subunit (n) with the ZP-C domain of the second to last one (n – 2) component of a growing polymer. Combined with the two-protofilament

(E) Native MS analysis of (D) material shows that peak 2 corresponds to uncleaved monomeric xZP2-N2N3, whereas peak 1 consists of dimeric and tetrameric forms of the cleaved protein.

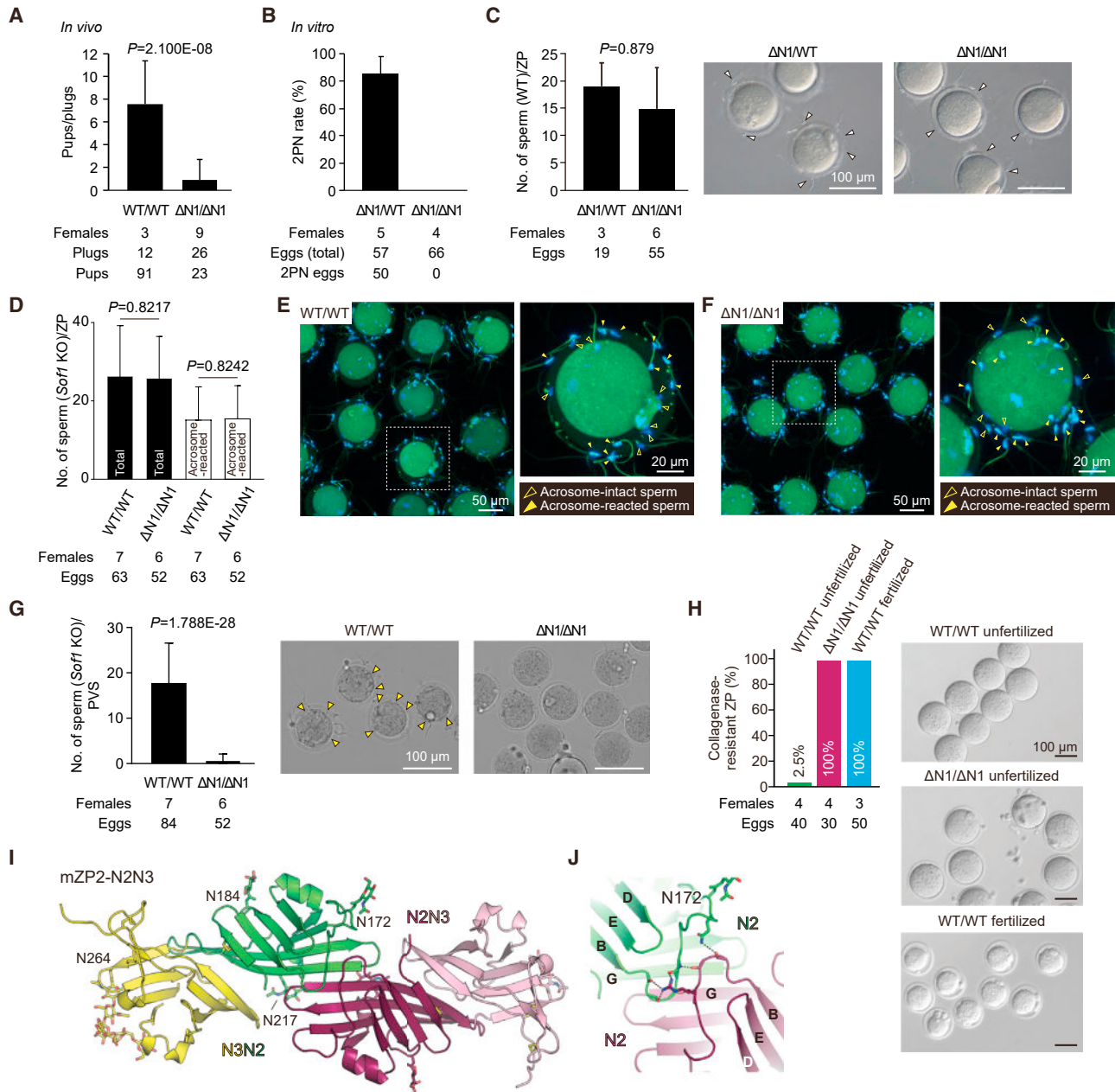
(F) SDS-PAGE analysis shows that (D) peaks 2 and 1 correspond to cleaved and uncleaved xZP2-N2N3, respectively.

(G–J) Details of ZP2 oligomeric interactions mediated by cleaved loops (G and H), G  $\beta$  strands (I), and C', E'  $\beta$  strands (J). In (H), the cleaved xZP2-N2 domain on the right is shown in surface representation and colored by hydrophobicity from white (low) to dark orange (high).

(K) Crystal structure of cleaved xZP2-N2N3. Right inset, xZP2-N2N3 tetramer scheme.

(L) Cryo-EM map of the xZP2-N2N3 tetramer.

See also Figures S2–S4.



**Figure 4. Deletion of ZP2-N1 does not affect sperm binding but leads to a semi-hardened ZP and subfertility in mice**

(A) Number of pups over plugs.

(B) *In vitro* fertilization percentage.

(C) Sperm-binding assay. White arrowheads indicate ZP-bound sperm.

(D–F) *Sof1* knockout (KO) sperm bind equally well to the ZP of WT and ZP2 ΔN1 homozygous eggs. (D) Number of *Sof1* KO sperm bound to WT or ΔN1/ΔN1 eggs.

*Sof1* KO sperm show comparable binding to the ZP of both kinds of oocytes, and their acrosome status does not influence this activity.

(E and F) Fluorescence images of *Sof1* KO sperm bound to WT or ΔN1/ΔN1 eggs. Sperm acrosome status was monitored by pre-staining with anti-IZUMO1

antibody. Open and close yellow arrowheads indicate acrosome-intact and -reacted sperm, respectively.

(G) ZP penetration assay. PVS, perivitelline space. Yellow arrowheads, sperm in the PVS.

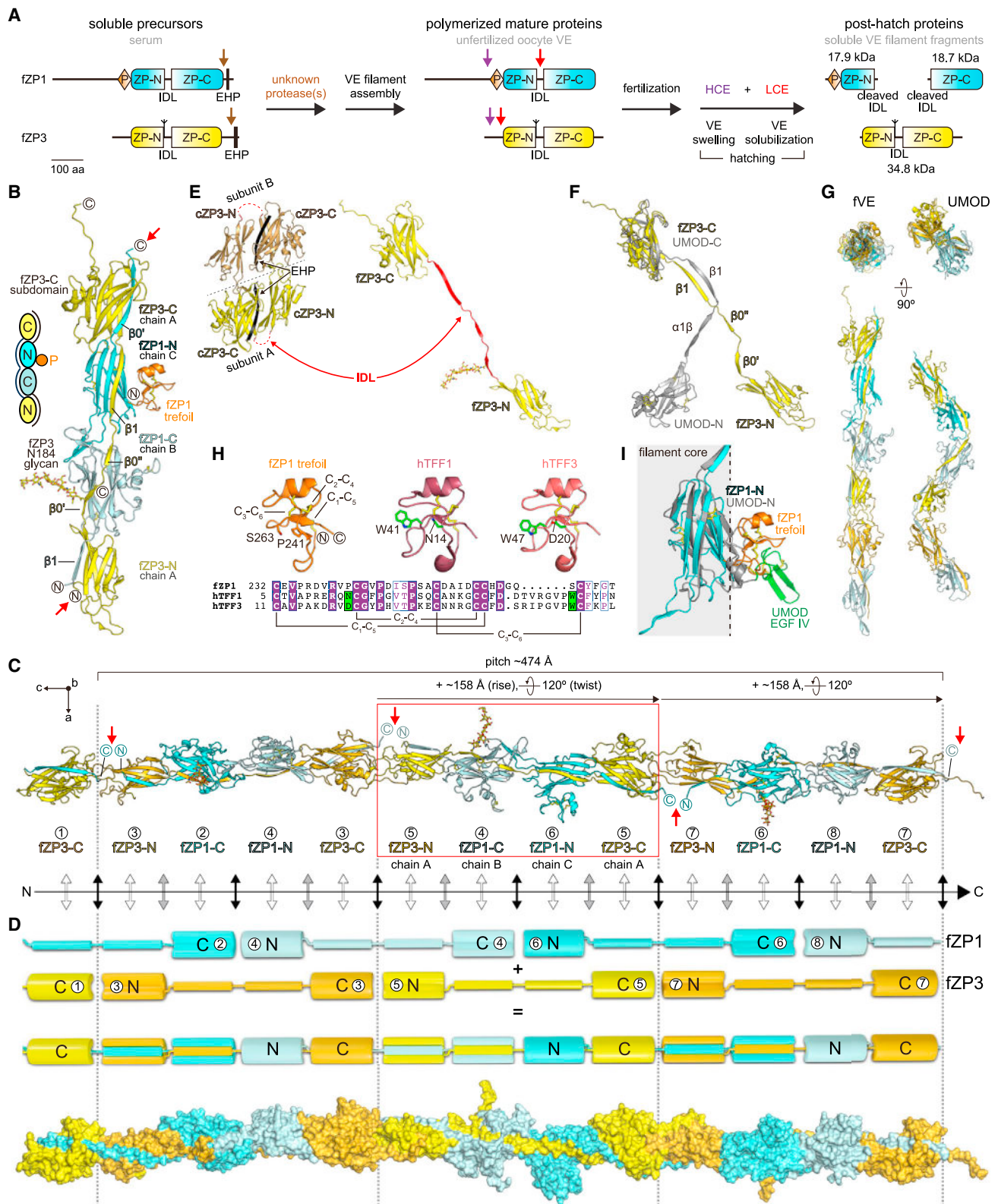
(H) Collagenase resistance of WT or ΔN1/ΔN1 eggs.

(I) The crystal structure of mZP2-N2N3 shows a homodimer.

(J) Detail of the interaction between uncleaved mZP2-N2 bc loops.

See also [Figures S2](#) and [S5](#).





(legend on next page)

architecture (Figure 5D), this gives rise to a linear alternation of homomeric (ZP1/ZP1 or ZP3/ZP3) ZP-C/ZP-N interfaces and heteromeric (ZP3/ZP1 or ZP1/ZP3) ZP-N/ZP-C interfaces (Figure 6A). These contacts, together with those between the IDL of one subunit and ZP-C and ZP-N domains from the other (Figure 6B), constitute a highly complex set of interactions that contribute to filament formation.

The homomeric interfaces share a set of six highly conserved residues (Figures 6C–6E), which are also present in UMOD despite a much more kinked ZP-C/ZP-N orientation (Figure 5G) stabilized by additional contacts (ZP-C ef loop/ZP-N cd loop and Y407) not found in the straight egg coat filament. Notably, the set includes the signature FXF motif of the ZP-C domain, as well as the invariant Arg of the ZP-N fg loop and its counterpart acidic residues on ZP-C. The interactions mediated by these elements are not only essential for UMOD polymerization,<sup>32</sup> but also required for the homodimerization and secretion of the cZP3 precursor.<sup>9</sup> This suggests how, following EHP release and the stochastic dissociation of one of their two ZP-C/ZP-N interfaces, both subunits of a ZP3 homodimer may be incorporated head-to-tail into a growing ZP3 protofilament.

Outside the six-residue interface core, fZP1 amino acids that are largely conserved in ZP1/2/4 homologs form a complex network of hydrogen bonds, salt bridges, and hydrophobic interactions. These include loops a'b, cc', and c''d of ZP-C, loop bc of ZP-N, and an extensive cross-hairpin interaction involving two pairs of  $\beta$  strands belonging to the ef and fg loops of ZP-C and ZP-N, respectively (Figures 6D and 6F). These contacts make the fZP1-C/fZP1-N interface structurally distinct and significantly larger than fZP3-C/fZP3-N (880 vs. 627 Å<sup>2</sup>). Together with differences in the latter at the level of the core (where the invariant Asp is supplied by ZP3-specific 3<sub>10</sub> helix c''d), the ef/fg loop interaction (where a much shorter ZP-N ef loop can only contribute a minimal  $\beta$  strand), and the C-terminal tail of fZP3-C (where a

stretch of 11 residues, oriented by ZP3 ZP-C subdomain, interacts with the subsequent subunit within the ZP3 protofilament) (Figures 6C, 6G, and 6H), these features explain the homomeric nature of ZP-C/ZP-N interactions.

The heteromeric ZP-N/ZP-C interfaces are smaller than homomeric ones (fZP1-N/fZP3-C: 494 Å; fZP3-N/fZP1-C: 571 Å) and less polar, as reflected by a distinct common core of five conserved hydrophobic residues (Figures 6I–6K). These include the signature Tyr within the ZP-N domain E'-F-G extension<sup>44</sup> (fZP1 Y339, fZP3 Y149), hydrogen-bonded to the main chain of ZP3 V245 or the side chain of invariant ZP1 D451, and a second near-invariant Tyr at its C terminus (fZP1 Y371, fZP3 Y179), involved in highly conserved intermolecular hydrogen/cation- $\pi$  and stacking interactions with ZP3 H244, L207, or fZP1 R449, L398 (Figures 6L and 6M). Mutation of the signature Tyr in human inner ear protein  $\alpha$ -tectorin disrupts the tectorial membrane and causes non-syndromic deafness<sup>45</sup>; moreover, both Tyr are implicated in UMOD homopolymerization<sup>32</sup> and replaced by non-hydrophobic amino acids in endoglin, a non-polymeric ZP module-containing protein.<sup>44</sup> Overall, the heteromeric interfaces appear to be less discriminating than the homomeric ones; however, some interactions outside their conserved core clearly favor the observed subunit combinations. These include contacts made by near-invariant fZP3 R181 (substituted by A373 in fZP1), as well as hydrophobic interactions between fZP3 M142 (E332 in fZP1) and the loop stapled by the conserved C<sub>a</sub>-C<sub>b</sub> disulfide of fZP1-C (Figures 6L and 6M).

Elements outside ZP-N and ZP-C also contribute to the specificity of subunit interaction. Rationalizing previous biochemical studies of the bovine (b) ZP3/ZP4 complex,<sup>46</sup> the near-invariant WVPF motif at the center of the fZP3 IDL engages in yet another cation- $\pi$  interaction with fZP1-C (fZP3 W194/fZP1 R528, corresponding to bZP3 W156/bZP4 K434) and hydrophobic interactions between its Phe (fZP3 F197, bZP3 F159) and a large patch

### Figure 5. Structure of a native fish VE fragment crystallized as a filament

(A) Domain structure of medaka VE major subunits fZP1 and fZP3 in their soluble precursor (left), polymerized (center), and hatching enzyme-digested (right) forms. Vertical arrows indicate the specific recognition sites of the proteases regulating VE subunit polymerization and hatching. Note that, because fZP1 and fZP3 are secreted into the serum by the liver, their precursors lack a C-terminal transmembrane domain.

(B) Overall structure of the fZP1/fZP3 complex (crystal form II). Red arrows at the top and bottom, fZP1 IDL residues D387 and S388 (P1 and P1' recognition sites of medaka's LCE hatching enzyme, respectively).

(C) ac plane view of crystal form II. fZP3-N and fZP3-C of complexes contained in adjacent asymmetric units interact head-to-tail, bringing the C- and N-terminal ends of cleaved fZP1 IDLs close to one another (red arrows). This regenerates pseudo-infinite helical egg coat filaments, only one of which is shown here, that run parallel along the c axis. Asymmetric unit boundaries are indicated by vertical dashed lines; the second of the three complexes that make up a complete helical turn, according to the parameters reported on top, is enclosed by a red rectangle and depicted as in (B). Protein subunits, colored in alternating yellow and gold (fZP3) or cyan and pale cyan (fZP1), are numbered according to their order in the filament. At the bottom, a black horizontal arrow indicates the filament polarity, with vertical black, white, and gray double arrows highlighting the alternation of homomeric ZP-C/ZP-N interfaces, heteromeric IDL/ZP(-N/C) interfaces, and heteromeric ZP-N/ZP-C interfaces, respectively.

(D) Schematic and surface representations of the egg coat filament shown in (C). The top scheme shows each ZP module as a solid object, with thick cylinders representing ZP-N and ZP-C domains and thin cylinders depicting the two halves of the IDL; cylinder lengths approximately match the extent of the corresponding 3D structures in (C), and fZP1 P was omitted for clarity. To highlight how the IDLs of each subunit wrap around the ZP-C and ZP-N domains of the other, the fZP1 linker is shown as intact in the solid scheme, and its cleaved moieties have been reconnected *in silico* in the surface model.

(E) Structural comparison of the homodimeric precursor of cZP3 (PDB: 3NK4)<sup>9</sup> and polymeric fZP3. The ZP-N/ZP-C linker (red) is largely disordered before polymerization, which requires cleavage and dissociation of the C-terminal EHP (black).<sup>12,41</sup>

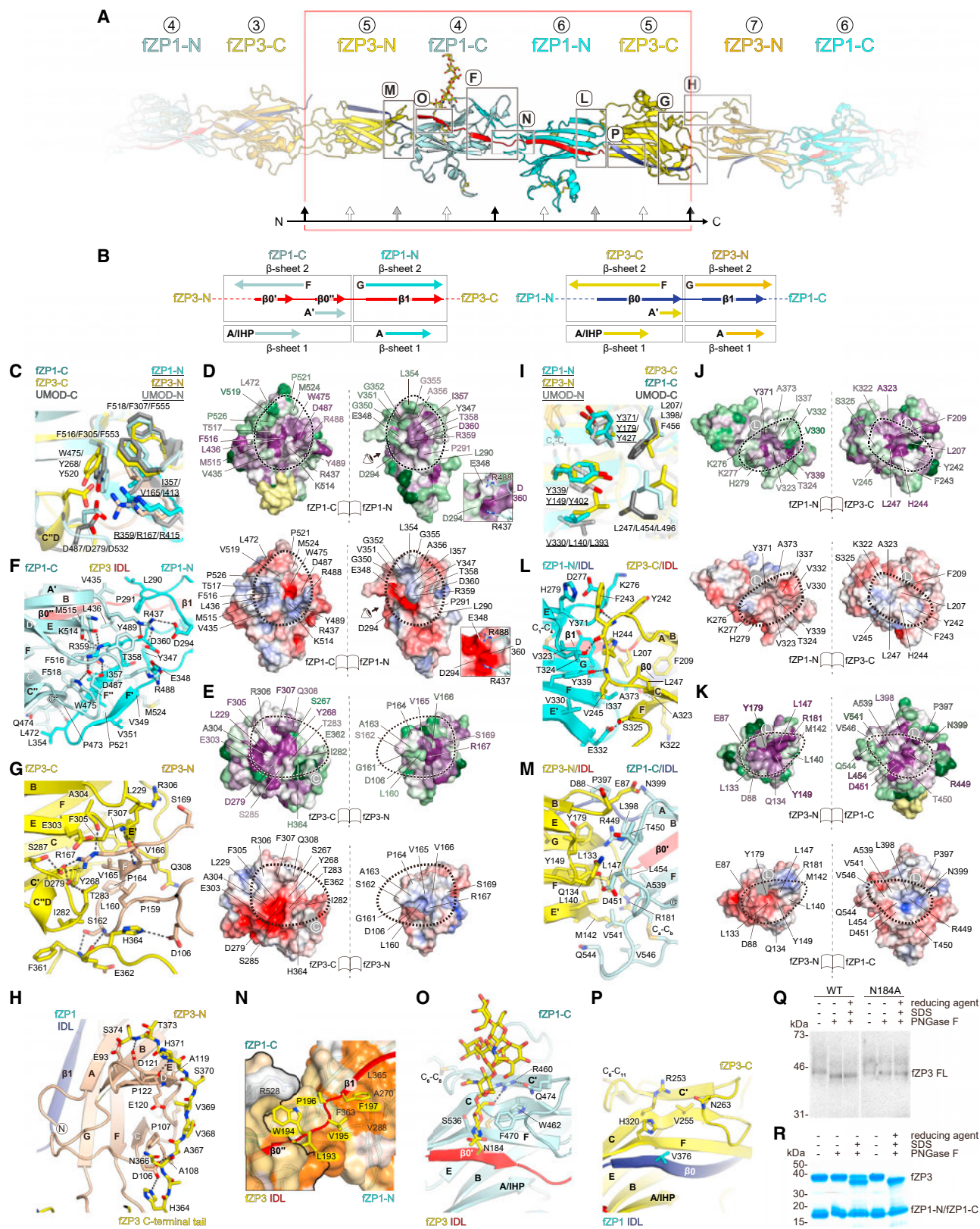
(F) Superposition of the fZP3 subunit from the heterocomplex and a subunit of the UMOD homopolymer (PDB: 6TQK).<sup>32</sup>

(G) Comparison of the fVE and UMOD filament cores in section (top) and side (bottom) views. fVE and corresponding UMOD subunits are depicted using the same color scheme as (C); non-ZP module domains are omitted.

(H) fZP1 trefoil domain structure and its comparison with hTFF1 (PDB: 6V1D) and hTFF3 (PDB: 6V1C)<sup>42</sup> (RMSD over 36 C $\alpha$  1.05 and 1.06 Å, respectively). While the fZP1 trefoil shares the same disulfide pattern as hTFF1 and hTFF3, it lacks both of their conserved sugar-binding residues (green).

(I) The NTRs of polymeric fZP1 and UMOD protrude similarly from the respective filaments.

See also Figure S6.



