Structural Conservation of Mouse and Rat Zona Pellucida Glycoproteins. Probing the Native Rat Zona Pellucida Proteome by Mass Spectrometry†

Emily S. Boja,‡,§ Tanya Hoodbhoy,§,|| Mark Garfield,⊥ and Henry M. Fales*,‡

Laboratory of Applied Mass Spectrometry, National Heart, Lung, and Blood Institute, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, and Research Technologies Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892

Received September 15, 2005; Revised Manuscript Received October 21, 2005

ABSTRACT: The mammalian zona pellucida is an egg extracellular matrix to which sperm bind. Mouse zonae are composed of three glycoproteins (ZP1, ZP2, and ZP3), while rat zonae contain four (ZP1, ZP2, ZP3, and ZP4/ZPB). Mouse sperm bind to zonae comprised solely of mouse ZP2 and ZP3. In this report, we show that rat sperm also bind to these zonae, indicating that ZP2 and ZP3 contain a "minimum structure(s)" to which rodent sperm can bind, and ZP1 and ZP4/ZPB are dispensable in these two rodents. These data are consistent with our mass spectrometric analysis of the native rat zona pellucida proteome (defined as the fraction of the total rat proteome to which the zonae glycoproteins contribute) demonstrating that the rat zonae glycoproteins share a high degree of conservation of structural features with respect to their mouse counterparts. The primary sequences of the rat zonae proteins have been deduced from cDNA. Each zona protein undergoes extensive co- and post-translational modification prior to its secretion and incorporation into an extracellular zona matrix. Each has a predicted N-terminal signal peptide that is cleaved off once protein translation begins and an anchoring C-terminal transmembrane domain from which the mature protein is released. Mass spectrometric analysis with a limited amount of native material allowed determination of the mature N-termini of rat ZP1 and ZP3, both of which are characterized by cyclization of glutamine to pyroglutamate; the N-terminus of ZP2 was identified by Edman degradation. The mature C-termini of ZP1 and ZP3 end two amino acids upstream of a conserved dibasic residue that is part of, but distinct from, the consensus furin cleavage sequence, while the C-terminus of ZP2 was not determined. Each zona protein contains a "zona domain" with eight conserved cysteine residues that is thought to play a role in the polymerization of the zona proteins into matrix filaments. Partial disulfide bond assignment indicates that the intramolecular disulfide patterns in rat ZP1, ZP2, and ZP3 are identical to those of their corresponding mouse counterparts. Last, nearly all potential N-glycosylation sites are occupied in the rat zonae glycoproteins (three of three for ZP1, six or seven of seven for ZP2, and four or five of six for ZP3). In comparison, potential O-glycosylation sites are numerous (59−83 Ser/Thr residues), but only two regions were observed to carry *O*-glycans in rat ZP3.

The mammalian zona pellucida is an extracellular matrix to which sperm bind with a certain degree of taxon specificity prior to fertilization. Fertilization triggers the exocytosis of cortically positioned secretory vesicles in the egg called cortical granules. The cortical granule contents, largely unidentified and uncharacterized in mammals, transform the zona into a formidable barrier to polyspermic egg penetration. In addition, the zona pellucida protects the pre-implantation embryo as it is transported down the oviduct to the uterus.

The glycoproteins comprising the zona pellucida are synthesized and secreted by oocytes during a 2−3 week growth period (*1*). The individual proteins that make up the mouse zona pellucida (ZP1,[1] ZP2, and ZP3) have been extensively characterized. Their primary sequences have been deduced from cDNA and consist of 623 (ZP1), 713 (ZP2), and 424 (ZP3) amino acids (*2*−*4*). The nascent polypeptide chains undergo extensive co- and post-translational modifications resulting in the secretion of mature glycoproteins that assemble into an extracellular layer to which sperm bind (*5*−*7*). Three homologous proteins (ZP1−ZP3) are found in rat zonae, and their nascent polypeptide chains are 617 (ZP1), 695 (ZP2), and 424 (ZP3) amino acids in length (*8*). Similar to human zonae (*9*), rat zonae also contain a fourth glycoprotein, ZP4/ZPB (*10*), the primary sequence of which, deduced from cDNA, is 545 amino acids in length.

The molecular basis of sperm−zona binding remains controversial (*11*). The mouse has provided an important

† This research was supported by the Intramural Research Program of the NIH, NHLBI, and NIDDK.

* To whom correspondence should be addressed: Laboratory of Applied Mass Spectrometry, NHLBI, NIH, 50 South Dr., Rm 3120, Bethesda, MD 20892. Telephone: (301) 496-2135. Fax: (301) 402-3404. E-mail: hmfales@helix.nih.gov.

‡ National Heart, Lung, and Blood Institute.

§ These authors contributed equally to this work.

|| National Institute of Diabetes and Digestive and Kidney Diseases.

⊥ National Institute of Allergy and Infectious Diseases.

[1] Abbreviations: ZP, zona pellucida; PNGase F, peptide *N*-glycosidase F; Gal, galactose; GalNAc, *N*-acetylgalactosamine; MS, mass spectrometry; CID, collision-induced dissociation; DTT, dithiothreitol; IAA, iodoacetamide.

10.1021/bi051883f
Published on Web 11/24/2005

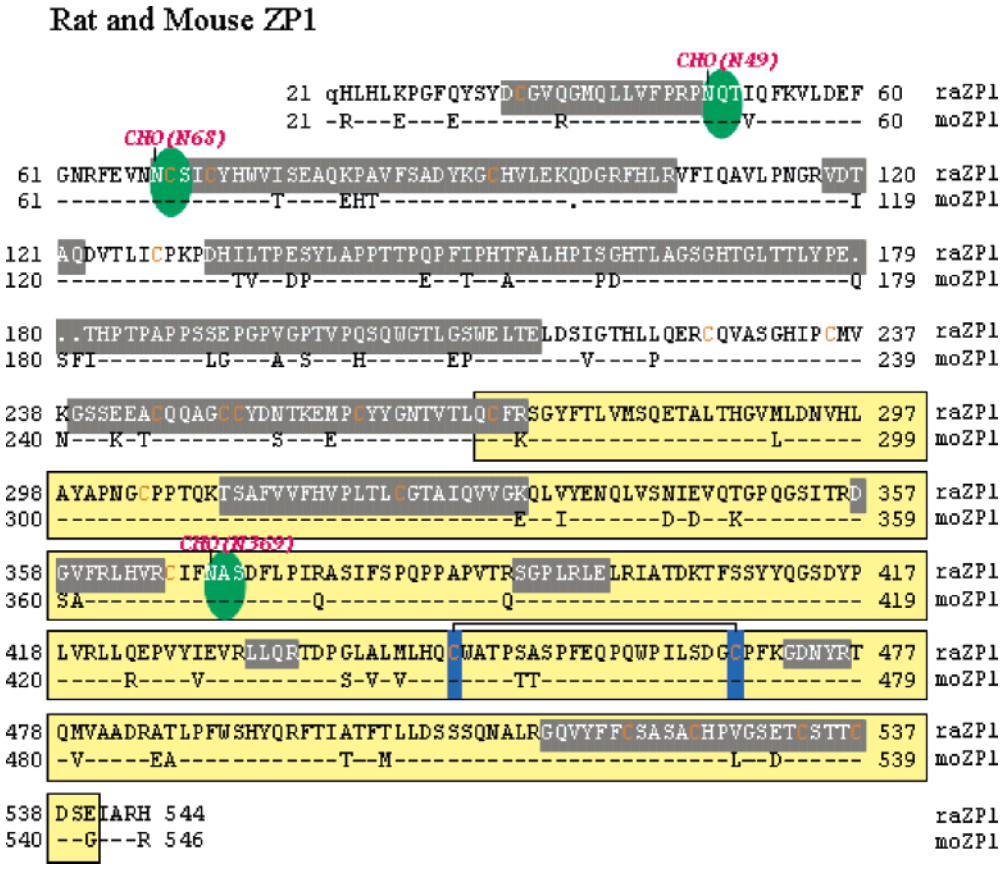

FIGURE 1: Summary of rat and mouse ZP1. The primary amino acid sequence (single-letter code) of mature secreted ZP1 extends from an N-terminal pyroglutamate (q=pyrQ21) to a C-terminal arginine (R544) immediately upstream of a conserved dibasic motif. There are 21 cysteine residues (orange); 10 are in the zona domain (yellow background), of which eight are conserved (C270, C304, C323, C366, C447, C468, C520, and C525). One disulfide bond was experimentally determined, C447−C468 (solid line with disulfide-linked cysteines in orange on a blue background). All three of the potential N-linked glycan sites (white on a green background) were glycosylated (N49, N68, and N369). Peptides representing ~48% of mature ZP1 were not identified (white on a gray background) because of a paucity of biological material. Within these sequences were multiple serine (S) or threonine (T) residues representing possible O-linked glycosylation sites.

model for investigating the mechanisms underlying gamete interactions. Its adequacy as a universal model of mammalian fertilization has recently been called into question given that mouse zonae contain only three zonae glycoproteins, while rat and human zonae contain four. The primary sequences of mouse and rat ZP1, ZP2, and ZP3 are highly conserved with 88% of the amino acids identical between ZP1 and ZP2 and 92% for ZP1 and ZP3. Mouse sperm bind promiscuously to zonae of almost all species examined to date, including to those of rat (*12*), and rat sperm bind to mouse zonae (*10*), despite the fact that mouse zonae are comprised of only three glycoproteins, compared to four in the rat. Mouse sperm, in fact, bind to zonae made up solely of mouse ZP2 and ZP3 (*13*). We seek here to define the zona structure required to support rodent sperm binding and looked for conservation of structural features between rat zonae glycoproteins and their cognate mouse counterparts.

MATERIALS AND METHODS

Proteomic Analysis of Rat Zonae Proteins. Vendors from which reagents were purchased and protocols used to isolate and process rat zonae pellucidae for subsequent mass spectrometric analyses have been described (*10*). Briefly, zonae were isolated from ovarian homogenates by density gradient ultracentrifugation. Rat ovaries contain far more connective tissue than mouse ovaries, resulting in less efficient homogenization, and poorer yields of isolated zonae following ultracentrifugation. A mixture containing solubilized native rat zonae proteins was reduced with dithiothreitol (DTT) and alkylated with iodoacetamide (IAA), prior to deglycosylation (*N*- or *N*/*O*-glycans) and proteolysis. Zonae were digested with trypsin, Asp-N, trypsin and Asp-N, or Glu-C. Nonreduced samples were similarly treated for identification of intramolecular disulfide bonds. Careful precautions were taken to minimize disulfide exchange by using an ammonium bicarbonate buffer at pH 7.2 as previously described for mouse zona preparation. After trypsin digestion, the sample was immediately acidified with formic acid before MS analysis.

