Species-restricted interaction between gametes at the beginning of fertilization is mediated by the extracellular coat of the egg, a matrix of cross-linked glycoprotein filaments called the zona pellucida (ZP) in mammals and the vitelline envelope in nonmammals. All egg coat subunits contain a conserved protein-protein interaction module—the “ZP domain”—that allows them to polymerize upon dissociation of a C-terminal propeptide containing an external hydrophobic patch (EHP). Recently, the first crystal structures of a ZP domain protein, sperm receptor ZP subunit zona pellucida glycoprotein 3 (ZP3), have been reported, giving a glimpse of the structural organization of the ZP at the atomic level and the molecular basis of gamete recognition in vertebrates. The ZP module is divided in two related immunoglobulin-like domains, ZP-N and ZP-C, that contain characteristic disulfide bond patterns and, in the case of ZP-C, also incorporate the EHP. This segment lies at the interface between the two domains, which are connected by a long loop carrying a conserved O-glycan important for binding to sperm in vitro. The structures explain several apparently contradictory observations by reconciling the variable disulfide bond patterns found in different homologues of ZP3 as well as the multiple ZP3 determinants alternatively involved in gamete interaction. These findings have implications for our understanding of ZP subunit biogenesis, egg coat assembly, architecture, and interaction with sperm; structural rearrangements leading to postfertilization hardening of the ZP and the block to sperm binding; and the evolutionary origin of egg coats.

INTRODUCTION

Approximately 30 years ago, it was discovered that the mouse zona pellucida (ZP) is composed of three glycoprotein subunits—ZP1 (~100 kDa), ZP2 (~120 kDa), and ZP3 (~83 kDa)—that are coordinately synthesized only by growing oocytes over a period of 2–3 wk (Fig. 1A) [1–3]. Based on analysis of solubilized filaments using electron microscopy (EM), it was suggested that ZP2 and ZP3 assemble into micron-long polymers, which are cross-linked into a three-dimensional (3D) matrix by disulfide-linked homodimers of ZP1 (Fig. 1B) [4, 5]. This model is consistent with the molar ratio of the assembled components [2] and has since been confirmed by the phenotype of mice lacking the genes for individual ZP subunits [6–9]. The ZP is thought to have a very similar structure in other mammals, where it often contains an additional ZP1-like subunit, ZP4, the gene for which is not expressed in the mouse [10]. Moreover, proteins homologous to ZPs 1–4 make up the vitelline envelope (VE) of nonmammalian vertebrates [11–14], and related molecules constitute the egg coat of marine invertebrates, such as mollusks and ascidians [15, 16]. Thus, the basic architecture of animal egg coats has been conserved over hundreds of millions of years of evolution [17].

