Cell

Insights into Egg Coat Assembly and Egg-Sperm Interaction from the X-Ray Structure of Full-Length ZP3

Ling Han,[1,5] Magnus Monné,[1,5,6] Hiroki Okumura,[1,2,5] Thomas Schwend,[1,7] Amy L. Cherry,[1] David Flot,[3,8] Tsukasa Matsuda,[4] and Luca Jovine[1,*]

[1]Department of Biosciences and Nutrition and Center for Biosciences, Karolinska Institutet, Hälsovägen 7, Huddinge SE-141 83, Sweden
[2]Department of Applied Biological Chemistry, Faculty of Agriculture, Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan
[3]EMBL Grenoble, 6 Rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France
[4]Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
[5]These authors contributed equally to this work
[6]Present address: Department of Pharmaco-Biology, University of Bari, Via E. Orabona 4, Bari I-70125, Italy
[7]Present address: Biomolecular Mass Spectrometry and Proteomics Group, Utrecht University, Padualaan 8, 3584 CH, Utrecht, The Netherlands
[8]Present address: ESRF, 6 Rue Jules Horowitz, BP 220, 38043 Grenoble Cedex 9, France
*Correspondence: luca.jovine@ki.se
DOI 10.1016/j.cell.2010.09.041

SUMMARY

ZP3, a major component of the zona pellucida (ZP) matrix coating mammalian eggs, is essential for fertilization by acting as sperm receptor. By retaining a propeptide that contains a polymerization-blocking external hydrophobic patch (EHP), we determined the crystal structure of an avian homolog of ZP3 at 2.0 Å resolution. The structure unveils the fold of a complete ZP domain module in a homodimeric arrangement required for secretion and reveals how EHP prevents premature incorporation of ZP3 into the ZP. This suggests mechanisms underlying polymerization and how local structural differences, reflected by alternative disulfide patterns, control the specificity of ZP subunit interaction. Close relative positioning of a conserved O-glycan important for sperm binding and the hypervariable, positively selected C-terminal region of ZP3 suggests a concerted role in the regulation of species-restricted gamete recognition. Alternative conformations of the area around the O-glycan indicate how sperm binding could trigger downstream events via intramolecular signaling.

INTRODUCTION

The first fundamental step of animal fertilization is binding between egg and sperm, whose fusion generates a zygote that will develop into a new individual. A specialized extracellular matrix of the egg, called zona pellucida (ZP) in mammals and vitelline envelope (VE) in nonmammals, is crucial for this process by directly mediating species-restricted recognition between gametes (Wassarman and Litscher, 2008). In the mouse, the ZP consists of glycoproteins ZP1 (100 kDa), ZP2 (120 kDa), and ZP3 (83 kDa). These components are coordinately secreted by growing oocytes and polymerize into μm-long filaments with a structural repeat of 14 nm. Pairs of filaments are then cross-linked by homodimers of the less abundant ZP1 subunit, giving rise to the three-dimensional (3D), 6.5 μm thick ZP matrix. In other mammals, the egg coat also contains a fourth subunit (ZP4) that is ~30% identical to ZP1; moreover, proteins homologous to mammalian ZP1–4 constitute the VE of other vertebrates, and highly related molecules comprise the egg coat of species evolutionarily very distant from mammals, like molluscs and ascidians. The basic structure of the ZP/VE has thus been conserved over more than 600 million years of evolution (Monné et al., 2006).

As indicated by in vitro sperm binding experiments (Bleil and Wassarman, 1980) and exemplified by the phenotype of ZP3 null mice, which produce eggs that lack a ZP and are completely infertile (Liu et al., 1996; Rankin et al., 1996), mouse ZP3 (mZP3) is essential for fertilization in vivo by acting as receptor for sperm (Wassarman and Litscher, 2008). This is supported by numerous studies in different mammalian species, including human (Barratt et al., 1993), as well as in other vertebrates such as chicken (Bausek et al., 2004) and *Xenopus* (Vo and Hedrick, 2000). However, the specific ZP3 determinants recognized by sperm are highly controversial, and the molecular basis of gamete interaction remains elusive (Gahlay et al., 2010; Wassarman and Litscher, 2008; Shur, 2008).

The domain structure of ZP3 reflects its dual biological function. Most of the protein consists of a polymerization module of 260 residues, the so-called ZP domain (Bork and Sander, 1992), followed by a C-terminal region of 40 amino acids that is specific to ZP3 and has been implicated in interaction with sperm (Wassarman and Litscher, 2008). The ZP module is not

only conserved in egg coat components but is also found in many other secreted eukaryotic proteins with variable architecture and biological function (Jovine et al., 2005; Bork and Sander, 1992). It is responsible for the incorporation of ZP3 and other subunits into the ZP (Jovine et al., 2002) and consists of two domains, ZP-N and ZP-C, that are separated by a protease-sensitive linker (Jovine et al., 2004). Whereas ZP-N is thought to constitute a basic building block of ZP filaments (Monné et al., 2008), ZP-C may mediate the specificity of interaction between subunits (Kanai et al., 2008; Sasanami et al., 2006). These processes are controlled by an external hydrophobic patch (EHP) contained within the C-terminal propeptide of ZP component precursors and an internal hydrophobic patch (IHP) inside the ZP module (Jovine et al., 2004).

A recent crystal structure of the ZP-N domain of mZP3 had important implications for the architecture of animal egg coats (Monné et al., 2008). However, it could not address the function of ZP3 as a sperm receptor, and, apart from a cryo-electron microscopy study of glycoprotein endoglin at 25 Å resolution (Llorca et al., 2007), no structural information is available on the complete ZP module and the regulation of its biological function. Here we present the high-resolution structure of full-length ZP3, providing crucial insights into both the mechanism of ZP module-mediated polymerization and the sperm binding activity of this key reproductive protein.

RESULTS

Protein Engineering and Structure Determination

Biogenesis of ZP3 requires processing of an N-terminal signal peptide, formation of six intramolecular disulfide bonds, and loss of a C-terminal propeptide that contains a polymerization-blocking EHP and a single-spanning transmembrane domain (TM). The latter event depends on cleavage of the protein precursor at a consensus furin-cleavage site (CFCS) located between the ZP-C domain and the EHP (Figure S1A and Figure S2A available online; Wassarman and Litscher, 2008). As a result of this complex maturation pathway, correctly folded recombinant ZP3 can only be efficiently expressed in mammalian cells. However, due to its heavy and heterogeneous glycosylation (accounting for ~50% of the total apparent mass of the mouse protein), as well as its tendency to aggregate when concentrated or enzymatically deglycosylated (Zhao et al., 2004; E. Litscher and P. Wassarman, personal communication), full-length ZP3 has eluded attempts at structure determination for over 25 years.

To overcome this impasse, we focused on chicken ZP3 (cZP3), a naturally hypoglycosylated homolog that contains a single N-glycosylation site and is 53% identical to human ZP3 (Takeuchi et al., 1999; Waclawek et al., 1998). A series of progressively modified, C-terminally histidine-tagged constructs (Figure S1A) were expressed in Chinese hamster ovary (CHO) cells (Figure S1B), which were previously shown to produce a recombinant avian ZP3 protein that is indistinguishable from its native counterpart (Sasanami et al., 2003). Deletion of the TM and inactivation of the N-glycosylation site and the CFCS resulted in construct cZP3-3 (Figure S1A), which was secreted from cells as a single homogeneous species of 41 kDa (Figure S1B, lane 9). Because it retained the EHP at its C terminus, this protein did not aggregate and could be purified by immobilized metal affinity chromatography (IMAC), followed by size-exclusion chromatography (SEC). The latter suggested that cZP3 exists as a dimer (Figure S1C), in agreement with crosslinking experiments (Figure S1D) and sedimentation equilibrium studies of human ZP3 (Zhao et al., 2004).

cZP3-3 had relatively low solubility and yielded only weakly diffracting crystals. However, its solubility could be significantly improved by limited trypsinization, which resulted in loss of an N-terminal fragment (residues Y21–R46; Figure S3) that is not conserved among ZP3s and is missing in the mature avian protein (Pan et al., 2000; Waclawek et al., 1998). Further mass spectrometric (MS) analysis of trypsinized forms of cZP3-3 and cZP3-4, a better expressed construct carrying a deletion of P23-H52 (Figures S1A and S1B, lane 11), revealed that the improvement in solubility was in fact due to proteolysis of a second fragment (R348–R358) immediately preceding the inactivated CFCS (Figure S3 and Figure S4). Trypsinized cZP3-3 and cZP3-4 (cZP3-3T/4T) produced tetragonal crystals that diffracted to high resolution despite 71% solvent content (Figures S5A and S5B). The structure of cZP3-4T was solved by molecular replacement using the ZP-N domain of mZP3 (Monné et al., 2008) as search model and refined against both a dataset at 2.0 Å resolution and an earlier 2.6 Å dataset that better resolved a functionally important O-linked carbohydrate (Table S1 and Figures S5C–S5F).

Overall Architecture of the ZP3 Homodimer

In the asymmetric unit, two molecules of ZP3 embrace each other in antiparallel orientation to form a flat, Yin-Yang-shaped homodimer (Figures 1A and 1B). Although part of the linker between ZP-N and ZP-C (E158–R166) is disordered in the electron density map, the connectivity between the two domains is unequivocally determined by their relative positions in the crystal. In this arrangement, the two ZP modules of the dimer are held together by interactions between ZP-N and ZP-C domains that belong to opposite subunits. On the other hand, no ZP-N/ZP-N or ZP-C/ZP-C contacts are observed within the dimer (Figure 1A).

Interaction with ZP-C Induces Local Rearrangements of Two Conserved ZP-N Domain Regions

The structure of a maltose-binding protein-mZP3 ZP-N fusion revealed that the ZP-N domain belongs to a distinct immunoglobulin (Ig) superfamily subtype, characterized by an E' strand and two invariant disulfides that link the first four Cys of the ZP module with C_1-C_4, C_2-C_3 connectivity (Monné et al., 2008). Consistent with the fact that the model of mZP3 ZP-N was sufficient to phase the structure of cZP3-4T despite representing only 28% of the scattering mass in the asymmetric unit, the secondary structure elements of cZP3 and mZP3 ZP-Ns can be superimposed with a C_α root-mean-square distance of 0.9 Å (Figure S6A). However, as further discussed below, contacts with ZP-C cause significant local differences in a conserved region within the long FG loop of the ZP-N domain, as well as around its invariant C_2-C_3 disulfide.

The ZP Module Is Internally Symmetric

As hinted by initial molecular replacement solutions that placed additional copies of ZP-N at the position of ZP-C, the latter

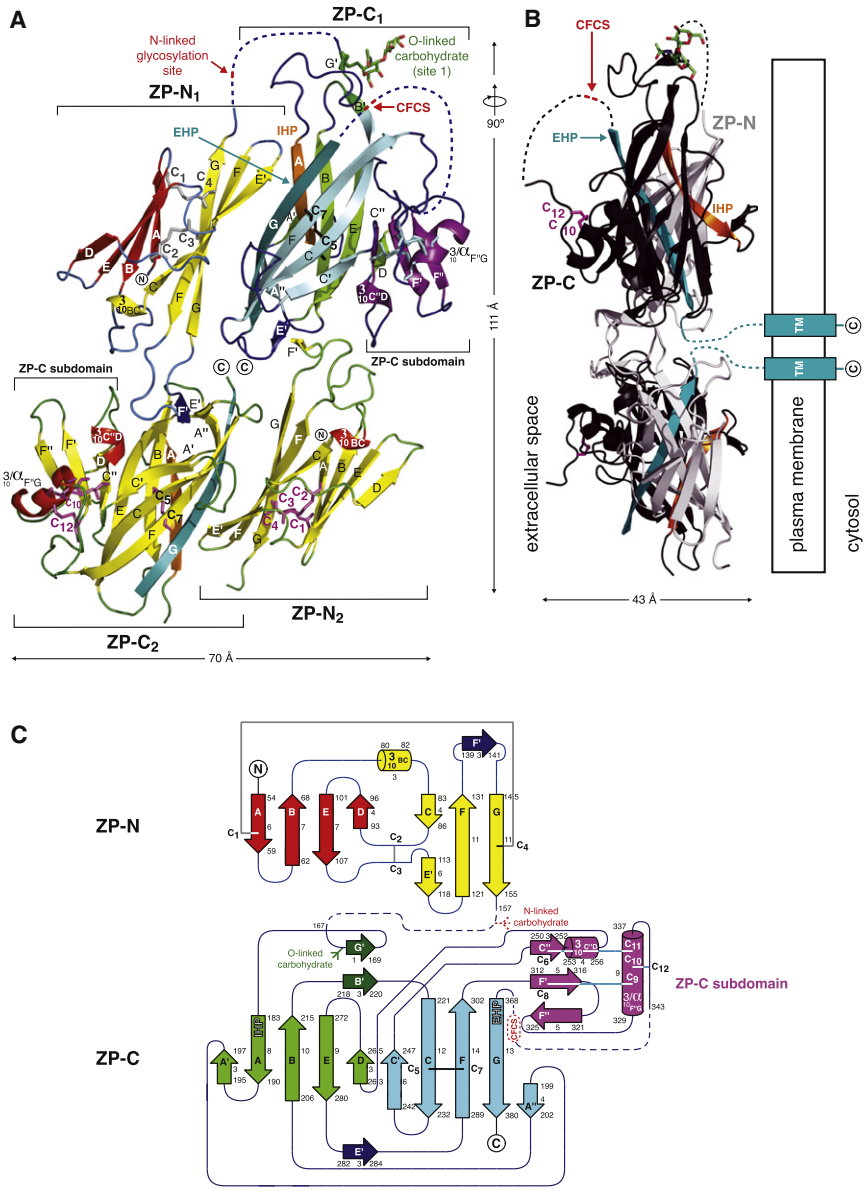

Figure 1. Overall Structure and Topology of cZP3

(A) Cartoon diagram of the cZP3 homodimer structure, formed by two ZP modules each consisting of a ZP-N and a ZP-C domain. In the upper molecule, β sheets and disulfides are colored according to the topology scheme in (C), except for ZP-C strands A (IHP; orange) and G (EHP; dark cyan). Dashed lines represent disordered loops. The lower ZP module is colored by secondary structure, with the IHP and EHP depicted as above and disulfides in magenta.

(B) Side view of the cZP3 homodimer with ZP-N and ZP-C domains in gray and black, respectively. The IHP and EHP lie at the domain interface. The C-terminal linkage from the EHP to the TM is indicated by a dark cyan dashed line.

(C) Topology scheme with secondary structure and disulfide connectivity.