(legend on next page)

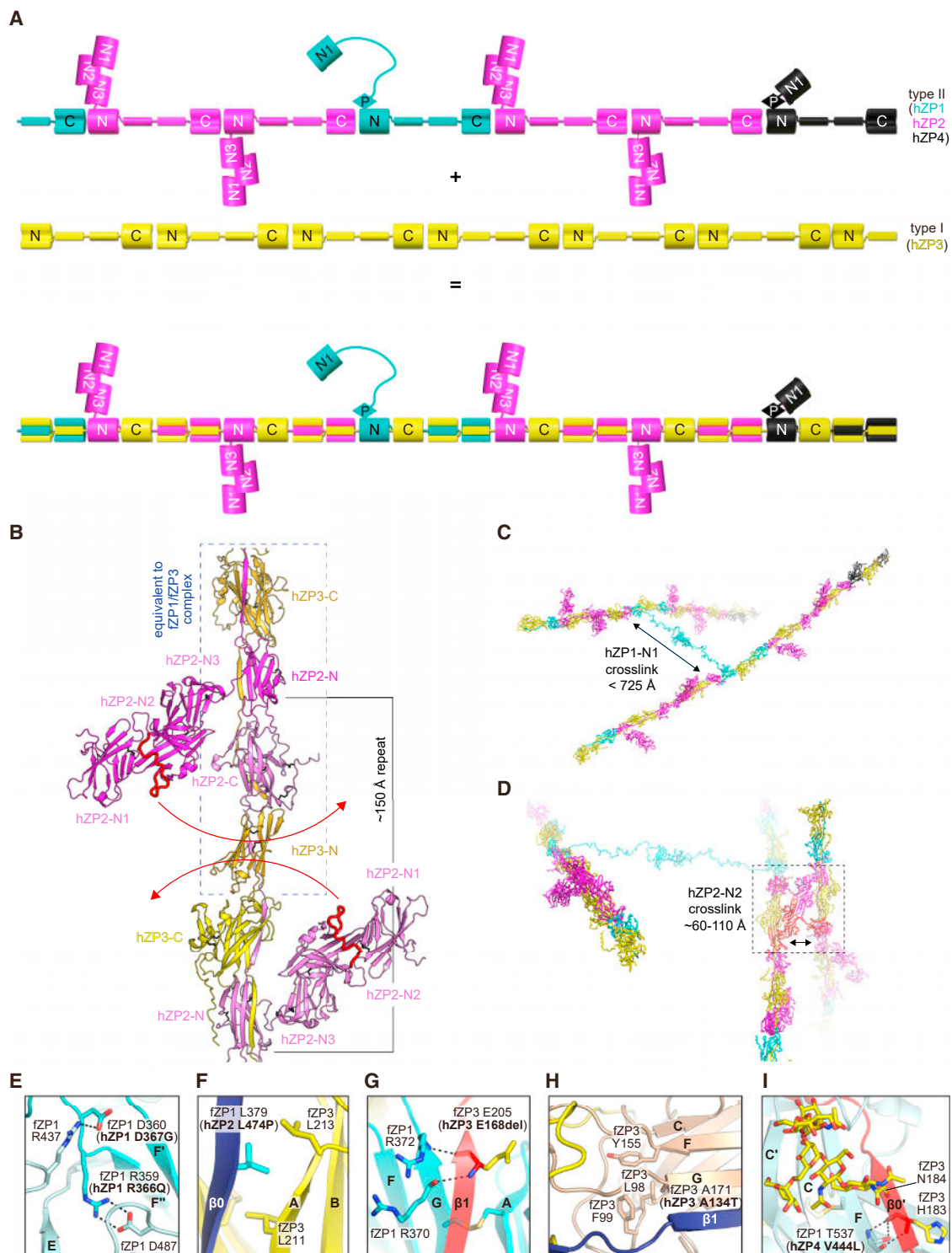
of fZP1-N residues highly conserved in ZP1/4 homologs, including bZP4 (Figure 6N). Interestingly, in animal classes other than fish (where sperm penetrates the egg coat through the micropyle and VE proteins have only a structural role), the residue that immediately precedes the WVPF motif (L193 in fZP3; Figure 6N) is almost always a Thr. O-glycosylation at this position was implicated in gamete recognition in bird and mammals<sup>9,47</sup> and the Thr is closely preceded by the only highly conserved N-glycosylation site of ZP3 (N184 in fZP3), important for sperm binding in bovine.<sup>46</sup> The map shows clear density for a corresponding N-glycan that contributes to the ZP3/ZP1 interface by packing against the  $\beta$ C- $\beta$ C' region of fZP1 ZP-C, including the invariant Trp (W462) that immediately follows conserved Cys 5 of ZP1/2/4 (Figures 6O and S6D). Substitution of the Trp with Val in ZP3 homologs and additional differences, such as fZP1 S536 being replaced by bulky H320 in fZP3, indicate that a comparable interaction between the ZP3 N-glycan and another ZP3 subunit would be less favored (Figure 6P). As in the case of the corresponding carbohydrate of cZP3,<sup>9</sup> the N184 glycan can be enzymatically cleaved from the fZP3 precursor and is not required for its secretion (Figure 6Q); on the contrary—consistent with its interaction with fZP1—it is protected from digestion in the context of the native fZP1/fZP3 complex (Figure 6R). These observations suggest that the conserved N-glycan of ZP3 contributes to egg coat filament architecture. Moreover, the packing of crystal form II (Figure S6H) hints at the possibility that, in the case of the compact VE of fish, the same glycan also facilitates the close juxtaposition of filaments observed by EM.<sup>32</sup>

### AlphaFold-based modeling of the hZP

Although AlphaFold models of monomeric proteins can generally facilitate crystallographic structure determination by MR,<sup>48</sup> this approach did not succeed in the case of the fZP1/fZP3 complex due to the major conformational changes involved in ZP subunit polymerization (Figures 5 and 6). However, the fact that we could use a multi-chain prediction to successfully phase the challenging fZP1/fZP3 X-ray diffraction data indicates that the multi-mer version of the neural network could accurately predict the structure of the corresponding ZP module complex. Based on this consideration, we extended our experimental studies by generating a comprehensive prediction matrix for possible interactions between hZP subunits (Figure S7). This computational analysis was first carried out for three-component systems consisting of one ZP module and two ZP-C and ZP-N moieties (based on the fZP1/fZP3 complex), and subsequently extended to systems encompassing four complete ZP modules. Because AlphaFold has no explicit knowledge of the concept of filament, top-scoring models from the latter set of predictions consisted of either circular or linear, but artificially stretched, arrangements of subunits. Nonetheless, the network's predictions consistently replicated the basic features of the experimentally observed filament architecture whenever copies of the hZP3 ZP module were combined with either partial or full ZP modules from hZP1, hZP2, or hZP4, or vice versa. By integrating this information with the crystallographic knowledge on fZP1/fZP3 (Figures 5 and 6), the structural data on ZP2 NTR (Figures 1, 3, and 4), and previous investigations of the ZP filament cross-link,<sup>18</sup> we were able to assemble molecular models of hZP filaments in their pre- and

### Figure 6. Analysis of protein-protein interactions regulating the assembly of the fZP1/fZP3 heteropolymer

- (A) Enlarged view of the central section of the egg coat filament shown in Figure 5C. Different contact areas between ZP subunits are enclosed by boxes, whose labels refer to the following panels.
- (B) Scheme of the  $\beta$  strand interactions between the IDL of one subunit and the ZP-C and ZP-N domains of the other. Note that  $\beta$  strand A of the ZP-C domain is also referred to as internal hydrophobic patch (IHP).<sup>9,10,12</sup>
- (C) Five highly conserved residues mediate homomeric ZP-C/ZP-N domain interactions in both the fZP1 and fZP3 protofilaments, as well as the UMOD polymer. ZP-N residues are underlined.
- (D) Surface representations of the fZP1-C/fZP1-N homomeric interface colored by evolutionary conservation of amino acid positions in ZP1 and the related ZP modules of ZP2 and ZP4 (using a spectrum ranging from deep magenta [highest conservation] to light green [lowest conservation], with yellow indicating residues with a low-confidence conservation level; top), or by electrostatic potential (using a spectrum ranging from red [-2.5 kT/e] to blue [+2.5 kT/e] through white [0 kT/e]; bottom). Panels are shown in an open book representation, with black dashed ovals indicating the moieties of the interface. Inset, close-up view of the fZP1-N1 surface from the direction of the eye symbol, highlighting charge-related interactions between fZP1-N D294, E348, D360, and fZP1-C R437, R488.
- (E) ZP3 homolog residue conservation and electrostatic potential surface of the homomeric fZP3-C/fZP3-N interface. Gray area, cross-section of the fZP3 C-terminal tail ("C").
- (F and G) fZP1-C/fZP1-N (F) and fZP3-C/fZP3-N (G) homomeric interface contacts. In these and subsequent panels, black dashed lines indicate hydrogen bonds and salt bridges.
- (H) The fZP3 C-terminal tail interacts with the ZP-N domain of the following subunit within the fZP3 protofilament. Notably, the same C-terminal residues are completely disordered in crystal form I, consistent with the idea that its packing is non-biologically relevant.
- (I) Heteromeric ZP-N/ZP-C interactions are mediated by five highly conserved residues sharing similar relative positions in the fZP1-N/fZP3-C, fZP3-N/fZP1-C, and UMOD-N/UMOD-C interfaces.
- (J and K) Amino acid conservation and electrostatic potential surface of the heteromeric fZP1-N/fZP3-C (J) and fZP3-N/fZP1-C (K) interfaces. IDL cross-sections are in gray and indicated by "L."
- (L and M) fZP1-N/fZP3-C (L) and fZP3-N/fZP1-C (M) heteromeric interface contacts.
- (N) Detail of part of the fZP3 IDL wrapping around the fZP1-C/fZP1-N homomeric interface. fZP1 subunit surfaces are colored by hydrophobicity from white (low) to dark orange (high), with a thick black outline marking the fZP1-C/fZP1-N boundary.
- (O) The N-glycan attached to the invariant Asn residue of fZP3 IDL's  $\beta$ 0' interacts with R460, W462, and F470 of the ZP-C domain of the preceding fZP1 subunit.
- (P) View of the fZP1  $\beta$ 0'/fZP3-C interface, inverse of the one shown in (O).
- (Q) Immunoblot analysis of secreted recombinant wild-type fZP3 and N-glycosylation site mutant N184A, treated with PNGase F under native or fully denaturing conditions.
- (R) PNGase F treatment of the fZP1/fZP3 complex under native, partially denaturing, or fully denaturing conditions.
- See also Figure S6.



**Figure 7. Modeling of human ZP filaments, before and after post-fertilization cleavage of ZP2, and infertility-associated mutations**

(A) Scheme of the hZP filament and its type I and II protofilament moieties. NTRs and ZP modules have arbitrary relative orientations.

(B) Part of a hZP filament 3D model encompassing two complete hZP3 and hZP2 subunits (gold and light magenta, respectively), as well as the NTR+ZP-N of another ZP2 subunit (magenta) and the ZP-C of another ZP3 subunit (yellow) plus part of the respective IDLs. The red arrows show that, although ZP2 NTR is predicted to adopt variable orientations relative to the core of the filament (Figure S7D), it cannot form an intramolecular cross-link with an adjacent copy of ZP2 upon post-fertilization cleavage of its ZP-N2 bc loop (red) by ovastacin.

(legend continued on next page)

post-fertilization state (Figures 7A–7D). As discussed in the following sections, this complemented our empirical findings to yield major insights into egg coat architecture, suggest how ZP2 cleavage blocks polyspermy at the molecular level, and rationalize ZP gene mutations linked to human infertility.

## DISCUSSION

### A conserved framework for egg coat assembly

Our findings complete the structural knowledge of ZP protein domains and, considering the phylogenetic relationship and architectural similarity between ZP4 and ZP1 and between ZPAX/ZPY and ZP2,<sup>26,49</sup> provide comprehensive structural information on the vertebrate egg coat.

Underlining the conservation of the ZP module as the building block of egg coats from mollusk to human,<sup>7</sup> the 158 Å helical rise of the fVE filament (Figures 5C and 5D) is remarkably consistent with the structural repeat of ~150 Å suggested by early EM studies of the mZP.<sup>8,17,50</sup> However, due to the distinctive architecture of the ZP module polymer, this conserved repeat does not correspond to a subunit heterodimer (as previously suggested for mZP2 and mZP3), but rather to a complete ZP module from one protofilament embracing two half-ZP modules from the other (Figures 5B–5D). A key feature of the repeat is the IDL of the ZP module, whose contact area with the interacting ZP-C/ZP-N unit is more than double the ZP-C/ZP-N interface itself. Because of this, the IDL acts as a molecular tape that strongly stabilizes the interaction between adjacent domains within the opposite protofilament (Figures 5 and 6). Combined with the fact that the ZP module of ZP1 is an integral component of the egg coat filament (as opposed to being peripherally attached, as previously hypothesized for mZP1), this explains how the fish hatching enzyme disrupts VE filaments by targeting an exposed site in the middle of fZP1 IDL<sup>38</sup> (Figures 5A, 5C, and 5D).

In the case of the crystallized fVE complex, the presence of two protofilaments is a direct product of the homomeric ZP-C/ZP-N interfaces that mediate the head-to-tail interaction between fZP3 or fZP1 subunits, respectively (Figures 5 and 6). However, AlphaFold modeling suggests that the same two-protofilament architecture is conserved across vertebrates, regardless of how many different subunits make up the respective VEs/ZPs. This is because, as a result of their ability to interact head-to-tail with each other, ZP1, ZP2, and ZP4 can coassemble into a common protofilament, which pairs with a ZP3 protofilament like the two sides of a zipper (Figures 7A–7D). Such an organization is consistent with the notion that—based on structural differences in their ZP-C domains—egg coat proteins can be divided into type I (ZP3) and type II (ZP1/2/4) subunits (with the latter type also including non-mammalian ZPAX and ZPY components)<sup>9,10,34,51</sup> and the observation that all vertebrate egg coats contain at least one subunit of each type.<sup>1,49</sup> More-

over, it agrees with the relative ~1:4:5 stoichiometry of mZP1–3 subunits.<sup>50</sup> Most importantly, because the protein-protein interactions within one protofilament are cemented by the IDLs of the other (Figures 5C, 5D, 6A, and 6B), the filament architecture straightforwardly explains the results of ZP gene ablation/replace-ment experiments in mouse, rat, rabbit and medaka.<sup>27,37,52–59</sup> For example, a mouse egg coat filament model where a type I mZP3 protofilament pairs with a type II protofilament containing 80% mZP2 and 20% mZP1 rationalizes why the ZP is completely absent and approximately half the normal width in *Zp3*<sup>-/-</sup> and *Zp3*<sup>+/-</sup> mice, respectively<sup>52,53,60</sup>; why the ZP is clearly present in *Zp1*<sup>-/-</sup> animals, despite being abnormally organized due to lack of covalent cross-links<sup>54</sup>; and why the oocytes of *Zp2*<sup>-/-</sup> mice have a transient, very thin ZP made up of mZP3 and mZP1.<sup>37</sup> At the same time, the existence of type I and type II protofilaments is compatible with the observation that, when incubated *in vitro* at relatively high concentrations, individual subunits purified from fish or mouse egg coats can form homooligomers.<sup>61,62</sup> Finally, the apparent interchangeability of type II subunits suggests that their incorporation into growing egg coat filaments is not only stochastic but also reflects their relative localization and expression levels; this explains how restricted areas of the egg coat, such as the germinal disc-facing part of the avian VE,<sup>63</sup> may in some cases become specialized to present a specific population of polymeric subunits implicated in sperm binding.

### Mechanism of the ZP2 block to polyspermy

The structural organization of the egg coat filament implies that, even when two copies of ZP2 immediately follow each other within a type II protofilament, their ZP-N2 domains are too far to interact between themselves (Figure 7B). Considering that ZP filaments are relatively straight and largely parallel to adjacent ones within ZP bundles,<sup>64</sup> this strongly suggests that post-fertilization cleavage triggers interaction between the processed NTRs of ZP2 subunits belonging to different filaments. Based on the relative abundance of ZP2<sup>3</sup> and our structural information (Figures 3K and 3L), this will extensively bridge ZP filaments and bring them within less than 110 Å from each other, approximately 7-fold closer than the maximum distance allowed by their intermolecular ZP1 cross-links<sup>17,18</sup> (Figures 7C and 7D). Consistent with the phenomena associated with hardening<sup>65</sup> and the phenotype of the ZP2 ΔN1/ΔN1 mice (Figure 4), this will in turn rigidify the overall structure of the matrix and prevent polyspermy by hindering sperm penetration. This mechanism provides a molecular basis for the observed coalescence of hZP filaments after fertilization,<sup>64</sup> as well as the polyspermy associated with female infertility mutations in the hOvst gene.<sup>66,67</sup> Moreover, it conceptually resembles how the fVE is hardened by transglutaminase-mediated cross-linking of its filaments (which lack ZP2),<sup>68</sup> and is compatible with the recent suggestion that cross-linking of ZP3 may also contribute to the block

(C) Before fertilization, hZP filaments are loosely cross-linked via covalent homodimerization of hZP1 ZP-N1, connected to the filament core by a long flexible linker. ZP subunits are colored as in (A).

(D) Post-fertilization cleavage triggers the homodimerization of hZP2 NTRs that protrude from different filaments, non-covalently cross-linking the ZP into a tight mesh.

(E–I) Close-up views of fZP1/fZP3 residues corresponding to positions affected by human ZP gene mutations associated with infertility.

See also Figure S7.

to polyspermy in the mouse.<sup>69</sup> On the other hand, in line with the observation that ZP2 cleavage is evolutionary conserved<sup>4,70–72</sup> whereas ZP2-N1 is post-translationally removed from amphibian ZP2 before egg-sperm interaction and fertilization,<sup>4,73</sup> the mechanism does not support the currently accepted model whereby ZP2-N1 contains an essential sperm-binding site, whose destruction or masking upon ZP2 processing abolishes gamete interaction.<sup>27,74</sup> Instead, our results suggest that mammalian ZP2-N1 plays important roles in both preventing the premature association of ZP2 and facilitating its post-fertilization cleavage. Consistent with the former function is the observation that, in the case of *Xenopus* ZP2 which lacks ZP2-N1, oligomerization appears to be strictly dependent on cleavage of the ZP2-N2 bc loop. Most importantly, our findings reformulate the supramolecular model of fertilization<sup>24,72</sup> by suggesting that the key role of ZP2 processing—which in mammals can be crucially controlled by ZP2-N1 itself (Figures 2 and 4)—is to alter the ZP architecture so that it physically prevents the penetration of sperm. Notably, as discussed below, this mechanism is compatible with data suggesting that, depending on the species, one or more binding sites other than ZP2-N1 may significantly contribute to the initial attachment of sperm to the egg coat. At the same time, again consistent with the evolutionary conservation of ZP2 cleavage, a physical barrier mechanism does not intrinsically depend on the nature of such interaction(s), or how gamete recognition is abolished following fusion. The basis of the ZP2 block to polyspermy is thus reminiscent of the extracellular matrix changes involved in the formation of the fertilization envelope of invertebrates<sup>75</sup> or the proposed cell wall block to polyspermy of plants.<sup>76</sup> More generally, the ZP2 cleavage-dependent remodeling of the ZP also exemplifies how the mosaic architecture of ZP module-containing proteins allows them to orchestrate a complex supramolecular rearrangement in response to a defined molecular signal.

### Implications for egg coat-sperm interaction

The phenotype of the ZP2  $\Delta$ N1 homozygous mice, which differ from previously reported animals lacking mZP2 residues C51–V149<sup>27</sup> by carrying a ZP2-N1 deletion that does not modify the first  $\beta$  strand of ZP2-N2 (Figures S1B and S5A–S5D), shows that mZP2-N1 is not essential for sperm binding to the mZP (Figures 4C–4F). However, our data do not exclude that ZP2-N1 contributes to gamete interaction in mammals, as indicated by *in vitro* binding experiments.<sup>27</sup> At the same time, our findings imply that egg coat regions other than ZP2-N1 play important (but, when present within a single species, possibly redundant) roles in mediating the initial, species-restricted attachment of sperm to the VE/ZP. Along these lines, vertebrate gamete recognition has long been suggested to also involve carbohydrates attached to different regions of egg coat subunits. These include a putative O-glycosylation site close to the N terminus of mature *Xenopus* ZP2<sup>4</sup>; glycosylation sites conserved within the IDL of avian and mammalian ZP3, such as the aforementioned cZP3 T168 and bZP3 N146 (L193 and N184 in fZP3; Figures 6N and 6O),<sup>9,46,47,77</sup> or located near the C terminus of mZP3<sup>78</sup>; and N-glycosylation sites of porcine ZP3 ZP-C and ZP4 ZP-N.<sup>79</sup> Additionally, the polypeptide moiety of the ZP-N1 domain of bovine ZP4, as well as a ZP module fragment encompassing the protein's IDL, were recently suggested to be involved in sperm binding.<sup>20</sup> Whereas the exact

role and relative importance of these elements remains unclear in the absence of a bona fide counterpart on sperm, the structure of the egg coat filament allows to put a large part of this knowledge into context, informing future studies of sperm-ZP interaction.

### Significance for human infertility

Because of the dramatic conformational changes that the ZP module undergoes during polymerization (Figure 5E), the previously available structures of isolated ZP subunits or domains thereof<sup>9,22,34</sup> could give only limited information about their mature filamentous form. The availability of detailed structural knowledge on the egg coat filament significantly improves our ability to interpret the growing number of human ZP gene mutations linked with female infertility.<sup>80</sup>

For example, hZP1 R366Q, associated with an oocyte maturation defect that leads to primary infertility due to atresia or empty follicles,<sup>81,82</sup> affects the key Arg residue engaged in the invariant salt bridge of homotypic interfaces (fZP1-C D487/fZP1-N R359; Figures 6D, 6F, and 7E). This suggests that the mutation impairs the assembly of hZP type II protofilaments, by disrupting the interaction between hZP1 R366 and the invariant ZP-C aspartic acid of hZP1 (D483), hZP2 (D571), or hZP4 (D393). Notably, mutation of the near-invariant Asp residue that immediately follows R366 has also been reported to cause infertility, when in compound heterozygosity with a frameshift mutation that produces a secretion-impaired truncated hZP1 variant.<sup>83</sup> The structure of the egg coat filament rationalizes the lack of a ZP associated with the hZP1 D367G mutation by showing that the corresponding fZP1-N residue (D360) makes an intersubunit hydrogen bond with another highly conserved residue, fZP1-C R437 (hZP1 R433) (Figures 6D, 6F, and 7E). This does not support a previous suggestion that the mutation hinders D367 from interacting intramolecularly with non-conserved residues R299, H306, and S368.<sup>83</sup>

On the other hand, the suggestion that hZP2 mutation L474P causes ZP abnormality and infertility not only by reducing protein secretion but also by affecting hZP2/hZP3 interaction<sup>84</sup> is supported by the insertion of the corresponding fZP1 IDL residue (L379) between two conserved Leu residues of ZP3 ZP-C (L211 and L213 in fZP3; Figure 7F). Similar considerations apply to hZP3 E168del, an IDL residue deletion found in patients with empty follicle syndrome<sup>85</sup> (Figure 7G); hZP3 A134T, a dominant negative heterozygous mutation that affects a ZP-N hydrophobic core residue packing against the hZP1/2/4 IDL (Figure 7H) and is associated with oocyte degeneration caused by lack of a ZP<sup>86</sup>; and hZP4 V444L, a mutation that is associated with thin and irregular ZPs<sup>87</sup> and affects a ZP-C residue (corresponding to fZP1 T537) that interfaces with the ZP3 IDL amino acid preceding the conserved N-glycosylation site (Figure 7I).

### Limitations of the study

Although all available evidence suggests that the NTRs and ZP modules of ZP subunits are structurally and functionally independent,<sup>11,17,18,27</sup> we cannot exclude that, within the context of the full-length polymerized proteins, there may be additional interactions that contribute to the architecture of the egg coat at different stages of development. For example, the VE filament fragment that we crystallized was obtained by digesting native material with hatching enzymes and thus lacked N-terminal protein regions

whose cross-linking by transglutaminase hardens the fish egg coat after fertilization.<sup>38,68</sup> Therefore, although the packing of the reconstituted filaments in crystal form II may provide some hints in this regard (Figure S6H), our study cannot address the overall organization of full-length filaments within the hardened fVE. A second question that remains open is whether, following fertilization, the ZP2-N1 domain is randomly oriented, contributes additional contacts to the cleaved ZP2-N2 homodimer, or interacts with other filament regions.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - DNA constructs
  - Protein expression and purification
  - *In vitro* cleavage of hZP2 with recombinant hOvst
  - *In vitro* cleavage of xZP2 with collagenase
  - *In vitro* cleavage of xZP2 with *Xenopus* oocyte exudate
  - Protein crystallization and derivatization
  - X-ray data collection and processing
  - X-ray structure determination of hZP2-N1N2N3
  - X-ray structure determination of hZP2-N1N2
  - X-ray structure determination of cleaved xZP2-N2N3
  - X-ray structure determination of mZP2-N2N3
  - X-ray structure determination of the fZP1/fZP3 complex (crystal form I)
  - X-ray structure determination of the fZP1/fZP3 complex (crystal form II) and analysis of its packing
  - Cryo-EM data collection and structure determination of cleaved xZP2-N2N3
  - Sequence-structure analysis
  - Size exclusion chromatography-multi angle light scattering (SEC-MALS)
  - Native mass spectrometry (MS)
  - Mouse experiments
  - Design and generation of ZP2 ΔN1 mutant mice
  - Mouse fertility test
  - *In vitro* fertilization
  - Sperm-ZP binding assays
  - Experiments with *Sof1* knockout mouse sperm
  - ZP hardening assay
  - Protein deglycosylation
  - Immunoblot analysis
  - Calculation of the relative stoichiometry of mouse ZP subunits
  - Modeling of the human ZP2 NTR/ovastacin complex
  - Modeling of human ZP filaments
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2024.02.013>.

## ACKNOWLEDGMENTS

This work was supported by the Knut and Alice Wallenberg Foundation (grant 2018.0042 to L.J.), the Swedish Research Council (grants 2016-03999, 2020-04936 to L.J.; 2019-01961 to M.L.), the Center for Innovative Medicine (grant 2-537/2014 to L.J.), the German Research Foundation (grant 401763100 to D.F.), NIH (grant R01 GM125638 to A.E.C.), and JSPS (grant 20K06774 to S.Y.; 19H05750, 21H05033 to M.I.). We thank the staff of the European Synchrotron Radiation Facility (ESRF), Diamond Light Source (DLS), Helmholtz-Zentrum Berlin (HZB)/BESSY II and the Cryo-EM Swedish National Facility (funded by the Knut and Alice Wallenberg, Family Erling Persson, and Kempe Foundations, SciLifeLab, Stockholm University) for help with X-ray and cryo-EM data collection; Kathryn Tunyasuvunakool (DeepMind) for initial AlphaFold models of monomeric fVE subunits; Gérard Bricogne and Clemens Vonrhein (Global Phasing Ltd.) for help with STARANISO-derived diffraction data deposition; Paul Wassarman (Icahn School of Medicine at Mount Sinai) for the 2C7 hybridoma line and discussion; and Franco Cotelli (University of Milano), Tsukasa Matsuda (Fukushima University), and Gertie van Pouderoeyen for comments.