In many cases, only enough zonae were obtained by ultracentrifugation for a single LC−MS/MS run and the Glu-C digestion was performed on only the N-deglycosylated sample. Lyophilized samples were reconstituted in 5 μL of 0.1% formic acid in 100% water before injection into an Agilent (Manchester, U.K.) HP1100 CapLC upfront column from a QTOF Ultima Global system. Data-dependent analysis (DDA) was performed on the top three most abundant multiply charged precursor ions ($\geq$+2) detected in the MS scan (*m*/*z* 300−1990; 1.0 s per scan). MS/MS scans (*m*/*z* 50−1990) were switched back to MS according to (1) the TIC threshold rising above 3000 counts or (2) an elapsed time of no more than 6 s (2.0 s per MS/MS scan). Operation

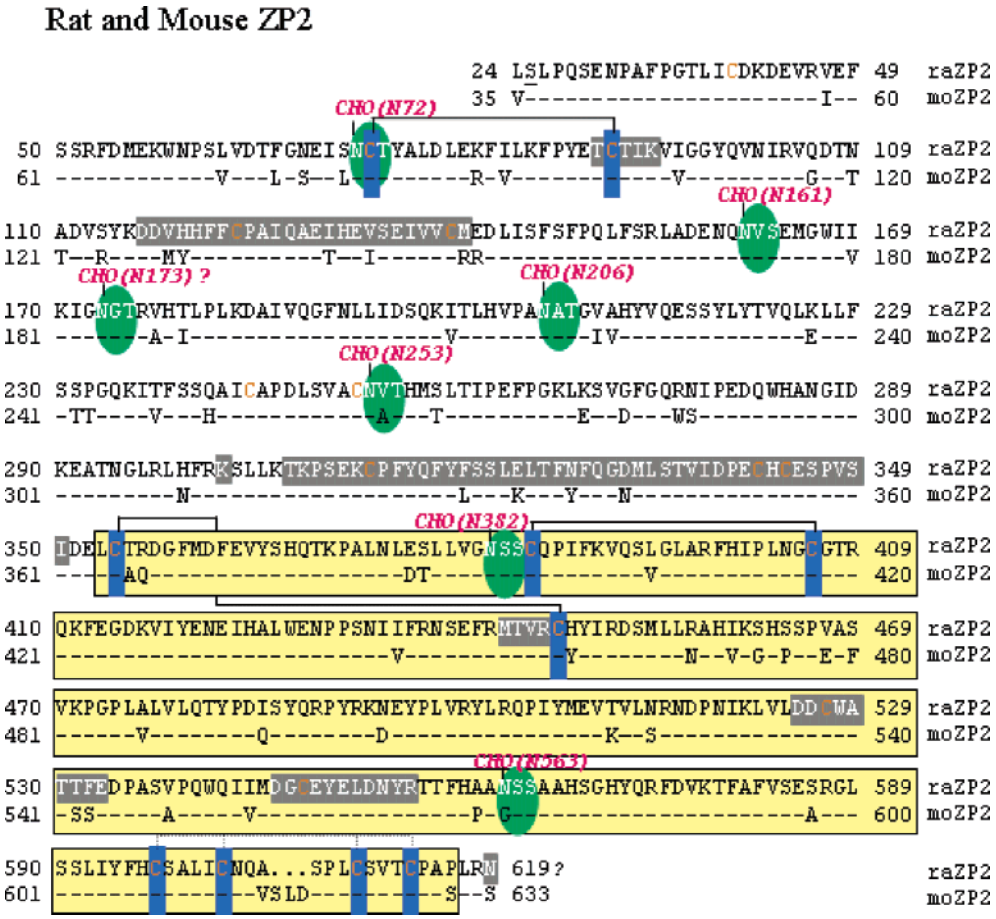

FIGURE 2: Summary of rat and mouse ZP2. The mature N-terminus of rat ZP2 begins at S25. The mature C-terminus was not found but extends until at least R618. There are 20 cysteine residues; 10 are in the zona domain, of which eight are conserved (C354, C385, C406, C447, C527, C597, C602, and C613). Two pairs of disulfide bonds in the zona domain were experimentally ascertained: C354−C447 and C385−C406. Four additional disulfide-bonded cysteines (C597, C602, C609, and C613) were found in a single tryptic peptide, but their precise linkage could not be determined because of the absence of appropriate proteolytic cleavage sites (dashed lines). Among the 10 cysteine residues in the N-terminus of ZP2, the disulfide linkage of one (C73−C91) was determined. Six or seven of seven potential N-linked glycan sites are glycosylated (N72, N161, N173, N206, N253, N382, and N563). Peptides representing ~18% of mature ZP2 were not identified. Color indications are the same as in Figure 1.

in this mode is no guarantee that the desired MS/MS scans will be obtained, particularly if there are other abundant extraneous ions in the region. Manual inspection of the raw data was always performed along with database searching.

Edman Microsequencing of Rat ZP2. Zonae isolated from 15 rat ovaries containing approximately 70 pmol of ZP2 were submitted to N-terminal Edman sequencing. Automated Edman sequencing was performed using a model Procise 494cLC sequanator connected to an on-line model 140D UV detector (Applied Biosystems). A C-18 capillary column (250 mm × 0.8 mm) was used for the separation of the phenylthiohydantoin (PTH) amino acids. Standard sequencing protocols were used according to the manufacturer's recommendations.

Sperm Binding Assays. Sperm binding assays were performed (*10*), and sperm binding numbers were expressed as the average number of sperm bound per egg ± the standard deviation.

RESULTS

Assessment of the Amino Termini of Rat ZP1, ZP2, and ZP3. Virtually all extracellular proteins have an N-terminal signal peptide that directs them into the secretory pathway and is subsequently removed by signal peptidases in the endoplasmic reticulum. A computer algorithm (*14*) predicts cleavage of rat ZP1, ZP2, and ZP3 immediately upstream of Gln21, Leu24, and Gln23, respectively (see Figures 1−3).

Peptide mapping of ZP1 from Asp-N digestion of the zona mixture followed by LC−MS indicated that the N-terminus starts at Gln21 (Figure 1), which had been converted to pyroglutamate. The +2 and +3 charged ions at *m*/*z* 800.90 and 534.27, respectively (Figure 4A, inset), represent the N-terminal peptide ^{21}qHLHLKPGFQYSY^{33}. The CID spectrum (Figure 4A) of the precursor ion at *m*/*z* 534.27^{3+} (inset) indicated the presence of both y and b ion series, including y_{1-2}, y_1-NH_3, b_{2-5}, and b_{10}. In addition, the a_4 ion, internal fragment ions of PG, PGFQ, and YS, and immonium ions of tyrosine were observed.

Unfortunately, MS data did not reveal any N-terminal information about rat ZP2 (see Table 4). Thus, Edman microsequencing was performed on the whole rat zona preparation on the basis of the assumption that ZP2 is one of the more abundant zona proteins and not N-terminally blocked as in the case of ZP1 (described above) and ZP3 (see below). Indeed, 20 cycles of microsequencing revealed that the N-terminus of rat ZP2 starts at Ser25 (at least in the majority of the population), one amino acid downstream from the predicted Leu24 site (i.e., SLPQSENPAFPGTLIC*DK,

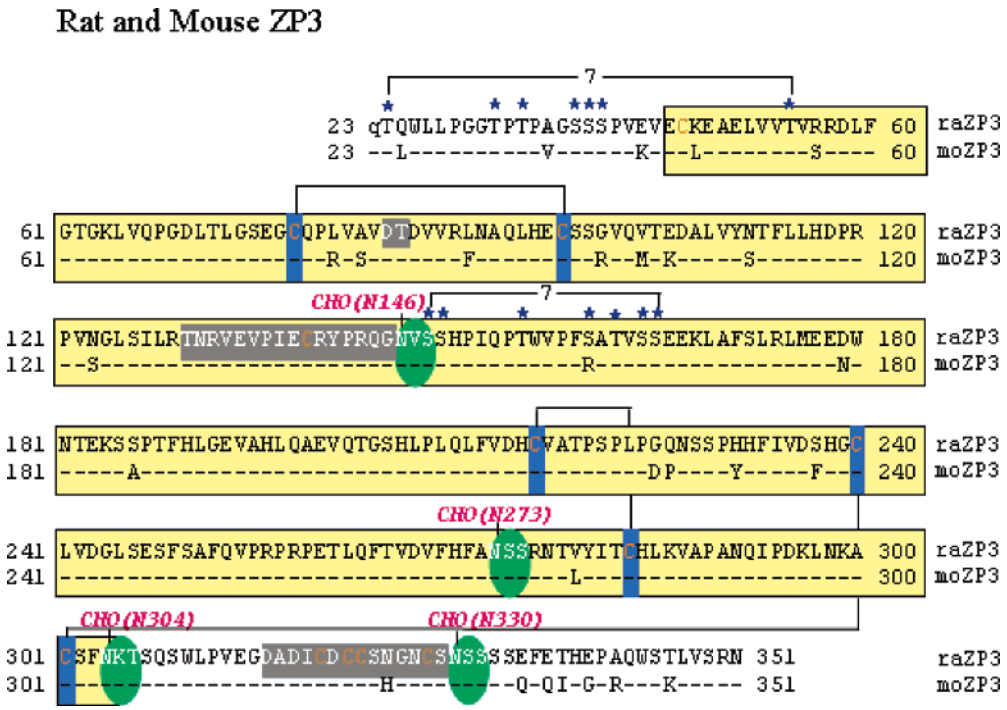

FIGURE 3: Summary of rat and mouse ZP3. The primary amino acid sequence of ZP3 extends from an N-terminal pyroglutamate (q=pyrQ23) to a C-terminal asparagine (N351) immediately upstream of a conserved dibasic motif. There are 12 cysteines in the mature form of ZP3, eight of which are conserved in the zona domain (C46, C78, C98, C139, C216, C240, C283, and C301). Three pairs of disulfide-linked cysteines were found (C78−C98, C216−C283, and C240−C301), as well as a peptide containing four disulfide-bonded cysteines (C320, C322, C323, and C328) that are C-terminal to the zona domain. The linkage of the latter was indeterminate (dashed line) due to clustering of cysteine residues and the absence of appropriate cleavage sites. Four of the six potential N-linked glycan sites are definitively glycosylated (N146, N273, N304, and N330), and one is not (N227). The glycosylation status of N327 could not be ascertained. There appear to be two clusters of O-linked glycans containing multiple potentially O-glycosylated serine or threonine residues: one at the N-terminus and one in the zona domain. Clusters are indicated with brackets and potential sites with asterisks (dark blue), and the numbers of glycans with Arabic numerals. Peptides representing ~9% of mature ZP3 were not identified. Color indications are the same as in Figure 1.

where C* is carbamidomethylated cysteine) (Figure 2, where Ser25 is the underlined N-terminus).

For rat ZP3 (Figure 3), proteolytic digestion of the PNGase F-treated zona mixture initially did not allow us to detect the N-terminal peptide. However, further exo-*O*-glycosidase digestion revealed a whole series of N-terminally carbamidomethylated peptides where Gln23 was cyclized to pyroglutamate with various core *O*-glycan attachments (i.e., HexNAc·Hex with additional HexNAc or Hex attached) (Table 5). However, a mass decrease of 71 Da was observed from its sequence predicted on the basis of cDNA. We speculate that this is due to the absence of Ala36 or a mutation of Lys47 to Gly, although fragment ions did not allow a decision between these sites. As shown in Figure 4B, the trypsin/Asp-N double digest gave rise to the peptide ^{23}qTQWLLPGGTPTP(A?)GSSSPVEVEC(K/G?)EAELVVTVR^{56} attached to four HexNAc·Hex residues and one HexNAc residue [*m*/*z* 1733.46^{3+} (not shown) and 1300.35^{4+} (inset)]. The CID spectrum of the latter showed the presence of b_{3-6} and y_{1-3}, ions clearly indicating the cyclization of Gln23 to pyroglutamate, in addition to previously described intense carbohydrate marker ions at *m*/*z* 204.09 and 366.17 (*6*) that reduce the apparent significance of the peptide ions. Examination of a Glu-C digest of the N-deglycosylated sample did not yield any information about this N-terminal peptide, suggesting that *O*-glycans may be inhibiting digestion.

Assessment of the Mature Carboxyl Termini of Rat ZP1 and ZP3. The generation of mature secreted zonae glycoproteins includes C-terminal processing to release them from an anchoring transmembrane domain. For rat ZP1, an Asp-N digestion revealed a MH$^+$ peptide of 827.42 Da corresponding to the ^{538}DSEIARH^{544} sequence as both a +1 and +2 charged ion at *m*/*z* 827.42 and 414.21 (Figure 5A). This suggests that the carboxyl terminus of rat ZP1 (His544) lies two amino acids upstream of a potential furin cleavage site. Unfortunately, the DDA conditions did not provide a CID spectrum of *m*/*z* 414.21 perhaps due to the multitude of peaks at this relatively low *m*/*z* value. Numerous attempts to identify the C-terminus of ZP2 by various enzymatic proteolysis and deglycosylation strategies failed.

For ZP3, in which there was no convenient aspartate residue, we digested with PNGase F, which releases protein-bound *N*-glycans and converts Asn330 to aspartic acid. Subsequent Asp-N digestion and LC−MS revealed the presence of the C-terminal peptide ^{330}NSSSSEFETHEPAQWSTLVSRN^{351} at *m*/*z* 832.04^{3+} (not shown) and 1247.57^{2+} (Figure 5B, inset), 1 Da higher than expected, indicating the conversion of Asn330 to aspartate. Its identity was confirmed by CID, including y_{2-14} (except y_9) and b_{2-11} (except b_8) ions with the expected abundance enhancement at proline. Glu-C digestion generated a ^{338}THEPAQWSTLVSRN^{351} peptide at *m*/*z* 813.41^{2+} with its CID spectrum further confirming the C-terminus of ZP3 at Asn351 (data not shown; see Table 5).