See also Figure S1, Figure S2, Figure S5, Figure S6, and Table S1.

domain also adopts an Ig-like fold, so that 50% of the residues of cZP3-4T are involved in β strands (Figure 1 and Figure S2A). ZP-N and ZP-C display no significant sequence similarity and have different disulfide connectivity (Boja et al., 2003). Nevertheless, despite replacement of the C and E' strands of ZP-N by a single C strand in ZP-C (which also contains additional A', A", and C' strands), the β sandwiches of the two domains share a common topology (Figure 1C). As a consequence, each ZP module has internal symmetry (Figures S6C and S6D).

The EHP Is Coupled to an Invariant ZP3 Disulfide at the Core of the ZP-C Domain

Analysis of purified cZP3-3T and -4T suggested that the C-terminal tryptic peptide produced by ZP3 cleavage at R358 remained noncovalently associated with the rest of the protein (Figure S3 and Figure S4). Surprisingly, the electron density map reveals that the EHP sequence contained in this peptide constitutes the G strand of the ZP-C domain and is thus an integral part of the ZP3 fold (Figure 1 and Figure 2A). Immediately next to the EHP, a C_5-C_7 disulfide staples the F strand of ZP-C to the neighboring C strand. This linkage is conserved in all ZP3 homologs (Kanai et al., 2008) and forms a short right-handed hook that is preceded by a β bulge in the C strand and protrudes toward the center of the ZP-C hydrophobic core (Figure 2A). To gain insights into the functional role of C_5-C_7 and other ZP3 disulfides, we individually mutated all Cys pairs of cZP3-4 (Figure 2B). As shown in lane 6, C_5-C_7 is the only disulfide whose mutation does not completely abolish secretion of ZP3. This result suggests that the invariant C_5-C_7 pair of ZP3 is involved in other functions besides protein folding, consistent with absence of both of these Cys in a subset of *Drosophila* ZP module proteins with a different biological function (Fernandes et al., 2010).

Cysteine Clustering in a Structurally Variable ZP3-Specific Subdomain

Insertions within the C'D and FG loops of ZP-C give rise to a C-terminal ZP-C subdomain (Figures 1A and 1C) that is conserved in the type I ZP module of ZP3 homologs but is not found in either the type II ZP module of other ZP subunits or unrelated Ig-like domains. The ZP-C subdomain has a remote similarity to EGF domains based on secondary structure and consists of a short 3_{10} helix C"D and a three-stranded β sheet that is connected to a longer, mixed $3_{10}/\alpha$ helix F"G through C_6-C_{11} and C_8-C_9 disulfides (Figure 2C and Figure S5F). This connectivity was confirmed by the anomalous signal of sulfur and is

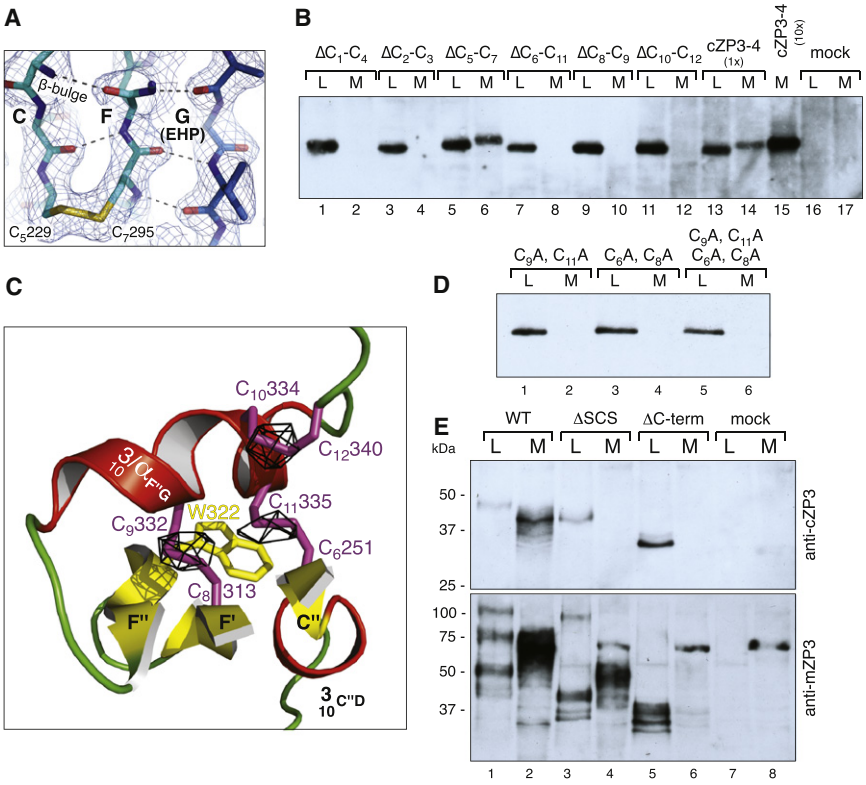

Figure 2. ZP-C Disulfide Connectivity

(A) $2F_{obs}$-F_{calc} map of the region around invariant disulfide C_5-C_7 and the EHP G strand, contoured at 1 σ. Dashed lines indicate hydrogen bonds.

(B) All disulfide-forming Cys pairs were individually substituted by Ala. The constructs were expressed and cell lysate (L) and conditioned medium (M; concentrated 10 times, unless otherwise indicated) were analyzed by immunoblot.

(C) C-terminal subdomain disulfide arrangement, showing the close proximity of C_6, C_8, C_9, and C_{11}. Black mesh is a 3.7 Å resolution phased anomalous difference map, calculated using diffraction data collected at 7.75 keV and contoured at 4 σ.

(D) Cys mutations preventing the native disulfide connectivity of the ZP-C subdomain abolish protein secretion. Medium was concentrated 5 times.

(E) Removal of C-terminal residues W322–R358 inhibited secretion of cZP3 whether the TM was present (ΔSCS) or not (ΔC-term). In corresponding mZP3 mutants lacking S309-K346 (Jovine et al., 2002), the TM rescued protein secretion.

See also Figure S2, Figure S3, and Figure S4.

consistent with partial disulfide bond assignments of pig ZP3 (C_8-C_9; C_6-C_{10}/C_{11}; C_{12}-C_{11}/C_{10}; Kanai et al., 2008). On the other hand, it differs from the disulfide pattern of fish, mouse, rat, and human ZP3, where the same Cys residues form a C_6-C_8 and, presumably, a C_9-C_{11} bridge (Figure S2B; Kanai et al., 2008; Darie et al., 2004; Boja et al., 2003). The structure reveals that, even though these Cys are spaced in sequence, they are closely clustered in space on top of a platform created by invariant W322 (Figure 2C). This 3D arrangement immediately suggests how the alternative C_6-C_8, C_9-C_{11} connectivity could be accommodated in the ZP-C subdomain. At the same time, the structure explains why cZP3 adopts the C_6-C_{11}, C_8-C_9 pattern, as the helical conformation of residues R329–T337 would not be compatible with a C_9-C_{11} disulfide.

Consistent with the latter observation, attempts to force partial formation of the alternative connectivity in cZP3 by mutating either C_6 and C_8 or C_9 and C_{11} resulted in nonsecreted protein products (Figure 2D, lanes 1–4). This suggests that formation of helix F"G is an early event in cZP3 folding that commits the C-terminal disulfides to the C_6-C_{11}, C_8-C_9 connectivity. Conversely, the same region of the protein probably adopts a different conformation in order to form the C_9-C_{11} disulfide observed in other homologs of ZP3. In support of this conclusion, mutations that either interfere with disulfide-mediated tethering of helix F"G to the rest of the subdomain (Figure 2B, lanes 7–10; Figure 2D, lanes 5–6) or delete the residues between loop F'F" and the CFCS (Figure 2E, top panel, lanes 3–6) are not tolerated by cZP3, whereas the corresponding amino acids are not required for secretion of mZP3 constructs when the TM is present (Figure 2E, bottom panel, lane 4).

ZP-N/ZP-C Contacts at the Homodimer Interface Are Essential for ZP3 Biogenesis

Electrostatic complementarity between the ZP-N and ZP-C domains of opposite ZP modules plays a major role at the interface of the homodimer, which buries 2450 Å^2 of surface area. The main interaction involves a positively charged protrusion formed by the long FG loop of the ZP-N domain of one molecule and a negatively charged cleft between ZP-C and the C-terminal subdomain of the other (Figure 3A). The tip of the ZP-N FG loop, which was loosely packed against maltose-binding protein in the ZP-N fusion crystals (Monné et al., 2008), forms a short F' β strand (Figure S6A) that generates an intermolecular antiparallel β sheet with the E' strand of ZP-C (Figure 3B). This involves a highly conserved FXF motif and is strengthened by hydrophobic contacts between the side chains of the F' strand and surrounding residues L204, Y243, and *cis*-P241. Additionally, conserved R142 forms a salt bridge with invariant D254 and an hydrogen bond with Y243. Deletion of the ZP-N F' strand or mutation of the neighboring R142 in ZP-C almost completely inhibits secretion (Figure 3C), indicating that dimer formation is a prerequisite for the biogenesis of ZP3.

Intramolecular Interaction between ZP-N and ZP-C Is Hydrophobically Mediated by the EHP

As described above, the protein used for crystallization retained a noncovalently bound C-terminal proteolytic fragment, whose EHP sequence forms the G strand of the ZP-C β sandwich (Figure 1 and Figure 2A). Notably, this positions the EHP not only next to the aforementioned invariant C_5-C_7 disulfide

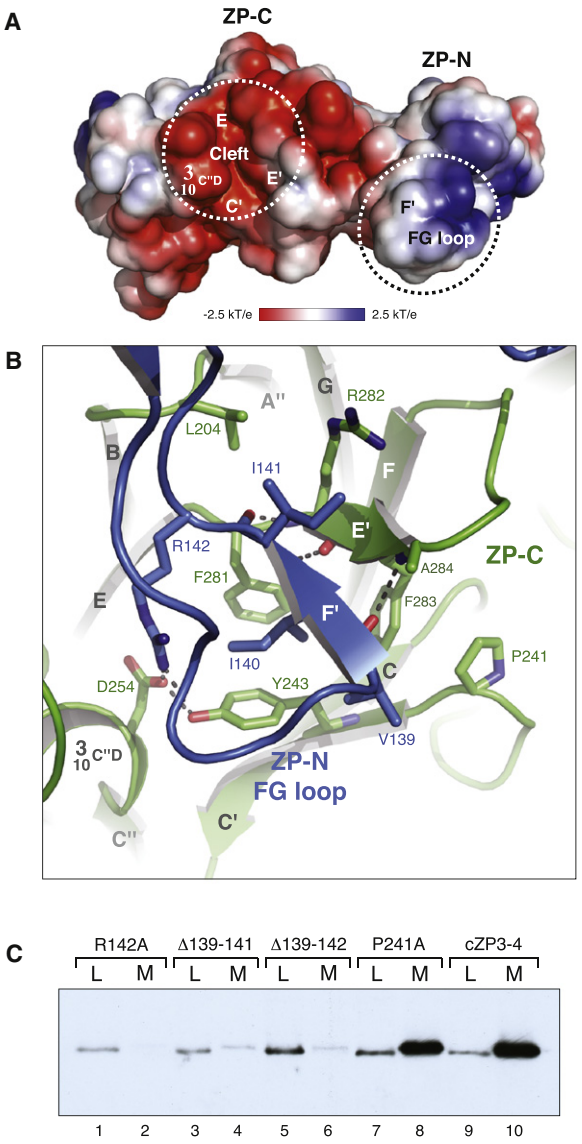

Figure 3. The Homodimer Interface

(A) Complementary electrostatic surface potential of the ZP-N FG loop and the cleft between ZP-C and the C-terminal subdomain.

(B) An intermolecular antiparallel β sheet is formed by the ZP-N F' strand of one monomer and ZP-C E' strand of the other.

(C) Whereas mutation of *cis*-P241 does not affect secretion, R142A and mutants lacking the ZP-N F' strand (Δ139–141 and Δ139–142) disrupt dimer formation and are essentially not secreted.

See also Figures S1C and S1D.

staple (Figure 2A) but also close to the E'-F-G extension of ZP-N, as well as the IHP sequence that constitutes the A strand of ZP-C (Figure 1, Figure 4A, and Figure S2A). Both of these elements have been implicated in ZP module-mediated polymerization (Schaeffer et al., 2009; Monné et al., 2008; Jovine et al., 2004).

Analysis of the 635/610 $Å^2$ interface between the adjacent ZP-N and ZP-C domains of the ZP module reveals a central role of hydrophobic interactions around absolutely conserved P376 in the EHP. This residue, which is in *cis* conformation and forms a β bulge together with invariant G375, is flanked by L290 and forms a stack of rings with highly conserved ZP-C amino acids Y292 and P235 (Figure 4A). The resulting surface interacts with V114, L127, V147, and P149 on the outside of the E'-F-G sheet of ZP-N, as well as with P87. This stretches the CD loop of ZP-N, causing the C_2-C_3 disulfide to adopt an unusual left-handed conformation and, in turn, to pull the underlying EE' region, which does not form the α helix observed in isolated mZP3 ZP-N (Figure S6B). Moreover, a Q116-E196 hydrogen bond and an R125-E196 ionic interaction are observed at one end of the interface, whereas variable contacts involving R288 are found at the other (Figure 4A). However, analysis of mutants shows that the hydrophobic contacts play a much more important role than these other kinds of interactions. In agreement with complementary mutational studies of invariant EHP residues (Schaeffer et al., 2009; Jovine et al., 2004), mutation of Y292 and P235 severely inhibits ZP3 secretion (Figure 4B, lanes 1–6), whereas an E196A mutant is secreted at levels comparable to the wild-type protein (Figure 4B, lane 8).

ZP3 Cleavage Causes Slow Spontaneous Dissociation of the EHP at Physiological Temperature

Apart from being involved in the ring stack and hydrogen-bonding to the neighboring F and A" β strands, the EHP makes many other interactions with the ZP-C domain. These include hydrophobic contacts with residues of the A, B, and F strands as well as F199 and conserved P171, F172, and F202 and a salt bridge between D371 and H296 on the F strand. Consistent with this array of interactions, our biochemical analysis of trypsinized cZP3 shows that the EHP is tightly bound to the core of the protein and is not removed by SEC or IMAC, even upon extensive washing. This raises the issue of whether the EHP can dissociate spontaneously, or if this is dependent on interaction between cZP3 and other ZP subunits. To answer this question, we incubated cZP3-4T at 39°C (the body temperature of the chicken) for 30 hr. As shown in Figure 4C (lanes 1–3), this resulted in loss of approximately 40% of the EHP from the sample, a reasonable proportion considering that avian VE assembly requires several weeks. SEC analysis (Figure 4D) revealed that—like mature native cZP3 (Bausek et al., 2004)—much of the remaining protein had formed different oligomeric states and large-molecular-weight species (gray profile) in comparison with an identical sample incubated at 4°C (violet profile), or with uncut protein incubated for the same time at 39°C (red profile). Consistent with the fact that this experiment was performed in the absence of other egg coat subunits, electron microscopy indicated that the material in the void volume peak of Figure 4D consisted of amorphous aggregates rather than polymers (data not shown).