## AUTHOR CONTRIBUTIONS

Protein expression and purification: S.N., D.F., E.D., L.H., and B.A.; X-ray crystallography: S.N., B.W., D.F., E.D., S.Z.-C., M.B., A.S., D.d.S., and L.J.; cryo-EM: B.W.; AlphaFold predictions: S.N., B.W., and L.J.; native MS: M.L.; *Xenopus* egg exudate collection: R.E.B., K.M.K., and A.E.C.; mouse experiments: C.E., Y.L., M.K., and M.I.; fVE complex preparation: S.Y. The project was conceived and directed by L.J., who wrote the manuscript together with S.N. and A.S., with contributions from the other authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 29, 2023

Revised: November 7, 2023

Accepted: February 9, 2024

Published: March 14, 2024

## REFERENCES

1. Wong, J.L., and Wessel, G.M. (2006). Defending the zygote: search for the ancestral animal block to polyspermy. *Curr. Top. Dev. Biol.* 72, 1–151.
2. Evans, J.P. (2020). Preventing polyspermy in mammalian eggs—Contributions of the membrane block and other mechanisms. *Mol. Reprod. Dev.* 87, 341–349.
3. Bleil, J.D., Beall, C.F., and Wassarman, P.M. (1981). Mammalian sperm-egg interaction: Fertilization of mouse eggs triggers modification of the major zona pellucida glycoprotein, ZP2. *Dev. Biol.* 86, 189–197.
4. Tian, J., Gong, H., and Lennarz, W.J. (1999). *Xenopus laevis* sperm receptor gp69/64 glycoprotein is a homolog of the mammalian sperm receptor ZP2. *Proc. Natl. Acad. Sci. USA* 96, 829–834.
5. 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.
6. Galindo, B.E., Moy, G.W., Swanson, W.J., and Vacquier, V.D. (2002). Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene* 288, 111–117.



7. Jovine, L., Darie, C.C., Litscher, E.S., and Wassarman, P.M. (2005). Zona pellucida domain proteins. *Annu. Rev. Biochem.* **74**, 83–114.
8. Litscher, E.S., and Wassarman, P.M. (2020). Zona Pellucida Proteins, Filbrils, and Matrix. *Annu. Rev. Biochem.* **89**, 695–715.
9. Han, L., Monné, M., Okumura, H., Schwend, T., Cherry, A.L., Flot, D., Matsuda, T., and Jovine, L. (2010). Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**, 404–415.
10. Bokhove, M., and Jovine, L. (2018). Structure of Zona Pellucida Module Proteins. *Curr. Top. Dev. Biol.* **130**, 413–442.
11. Jovine, L., Qi, H., Williams, Z., Litscher, E., and Wassarman, P.M. (2002). The ZP domain is a conserved module for polymerization of extracellular proteins. *Nat. Cell Biol.* **4**, 457–461.
12. Jovine, L., Qi, H., Williams, Z., Litscher, E.S., and Wassarman, P.M. (2004). A duplicated motif controls assembly of zona pellucida domain proteins. *Proc. Natl. Acad. Sci. USA* **101**, 5922–5927.
13. Schaeffer, C., Santambrogio, S., Perucca, S., Casari, G., and Rampoldi, L. (2009). Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly. *Mol. Biol. Cell* **20**, 589–599.
14. Litscher, E.S., Qi, H., and Wassarman, P.M. (1999). Mouse Zona Pellucida Glycoproteins mZP2 and mZP3 Undergo Carboxy-Terminal Proteolytic Processing in Growing Oocytes. *Biochemistry* **38**, 12280–12287.
15. Brunati, M., Perucca, S., Han, L., Cattaneo, A., Consolato, F., Andolfo, A., Schaeffer, C., Olinger, E., Peng, J., Santambrogio, S., et al. (2015). The serine protease hepsin mediates urinary secretion and polymerisation of zona Pellucida domain protein uromodulin. *eLife* **4**, e08887.
16. Bleil, J.D., and Wassarman, P.M. (1980). Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. *Cell* **20**, 873–882.
17. Greve, J.M., and Wassarman, P.M. (1985). Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *J. Mol. Biol.* **187**, 253–264.
18. Nishimura, K., Dioguardi, E., Nishio, S., Villa, A., Han, L., Matsuda, T., and Jovine, L. (2019). Molecular basis of egg coat cross-linking sheds light on ZP1-associated female infertility. *Nat. Commun.* **10**, 3086.
19. Lefièvre, L., Conner, S.J., Salpekar, A., Olufowobi, O., Ashton, P., Pavlovic, B., Lenton, W., Afnan, M., Brewis, I.A., Monk, M., et al. (2004). Four zona pellucida glycoproteins are expressed in the human. *Hum. Reprod.* **19**, 1580–1586.
20. Diliumulati, K., Orita, M., Yonahara, Y., Imai, F.L., and Yonezawa, N. (2022). Identification of Sperm-Binding Sites in the N-Terminal Domain of Bovine Egg Coat Glycoprotein ZP4. *Int. J. Mol. Sci.* **23**, 762.
21. Callebaut, I., Mornon, J.P., and Monget, P. (2007). Isolated ZP-N domains constitute the N-terminal extensions of Zona Pellucida proteins. *Bioinformatics* **23**, 1871–1874.
22. Monné, M., Han, L., Schwend, T., Burendahl, S., and Jovine, L. (2008). Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456**, 653–657.
23. Raj, I., Sadat Al Hosseini, H., Dioguardi, E., Nishimura, K., Han, L., Villa, A., de Sanctis, D., and Jovine, L. (2017). Structural Basis of Egg Coat-Sperm Recognition at Fertilization. *Cell* **169**, 1315–1326.e17.
24. Rankin, T.L., Coleman, J.S., Epifano, O., Hoodbhoy, T., Turner, S.G., Castle, P.E., Lee, E., Gore-Langton, R., and Dean, J. (2003). Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Dev. Cell* **5**, 33–43.
25. Lindsay, L.L., and Hedrick, J.L. (2004). Proteolysis of *Xenopus laevis* egg envelope ZPA triggers envelope hardening. *Biochem. Biophys. Res. Commun.* **324**, 648–654.
26. Goudet, G., Mugnier, S., Callebaut, I., and Monget, P. (2008). Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biol. Reprod.* **78**, 796–806.
27. Avella, M.A., Baibakov, B., and 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.
28. Tokuhiro, K., and Dean, J. (2018). Glycan-Independent Gamete Recognition Triggers Egg Zinc Sparks and ZP2 Cleavage to Prevent Polyspermy. *Dev. Cell* **46**, 627–640.e5.
29. Schroeder, A.C., Schultz, R.M., Kopf, G.S., Taylor, F.R., Becker, R.B., and Eppig, J.J. (1990). Fetuin inhibits zona pellucida hardening and conversion of ZP2 to ZP2f during spontaneous mouse oocyte maturation in vitro in the absence of serum. *Biol. Reprod.* **43**, 891–897.
30. Wassarman, P.M., Florman, H.M., and Greve, J.M. (1985). Receptor-Mediated Sperm–Egg Interactions in Mammals. In *Biology of Fertilization*, C.B. Metz and A. Monroy, eds. (Academic Press), pp. 341–360.
31. Wassarman, P.M. (1988). Zona pellucida glycoproteins. *Annu. Rev. Biochem.* **57**, 415–442.
32. Stsiapanava, A., Xu, C., Brunati, M., Zamora-Caballero, S., Schaeffer, C., Bokhove, M., Han, L., Hebert, H., Carroni, M., Yasumasu, S., et al. (2020). Cryo-EM structure of native human uromodulin, a zona pellucida module polymer. *EMBO J.* **39**, e106807.
33. Stanisich, J.J., Zyla, D.S., Afanasyev, P., Xu, J., Kipp, A., Olinger, E., Devuyt, O., Pilhofer, M., Boehringer, D., and Glockshuber, R. (2020). The cryo-EM structure of the human uromodulin filament core reveals a unique assembly mechanism. *eLife* **9**, e60265.
34. Bokhove, M., Nishimura, K., Brunati, M., Han, L., de Sanctis, D., Rampoldi, L., and Jovine, L. (2016). A structured interdomain linker directs self-polymerization of human uromodulin. *Proc. Natl. Acad. Sci. USA* **113**, 1552–1557.
35. Tian, J., Gong, H., Thomsen, G.H., and Lennarz, W.J. (1997). Gamete interactions in *Xenopus laevis*: identification of sperm binding glycoproteins in the egg vitelline envelope. *J. Cell Biol.* **136**, 1099–1108.
36. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv.
37. Rankin, T.L., O'Brien, M., Lee, E., Wigglesworth, K., Eppig, J., and Dean, J. (2001). Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. *Development* **128**, 1119–1126.
38. Yasumasu, S., Kawaguchi, M., Ouchi, S., Sano, K., Murata, K., Sugiyama, H., Akema, T., and Iuchi, I. (2010). Mechanism of egg envelope digestion by hatching enzymes, HCE and LCE in medaka, *Oryzias latipes*. *J. Biochem.* **148**, 439–448.
39. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
40. Jovine, L. (2021). Using machine learning to study protein-protein interactions: From the uromodulin polymer to egg zona pellucida filaments. *Mol. Reprod. Dev.* **88**, 686–693.
41. Sugiyama, H., Murata, K., Iuchi, I., Nomura, K., and Yamagami, K. (1999). Formation of mature egg envelope subunit proteins from their precursors (choriogenins) in the fish, *Oryzias latipes*: loss of partial C-terminal sequences of the choriogenins. *J. Biochem.* **125**, 469–475.
42. Järnvå, M.A., Lingford, J.P., John, A., Soler, N.M., Scott, N.E., and Goddard-Borger, E.D. (2020). Trefoil factors share a lectin activity that defines their role in mucus. *Nat. Commun.* **11**, 2265.
43. Zhao, M., Boja, E.S., Hoodbhoy, T., Nawrocki, J., Kaufman, J.B., Kresge, N., Ghirlando, R., Shiloach, J., Pannell, L., Levine, R.L., et al. (2004). Mass spectrometry analysis of recombinant human ZP3 expressed in glycosylation-deficient CHO cells. *Biochemistry* **43**, 12090–12104.
44. Saito, T., Bokhove, M., Croci, R., Zamora-Caballero, S., Han, L., Letarte, M., de Sanctis, D., and Jovine, L. (2017). Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Rep.* **19**, 1917–1928.

45. Legan, P.K., Lukashkina, V.A., Goodyear, R.J., Lukashkin, A.N., Verhoeven, K., Van Camp, G., Russell, I.J., and Richardson, G.P. (2005). A deafness mutation isolates a second role for the tectorial membrane in hearing. *Nat. Neurosci.* **8**, 1035–1042.
46. Suzuki, K., Tatebe, N., Kojima, S., Hamano, A., Orita, M., and Yonezawa, N. (2015). The Hinge Region of Bovine Zona Pellucida Glycoprotein ZP3 Is Involved in the Formation of the Sperm-Binding Active ZP3/ZP4 Complex. *Biomolecules* **5**, 3339–3353.
47. Chalabi, S., Panico, M., Sutton-Smith, M., Haslam, S.M., Patankar, M.S., Lattanzio, F.A., Morris, H.R., Clark, G.F., and Dell, A. (2006). Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. *Biochemistry* **45**, 637–647.
48. Terwilliger, T.C., Afonine, P.V., Liebschner, D., Croll, T.I., McCoy, A.J., Oeffner, R.D., Williams, C.J., Poon, B.K., Richardson, J.S., Read, R.J., et al. (2023). Accelerating crystal structure determination with iterative *AlphaFold* prediction. *Acta Crystallogr. D Struct. Biol.* **79**, 234–244.
49. Feng, J.M., Tian, H.F., Hu, Q.M., Meng, Y., and Xiao, H.B. (2018). Evolution and multiple origins of zona pellucida genes in vertebrates. *Biol. Open* **7**, bio036137.
50. Wassarman, P.M., and Mortillo, S. (1991). Structure of the mouse egg extracellular coat, the zona pellucida. *Int. Rev. Cytol.* **130**, 85–110.
51. Boja, E.S., Hoodbhoy, T., Fales, H.M., and Dean, J. (2003). Structural characterization of native mouse zona pellucida proteins using mass spectrometry. *J. Biol. Chem.* **278**, 34189–34202.
52. Liu, C., Litscher, E.S., Mortillo, S., Sakai, Y., Kinloch, R.A., Stewart, C.L., and Wassarman, P.M. (1996). Targeted disruption of the *mZP3* gene results in production of eggs lacking a zona pellucida and infertility in female mice. *Proc. Natl. Acad. Sci. USA* **93**, 5431–5436.
53. Rankin, T., Familiari, M., Lee, E., Ginsberg, A., Dwyer, N., Blanchette-Mackie, J., Drago, J., Westphal, H., and Dean, J. (1996). Mice homozygous for an insertional mutation in the *Zp3* gene lack a zona pellucida and are infertile. *Development* **122**, 2903–2910.
54. Rankin, T., Talbot, P., Lee, E., and Dean, J. (1999). Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. *Development* **126**, 3847–3855.
55. Zeng, M.H., Wang, Y., Huang, H.L., Quan, R.P., Yang, J.T., Guo, D., Sun, Y., Lv, C., Li, T.Y., Wang, L., et al. (2021). *Zp4* is completely dispensable for fertility in female rats. *Biol. Reprod.* **104**, 1282–1291.
56. Lamas-Toranzo, I., Fonseca Balvís, N., Querejeta-Fernández, A., Izquierdo-Rico, M.J., González-Brusi, L., Lorenzo, P.L., García-Rebollar, P., Avilés, M., and Bermejo-Álvarez, P. (2019). ZP4 confers structural properties to the zona pellucida essential for embryo development. *eLife* **8**, e48904.
57. Murata, K., and Kinoshita, M. (2022). Targeted deletion of liver-expressed Choriogenin L results in the production of soft eggs and infertility in medaka, *Oryzias latipes*. *Zool. Lett.* **8**, 1.
58. Birk, D.S., Onose, S., Kinoshita, M., and Murata, K. (2022). Medaka, *Oryzias latipes*, egg envelopes are created by ovarian-expressed ZP proteins and liver-expressed choriogenins. *Zool. Lett.* **8**, 11.
59. Yokokawa, R., Watanabe, K., Kanda, S., Nishino, Y., Yasumasu, S., and Sano, K. (2023). Egg envelope formation of medaka *Oryzias latipes* requires ZP proteins originating from both the liver and ovary. *J. Biol. Chem.* **299**, 104600.
60. Wassarman, P.M., Qi, H., and Litscher, E.S. (1997). Mutant female mice carrying a single *mZP3* allele produce eggs with a thin zona pellucida, but reproduce normally. *Proc. Biol. Sci.* **264**, 323–328.
61. Darie, C.C., Janssen, W.G., Litscher, E.S., and Wassarman, P.M. (2008). Purified trout egg vitelline envelope proteins VE $\beta$  and VE $\gamma$  polymerize into homomeric fibrils from dimers *in vitro*. *Biochim. Biophys. Acta* **1784**, 385–392.
62. Litscher, E.S., Janssen, W.G., Darie, C.C., and Wassarman, P.M. (2008). Purified mouse egg zona pellucida glycoproteins polymerize into homomeric fibrils under non-denaturing conditions. *J. Cell. Physiol.* **214**, 153–157.
63. Nishio, S., Kohno, Y., Iwata, Y., Arai, M., Okumura, H., Oshima, K., Nandano, D., and Matsuda, T. (2014). Glycosylated chicken ZP2 accumulates in the egg coat of immature oocytes and remains localized to the germinal disc region of mature eggs. *Biol. Reprod.* **91**, 107.
64. Familiari, G., Nottola, S.A., Macchiarelli, G., Micara, G., Aragona, C., and Motta, P.M. (1992). Human zona pellucida during *in vitro* fertilization: an ultrastructural study using saponin, ruthenium red, and osmium-thiocarbonylhydrazide. *Mol. Reprod. Dev.* **32**, 51–61.
65. Fahrenkamp, E., Algarrá, B., and Jovine, L. (2020). Mammalian egg coat modifications and the block to polyspermy. *Mol. Reprod. Dev.* **87**, 326–340.
66. Maddirevula, S., Coskun, S., Al-Qahtani, M., Aboyousee, O., Alhassan, S., Aldeery, M., and Alkuraya, F.S. (2022). *ASTL* is mutated in female infertility. *Hum. Genet.* **141**, 49–54.
67. Wu, H., Che, J., Zheng, W., Cheng, D., Gong, F., Lu, G., Lin, G., and Dai, C. (2023). Novel biallelic *ASTL* variants are associated with polyspermy and female infertility: A successful live birth following ICSI treatment. *Gene* **887**, 147745.
68. Sugiyama, H., and Iuchi, I. (2000). Molecular structure and hardening of egg envelope in fish. *Recent Res. Devel. Comp. Biochem. Physiol.* **1**, 139–161.
69. Cui, Z., Lu, Y., Miao, Y., Dai, X., Zhang, Y., and Xiong, B. (2022). Transglutaminase 2 crosslinks zona pellucida glycoprotein 3 to prevent polyspermy. *Cell Death Differ.* **29**, 1466–1473.
70. Hasegawa, A., Koyama, K., Okazaki, Y., Sugimoto, M., and Isojima, S. (1994). Amino acid sequence of a porcine zona pellucida glycoprotein ZP4 determined by peptide mapping and cDNA cloning. *J. Reprod. Fertil.* **100**, 245–255.
71. Iwamoto, K., Ikeda, K., Yonezawa, N., Noguchi, S., Kudo, K., Hamano, S., Kuwayama, M., and Nakano, M. (1999). Disulfide formation in bovine zona pellucida glycoproteins during fertilization: evidence for the involvement of cystine cross-linkages in hardening of the zona pellucida. *J. Reprod. Fertil.* **117**, 395–402.
72. Gahlay, G., Gauthier, L., Baibakov, B., Epifano, O., and Dean, J. (2010). Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. *Science* **329**, 216–219.
73. Scarpecci, S.L., Sanchez, M.L., and Cabada, M.O. (2008). Cellular origin of the *Bufo arenarum* sperm receptor gp75, a ZP2 family member: its proteolysis after fertilization. *Biol. Cell* **100**, 219–230.
74. Baibakov, B., Boggs, N.A., Yauger, B., Baibakov, G., and Dean, J. (2012). Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *J. Cell Biol.* **197**, 897–905.
75. Wessel, G.M., and Wong, J.L. (2009). Cell surface changes in the egg at fertilization. *Mol. Reprod. Dev.* **76**, 942–953.
76. Tekleyohans, D.G., Mao, Y., Kägi, C., Stierhof, Y.D., and Groß-Hardt, R. (2017). Polyspermy barriers: a plant perspective. *Curr. Opin. Plant Biol.* **35**, 131–137.
77. Barisone, G.A., Krapf, D., Correa-Fiz, F., Arranz, S.E., and Cabada, M.O. (2007). Glycoproteins of the vitelline envelope of Amphibian oocyte: biological and molecular characterization of ZPC component (gp41) in *Bufo arenarum*. *Mol. Reprod. Dev.* **74**, 629–640.
78. Chen, J., Litscher, E.S., and Wassarman, P.M. (1998). Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proc. Natl. Acad. Sci. USA* **95**, 6193–6197.
79. Yonezawa, N. (2014). Posttranslational modifications of zona pellucida proteins. *Adv. Exp. Med. Biol.* **759**, 111–140.
80. Litscher, E.S., and Wassarman, P.M. (2020). Zona pellucida genes and proteins and human fertility. *Trends Dev. Biol.* **13**, 21–33.
81. Loeuillet, C., Dhellemmes, M., Cazin, C., Kherraf, Z.E., Fourati Ben Moustapha, S., Zouari, R., Thierry-Mieg, N., Arnoult, C., and Ray, P.F. (2022). A

- recurrent *ZP1* variant is responsible for oocyte maturation defect with degenerated oocytes in infertile females. *Clin. Genet.* *102*, 22–29.
82. Pujalte, M., Camo, M., Celton, N., Attencourt, C., Lefranc, E., Jedraszak, G., and Scheffler, F. (2023). A *ZP1* gene mutation in a patient with empty follicle syndrome: A case report and literature review. *Eur. J. Obstet. Gynecol. Reprod. Biol.* *280*, 193–197.
  83. Chu, K., He, Y., Wang, L., Ji, Y., Hao, M., Pang, W., Liu, Y., Sun, N., Yang, F., and Li, W. (2020). Novel *ZP1* pathogenic variants identified in an infertile patient and a successful live birth following ICSI treatment. *Clin. Genet.* *97*, 787–788.
  84. Hou, M., Zhu, L., Jiang, J., Liu, Z., Li, Z., Jia, W., Hu, J., Zhou, X., Zhang, D., Luo, Y., et al. (2022). Novel Heterozygous Mutations in *ZP2* Cause Abnormal Zona Pellucida and Female Infertility. *Reprod. Sci.* *29*, 3047–3054.
  85. Zhang, Z., Guo, Q., Jia, L., Zhou, C., He, S., Fang, C., Zhang, M., Sun, P., Zeng, Z., Wang, M., et al. (2022). A novel gene mutation in *ZP3* loop region identified in patients with empty follicle syndrome. *Hum. Mutat.* *43*, 180–188.
  86. Chen, T., Bian, Y., Liu, X., Zhao, S., Wu, K., Yan, L., Li, M., Yang, Z., Liu, H., Zhao, H., et al. (2017). A Recurrent Missense Mutation in *ZP3* Causes Empty Follicle Syndrome and Female Infertility. *Am. J. Hum. Genet.* *101*, 459–465.
  87. Wei, X., Li, Y., Liu, Q., Liu, W., Yan, X., Zhu, X., Zhou, D., Tian, Y., Zhang, F., Li, N., et al. (2022). Mutations in *ZP4* are associated with abnormal zona pellucida and female infertility. *J. Clin. Pathol.* *75*, 201–204.
  88. East, I.J., and Dean, J. (1984). Monoclonal antibodies as probes of the distribution of ZP-2, the major sulfated glycoprotein of the murine zona pellucida. *J. Cell Biol.* *98*, 795–800.
  89. Noda, T., Lu, Y., Fujihara, Y., Oura, S., Koyano, T., Kobayashi, S., Matzuz, M.M., and Ikawa, M. (2020). Sperm proteins SOF1, TMEM95, and SPACA6 are required for sperm-oocyte fusion in mice. *Proc. Natl. Acad. Sci. USA* *117*, 11493–11502.
  90. Sasado, T., Tanaka, M., Kobayashi, K., Sato, T., Sakaizumi, M., and Naruse, K. (2010). The National BioResource Project Medaka (NBRP Medaka): an integrated bioresource for biological and biomedical sciences. *Exp. Anim.* *59*, 13–23.
  91. Hekkelman, M.L., de Vries, I., Joosten, R.P., and Perrakis, A. (2023). AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat. Methods* *20*, 205–213.
  92. Arnold, M.J. (2021). *mattarnoldbio/alphapickle: Release v1.4.0.* Zenodo. <https://zenodo.org/records/5708709>.
  93. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmering, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* *11*, 3696–3713.
  94. Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011). Data processing and analysis with the *autoPROC* toolbox. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 293–302.
  95. Vonrhein, C., Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., and Bricogne, G. (2018). Advances in automated data analysis and processing within *autoPROC*, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr. A Found. Adv.* *74*, a360.
  96. Vonrhein, C., Blanc, E., Roversi, P., and Bricogne, G. (2007). Automated structure solution with *autoSHARP*. *Methods Mol. Biol.* *364*, 215–230.
  97. Barson, G., and Griffiths, E. (2016). SeqTools: visual tools for manual analysis of sequence alignments. *BMC Res. Notes* *9*, 39.
  98. Smart, O.S., Womack, T.O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. (2012). Exploiting structure similarity in refinement: automated NCS and target-structure restraints in *BUSTER*. *Acta Crystallogr. D Biol. Crystallogr.* *68*, 368–380.
  99. Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. *Nat. Protoc.* *2*, 2728–2733.
  100. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* *19*, 679–682.
  101. Casañal, A., Lohkamp, B., and Emsley, P. (2020). Current developments in *Coot* for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci.* *29*, 1069–1078.
  102. Naito, Y., Hino, K., Bono, H., and Ui-Tei, K. (2015). CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* *31*, 1120–1123.
  103. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* *14*, 290–296.
  104. Delgado, J., Radusky, L.G., Cianferoni, D., and Serrano, L. (2019). FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* *35*, 4168–4169.
  105. Bohne-Lang, A., and von der Lieth, C.W. (2005). GlyProt: *in silico* glycosylation of proteins. *Nucleic Acids Res.* *33*, W214–W219.
  106. Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., et al. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* *49*, D344–D354.
  107. Croll, T.I. (2018). *ISOLDE*: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D Struct. Biol.* *74*, 519–530.
  108. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* *30*, 772–780.
  109. Necci, M., Piovesan, D., Clementel, D., Dosztányi, Z., and Tosatto, S.C.E. (2021). MobiDB-lite 3.0: fast consensus annotation of intrinsic disorder flavors in proteins. *Bioinformatics* *36*, 5533–5534.
  110. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Protein Sci.* *86*, 2.9.1–2.9.37.
  111. Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 12–21.
  112. Brown, N.P., Leroy, C., and Sander, C. (1998). MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* *14*, 380–381.
  113. Edgar, R.C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* *13*, 6968.
  114. de Beer, T.A.P., Berka, K., Thornton, J.M., and Laskowski, R.A. (2014). PDBsum additions. *Nucleic Acids Res.* *42*, D292–D296.
  115. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. *J. Appl. Crystallogr.* *40*, 658–674.
  116. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. D Struct. Biol.* *75*, 861–877.
  117. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J., and Adams, P.D. (2008). Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. *Acta Crystallogr. D Biol. Crystallogr.* *64*, 61–69.
  118. Terwilliger, T.C., Adams, P.D., Read, R.J., McCoy, A.J., Moriarty, N.W., Grosse-Kunstleve, R.W., Afonine, P.V., Zwart, P.H., and Hung, L.W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the *PHENIX AutoSol* wizard. *Acta Crystallogr. D Biol. Crystallogr.* *65*, 582–601.

119. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367.
120. Terwilliger, T.C., Sobolev, O.V., Afonine, P.V., and Adams, P.D. (2018). Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr. D Struct. Biol.* **74**, 545–559.
121. Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797.
122. Agirre, J., Iglesias-Fernández, J., Rovira, C., Davies, G.J., Wilson, K.S., and Cowtan, K.D. (2015). Privateer: software for the conformational validation of carbohydrate structures. *Nat. Struct. Mol. Biol.* **22**, 833–834.
123. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51.
124. Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579.
125. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
126. Meng, E.C., Pettersen, E.F., Couch, G.S., Huang, C.C., and Ferrin, T.E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* **7**, 339.
127. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82.
128. Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
129. Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., and Karplus, K. (2009). Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77** (Suppl 9), 114–122.
130. Cuppari, A., Körschgen, H., Fahrenkamp, D., Schmitz, C., Guevara, T., Karmilín, K., Kuske, M., Olf, M., Dietzel, E., Yiallourous, I., et al. (2019). Structure of mammalian plasma fetuin-B and its mechanism of selective metalloproteinase inhibition. *IUCrJ* **6**, 317–330.
131. Qi, H., Williams, Z., and Wassarman, P.M. (2002). Secretion and assembly of zona pellucida glycoproteins by growing mouse oocytes microinjected with epitope-tagged cDNAs for mZP2 and mZP3. *Mol. Biol. Cell* **13**, 530–541.
132. Aricescu, A.R., Lu, W., and Jones, E.Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1243–1250.
133. Bokhove, M., Sadat Al Hosseini, H., Saito, T., Dioguardi, E., Gegen-schatz-Schmid, K., Nishimura, K., Raj, I., de Sanctis, D., Han, L., and Jovine, L. (2016). Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins. *J. Struct. Biol.* **194**, 1–7.
134. Briggs, D.C., and Hohenester, E. (2018). Structural Basis for the Initiation of Glycosaminoglycan Biosynthesis by Human Xylosyltransferase 1. *Structure* **26**, 801–809.e3.
135. Pulido, D., Hussain, S.A., and Hohenester, E. (2017). Crystal Structure of the Heterotrimeric Integrin-Binding Region of Laminin-111. *Structure* **25**, 530–535.
136. Han, K. (1996). An efficient DDAB-mediated transfection of *Drosophila* S2 cells. *Nucleic Acids Res.* **24**, 4362–4363.
137. Karmilín, K., Schmitz, C., Kuske, M., Körschgen, H., Olf, M., Meyer, K., Hildebrand, A., Felten, M., Fridrich, S., Yiallourous, I., et al. (2019). Mammalian plasma fetuin-B is a selective inhibitor of ovastacin and mep- rin metalloproteinases. *Sci. Rep.* **9**, 546.
138. Gorrec, F. (2015). The MORPHEUS II protein crystallization screen. *Acta Crystallogr. F Struct. Biol. Commun.* **71**, 831–837.
139. Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., and Perrakis, A. (2005). Towards rationaliza-tion of crystallization screening for small- to medium-sized academic lab-oratories: the PACT/JCSG+ strategy. *Acta Crystallogr. D Biol. Crystallogr.* **61**, 1426–1431.
140. Mueller, U., Förster, R., Hellmig, M., Huschmann, F.U., Kastner, A., Mal-ecki, P., Pühringer, S., Röwer, M., Sparta, K., Steffien, M., Uhlein, M., Wilk, P., and Weiss, M.S. (2015). The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur. Phys. J. Plus* **130**, 141–150.
141. Nurizzo, D., Mairs, T., Guijarro, M., Rey, V., Meyer, J., Fajardo, P., Cha-vanne, J., Biasci, J.C., McSweeney, S., and Mitchell, E. (2006). The ID23-1 structural biology beamline at the ESRF. *J. Synchrotron Radiat.* **13**, 227–238.
142. de Sanctis, D., Beteva, A., Caserotto, H., Dobias, F., Gabadinho, J., Gir-aud, T., Gobbo, A., Guijarro, M., Lentini, M., Lavault, B., et al. (2012). ID29: a high-intensity highly automated ESRF beamline for macromolecu-lar crystallography experiments exploiting anomalous scattering. *J. Synchrotron Radiat.* **19**, 455–461.
143. McCarthy, A.A., Barrett, R., Beteva, A., Caserotto, H., Dobias, F., Felisaz, F., Giraud, T., Guijarro, M., Janocha, R., Khadrouche, A., et al. (2018). ID30B - a versatile beamline for macromolecular crystallography exper-iments at the ESRF. *J. Synchrotron Radiat.* **25**, 1249–1260.
144. Allan, D.R., Collins, S.P., Evans, G., Hall, D., McAuley, K., Owen, R.L., Sorensen, T., Tang, C.C., von Delft, F., Wagner, A., et al. (2015). Status of the crystallography beamlines at Diamond Light Source. *Eur. Phys. J. Plus* **130**, 56.
145. Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F<sub>1</sub> ATPase. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42.
146. Cowtan, K. (2006). The Buccaneer software for automated model build-ing. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011.
147. Cowtan, K. (2010). Recent developments in classical density modifica-tion. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470–478.
148. Murata, K., Sugiyama, H., Yasumasu, S., Iuchi, I., Yasumasu, I., and Ya-magami, K. (1997). Cloning of cDNA and estrogen-induced hepatic gene expression for choriogenin H, a precursor protein of the fish egg envelope (chorion). *Proc. Natl. Acad. Sci. USA* **94**, 2050–2055.
149. Schröder, G.F., Levitt, M., and Brunger, A.T. (2014). Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 2241–2255.
150. DeLaBarre, B., and Brunger, A.T. (2006). Considerations for the refine-ment of low-resolution crystal structures. *Acta Crystallogr. D Biol. Crys-tallogr.* **62**, 923–932.
151. Sugiyama, H., Yasumasu, S., Murata, K., Iuchi, I., and Yamagami, K. (1998). The third egg envelope subunit in fish: cDNA cloning and analysis, and gene expression. *Dev. Growth Differ.* **40**, 35–45.
152. Punjani, A., Zhang, H., and Fleet, D.J. (2020). Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221.
153. Kidmose, R.T., Juhl, J., Nissen, P., Boesen, T., Karlsen, J.L., and Peder-sen, B.P. (2019). *Namdinator* - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experi-mental maps. *IUCrJ* **6**, 526–531.
154. Robert, X., and Guet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324.
155. Wilson, C.J., Choy, W.Y., and Karttunen, M. (2022). AlphaFold2: A Role for Disordered Protein/Region Prediction? *Int. J. Mol. Sci.* **23**, 4591.

156. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L.E., Brookes, D.H., Wilson, L., Chen, J., Liles, K., et al. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* *27*, 112–128.
157. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* *44*, W344–W350.
158. Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* *33*, D501–D504.
159. Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* *179*, 125–142.
160. Abbasi, F., Miyata, H., Shimada, K., Morohoshi, A., Nozawa, K., Matsu-mura, T., Xu, Z., Pratiwi, P., and Ikawa, M. (2018). RSPH6A is required for sperm flagellum formation and male fertility in mice. *J. Cell Sci.* *131*, jcs221648.
161. Kaneko, T., and Mashimo, T. (2015). Simple Genome Editing of Rodent Intact Embryos by Electroporation. *PLoS One* *10*, e0142755.
162. Toyoda, Y., Yokoyama, M., and Hoshi, T. (1971). Studies on the fertiliza-tion of mouse *in vitro*. I. In vitro fertilization of mouse eggs by fresh epidid-ymal sperm. *Japan. J. Anim. Reprod.* *16*, 147–151.
163. Okada, A., Sano, K., Nagata, K., Yasumasu, S., Ohtsuka, J., Yamamura, A., Kubota, K., Iuchi, I., and Tanokura, M. (2010). Crystal structure of ze-brafish hatching enzyme 1 from the zebrafish *Danio rerio*. *J. Mol. Biol.* *402*, 865–878.
164. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yorda-nova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). Alpha-Fold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nu-cleic Acids Res.* *50*, D439–D444.
165. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly ac-curate protein structure prediction for the human proteome. *Nature* *596*, 590–596.
166. Jones, D.T., and Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformat-ics* *31*, 857–863.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rat monoclonal anti-mZP2-N1 (IE-3)	East and Dean <sup>88</sup>	Cat# CRL-2463; RRID: CVCL_9183
Rabbit polyclonal anti-mZP2 cleavage site Ab164	This paper	N/A
Mouse monoclonal anti-mZP2 ZP-C 2C7	This paper	N/A
Mouse monoclonal anti-β-actin	Abcam	Cat# ab6276; RRID: AB_2223210
Mouse monoclonal anti-His-tag	Proteintech	Cat# 66005-1-Ig; RRID: AB_11232599
Goat polyclonal anti-mouse IgG-HRP	Jackson ImmunoResearch	Cat# 115-035-003; RRID: AB_10015289
Goat polyclonal anti-rabbit IgG-HRP	Jackson ImmunoResearch	Cat# 111-035-144; RRID: AB_2307391
Goat polyclonal anti-mouse IgG (H+L)-HRP	Proteintech	Cat# SA00001-1; RRID: AB_2722565
<b>Bacterial and virus strains</b>		
<i>E. coli</i> : One Shot™ TOP10 Chemically Competent <i>E. coli</i>	Thermo Fisher Scientific	Cat# C404010
<i>E. coli</i> : HST08 Premium Competent Cells	Takara Bio	Cat# 9128
<b>Chemicals, peptides, and recombinant proteins</b>		
4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF)	Roche	Cat# 11427393103; CAS: 34284-75-8
BioLock biotin blocking solution	IBA Lifesciences	Cat# 2-0205
Chemi-Lumi One Super	Nacalai Tesque	Cat# 02230
Collagenase type I	Sigma-Aldrich	Cat# SCR103; CAS: 9001-12-1
Desthiobiotin	Sigma-Aldrich	Cat# D1411; CAS: 533-48-2
Dimethyldioctadecylammonium bromide (DDAB)	Sigma-Aldrich	Cat# D2779; CAS: 3700-67-2
Ethyl 3-aminobenzoate methanesulfonate	Sigma-Aldrich	Cat# E10521; CAS: 886-86-2
EzWestLumi One	ATTO	Cat# 2332632
Ni-NTA Agarose	QIAGEN	Cat# 30230
Ni-NTA Agarose HP	FUJIFILM Wako	Cat#149-09684
Ni Sepharose excel	GE Healthcare	Cat# 17-3712-02
Polyethyleneimine, branched	Sigma-Aldrich	Cat# 408727; CAS: 9002-98-6
PEI MAX	Polysciences	Cat# 24765; CAS: 49553-93-7
PNGase F	New England Biolabs	Cat# P0704L
PNGase F	Takara Bio	Cat# 4450
Sf-900 II SFM medium	Thermo Fisher Scientific	Cat# 10902088
Trypsin	Sigma-Aldrich	Cat# T1426; CAS: 9002-07-7
Cys-Ahx-SRLADEN (mZP2 post- fertilization cleavage site) peptide	This paper	N/A

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Continued</b>		
<b>Critical commercial assays</b>		
Reductive Alkylation Kit	Hampton Research	Cat# HR2-434
<b>Deposited data</b>		
hZP2-N1N2N3 coordinates and structure factors	This paper	PDB: 8RKE
hZP2-N1N2 coordinates and structure factors	This paper	PDB: 8RKF
Cleaved xZP2-N2N3 coordinates and structure factors	This paper	PDB: 8RKG
mZP2-N2N3 coordinates and structure factors	This paper	PDB: 8RKH
fZP1/fZP3 complex coordinates and structure factors (crystal form I)	This paper	PDB: 8RKI
fZP1/fZP3 complex coordinates and structure factors (crystal form II)	This paper	PDB: 8BQU
Cleaved xZP2-N2N3 cryo-EM maps (C1)	This paper	EMDB: EMD-19276
Cleaved xZP2-N2N3 cryo-EM maps (C2)	This paper	EMDB: EMD-19277
Crystal structure of full-length sperm receptor ZP3 at 2.0 Å resolution	Han et al. <sup>9</sup>	PDB: 3NK4
Crystal structure of the ZP-N1 domain of mouse sperm receptor ZP2 at 0.95 Å resolution	Raj et al. <sup>23</sup>	PDB: 5II6
Cryo-EM of native human uromodulin (UMOD)/Tamm-Horsfall protein (THP) filament	Stsiapanava et al. <sup>32</sup>	PDB: 6TQK EMDB: EMD-10553
Crystal structure of human trefoil factor 1	Järvå et al. <sup>42</sup>	PDB: 6V1D
Crystal structure of human trefoil factor 3 in complex with its cognate ligand	Järvå et al. <sup>42</sup>	PDB: 6V1C
Raw data for Figures 2B, 2C, 2E, 3B, 3C, 3E, 3F, 6Q, 6R, S4C, S5B, and S5E–S5H	This paper	Mendeley Data: <a href="https://doi.org/10.17632/89m5vgrg7c.1">https://doi.org/10.17632/89m5vgrg7c.1</a>
<b>Experimental models: Cell lines</b>		
Human: HEK293T cells	ATCC	Cat# CRL-3216; RRID: CVCL_0063
Human: HEK293S GnTI- cells	ATCC	Cat# CRL-3022; RRID: CVCL_A785
Human: Expi293F GnTI- cells	Thermo Fisher Scientific	Cat# A39240; RRID: CVCL_B0J7
<i>D. melanogaster</i> : S2 cells	Thermo Fisher Scientific	Cat# R69007; RRID: CVCL_Z232
<b>Experimental models: Organisms/strains</b>		
<i>Mus musculus</i> : B6D2F1/Slc	Japan SLC	N/A
<i>Mus musculus</i> : B6D2F1/Jcl	CLEA Japan	N/A
<i>Mus musculus</i> : B6D2-Llcfc1 <sup>em2Osb</sup>	Noda et al. <sup>89</sup>	RBRC: 10352 CARD: 2722
<i>Mus musculus</i> : B6D2-Zp2 <sup>em4Osb</sup>	This paper	RBRC: 11009 CARD: 2916
<i>Oryzias latipes</i> : orange-red variety	National BioResource Project Medaka <sup>90</sup>	N/A
<i>Xenopus laevis</i> : Xla.NXR-WT <sup>NXR</sup>	Nasco	Cat# NXR_0.0031
<b>Oligonucleotides</b>		
Oligonucleotide: ssDNA for generating <i>mZp2</i> ΔN1: TCCACATCAGATTACAGATATTAGATCTCTCCTGCA GACAACAATTTCTGACTGAGGAAGGCTTACTGAG TTCACTGAAGTCACAAGGGTGAATAAGAGGG	Sigma-Aldrich	N/A
Primer: genotyping Fw primer for <i>mZp2</i> ΔN1: CGAGGTGGCAGAGGAAAGCATCTGT	Sigma-Aldrich	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer: genotyping Rv primer for <i>mZp2</i> ΔN1: ATTCATTCATCTGCTAAGCCAGCT	Sigma-Aldrich	N/A
Primer: genotyping Fw primer for <i>mZp2</i> -N1 (WT): CCACTCTAATGGTCAGATGG	Sigma-Aldrich	N/A
Primer: genotyping Rv primer for <i>mZp2</i> -N1 (WT): ATTCATTCATCTGCTAAGCCAGCT	Sigma-Aldrich	N/A

**Recombinant DNA**

Plasmid: pHLsec2-hZP2-N1N2N3 (M1-S365)	This paper	N/A
Plasmid: pHLsec2-hZP2-N1N2 (M1-P265)	This paper	N/A
Plasmid: pHLsec2-mZP2-N2N3 (P135-S360)	This paper	N/A
Plasmid: pLJ6-xZP2-N2N3 (M1-A33, D130-N338)	This paper	N/A
Plasmid: pIEx-5-hOvst (A24-D431)	This paper	N/A
Plasmid: pLJ6-fZP3 (M1-V420)	This paper	N/A
Plasmid: pLJ6-fZP3-N184A (M1-V420 [N184A])	This paper	N/A

**Software and algorithms**

AlphaFill	Hekkelman et al. <sup>91</sup>	<a href="https://alphafill.eu">https://alphafill.eu</a>
AlphaFold	Jumper et al. <sup>39</sup>	<a href="https://github.com/google-deepmind/alphafold">https://github.com/google-deepmind/alphafold</a>
AlphaPickle	Arnold <sup>92</sup>	<a href="https://github.com/mattarnoldbio/alphapickle">https://github.com/mattarnoldbio/alphapickle</a>
AMBER14	Maier et al. <sup>93</sup>	<a href="https://ambermd.org">https://ambermd.org</a> RRID: SCR_014230
ASTRA	Wyatt Technology	<a href="https://www.wyatt.com/products/software/astra.html">https://www.wyatt.com/products/software/astra.html</a> RRID: SCR_016255
autoPROC/STARANISO	Vonrhein et al. <sup>94</sup> Vonrhein et al. <sup>95</sup>	<a href="https://www.globalphasing.com/autoproc">https://www.globalphasing.com/autoproc</a> RRID: SCR_015748
autoSHARP	Vonrhein et al. <sup>96</sup>	<a href="https://www.globalphasing.com/sharp">https://www.globalphasing.com/sharp</a>
Belvu	Barson and Griffiths <sup>97</sup>	<a href="https://www.sanger.ac.uk/tool/seqtools">https://www.sanger.ac.uk/tool/seqtools</a> RRID: SCR_015989
Blender	Blender Foundation	<a href="https://www.blender.org">https://www.blender.org</a> RRID: SCR_008606
BUSTER	Smart et al. <sup>98</sup>	<a href="https://www.globalphasing.com/buster">https://www.globalphasing.com/buster</a> RRID: SCR_015653
CNS	Brunger <sup>99</sup>	<a href="http://cns-online.org">http://cns-online.org</a> RRID: SCR_014223
ColabFold	Mirdita et al. <sup>100</sup>	<a href="https://github.com/sokrypton/ColabFold">https://github.com/sokrypton/ColabFold</a>
Coot	Casañal et al. <sup>101</sup>	<a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot</a> RRID: SCR_014222
CRISPRdirect	Naito et al. <sup>102</sup>	<a href="https://crispr.dbcls.jp">https://crispr.dbcls.jp</a> RRID: SCR_018186
cryoSPARC	Punjani et al. <sup>103</sup>	<a href="https://cryosparc.com">https://cryosparc.com</a> RRID: SCR_016501
FoldX	Delgado et al. <sup>104</sup>	<a href="http://foldx.crg.es">http://foldx.crg.es</a> RRID: SCR_008522

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GlyProt	Bohne-Lang and von der Lieth <sup>105</sup>	<a href="http://www.glycosciences.de/modeling/glyprot">http://www.glycosciences.de/modeling/glyprot</a> RRID: SCR_001560
InterProScan	Blum et al. <sup>106</sup>	<a href="https://www.ebi.ac.uk/interpro/search">https://www.ebi.ac.uk/interpro/search</a> RRID: SCR_005829
ISOLDE	Croll <sup>107</sup>	<a href="https://isolde.cimr.cam.ac.uk">https://isolde.cimr.cam.ac.uk</a>
MAFFT	Katoh and Standley <sup>108</sup>	<a href="https://mafft.cbrc.jp/alignment/server">https://mafft.cbrc.jp/alignment/server</a> RRID: SCR_011811
MassLynx	Waters	<a href="https://www.waters.com/waters/en_US/MassLynx-MS-Software/nav.htm?cid=513662">https://www.waters.com/waters/en_US/MassLynx-MS-Software/nav.htm?cid=513662</a> RRID: SCR_014271
MobiDB-lite	Necci et al. <sup>109</sup>	<a href="https://github.com/BioComputingUP/MobiDB-lite">https://github.com/BioComputingUP/MobiDB-lite</a>
MODELLER	Webb and Sali <sup>110</sup>	<a href="https://salilab.org/modeller">https://salilab.org/modeller</a> RRID: SCR_008395
MolProbity	Chen et al. <sup>111</sup>	<a href="http://molprobity.biochem.duke.edu">http://molprobity.biochem.duke.edu</a> RRID: SCR_014226
Mview	Brown et al. <sup>112</sup>	<a href="https://desmid.github.io/mview">https://desmid.github.io/mview</a> RRID: SCR_024129
Muscle5	Edgar <sup>113</sup>	<a href="https://www.drive5.com/muscle">https://www.drive5.com/muscle</a> RRID: SCR_011812
PDBsum	de Beer et al. <sup>114</sup>	<a href="https://www.ebi.ac.uk/pdbsum">https://www.ebi.ac.uk/pdbsum</a> RRID: SCR_006511
Phaser	McCoy et al. <sup>115</sup>	<a href="https://phenix-online.org/documentation/reference/phaser.html">https://phenix-online.org/documentation/reference/phaser.html</a> RRID: SCR_014219
PHENIX	Liebschner et al. <sup>116</sup> Terwilliger et al. <sup>117</sup> Terwilliger et al. <sup>118</sup> Afonine et al. <sup>119</sup> Terwilliger et al. <sup>120</sup>	<a href="https://phenix-online.org">https://phenix-online.org</a> RRID: SCR_014224
PISA	Krissinel and Henrick <sup>121</sup>	<a href="https://www.ebi.ac.uk/pdbe/pisa">https://www.ebi.ac.uk/pdbe/pisa</a> RRID: SCR_015749
Privateer	Agirre et al. <sup>122</sup>	<a href="https://github.com/glycojones/privateer">https://github.com/glycojones/privateer</a>
PyMOL	Schrödinger	<a href="https://pymol.org">https://pymol.org</a> RRID: SCR_000305
SerialEM	Mastronarde <sup>123</sup>	<a href="https://bio3d.colorado.edu/SerialEM">https://bio3d.colorado.edu/SerialEM</a> RRID: SCR_017293
SnapGene	GSL Biotech LLC / Dotmatics	<a href="https://www.snapgene.com">https://www.snapgene.com</a> RRID: SCR_015052
STARANISO	Vonrhein et al. <sup>95</sup>	<a href="https://staraniso.globalphasing.org">https://staraniso.globalphasing.org</a> RRID: SCR_018362
STRIDE	Frishman and Argos <sup>124</sup>	<a href="http://webclu.bio.wzw.tum.de/stride">http://webclu.bio.wzw.tum.de/stride</a>
stride_consensus	This study	<a href="https://doi.org/10.5281/zenodo.10040747">https://doi.org/10.5281/zenodo.10040747</a> <a href="https://github.com/lucajovine/stride_consensus">https://github.com/lucajovine/stride_consensus</a>
UCSF Chimera	Pettersen et al. <sup>125</sup> Meng et al. <sup>126</sup>	<a href="https://www.cgl.ucsf.edu/chimera">https://www.cgl.ucsf.edu/chimera</a> RRID: SCR_004097
UCSF ChimeraX	Pettersen et al. <sup>127</sup>	<a href="https://www.cgl.ucsf.edu/chimerax">https://www.cgl.ucsf.edu/chimerax</a> RRID: SCR_015872
XDS	Kabsch <sup>128</sup>	<a href="https://xds.mr.mpg.de">https://xds.mr.mpg.de</a> RRID: SCR_015652
YASARA Structure	Krieger et al. <sup>129</sup>	<a href="http://www.yasara.org">http://www.yasara.org</a> RRID: SCR_017591

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other		
Streptactin Sepharose High Performance beads	GE Healthcare	Cat# 28935599
SulfoLink Coupling Resin	Thermo Fisher Scientific	Cat# 20401
Superdex 200 Increase 10/300 GL column	GE Healthcare	Cat# 28990944
Quantifoil R2/2-Au300	Jena Bioscience	Cat# X-103-Au300

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Luca Jovine ([luca.jovine@ki.se](mailto:luca.jovine@ki.se)).