Taken together, these mass spectrometric data indicated that the primary cleavage sites of native rat ZP1 and ZP3 lie N-terminal to a dibasic motif that is part of, but distinct

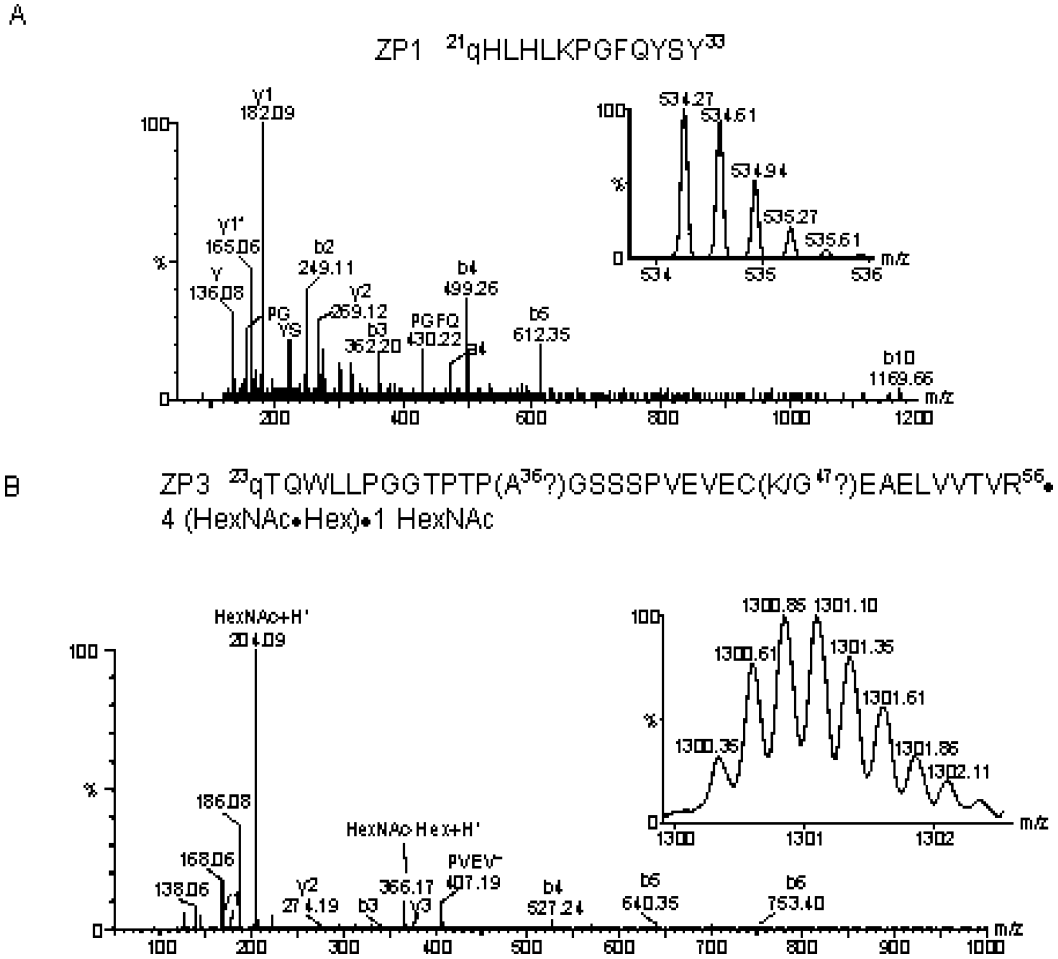

FIGURE 4: Identification of the amino termini of mature secreted rat ZP1 and ZP3. The N-termini were mapped by digesting rat zonae with Asp-N alone or Asp-N and trypsin followed by microscale LC−MS analysis. (A) The N-terminus of ZP1 defined by the Asp-N peptide ^{21}qHLHLKPGFQYSY^{33} with pyroglutamate (q) in place of Gln21 exhibits the +3 charged ion at *m*/*z* 534.27 (inset). The CID spectrum confirms the sequence. (B) The CID spectrum of the +4 charged precursor ion at *m*/*z* 1300.35 corresponding to the amino-terminal peptide ^{23}qTQWLLPGGTPTP(A?)GSSSPVEVEC(K/G?)EAELVVTVR^{56} of rat ZP3 from sequential trypsin and Asp-N cleavage confirms that a pyroglutamate replaced a glutamine and a unit of five HexNAc molecules and one Hex molecule was attached. In addition, Ala36 was missing or Lys47 was replaced with a Gly, resulting in a mass decrease of 71 Da (compared to cDNA data).

from, the proprotein convertase (furin) cleavage site. We have reported a similar cleavage site in rat ZP4 (*10*), and it is likely that C-terminal processing in rat ZP2 is similarly conserved.

Disulfide Linkage Mapping. Using the same fragmentation nomenclature reported previously (*6*), we have identified a disulfide bridging pattern in rat that is identical to that of mouse zona proteins (see Table 1). In this, the two disulfide-bonded peptide chains are arbitrarily designated as P1 and P2, priming fragmentations that arise from the latter, i.e., y′. Since the disulfide bridge is sometimes "reductively" cleaved either between the sides or on either side of S, peptide fragment ions will appear carrying either an SH or SSH at the cysteine site, and these are called y^r (or y'^r) or y^d (or y'^d), respectively. In cases where y^d (or y'^d) ions were not observed, we simply labeled the fragment ions as y and y′.

Mature ZP1 has 21 cysteine residues and the potential to form 11 disulfide bonds, including at least one intermolecular disulfide bridge (ZP1 runs as a dimer on nonreduced SDS−PAGE gels). However, only one disulfide-bonded peptide was detected, perhaps due to the low abundance of ZP1 in the zona protein mixture and the paucity of native material. The +3 and +4 charged ions at *m*/*z* 1322.96 and 992.45 from a trypsin digest corresponded to the intramolecularly disulfide-bonded peptide ^{436}TDPGLALMLHQCWATP-SASPFEQPQWPILSDGCPFK^{471} between Cys447 and Cys468 (Table 1 and Figure 6A, calcd $[M + 3H]^{3+}$ 1322.96), corresponding to the bridge between Cys449 and Cys470 in mouse ZP1.

Mature rat ZP2 has 20 cysteine residues capable of forming 10 intramolecular disulfide bonds. We found a disulfide bond between Cys73 and Cys91, near the N-terminus of ZP2, outside the "zona domain", as in mouse ZP2. The +3 and +4 charged ions observed at *m*/*z* 1276.89 and 957.92, respectively (MH^+ 3828.68), correspond to the calculated mass of the S−S-linked peptides ^{58}WNPSLVDTFGNEISN-CTYALDLEK^{81} (P1; Asn72 → Asp72 conversion) and ^{86}FPYETCTIK^{94} (P2) (calcd MH^+ 3828.76). Moreover, the CID spectra of 957.92^{4+} showed fragment ions corresponding to the sequence of both peptides linked via disulfides (spectra not shown).

Within the zona domain containing 10 cysteines, eight of which are conserved, three of five possible disulfide bonds were identified. Cys354 and Cys447 formed a disulfide pair as observed by the +3 and +4 ions at *m*/*z* 475.53 and

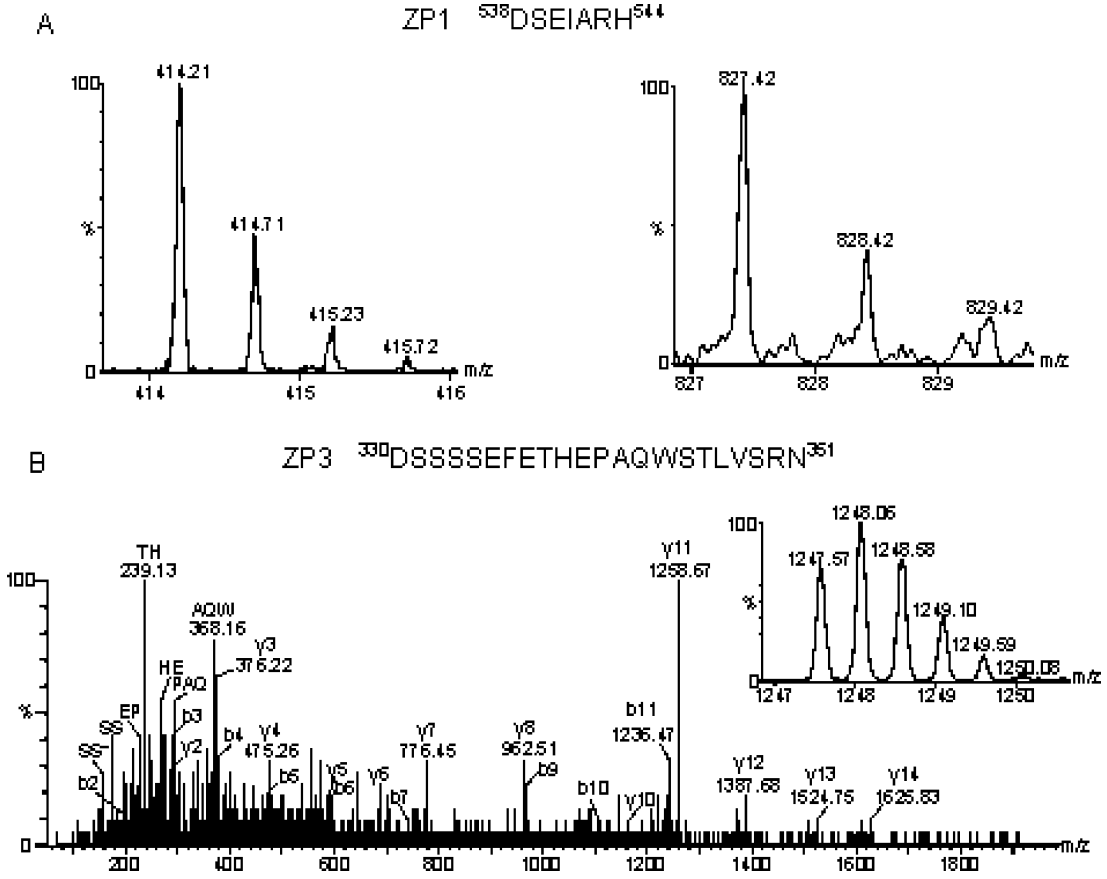

FIGURE 5: Identification of the carboxyl termini of mature secreted rat ZP1 and ZP3. Digestion with Asp-N (ZP1) or Glu-C (ZP3) identified the C-termini of ZP1 and ZP3 as ending immediately preceding a conserved dibasic peptide motif that is part of, but distinct from, the furin consensus cleavage sequence. (A) The C-terminus of ZP1 suggested by the +1 and +2 charged ions at *m/z* 827.42 and 414.21, respectively, corresponding to the peptide ^{538}DSEIARH^{544}. (B) The carboxyl terminal peptide of rat ZP3 ^{330}DSSSSEFETHEPAQWSTLVSRN^{351} was detected at *m/z* 1247.56 (+2 charged, inset) as well as *m/z* 832.04 (+3 charged), 0.98 Da higher than expected, demonstrating that Asn330 was replaced with Asp. The CID spectrum of the ion at *m/z* 1247.56 confirmed the sequence identity of this peptide (see the text).

356.90, respectively, derived from the trypsin/Asp-N digest (Table 1 and Figure 6B). The calculated MH^+ of the S−S-linked peptides ^{351}DEL$\underline{C}$TR^{356} (P1) and $^{447}\underline{C}$HYIR^{451} (P2) is 1424.63 Da, which is in good agreement with these experimental values. The CID spectrum of 475.53^{3+} generated y_{1-4} and b_2 from P1, as well as y'_{1-4}, b'_2, and immonium ions of tyrosine and arginine residues from P2 (Figure 6B).

The Cys385−Cys406 disulfide pair in ZP2 was observed by relatively lower abundance +4 and +5 charged ions at *m/z* 836.41 and 669.32, respectively (Table 1, MH^+ 3342.62, spectra not shown). This ion derived from trypsin digestion corresponds fairly well to the peptides ^{371}PALNLESLLVGNS-S$\underline{C}$QPIFK^{390} (Asn to Asp conversion at position 382 after PNGase F treatment) joined with ^{399}FHIPLNG$\underline{C}$GTR^{409} via a S−S bond (combined masses of two peptides minus 2 Da, calcd MH^+ 3342.70).

Two more disulfide links in ZP2 were contained in a single peptide (Table 1, spectra not shown), observed as a +3 charged ion at *m/z* 1087.52 (MH^+ 3260.56) which correspond to the intramolecularly disulfide-bonded peptide ^{588}GLSS-LIYFH$\underline{C}$SALI$\underline{C}$NQASPL$\underline{C}$SVT$\underline{C}$PAPLR^{618} formed among the four cysteines within the same tryptic peptide (two disulfide bonds with a loss of 4 Da, calcd MH^+ 3260.57). The CID spectrum of 1087.52^{3+} was not detected, nor were any peptides available to determine the precise disulfide pairing among these four cysteines.