An Evolutionarily Conserved O-Glycan Plays a Major Role in Sperm Binding

In the structure of cZP3-4T, one molecule in the homodimer has visible density for part of the ZP-N/ZP-C linker region, which can be modeled from residue P167 onward (Figure 1). Additional electron density was found next to T168, which belongs to

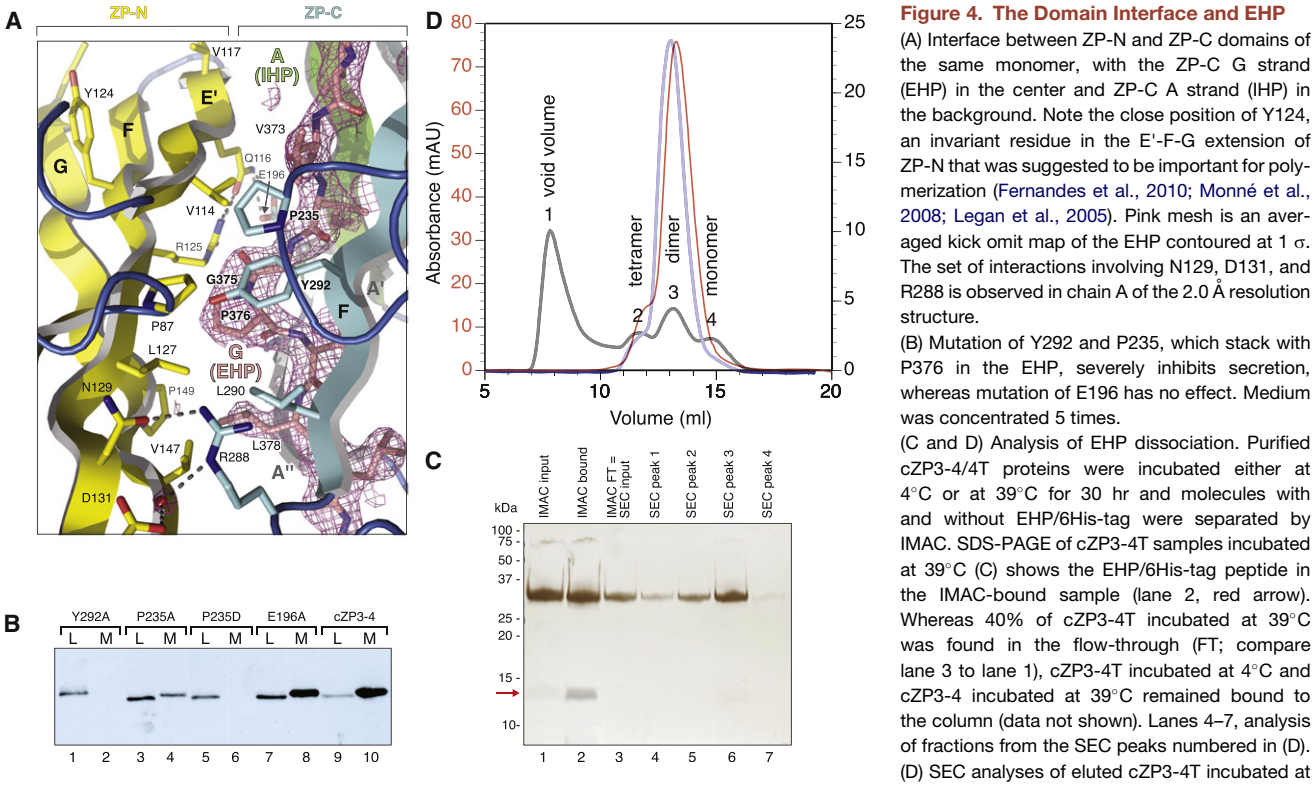

Figure 4. The Domain Interface and EHP

(A) Interface between ZP-N and ZP-C domains of the same monomer, with the ZP-C G strand (EHP) in the center and ZP-C A strand (IHP) in the background. Note the close position of Y124, an invariant residue in the E'-F-G extension of ZP-N that was suggested to be important for polymerization (Fernandes et al., 2010; Monné et al., 2008; Legan et al., 2005). Pink mesh is an averaged kick omit map of the EHP contoured at 1 σ. The set of interactions involving N129, D131, and R288 is observed in chain A of the 2.0 Å resolution structure.

(B) Mutation of Y292 and P235, which stack with P376 in the EHP, severely inhibits secretion, whereas mutation of E196 has no effect. Medium was concentrated 5 times.

(C and D) Analysis of EHP dissociation. Purified cZP3-4/4T proteins were incubated either at 4°C or at 39°C for 30 hr and molecules with and without EHP/6His-tag were separated by IMAC. SDS-PAGE of cZP3-4T samples incubated at 39°C (C) shows the EHP/6His-tag peptide in the IMAC-bound sample (lane 2, red arrow). Whereas 40% of cZP3-4T incubated at 39°C was found in the flow-through (FT; compare lane 3 to lane 1), cZP3-4T incubated at 4°C and cZP3-4 incubated at 39°C remained bound to the column (data not shown). Lanes 4–7, analysis of fractions from the SEC peaks numbered in (D). (D) SEC analyses of eluted cZP3-4T incubated at 4°C and cZP3-4 incubated at 39°C are shown in violet and red, respectively (left-hand scale), and that of FT of cZP3-4T incubated at 39°C is shown in gray with four distinct peaks corresponding to different oligomeric forms (right-hand scale).

See also Figure S1C, Figure S3, Figure S4, and Figure S6B.

a highly conserved PTWXPF ZP3 motif (Figure S2A). MS analysis identified a E158–R181 peptide containing a 365 Da Hex-HexNAc modification (Figure S7A) that, based on carbohydrate composition analysis of cZP3 (Takeuchi et al., 1999) and lectin-binding experiments (Figure S7B, lanes 2 and 5), was interpreted as Galβ1-3GalNAc (T antigen). This disaccharide could be fitted into the electron density map of the 2.6 Å structure (Figure 5A), whereas density for the second carbohydrate residue was not as well defined in the 2.0 Å crystal.

Considering the evolutionary conservation of this site, which has been denominated "site 1" and is also modified with core 1-related glycans in native mZP3, native rat ZP3, and human ZP3 expressed in transgenic mice or CHO-Lec3.2.8.1 cells (Chalabi et al., 2006; Boja et al., 2005; Zhao et al., 2004; Boja et al., 2003), T168 was mutated to Ala in order to assess the carbohydrate function. The mutant protein was expressed and secreted as efficiently as the wild-type, excluding a role for the T168 O-glycan in ZP3 biogenesis (Figure 5B). This is consistent with the observation that the Thr is substituted by other amino acids in a subset of ZP3 sequences from fish, where the protein has an equivalent structural, but not receptorial, role. As expected from the lack of a single O-linked sugar chain, a small change was observed in the migration of the mutant protein (Figure 5B), which no longer bound to either jacalin or peanut agglutinin (Figures S7B and S7C). This hinted at the lack of additional O-glycans, which was confirmed by both inspection of electron density maps and extensive MS analysis of both cZP3-4 and native cZP3. The fact that this protein carries a single O-linked carbohydrate at T168 allowed us to conclusively evaluate the role of this particular sugar chain in sperm binding, in the absence of possible compensatory effects from other glycans. Quantification of protein binding to the tip of chicken sperm head (Figure 5C) showed that the T168A mutation caused a decrease of ~80% in binding relative to wild-type cZP3-4 (Figure 5D), indicating an important role of the conserved O-glycan in avian gamete interaction.

DISCUSSION

Thirty years after ZP3 was identified (Bleil and Wassarman, 1980), this work yields structural information on an egg protein region directly recognized by sperm at the beginning of fertilization. Combined with mutational and in vitro binding studies, the structure provides insights into many aspects of ZP3 biology, ranging from secretion and polymerization to interaction with sperm. Moreover, it has important implications for human reproductive medicine.

Evolution of the ZP and Role of the ZP Module Dimer Interface

Our previous crystal structure of the ZP-N domain of mZP3 (Monné et al., 2008) strongly supported the suggestion that additional copies of ZP-N are found within N-terminal extensions of ZP1, ZP2, and ZP4 (Callebaut et al., 2007). By showing

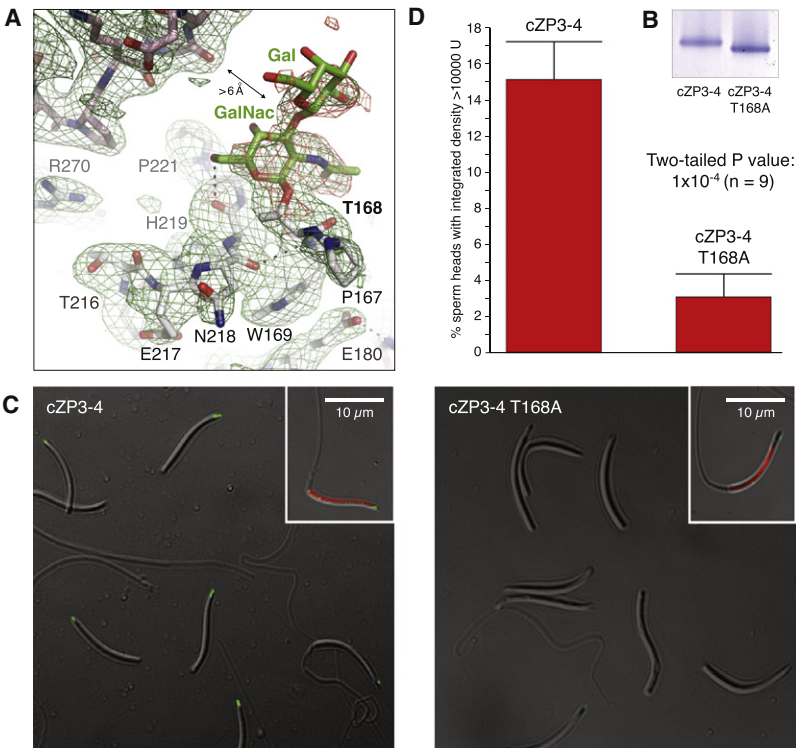

Figure 5. T168 Carries an O-Glycan Important for Sperm Binding

(A) Averaged kick omit map (0.8 σ; green mesh) and composite omit map (0.9 σ; red mesh) of the Galβ1-3GalNAc chain attached to T168.

(B) cZP3-4 T168A mutant protein shows a migration shift relative to the wild-type during SDS-PAGE.

(C) Chicken sperm were incubated with cZP3-4 and its mutant T168A and bound protein were detected by immunofluorescence (green). Corner inserts show TOTO3-stained (red) sperm.

(D) Statistically highly significant difference in the sperm-binding activity of cZP3-4 and T168A. Data are represented as mean ± standard error of the mean (SEM).

See also Figure S7.

that the ZP-C domain adopts an Ig-like fold with the same topology as ZP-N (Figures 1A and 1C), the X-ray map of full-length ZP3 reveals that the ZP module contains internal symmetry (Figures S6C and S6D). Considering that the ZP-C domain has so far been found only within the context of a complete ZP module, this observation raises the possibility that ZP-N and ZP-C—and thus essentially the whole mammalian egg coat—originated by duplication of a common ancestral Ig-like domain. Moreover, conservation of ZP-N and ZP-C residues that mediate formation of the antiparallel ZP module homodimer (Figure 1A and Figure 3B), which is essential for ZP3 secretion (Figure 3C), suggests that this quaternary structure is also important for the function of other ZP subunits and unrelated ZP module proteins. In agreement with this conclusion, a C582-C582 interchain disulfide that characterizes human endoglin (Llorca et al., 2007) can be readily modeled on the basis of the ZP module arrangement observed in the ZP3 crystal. The biological importance of the dimer interface is further highlighted by a recent study of fish embryo hatching, identifying R167 of medaka ZP3 as a target cleavage site of hatching enzymes (Yasumasu et al., 2010). Because this residue corresponds to cZP3 R142 (Figure S2A), which plays an essential role at the interface (Figure 3B and Figure 3C, lane 2), the structure immediately suggests how hatching enzymes could solubilize egg coat filaments by disrupting the stability of ZP module dimers. Considering that a mammalian homolog of fish hatching enzymes is expressed in unfertilized oocytes and preimplantation embryos (Quesada et al., 2004), conservation of the R|T cleavage site in human ZP3 (Figure S2A) might indicate that a similar mechanism is involved in human embryo hatching and implantation.

Mechanism of Protein Polymerization Inhibition by the EHP and Implications for ZP Assembly

Previous mutational studies suggested that cleavage of the membrane-bound precursors of ZP module proteins around the CFCS releases a block to polymerization by causing dissociation of the EHP (Schaeffer et al., 2009; Jovine et al., 2004). However, how the EHP inhibits polymer assembly at the molecular level, and what is its relationship with other elements involved in polymerization such as the IHP (Schaeffer et al., 2009; Jovine et al., 2004) and the ZP-N E'-F-G extension (Fernandes et al., 2010; Monné et al., 2008; Legan et al., 2005), was unknown.

The structure of ZP3 reveals that, rather than simply shielding a surface-exposed polymerization interface, the EHP penetrates through the core of the molecule by constituting β strand G of the ZP-C domain (Figure 1). This strand directly faces the IHP (ZP-C β strand A) and makes contacts with the E'-F-G face of ZP-N (Figure 4A). Although stable within the context of the uncleaved protein precursor, the resulting ZP-N/ZP-C interface is dominated by hydrophobic contacts involving the EHP. This suggests that the two domains must undergo significant rearrangements upon cleavage of ZP3 at the CFCS and dissociation of the C-terminal propeptide. Thus, the EHP blocks premature protein polymerization by acting as a "molecular glue" that keeps the ZP module in a conformation that is essential for secretion (Figure 4B) but not compatible with formation of higher-order structures.