**Materials availability**

Unique reagents generated in this study are available from the [lead contact](#) with a completed Materials Transfer Agreement.

**Data and code availability**

Structure factors and coordinates are deposited in the Protein Data Bank under accession numbers PDB: 8RKE (hZP2-N1N2N3), PDB: 8RKF (hZP2-N1N2), PDB: 8RKG (cleaved xZP2-N2N3), PDB: 8RKH (mZP2-N2N3), PDB: 8RKI and PDB: 8BQU (fZP1/fZP3 complex crystal forms I and II); cryo-EM maps of cleaved xZP2-N2N3 are deposited in the Electron Microscopy Data Bank under accession code EMDB: EMD-19276 (C1) and EMDB: EMD-19277 (C2). Raw data for [Figures 2B, 2C, 2E, 3B, 3C, 3E, 3F, 6Q, 6R, S4C, S5B, and S5E–S5H](#) were deposited as Mendeley Data: <https://doi.org/10.17632/89m5vgrg7c.1>. The code for stride\_consensus has been deposited in Zenodo and GitHub at the DOIs listed in the [key resources table](#). All deposited data is publicly available as of the date of publication.

Other relevant data and materials, as well as any additional information required to reanalyze the data reported in this paper, are available from the [lead contact](#) upon request.

**EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

*Escherichia coli* cells were cultured at 37°C in 2xYT medium containing 50 µg/mL ampicillin or 12.5 µg/mL kanamycin.

Sexually mature, oocyte-positive *Xenopus laevis* females (>3 years old) were housed at 18°C with a 12 h/12 h light/dark cycle, following procedures conducted using accepted standards of humane animal care and approved by the Animal Care and Use Committee at the University of Pittsburgh (protocol 20026806).

Mice were housed with a 12 h/12 h light/dark cycle and food/water available *ad libitum*. All experiments involving mice were approved by the Institutional Animal Care and Use Committees of Research Institute for Microbial Diseases of Osaka University (approval ID Biken-AP-H30-01) and were conducted in compliance with the guidelines and regulations for animal experimentation. For mouse fertility test, sexually mature ZP2 ΔN1 female mice (5–16-week-old) and WT B6D2F1 mice were caged with 8-week-old WT male mice for 2 months.

Sexually mature adult individuals of the orange-red variety of medaka (*Oryzias latipes*) were housed in tanks either outdoors from late spring to early autumn or indoors at 27°C with a 14 h/10 h light/dark cycle to keep them in breeding condition. Spawning individuals were spinalized under anesthesia with 200 µg/mL ethyl 3-aminobenzoate methanesulfonate (Sigma-Aldrich) and dissected to obtain unfertilized eggs, according to procedures approved by the Ethics Committee of Sophia University (approval ID 2016-006).

HEK293T cells (a kind gift of R. Aricescu and Y. Zhao, University of Oxford) and HEK293S GnTI- cells (ATCC) were cultivated at 37°C, 5% CO<sub>2</sub> in DMEM (Thermo Fisher Scientific) supplemented with 10% FBS (Biological Industries) and 4 mM L-glutamine (Thermo Fisher Scientific). Expi293F GnTI- cells (Thermo Fisher Scientific) were grown at 37°C, 8% CO<sub>2</sub> in FreeStyle 293 Expression Medium (Thermo Fisher Scientific). *Drosophila melanogaster* S2 cells (Thermo Fisher Scientific) were maintained at 27°C and ambient CO<sub>2</sub> in Sf-900 II SFM medium (Thermo Fisher Scientific).

**METHOD DETAILS**

**DNA constructs**

cDNAs encoding hZP2-N1N2N3 (M1-S365) and hZP2-N1N2 (M1-P265) were amplified by PCR from I.M.A.G.E. Consortium clone 40007473 (ATCC) and subcloned into mammalian expression vector pHLsec2,<sup>130</sup> in frame with a C-terminal Leu+Glu+6His (LEH<sub>6</sub>)-tag-encoding sequence. A cDNA fragment encoding mZP2-N2N3 (P135-S360) cDNA, amplified from a full-length *Zp2*

cDNA,<sup>131</sup> was subcloned into the same vector, in frame with 5' and 3' sequences encoding an N-terminal chicken Cryp $\alpha$  signal peptide and the C-terminal LEH<sub>6</sub>-tag, respectively.

A synthetic gene encoding xZP2-N2N3 (D130-N338) (ATUM) was subcloned into mammalian expression vector pLJ6, in frame with sequences that encoded the native xZP2 signal peptide (M1-A33) and a C-terminal 8His-tag.

A cDNA fragment encoding hOvst without signal peptide (A24-D431) was amplified from ASTL (NM\_001002036) cDNA (GenScript) and subcloned into insect cell expression vector pEX-5, in frame with 5' and 3' sequences encoding an AKH signal peptide and a C-terminal Strep-tag II (PGWSHPQFEK), respectively.

For evaluating the contribution of the conserved ZP3 N-glycan to protein secretion, synthetic genes encoding full-length WT fZP3 (M1-V420) and its N184A counterpart (ATUM) were cloned into pLJ6, in frame with the sequence encoding the C-terminal LEH<sub>6</sub> tag.

Oligonucleotides were purchased from Sigma-Aldrich and all constructs were verified by DNA sequencing (Eurofins Genomics).

### Protein expression and purification

For expression construct evaluation, biochemical studies and cryo-EM, proteins were expressed in HEK293T cells (ATCC) transiently transfected with 25 kDa branched polyethyleneimine (PEI; Sigma-Aldrich).<sup>132,133</sup> The same method was used to transfect HEK293S GnTI- cells (ATCC) in order to produce hZP2-N1N2N3, hZP2-N1N2 and mZP2-N2N3 proteins for X-ray crystallography. After capture from the conditioned media of transfected cells by immobilized metal affinity chromatography (IMAC) with Ni-NTA Agarose (QIAGEN) or Ni Sepharose excel (GE Healthcare) beads, proteins were subjected to size-exclusion chromatography (SEC), concentrated and stored at -80°C until use. For crystallographic studies, a deglycosylation step with Endoglycosidase H (Endo H) was also included before SEC. Reductive methylation of hZP2-N1N2N3 was carried out using a Reductive Alkylation Kit (Hampton Research); production of SeMet-hZP2-N1N2 was performed following an established protocol.<sup>132</sup>

xZP2-N2N3 was expressed by transiently transfecting Expi293F GnTI- cells with PEI MAX (Polysciences) and culturing them for 5 days.<sup>134,135</sup> The secreted protein was purified and deglycosylated as detailed above.

For expression of hOvst, *Drosophila* Schneider 2 (S2) cells were cultured in Sf-900 II SFM medium (Thermo Fisher Scientific) and transiently transfected using dimethyldioctadecylammonium bromide (DDAB; Sigma-Aldrich).<sup>136</sup> 90-96 h after transfection, the conditioned medium was harvested and cell debris was removed by centrifugation at 4,000 x g. The supernatant was adjusted to 0.1 M Tris-HCl pH 8.1, 150 mM NaCl (buffer A) and 240  $\mu$ L BioLock (IBA Lifesciences) was added to 100 mL medium. After 20 min incubation at room temperature (RT), the medium was further centrifuged at 4,000 x g for 15 min. The cleared medium was incubated with 1/100 volume of Streptactin Sepharose High Performance beads (GE Healthcare) pre-equilibrated with buffer A for 1 h at RT. Streptactin beads were collected to an open column by gravity flow and washed with 20 volumes buffer A. Bound protein was batch-eluted 5 times with 1 mL buffer A supplemented with 5 mM desthiobiotin. Fractions containing hOvst were concentrated with 10 kDa-cutoff centrifugal filtration devices and stored at -80°C until use.

The native fZP1/fZP3 complex, corresponding to fraction F2 of the medaka VE hatching enzyme digest, was prepared from unfertilized eggs isolated from spawning adult female Japanese rice fish following an established procedure.<sup>32</sup> Briefly, the eggs were crushed in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA. After centrifugation at 2,900 g for 30 s, the supernatant was decanted and this procedure was repeated several times to completely remove yolk protein and cell debris. Fifty VEs were incubated for 1 h at 30°C in 300  $\mu$ L 50 mM Tris-HCl pH 8.0 containing 2.3  $\mu$ g purified HCE and 1.2  $\mu$ g purified LCE. The resulting material was fractionated using a HiLoad 16/60 Superdex 75 SEC column (GE Healthcare) equilibrated with 20 mM phosphate buffer pH 7.2, 150 mM NaCl. F2, the second of the two protein peaks eluting near the void volume, was collected, concentrated and re-chromatographed using a Superdex 75 5/150 GL column (GE Healthcare).

C-terminally His-tagged recombinant fZP3 and its N184 mutant counterpart were expressed in HEK293T cells transiently transfected by 25 kDa branched PEI in 6-well plates. 72 h after transfection, conditioned media were collected and secreted proteins were captured using Ni-NTA Agarose HP (FUJIFILM Wako).

### In vitro cleavage of hZP2 with recombinant hOvst

For cleavage of hZP2-N1N2N3, hOvst (0.2 mg mL<sup>-1</sup>) was activated using 1/100 (v/v) volume of trypsin (Sigma-Aldrich) for 1 h at 37°C. Trypsin activity was quenched with 10 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (AEBSF, Roche).<sup>137</sup> hZP2-N1N2N3 (1 mg mL<sup>-1</sup>) was treated for 36-40 h at RT with trypsinized hOvst (hOvst-T) or untreated hOvst (100:1 substrate:protease ratio) and subjected to SEC-MALS. To confirm the effect of metal ions — and thus the identity of the protease responsible for cleaving hZP2 —, 5 mM EDTA was added to the reaction mixture. hZP2-N1N2N3 or hZP2-N2N3 (1 mg mL<sup>-1</sup>) were incubated with hOvst (100:1 substrate:protease ratio) and reaction mixture aliquots harvested at different times (0-40 h) were subjected to SDS-PAGE under reducing and non-reducing conditions to analyze the cleavage products.

### In vitro cleavage of xZP2 with collagenase

Purified xZP2-N2N3 was adjusted to 2 mg mL<sup>-1</sup> in 20 mM Na-HEPES pH 7.8, 150 mM NaCl, 5 mM CaCl<sub>2</sub> and incubated for 2 h at RT with 200  $\mu$ g mL<sup>-1</sup> collagenase type I (Sigma-Aldrich). The cleaved protein was then further purified by SEC, using a Superdex 200 Increase 10/300 GL column (GE Healthcare) pre-equilibrated with 20 mM Na-HEPES pH 7.8, 150 mM NaCl.

### **In vitro cleavage of xZP2 with *Xenopus* oocyte exudate**

To collect eggs, laying was stimulated by injection of 1,000 IU human chorionic gonadotropin into the dorsal lymph sac of sexually mature *X. laevis* females. Following injection, animals were housed overnight for 12–16 h at 14–16°C. Typically, egg laying began within 2 h of moving to RT, and batches of 12–155 eggs from 8 frogs were collected on dry petri dishes and used within 10 min of being laid. Exudate collection was performed in two steps. First, the jelly coat was removed by agitating the eggs in 45 mM  $\beta$ -mercaptoethanol in dilute modified Ringer's solution (33.3 mM NaCl, 0.6 mM KCl, 0.67 mM CaCl<sub>2</sub>, 0.33 mM MgCl<sub>2</sub>, 1.33 mM Na-HEPES) pH 8.5 for 2–3 min; once de-jellying was confirmed by visual observation, eggs were neutralized in dilute modified Ringer's solution pH 6.5, following by three washes in dilute modified Ringer's pH 7.8. Second, de-jellied eggs were incubated in a 35 mm plastic petri dish containing 2  $\mu$ M ionomycin in dilute modified Ringer's solution pH 7.8 for 60 min on a flat rotating platform at 27 rpm. Following incubation, dish solution was collected into 1.7 mL Eppendorf tubes while carefully avoiding cell lysis. Samples were held on ice, spun at 6000 rcf for 30 min at 4°C using 0.5 mL Pierce protein concentrators with a polyethersulfone (PES) membrane and a 3 kDa-cutoff (Thermo Fisher Scientific), and finally transferred to clean Eppendorf tubes. After evaluating total protein concentration with a Bradford solution (1:4 dilution), samples were frozen in liquid nitrogen and stored at -80°C or on dry ice.

For *in vitro* cleavage experiments, thawed exudate samples were further concentrated 10 times using 10 kDa-cutoff centrifugal filtration devices (Amicon) and then cleared from precipitate using 0.2  $\mu$ m centrifugal filtration devices (Millipore). 10  $\mu$ L concentrated exudate was added to 90  $\mu$ L purified xZP2-N2N3 (2 mg mL<sup>-1</sup>) and, after 10 days of incubation at RT, the sample was subjected to SEC-MALS.

### **Protein crystallization and derivatization**

Crystallization trials by hanging or sitting drop vapor diffusion were set up at 20°C or 4°C using a mosquito crystallization robot (TTP Labtech).

Thick plates of methylated hZP2-N1N2N3 (6.4 mg mL<sup>-1</sup>) grew at 4°C in 10% (w/v) PEG 4000, 0.1 M MES/imidazole pH 6.5, 0.1 M carboxylic acids (MD2-100-76, Molecular Dimensions), 20% (v/v) glycerol. For experimental phasing, 0.1  $\mu$ L drops containing crystals of the protein were mixed with 0.5  $\mu$ L 15% (w/v) PEG 4000, 0.1 M MES/imidazole pH 6.5, 0.1 M carboxylic acids, 30% (v/v) glycerol, 1 mM (NH<sub>4</sub>)PtCl<sub>2</sub> for 4 h at 4°C.

hZP2-N1N2 and SeMet-hZP2-N1N2 (5.0 mg mL<sup>-1</sup>) formed crystals with a hexagonal bipyramid shape at 20°C in 10–30% (v/v) MPD, 0.1 M MES pH 6.0 or 0.1 M Na-HEPES pH 7.5.

Thin plate crystals of collagenase-cleaved xZP2-N2N3 (6.9 mg mL<sup>-1</sup>) were obtained at 20°C by mixing 0.1  $\mu$ L protein solution with 0.2  $\mu$ L reservoir solution containing 12.5% (v/v) MPD, 12.5% (v/v) PEG 1000, 12.5% (w/v) PEG 3350, 0.1 M Tris base/Bicine pH 8.5, 0.09 M NPS (nitrate, phosphate, sulfate; MD2-100-72, Molecular Dimensions).

mZP2-N2N3 (7.9 mg mL<sup>-1</sup>) formed prismatic crystals at 20°C in 20% (w/v) PEG 3350, 0.2 M Na-nitrate.

The medaka ZP1/ZP3 complex (6 mg mL<sup>-1</sup>) yielded two distinct trigonal crystal forms. The first had a hexagonal plate morphology and grew at 20°C in 5% (w/v) PEG 20000, 25% (w/v) trimethylpropane, 0.1 M N,N-bis(2-hydroxyethyl)-2-amino ethanesulfonic acid (BES)/triethanolamine pH 7.5, 1% (w/v) non-detergent sulfobetaine (NDSB) 195, 0.5 mM YCl<sub>3</sub>, 0.5 mM ErCl<sub>3</sub>, 0.5 mM TbCl<sub>3</sub>, 0.5 mM YbCl<sub>3</sub> (condition E8 of the MORPHEUS II crystallization screen<sup>138</sup> [Molecular Dimensions]). The second crystal form, with a bipyramidal prismatic morphology, grew at 4°C in 22% (w/v) PEG 3350, 0.2 M sodium/potassium phosphate, 0.1 M Tris-HCl pH 8.0 (a condition obtained by optimizing an initial hit in condition 94 of the PACT suite<sup>139</sup> [QIAGEN]).

### **X-ray data collection and processing**

For data collection at synchrotron, single crystals were harvested using MicroMounts, MicroLoops or MicroMeshes (MiTeGen) and flash frozen in liquid nitrogen.

X-ray diffraction data was collected at Helmholtz-Zentrum Berlin (HZB)/BESSY II beamline BL14.1<sup>140</sup> (hZP2-N1N2N3); European Synchrotron Radiation Facility (ESRF) beamlines ID23-1<sup>141</sup> (hZP2-N1N2), ID29<sup>142</sup> (mZP2-N2N3 and fZP1/fZP3 complex crystal form I) and ID30B<sup>143</sup> (cleaved xZP2-N2N3); and Diamond Light Source beamlines I04-1<sup>144</sup> (SeMet-hZP2-N1N2) and I02<sup>144</sup> (fZP1/fZP3 complex crystal form II).

All datasets were processed using XDS,<sup>128</sup> except for the fZP1/fZP3 complex crystal form II data which — as further detailed below — was processed using autoPROC/STARANISO<sup>95</sup> (Tables S1 and S3).

### **X-ray structure determination of hZP2-N1N2N3**

Initial attempts to solve the structure of hZP2-N1N2N3 by MR using search models derived from the crystal structure of mZP2-N1<sup>23</sup> failed. Data collected from Pt-soaked crystals could however be phased by single isomorphous replacement with anomalous scattering using autoSHARP.<sup>96</sup> This identified 8 Pt sites that resulted in a phasing power for isomorphous differences of 0.688/0.759 (acentric/centric reflections) and a phasing power for anomalous differences of 0.524. Despite the very low figure of merit (FOM) of the initial phases (0.087/0.108), density modification<sup>145</sup> showed a clear difference between the two hands that had scores of 0.1576 and 0.0855, respectively; moreover, alternated cycles of density modification and automated model building<sup>146,147</sup> produced an initial model ( $R_{\text{free}}$  0.51) that showed some evidence for the presence of  $\beta$ -sheets. When this set of coordinates was used as a search model for MR with Phaser,<sup>115</sup> it produced a single solution (LLG 468, TFZ 18.7) that could be improved to  $R_{\text{free}}$  0.40 by PHENIX AutoBuild.<sup>117</sup> The structure, whose asymmetric unit consists of three molecules arranged in ring-like fashion, was completed

by alternating cycles of manual rebuilding in Coot,<sup>101</sup> refinement with non-crystallographic symmetry (NCS) restraints using phenix.refine<sup>119</sup> and validation with MolProbity<sup>111</sup> and Privateer<sup>122</sup> (Table S1). Notably, chain C of the model has a significantly higher average B factor (166 Å<sup>2</sup>) than the other two chains (chain A: 93 Å<sup>2</sup>; chain B: 98 Å<sup>2</sup>), largely due to local disorder at the level of its ZP-N3 domain (219 Å<sup>2</sup>); however, because omission of the latter consistently increased  $R_{\text{free}}$  by ~0.5%, its coordinates were kept in the model but only subjected to rigid-body refinement.

### X-ray structure determination of hZP2-N1N2

As in the case of hZP2-N1N2N3, the structure of hZP2-N1N2 could not be phased by MR using mZP2-N1<sup>23</sup> as a search model. However, MR was successful using an ensemble consisting of the latter (residues N42-Q138 of PDB: 5II6) and the three copies of the ZP-N1 domain from a partially refined model of hZP2-N1N2N3 (residues N46-V143), trimmed using a 1.5 Å threshold. This produced a single solution with two molecules of hZP2-N1N2 per asymmetric unit (LLG 282, TFZ 18.9), whose correctness was verified by its ability to locate 9 Se sites (chain A M116, M160, M203, M239; chain B M85, M116, M160, M203, M239) by cross-Fourier analysis in PHENIX AutoSol<sup>118</sup> (FOM 0.694, Bayesian correlation coefficient [BAYES-CC] 75.49), using a 3.5 Å-resolution dataset collected at 13.477 keV from crystals of SeMet-hZP2-N1N2 (Figure S2A). The structure was completed, refined and validated as described for hZP2-N1N2N3, except for the use of reference restraints derived from AlphaFold models generated without multiple sequence alignment information or models that were energy-minimized with YASARA Structure (using the YASARA2 force field),<sup>129</sup> as well as the fact that refinement was also carried out with BUSTER<sup>98</sup> (Table S1).

Notably, the ZP-N2 domains from mZP2-N2N3 and cleaved xZP2-N2N3 can be superimposed on the ZP-N2 domain of hZP2-N1N2N3 with average RMSDs of 2.1 Å (over 95 C $\alpha$ ) and 2.2 Å (over 90 C $\alpha$ ), respectively (Figures S2B and S2C). On the contrary, the ZP-N2 domain of hZP2-N1N2 shows significant differences in the orientation of the edges of its C/E'/F/G  $\beta$ -sheet relative to the A/B/E/D  $\beta$ -sheet, so that its average RMSD from the ZP-N2 domain of hZP2-N1N2N3 is 3.9 Å (over 95 C $\alpha$ ). Despite these differences in the hZP2-N1N2 structure, which are likely to reflect the lack of an interacting ZP-N3 domain, the relative conformational variability of the region from  $\beta$ C to  $\alpha$ <sub>C'D</sub> and different crystal packing, the ZP-N1/ZP-N2 domain interface of hZP2-N1N2 is essentially the same as that of hZP2-N1N2N3 (Figures 1D and S2D).

### X-ray structure determination of cleaved xZP2-N2N3

The structure of cleaved xZP2-N2N3 was solved by MR, using as search model the top ranked AlphaFold prediction generated by ColabFold<sup>39,100</sup> for the sequence D130-N338, trimmed to exclude two low-confidence regions corresponding to the N terminus of mature xZP2 (D130-V137) and the cleavage loop of ZP-N2 (P149-A163). Phaser found a single unambiguous solution (LLG=427, TFZ=18.4) that consisted of 4 copies of xZP2-N2N3, corresponding to the biologically relevant tetramer, in the asymmetric unit of a P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> cell. After autobuilding with PHENIX AutoBuild, several cycles of manual rebuilding and refinement could not improve  $R_{\text{free}}$  beyond ~0.31; however, twin refinement of two copies of the xZP2-N2N3 tetramer against data reprocessed in space group P2<sub>1</sub>, using operator h, -k, -l, allowed structure determination to proceed to a final  $R_{\text{free}}$ =0.25 (Table S1).

### X-ray structure determination of mZP2-N2N3

The structure of mZP2-N2N3 was solved by native single-wavelength anomalous diffraction (SAD) using a 2.0 Å-resolution dataset collected at 7 keV. PHENIX AutoSol identified a 23-atom substructure (FOM 0.232, BAYES-CC 27.69) whose phases resulted in an initial model of 299 residues ( $R$ =0.33,  $R_{\text{free}}$ =0.34). After further autobuilding using PHENIX AutoBuild, which significantly improved the solution (383 residues,  $R$ =0.26,  $R_{\text{free}}$ =0.29), manual rebuilding, refinement and validation were performed as detailed above (Table S1).