The precursor forms of all ZP/VE proteins contain an N-terminal secretory signal peptide and share a “ZP domain,” a C-terminal block of sequence homology that spans approximately 260 amino acids and is distinguished by a few conserved hydrophobic positions (e.g., as many as eight times in type I ZP domain, found in ZP3) or 10 (type II ZP domain, found in ZP1/2/4) essentially invariant Cys residues (Fig. 1A) [18, 19]. With the exception of the Cys pair linking the two subunits of the mouse ZP1, all these residues are engaged in intramolecular disulfide bonds [20–24]. The conserved ZP domain, which was suggested to consist of two moieties (ZP-N/ZP-C) [25] and is preceded by a single trefoil/P domain in ZP1 and ZP4 [26], is then followed by a consensus furin cleavage site (CFCS) [27–29]. This separates the mature part of the proteins from a C-terminal propeptide (CTP) that in most cases contains a single-spanning transmembrane domain (TM) and a short cytoplasmic tail (Fig. 1A). A short external hydrophobic patch (EHP), located between the CFCS and the TM, plays a crucial role in the secretion of egg subunit precursors [25, 30]. Dissociation of the EHP upon cleavage of the precursors at the CFCS in the trans-Golgi or at the plasma membrane is thought to activate the mature subunits [25], which polymerize into extracellular filaments using their ZP domain [31]. Notably, many different eukaryotic proteins unrelated to fertilization also contain a ZP domain, and parallel studies on some of these systems have provided additional support for the polymerization function of this protein region as well as its control by the EHP-containing CTP [18, 19, 31, 32]. Nevertheless, many questions remain...
open concerning the mechanism of ZP subunit polymerization as well as the biological function of individual ZP components. Unlike ZP1-null mice, which produce eggs with a loose ZP and are subfertile [7], the ovulated oocytes of ZP3- and ZP2-knockout mice totally lack a ZP, resulting in complete infertility [6, 8, 9]. Consistent with this observation, both ZP3 and ZP2 bind sperm in vitro, and they have long been considered to act as receptors for non-acrosome-reacted and for acrosome-reacted sperm, respectively [33–36]. Moreover, ZP3 can stimulate the acrosome reaction on sperm in vitro [37, 38], although recent data suggest that cumulus cells might be mainly responsible for inducing this process in vivo [39]. Nevertheless, significant disagreement remains over the precise functions of ZP3 and ZP2 [40, 41], and whether sperm binding to ZP3 is mainly dependent on carbohydrates, the polypeptide, or requires both is unclear [1, 41, 42]. Structural biology could potentially contribute to the solution of this problem, but its application to egg coat subunits—as well as to ZP domain proteins in general—has been hindered by the difficulty of overproducing these molecules in properly folded form at the milligram scale, their very heavy and heterogeneous glycosylation, and their tendency to aggregate at high concentrations. However, x-ray crystallographic structures of both the ZP-N region of mouse ZP3 (Fig. 1C) and full-length chicken ZP3 (Fig. 2) bound to the EHP propeptide have recently been determined [43, 44]. In this review, we discuss the implications that these first atomic-resolution structures of a ZP domain protein have for the architecture, function, and evolution of the ZP.

3D STRUCTURE OF THE ZP DOMAIN

The crystal structures of ZP3 demonstrated that the ZP domain first recognized by pattern-based sequence analysis [19] is in fact a module consisting of two distinct domains, an N-terminal ZP-N domain and a C-terminal ZP-C domain, the latter of which also includes the EHP (Fig. 2) [43, 44]. Although this was not apparent from their sequence, the structure of full-length chicken ZP3 [43] revealed that both of these domains adopt an immunoglobulin (Ig)-like fold, which consists of a sandwich of two β-sheets, the strands A-B-E and C-F-G of which enclose a hydrophobic core [45]. This accommodates most of the conserved hydrophobic residues originally identified in the ZP domain sequence, the eight invariant Cys residues of which are equally distributed between the two domains [19]. Despite having significantly different primary structure and disulfide bonds, ZP-N and ZP-C share a common basic topology; moreover, structural features both inside and outside their core make them distinct from previously studied Ig-like domains [43].

The ZP-N domain defines a new subtype of the Ig superfamily of proteins by containing a characteristic E′-strand within an E′-F-G β-sheet that extends from the hydrophobic core as well as two almost completely buried disulphide bonds that clamp both sides of the β-sandwich (Figs. 1C and 2A) [43, 44]. These linkages are formed by the first four conserved Cys (C) residues of the ZP module, which adopt a C₁-C₄; C₃-C₆ pattern that is invariant in all ZP subunits so far characterized by mass spectrometry (MS) [20–24]. Mutation of ZP-N disulphide bonds abolishes secretion of ZP3 by transfected CHO cells [43], and their functional importance in vivo is highlighted by the phenotype of substitution Y1870C in the ZP-N domain of inner ear protein tectorin alpha [46]. As discussed below, this mutation causes hearing loss in humans by affecting formation of the canonical C₅-C₆ disulphide [44].