The mature rat ZP3 amino acid sequence is essentially a compact zona domain. There are 12 cysteines in the mature form with four of them clustered near the C-terminus outside the zona domain. We were able to observe three disulfide bridges by MS. In the first pair, masses corresponding to the peptide ^{70}DLTLGSEG$\underline{C}$QPLVAV^{84} (P1) disulfide-linked to ^{91}LNAQLHE$\underline{C}$SSGVQ$\underline{C}$VTE^{106} (P2) were observed at *m/z* 1607.27^{2+} and 1071.86^{3+} (Table 1 and Figure 6C, inset) from a trypsin/Asp-N double digest. The CID spectrum of the 1071.86^{3+} species revealed the presence of y_{1-5}, a_{2-3}, and b_{2-6} ions from P1, together with y'_{1-4}, y'_6, a'_2, a'_4, and b'_{2-7} ions from P2 (Figure 6C). In the second pairing (Table 1, spectra not shown), a precursor ion of MH^+ 3528.69 as detected by its +4 and +5 charged ions at *m/z* 882.92 and 706.54, respectively, corresponds to ^{214}DH$\underline{C}$VATPSPLP-GQNSSPHHFIV^{235} (P1; Asn227 was not N-glycosylated) disulfide-bridged to ^{277}NTVYIT$\underline{C}$HLK^{286} (P2). The last pair was observed to occur between ^{236}DSHG$\underline{C}$LV^{242} and ^{300}A$\underline{C}$SFNK^{305} from a trypsin/Asp-N digest. The +2 and +3 charged ions at *m/z* 698.85 and 466.24, respectively, corresponding to this linkage (calcd MH^+ 1396.60 Da) was detected in the double digest sample. This observation shows that Cys240 was originally linked to Cys301. These three disulfide linkages were exactly the same as those previously reported in mouse ZP3 (*6*).

N-Linked Glycosylation Sites. N-Glycosylation of proteins occurs at only asparagine residues within the NXS/T

Table 1: Disulfide Bond Linkage Mapping of Native Rat Zona Proteins 1−3

ZP	Residue	Sequences	Enzymes	*m/z* exp.	*m/z* calc.[c]
1	436-471[a]	TDPGLALMLHQCWATPSASPFEQP QWPILSDGCPFK	Trypsin	1322.96^{3+} 992.45^{4+}	1322.96^{3+} 992.47^{4+}
2	58-81[b] 86-94	WNPSLVDTFGNEISN*CTYALDLEK FPYETCTIK	Trypsin	1276.89^{3+} 957.92^{4+}	1276.59^{3+} 957.69^{4+}
2	351-356 447-451	DELCTR CHYIR	Trypsin +Asp-N	475.53^{3+} 356.90^{4+}	475.55^{3+} 356.92^{4+}
2	371-390[b] 399-409	PALNLESLLVGN*SSCQPIFK FHIPLNGCGTR	Trypsin	836.41^{4+} 669.32^{5+}	836.18^{4+} 669.15^{5+}
2	588-618[a]	GLSSLIYFHCSALICNQASPLCSVTCPA PLR	Trypsin	1087.52^{3+}	1087.53^{3+}
3	70-84 91-106	DLTLGSEGCQPLVAV LNAQLHECSSGVQVTE	Trypsin +Asp-N	1071.86^{3+}	1071.85^{3+}
3	214-235 277-286	DHCVATPSPLPGQNSSPHHFIV NTVYITCHLK	Trypsin +Asp-N	882.92^{4+} 706.54^{5+}	882.93^{4+} 706.55^{5+}
3	236-242 300-305	DSHGCLV ACSFNK	Trypsin +Asp-N	698.85^{2+}	698.80^{2+}

[a] Intramolecular disulfide bonds within the same proteolytic fragment. [b] N-Glycosylated peptides (where N* represents an originally *N*-glycosylated asparagine residue converted to an aspartic acid upon PNGase F treatment). [c] Note that the calculated masses for N-glycosylated peptides are based on Asn instead of Asp to indicate an observed mass increase of 0.98 Da after enzyme treatment, and this nomenclature is used throughout the entire text unless otherwise stated.

consensus sequence, where X cannot be a proline. The endoglycosidase PNGase F releases protein-bound N-linked glycans and, as mentioned earlier, converts the involved asparagine residue to an aspartic acid, providing a signature mass increase (0.98 Da). Although other situations during sample preparation, for instance, deamidation in the Asn-Gly sequence motif resulting in Asn to Asp conversion (see Tables 3−5), make a control experiment by trypsin digestion without any prior PNGase F treatment very desirable, our results under such conditions failed to yield good sequence coverage due to the interference from bulky N-linked oligosaccharides.

Assuming our determination of the C-terminus in ZP1 and ZP3 (see above) was correct, there are three predicted N-linked glycosylation sites that follow the NXS/T sequence motif in native secreted ZP1, seven in ZP2 (where the C-terminus is not defined, but is likely the same as that in mouse ZP2), and six in ZP3 (Figures 1−3). In ZP1, all three predicted asparagines at positions 49, 68, and 369 were N-glycosylated within the mature protein (Table 2). Figure 7A is an example of the CID spectrum of a +2 charged peptide (^{49}NQTIQFKVL^{57}) at *m/z* 546.32 derived from Asp-N digestion. The MH^+ ion of this peptide is 1.02 Da higher than the expected value (MH^+ 1090.62), showing that Asn49 was converted to Asp. Fragmentation generated b_{2-4} as well as y_{1-6} and y_5-NH_3, confirming the peptide sequence. The b_{2-4} ions along with their H_2O and NH_3 losses clearly demonstrated a change of Asn to Asp at position 49 upon PNGase F treatment.

In ZP2, at least six out of seven N-glycosylation sites were occupied (Table 2). Trypsin, Asp-N, and trypsin/Asp-N digestion after PNGase F treatment clearly showed that six Asn residues at positions 72, 161, 206, 253, 382, and 563 and possibly a seventh residue at position 173 were converted to Asp. The possibility of deamidation at the Asn-Gly sequence motif at this site (Asn173) cannot be ruled out due to its occurrence at one other NG sequence in ZP2 (residues 281−297, Table 4). In Figure 7B, the glycosylation site identification by CID is illustrated for the +2 charged ion of a tryptic glycopeptide ^{155}LADENQNVSEMGWIIK^{170} at *m/z* 924.44 (inset) as an example. Again, the experimental precursor ion MH^+ 1847.88 is 0.98 Da higher than the calculated value of this peptide (MH^+ 1846.90). The presence of y_{1-9} and b_{2-5} ions confirms the sequence identity. Interestingly, some of the predicted sites have incomplete N-glycosylation (see Table 4, e.g., Asn72 in residues 58−81 and Asn161 in residues 161−170) as ions corresponding to their predicted masses with Asn rather than Asp were also observed (*m/z* in parentheses).

Similarly, proteolytic digestion of ZP3 generated four of six Asp-containing peptides after PNGase F deglycosylation (Table 2). A +3 charged ion at *m/z* 699.35 (Figure 7C) indicates that the tryptic peptide ^{259}PETLQFTVDVFH-FANSSR^{276} was N-glycosylated at Asn273. Its CID spectrum

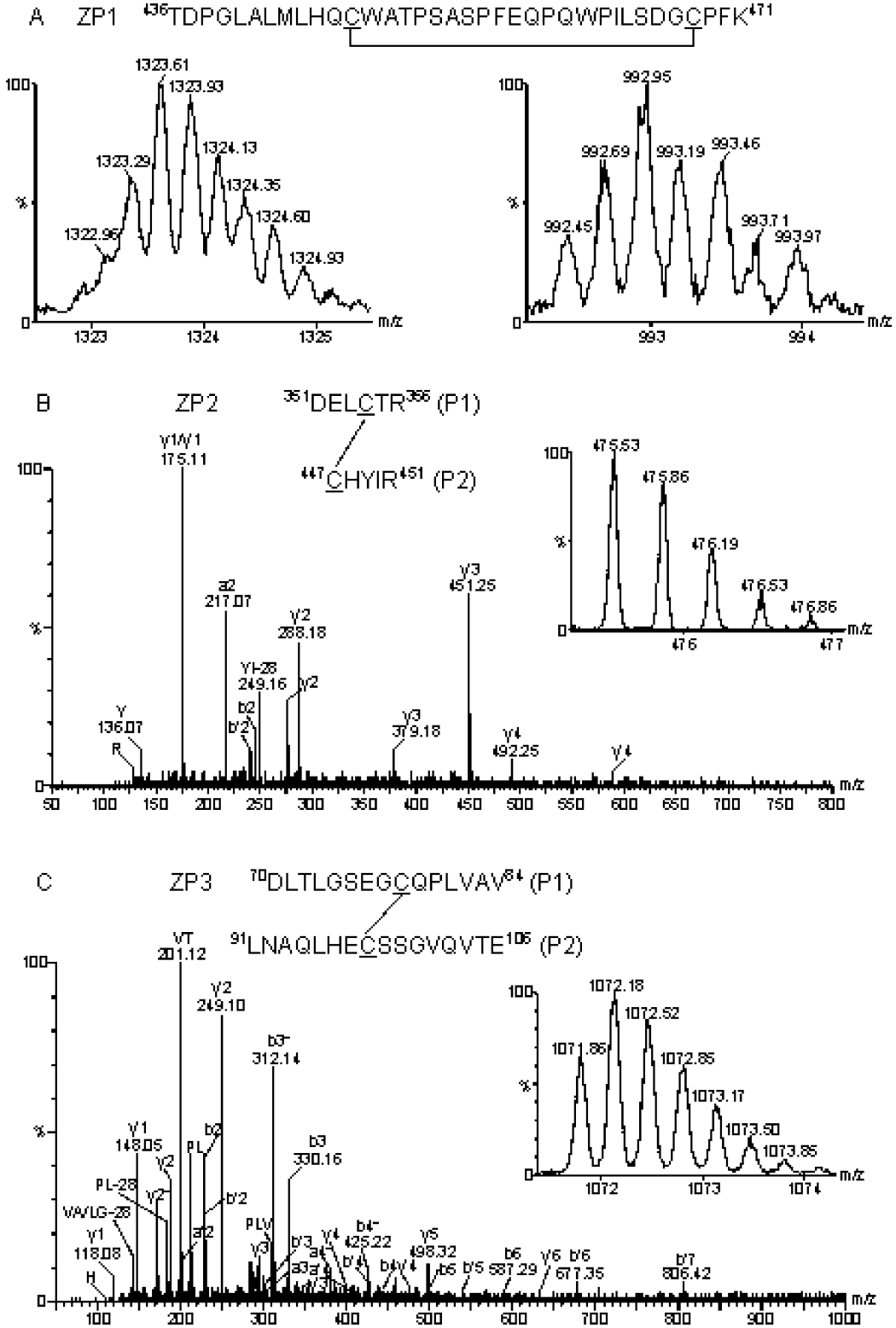

FIGURE 6: Disulfide bond localization in rat ZP1−ZP3. According to our nomenclature, wherever y^d (or y'^d) ions were not observed, we simply labeled fragment ions as y and y′. (A) An intramolecular disulfide-linked peptide ^{436}TDPGLALMLHQCWATPSASPFEQPQWPILSDGCPFK^{471} from ZP1 derived from trypsin digestion is shown at *m*/*z* 1322.96^{3+} and 992.45^{4+} without being selected for CID. (B) The CID spectrum of an ion at *m*/*z* 475.53^{3+} from ZP2 shows fragment ions corresponding to the sequences ^{351}DELCTR^{356} (P1) and ^{447}CHYIR^{451} (P2) disulfide-bonded to each other. Many ions are formed from "reductive" processes and contain either CysSH or CysSSH (see the text). (C) The disulfide linkage formed between ^{70}DLTLGSEGCQPLVAV^{84} (P1) and ^{91}LNAQLHECSSGVQVTE^{106} (P2) of ZP3 was detected by ions at *m*/*z* 1071.86^{3+} (inset) and 1607.27^{2+}. The CID spectrum as shown here clearly indicated the fragment ions derived from both peptides connected via a disulfide bond in addition to internally cleaved ions.