In agreement with studies on soluble fish egg coat protein precursors secreted by the liver (Sugiyama et al., 1999), our in vitro analysis of EHP ejection shows that, even in constructs lacking a TM, the propeptide containing the EHP must be physically cleaved before the latter is released from the protein (Figures 4C and 4D). This implies that, regardless of the presence of C-terminal membrane-anchoring elements, the patch can only be ejected from the side of the homodimer opposite to where the CFCS lies; this is where the C-terminal ends of the two ZP3 subunits come almost in contact with each other (Figure 1B). Coupling of this structural constraint, probably deriving from the sharp kink made by the invariant GP sequence of the EHP

(Figure 4A), with membrane anchoring may play an important role in ZP assembly by orienting the ZP3 precursor so that it can properly interact with other subunits upon cleavage at the CFCS (Figure 1B). This would explain why, although the TM is not required for secretion, it is essential for incorporation of ZP3 into the mouse ZP (Jovine et al., 2002).

Following cleavage, dissociation of the EHP must cause exposure of a large hydrophobic region on ZP-C (Figure 4A), triggering interaction with its cognate ZP-N or another ZP module. This might depend on strand- or domain-swapping events involving the exposed IHP and the E'-F-G extension of ZP-N, which in the structure does not interact with other parts of the homodimer (Figure 1A and Figure 4A). Moreover, because of the direct structural relationship between the EHP and the F strand of ZP-C (Figure 2A), rearrangements connected with ZP assembly may also involve the C_5-C_7 disulfide staple, which is conserved in ZP3 homologs from fish to human despite not being essential for secretion (Figure 2B). Notably, a similar disulfide has been found in CD4 and implicated in domain swapping, CD4 dimerization, and entry of HIV-1 into $CD4^+$ cells (Sanejouand, 2004).

A Structural Basis for the Specificity of Egg Coat Subunit Interaction

Even though several ZP module-containing proteins can homopolymerize, formation of egg coat filaments requires ZP3 (a type I subunit) and at least one type II (ZP1/ZP2/ZP4-like) component (Jovine et al., 2005; Boja et al., 2003). Furthermore, in spite of very high sequence identity, only certain combinations of heterologous ZP subunits can productively interact to form a ZP (Hasegawa et al., 2006). How is the specificity of ZP assembly regulated at the molecular level?

Our crystallographic and mutational analysis indicates that, although clustering of conserved Cys within the ZP-C subdomain of ZP3 (Figure 2C) can account for the two disulfide connectivities observed in different ZP3 homologs (Figure S2B), these alternative patterns must be accommodated by local differences in the surrounding structural elements (Figures 2C–2E). Because the ZP-C domain mediates interaction between type I and type II ZP subunits (Okumura et al., 2007; Sasanami et al., 2006), and because different ZP3 disulfide connectivities are reflected by changes in the disulfide patterns of cognate type II proteins (Kanai et al., 2008), this suggests that the tertiary structure of the ZP-C subdomain of ZP3 determines the specificity of egg coat assembly. Considering that pig and mouse ZP3 adopt different disulfide patterns (Kanai et al., 2008), this conclusion explains why pig ZP2 does not incorporate into the mouse ZP when secreted by transgenic animals (Hasegawa et al., 2006).

Sperm Binding and Modulation of the Specificity of Gamete Interaction

Carbohydrates of ZP3 have been repeatedly implicated in binding to sperm, but there is highly conflicting evidence about the chemical nature and location of the bioactive glycans, as well as about their functional importance relative to the polypeptide moiety of the protein (Wassarman and Litscher, 2008; Shur, 2008). Nevertheless, many studies from different laboratories support the idea that initial species-restricted binding between mammalian gametes is mediated by ZP3 O-glycans (Florman and Wassarman, 1985) and involves a C-terminal region of the molecule that, in the mouse, is encoded by exon 7 of the *Zp3* gene (Figure S2A; Wassarman and Litscher, 2008; Kinloch et al., 1995). This region varies between species as a result of positive Darwinian selection (Swann et al., 2007; Turner and Hoekstra, 2006; Jansa et al., 2003; Swanson et al., 2001) and, based on mZP3 mutants expressed in embryonal carcinoma cells, was suggested to contain a sperm-combining site (SCS; Figure S2A) carrying active O-glycans at S332 and S334 (Chen et al., 1998). This hypothesis was challenged by MS analysis of purified ZP material, which indicated that the same sites are not glycosylated in native mZP3 (Boja et al., 2003). A suggestion was thus made that the functional O-glycans of the native protein are instead located at site 1 and/or a downstream Ser/Thr-rich region called "site 2" (Figure S2A; Chalabi et al., 2006). More recently, the biological importance of S332 and S334 in vivo was excluded based on the fertility of $ZP3^{-/-}$ mice expressing a ZP3 transgene where these residues are mutated, although alternative binding sites were not identified (Gahlay et al., 2010). How can these results be reconciled with the strong evidence for a role of O-linked carbohydrates in binding to sperm (Florman and Wassarman, 1985)?

The data presented in Figure 5 provide direct evidence in favor of the importance of ZP3 site 1 O-glycans in gamete interaction. At the same time, they allow evaluation of the relationship between the various ZP3 sites that have been implicated in sperm binding, by projecting them on top of the structure of cZP3. As shown in Figures 1A and 1B, the interdomain loop carrying T168 folds back onto itself, positioning site 2 next to site 1 on top of ZP-C (Figure 6A). On the other side of the β sandwich, disulfide C_{10}-C_{12} in the ZP-C subdomain, which partly overlaps with the exon 7/SCS region (Figure S2A), fastens the C-terminal region of mature ZP3 to helix F"G (Figure 2C) so that it bends toward the interdomain loop (Figures 1A and 1B). The resulting ~120° inversion in chain direction is necessary for inserting the EHP at the core of the ZP module, explaining why the C_{10}-C_{12} connectivity is invariant between ZP3 homologs (Figure S2B). At the same time, this has the effect of positioning the C-terminal half of the SCS on the same surface of the molecule as sites 1 and 2 (Figure 6A). Although this region and the CFCS that follows it are disordered in the electron density, the approximate positions of mZP3 S332 and S334 can be easily inferred because these residues would immediately follow P343, the last visible SCS residue in the cZP3 map. By revealing that site 1, site 2, and the SCS are all exposed within a restricted area on the same surface of ZP3, our structure suggests that any of them could in principle contribute to carbohydrate-mediated sperm binding, as long as it is modified with the correct type of sugar chain in either native ZP3 (sites 1 and/or 2) or recombinant ZP3 produced in embryonal carcinoma cells (SCS). As shown in Figures 6B and 6C, spatial clustering of the sites also immediately suggests how—regardless of glycosylation—the hypervariable SCS and very C-terminal part of mature ZP3 could affect the specificity of gamete interaction by modulating the recognition of sites 1 and 2. At the same time, the conformational flexibility of the C-terminal region of ZP3, which could be amplified by the

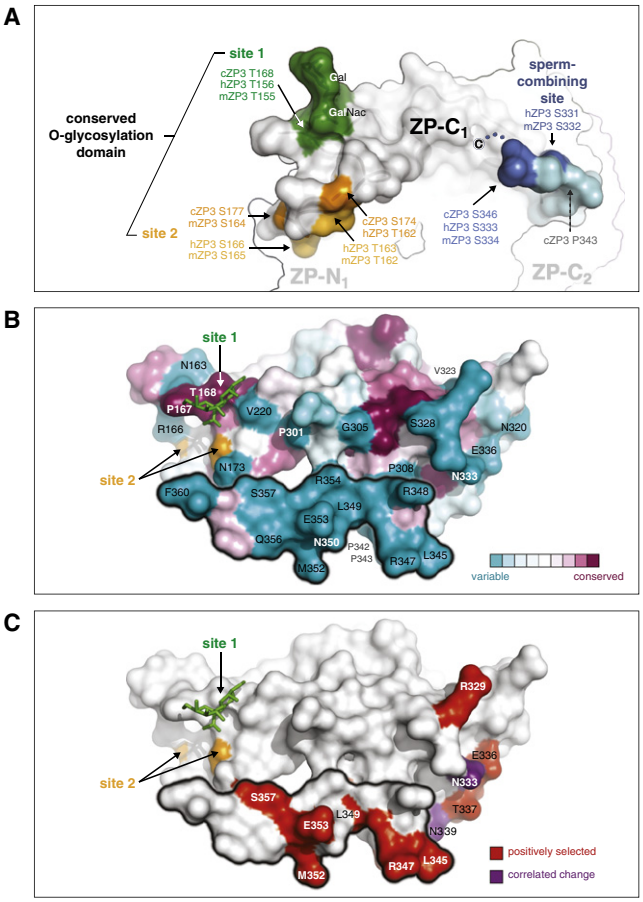

Figure 6. Conserved O-Glycosylation Sites are Clustered on the Same Protein Surface as Hypervariable, Positively Selected Regions of ZP3

(A) Conserved O-glycosylation sites 1 and 2 and the SCS are exposed on the same surface of ZP3.

(B) Top view of the ZP-C domain, colored according to amino acid conservation of ZP3 homologs from amphibian to human. Approximately 70% of the most variable residues in ZP-C are located in the depicted area. The figure includes a model of disordered C-terminal residues L345–F360 (black outline), which were added to the crystal structure and relaxed by molecular dynamics. Statistically significant variable residues, as well as invariant P167 and T168, are marked. cZP3 sites 1 and 2 are indicated, with the conserved site 1 O-glycan shown in stick representation.

(C) Mapping of positively selected sites (red) onto the model of mature cleaved ZP3, oriented as in (B). Two sites showing correlated changes are colored in violet.

aforementioned species-dependent variations in the local structure of the ZP-C subdomain, could clearly provide opportunities for protein-based recognition. This is consistent with the observation that sperm binding is highly reduced but not completely abolished in the T168A mutant (Figure 5D) and agrees with the growing evidence that gamete interaction probably relies on multiple distinct binding events (Shur, 2008). With relation to this point, it is interesting to notice how the conserved glycosylation sites of ZP3 are located within a linker region whose flexibility is probably important for ZP-N/ZP-C domain rearrangements during polymerization, and the orientation of the hypervariable C-terminal region results from the position of the EHP within the protein precursor. These structural considerations suggest how the sperm recognition function of ZP3 might have arisen during evolution as a specialization of its polymerization activity.

Regardless of which exact ZP3 epitope(s) are recognized by sperm in the mouse, the data by Gahlay et al. (2010) suggest that lack of sperm binding to the murine ZP following fertilization is not the result of ZP3 carbohydrate cleavage or modification but rather depends on proteolytic processing of ZP2. These results are still compatible with an important role of ZP3 carbohydrates in sperm binding, as ZP2 cleavage could act indirectly by causing structural rearrangements that ultimately shield the ZP3-binding surface identified by our structural and functional studies.

Downstream Transmission of Sperm Binding Information

Ordering of the O-glycosylated interdomain linker region, which is not involved in crystal contacts with symmetry-related molecules, has remarkable effects on underlying ZP-C domain residues (Figure 7A). In the ZP3 monomer where T168 is ordered, the conserved neighboring residue W169 stacks against the side chain of E180 and forms a short β sheet by inducing the formation of a B' β strand within the ZP-C BC loop. Consequently, an invariant residue in this loop, H219, flips inwards making hydrogen bonds with main chain carbonyl oxygens of S215 in strand B and V220 in strand B'. The presence of two different conformations within the crystal allows us to hypothesize how information about sperm binding might be transmitted through ZP3. It is possible that in the unbound protein the linker region around T168 is highly flexible. However, upon sperm binding this zone assumes a more ordered conformation that is stable (Figure 7B) and transmits a signal through the molecule as a result of H219 flipping. This may lead to stimulation of the acrosome reaction, a process that depends on the polypeptide moiety of ZP3 (Wassarman and Litscher, 2008; Shur, 2008). Alternatively, the conformational switch could be part of the structural changes of the ZP that take place during the block to polyspermy, and regulate the accessibility of the O-glycan before and after fertilization.

Relevance for Human Reproductive Medicine

Antibodies against ZP proteins, and in particular the C-terminal region of ZP3, have been shown to be powerful tools for inhibiting fertilization of domestic animals and wildlife, including primates (Kaul et al., 2001; Millar et al., 1989). However, variable efficiency and safety concerns suggest that immunocontraception is unlikely to become a feasible option for humans. At the same time, no completely novel contraceptive method has been introduced in the last 50 years to address the continuous growth of the world population (McLaughlin and Aitken, 2010). By allowing the development of small-molecule compounds that specifically target the sperm binding surface shown in Figure 6A, the structure of ZP3 could pave the way to the rational design of nonhormonal contraceptives. Moreover, structural information on the molecule will be essential for understanding ZP mutations linked to human infertility at the molecular level.

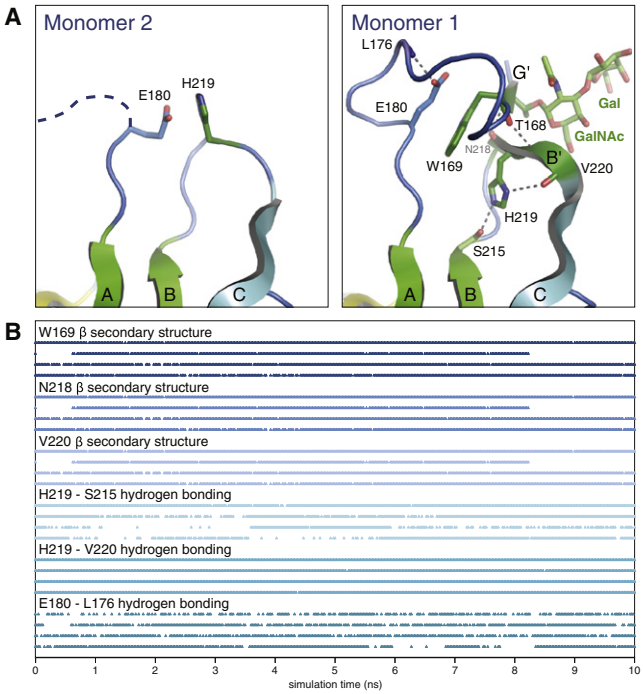

Figure 7. Alternative Conformations of the Conserved O-Linked Site Region

(A) The interdomain loop containing O-glycosylated T168 is disordered in monomer 2 (left) but adopts an ordered structure in monomer 1 by interacting with the BC loop of ZP-C (right).

(B) Key elements of the ordered loop conformation are stable during the course of independent 10 ns molecular dynamics simulations.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

Protocols used for DNA construct generation, protein expression in CHO cells, and protein purification are outlined in the Extended Experimental Procedures.

Protein and Carbohydrate Analysis

Methods used for immunoblot analysis, oligomeric state determination, cross-linking in solution, mass spectrometry, and lectin binding are described in the Extended Experimental Procedures.