Unlike ZP-N, the ZP-C domain of ZP3 includes an invariant C₅-C₆ disulphide that is located in the hydrophobic core, between residues on neighboring antiparallel β-strands C and F. Previously, such a feature had been found only in the second Ig-like domain of glycoprotein CD4 [47]. Furthermore, two insertions within ZP-C give rise to a ZP3-specific C-terminal subdomain, which contains three additional disulphide bonds and is positioned on the outside of the Ig-like fold (Fig. 2, A and B) [43]. As in the case of their ZP-N counterparts, pairwise mutation of the Cys residues forming the evolutionarily conserved disulphides of ZP-C either severely reduced (ΔC₅-C₆) or completely abolished secretion of ZP3 [43]. Moreover, the spatial arrangement of these Cys residues provides an explanation to the puzzling observation that different ZP3 homologues can harbor two alternative ZP-C disulphide bond configurations [20–24]: Although chicken ZP3 adopts the same C₆-C₁₁; C₈-C₉; C₁₀-C₁₂ pattern as pig ZP3, 3D clustering of the
The recombinant chicken ZP3 protein used for structure determination retained a CTP truncated before the TM and was found to be a dimer both in solution and in the crystals (Fig. 2) [43]. At the center of the dimer interface, an antiparallel β-sheet is formed by a short F'-strand in the FG-loop of ZP-N and an E'-strand located below the β-sandwich of the ZP-C domain. This intermolecular interaction is ensured by the high electrostatic complementarity of the two binding partners [43], and it relies on induced fit of the ZP-N moiety of the ZP module, which lacks the F'-strand when crystallized in the absence of ZP-C [44]. As underlined by the high conservation of residues at the interface, the mutation of which abolishes protein secretion [43], formation of the homodimer is essential for biogenesis of ZP3 and, thus, egg coat assembly.

PREVENTION OF PREMATURE POLYMERIZATION AND EVENTS LEADING TO ZP FILAMENT ASSEMBLY

Secretion and incorporation of ZP3 into the ZP are dependent on the EHP, which is conserved in the CTP of all egg coat subunits as well as ZP domain proteins unrelated to fertilization [25, 30]. Analysis of ZP3 proteins mutated in either the EHP or a related internal hydrophobic patch (IHP) region within the ZP module suggested that the presence of both patches prevents premature polymerization of the ZP3 precursor, possibly due to interaction between the patches themselves [25]. This hypothesis was supported by parallel studies on the ZP domain-containing protein uromodulin [32], but other possible functions for the EHP were also suggested [30].

Unexpectedly, the structure of chicken ZP3 revealed that the EHP constitutes the G-strand of the Ig-like fold of ZP-C, so the patch is an integral part of the domain. Moreover, the EHP G-strand is positioned opposite to the IHP, which corresponds to the A-strand of the β-sandwich, and lies at the intramolecular interface between the ZP-C and ZP-N domains (Fig. 2A). The EHP is mainly held together by interactions between hydrophobic residues, the mutation of which abolishes protein secretion [25, 43]. Consequently, four important interfaces exist within the chicken ZP3 structure that both ensure protein dimerization—a prerequisite for secretion—and mask areas that take part into polymerization: two direct intermolecular interfaces between the ZP-N and ZP-C domains of opposite subunits, and two EHP-mediated intramolecular interfaces between the ZP-N and ZP-C domains within each monomer (Fig. 2A) [43]. Although burial of the EHP makes it unlikely that this binds shuttling chaperone(s) important for intracellular trafficking of ZP3 [30], a recent study on ZP domain protein deleted in malignant brain
monocytes and monocyte/macrophage tissue 1 (also known as hensin) [48] suggests that the essential cis-Pro residue in the patch might perhaps be targeted by a secreted peptidyl-prolyl cis-trans isomerase during early stages of ZP3 folding. On the other hand, consistent with the finding that ZP subunits traffic independently within oocytes before incorporation into the ZP [49], the position of the EHP does not support the idea that the patch might allow the precursors of ZP subunits to form complexes inside the oocyte [30]; rather, EHP appears to lock the ZP-N and ZP-C domains of the ZP3 precursor in a conformation that is incompatible with polymerization [25, 43]. So what could happen when the CTP, which contains the EHP and the membrane anchor, is cleaved and leaves the dimer? What structural rearrangements could this trigger in ZP3 that lead to interaction with other ZP components and polymerization?