showed the sequential y_{1-7} and y_9 ions, demonstrating the replacement of Asn273 with Asp, as well as b_{2-6} and a_2 ions. Analysis of an Asp-N digest (Table 2) revealed a mass shift of +0.98 Da for peptides ^{304}NKTSQSWLPVEG^{315} and ^{330}NSSSSEFETHEPAQWSTLVSRN^{351}, indicating N-glycosylation at Asn304 and Asn330. An Asn residue at position

Table 2: LC−MS Analysis of Rat ZP1−ZP3 N-Linked Glycosylation Sites

ZP	residues[b]	sequence[f]	enzyme	*m*/*z*(exp)	*m*/*z*(calcd)[g]
1	49−57[c]	N*QTIQFKVL	Asp-N	546.32^{2+}	545.82^{2+}
1	58−67	DEFGNRFEVN(N*C*S)	Asp-N	613.78^{2+}	613.78^{2+}
1	366−377[c]	C*IFN*ASDFLPIR	trypsin	727.34^{2+}	726.87^{2+}
2	58−81[c]	WNPSLVDTFGNEISN*CTYALDLEK	trypsin	910.78^{3+}	910.43^{3+}
2	155−170[c]	LADENQN*VSEMGWIIK	trypsin	924.44^{2+}	923.95^{2+}
				616.63^{3+}	616.31^{3+}
2	161−170[c]	N*VSEMGWIIK	trypsin and Asp-N	589.30^{2+}	588.81^{2+}
2	173−183[d]	N?GTRVHTLPLK	trypsin and Asp-N	618.87^{2+}	618.37^{2+}
				412.91^{3+}	412.58^{3+}
				309.92^{4+}	309.69^{4+}
2	199−226[c]	ITLHVPAN*ATGVAHYVQESSYLYTVQLK	trypsin	1035.21^{3+}	1034.88^{3+}
				776.65^{4+}	776.41^{4+}
2	206−226[c]	N*ATGVAHYVQESSYLYTVQLK	trypsin and Asp-N	1186.59^{2+}	1186.10^{2+}
				791.40^{3+}	791.07^{3+}
2	253−280[c]	N*VTHMSLTIPEFPGKLKSVGFGQRNIPE	Asp-N	775.15^{4+}	774.91^{4+}
				620.32^{5+}	620.13^{5+}
2	382−390[c]	N*SSC*QPIFK	trypsin and Asp-N	541.26^{2+}	540.76^{2+}
2	557−574[c]	TTFHAAN*SSAAHSGHYQR	trypsin	648.62^{3+}	648.30^{3+}
				486.71^{4+}	486.48^{4+}
3[a]	146−168[c,e]	N*VSSHPIQPTWVPFSATVSSEEK·2(HexNAc·Hex)	trypsin and Asp-N	1086.83^{3+}	1086.51^{3+}
				1629.78^{2+}	1629.26^{2+}
3	259−276[c]	PETLQFTVDVFHFAN*SSR	trypsin	699.35^{3+}	699.01^{3+}
3	304−315[c]	N*KTSQSWLPVEG	Asp-N	673.82^{2+}	673.34^{2+}
3	330−351[c]	N*SSSSEFETHEPAQWSTLVSRN	Asp-N	1247.57^{2+}	1247.07^{2+}
				832.04^{3+}	831.71^{3+}

[a] Refer to Table 3 for details on glycosylation heterogeneity. [b] The residue numbers based on the primary sequences of full-length ZP1−ZP3. [c] N-Glycosylated peptides. [d] The possibility of either N-deglycosylation (N*GS/T) or deamidation (N#G). [e] O-Glycosylated peptides. [f] C* is carbamidomethylated cysteine. M* is methionine sulfoxide. N*XS/T is an N-substituted asparagine converted to aspartate after PNGase treatment (+0.98 Da). N is a non-N-substituted asparagine site. [g] The calculated masses for N-glycosylated peptides are based on Asn instead of Asp to indicate an observed mass increase of 0.98 Da after enzyme treatment. Some of these N-linked peptides are partially glycosylated (see Table 1C).

227 was found unglycosylated presumably due to the adjacent proline residue as in the case of mouse ZP3, while the status of Asn327 was not determined because of the lack of sequence coverage from Asp316 to Ser329. It is possible that Asn327 is N-glycosylated in rat ZP3 on the basis of our general observation that rat and mouse ZP3 share a great degree of structural similarity.

O-Linked Glycosylation Sites. *O*-Glycans attach to threonines and serines; however, one cannot predict accurately which residues are O-glycosylated, due to the absence of an O-glycosylation consensus sequence. Instead, monosaccharides must be removed by a series of exo-*O*-glycosidases [sialidase A, β(1−4)-galactosidase, β-*N*-acetylglucosaminidase] until only the Hexβ(1−3)HexNAc core remains attached to the serine and threonine residues, which can then be detected as a "signature" of O-glycosylation, i.e., a mass increase of 365.13 Da/core glycan over the peptide backbone. Further O-deglycosylation with endo-*O*-glycosidase removes the core sugar, leaving serine and threonine residues unmodified.

There are no O-glycosylation sites predicted for rat ZP1 (http://us.expasy.org/sprot), and indeed, we did not detect any. Our observation confirms the program prediction (written for mucin-type glycoproteins), or ZP1 glycopeptides were below our detection limit because of the paucity of biological material, as described in Materials and Methods, and/or the existence of sparse amounts of rat ZP1 in the rat zona mixture.

In ZP2, Thr444 was the only predicted O-glycosylation site. We did not observe the ^{443}MTVR^{446} tryptic peptide with or without *O*-glycans, which is consistent with a mass spectrometric analysis of mouse ZP2 (*6*). One explanation for missing this unglycosylated peptide is that the +2 charged ion (calcd *m*/*z* 253.64) lies outside our set MS scan range. Alternatively, this peptide may have not been retained on the C_{18} column especially if there are *O*-glycans attached.

The prediction program suggests five potential O-glycosylation sites, including Thr32, Thr34, Ser39, Thr155, and Thr162 in rat ZP3. Two domains were identified that contain one or more O-linked oligosaccharide side chains: one at the N-terminus (residues 23−56 with seven potential sites) and the other within the zona domain (residues 146−168 with seven potential sites). As previously described, multiply charged ions (mostly +3 and +4) of the N-terminal ZP3 peptide ^{23}qTQWLLPGGTPTP(A?)GSSSPVEVEC(K/G?)-EAELVVTVR^{56} with a series of sugar attachments (Table 5) were detected in the N/exo-O-deglycosylated sample (refer to Figure 4C for an example), but not in the N-deglycosylated sample. This observation indicates the heterogeneity of O-glycosylation at multiple sites (up to a maximum of seven HexNAc and seven Hex residues among seven potential sites at Thr24, Thr32, Thr34, Ser38, Ser39, Ser40, and Thr54).

In the N/exo-O-deglycosylated sample, the middle region between residues 146 and 168 was also heterogeneously O-glycosylated with multiple sugar residues (Table 5). As an example, the +2 and +3 charged ions of the doubly glycosylated [i.e., HexNAcβ(1−3)Hex] ZP3 trypsin/Asp-N peptide ^{144}NVSSHPIQPTWVPFSATVSSEEK^{168} were detected at *m*/*z* 1629.78 and 1086.83, respectively (a mass increase of 0.98 as a result of the Asn-to-Asp conversion at position 146). The CID spectrum of the ion at *m*/*z* 1086.83^{3+} (Figure 8A) confirmed the sequence by the presence of y_{1-13}, y_{15}, $y_{16}-NH_3$, b_{2-8}, and b_{10-12} ions, all of which bear no core sugars. The b ion series demonstrated that Asn146 was converted to an aspartic acid upon PNGase F deglycosylation. The presence of a Hexβ(1−3)HexNAc core was again

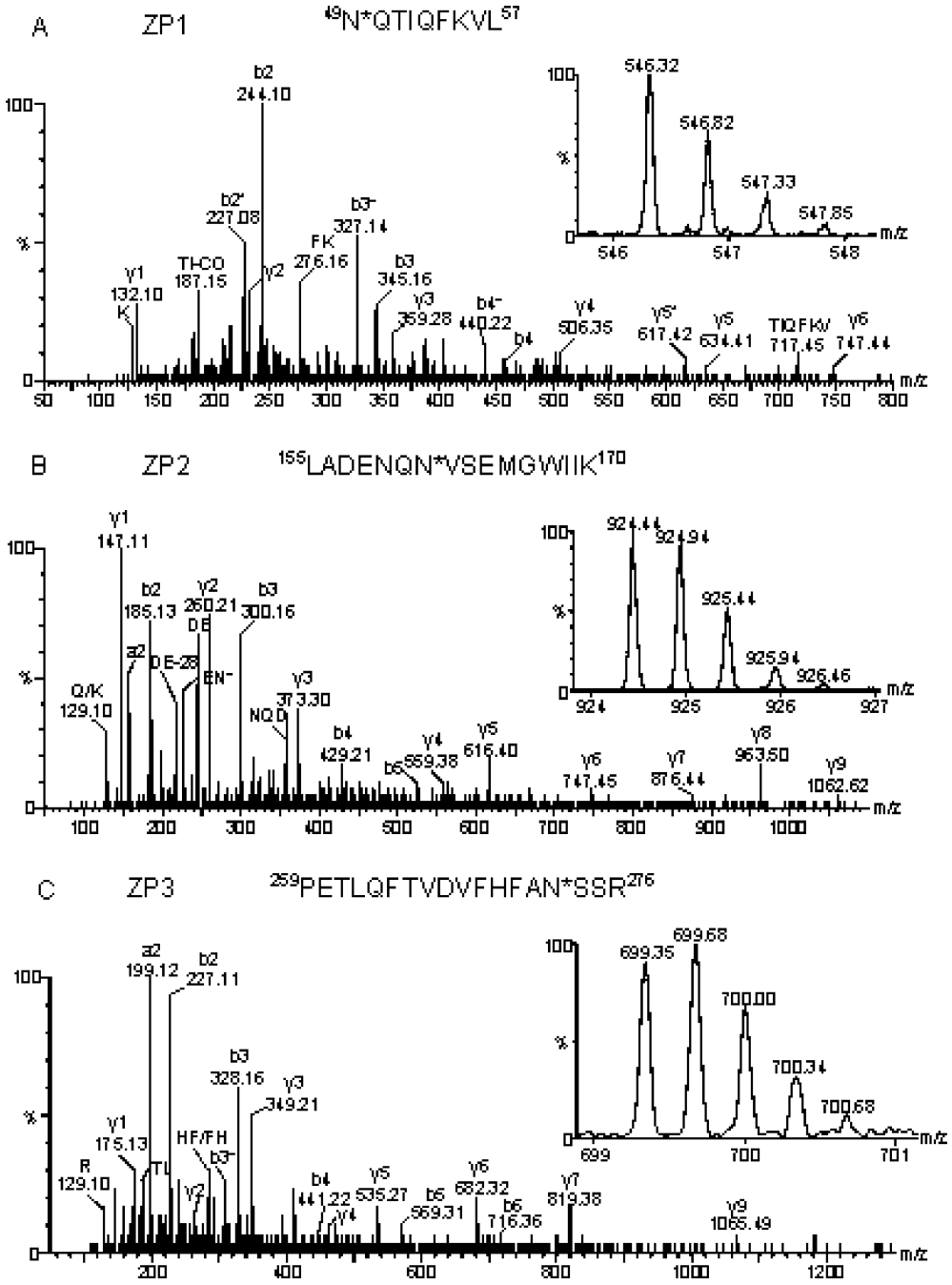

FIGURE 7: Localization of N-glycosylation sites in rat ZP1−ZP3. (A) CID spectrum of ZP1 N-linked glycopeptide ^{49}NQTIQFKVL^{57} at *m*/*z* 546.32^{2+}. (B) ZP2 N-linked glycopeptide ^{155}LADENQNVSEMGWIIK^{170} at *m*/*z* 924.44^{2+}. (C) ZP3 N-linked glycopeptide ^{259}PETLQFTVDVFHFANSSR^{276} at *m*/*z* 699.35^{3+} with an Asn-to-Asp conversion at position 273 resulting in a mass increase of 0.98 Da.

detected by the base peak at *m*/*z* 204.09 and the less abundant peak at *m*/*z* 366.15. Unfortunately, CID data did not allow us to determine the exact location of the sugar linkage among the seven potential sites (Ser148, Ser149, Thr155, Thr160, Thr162, Thr164, and Thr165) because of the easy loss of the sugar moiety upon CID. Figure 8B is the CID spectrum of the same peptide carrying two (HexNAc·Hex)·Hex sugar residues, which showed virtually the same fragmentation pattern as the previous peptide (Figure 8A). Our data in Table 5 showed a maximum of seven Hexβ(1−3)HexNAc sugars in addition to this peptide (at *m*/*z* 1695.40^{3+}), indicating a population of full occupancy of *O*-glycans in this region.