Crystallization and Data Collection

Crystals of cZP3-4T (25 mg/ml) were grown in 0.1 M Na citrate (pH 5.0), 10 mM Tris-HCl (pH 8.0), 3%–13% PEG 6000, 50 mM NaCl (Figure S5A). They appeared in 1–5 days at 4°C and were cryoprotected by stepwise addition of PEG 6000 and PEG MME 550 to a final solution of 0.1 M Na citrate (pH 5.0), 10 mM Tris-HCl (pH 8.0), 6% PEG 6000, 30% PEG MME 550, 50 mM NaCl, after which they were flash frozen in liquid nitrogen. Datasets were collected at the European Synchrotron Radiation Facility (ESRF), Grenoble (Table S1). Details of structure determination and refinement, as well as structure analysis and molecular dynamics simulation, are provided in the Extended Experimental Procedures.

Sperm Binding Assays

Semen collected from 15 White Leghorn cocks was frozen in liquid nitrogen as described (Japanese patent No. 2942822). Sperm ($1.5 \times 10^4/\mu l$) were incubated with protein (5 ng/μl = 134 nM) in 20 mM Na-HEPES (pH 7.4), 150 mM NaCl at 37°C for 15 min. They were then fixed onto a glass slide with 3% paraformaldehyde, blocked with 2% BSA, and incubated with anti-5His (QIAGEN; 1:1,000), followed by Alexa Fluor-488 goat anti-mouse IgG (Invitrogen, 1:300). Imaging was performed on an Axioplan2 microscope equipped with LSM5 PASCAL laser scanning confocal optics (Zeiss) in multitrack mode. 488 nm excitation and 505–530 nm band-pass emission filters were used for imaging Alexa-Fluor 488. Stacks of 7–11 images taken at 0.5 μm intervals along the Z axis were merged, and signal intensities of the tip region of sperm heads were measured. Differential interference contrast images were taken by the same system. Analysis was performed with ImageJ (http://imagej.nih.gov/ij/), using a negative control-based integrated density cutoff of 10,000. t test statistical analysis was performed with InStat (GraphPad Software, Inc.). Animal procedures were approved by the Nagoya University Institutional Animal Care and Use Committee.

ACCESSION NUMBERS

Atomic coordinates and structure factors are deposited in the Protein Data Bank with accession codes 3NK3 (2.6 Å resolution) and 3NK4 (2.0 Å resolution).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at doi:10.1016/j.cell.2010.09.041.

ACKNOWLEDGMENTS

We thank CMC Biologics for expression plasmids pDEF38 and pNEF38; the ESRF for provision of synchrotron radiation facilities and Joanne McCarthy for assistance; Pavel Afonine, Ralf Grosse-Kunstleve, and Tom Terwilliger for help with PHENIX; Elmar Krieger and Alessandra Villa for help with molecular dynamics simulations; Hans Hebert for electron microscopy; Hisako Watanabe for help with sperm preparation; Franco Cotelli, Eveline Litscher, Rune Toftgård, and Paul Wassarman for discussions and comments. This work was supported by the Center for Biosciences; the Swedish Research Council (grants 2005-5102 and 2007-6068); the European Community (Marie Curie ERG 31055); the Scandinavia-Japan Sasakawa Foundation; Grant-in-aids from the Japan Society for the Promotion of Science and MEXT; and an EMBO Young Investigator award to L.J. Author Contributions: L.H. expressed proteins and analyzed mutants; M.M. generated constructs, purified and crystallized proteins, carried out model building, and refined the structures; H.O. generated and characterized constructs, expressed proteins, and performed and analyzed sperm binding assays; T.S. performed mass spectrometric analysis; A.L.C. analyzed crystallographic data; D.F. assisted data collection at ESRF; T.M. performed and analyzed sperm binding assays; L.J. directed the research, solved the structure, took part in structure refinement, ran molecular dynamics simulations, and wrote the manuscript with contributions from all other authors. L.J. dedicates this work to Marta, Smilla, and Sofia.

Received: June 23, 2010
Revised: August 11, 2010
Accepted: August 24, 2010
Published online: October 21, 2010

REFERENCES

Barratt, C.L.R., Andrews, P.A., McCann, C.T., Hornby, D.P., and Cooke, I.D. (1993). Recombinant human ZP3 expressed in Chinese hamster ovary cells (CHO) is a potent inducer of the acrosome reaction. Hum. Reprod. (8 Suppl.), Abstr. no. 407.

Bausek, N., Ruckenbauer, H.H., Pfeifer, S., Schneider, W.J., and Wohlrab, F. (2004). Interaction of sperm with purified native chicken ZP1 and ZPC proteins. Biol. Reprod. *71*, 684–690.

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.

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.

Boja, E.S., Hoodbhoy, T., Garfield, M., and Fales, H.M. (2005). Structural conservation of mouse and rat zona pellucida glycoproteins. Probing the native rat zona pellucida proteome by mass spectrometry. Biochemistry *44*, 16445–16460.

Bork, P., and Sander, C. (1992). A large domain common to sperm receptors (Zp2 and Zp3) and TGF-β type III receptor. FEBS Lett. *300*, 237–240.

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.

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.

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.

Darie, C.C., Biniossek, M.L., Jovine, L., Litscher, E.S., and Wassarman, P.M. (2004). Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. Biochemistry *43*, 7459–7478.

Fernandes, I., Chanut-Delalande, H., Ferrer, P., Latapie, Y., Waltzer, L., Affolter, M., Payre, F., and Plaza, S. (2010). Zona pellucida domain proteins remodel the apical compartment for localized cell shape changes. Dev. Cell *18*, 64–76.

Florman, H.M., and Wassarman, P.M. (1985). O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. Cell *41*, 313–324.

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.

Hasegawa, A., Kanazawa, N., Sawai, H., Komori, S., and Koyama, K. (2006). Pig zona pellucida 2 (pZP2) protein does not participate in zona pellucida formation in transgenic mice. Reproduction *132*, 455–464.

Jansa, S.A., Lundrigan, B.L., and Tucker, P.K. (2003). Tests for positive selection on immune and reproductive genes in closely related species of the murine genus *Mus*. J. Mol. Evol. *56*, 294–307.

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.

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.

Jovine, L., Darie, C.C., Litscher, E.S., and Wassarman, P.M. (2005). Zona pellucida domain proteins. Annu. Rev. Biochem. *74*, 83–114.

Kanai, S., Kitayama, T., Yonezawa, N., Sawano, Y., Tanokura, M., and Nakano, M. (2008). Disulfide linkage patterns of pig zona pellucida glycoproteins ZP3 and ZP4. Mol. Reprod. Dev. *75*, 847–856.

Kaul, R., Sivapurapu, N., Afzalpurkar, A., Srikanth, V., Govind, C.K., and Gupta, S.K. (2001). Immunocontraceptive potential of recombinant bonnet monkey (*Macaca radiata*) zona pellucida glycoprotein-C expressed in *Escherichia coli* and its corresponding synthetic peptide. Reprod. Biomed. Online *2*, 33–39.

Kinloch, R.A., Sakai, Y., and Wassarman, P.M. (1995). Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. Proc. Natl. Acad. Sci. USA *92*, 263–267.

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.

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.

Llorca, O., Trujillo, A., Blanco, F.J., and Bernabeu, C. (2007). Structural model of human endoglin, a transmembrane receptor responsible for hereditary hemorrhagic telangiectasia. J. Mol. Biol. *365*, 694–705.

McLaughlin, E.A., and Aitken, R.J. (2010). Is there a role for immunocontraception? Mol. Cell. Endocrinol. Published online April 20, 2010. 10.1016/j.mce.2010.04.004.

Millar, S.E., Chamow, S.M., Baur, A.W., Oliver, C., Robey, F., and Dean, J. (1989). Vaccination with a synthetic zona pellucida peptide produces long-term contraception in female mice. Science *246*, 935–938.

Monné, M., Han, L., and Jovine, L. (2006). Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. Semin. Reprod. Med. *24*, 204–216.

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.

Okumura, H., Aoki, N., Sato, C., Nadano, D., and Matsuda, T. (2007). Heterocomplex formation and cell-surface accumulation of hen's serum zona pellucida B1 (ZPB1) with ZPC expressed by a mammalian cell line (COS-7): a possible initiating step of egg-envelope matrix construction. Biol. Reprod. *76*, 9–18.

Pan, J., Sasanami, T., Nakajima, S., Kido, S., Doi, Y., and Mori, M. (2000). Characterization of progressive changes in ZPC of the vitelline membrane of quail oocyte following oviductal transport. Mol. Reprod. Dev. *55*, 175–181.

Quesada, V., Sanchez, L.M., Alvarez, J., and Lopez-Otin, C. (2004). Identification and characterization of human and mouse ovastacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. J. Biol. Chem. *279*, 26627–26634.

Rankin, T., Familari, 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.

Sanejouand, Y.H. (2004). Domain swapping of CD4 upon dimerization. Proteins *57*, 205–212.

Sasanami, T., Ohtsuki, M., Ishiguro, T., Matsushima, K., Hiyama, G., Kansaku, N., Doi, Y., and Mori, M. (2006). Zona Pellucida Domain of ZPB1 controls specific binding of ZPB1 and ZPC in Japanese quail (*Coturnix japonica*). Cells Tissues Organs *183*, 41–52.

Sasanami, T., Toriyama, M., and Mori, M. (2003). Carboxy-terminal proteolytic processing at a consensus furin cleavage site is a prerequisite event for quail ZPC secretion. Biol. Reprod. *68*, 1613–1619.

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.

Shur, B.D. (2008). Reassessing the role of protein-carbohydrate complementarity during sperm-egg interactions in the mouse. Int. J. Dev. Biol. *52*, 703–715.

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.

Swann, C.A., Cooper, S.J., and Breed, W.G. (2007). Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. Reproduction *133*, 697–708.

Swanson, W.J., Yang, Z., Wolfner, M.F., and Aquadro, C.F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc. Natl. Acad. Sci. USA *98*, 2509–2514.

Takeuchi, Y., Nishimura, K., Aoki, N., Adachi, T., Sato, C., Kitajima, K., and Matsuda, T. (1999). A 42-kDa glycoprotein from chicken egg-envelope, an avian homolog of the ZPC family glycoproteins in mammalian zona pellucida.

Its first identification, cDNA cloning and granulosa cell-specific expression. Eur. J. Biochem. *260*, 736–742.

Turner, L.M., and Hoekstra, H.E. (2006). Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). Mol. Biol. Evol. *23*, 1656–1669.

Vo, L.H., and Hedrick, J.L. (2000). Independent and hetero-oligomeric-dependent sperm binding to egg envelope glycoprotein ZPC in *Xenopus laevis*. Biol. Reprod. *62*, 766–774.

Waclawek, M., Foisner, R., Nimpf, J., and Schneider, W.J. (1998). The chicken homologue of zona pellucida protein-3 is synthesized by granulosa cells. Biol. Reprod. *59*, 1230–1239.

Wassarman, P.M., and Litscher, E.S. (2008). Mammalian fertilization: the egg's multifunctional zona pellucida. Int. J. Dev. Biol. *52*, 665–676.

Yasumasu, S., Kawaguchi, M., Ouchi, S., Sano, K., Murata, K., Sugiyama, H., Akama, 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.

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.

EXTENDED EXPERIMENTAL PROCEDURES

DNA Constructs

Construct cZP3-WT (cZP3 residues 1-437) was generated by PCR cloning of the full-length chicken ZP3 cDNA (accession number D89097; Takeuchi et al., 1999) into mammalian expression vector pSI (Promega) or Chinese Hamster Elongation Factor 1 (CHEF1) plasmids pDEF38 and pNEF38 (CMC Biologics; Running Deer and Allison, 2004). Subsequent constructs and mutations were obtained by overlap extension PCR or using a QuickChange Site-Directed Mutagenesis Kit (Stratagene). All constructs were confirmed by DNA sequencing.

Protein Expression

CHO-K1 cells (American Type Culture Collection) were cultured in F12 medium (Invitrogen) and transiently transfected as described (Jovine et al., 2002); cell lysate and conditioned medium were collected 48 hr after transfection. For stable expression, dihydrofolate reductase-deficient CHO DG44 cells (Invitrogen) grown in α-MEM medium supplemented with 10% fetal bovine serum (Invitrogen) were co-transfected with CHEF1 vectors carrying *cZP3* genes, and stably transfected clones were identified by double selection using α-MEM medium without hypoxanthine and thymidine, supplemented with 8% dialyzed fetal bovine serum and 0.8 mg/ml G418 (Invitrogen). Clones expressing high levels of C-terminally histidine-tagged cZP3 constructs were identified by immunoblot. For large-scale expression, stable cell lines were adapted to serum-free medium HyQ SFM4CHO-utility (HyClone) and cultivated for over 2 months in hollow fiber bioreactors with a 5 kDa MWCO cartridge (FiberCell Systems Inc.), harvesting 15 ml of protein concentrate per day. Alternatively, batch suspension cultures were run in 3-l spinner flasks for 2 weeks.

Protein Purification

Secreted cZP3 proteins were captured from conditioned medium by batch IMAC with Ni-NTA Superflow (QIAGEN). Subsequently, they were subjected to SEC using a HiLoad 26/60 Superdex 200-HiPrep 26/60 Sephacryl 200 double column system (GE Healthcare) equilibrated against GF buffer (10 mM Tris-HCl [pH 8.0] at 4°C, 50 mM NaCl), and concentrated to 4 mg/ml. Trypsinized cZP3 was generated by digesting purified protein with sequencing grade trypsin (Promega) at a weight ratio of 1,000:1 in 50 mM Na-HEPES (pH 7.4) for 2 hr at 37°C. The cleaved protein was then purified by cation exchange chromatography and SEC and concentrated to 25 mg/ml.

Immunoblot Analysis

Dot blot and Western blot experiments were performed using primary antibodies mouse monoclonal anti-5His (QIAGEN; 1:1,000), mouse polyclonal anti-cZP3 (Okumura et al., 2007; 1:5000) and rabbit polyclonal anti-mZP3 8818 (Pocono Rabbit Farm; 1:10,000). Secondary antibodies were peroxidase-conjugate AffiniPure goat anti-mouse or goat anti-rabbit IgG (Jackson ImmunoResearch Lab, Inc.; 1:10,000 and 1:5,000, respectively).