### X-ray structure determination of the fZP1/fZP3 complex (crystal form I)

Initial attempts to phase the data collected from crystal form I of the native fZP1/fZP3 complex, using AlphaFold<sup>39</sup> models of individual egg coat subunits or domains thereof as MR search models (either by themselves or together in the context of sequential searches), failed. However, it was possible to solve the structure by using a combination of two AlphaFold-generated subcomplex models (whose fZP1 coordinates matched the sequence of the most abundant medaka VE fZP1 isoform, derived from choriogenin H [Chg H]<sup>148</sup>). The first AlphaFold prediction was generated by submitting to ColabFold<sup>100</sup> two sequences: the first was a hybrid sequence consisting of (a) the ZP-N domain and IDL of fZP3 (residues Y74-L207, with Y74 being the first residue of the fZP3 37 kDa low choriolytic enzyme [LCE] digestion product<sup>38</sup>), (b) a 10 glycine-spacer and (c) the trefoil and ZP-N domains of fZP1, plus part of the protein's IDL (residues T221-D387, corresponding to the fZP1 17 kDa LCE digestion product<sup>38</sup>); the second was the sequence of fZP1 ZP-C and part of the preceding IDL (residues S388-G557, corresponding to the fZP1 16 kDa LCE digestion fragment<sup>38</sup>). The top 3 ranked models of this subcomplex were then aligned into an ensemble (ensemble I) and trimmed to remove low confidence regions. The second prediction was also carried out by submitting two sequences to ColabFold: the first corresponded to the IDL and ZP-C domain of fZP3 (residues H183-T393), and the second was the same fZP1 trefoil/ZP-N/IDL fragment described above (residues T221-D387). A corresponding ensemble (ensemble II) was then generated by aligning 4 of the top 5 ranked models of this subcomplex and subsequently trimming them to exclude the trefoil and ZP-N domains of ZP3 as well as remove low confidence regions.

Considering that the Matthews coefficient suggested the presence of 1 or 2 complexes per asymmetric unit and taking into account the relatively weak diffraction of the fZP1/fZP3 crystals, ensemble I was first used to search for a single molecule in the asymmetric unit using Phaser (~80% solvent content). Using RMSD 1.5 Å as search model variance from the expected structure, this resulted in a single MR solution (LLG=95, TFZ=11.5) that was refined to  $R_{\text{free}}=0.49$ . MR was then repeated with the resulting coordinates (using 100% sequence identity as search model variance), yielding the same solution but with increased LLG=203 and TFZ=17.0. Finally, a search for a single copy of ensemble II (with RMSD 1.5 Å as search model variance) was successfully carried out using the placed and refined ensemble I coordinates as a partial solution (notably, a parallel search with placed but not refined ensemble I coordinates failed). This resulted in a single solution with LLG=558 and TFZ=22.9 that was automatically refined to  $R_{\text{free}}=0.44$ , producing a difference map that showed clear density for the N-glycan attached to fZP3 N184 (a post-translational modification that was not included in the search model) (Figure S6A) but no evidence for any additional copy of the complex within the asymmetric unit of the crystal. Although the fZP1/fZP3 complex data was collected at 11.563 keV, we could still take advantage of the fact that — as detailed above — the protein was crystallized in the presence of Y/Er/Tb/Yb to further confirm the correctness of the solution by its ability to locate heavy atom sites by MR-SAD in PHENIX AutoSol (FOM 0.419, overall score 75.49±3.41). Based on this result, we added to the model coordinates four sites with occupancy >0.5 (all of which also matched large positive difference density peaks), tentatively assigned as Yb by AutoSol (Figure S6A). The model was improved through iterative rounds of manual model building, taking advantage of maps generated by deformable elastic network (DEN) refinement<sup>149</sup> in CNS<sup>99</sup> (Figure S6B) as well as map sharpening<sup>150</sup> in PHENIX,<sup>120</sup> and refinement and validation as detailed above (Table S3).

Notably, inspection of the map region between the ZP-C sequences FKMFTFV (identical in both the Chg H- and Chg Hminor-derived fZP1 subunits of the VE<sup>148,151</sup>) and PL<sup>R</sup>/<sub>K</sub>E<sup>K</sup>/<sub>H</sub>VYIHC (Chg H/Chg Hminor) showed that these are connected by a relatively short loop, consistent with the sequence of Chg H (6 intervening residues) but not that of Chg Hminor (21 intervening residues). This suggested that the crystallized VE fragments contained only the higher abundant Chg H-derived fZP1 subunit, a hypothesis that was later confirmed by the higher resolution map of crystal form II.

### X-ray structure determination of the fZP1/fZP3 complex (crystal form II) and analysis of its packing

The second crystal form of the medaka VE filament fragment was characterized by a very long unit cell c axis (>470 Å) and extremely anisotropic diffraction, combined with variable diffracting power and significant anisomorphism. Although some specimens showed reflections beyond 3 Å resolution, these features precluded early attempts at structure determination by either experimental phasing or MR. The subsequent availability of a refined model for crystal form I of the complex allowed Phaser to identify a convincing single solution (LLG=222, TFZ=16.3), using a crystal form II dataset that had been conservatively processed to 3.5 Å resolution with autoPROC<sup>94</sup> (Table S3, column “isotropic, standard (TRUNCATE)”). Initial refinement of the placed model, using two rigid body groups ((1) fZP3 ZP-N + IDL residues K182-L193, fZP1 IDL residues P389-L398 + ZP-C; (2) fZP1 trefoil + ZP-N + IDL residues T374-P386, fZP3 IDL residues W194-L207 + ZP-C) as well as four TLS groups (matching fZP3 ZP-N, fZP1 ZP-C, fZP1 trefoil + ZP-N and fZP3 ZP-C, plus interacting IDL moieties from the other subunits), resulted in  $R=0.36$ ,  $R_{\text{free}}=0.42$ . As in the case of crystal form I, the corresponding  $2mFo-DFc$  and  $mFo-DFc$  maps also confirmed the correctness of the MR solution by showing prominent density for the fZP3 N184 N-glycan, which had been omitted from the search model coordinates; moreover, the maps also showed density for fZP3 C-terminal residues G359-S374, which were disordered in crystal form I (whose packing prevented the interaction between the ZP-C and ZP-N domains of different ZP3 molecules). Consistent with visual inspection of crystal form II's diffraction, analysis of the corresponding data with STARANISO<sup>95</sup> identified an anisotropic cut-off surface that could be fit with an ellipsoid having approximate diffraction limits of 4.0 Å, 4.0 Å and 2.3 Å along its principal axes. Whereas conventional isotropic processing could not cope with such strong anisotropy beyond ~3.0 Å resolution (Table S3, column “isotropic, extended (AIMLESS)”), we were able to include a significant amount of higher resolution information by carrying out anisotropic processing with autoPROC<sup>95</sup> (Table S3, column “Anisotropic (STARANISO)”). Refinement against anisotropically processed data lowered  $R_{\text{free}}$  by >5% and yielded maps with overall improved density (an exception being fZP1 Q410, which was not resolved but whose position was however constrained by well-defined adjacent residues). The model was completed by alternating cycles of rebuilding using both Coot and ISOLDE,<sup>107</sup> refinement with phenix.refine and validation with MolProbity and Privateer; based on visual inspection of maps calculated using different cutoffs during the course of this process, a high-resolution limit of 2.7 Å was ultimately chosen as optimal for the crystal form II dataset (Figures S6D–S6G; Table S3).

To assess whether the packing of crystal form II of the fZP1/fZP3 complex was likely to recapitulate the conformation of intact fZP1 within the uncleaved native VE filament (and thus mirror the structure of the latter), the refined coordinates of the complex and those of its coaxially stacked  $-X+Y$ ,  $-X$ ,  $Z+1/3$  + (1 0 0) & {0 0 0} and  $-Y$ ,  $X-Y$ ,  $Z+2/3$  + (0 -1 -1) & {0 0 0} symmetry-related copies were imported into YASARA Structure, which was then used to introduce covalent bonds between the D387 and S388 residues of adjacent fZP1 moieties. This did not introduce any clash with spatially close non-linker residues and was followed by AMBER14<sup>93</sup> energy minimization of residues P385-P389 (to relax the linker region immediately around 387|388) and then of all residues within 10 Å of 387|388. Notably, this allowed S388 and P389 to make main chain hydrogen bonds with ZP3 V172 and thereby extend the  $\beta 2$  strand of ZP1's IDL. Together with the formation of additional intersubunit hydrogen bonds (hydroxyl group of fZP1 S388/carbonyl oxygen of ZP3 Q170; hydroxyl group of fZP3 S169/carbonyl oxygen of fZP1 P386), this suggests that the packing of proteolytic fragments observed in crystal form II results in a supramolecular arrangement that closely resembles the structure of the intact VE filament.

### Cryo-EM data collection and structure determination of cleaved xZP2-N2N3

Purified cleaved xZP2-N2N3 protein at 0.9 mg mL<sup>-1</sup> (3 μL) was applied to cryo-EM grids (Quantifoil R2/2-Au300) previously glow-discharged under vacuum for 120 s at 20 mA (PELCO easiGLOW), incubated for 30 s, blotted for 3 s, and plunge-frozen in liquid ethane using a Vitrobot Mark4 grid freezing device (FEI) with the chamber maintained at 4°C and 100% relative humidity. An optimized grid was imaged at the Cryo-EM Swedish National Facility node at Stockholm SciLifeLab, using a Talos Arctica electron microscope (Thermo Scientific) equipped with a K2 camera, operating at 200 kV. 1347 movies of 40 frames each were acquired in electron counting mode at 0.9459 Å/pixel, with a total dose of 52 electrons/Å<sup>2</sup> and stacked into a single TIFF stack using SerialEM<sup>123</sup> control software with defocus values ranging from 1.4 to 3.0 μm (Table S2).

Contrast transfer function (CTF) parameters were estimated from averaged movies using Patch CTF and beam-induced particle motion between fractions was corrected using Patch Motion within cryoSPARC.<sup>103</sup> Automatic particle selection was performed with templates from the initial 2D classification. The number of particle images were reduced by further 2D/3D classification and heterogeneous refinement (Figure S3). Initial maps were calculated *ab initio* and the final maps were refined using cryoSPARC's non-uniform refinement feature,<sup>152</sup> with and without the application of C2 symmetry. The final raw maps have an overall nominal resolution of 4.4 Å (C2)/4.6 Å (C1) according to cryoSPARC (Table S2), or 4.9 Å (C2)/5.3 Å (C1) according to the EMD Fourier Shell Correlation (FSC) server. Rigid-body fitting of the crystallographic model of cleaved xZP2-N2N3 into the maps using ChimeraX<sup>127</sup> yielded correlation values of 0.75 (C1) and 0.74 (C2), which could be improved to 0.81 (C1) and 0.82 (C2) using flexible fitting in Namdinator.<sup>153</sup>

### Sequence-structure analysis

Alignments were generated using MAFFT<sup>108</sup> as implemented in SnapGene (GSL Biotech), MUSCLE5<sup>113</sup> or UCSF Chimera<sup>125</sup> MatchMaker/Match → Align<sup>126</sup>; edited and analyzed with Belvu from the SeqTools package<sup>97</sup>; and rendered using ESPript<sup>154</sup> or MView.<sup>112</sup> Disorder predictions were carried out using MobiDB-lite<sup>109</sup> and AlphaFold.<sup>155</sup> Crystallographic statistics reported in Tables S1 and S3 were generated with phenix.table\_one.<sup>116</sup> Maps and models were visualized and analyzed using Coot, PyMOL (Schrödinger) and UCSF Chimera/ChimeraX. Secondary structure assignments were made using STRIDE<sup>124</sup> and, when multiple set of coordinates were available for the ZP-N domains shown in Figure S1, averaged with stride\_average (available on Zenodo at <https://doi.org/10.5281/zenodo.10040747>, or GitHub at [https://github.com/lucajovine/stride\\_consensus](https://github.com/lucajovine/stride_consensus)); β strands were indicated by letters following standard ZP-N nomenclature.<sup>10</sup> Interfaces were analyzed with PISA,<sup>121</sup> PDBsum<sup>114</sup> and FoldX.<sup>104</sup> Electrostatic surface potential calculations were carried out using the APBS Tools plugin of PyMOL.<sup>156</sup> Structural mapping of the evolutionary conservation of amino acid positions was performed with ConSurf,<sup>157</sup> using manually curated sequence alignments derived from NCBI RefSeq.<sup>158</sup> The color gradient used for hydrophobicity coloring follows Eisenberg's normalized consensus hydrophobicity scale.<sup>159</sup> Structural figure panels were generated using PyMOL, except for those of Figures 3L and S3A, E which were created with UCSF ChimeraX. AlphaFold predicted alignment error (PAE) plots were generated with AlphaPickle.<sup>92</sup>

### Size exclusion chromatography-multi angle light scattering (SEC-MALS)

Uncleaved or cleaved xZP2-N2N3 or hZP2-N1N2N3 (100–150 μg) were measured using an Ettan LC high-performance liquid chromatography system with a UV-900 detector (Amersham Pharmacia Biotech; λ = 280 nm), coupled with a miniDawn Treos MALS detector (Wyatt Technology; λ = 658 nm) and an Optilab T-rEX dRI detector (Wyatt Technology; λ = 660 nm). Separation was performed at 20°C using a Superdex 200 Increase 10/300 GL column (GE Healthcare), with a flow rate of 0.5 mL min<sup>-1</sup> and a mobile phase consisting of 20 mM Na-HEPES pH 7.8, 150 mM NaCl. Data processing and weight-averaged molecular mass calculation were performed using the ASTRA software (Wyatt Technology). The percentages of tetrameric and dimeric species reported in Figures 3A and 3D were calculated using the formula  $(4 \times MW_{\text{monomer}})X + (2 \times MW_{\text{monomer}})(1 - X) = MW_{\text{peak}}$ .

### Native mass spectrometry (MS)

Protein samples (0.50–0.75 mg mL<sup>-1</sup>) were buffer-exchanged to 100 mM ammonium acetate buffer using ZebaSpin columns (Thermo Scientific) and immediately subjected to native MS analysis. Mass spectra were acquired on a Waters Synapt G1 TWIMS MS modified for analysis of intact protein complexes (MS Vision) and equipped with an offline nanospray source. The capillary voltage was 1.5 kV. The cone voltage was set to 50 V, the trap voltage was 20 V and the source temperature was maintained at 20°C. The source pressure was adjusted to 8 mbar. Data were analyzed using MassLynx (Waters).

### Mouse experiments

WT mice were purchased from CLEA Japan, Inc. or Japan SLC, Inc. The mutant mice used in this study are available through the Riken BioResource Center (RBRC; <https://web.brc.riken.jp/en>) under ID *in-frame del*, 11009 or via the Center for Animal Resources and Development (CARD; <http://card.medic.kumamoto-u.ac.jp/card/english>) under ID *in-frame del*, 2916.

### Design and generation of ZP2 ΔN1 mutant mice

The ΔN1 deletion, corresponding to mZP2 residues S40–I144, was designed based on (a) the crystal structure of the mZP2-N1 domain<sup>23</sup>; (b) the predicted (and subsequently experimentally confirmed, Figure S1B) N terminus of the mZP2-N2 domain, based on secondary structure predictions and homology modeling; and (c) the observation that the two residues immediately before the predicted beginning of mZP2-N2 match the sequence 40-SE-41 that closely follows the signal peptide cleavage site of mZP2 (S34|V35). As

a result, the mZP2-N1 deletion of the mice described in this manuscript is different from that of previously reported animals expressing a mutant ZP2 protein that lacked C51-V149 (which replaced the first  $\beta$  strand of mZP2-N2 with the first strand of mZP2-N1).<sup>27</sup>

Mutant lines were generated by genetically modifying WT zygotes *in vitro*, using CRISPR/Cas9 genome editing technology. Single guide RNAs (sgRNAs) were designed with CRISPRdirect.<sup>102</sup> Zygotes were electroporated using a NEPA21 Super Electroporator (Nepa Gene) with crRNA/tracrRNA/Cas9 ribonucleoprotein complexes and single strand DNA (ssDNA) with homologous sequence targeting sites.<sup>160,161</sup> The eggs were incubated in KSOM media overnight and two-cell stage embryos were transferred into the ampulla of pseudo-pregnant ICR females. 19 days after egg implantation, offspring (F0) were obtained by natural birth or caesarean section. Pups carrying a mutant genome were screened by PCR and Sanger sequencing. PCR was performed using KOD Fx Neo (KFX-201). The sequence of the ssDNA used for ZP2  $\Delta$ N1 is: 5'-TCCACATCACTTACAGATATTAGATCTCTCCTGCAGACAACA ATTTCTGAC TGAGGAAGGCTTACTGAGTTCCTGAAAGTCACAAGGGTGAATAAGAGGG-3'. The sequences of the primers used for genotyping are: ZP2  $\Delta$ N1: Fw 5'-CGAGGTGGCAGA GGAAAGCATCTGT-3', Rv 5'-ATTCATTCATCTGCTAAGCCCAGCT -3'; ZP2-N1 (WT): Fw 5'-CCACTCTAATGGTCAGATGG-3', Rv 5'- ATTCATTCATCTGCTAAGCCCAGCT-3'.

### Mouse fertility test

Sexually mature ZP2  $\Delta$ N1 female mice (5–16-week-old) and WT B6D2F1 mice were caged with 8-week-old WT male mice for 2 months. During the fertility test, the vaginal plugs and the number of pups at birth were counted every day. The average number of pups/plugs for each mouse line was calculated by dividing the total number of pups by the number of plugs.

### In vitro fertilization

Spermatozoa were obtained from the cauda epididymis of adult B6D2F1 males and cultured in Toyoda, Yokoyama, Hoshi (TYH) medium<sup>162</sup> for 2 h at 37 °C under 5% CO<sub>2</sub>. Oocytes at the second metaphase (MII) stage were collected from the oviducts of females 16 h after injection with pregnant mare's serum gonadotropin (PMSG) (5 units), followed 48 h later by human chorionic gonadotropin (hCG) (5 units). WT sperm incubated for 2 h in TYH medium were added to a drop containing cumulus-intact oocytes at a final density of  $2 \times 10^5$  spermatozoa mL<sup>-1</sup>. 8 h after adding the spermatozoa, two-pronuclear (2-PN) eggs were counted.

### Sperm-ZP binding assays

Spermatozoa and MII stage oocytes were obtained as described above. Cumulus cells were removed from oocytes by treatment with 330  $\mu$ g/ml hyaluronidase (Sigma-Aldrich) for 5 min and pipetting. Spermatozoa were incubated for 2 h at 37 °C, 5% CO<sub>2</sub> and added to a drop of TYH medium containing cumulus-free oocytes at a final density of  $2 \times 10^5$  spermatozoa mL<sup>-1</sup>. 30 min after mixing the spermatozoa, oocytes were fixed with 0.25% glutaraldehyde. Sperm-bound eggs were observed with an IX73 microscope (Olympus).

### Experiments with *Sof1* knockout mouse sperm

To distinguish whether the infertile phenotype of ZP2  $\Delta$ N1 mice was due to impairment of sperm binding or penetration through the ZP, we have taken advantage of recently described *Sof1*<sup>-/-</sup> mice.<sup>89</sup> These KO animals produce sperm that can bind to the ZP and penetrate it, but cannot fuse with the plasma membrane of the oocyte. As a result, sperm accumulates into the perivitelline space and can thus be used to follow and quantify ZP penetration.

For ZP binding experiments, *Sof1* KO sperm were preincubated in TYH medium for 2 h at 37 °C, 5% CO<sub>2</sub>. The capacitated sperm were incubated with anti-IZUMO1 antibody and anti-rat IgG Alexa Fluor 488-conjugated antibody (both at a dilution of 1:100) for another 30 min. The eggs harvested from WT and ZP2  $\Delta$ N1 mutant female mice were denuded by treating with 0.3 mg mL<sup>-1</sup> hyaluronidase at 37 °C for 5 min. The antibody-probed sperm were added to medium drops containing cumulus-free eggs at a sperm density of  $2 \times 10^5$  spermatozoa mL<sup>-1</sup> and incubated for 30 min at 37 °C, 5% CO<sub>2</sub>. Sperm-egg complexes were fixed in 0.25% glutaraldehyde for 15 min on ice, rinsed in fresh FHM medium drops for three times, and examined under a Keyence microscope. The total number of sperm and the number of acrosome-reacted sperm bound to the ZP of each egg were recorded. Z-stack images were captured by an Eclipse Ti microscope equipped with a C2 confocal module (Nikon).

For ZP penetration assays, *Sof1* KO sperm were preincubated in TYH medium for 2 h at 37 °C, 5% CO<sub>2</sub>. Capacitated sperm were then incubated with cumulus-intact eggs harvested from WT and ZP2  $\Delta$ N1 mutant female mice at a sperm concentration of  $2 \times 10^5$  spermatozoa mL<sup>-1</sup> for 6 h at 37 °C, 5% CO<sub>2</sub>. Sperm-egg complexes were fixed in 0.25% glutaraldehyde for 15 min on ice, rinsed in fresh FHM medium drops for three times, and observed under an Olympus inverted microscope. The number of sperm in the perivitelline space of each egg was counted. Images and videos were captured using a Keyence microscope.

### ZP hardening assay

Oocytes at the second metaphase (MII) stage were collected as described above. Oocytes and embryos were treated with 10 mg mL<sup>-1</sup> collagenase (type I; Sigma-Aldrich) to remove the ZP. 10 min after the addition of the collagenase, eggs were repeatedly pipetted and their morphology was observed as described above.

### Protein deglycosylation

The native fZP1/fZP3 complex and recombinant fZP3 proteins were treated with PNGase F (New England Biolabs and Takara Bio, respectively) for 1 h at 37 °C (native fZP1/fZP3 complex) or overnight at 37 °C (recombinant fZP3 proteins), in three different



conditions: native (New England Biolabs: 50 mM sodium phosphate pH 7.5; Takara Bio: 100 mM Tris-HCl pH 8.6), semi-denaturing (New England Biolabs: 50 mM sodium phosphate pH 7.5, 0.5% SDS) and fully denaturing (New England Biolabs: 50 mM sodium phosphate pH 7.5, 0.5% SDS, 40 mM DTT; Takara Bio: 100 mM Tris-HCl pH 8.6, 0.1% SDS, 0.15%  $\beta$ -mercaptoethanol).

### Immunoblot analysis

Protein lysates from MII or fertilized eggs from WT or ZP2  $\Delta$ N1 mice were resolved by SDS-PAGE and transferred to PVDF membranes. After blocking, blots were incubated with primary antibodies overnight at 4°C and then incubated with secondary antibodies conjugated to horseradish peroxidase (Jackson ImmunoResearch; 1:10000) for 1 h at RT. Antibodies used: anti-mZP2-N1 monoclonal IE-3<sup>88</sup> (ATCC CRL-2463; 1:1000); anti-mZP2 cleavage site polyclonal Ab164 (1:1000); anti-mZP2 ZP-C monoclonal 2C7 (1:1000), anti- $\beta$ -actin (ACTB) (Abcam ab6276; 1:1000). Detection was performed using Chemi-Lumi One Super (Nacalai Tesque). The rabbit polyclonal Ab164 antibody was produced by immunization with a synthetic peptide corresponding to the mZP2 post-fertilization cleavage site (Cys-Ahx-SRLADEN) and purified using the same peptide coupled to SulfoLink resin (Thermo Fisher Scientific). Recombinant fZP3 proteins were detected using anti-His-tag monoclonal (Proteintech) and HRP-conjugated goat anti-mouse IgG (H+L) secondary antibody (Proteintech); EzWestLumi One (ATTO) was used as chemiluminescent substrate.

### Calculation of the relative stoichiometry of mouse ZP subunits

After digitizing a previously reported high performance liquid chromatograph SEC elution profile of a preparation of  $\sim$ 30,000 mZPs,<sup>50</sup>  $A_{280\text{ nm}}$  peak areas corresponding to mZP1-3 were integrated and converted to a relative subunit ratio by taking into account extinction coefficients calculated from the sequences of the respective mature polypeptides.