In the structure of chicken ZP3, the CTP is positioned so that the CFCS sits within a surface loop facing the extracellular space. This is followed by the EHP, which penetrates through the core of the homodimer toward the TM on the other side of the molecule (Fig. 2A). Such an arrangement makes the cleavage site highly accessible to protease-mediated hydrolysis, in agreement with the observation that chicken ZP3 can be cleaved by trypsin immediately before the CFCS, which was mutated in the recombinant protein used for crystallographic analysis [43]. Anchoring of the CTP to the plasma membrane suggests that following the initial cleavage event, the EHP must exit the structure in the direction of the TM. Indeed, this directionality is likely to also be maintained in ZP protein precursors that lack a membrane-anchoring element, because processing of ZP3 constructs truncated before the TM is still required for EHP dissociation [43]. Exit of the patch must require significant structural rearrangements (Fig. 2A): The intramolecular interface between ZP-N and ZP-C must open up, and because the EHP is fully incorporated into the ZP-C fold, many hydrogen bonds between the patch and its neighboring F-strand, as well as its interactions with the hydrophobic core, must be broken. Analysis of EHP dissociation after cleavage suggests that this process is slow but can occur spontaneously in vitro [43]; nevertheless, it is possible that interaction between cognate ZP subunits at polymerization facilitates EHP ejection so that the two processes are strictly coordinated in vivo.

**POTENTIAL POLYMERIZATION INTERFACES**

As a result of EHP dissociation, protein-protein interaction regions must be exposed that trigger polymerization of ZP subunits. Structural analysis of mouse ZP3 ZP-N and full-length chicken ZP3 has identified a number of potential elements that could take part in these contacts. An obvious candidate is the IHP β-strand that has lost its EHP counterpart on the opposite side of the β-sandwich and is therefore exposing its hydrophobic surface. Another potential player could be the EHP neighboring F-strand, which is linked to the C-strand through a Cys-Cys disulfide bond conserved in all ZP3 homologues so far characterized [20–24, 43]. Interestingly, the analogous disulfide in the second Ig-like domain of extracellular protein CD4 has been suggested to take part in domain swapping and dimerization events [50]. Next to the EHP exit pathway also lies the ZP-N domain E'-F-G extension that contains a highly conserved Tyr residue (Figs. 1C and 2A) implicated in polymerization by human deafness-associated tectorin alpha mutation Y1870C [46]. Introduction of a corresponding Y111C mutation in the ZP-N domain of mouse ZP3 interfered with correct disulfide bond formation, suggesting that tectorin alpha Y1870C acts as a dominant negative by incorporating into filaments of the tectorial membrane—an important extracellular matrix of the inner ear—and disrupting their further elongation [44]. It is possible that structural changes resulting from EHP ejection are transmitted to the E'-F-G extension through the loops around the Cys-Cys disulfide of ZP-N, because this adopts different conformations in the structure of isolated mouse ZP3 ZP-N and of full-length chicken ZP3, where it interacts with the ZP-C domain [43, 44].

In addition to the potential interaction interfaces mentioned so far, all of which could be inhibited from taking part in polymerization because of spatial proximity to the EHP, other potential regions could be involved in ZP assembly. Between the A- and G-strands around Cys-Cys in the ZP-N domain (Figs. 1C and 2A) is a conserved hydrophobic surface that in the structure of mouse ZP3 ZP-N packs against maltose-binding protein (MBP)—a fusion partner that was used to reduce aggregation of ZP-N and facilitate its crystallization [44]. Because the same region is exposed in the crystals of full-length chicken ZP3 [43], it might be available for interaction with other ZP subunits during polymerization. Finally, structural differences reflected by the alternative disulfide bond patterns of the ZP3-specific ZP-C subdomain appear to have a role in determining the specificity of ZP subunit interaction. This is because alternative disulfide connectivities have been found in type II ZP subunits (ZP1/ZP2/ZP4) that seem to correlate with those in type I subunit ZP3 [20]. Accordingly, some experimental evidence suggests that only certain combinations of type I/type II disulfide patterns might be compatible with productive interaction between the respective subunits and incorporation into the ZP [51].

Because no high-resolution structural information is available on the final result of polymerization—the ZP filament—it is still a matter of speculation which of the aforementioned potential interfaces interact with each other during egg coat assembly. Similarly, it remains to be established whether ZP3 incorporates into the filament as a homodimer and if the same quaternary structure also is conserved in other ZP subunits. In this regard, it is notable that the dimensions of the chicken ZP3 dimer approximately match those of the structural repeat found in ZP filaments [5]. Moreover, a recent analysis of the molecular basis of hatching in fish [52] suggests that the dimer interface of chicken ZP3, the structure of which represents a snapshot of the protein after CFCS cleavage but before EHP dissociation (Fig. 2A), could be maintained in the egg coat. However, these data do not exclude the possibility that a comparable heterodimeric interface might be established in the filaments between ZP3 and ZP1/ZP2/ZP4 through domain swapping of activated ZP subunits.