Oligomeric State Determination by Analytical SEC

cZP3 samples and protein markers of known structure (MW 13.7–232 kDa; GE Healthcare) were run on a Superdex 200 10/300 GL column (GE Healthcare) equilibrated with GF buffer. In order to take into account the non-globular shape of cZP3, the experimental sqrt(-log(K_{av})) values of the markers were plotted against the corresponding radii of gyration (Cutler, 2004), calculated using HYDROPRO (García De La Torre et al., 2000). A calibration curve was then obtained with MacCurveFit (Kevin Raner Software), resulting in a linear function that fit the data with correlation coefficient $R^2 = 0.9446$ and was used to calculate the radii of gyration of cZP3 species from their K_{av}.

Crosslinking in Solution

Crosslinking of cZP3-4 (3 mg/ml) was performed in 100 mM NaPi (pH 7.2) with 10 mM $BS(PEG)_9$ (Thermo Scientific) for 2 hr on ice.

Mass Spectrometry

Samples were analyzed by MALDI–TOF and MALDI–TOF/TOF (Ultraflex II and Autoflex III, Bruker Daltonics), and spectra annotated with the Flexanalysis software (Bruker Daltonics).

Lectin Binding

Ten micrograms of protein were incubated, in a total volume of 0.4 ml for 1 hr at room temperature, with 20 μl of a prewashed slurry of jacalin agarose beads (Vector Laboratories) in 175 mM Tris-HCl (pH 7.4), 50 mM NaCl, or PNA beads (Vector Laboratories) in 10 mM Na-HEPES (pH 7.4), 50 mM NaCl, 0.1 mM $CaCl_2$, 0.01 mM $MnSO_4$. Beads were washed 3 times with 0.5 ml of the corresponding buffers and subsequently captured on pre-equilibrated 0.2 μm Ultrafree-MC filters (Millipore) by centrifugation at 1,000 g. Finally, proteins bound to jacalin or PNA beads were eluted with 175 mM Tris-HCl (pH 7.4), 50 mM NaCl, 1 M galactose, or a solution of 10 mM Na-HEPES (pH 7.4), 50 mM NaCl, 0.1 mM $CaCl_2$, 0.01 mM $MnSO_4$, 1 M galactose adjusted to (pH 4.0) with CH_3COOH, respectively.

Structure Determination and Refinement

Data was processed with iMosflm (Leslie, 1992) and integrated in space group $P4_12_12$ using SCALA (Evans, 2006) and TRUNCATE (French and Wilson, 1978). The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) and model building was performed with PHENIX AutoBuild (Terwilliger et al., 2008) and Coot (Emsley et al., 2010). Refinement was carried out with phenix.refine (Afonine et al., 2005), using riding hydrogens and partial noncrystallographic symmetry restraints. Models were validated with MolProbity (Chen et al., 2010) and the Carbohydrate Structure Suite (Lütteke et al., 2005). σA-weighted $2F_{obs}$-F_{calc} and averaged kick omit maps were calculated with phenix.refine (Praznikar et al., 2009); composite annealed omit maps were calculated with CNS (Brünger, 2007), using a starting temperature of 4500 K to minimize model bias.

Structure Analysis

Secondary structure assignments were based on DSSPcont (Carter et al., 2003), β-Spider (Parisien and Major, 2005) and YASARA (Krieger et al., 2002). Evolutionary conservation was assessed using ConSurf (Ashkenazy et al., 2010), structural similarity was evaluated with Dali (Holm and Sander, 1995), GANSTA+ (Guerler and Knapp, 2008) and TM-score (Zhang and Skolnick, 2004), and internal symmetry was detected using SymD (Kim et al., 2010). Interfaces were examined with PISA (Krissinel and Henrick, 2007), PreBI (Tsuchiya et al., 2006a) and ClassPPI (Tsuchiya et al., 2006b), residue interactions were identified with PIC (Tina et al., 2007), hydrogen bonds were analyzed with HBPLUS (McDonald and Thornton, 1994) and YASARA (Krieger et al., 2002), and Poisson-Boltzmann electrostatic calculations were performed using APBS (Baker et al., 2001). Disulfide bond energy and possibility to form the alternative disulfide bond pattern of ZP3 were analyzed with SSBOND (Hazes and Dijkstra, 1988), MODIP (Dani et al., 2003) and Disulfide by Design (Dombkowski, 2003). Figures were made with PyMOL (Schrödinger, LLC; http://www.pymol.org/).

Molecular Dynamics Simulations

A homodimeric model of cZP3 residues A47-G393, including the T antigen epitope linked to T168, was generated from the refined crystallographic structures of cZP3-4T by automatic loop building and energy minimization in YASARA (Krieger et al., 2009); subsequently, a core $Man_3GlcNac_2$ N-glycan was added to the single N-glycosylation site N159 using GlycoProt (Bohne-Lang and von der Lieth, 2005), and the resulting structure was again energy minimized in YASARA using the YAMBER3 force field (Krieger et al., 2004). The model was put into a cubic simulation cell with a 15 Å extension on each side of the elongated protein, in order to allow the latter to rotate freely during long-term simulations without crossing periodic boundaries. The YAMBER3 force field was then used to run 10 ns molecular dynamics simulations in explicit solvent (0.9% NaCl) at pH 7.0, 311-312 K (within the body temperature range of the chicken; Kadono et al., 1981).

SUPPLEMENTAL REFERENCES

Afonine, P. V., Grosse-Kunstleve, R. W., and Adams, P. D. (2005). The Phenix refinement framework. CCP4 Newsletter *42*, contribution 8.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., and Ben-Tal, N. (2010). ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. *38* (*Suppl.*), W529–W533.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA *98*, 10037–10041.

Bohne-Lang, A., and von der Lieth, C.W. (2005). GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. *33*, W214–W219.

Brünger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat. Protoc. *2*, 2728–2733.

Carter, P., Andersen, C.A., and Rost, B. (2003). DSSPcont: Continuous secondary structure assignments for proteins. Nucleic Acids Res. *31*, 3293–3295.

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.

Cutler, P. (2004). Size-exclusion chromatography. Methods Mol. Biol. *244*, 239–252.

Dani, V.S., Ramakrishnan, C., and Varadarajan, R. (2003). MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins. Protein Eng. *16*, 187–193.

Dombkowski, A.A. (2003). Disulfide by Design: a computational method for the rational design of disulfide bonds in proteins. Bioinformatics *19*, 1852–1853.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, C. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. *62*, 72–82.

French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. Acta Crystallogr. A *34*, 517–525.

García De La Torre, J., Huertas, M.L., and Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. *78*, 719–730.

Gille, C., and Frömmel, C. (2001). STRAP: editor for STRuctural Alignments of Proteins. Bioinformatics *17*, 377–378.

Guerler, A., and Knapp, E.W. (2008). Novel protein folds and their nonsequential structural analogs. Protein Sci. *17*, 1374–1382.

Hazes, B., and Dijkstra, B.W. (1988). Model building of disulfide bonds in proteins with known three-dimensional structure. Protein Eng. *2*, 119–125.

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.

Kadono, H., Besch, E.L., and Usami, E. (1981). Body temperature, oviposition, and food intake in the hen during continuous light. J. Appl. Physiol. *51*, 1145–1149.

Kim, C., Basner, J., and Byungkook, L. (2010). Detecting internally symmetric protein structures. BMC Bioinformatics *11*, 303.

Krieger, E., Koraimann, G., and Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field. Proteins *47*, 393–402.

Krieger, E., Darden, T., Nabuurs, S.B., Finkelstein, A., and Vriend, G. (2004). Making optimal use of empirical energy functions: force-field parameterization in crystal space. Proteins *57*, 678–683.

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.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. *372*, 774–797.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography *26*.

Lütteke, T., Frank, M., and von der Lieth, C.W. (2005). Carbohydrate Structure Suite (CSS): analysis of carbohydrate 3D structures derived from the PDB. Nucleic Acids Res. *33*, D242–D246.

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. Cryst. *40*, 658–674.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.

Parisien, M., and Major, F. (2005). A new catalog of protein β-sheets. Proteins *61*, 545–558.

Praznikar, J., Afonine, P.V., Guncar, G., Adams, P.D., and Turk, D. (2009). Averaged kick maps: less noise, more signal… and probably less bias. Acta Crystallogr. D Biol. Crystallogr. *65*, 921–931.

Running Deer, J., and Allison, D.S. (2004). High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1α gene. Biotechnol. Prog. *20*, 880–889.

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.

Tina, K.G., Bhadra, R., and Srinivasan, N. (2007). PIC. Protein Interactions Calculator. Nucleic Acids Res. *35*, W473–W476.

Tsuchiya, Y., Kinoshita, K., Ito, N., and Nakamura, H. (2006a). PreBI: prediction of biological interfaces of proteins in crystals. Nucleic Acids Res. *34*, W320–W324.

Tsuchiya, Y., Kinoshita, K., and Nakamura, H. (2006b). Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity. Protein Eng. Des. Sel. *19*, 421–429.

Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. Proteins *57*, 702–710.

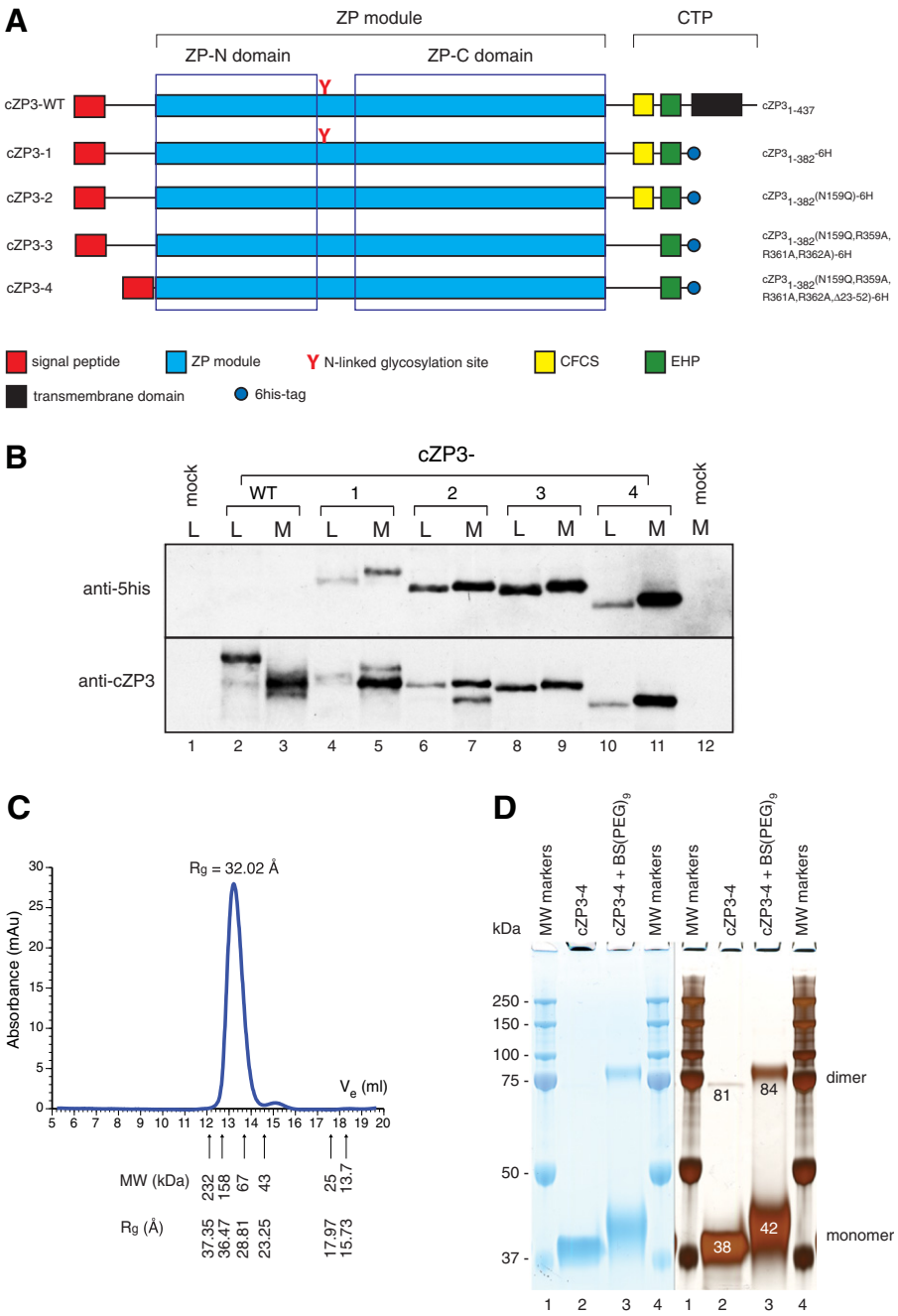

Figure S1. Construct Design and Determination of cZP3 Oligomeric State, Related to Figure 1, Figure 3, and Figure 4

(A) Knowledge of cZP3 biogenesis was used to optimize constructs for expression and crystallization. The signal peptide directs the nascent protein to the endoplasmic reticulum and the secretory pathway, and a C-terminal TM keeps the protein precursor anchored to the membrane. After glycosylation and disulfide bond formation in the endoplasmic reticulum, the C-terminal propeptide (CTP), containing the TM and the EHP that inhibits polymerization, is proteolytically processed at the CFCS. In the construct design, the TM was deleted and replaced with a C-terminal 6His-tag. The single N-glycosylation site was mutated and the CFCS was inactivated so that the protein would retain the EHP, thus inhibiting polymerization.

(B) Constructs were expressed in CHO cells and expression levels determined by immunoblot analysis of cell lysate (L) and conditioned media (M) using antibodies against the 6His-tag or cZP3. When using anti-cZP3 antibodies, the M fractions of cZP3-1 and cZP3-2 contained two discrete bands corresponding to full-length and CFCS-cleaved protein which lacks the C-terminal propeptide and 6His-tag.

(C) Purified cZP3-4 migrated slightly faster than a 67 kDa standard marker during SEC, consistent with a radius of gyration (R_g) of 32.02 Å. This is close to the value calculated from the homodimeric crystal structure (31.29 Å), indicating that cZP3-4 is also a dimer in solution.

(D) Coomassie blue and silver stained SDS-PAGE gels show an 84 kDa band after crosslinking of cZP3-4 in solution with $BS(PEG)_9$, equivalent to a dimeric species.