### Modeling of the human ZP2 NTR/ovastacin complex

The structure of the complex between hZP2 NTR (residues I39-I366) and the mature form of hOvst (L86-D431) was predicted with AlphaFold-Multimer,<sup>36</sup> using version 3 of its model parameters (released on 13 December 2022) and standard run settings. AlphaFill<sup>91</sup> was then used to transplant the catalytic  $\text{Zn}^{2+}$  ion from the crystallographic model of zebrafish hatching enzyme ZHE1 (another member of the astacin-like metalloprotease family)<sup>163</sup> to the AlphaFold Protein Structure Database<sup>164</sup> entry for hOvst (<https://www.alphafold.ebi.ac.uk/entry/Q6HA08>) (optimized transplant clash score 0.14 Å). This was then superimposed on the corresponding chain of the top-ranked complex prediction (ranking score  $[0.8 \times \text{predicted interface template modeling score \{ipTM\}} + 0.2 \times \text{predicted template modeling score \{pTM\}}]$  0.76; ipTM 0.75; predicted local distance difference test [pLDDT] 85), with RMSD 0.54 Å over 181 C $\alpha$  atoms. The resulting hZP2 NTR/hOvst+ $\text{Zn}^{2+}$  model was finally energy minimized in YASARA Structure, using the YASARA2 force field. C-terminal amino acids P283-D431 of hOvst were predicted with low confidence and largely adopted the typical ribbon-like appearance of disordered protein regions<sup>165</sup>; since these residues also did not contact hZP2 NTR, they were not shown in Figure 2F.

### Modeling of human ZP filaments

As described above, we initially determined the structure of the fZP1/fZP3 complex by MR using two subcomplex models generated by a ColabFold implementation of AlphaFold. When AlphaFold-Multimer became subsequently available, we found that we could use a single prediction job to generate models of the whole VE fragment that phased the crystallographic data with improved statistics relative to the original two-step procedure. Specifically, a 5 model-ensemble of the fZP1/fZP3 complex produced a single Phaser solution (LLG=274, TFZ=20.2) that could be automatically refined to  $R_{\text{free}}=0.40$ . Based on this result and the experimental evidence that homomeric and heteromeric filaments made by ZP module-containing proteins share the same basic architecture (Figure 5G), we used the AlphaFold-Multimer model parameters to carry out a comprehensive computational analysis of possible subunit-subunit interactions between the four components of the hZP (hZP1-4), in three steps (Figure S7).

In the first step, we ran a matrix of predictions for ternary complexes consisting of (I-1) a complete hZP subunit; (I-2) a fragment corresponding to the second half of the ZP-N/ZP-C linker and ZP-C domain of any hZP subunit (including the one in [I-1]), up to its C-terminal consensus furin-cleavage site; and (I-3) a fragment encompassing the full N-terminal region, ZP-N domain and first half of the ZP-N/ZP-C linker of the same subunit (Figure S7A, step I). For type II subunits (hZP1/2/4), this showed that complexes consistent with the filament architecture shared by the VE fragment and the UMOD polymer were obtained only in combination with type I subunit hZP3 (green boxes); moreover, all these filament predictions had significantly higher ranking scores (average ranking score 0.75) than the non-filamentous arrangements obtained when type II subunits were combined with each other (gray boxes; average ranking score 0.30), as well as much more convincing PAE heatmaps (compare Figure S7B with Figure S7C). At the same time, it is interesting to notice that cross-hairpin interactions like that observed in the fZP1/fZP3 crystal structures were also found in low-scoring, non-filamentous complex predictions (Figure S7C). In the case of the hZP3-only combination (orange box), 3 models out of 5 predicted a filamentous arrangement, whereas the remaining 2 yielded inferences that matched the homodimeric structure of the cZP3 precursor.<sup>9</sup>

In the second step, we generated an additional matrix of predictions for ternary complexes consisting of (II-1) a complete hZP3 subunit; (II-2) a fragment corresponding to the second half of the ZP-N/ZP-C linker and ZP-C domain of any subunit; and (II-3) a fragment encompassing the full N-terminal region, ZP-N domain and first half of the ZP-N/ZP-C linker of any subunit other than the one in (II-2) (Figure S7A, step II). This second analysis extended the previous by indicating that hZP1, hZP2 and hZP4 are in principle able to

form all possible mixed ZP-C/ZP-N interfaces between themselves, whereas the majority of combinations involving the domains of the hZP3 ZP module resulted in a lower fraction of filamentous complex predictions.

In the third step, we predicted larger quaternary complexes consisting of 2 copies of hZP3 and: (III-1) 1 copy of the hZP1 trefoil domain+ZP module, 1 copy of the hZP2 ZP module; (III-2) 1 copy of the hZP1 trefoil domain+ZP module, 1 copy of the hZP4 trefoil domain+ZP module; or (III-3) 1 copy of the hZP2 ZP module, 1 copy of the hZP4 trefoil domain+ZP module. Even when using 5 seeds per deep learning model, all predictions resulting from these combinations showed filamentous arrangements that were either circular (reminiscent of early prediction studies of UMOD<sup>40</sup>) or straight (by disconnecting the ZP-N and ZP-C moieties of one subunit) (Figure S7A, step III). Moreover, in both cases, all the predicted filaments (amounting to a total of 75 models) consisted of a hZP3 protofilament combined with a hZP1/2/4 protofilament. In other words, in complete agreement with the crystal structure of the fZP1/fZP3 complex (Figures 5 and 6), homomeric ZP-N/ZP-C interfaces were never observed. Accordingly, no filament-like arrangement were observed in parallel predictions of (III-4) 4 copies of the hZP1 trefoil domain+ZP module; (III-5) 4 copies of the hZP2 module; (III-6) 4 copies of the hZP3 module; (III-7) 4 copies of the hZP4 trefoil domain+ZP module; and (III-8) 2 copies of the hZP2 module, 1 copy of the hZP1 trefoil domain+ZP module and 1 copy of the hZP4 trefoil domain+ZP module (for a total of 125 models).

The information obtained from steps I-III was used for assembling filament models like the ones shown in Figure 7C. To do so, fragments corresponding to ternary complexes generated as described above were superimposed over common protein domains, after which all half-protein copies with truncated ZP-N/ZP-C IDLs were deleted (except the ones at either end of the filament). Despite its simplicity, this procedure did not introduce significant clashes between hZP subunits; as a result, relieving minor unfavorable contacts by standard energy minimization in YASARA Structure was sufficient to achieve very good geometry (for the individual filaments in Figure 7C: MolProbity score 1.06, clashscore 0.83, Ramachandran favored 96.3%, Rama distribution Z-score  $0.01 \pm 0.10$ ). Moreover, inspection of the polymeric models in PyMOL and analysis with GlyProt<sup>105</sup> showed that all their possible N-glycosylation sites (corresponding to a total of 116 sites in the case of the filaments shown in Figure 7C) are, as expected, exposed to the solvent.

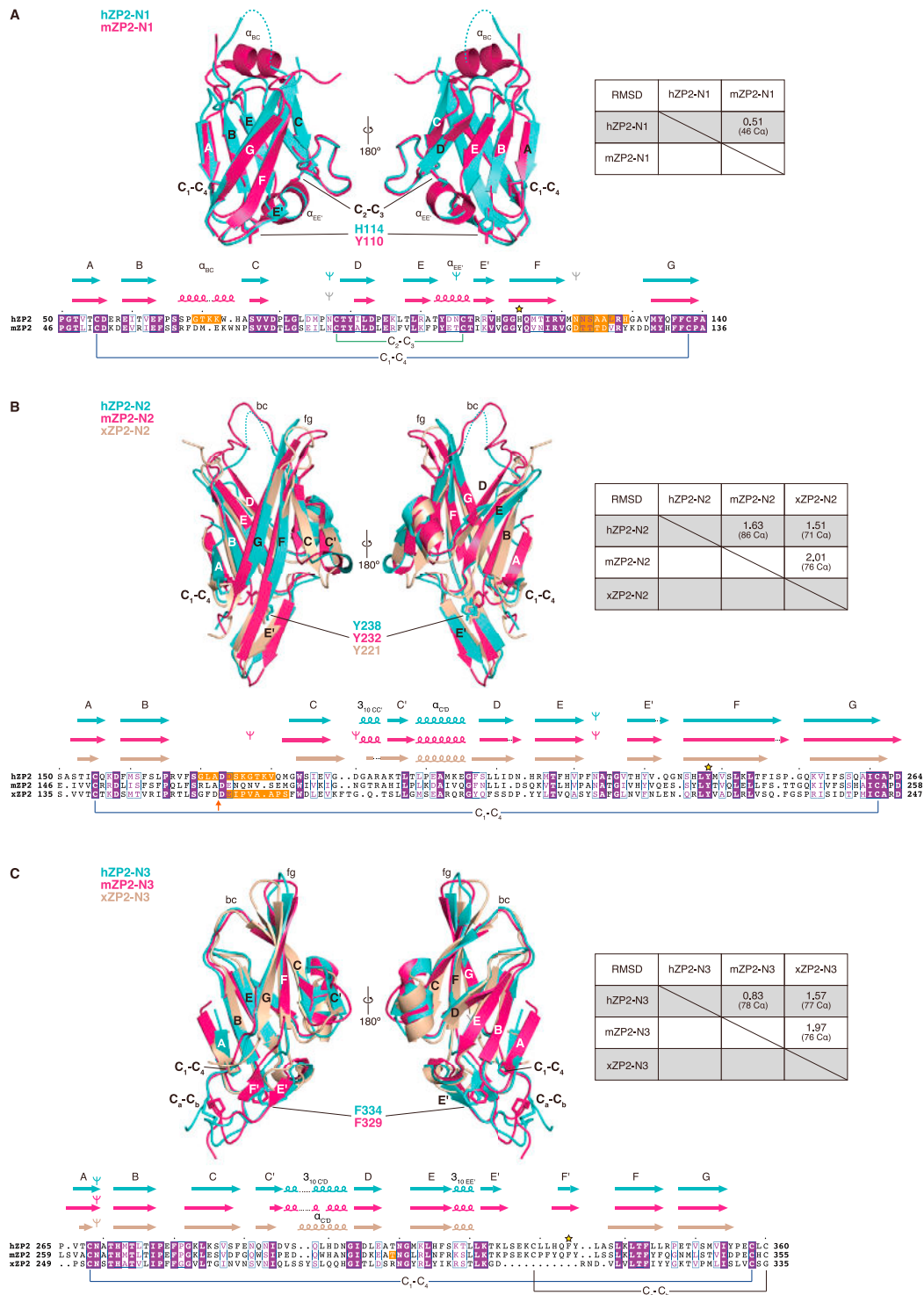
Like in the mouse and the chicken, human egg coat filaments are covalently cross-linked into a 3D matrix by intermolecular disulfides involving the ZP-N1 domains of hZP1 subunits from different filaments.<sup>17,18</sup> Each of these N-terminal hZP1-N1 domains is spaced from the corresponding trefoil domain and C-terminal ZP module by a  $\sim 90$ -residue linker, which is recognized as unstructured by both dedicated predictors<sup>166</sup> and an analysis of AlphaFold model pLDDT values.<sup>39,165</sup> Based on this information, we modeled hZP1 cross-links between hZP filaments by (1) generating a random coil model of the hZP1 disordered spacer region (which, when fully stretched, could in principle extend over  $\sim 3.6 \text{ \AA} \times 90 = 324 \text{ \AA}$ ); (2) connecting the N-termini of two copies of this spacer to the C-termini ( $\sim 52 \text{ \AA}$  apart) of a model of a disulfide-bonded dimer of hZP1-N1, generated using a combination of MODELLER<sup>110</sup> and AlphaFold; (3) connecting the C-termini of the resulting  $2 \times (\text{hZP1-N1} + \text{spacer})$  assembly to the N-termini of the trefoil domains of hZP1 subunits belonging to two hZP filaments (with the N-termini of the trefoil and ZP-N domains of each hZP1 subunit being  $\sim 12 \text{ \AA}$  apart). Clearly, given the intrinsic flexibility of the hZP1 spacer, the semi-extended cross-link of the model depicted in Figure 7C is only one of the many possible conformations that this region of the protein adopts while cross-linking hZP filaments with variable relative orientations. For this reason, the model is strictly meant to provide a chemically realistic example that can be used to draw general conclusions about hZP architecture, rather than recapitulate the variety of filament cross-link conformations that are likely to exist in nature.

To model the non-covalent filament cross-links resulting from post-fertilization processing of hZP2 (Figure 7D), we first used AlphaFold-Multimer to model cleaved hZP2-N2N3. The top ranked prediction had a clearly superior PAE heatmap and showed the same relative subunit arrangement observed in our experimental structures of cleaved xZP2-N2N3 (Figures 3K and 3L) and un-cleaved mZP2-N2N3 (Figure 4I), with the  $\beta$ G strand of the hZP2-N2 domain mediating homodimer formation. The two moieties of this model were then superimposed on the ZP-N2 and -N3 domains from two hZP2 subunits belonging to different filaments. To complete the assembly, two copies of the hZP2-N1 domain from the hZP2-N1N2N3 crystal structure were finally linked to the hZP2-N2 N-termini in random relative orientation. Although AlphaFold can accurately predict the structure of the NTR of hZP2, analysis of hZP filament fragment models that include this subunit suggests that the NTR does not appreciably interact with the rest of the protein and can adopt multiple orientations relative to the core of the filament, due to flexibility of the linker between ZP-N3 and the ZP module (Figure S7D). One possible set of orientations was chosen to generate the model depicted in Figure 7D but, as in the case of the hZP1 cross-link, the presence of such a hinge within hZP2 is likely to be important in order to establish non-covalent cross-links between variably oriented filaments. However, even in the extreme case where the longest axis of a cleaved hZP2 NTR homodimer is oriented perpendicular to the filaments, the latter could not be separated by more than the longest dimension of the NTR dimer itself ( $\sim 110 \text{ \AA}$ ).

## QUANTIFICATION AND STATISTICAL ANALYSIS

Biochemical and mouse experiments were repeated at least two and three times, respectively; in all cases, results were reproducible. The statistical analyses whose outcomes are reported in Figure 4 were performed using Student's t-test (two-tailed). Data represent the means  $\pm$  standard deviation (SD) and error bars indicate SD.

# Supplemental figures



(legend on next page)

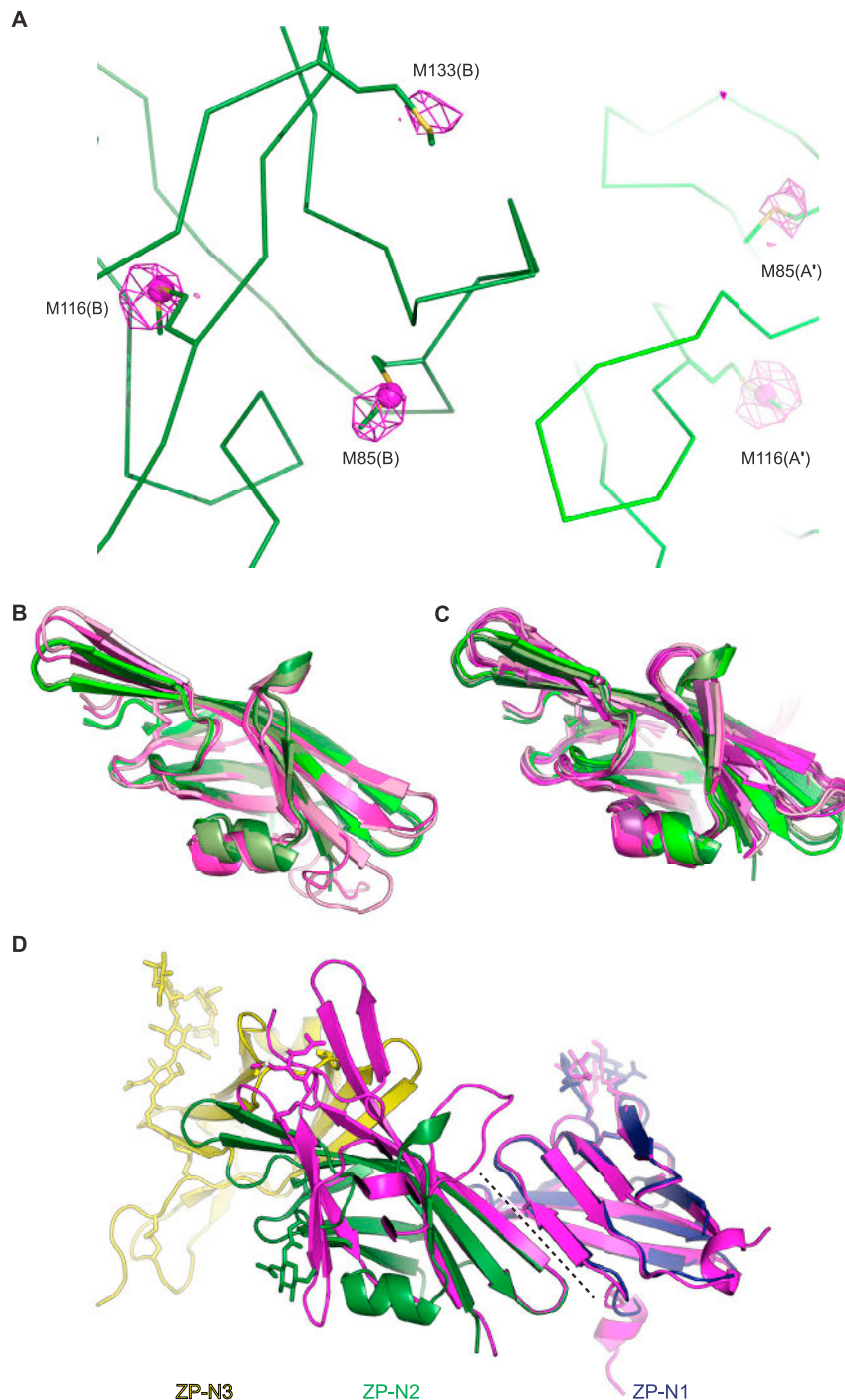
---

**Figure S1. Structural comparisons of the ZP-N domains of ZP2 NTR, related to Figure 1**

(A) The ZP-N1 domains of hZP2 and mZP2 (PDB: 5II6) adopt a typical ZP-N fold, characterized by 8  $\beta$  strands and two intramolecular disulfides with 1-4, 2-3 connectivity. Inverted tripods, N-glycan sites (with the sites observed in the experimental maps shown in color); yellow stars, positions corresponding to the signature Tyr of the ZP-N domain. Amino acids highlighted in orange are disordered in the crystals structures and were thus aligned based on sequence only.

(B) The ZP-N2 domains of hZP2, mZP2, and xZP2 lack the C<sub>2</sub>-C<sub>3</sub> disulfide. The vertical orange arrow below the sequence alignment indicates the position of the post-fertilization cleavage site.

(C) ZP-N3 domains of hZP2, mZP2, and xZP2. Like ZP-N2, ZP-N3 also lacks the second classic ZP-N disulfide. Note that the mammalian ZP2 ZP-N3 domain has a lower secondary structure content at the level of its E'FG extension but is stabilized by an additional C<sub>a</sub>-C<sub>b</sub> disulfide bond.



**Figure S2. Structure determination and analysis of hZP2-N1N2, related to Figures 1, 3, and 4**

(A) Validation of the MR solutions for hZP2-N1N2. A section of the hZP2-N1N2 model is shown in C $\alpha$  trace representation, with the side chains of Met residues shown as sticks. Chain B and a symmetry-related copy of chain A (A' = Y, X, -Z + (-1 -1 0) & {-1 0 0}) are dark and light green, respectively, while other symmetry-related molecules have been omitted for simplicity. The violet mesh is an anomalous difference map contoured at 3.5  $\sigma$ , calculated from SeMet-hZP2-N1N2 data collected at  $\lambda = 0.92$  Å. Note how the map, calculated with phases from the experiment and the refined model, also shows clear peaks for Se atoms that were not included among the sites initially identified by PHENIX AutoSol (depicted as magenta spheres).

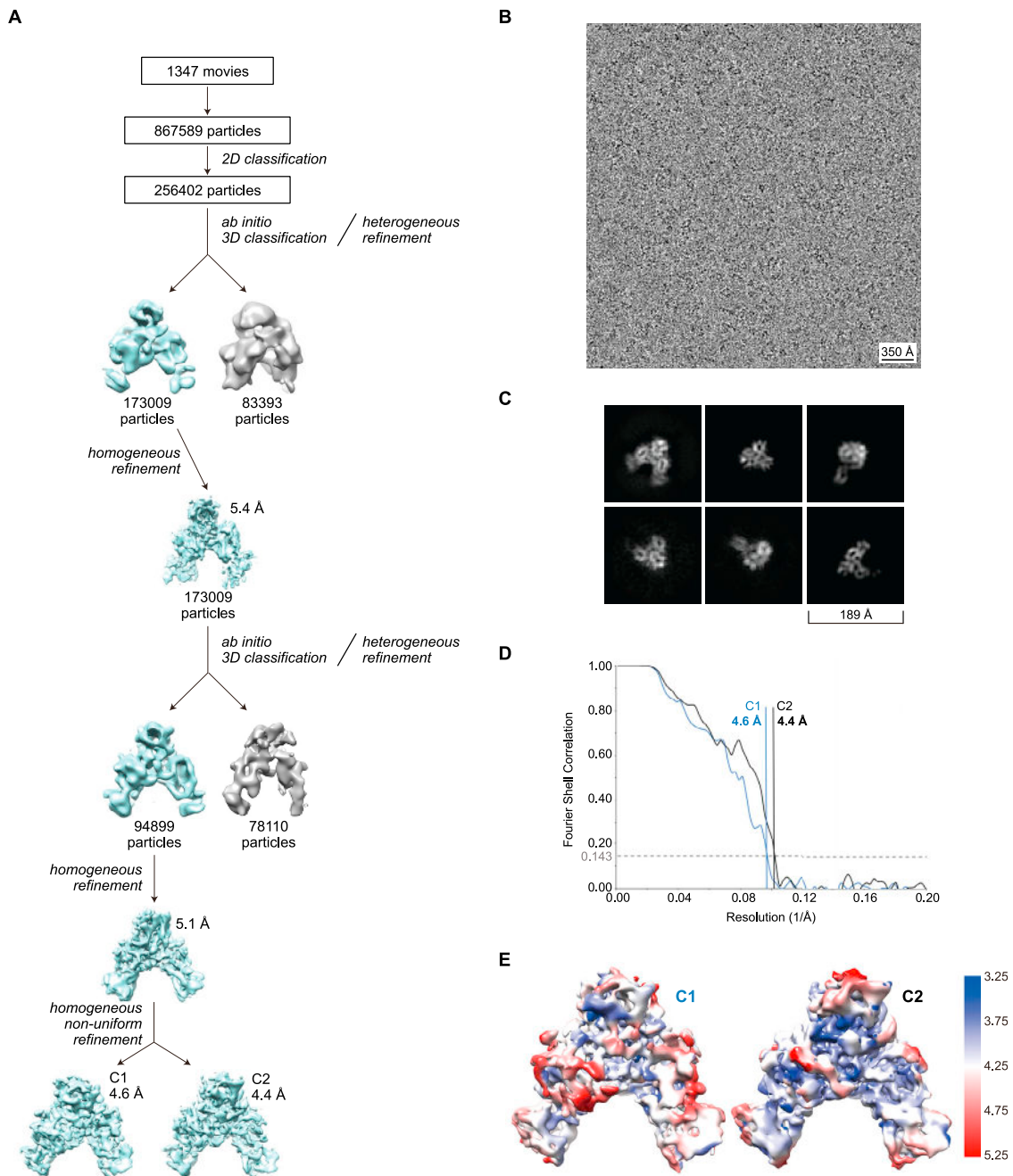
(B) Superposition of the ZP-N2 domains of hZP2-N1N2N3 and mZP2-N2N3 (average RMSD 2.1 Å). The three chains in the asymmetric unit of the hZP2-N1N2N3 are shown in cartoon representation and colored in shades of green, whereas the two chains in the asymmetric unit of the mZP2-N2N3 are colored in shades of magenta. Disulfide bonds are shown as sticks.

(legend continued on next page)

---

(C) Superposition of the ZP-N2 domains of hZP2-N1N2N3 and cleaved xZP2-N2N3 (average RMSD 2.2 Å). Molecules are represented as in (B), with the hZP2-N1N2N3 chains colored in shades of green and the eight chains in the asymmetric unit of the cleaved xZP2-N2N3 crystal colored in shades of magenta.