Finally, as in the case of mouse and human ZP3 (*6*, *7*), there was no evidence of O-glycosylation at serines 332 and 334 in rat ZP3. O-Glycosylation of these residues has been proposed to provide the basis for sperm binding (*15*). The Asp-N digest of the native N-deglycosylated ZP mixture generated the masses at *m*/*z* 1247.57^{2+} and 832.04^{3+} which correspond to the peptide ^{330}NSSSSEFETHEPAQWSTLVSRN^{351}, where Asp-N cleavage took place at Asn330 due to the Asn-to-Asp conversion. Since these masses match the calculated masses of this peptide with the replacement of Asn with Asp without any prior O-deglycosylation treatment (N-deglycosylated sample), and since the peptide identity was confirmed by CID sequence data, it indicates that neither Ser332 nor Ser334 is O-glycosylated at a measurable level. We again looked specifically for the masses corresponding to various combinations of glycosylation sites in this serine-rich region using extracted ion chromatograms in the N/O-deglycosylated samples without any success. Hence, with

Table 3: LC−MS and MS/MS Analysis of Rat ZP1

residues[a]	sequence[c]	enzymes	*m*/*z*(exp)	*m*/*z*(calcd)[d]	elution time (min)
21−33	qHLHLKPGFQYSY	Asp-N	800.90^{2+}	800.90^{2+}	52.0
			534.27^{3+}	534.27^{3+}	
49−57[b]	N*QTIQFKVL	Asp-N	546.32^{2+}	545.82^{2+}	57.7
58−67	DEFGNRFEVN(N*C*S)	Asp-N	613.78^{2+}	613.78^{2+}	48.0
107−117	VFIQAVLPNGR	trypsin and Asp-N	607.36^{2+}	607.36^{2+}	54.3
123−131	DVTLIC*PKP	trypsin and Asp-N	521.79^{2+}	521.78^{2+}	33.8
214−224	LDSIGTHLLQE	Glu-C	613.32^{2+}	613.33^{2+}	53.2
215−225	DSIGTHLLQER	trypsin and Asp-N	634.84^{2+}	634.83^{2+}	45.6
			423.56^{3+}	423.56^{3+}	
226−238	C*QVASGHIPC*MVK	trypsin and Asp-N	496.24^{3+}	496.24^{3+}	41.1
273−292	SGYFTLVMSQETALTHGVML	trypsin and Asp-N	1093.02^{2+}	1093.04^{2+}	78.3
			729.01^{3+}	729.03^{3+}	
293−301	DNVHLAYAP($\underline{N^{\#}G}$)[e]	trypsin and Asp-N	500.26^{2+}	500.25^{2+}	45.0
302−309	$\underline{N^{\#}G}$C*PPTQK[e]	trypsin and Asp-N	451.70^{2+}	451.71^{2+}	40.4
333−356	$\overline{QL}$VYENQLVSNIEVQTGPQGSITR	trypsin and Asp-N	1337.20^{2+}	1337.20^{2+}	62.1
		trypsin	891.80^{3+}	891.80^{3+}	
366−377[b]	C*IFN*ASDFLPIR	trypsin	727.34^{2+}	726.87^{2+}	63.7
378−392	ASIFSPQPPAPVTR	trypsin	734.40^{2+}	734.40^{2+}	51.2
399−429	LRIATDKTFSSYYQGSDYPLVRLLQEPVYIE	Glu-C	1222.30^{3+}	1222.30^{3+}	76.4
			916.97^{4+}	916.98^{4+}	
406−420	TFSSYYQGSDYPLVR	trypsin	891.91^{2+}	891.92^{2+}	56.3
421−431	LLQEPVYIEVR	trypsin	679.88^{2+}	679.89^{2+}	56.8
466−471	DGC*PFK	trypsin and Asp-N	362.18^{2+}	362.16^{2+}	40.5
477−484	TQMVAADR	trypsin	446.23^{2+}	446.22^{2+}	6.6
485−495	ATLPFWSHYQR	trypsin	703.36^{2+}	703.35^{2+}	55.5
			469.24^{3+}	469.24^{3+}	
496−513	FTIATFTLLDSSSQNALR	trypsin	993.02^{2+}	993.02^{2+}	73.4
538−544	DSEIARH	Asp-N	827.42^{+}	827.40^{+}	5.3
			414.21^{2+}	414.20^{2+}	

[a] Residue numbers based on the primary sequence of full-length ZP1. [b] N-Glycosylated peptides. [c] q is pyroglutamic acid. C* is carbamidomethylated cysteine. C is uncarbamidomethylated cysteine. M* is methionine sulfoxide. N* is N-linked asparagine converted to aspartate after PNGase treatment (0.98 Da). N is non-N-linked asparagine. [d] Note that the calculated masses for N-glycosylated peptides are based on Asn instead of Asp to indicate an observed mass increase of 0.98 Da after enzyme treatment. In cases where deamidation or N-glycosylation is the possible cause for a mass shift of +0.98 Da, a question mark is placed next to the corresponding Asn residue. Furthermore, masses with a shift of +0.98 Da inside the parentheses represent partial N-glycosylation or deamidation. This nomenclature is used for all tables unless otherwise stated. [e] ($\underline{N^{\#}G}$) represents the deamidation sequence motif.

our MS detection capabilities (∼1−2 fmol), the only Ser and Thr residues we found to be O-glycosylated out of 59 potential sites were in an N-terminal cluster and a second cluster in the zona domain.

Sperm Binding Assays. Previous sperm binding studies have demonstrated that a zona pellucida matrix comprised solely of mouse ZP2 and ZP3 is sufficient to support mouse sperm binding (*13*). This paper shows, using mass spectrometry, that the major structural features of ZP1, ZP2, and ZP3 are highly conserved between rats and mice. Taken together, these data led us to hypothesize that rat sperm would bind to a mouse zona matrix containing only ZP2 and ZP3, even though rat zonae are comprised of four zona glycoproteins (*10*). That is, ZP2 and ZP3 contain a minimum structural unit to which rodent sperm bind. Sperm binding assays were conducted by "challenging" normal mouse eggs and Zp1 null mouse eggs with rat or mouse sperm. A few two-cell embryos were also included in these assays. Washing eggs and two cell embryos together with a wide-bore pipet until only one to five sperm remained attached to two cell embryos permitted the counting of sperm that were tightly bound to eggs since sperm only loosely adhere to zonae surrounding two cell embryos. The eggs were fixed and stained with Hoechst to visualize sperm heads, and the total number of sperm bound per egg was quantified using an LSM-510 Zeiss confocal microscope (Figure 9). Each assay was repeated in triplicate with 20−30 eggs. As expected, mouse sperm bound both to normal mouse eggs (71.8 ± 14 sperm/egg) and null mouse eggs for ZP1 (78.55 ± 14 sperm/egg). Similarly, rat sperm bound to normal mouse eggs (43.30 ± 13 sperm/egg) and Zp1 null mouse eggs (52.35 ± 7 sperm/egg).

DISCUSSION

Molecular Basis of Sperm−Zona Binding. The mechanism by which sperm bind to the zona pellucida remains elusive despite decades of investigation (*11*). A model in which a sperm receptor binds to a zona glycan ligand that is removed by a cortical granule glycosidase postfertilization has received considerable attention. Both terminal *O*-glycan residues on Ser332 and Ser334 in ZP3, as well as polylactosamine containing *N*-glycan structures, have been proposed as sperm binding ligands (*15*, *16*). However, mass spectrometric analyses of native mouse ZP3 (*6*), recombinant human ZP3 (*7*), and rat ZP3 (this work) have not detected *O*-glycans on Ser332 or Ser334. Moreover, mouse genetic models lacking specific candidate zona carbohydrate moieties or their proposed sperm receptors, as well as zonae lacking type 2 *O*-glycans or complex and hybrid *N*-glycans, remain fertile in vivo (*17*−*20*). The only observed molecular change in the zona pellucida constituents postfertilization is the proteolytic cleavage of ZP2, attributed to a putative secreted cortical granule protease (*21*, *22*). It has been shown that mouse sperm bind persistently to the zona pellucida of genetically engineered mice in which mouse ZP2 is replaced

Table 4: LC−MS and MS/MS Analysis of Rat ZP2

residues	sequence	enzyme	*m*/*z*(exp)	*m*/*z*(calcd)	elution time (min)
47−52	VEFSSR	trypsin	362.67^{2+}	362.68^{2+}	41.5
53−57	FDMEK	trypsin	335.15^{2+}	335.15^{2+}	28.1
54−63	DMEKWNPSLV	Asp-N	609.80^{2+}	609.80^{2+}	55.3
58−81	WNPSLVDTFGNEISN*CTYALDLEK	trypsin	910.44^{3+} (910.78^{3+})	910.43^{3+}	63.8
81−89	KFIIKFPYE	Glu-C	592.84^{2+}	592.84^{2+}	59.2
95−104	VIGGYQVNIR	trypsin	559.82^{2+}	559.82^{2+}	47.8
105−115	VQDTNADVSYK	trypsin	620.30^{2+}	620.30^{2+}	33.0
141−154	EDLISFSFPQLFSR	trypsin and Asp-N	843.42^{2+}	843.43^{2+}	78.5
155−170	LADENQN*VSEMGWIIK	trypsin	924.44^{2+}	923.95^{2+}	63.7
			616.63^{3+}	616.31^{3+}	
157−172	DENQNVSEMGWIIKIG	Asp-N	916.95^{2+}	916.95^{2+}	71.5
161−170	N*VSEMGWIIK	trypsin and Asp-N	588.83^{2+} (589.30^{2+})	588.81^{2+}	60.6
171−176	IGN?GTR	trypsin	309.66^{2+}	309.17^{2+}	3.7
173−183	N?GTRVHTLPLK	trypsin and Asp-N	618.87^{2+}	618.37^{2+}	40.0
			412.91^{3+}	412.58^{3+}	
			309.92^{4+}	309.69^{4+}	
177−183	VHTLPLK	trypsin	404.25^{2+}	404.26^{2+}	32.4
184−198	DAIVQGFNLLIDSQK	trypsin	830.95^{2+}	830.95^{2+}	71.2
199−226	ITLHVPAN*ATGVAHYVQESSYLYTVQLK	trypsin	1035.21^{3+}	1034.88^{3+}	72.2
			776.65^{4+}	776.41^{4+}	
206−226	N*ATGVAHYVQESSYLYTVQLK	trypsin and Asp-N	1186.59^{2+}	1186.10^{2+}	64.9
			791.40^{3+}	791.07^{3+}	
227−235	LLFSSPGQK	trypsin	488.77^{2+}	488.78^{2+}	45.1
236−267	ITFSSQAIC*APDLSVAC*N*VTHMSLTIPEFPGK	trypsin	1164.88^{3+}	1164.57^{3+}	65.5
253−280	N*VTHMSLTIPEFPGKLKSVGFGQRNIPE	Asp-N	775.15^{4+}	774.91^{4+}	64.7
			620.32^{5+}	620.13^{5+}	
270−276	SVGFGQR	trypsin	375.69^{2+}	375.70^{2+}	27.6
281−297	DQWHAN#GIDKEATN#GLR	trypsin	963.00^{2+}	962.97^{2+}	45.3
			(963.96^{2+})	642.31^{3+}	
			642.31^{3+}		
			(642.98^{2+})		
291−297	EATNGLR	trypsin	380.69^{2+}	380.70^{2+}	34.8
298−301	LHFR	trypsin	572.34^{+}	572.33^{+}	8.2
303−306	SLLK	trypsin	460.31^{+}	460.31^{+}	7.0
357−381	DGFMDFEVYSHQTKPALNLESLLVG(N*SS)	trypsin and Asp-N	1405.69^{2+}	1405.69^{2+}	79.9
		Asp-N	937.45^{3+}	937.46^{3+}	
361−381	DFEVYSHQTKPALNLESLLVG(N*SS)	Asp-N	1180.60^{2+}	1180.61^{2+}	70.8
			787.41^{3+}	787.41^{3+}	
382−390	N*SSC*QPIFK	trypsin and Asp-N	541.26^{2+}	540.76^{2+}	38.2
391−398	VQSLGLAR	trypsin	422.26^{2+}	422.26^{2+}	37.6
399−409	FHIPLNGC*GTR	trypsin and Asp-N	424.54^{3+}	424.55^{3+}	34.8
410−437	QKFEGDKVIYENEIHALWENPPSNIIFR	trypsin	1129.58^{3+}	1129.58^{3+}	72.1
			847.43^{4+}	847.44^{4+}	
417−437	VIYENEIHALWENPPSNIIFR	trypsin	852.10^{3+}	852.11^{3+}	74.5
423−440	IHALWENPPSNIIFRNSE	Glu-C	1069.04^{2+}	1069.04^{2+}	63.1
			713.02^{3+}	713.03^{3+}	
438−442	NSEFR	trypsin	326.66^{2+}	326.66^{2+}	4.0
447−457	C*HYIRDSMLLR	trypsin	732.35^{2+}	732.37^{2+}	39.7
			488.57^{3+}	488.58^{3+}	
451−483	DSMLLRAHIKSHSSPVASVKPGPLALVLQTYP	Asp-N	1138.28^{3+}	1138.30^{3+}	64.6
			853.96^{4+}	853.98^{4+}	
462−492	SHSSPVASVKPGPLALVLQTYPDISYQRPYR	trypsin	1142.96^{3+}	1142.94^{3+}	68.4
			857.45^{4+}	857.46^{4+}	
493−500	KNEYPLVR	trypsin	509.79^{2+}	509.79^{2+}	34.8
			340.19^{3+}	340.19^{3+}	
494−500	NEYPLVR	trypsin	445.74^{2+}	445.74^{2+}	39.3
501−503	YLR	trypsin	451.27^{+}	451.27^{+}	5.0
504−515	QPIYMEVTVLNR	trypsin	731.89^{2+}	731.89^{2+}	60.9
516−521	NDPNIK	trypsin	350.69^{2+}	350.69^{2+}	3.7
517−524	DPNIKLVL	Asp-N	456.28^{2+}	456.28^{2+}	57.3
534−545	DPASVPQWQIIM	Asp-N	692.85^{2+}	692.85^{2+}	68.8
557−574	TTFHAAN*SSAAHSGHYQR	trypsin	648.62^{3+}	648.30^{3+}	31.2
			486.71^{4+}	486.48^{4+}	
575−578	FDVK	trypsin	508.28^{+}	508.28^{+}	7.1
579−587	TFAFVSESR	trypsin	522.25^{2+}	522.26^{2+}	47.8