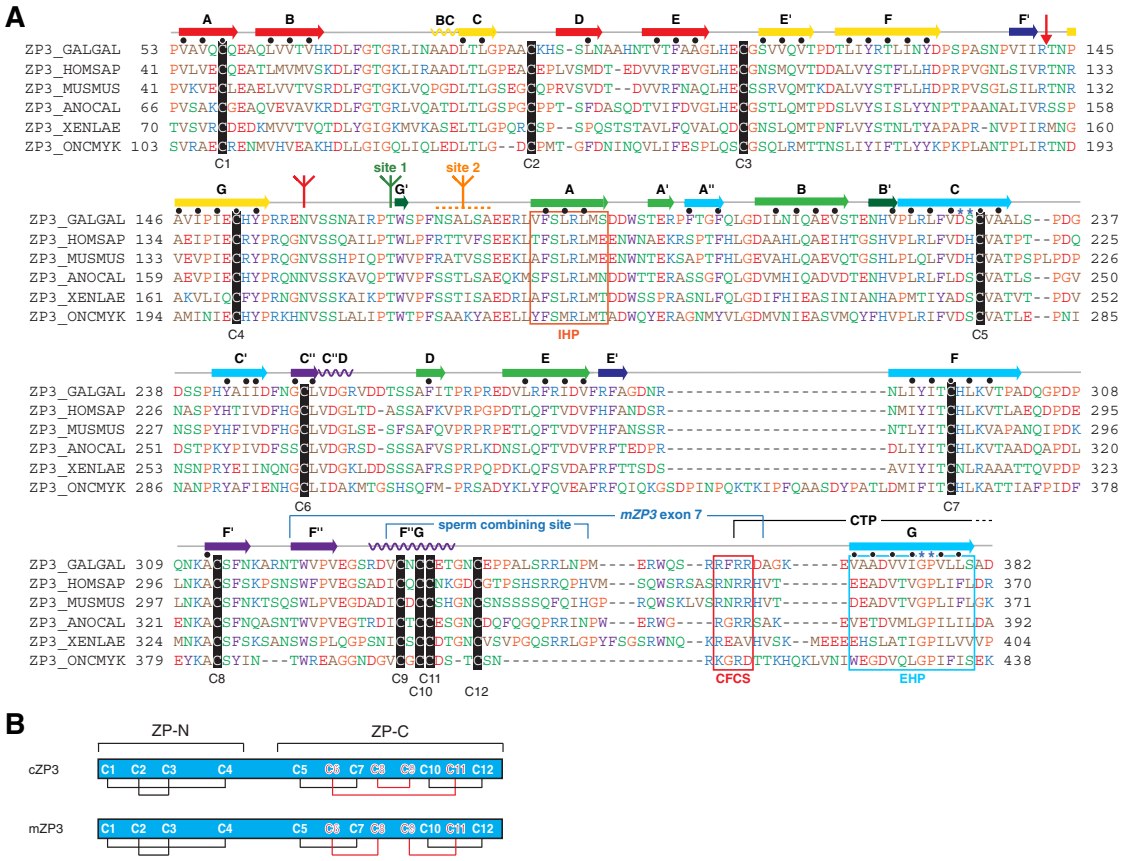

Figure S2. Structure-Based Sequence Alignment of ZP3 Orthologs, Related to Figure 1 and Figure 2

(A) ZP3 sequences from *Gallus gallus* (BAA13760), *Homo sapiens* (NP_001103824), *Mus musculus* (P10761), *Anolis carolinensis* (Lizard.scaffold_728), *Xenopus laevis* (Q91728), and *Oncorhynchus mykiss* (Q9I9M6) were aligned on the basis of the structure of cZP3 using STRAP (Gille and Frömmel, 2001). Residues are colored according to their properties: brown, hydrophobic; purple, aromatic; orange, [Pro, Gly]; white, Cys; green, polar; blue, positively charged; red, negatively charged. Secondary structure elements are colored as in Figure 1C, black dots indicate positions with side chains pointing toward the core of the structure, and blue stars indicate β bulges. IHP, EHP, and CFCS are boxed. Conserved N-linked and O-linked glycosylation sites 1 and 2 are indicated schematically in red, green, and orange respectively. A vertical red arrow marks the ZP3 site cleaved during embryo hatching in fish (Yasumasu et al., 2010).

(B) Schematic diagram of the two alternative disulfide bond patterns of ZP3 (Kanai et al., 2008), exemplified by the chicken and mouse proteins. Invariant bonds are black and varying bonds are red.

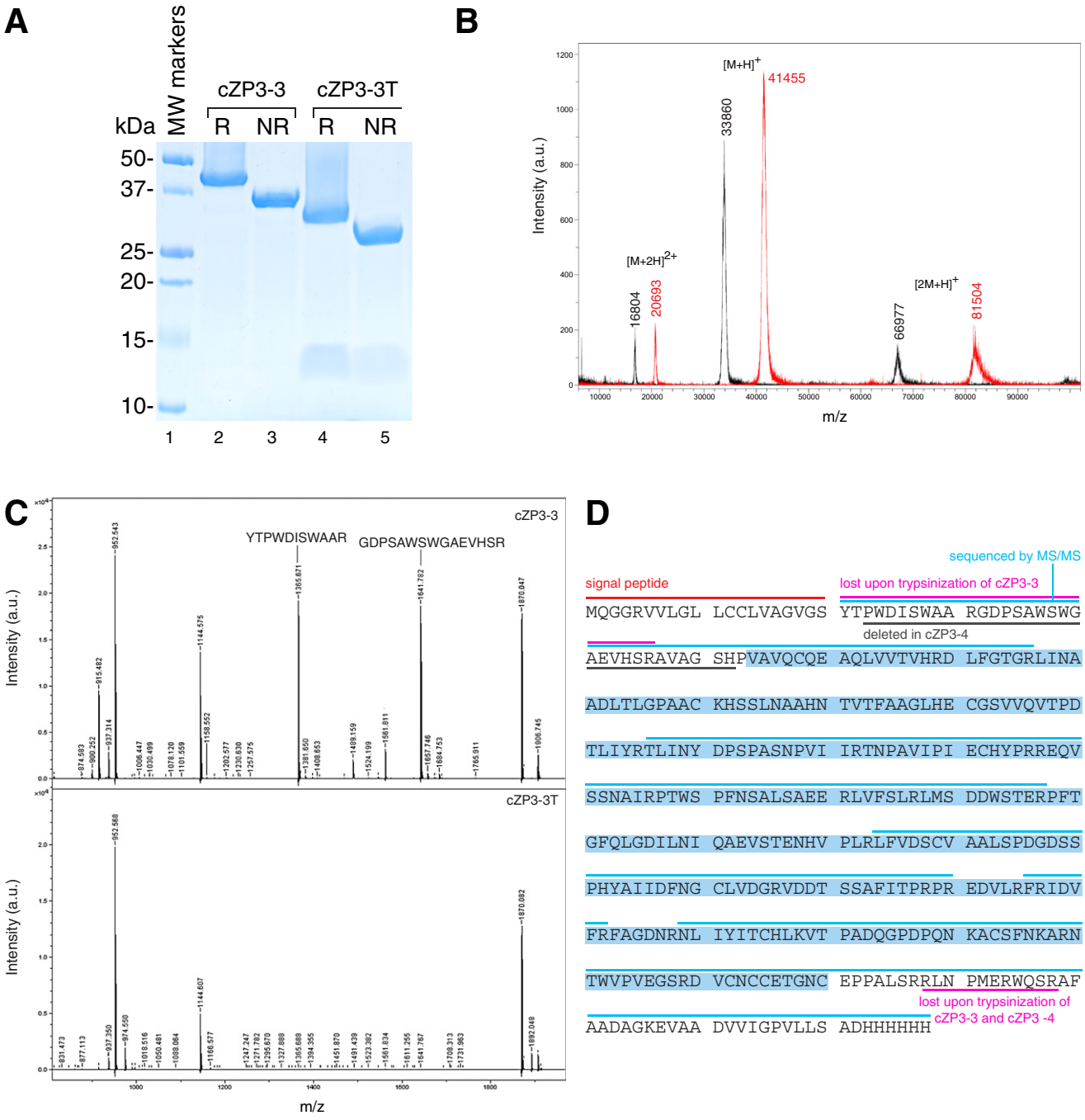

Figure S3. Biochemical and Mass Spectrometric Analysis of cZP3-3 and cZP3-3T, Related to Figure 2A and Figure 4

(A) cZP3-3 treated with trypsin (cZP3-3T; lanes 4-5) migrates faster during SDS-PAGE than the untreated protein (lanes 2 and 3). R, reducing conditions; NR, nonreducing conditions.

(B) Mass spectrometry (MS) analysis of cZP3-3 (red) and cZP3-3T (black) shows a decrease in mass following trypsinization.

(C) Comparison of MS spectra for cZP3-3 and cZP3-3T identified an N-terminal peptide (Y21-R46) that is no longer present in the trypsin-treated sample.

(D) Summary of MS analysis of the cZP3-3 and cZP3-4 constructs. The blue line indicates sequence coverage by MS analysis. Magenta lines indicate peptides lost by trypsinization of cZP3-3 and cZP3-4 and the black line indicates the sequence removed in the design of the cZP3-4 construct. Signal peptide is indicated with a red line and the ZP module is highlighted in blue.

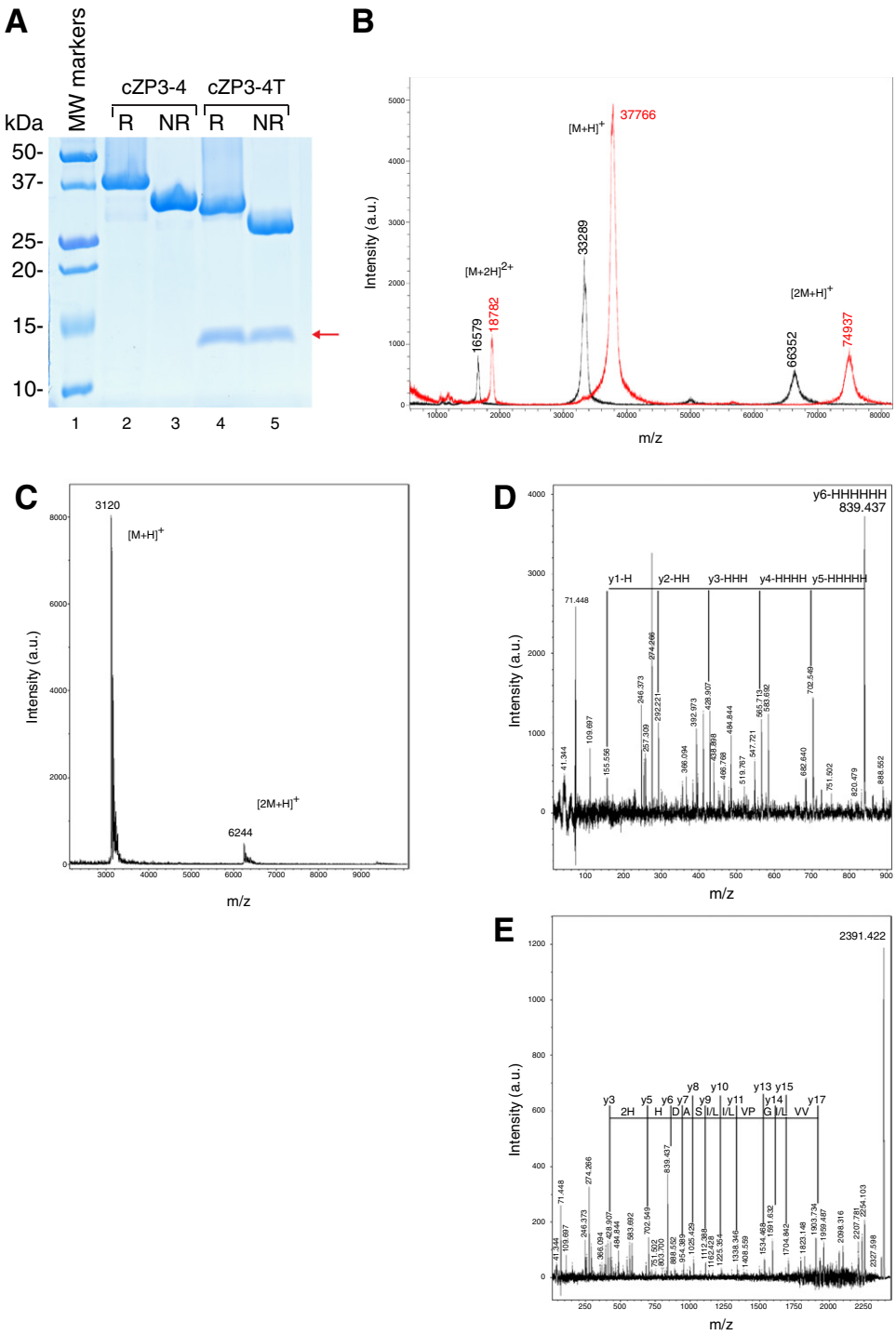

Figure S4. Biochemical and Mass Spectrometric Analysis of cZP3-4 and cZP3-4T, Related to Figure 2A and Figure 4

(A) cZP3-4 treated with trypsin (cZP3-4T; lanes 4 and 5) has a faster migration than the untreated protein (lanes 2 and 3) during SDS-PAGE.

(B) MS analysis of cZP3-4 (red) and cZP3-4T (black) shows a mass decrease following trypsinization.

(C) In the low molecular weight range, a cZP3-4T peak of m/z 3120 corresponds to the C-terminal peptide AFAADAGKEVAADVVIGPVLLSADHHHHHH, including the EHP and 6His-tag. This peptide migrated as an ~14 kDa species in SDS-PAGE (indicated by a red arrow in A), probably due to SDS-micelle aggregates.

(D and E) The peptide from (C) was further digested with trypsin and identified as the very C-terminal peptide EVAADVV(I/L)GPV(I/L)(I/L)SADHHHHHH of m/z 2391.