**STRUCTURE OF ZP3 RECONCILE APPARENTLY CONTRASTING SPERM-BINDING DATA**

A large body of experimental evidence from a number of independent laboratories suggests that the very C-terminal part of mature ZP3 plays a major role in its ability to bind sperm in a species-restricted manner. In particular, this function has been ascribed to the region corresponding to exon 7 of the mouse Zp3 gene, the sequence of which has relatively high variability between species due to positive Darwinian selection [53]. Following seminal biochemical studies that strongly implicated mouse ZP3 O-linked oligosaccharides in sperm binding [54], experiments with recombinant protein expressed in embryonal carcinoma (EC) cells suggested that the functional carbohydrates were linked to S332 and S334, located within a so-called sperm combining site (SCS) in the peptide encoded by exon 7 [55]. However, MS analysis of native mouse ZP material revealed that these sites are not modified in vivo [24], and...
replacement of wild-type Zp3 with a transgene expressing a protein that lacks O-glycans at S332 and S334 (as well as N-glycans at N327 and N330) was recently reported not to impair mouse fertility [40]. On the other hand, the finding that native mouse ZP3 as well as human ZP3 expressed by transgenic mice carry two conserved O-glycosylation sites within the ZP domain (site 1, mouse ZP3 T155; site 2, mouse ZP3 T162/S164/S165) led to the suggestion that these might be the authentic attachment sites of the sperm-binding carbohydrates on ZP3 [56]. As in the case of the alternative disulfide bond patterns of different ZP3 homologues, the crystal structure of the protein provides a possible explanation for these seemingly incompatible observations (Fig. 3).

In the electron-density map of chicken ZP3, clear evidence was found for a single O-linked oligosaccharide chain attached to T168, a residue that corresponds to mouse/human ZP3 site 1 [43]. The carbohydrate, which based on both MS and lectin-binding analyses is thought to have a T-antigen (Galβ1-3GalNAc) structure, is located in a region of the flexible loop between ZP-N and ZP-C that is ordered in the crystals and packs on top of ZP-C (Fig. 2A). In agreement with the exposed position of the sugar, which does not make significant contacts with the polypeptide moiety of the protein, a chicken ZP3 T168A mutant was indistinguishable from the wild type in terms of secretion and biochemical behavior. However, binding of the mutant protein to avian sperm was reduced by approximately 80% compared with the wild type. Because no other O- or N-linked carbohydrate chain was attached to the chicken ZP3 protein used for the binding assays, this result strongly suggests that the O-glycan attached to the evolutionarily conserved site 1 plays a major role in gamete interaction. At the same time, T168 is located on the same protein surface of the chicken ZP3 residues corresponding to mouse ZP3 T162/S164/S165 (site 2) and S332/S334 (SCS) (Fig. 3). Thus, when properly O-glycosylated in other ZP3 homologues, these sites could also be involved in carbohydrate-mediated binding, either together with site 1 or, as in the case of EC cell-derived ZP3, instead of it. Moreover, a conserved N-glycosylation site (N159 in chicken ZP3), which was inactivated in the protein used for crystallography, is located in a disordered region of the interdomain loop that precedes T168 by a few residues. As in the case of the O-glycan attached to T168, this N-glycan is not required for secretion in either chicken ZP3 [43] or mouse ZP3 [49, 57], and it might also contribute to sperm binding, at least in some species [41, 42]. Notably, the structure suggests that this and all other N-glycans of mouse and human ZP3 are exposed on the extracellular side of the molecule. Finally, modeling of the hypervariable C-terminal region of mature ZP3, which is not defined in the crystal structure due to high conformational flexibility, suggests that this part of the protein could also be exposed in proximity to the glycosylated sites. As a result, the C-terminal region could modulate the function of the glycans as well as provide polypeptide moieties that might be recognized by sperm. This would explain the residual binding activity of chicken ZP3 T168A as well as data suggesting that core-1-derived O-glycans may not be strictly required for fertilization in the mouse [58]. Taken together, these considerations are in line with the hypothesis that multiple distinct binding events might be involved in the recognition of ZP3 by different counterpart proteins on sperm [41, 59].