with human ZP2, which remains uncleaved postfertilization despite normal cortical granule exocytosis (*23*). Thus, although the structure(s) (protein and/or carbohydrate) to which sperm bind in the zona is unknown at present, it appears that ZP2 must be intact for sperm binding to occur. These observations combined with the ability of mouse sperm to bind to mouse zonae lacking ZP1 (*13*) suggest that mouse sperm bind to a structure(s) that requires only the participa-

Table 5: LC−MS and MS/MS Analysis of Rat ZP3

residues	sequence	enzyme	*m*/*z*(exp)	*m*/*z*(calcd)	elution time (min)
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·5(HexNAc·Hex)·1 HexNAc	trypsin and Asp-N	1855.16^{3+}	1855.17^{3+}	45.8
			1391.64^{4+}	1391.63^{4+}	
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·4(HexNAc·Hex)·3 HexNAc	trypsin and Asp-N	1401.92^{4+}	1401.89^{4+}	45.8
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·4(HexNAc·Hex)·2 HexNAc	trypsin and Asp-N	1801.15^{3+}	1801.15^{3+}	46.4
			1351.11^{4+}	1351.11^{4+}	
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·4(HexNAc·Hex)·1 HexNAc	trypsin and Asp-N	1733.46^{3+}	1733.46^{3+}	47.2
			1300.35^{4+}	1300.35^{4+}	
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·4(HexNAc·Hex)	trypsin and Asp-N	1665.79^{3+}	1665.77^{3+}	48.2
			1249.58^{4+}	1249.58^{4+}	
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·3(HexNAc·Hex)·1 Hex	trypsin and Asp-N	1598.09^{3+}	1598.07^{3+}	48.5
			1198.80^{4+}	1198.80^{4+}	
23−56	qTQWLLPGGTPTP(A?)GSSSPVEVEC*(K/G?)EAELVVTVR·2(HexNAc·Hex)·2 Hex	trypsin and Asp-N	1530.40^{3+}	1530.38^{3+}	48.7
			1148.04^{4+}	1148.04^{4+}	
51−76	LVVTVRRDLFGTGKLVQPGDLTLGSE	Glu-C	924.18^{3+}	924.19^{3+}	64.8
			693.38^{4+}	693.39^{4+}	
57−64	RDLFGTGK	trypsin	447.25^{2+}	447.25^{2+}	30.1
58−69	DLFGTGKLVQPG	Asp-N	616.33^{2+}	616.34^{2+}	55.3
70−84	DLTLGSEGC*QPLVAV	Asp-N	779.87^{2+}	779.89^{2+}	44.0
87−96	DVVRLNAQLH	Asp-N	582.84^{2+}	582.83^{2+}	43.6
			388.89^{3+}	388.89^{3+}	
91−96	LNAQLH	trypsin and Asp-N	348.20^{2+}	348.20^{2+}	29.8
97−105	EC*SSGVQVT	Asp-N	483.73^{2+}	483.71^{2+}	32.0
107−117	DALVYSTFLLH	trypsin and Asp-N	639.84^{2+}	639.84^{2+}	64.3
118−129	DPRPV$\underline{N^{\#}}$GLSILR	trypsin	669.39^{2+}	668.89^{2+}	50.8
			446.59^{3+}	446.26^{3+}	
146−168	N*VSSHPIQPTWVPFSATVSSEEK·2 (HexNAc·Hex)	trypsin and Asp-N	1086.83^{3+}	1086.51^{3+}	43.1
			1629.78^{2+}	1629.26^{2+}	
146−168	N*VSSHPIQPTWVPFSATVSSEEK·2 (HexNAc·Hex)·1 HexNAc	trypsin and Asp-N	1154.53^{3+}	1154.21^{3+}	42.1
146−168	N*VSSHPIQPTWVPFSATVSSEEK·2 (HexNAc·Hex)·2 HexNAc	trypsin and Asp-N	1222.23^{3+}	1221.90^{3+}	41.1
			1832.86^{2+}	1832.34^{2+}	
146−168	N*VSSHPIQPTWVPFSATVSSEEK·2 (HexNAc·Hex)·3 HexNAc	trypsin and Asp-N	1289.93^{3+}	1289.59^{3+}	40.8
			1934.39^{2+}	1933.88^{2+}	
146−168	N*VSSHPIQPTWVPFSATVSSEEK·3 (HexNAc·Hex)	trypsin and Asp-N	1208.55^{3+}	1208.22^{3+}	41.2
146−168	N*VSSHPIQPTWVPFSATVSSEEK·4 (HexNAc·Hex)·1 HexNAc	trypsin and Asp-N	1397.97^{3+}	1397.62^{3+}	40.6
146−168	N*VSSHPIQPTWVPFSATVSSEEK·5 (HexNAc·Hex)	trypsin and Asp-N	1451.96^{3+}	1451.64^{3+}	40.6
146−168	N*VSSHPIQPTWVPFSATVSSEEK·6 (HexNAc·Hex)	trypsin and Asp-N	1573.69^{3+}	1573.35^{3+}	40.3
146−168	N*VSSHPIQPTWVPFSATVSSEEK·7 (HexNAc·Hex)	trypsin and Asp-N	1695.40^{3+}	1695.06^{3+}	40.3
169−174	LAFSLR	trypsin	353.72^{2+}	353.72^{2+}	47.6
175−184	LMEEDWNTEK	trypsin	647.79^{2+}	647.79^{2+}	43.1
185−213	SSPTFHLGEVAHLQAEVQTGSHLPLQLFV	trypsin and Asp-N	1048.19^{3+}	1048.20^{3+}	75.7
			786.39^{4+}	786.41^{4+}	
214−235	DHC*VATPSPLPGQNSSPHHFIV	trypsin and Asp-N	1199.07^{2+}	1199.07^{2+}	53.7
			779.70^{3+}	779.72^{3+}	
236−242	DSHGC*LV	Asp-N	394.17^{2+}	394.17^{2+}	32.7
243−256	DGLSESFSAFQVPR	trypsin and Asp-N	770.38^{2+}	770.38^{2+}	61.5
243−266	DGLSESFSAFQVPRPRPETLQFTV	Asp-N	903.44^{3+}	903.46^{3+}	69.7
259−276	PETLQFTVDVFHFAN*SSR	trypsin	699.35^{3+}	699.01^{3+}	70.8
277−286	NTVYITCHLK	trypsin	596.32^{2+}	596.31^{2+}	46.3
287−296	VAPANQIPDK	trypsin	526.78^{2+}	526.79^{2+}	31.7
287−299	VAPANQIPDKLNK	trypsin	704.40^{2+}	704.40^{2+}	39.0
			469.94^{3+}	469.94^{3+}	
304−315	N*KTSQSWLPVEG	Asp-N	673.82^{2+}	673.34^{2+}	52.8
330−351	N*SSSSEFETHEPAQWSTLVSRN	Asp-N	1247.57^{2+}	1247.07^{2+}	56.1
			832.04^{3+}	831.71^{3+}	
338−351	THEPAQWSTLVSRN	Glu-C	813.42^{+}	813.41^{2+}	35.6
			542.61^{3+}	542.61^{3+}	

tion of ZP2 and ZP3 in zona matrix formation and that ZP2 be intact. Despite the fact that rat zonae are composed of four glycoproteins (*10*), our work shows that rat sperm also bind to mouse zonae comprised solely of ZP2 and ZP3, indicating that one or both of these zonae glycoproteins contain the "minimum structure(s)" to which rodent sperm bind. Whether these structures include the entire glycoprotein or just domains within it remains to be determined.

N- and C-Terminal Processing of the Zona Pellucida Glycoproteins. The N-terminal signal peptide is clipped off the nascent polypeptide chain after the polysome translating the zona transcript docks on the surface of the endoplasmic