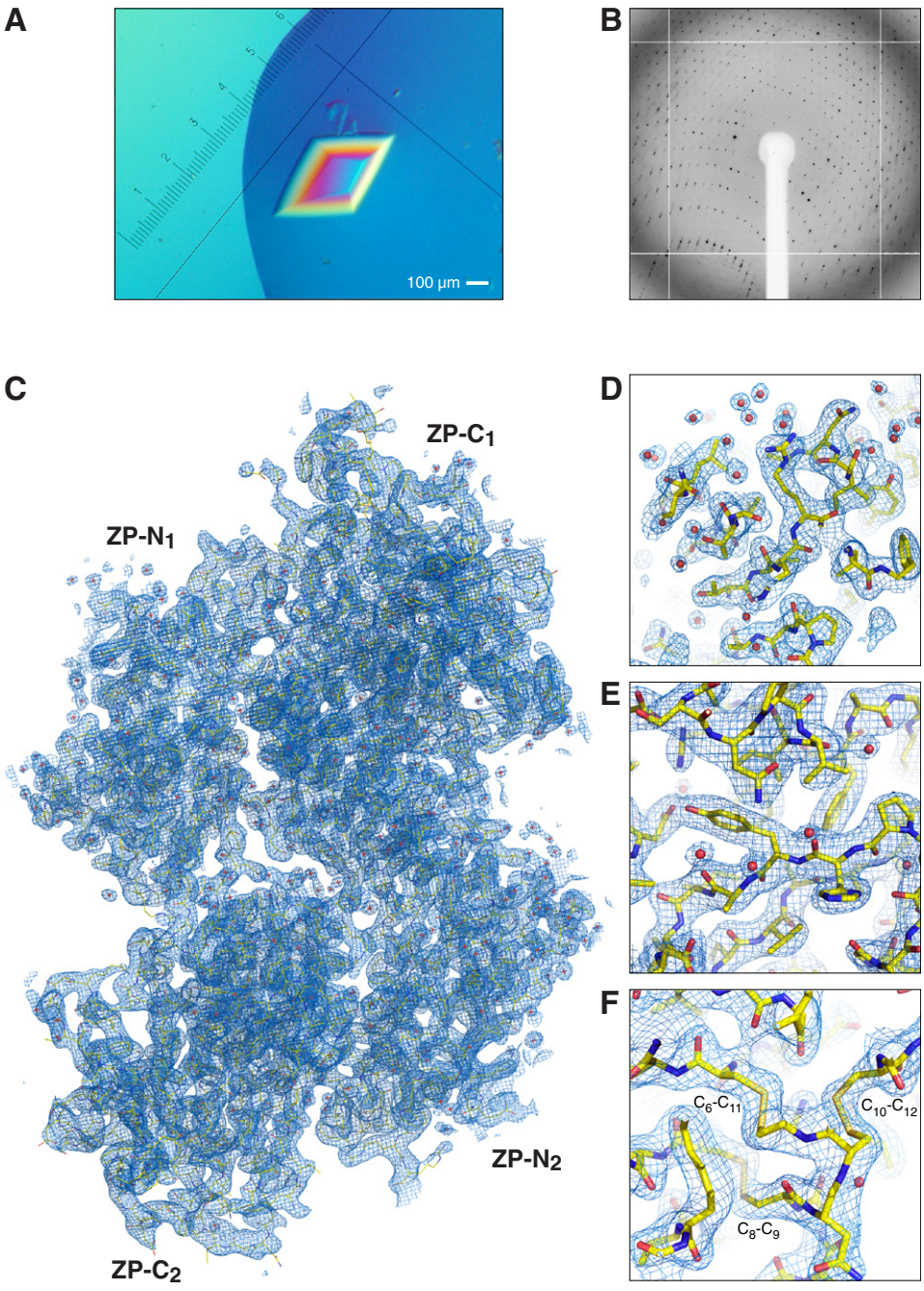

Figure S5. X-Ray Crystallography of cZP3-4T, Related to Figure 1 and Table S1

(A) Crystal of cZP3-4T grown in a hanging drop at a protein concentration of 25 mg/ml in 0.1 M Na citrate (pH 5.0), 10 mM Tris-HCl (pH 8.0), 4% PEG 6000 and 50 mM NaCl at 4°C. This is the actual crystal that yielded the 2.6 Å resolution native 1 dataset (PDB ID 3NK3).

(B) Diffraction image from the 2.0 Å resolution native 2 dataset (PDB ID 3NK4).

(C) Refined $2F_{obs}$-F_{calc} electron density map of the whole cZP3 homodimer, contoured at 1 σ. The molecule is oriented as in Figure 1A.

(D–F) Details of the electron density map in (C), showing different regions of cZP3. (D) ZP-N domain; (E) ZP-C domain; (F) ZP-C subdomain disulfide bonds.

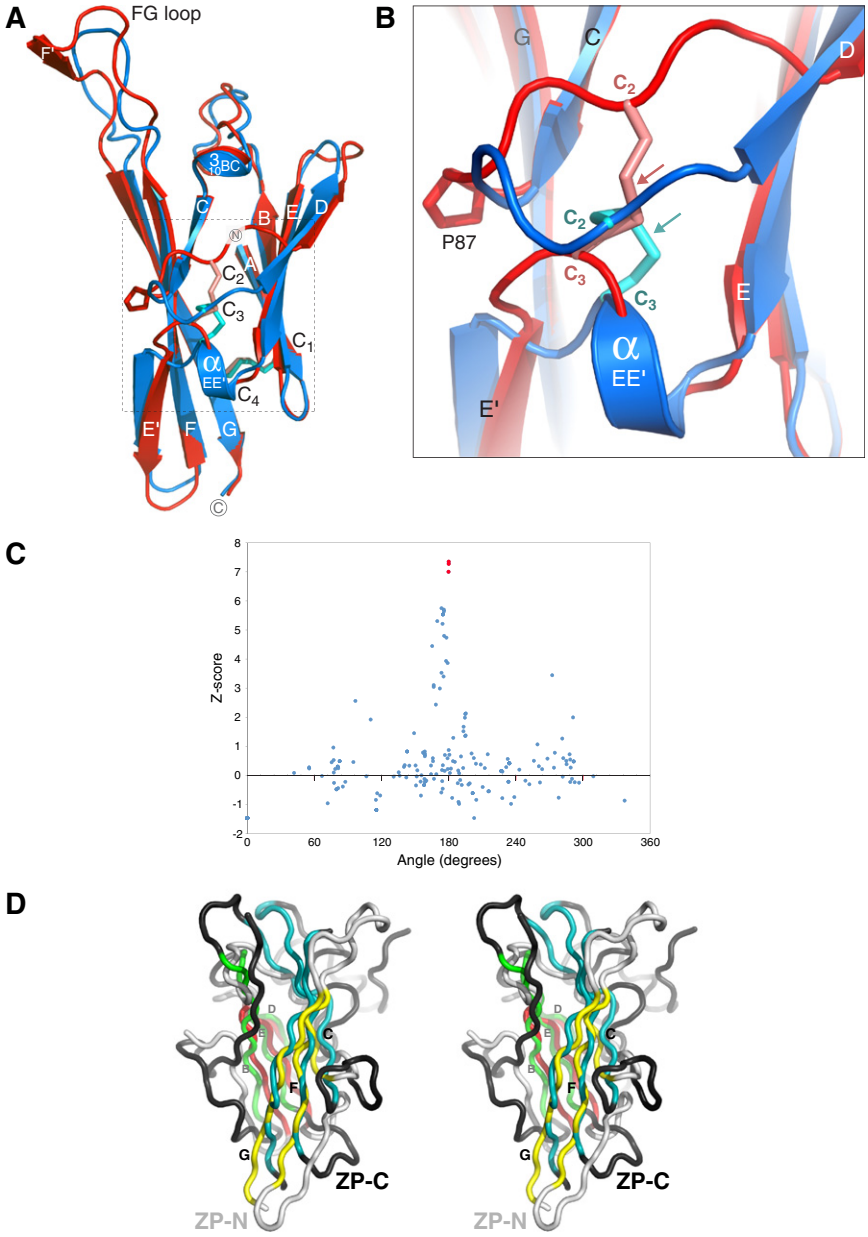

Figure S6. Comparison of mZP3 and cZP3 ZP-N Domains, and of the ZP-N and ZP-C Domains of cZP3, Related to Figure 1 and Figure 4A

(A) Superposition of the structures of mZP3 ZP-N (Monné et al., 2008; blue) and cZP3 ZP-N (red). The structural alignment has a TM-score of 0.8 over 89 C_α, with the core β sandwiches superimposing with a C_α root-mean-square distance of 0.9 Å. The major differences are found in the FG loop, where cZP3 ZP-N forms a β strand F' that interacts with the ZP-C domain of the other monomer at the dimer interface, and the loops around the invariant C_2-C_3 disulfide.

(B) Details of the C_2-C_3 disulfide. Whereas in isolated mZP3 ZP-N the C_2-C_3 bond has a right-handed hook conformation (C_α-C_α distance 4.4 Å; cyan arrow), in cZP3 ZP-N it adopts an unusual left-handed conformation (C_α-C_α distance 6.1 Å; red arrow) due to stretching of the CD loop as a result of P87-mediated interactions with the ZP-C domain.

(C) SymD analysis of cZP3 ZP module internal symmetry. High-scoring data points are indicated by red dots.

(D) Stereo view of a superposition of cZP3 ZP-N and ZP-C based on internal symmetry (TM-score over 69 C_α = 0.6). The B/E/D and C/F/G β sheets of ZP-N (red and yellow strands) and ZP-C (green and cyan strands) are superimposed; loops not used for superposition are shown in gray for ZP-N and black for ZP-C.

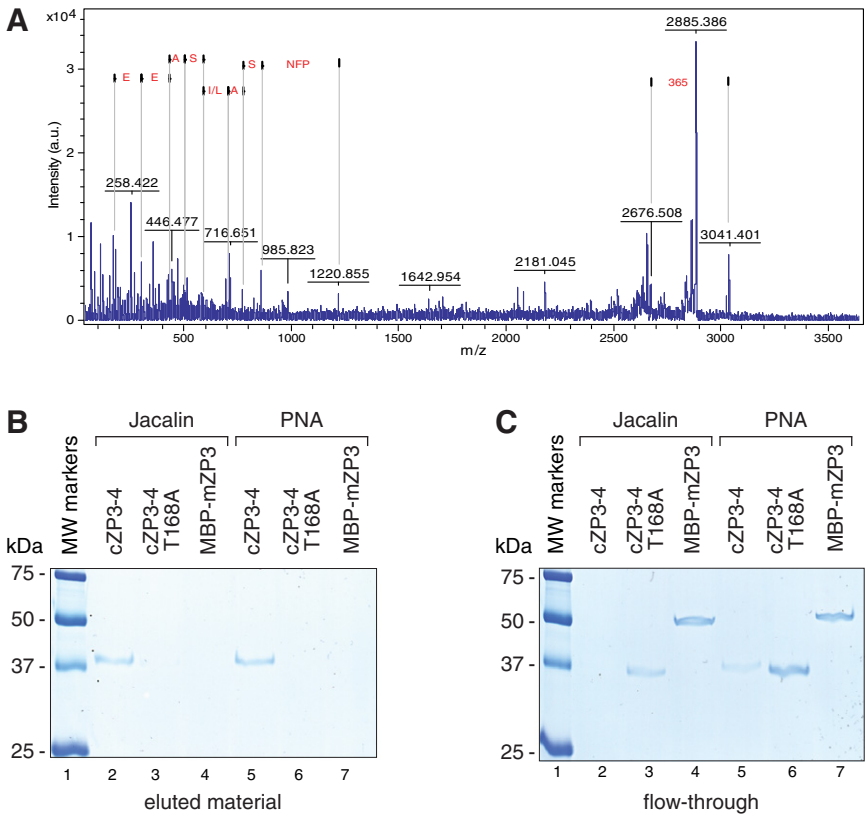

Figure S7. O-Linked Glycosylation of T168, Related to Figure 5

(A) Electrospray ionization mass spectrometry analysis shows that peptide EQVSSNAIRPTWSPFNSALSAEER has a monoisotopic mass of 3041. Fragmentation of the peptide results in a neutral loss of 365 Da (m/z 2676), indicating the loss of a Hex-HexNAc.

(B and C) Lectin-binding analysis. cZP3-4, cZP3-4 T168A, and the negative control maltose-binding protein (MBP)-mZP3$_{42\text{-}143}$ (Monné et al., 2008) were tested for binding to jacalin that recognizes the T antigen (Galβ1-3GalNAc) with or without sialic acid modification and peanut agglutinin (PNA) that recognizes the T antigen devoid of sialic acid. Neither cZP3-4 T168A nor MBP-mZP3$_{42\text{-}143}$ bound to either jacalin or PNA, whereas cZP3-4 bound to both lectins. Notably, a small portion of cZP3-4 was also found in the unbound fraction for PNA, which could indicate that a small fraction of the O-glycan attached to T168 is modified with sialic acid.

Table S1. Data Collection and Refinement Statistics, Related to Figure 1 and Figure S5

	Native 1 (PDB ID 3NK3)	Native 2 (PDB ID 3NK4)
Data Collection [a]		
Beamline	ESRF ID23-2	ESRF ID23-1
Wavelength (Å)	0.8726	0.9757
Space group	$P4_12_12$ (92)	$P4_12_12$ (92)
Cell dimensions		
a, *b*, *c* (Å)	98.385, 98.385, 257.369	97.700, 97.700, 256.419
α, β, γ (°)	90, 90, 90	90, 90, 90
Mosaicity (°)	0.52	0.28
Resolution range (Å)	61.20-2.60 (2.74-2.60)	38.90-2.00 (2.11-2.00)
Number of reflections (total/unique)	236350/37990 (32171/5090)	1125760/84686 (169301/12166)
Completeness (%)	95.4 (89.5)	99.9 (100.0)
Redundancy	6.2 (6.3)	13.3 (13.9)
Mean (I/σ(I))	10.8 (2.0)	12.6 (2.0)
$R_{p.i.m.}$ (%) [b]	5.9 (46.6)	2.7 (38.7)
Refinement [a]		
Resolution (Å)	48.32-2.60 (2.74-2.60)	36.34-2.00 (2.08-2.00)
Number of reflections (work/test)	35903/1903 (4680/238)	80278/4209 (8796/456)
R_{work}/R_{free}	21.9/24.9 (31.5/36.1)	20.8/22.6 (38.5/40.4)
Number of atoms		
Protein	4555	4493
Ligand	13	26
Water	127	208
Mean B-factors (Å^2)		
Protein	68.0	71.2
Ligand	73.1	94.9
Water	55.8	62.3
RMS deviations		
Bond lengths (Å)	0.011	0.010
Bond angles (°)	1.123	1.224
Chirality (Å^3)	0.062	0.079
Planarity (Å)	0.007	0.007
Dihedral angle (°)	16.592	17.388
ML estimate for coordinate error (Å)	0.45	0.26
Ramachandran plot (%) [c]		
Favored	98.4	98.4
Allowed	1.6	1.6
Outlier	0.0	0.0

[a] Values in parenthesis correspond to the highest resolution shell.

[b] $R_{p.i.m.} = \sum_{\mathbf{h}} \sqrt{\frac{1}{n_h - 1}} \sum_{l} \left| I_{hl} - \langle I_{\mathbf{h}} \rangle \right| / \sum_{\mathbf{h}} \sum_{l} \langle I_{\mathbf{h}} \rangle$

[c] Values are according to the program MolProbity (Chen et al., 2010).