**IMPLICATIONS FOR THE OVERALL DOMAIN ARCHITECTURE OF THE ZP**

Threading of sequences N-terminal to the ZP module of other egg coat components onto the 3D structure of the mouse ZP3 ZP-N domain revealed that in Tetrapoda, ZP1 and ZP4 contain an additional copy of this fold [44]. In the case of ZP1, this is spaced from the central trefoil domain and C-terminal ZP module by a Pro/Thr/Leu-rich region, which in chicken ZP1 is expanded into an approximately 450-residue linker containing a highly repetitive sequence pattern [60]. Moreover, in a subset of ZP1 homologues, the N-terminal ZP-N domain contains an additional unpaired Cys residue within the loop between β-strands C and D, which is presumably responsible for forming an intermolecular disulfide bond that underlies the observed homodimerization of chicken, mouse, and human ZP1 [2, 61, 62].

Similarly, the whole N-terminal region of ZP2 consists of three ZP-N domains (ZP-N1 to ZP-N3) arranged in tandem and connected by short linkers, suggestive of an extended rod-like structure (Figs. 1A and 4A) [44]. Because polymerization of egg coat subunits is mediated by their C-terminal ZP modules [31, 63], and because ZP assembly is not hindered by fusion of large protein tags N-terminal to the ZP module itself [49], the ZP-N repeats of ZP2 likely project out from the ZP module core of ZP filaments (Fig. 1B). Such a supramolecular arrangement is supported by early EM studies of the filaments [4], showing protrusions with a length of approximately 110 Å that matches the span of a homology model of ZP2 ZP-N1 to ZP-N3 (Fig. 4A) [44]. This model, which is consistent with all biochemical data currently available on ZP2, not only visualizes protein regions thought to be important in binding to acrosome-reacted sperm [34] but also provides information on its postfertilization function.

**FIG. 3.** Different ZP3 regions implicated in binding to sperm are exposed on the same surface of the protein. Chicken ZP3 O-glycosylation site 1 (T168) is depicted in red, with the crystallographically resolved carbohydrate [43] shown in a stick representation; chicken ZP3 amino acids corresponding to mouse/human ZP3 site 2 residues [56] are colored in orange. Mouse ZP3 SCS residues S329 to S334 are modeled on top of the structure of the C-terminal part of mature chicken ZP3 (the last visible residue of which, P343, corresponds to S331), with S332 and S334 [55] shown in green.
Upon gamete fusion or artificial activation of unfertilized mouse oocytes with ionophore A23187, ZP2 is specifically cleaved into two proteolytic fragments of approximately 23 kDa (ZP2-N-Terminal Fragment [ZP2-NTF]) and 90 kDa (ZP2f) that remain linked by disulfide bridges [64]. The conversion is catalyzed by an as-yet-unidentified protease that is released into the perivitelline space upon fusion of the egg cortical granules with the plasma membrane [65]. ZP2 cleavage correlates with hardening of the ZP, an alteration of the mechanical properties of the matrix that results from the compaction of filaments in the inner layer of the egg coat and that is thought to contribute to the block to polyspermy by hindering sperm penetration [66, 67]. For simplicity, ZP subunits within each filament are depicted as monomers.

Mapping of the conserved ZP2 cleavage site [40, 70–72] onto the homology model of the protein indicates that the ZP2-specific protease severs the loop that connects β-strands B and C of the second ZP-N repeat (ZP-N2) (Fig. 4A). Thus, ZP2-NTF largely consists of the first ZP-N domain of ZP2, which following cleavage remains attached to the rest of the protein via the canonical C1-C4 disulfide bond of ZP-N2 [44]. Because the ZP-N2 BC loop is at the interface with ZP-N1, cleavage could cause a change in the relative position of the two repeats. This might in turn expose previously shielded protein surfaces, triggering interaction between N-terminal regions of ZP2 molecules that belong to different filaments (Fig. 4B) or, alternatively, between cleaved ZP2 molecules on one filament and the ZP module core of another. By analogy with Ig domain zippers formed by other extracellular matrix proteins [73, 74], the establishment of these new contacts would bring together adjacent ZP filaments and tighten the structure of the egg coat [44]. This may both shield recognition elements on the ZP, causing loss of sperm binding upon ZP2 cleavage [40], as well as contribute to creating a physical block to sperm penetration. Moreover, it would explain the recent observation that ZP2-NTF remains associated with the ZP of two-cell embryos even upon reduction of disulfide bonds [40].