(D) Superposition of the structures of hZP2-N1N2N3 (chain A, colored by domain as in [Figures 1B and 1C](#)) and hZP2-N1N2 (chain A, colored magenta) over ZP-N1 and part of β strands F and G of ZP-N2 (RMSD 0.7 Å over 94 Cα). N-glycans and disulfides are depicted in stick representation, and the conserved interface between ZP-N1 and ZP-N2 domains is indicated by a dashed black line.



**Figure S3. Workflow and data assessment for the cryo-EM structure of cleaved xZP2-N2N3, related to Figure 3**

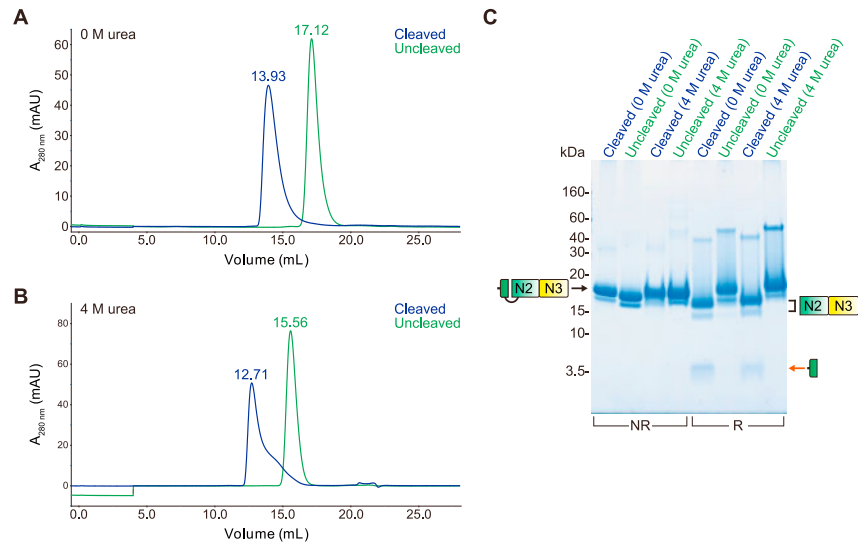
(A) Cryo-EM particle processing workflow. Auto-picked particles were filtered through one round of 2D classification followed by two rounds of *ab initio* 3D classification. The use of homogeneous non-uniform refinement as well as the application of C2 symmetry to the final reconstruction resulted in a modest improvement in resolution.

(B) Representative micrograph from a single data collection of 1,347 micrographs used for automated particle picking for 2D classification.

(C) 2D class averages of cleaved xZP2-N2N3.

(D) Fourier shell correlation (FSC) of the final volumes.

(E) Local resolution estimation of the C1 and C2 density maps.

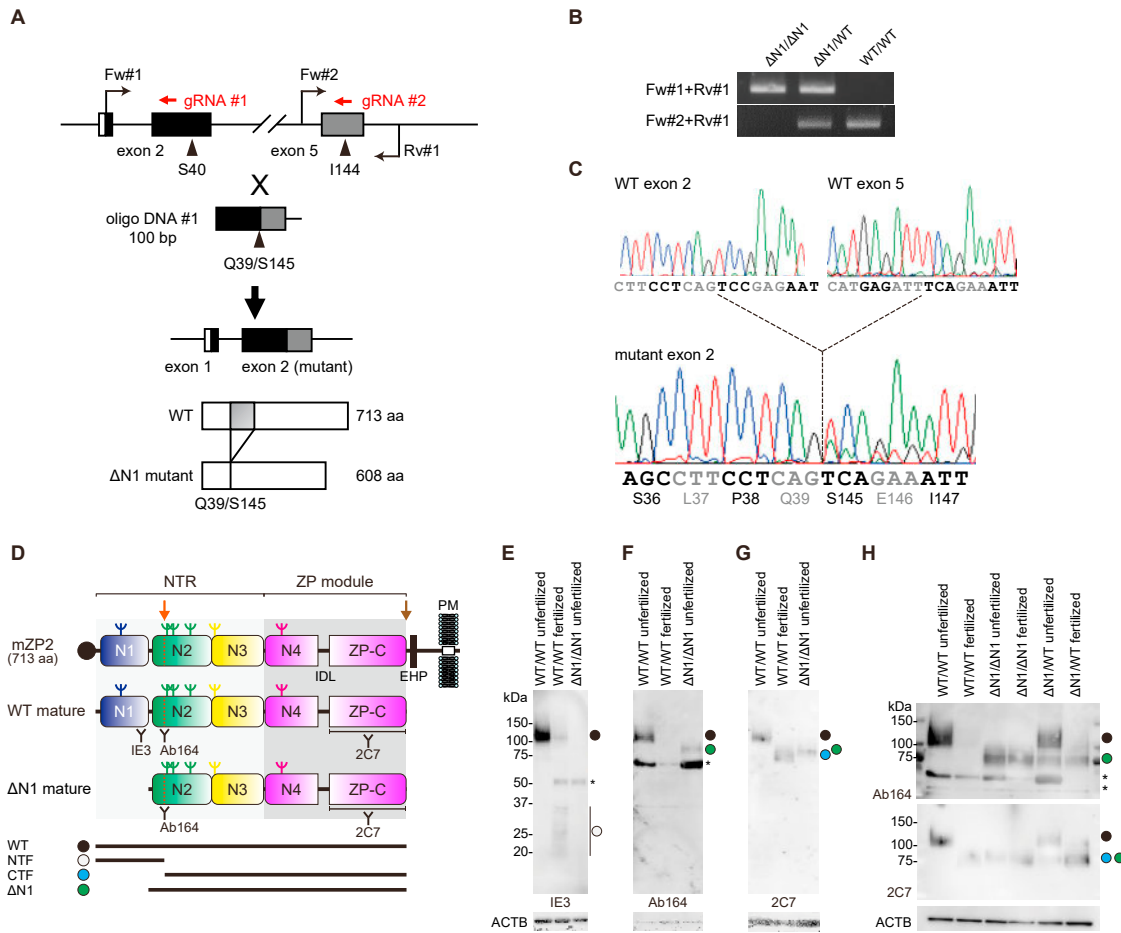


**Figure S4. Stability of the xZP2 dimer of dimers, related to Figure 3**

(A and B) SEC profiles of uncleaved and cleaved xZP2-N2N3 in the absence (A) and presence (B) of 4 M urea.

(C) SDS-PAGE analysis of peaks from the SEC runs shown in (A) and (B). The orange arrow indicates the N-terminal fragment of xZP2.

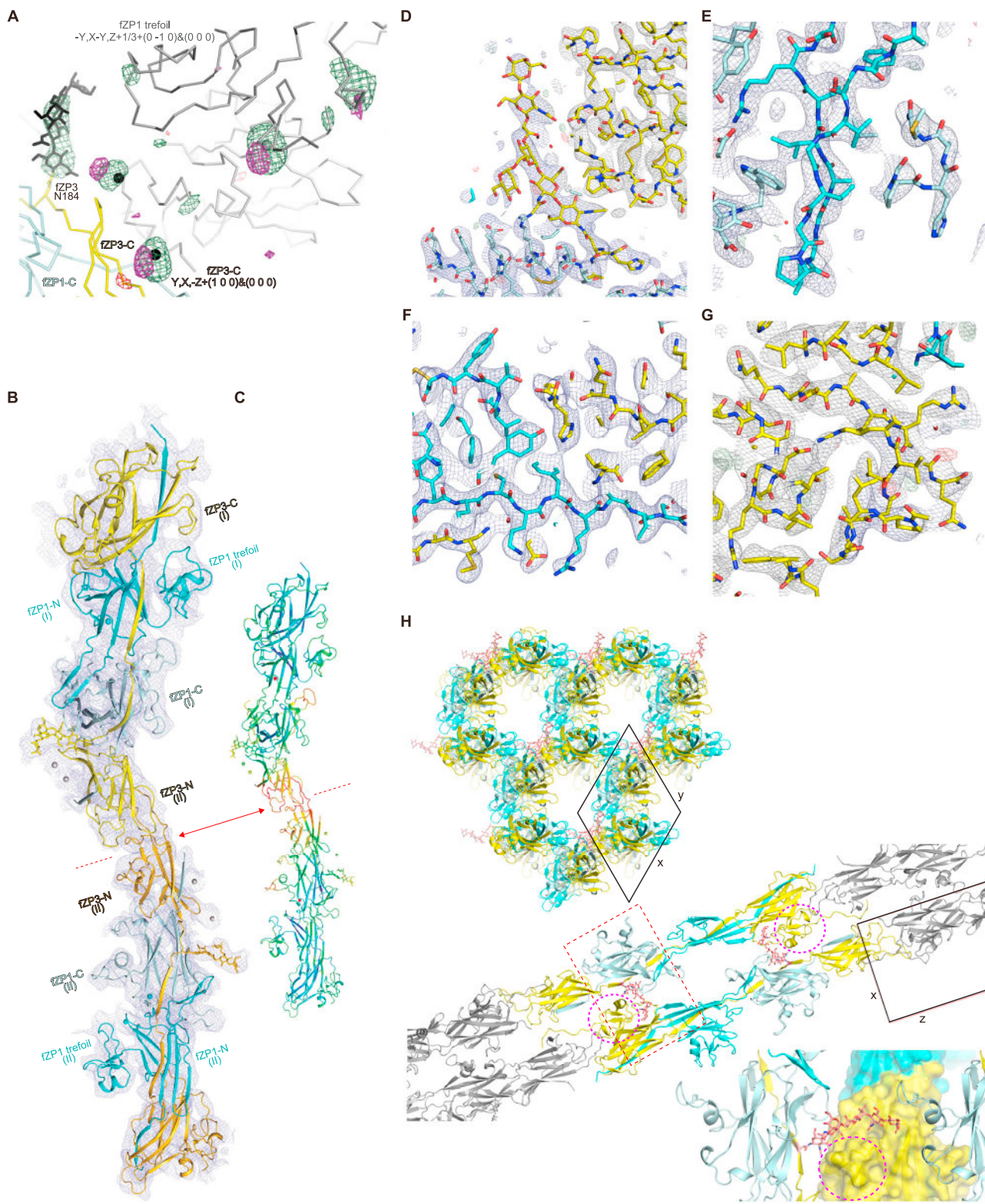




**Figure S5. Generation of ZP2 ΔN1 oocytes and analysis of their ZP2 cleavage status, related to Figure 4**

(A–C) Generation of ZP2 ΔN1 mice: strategy for generating the ZP2 ΔN1 mutant mice (A), PCR amplification of the mutated genomic region (B), and DNA sequencing of ZP2 ΔN1 mice (C).

(D–H) Anti-mZP2 immunoblotting of WT and ΔN1 oocytes. (D) Schematic diagram of the mZP2 precursor and the mature forms of the WT protein and its ΔN1 mutant. Elements are depicted and labeled as in Figures 1A and 1B. Black circle, signal peptide; NTF and CTF, N- and C-terminal fragments generated by the post-fertilization cleavage of mZP2. Monoclonal antibodies IE-3 and 2C7 recognize the ZP-N1 and ZP-C domains of mZP2, respectively. Rabbit polyclonal antibody Ab164 recognizes the uncleaved bc loop of mZP2-N2 (S164-N170). (E–G) Reducing immunoblot analysis of oocytes from WT and ΔN1 homozygous mice shows that the mZP2-N2 cleavage loop of mutant animals is intact before fertilization. mZP2 is detected by IE-3 (E), Ab164 (F), and 2C7 (G). ACTB, β-actin internal control. Asterisks indicate non-specific bands. (H) Reducing immunoblot analysis of WT and ΔN1 homozygous or heterozygous oocytes with Ab164 and 2C7. Note how the mZP2-N2 cleavage loop remains intact even after ΔN1/ΔN1 animals are fertilized by *in vitro* fertilization (by making holes in the ZP) and how, in heterozygous mice that express both WT and ΔN1 mutant protein, mOvst only cleaves the former. These observations agree with the fact that hZP2-N1 is required for efficient cleavage of hZP2 by hOvst *in vitro* (Figure 2E), and are also consistent with the idea that the loop/loop interaction observed in the structure of homodimeric mZP2-N2N3 (Figures 4I and 4J) makes the mZP2 cleavage site less accessible to the protease.



**Figure S6. Structure determination and analysis of the native fish VE fragment, related to Figures 5 and 6**

(A and B) Validation of the single Phaser MR solution for crystal form I of the fZP1/fZP3 complex. (A) Inspection of the sigma-A-weighted  $F_{\text{obs}} - F_{\text{calc}}$  map obtained upon initial refinement of the solution (green mesh,  $3\sigma$ ; red mesh,  $-3\sigma$ ) revealed strong positive difference density for the glycan chain attached to fZP3 N184,

(legend continued on next page)

which is shown here (black sticks) but was not included in the search model used for MR and initial refinement. Four additional large positive difference density peaks were also observed, which have been tentatively assigned to  $\text{Yb}^{3+}$  ions from the mother liquor (spheres) based on the fact that they match peaks in a phased anomalous difference map calculated from the same data (magenta mesh,  $3.8 \sigma$ ). Molecules are shown as  $C\alpha$  traces, with the copy of the fZP1/fZP3 complex contained in the reference asymmetric unit of the crystal (XYZ) colored as in Figure 5B and two symmetry-related molecules colored light and dark gray.

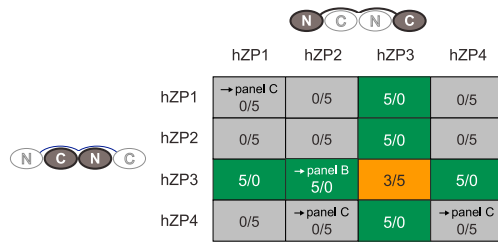
(B) 4.2 Å resolution sigma-A-weighted  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map after DEN refinement, contoured at  $1 \sigma$  and carved at 4 Å around the complex model in the reference asymmetric unit (complex copy [I], above the red dotted line) as well as a copy related by symmetry operation  $-X, -X+Y, -Z+1/3+(2\ 1\ 0)$  &  $(0\ 0\ 0)$  (complex copy [II], below the red dotted line). Note how the relative arrangement of the two copies of the complex mimics a filament, although their relative orientation is inverted compared with the one that these fragments must necessarily adopt in the native VE filament (due to the covalent continuity of the fZP1 IDL that connects them before hatching enzyme cleavage).

(C) Cartoon model of the same complexes depicted in (B), shown at a smaller scale and colored by B factor from blue (low) to red (high). Consistent with the above considerations and reflecting the poor quality of this part of the map, high model B factors at the level of the fZP3 ZP-N/ZP-N interaction that connects the two complexes (red double-headed arrow) clearly identify the latter as a non-native crystal contact.

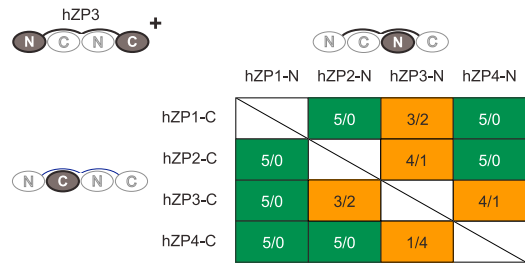
(D–G) Representative sections of the refined sigma-A-weighted  $2F_{\text{obs}} - F_{\text{calc}}$  and  $F_{\text{obs}} - F_{\text{calc}}$  electron density maps at 2.7 Å resolution of crystal form II of the fZP1/fZP3 complex. The  $2F_{\text{obs}} - F_{\text{calc}}$  map with unfilled missing reflections (phenix.refine output MTZ columns 2FOFCWT\_no\_fill PHFOFCWT\_no\_fill), contoured at  $1.25 \sigma$  and carved at 4 Å around the molecular model (in stick representation and colored as in Figure 5B), is colored blue and dark gray for the complex in the reference asymmetric unit and symmetry-related ones, respectively; the  $F_{\text{obs}} - F_{\text{calc}}$  map is also carved at 4 Å around the model and colored green ( $3.5 \sigma$ ) and red ( $-3.5 \sigma$ ). (D) The fZP3 IDL N-glycan (center) packs against the ZP-C subdomain of a symmetry-related fZP3 molecule from a different filament (top right). (E) View of the ZP-C/ZP-N interface of two fZP1 subunits, centered on the ZP-N  $\beta\text{F}'$  and  $\beta\text{F}''$  strands packing against ZP-C W475 and P521. (F) Detail of the fZP1 ZP-N/fZP3 ZP-C interface, involving ZP-N Tyr 339 ( $\beta\text{F}$ ) and 371 ( $\beta\text{G}$ ). (G) The fZP3 ZP-C/fZP3 ZP-N interface, centered on the ZP-C D279/ZP-N R167 salt bridge.

(H) Cross-section (XY plane; top) and longitudinal (XZ plane; middle) views of the packing of crystal form II, highlighting how the fZP3 glycans of one pseudo-infinite filament (salmon sticks) interact with the fZP3 ZP-C subdomains of another (dashed magenta circles). fZP1/fZP3 complex subunits are colored as in Figure 5B, except in the middle panel where this is only done for two neighboring asymmetric units. The black box represents the trigonal unit cell, and the dashed red rectangle indicates the fZP3 N-glycan/ZP-C subdomain contact region magnified in the bottom panel. The latter depicts one of the filaments in mixed surface/cartoon representation and shows how the tip of the fZP3 N-glycan also packs against the fZP1 ZP-C loop that contains the  $\text{C}_x\text{-C}_y$  disulfide.

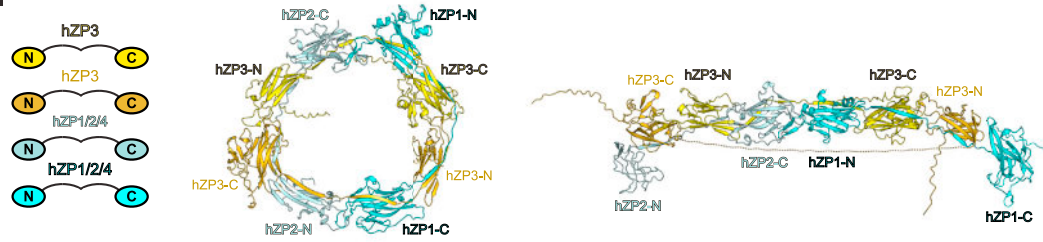
**A** Step I



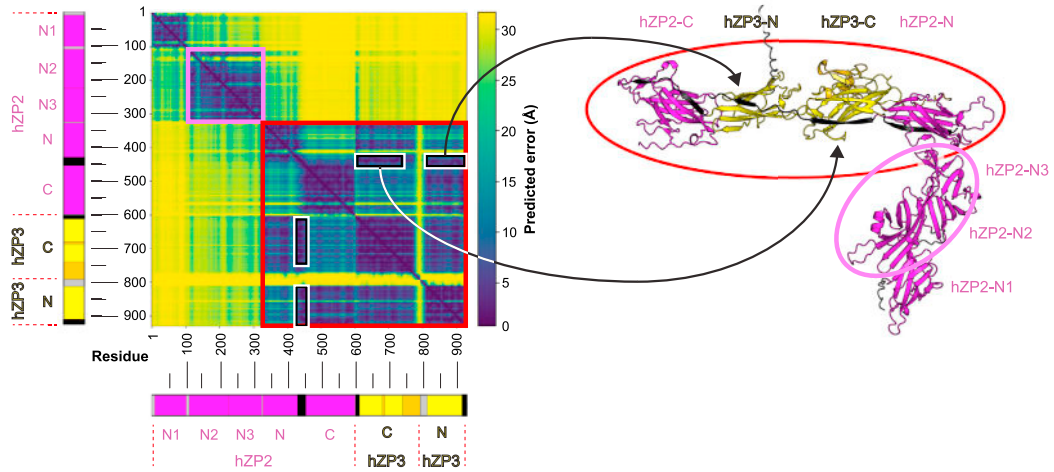
Step II



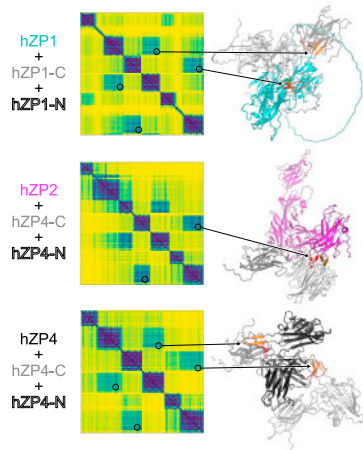
Step III



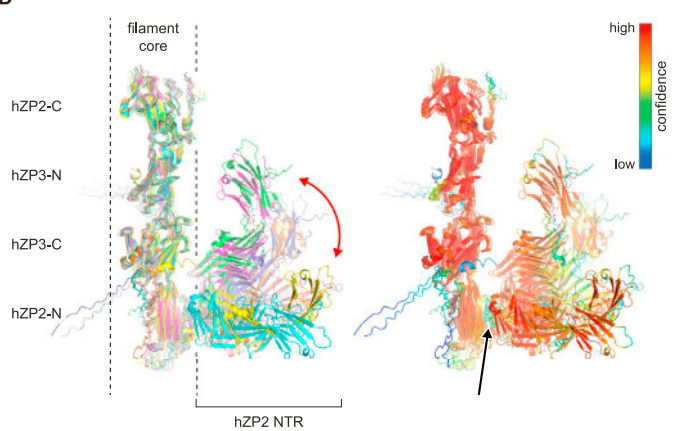
**B**



**C**



**D**



**Figure S7. AlphaFold interaction matrix of human ZP subunits, related to Figure 7**

(A–C) Schematic summary of the three sets of AlphaFold prediction runs used to probe possible subunit interactions within hZP filaments. (A) Based on the experimental architecture of the fVE fragment (Figure 5), step I tested all possible combinations between any full-length hZP component and the two moieties that are generated by dividing every subunit (including itself) in the center of its ZP-N/ZP-C linker (dark ovals). Each prediction was performed using 5 distinct deep learning models; combinations that resulted in a filament conformation in all cases, in some cases only, and in no cases are indicated by green, orange, and gray boxes, respectively. Step II tested all possible combinations between a full-length hZP3 subunit and mixed moieties from all subunits, whereas step III predicted quaternary complexes consisting of two type I (hZP3) and two type II (hZP1/2/4) components. Notably, the latter adopted a circular filament arrangement (middle panel) or a straight filament conformation (right panel) resulting from the non-physical stretching of the ZP-N/ZP-C linker of one subunit (dotted line). For more details, please refer to the corresponding section of the [method details](#). (B) Step I AlphaFold prediction for a type I/type II hZP subunit combination, the ternary complex of a full-length hZP3 subunit, and two moieties from hZP2. The left half of the panel shows the PAE heatmap of the top-ranked prediction of the complex, whose model is shown on the right, in relation to the domain structure of its components (depicted as bars with bands color-matched to the model). By reporting the predicted error (in Å) between all pairs of residues, AlphaFold's PAE heatmaps can be used to identify domain interactions as areas of low score (colored dark blue in this panel's maps).<sup>39</sup> The white boxes and corresponding arrows indicate the clearly recognizable interactions between the N- and C-terminal halves of the IDL of hZP3 and the ZP-C and ZP-N domains of hZP2, respectively; the light magenta square and oval show how the hZP2 ZP-N2 and -N3 domains extensively interact with each other (as observed in the crystal structures of hZP2 [Figure 1C], xZP2 [Figures 3K and 3L], and mZP2 [Figure 4I]) but not with the rest of the complex; the red box and oval encompass the large number of interactions at the level of the filament core. (C) Top ranking step I predictions for different combinations of type II hZP components (hZP1/2/4). Despite having low ranking scores and lacking a filament conformation, these models include cross-hairpin ZP-N/ZP-C interactions between subunits (indicated by arrows and orange/red colors for the ZP-N and ZP-C hairpins, respectively). (D) Superposition of multiple models of the complex between full-length hZP2 and two hZP3 moieties, colored by model (left) or by pLDDT (right). As also shown in (B), the NTR of hZP2 is predicted to make little interaction with the protein's ZP module or the rest of the complex; instead, the NTR adopts a range of different orientations relative to the core of the filament (red double-headed arrow) due to flexibility (and thus relatively low prediction confidence) of the short linker between hZP2-N3 and the ZP module (black arrow).