ORIGIN OF THE ZP AND PARALLELS WITH REPRODUCTIVE PROTEINS FROM MOLLUSK AND YEAST

The structure-based identification of isolated ZP-N domains lacking a ZP-C counterpart within ZP subunits of Tetrapoda strongly supports a suggestion derived from hydrophobic cluster analysis [75]. Together with the presence of single

FIG. 4. Postfertilization cleavage of ZP2 and hardening of the ZP. **A** Homology model of the three ZP-N repeats that constitute the N-terminal region (NTR) of mammalian ZP2, followed by the canonical ZP-N domain of its C-terminal ZP module [44]. The location of the postfertilization cleavage site [40, 70–72] is marked in magenta. **B** Possible mechanism of ZP hardening. Cleavage of the ZP-N2 domain of ZP2 by an egg cortical granule protease (left; green arrows) [64, 65] could cause a conformational change of the ZP2 NTR (middle) that leads to interaction between adjacent ZP filaments and compaction of matrix structure (right) [66, 67]. For simplicity, ZP subunits within each filament are depicted as monomers.

STRUCTURAL INSIGHTS INTO ZP2 CLEAVAGE-ASSOCIATED EVENTS

Upon gamete fusion or artificial activation of unfertilized mouse oocytes with ionophore A23187, ZP2 is specifically cleaved into two proteolytic fragments of approximately 23 kDa (ZP2-N-Terminal Fragment [ZP2-NTF]) and 90 kDa (ZP2f) that remain linked by disulfide bridges [64]. The conversion is catalyzed by an as-yet-unidentified protease that is released into the perivitelline space upon fusion of the egg cortical granules with the plasma membrane [65]. ZP2 cleavage correlates with hardening of the ZP, an alteration of the mechanical properties of the matrix that results from the compaction of filaments in the inner layer of the egg coat and that is thought to contribute to the block to polyspermy by hindering sperm penetration [66, 67]. Experiments in the *Xenopus* system showed that the ZP2-cleaving protease is responsible for hardening of the frog VE [68]. On the other hand, recent transgenesis experiments suggest that processing of ZP2 may not be essential for preventing polyspermy in the mouse but, rather, plays an important role in regulating the interaction between gametes [40, 69]. How can cleavage of ZP2 alter the overall conformation of the egg coat so that this becomes refractory to sperm binding and penetration?

Mapping of the conserved ZP2 cleavage site [40, 70–72] onto the homology model of the protein indicates that the ZP2-specific protease severs the loop that connects β-strands B and C of the second ZP-N repeat (ZP-N2) (Fig. 4A). Thus, ZP2-NTF largely consists of the first ZP-N domain of ZP2, which following cleavage remains attached to the rest of the protein via the canonical C1-C4 disulfide bond of ZP-N2 [44]. Because the ZP-N2 BC loop is at the interface with ZP-N1, cleavage could cause a change in the relative position of the two repeats. This might in turn expose previously shielded protein surfaces, triggering interaction between N-terminal regions of ZP2 molecules that belong to different filaments (Fig. 4B) or, alternatively, between cleaved ZP2 molecules on one filament and the ZP module core of another. By analogy with Ig domain zippers formed by other extracellular matrix proteins [73, 74], the establishment of these new contacts would bring together adjacent ZP filaments and tighten the structure of the egg coat [44]. This may both shield recognition elements on the ZP, causing loss of sperm binding upon ZP2 cleavage [40], as well as contribute to creating a physical block to sperm penetration. Moreover, it would explain the recent observation that ZP2-NTF remains associated with the ZP of two-cell embryos even upon reduction of disulfide bonds [40].
isolated ZP-N domains in several non-egg coat proteins [76], this finding provides further evidence that—as implied by its 3D structure [44]—ZP-N should be considered a domain of its own. Moreover, considering the aforementioned structural similarity between ZP-N and ZP-C, it raises the possibility that the current domain architecture of the egg coat results from duplication events that involved an ancestral ZP-N protein-encoding gene [43]. Interestingly, with the exception of some homologues of the amphibian ZPAX VE subunit [11], ZP-N repeats are not detected in VE proteins from fish, which have a structural role comparable to that of their counterparts from Tetrapoda but are not involved in sperm binding. Similarly, despite the fact that a C-terminal ZP module is also conserved in hundreds of extracellular eukaryotic proteins unrelated to fertilization [18], none of these molecules appears to contain ZP-N repeats. These considerations suggest that—in agreement with data available on ZP2 [34, 35, 40]—the repeats have functional roles specific to fertilization. This is highly reminiscent of the mollusk and ascidian egg coat proteins VERL and VC70, which also contain a C-terminal ZP module preceded by repeats that have been implicated in sperm binding [77–80]. Whereas the 12 repeats of VC70 could be easily recognized as epidermal growth factor (EGF)-like domains [78], the 22 tandem repeats of abalone VERL did not display obvious sequence similarities to known protein domains [79]. However, a recent threading analysis of VERL repeats suggested that just like their ZP2 counterparts, these probably also adopt a ZP-N fold, consistent with the absolute conservation of four Cys residues within each repeat [81]. The presence of a single VERL-like ZP-N domain within the N-terminal region of VEZP14, another abalone egg coat protein thought to be involved in gamete fusion [15], subsequently acted as a bridge toward the identification of a further putative ZP-N domain within yeast protein alpha-agglutinin, Sag1p [81]. This is consistent with significant experimental data available on Sag1p, which was already known to contain a functionally essential and structurally unusual Ig-like domain (Ig III) [82–84]. The surprising aspect of this finding comes from the fact that Sag1p Ig III/ZP-N is directly involved in the interaction between yeast haploid cells at mating [85]—a process that mirrors the encounter of egg and sperm in higher eukaryotes. These considerations raise the possibility that despite being separated by 1 billion years of evolution, reproductive processes like mammalian gamete recognition and yeast mating share common structural features [81]. At the same time, they imply that the biological role of the ZP fold in fertilization is not limited to polymerization, but can also extend to the specific recognition of molecules that do not contain a ZP module. Indeed, this is consistent with findings from a different biological system, showing that the ZP module of transforming growth factor (TGF)-beta receptor III (TGFBR3; also known as betaglycan) is directly responsible for its binding to both TGF-beta and inhibin [86–88].

CONCLUSIONS AND FUTURE DIRECTIONS

The recently reported crystal structures of ZP3 have provided information that goes well beyond the function of this particular molecule: Within just a few years, we have moved from a situation in which the fold of ZP proteins was completely unknown to one in which, due to the presence of ZP-N repeats in subunits ZP1/2/4, we can generate approximate models of all egg coat components. Of course, this does not obviate the need for additional experimental structures, but it does allow us to start interpreting a large amount of available biochemical data in structural terms. At the same time, many important questions remain to be answered. How do the ZP-C domains of type II subunits differ from that of ZP3? How do type I and type II subunits interact, and what is their relative arrangement within ZP filaments? How do ZP components bind to their sperm counterparts at the atomic level, and what is the structural basis of the species-specificity of gamete recognition? What conformational changes do ZP filaments undergo during postfertilization hardening, and how are they unraveled at hatching? Future structural studies addressing these issues have the potential to make fundamental contributions to our understanding of fertilization and early development. Moreover, structural information on ZP3 and other ZP subunits could be used to inform the rational design of targeted nonhormonal contraceptives as well as to interpret the effect of mutations in human ZP genes associated with infertility. In this regard, it is notable that human ZP3 sequence variation P221R, which was very recently identified in a single in vitro fertilization patient with an anomalous ZP [89], affects a residue located next to ZP-C β-strands E’ and F. Although the significance of this variant remains to be established using larger sample sets, P221R might destabilize dimer formation or impair the ability of strand F to take part in polymerization following EHP ejection.

ACKNOWLEDGMENTS

We are grateful to present and past members of our laboratory and to Jan Aagaard, Franco Cotelli, Harvey Florman, Eveline Litscher, Tsukasa Matsuda, Luca Ramboldi, Willie Swanson, Vic Vaccquier, and Paul Wassarman for many helpful discussions throughout the years.

REFERENCES


75. Swanson WJ, Vacquier WD. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalence molecule. Proc Natl Acad Sci U S A 1997; 94:6724–6729.