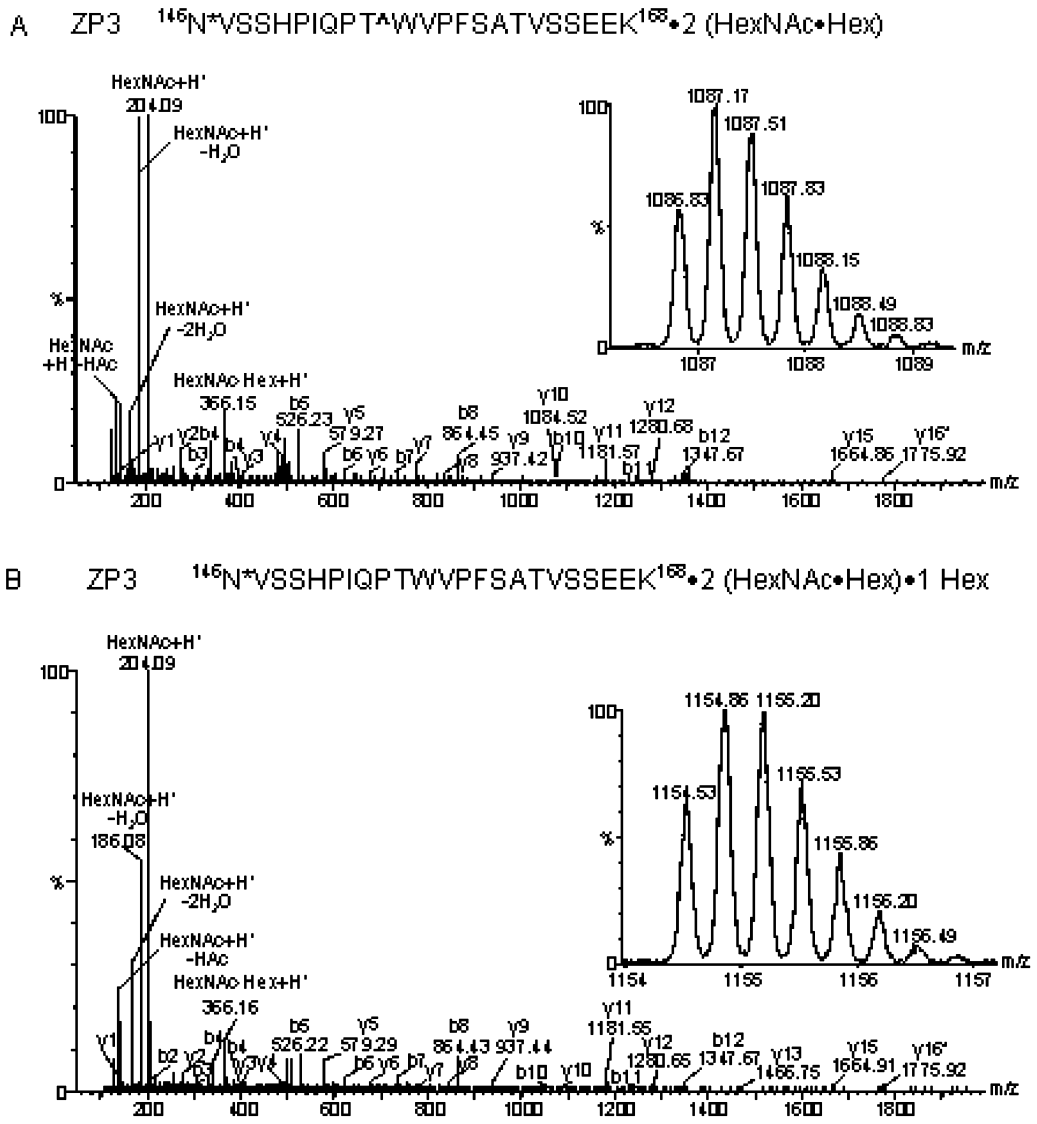

FIGURE 8: O-Glycosylation of rat ZP3. A ^{146}NVSSHPIQPT∧WVPFSATVSSEEK^{168} peptide with multiple sugar attachment was observed in the N/O-deglycosylated sample. (A) The ions at *m/z* 1086.83^{3+} (inset) and 1629.78^{2+} corresponding to the above peptide with an Asn-to-Asp conversion at positions 146 and two HexNAc·Hex molecules at any seven potential sites (a mass increase of 365.13 Da/HexNAc·Hex). The CID spectrum of *m/z* 1086.83^{3+} confirmed the sequence of this peptide, together with sugar signature ions for HexNAc·Hex. (B) The same peptide as in panel A with one additional Hex molecule attached is shown at *m/z* 1154.53^{3+} (inset) and 1731.29^{2+}. The CID spectrum of 1154.53^{3+} showed virtually the same fragmentation pattern as in panel A with carbohydrate marker ions.

reticulum (ER). The lengths of the signal peptides of the zona glycoproteins examined to date by mass spectrometry are 20 (ZP1; rat or mouse), 24 or 34 (ZP2; rat or mouse), 22 (ZP3; rat, mouse, or human), and 28 (ZP4; rat) amino acids. The mature N-terminus of rat ZP2 (Ser25) was not detected by mass spectrometry but was found by Edman degradation sequencing. No other zona proteins were detected by Edman degradation either because their N-termini were blocked by cyclization of glutamine to pyroglutamic acid (ZP1/q[21], ZP3/q[23]) or because the protein was present in a limited quantity (ZP4; q[29]). Moreover, the signal peptide of rat ZP2 (Met1−Leu24) was the only one not predicted by the von Heijne algorithm (i.e., Met1−Ser23).

After synthesis is completed in the ER, the zona proteins are transported through the stacks of the Golgi apparatus. Membrane-anchored zona glycoproteins are released following proteolytic cleavage N-terminal to their transmembrane domain. The mechanism by which the mature C-terminus of the zona glycoproteins is generated remains unknown. What is known is that each C-terminus examined to date ends two amino acids upstream of a dibasic residue that is part of, but distinct from, the consensus cleavage sequence for the proprotein convertase furin [RX(K/R)R]. The C-termini of the secreted rat zona glycoproteins are His544 (ZP1; RH^{544}∧RR), Asn351 (ZP3; RN^{351}∧RR), and Arg473 (ZP4; RR^{473}∧RR). Although the peptide containing the mature C-terminus of rat ZP2 (predicted at Asn619; RN^{619}∧KR) was not found, the C-terminus of rat ZP2 extends until at least Arg618 since the furthest C-terminal tryptic peptide fragment found ends with this residue. These data correspond well with those of native mouse ZP1 (Arg546), ZP2 (Ser633), and ZP3 (Asn351), and recombinant human ZP3 (Asn350) (*6*, *7*).

Intramolecular Disulfide Bond Formation in the Zona Pellucida Glycoproteins. Disulfide bonds are one of the major determinants of stable native conformations in secreted proteins (*24*, *25*). Zonae glycoproteins contain an ∼260-amino acid zona domain. This domain is defined by the

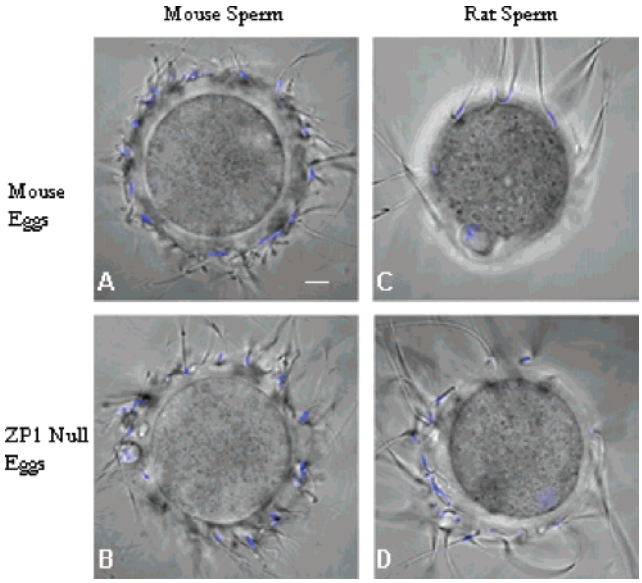

FIGURE 9: Binding of heterologous sperm to normal mouse eggs and mouse ZP1 null eggs. Capacitated mouse or rat sperm (1×10^6 per milliliter) were incubated with 20−30 ovulated normal mouse eggs or ZP1 null mouse eggs for 30 min and then washed with a wide-bore pipet until no more than one to five sperm bound to control mouse two-cell embryos (insets): (A) binding of mouse sperm to mouse eggs, (B) binding of mouse sperm to ZP1 null mouse eggs, (C) binding of rat sperm to mouse eggs, and (D) binding of rat sperm to ZP1 null mouse eggs. The scale bar is 10 μm.

presence of eight conserved cysteine residues, has been observed in multiple proteins, and has been implicated in the polymerization of extracellular matrices (*26*, *27*). The zona pellucida proteins fall into two distinct classes of disulfide-bonded zona domain proteins (*6*, *7*, *10*). Members of zona domain type I have eight conserved cysteine residues followed C-terminally by 45−50 amino acids, while members of zona domain type II contain 10 cysteine residues that end three to five amino acids short of the C-terminus. The first four cysteines in each isoform have similar linkages [Cys1−Cys4 and Cys2−Cys3 (loop within loop)] while the remaining linkages differ. In zona domain I, it is Cys5−Cys7 and Cys6−Cys8 (crossover motif), whereas in zona domain II, it is Cys5−Cys6, Cys7−Cys8, and Cys9−Cys10 (sequential motif). Zona pellucida formation appears to depend on the presence of at least one of each type of zona glycoprotein (*13*, *28*, *29*).

Disulfide mapping of rat ZP1−ZP3 indicates that the intramolecular disulfide bonds formed are the same as those in mouse and human counterparts (*6*, *7*). Rat ZP3 is a type I zona domain protein, while rat ZP1 and ZP2 can be categorized as type II proteins. Only one disulfide-bonded pair of cysteines [Cys447−Cys468 (5−6 sequential motif)] was found in rat ZP1 due to its low abundance compared to those of the other zona glycoproteins. Three pairs of disulfide bonds were located in rat ZP2 [Cys73−Cys91 (N-terminal to the zona domain), Cys354−Cys447 (1−4 loop within loop), and Cys385−Cys406 (2−3 loop within loop)], and two additional pairs were present in another peptide [^{588}GLSSLIYFHC^{597}SALIC^{602}NQASPLC^{609}SVTC^{613}PAPLR^{618} (7−8 and 9−10 sequential motifs)], although their precise linkage pattern was not resolved due to the absence of an appropriate proteolytic cleavage site. A cognate disulfide pair was, however, resolved in mouse ZP2 (Cys608−Cys613) (*6*). Last, three pairs of disulfide-bonded cysteines were found in rat ZP3 [Cys78−Cys98 (2−3 loop-within-loop linkage), Cys216−Cys283 (5−7 crossover linkage), and Cys240−Cys301 (6−8 crossover linkage)]. Four additional cysteine residues that are clustered C-terminal to the zona domain (^{316}DADIC^{320}DC^{322}C^{323}SNGNC^{328}S^{329}) and form intramolecular disulfide linkages in mouse and human ZP3 (*6*, *7*) were not found.

N- and O-Glycosylation of the Zona Pellucida Glycoproteins. Secreted proteins can be cotranslationally N-glycosylated in the endoplasmic reticulum on asparagine residues that reside within a consensus sequence for N-linked glycosylation (NXS/T, where X is any amino acid except proline). Complex and hybrid *N*-glycan chains function in the folding and trafficking of glycoproteins as well as in their protection against spurious proteolytic damage (*30*). All except for one of the potential N-glycosylation sites (Asn227 in mouse ZP3 and Asn226 in human ZP3) are N-glycosylated in the zona pellucida glycoproteins examined thus far (*6*, *7*). As a result, up to a third of the mass of each zona pellucida glycoprotein is due to extensive N-glycosylation of the polypeptide backbone. All asparagines in rat ZP1 (Asn49, Asn68, and Asn369) and ZP2 (Asn72, Asn161, Asn173, Asn206, Asn253, Asn382, and Asn563) appear to be N-glycosylated. Rat ZP3 has four of six asparagines that are definitely N-glycosylated (Asn146, Asn273, Asn304, and Asn330), one asparagine that is not glycosylated (Asn227), and an asparagine whose glycosylation status was not determined (Asn327) since it was in a peptide that was not detected by mass spectrometry. Last, all asparagines (Asn50, Asn74, Asn228, and Asn336) are glycosylated in rat ZP4 (*10*).

O-Linked glycosylation has also been implicated in the processes of protein folding, sorting, and secretion in addition to mediation of cellular adhesion (*31*). No *O*-glycans were detected on rat ZP1, possibly due to sequence coverage of only 52% in this least abundant zona protein. Very little or no O-glycosylation of rat ZP2 occurs despite the presence of 83 potential sites and 82% sequence coverage. Similarly, no O-glycosylation was detected on mouse ZP2, with 94% sequence coverage and only one "unaccounted for" threonine (*6*). As seen in the mouse (*6*), rat ZP3 (59 S/T residues; 91% sequence coverage) contained two patches of O-glycosylation: one at the mature N-terminus and one in the zona domain. The N-terminal patch had a series of sugar attachment sites (six potential sites at Thr24, Thr32, Thr34, Thr38, Thr39, and Thr40). In the zona domain, a peptide containing seven potential sites for O-glycosylation (^{146}NVSSHPIQPT^{155}WVPFSATVSSEEK^{168}) was found to be glycosylated with several glycan chains. A series of up to seven sugar chains (two to seven HexNAc·Hex sugars plus additional sugar residues) were observed, indicating heterogeneous O-glycosylation. Although definitive assignment of rat ZP3 *O*-glycan chains could not be made to specific serine or threonine residues, it is very likely that Thr155 is glycosylated as in mouse ZP3, and its cognate residue (Thr156) in human ZP3 (*6*, *7*). Terminal O-linked glycan residues on Ser332 and Ser334 in ZP3 have been implicated in mediating species-specific sperm−egg binding (*15*). However, just as in previous studies with mouse and human ZP3 (*6*, *7*), no *O*-glycans were detected on these residues in rat ZP3.

In summary, although the molecular basis underlying sperm−zona binding remains elusive, ZP2 and ZP3 appear to contain the minimal domains that are required for mouse and rat sperm binding. We speculate that similarly conserved domains may be involved in sperm−egg binding of other mammalian taxa, including humans.

NOTE ADDED AFTER PRINT PUBLICATION

The graphics and legends of Figures 4−8 were scrambled in the version published on the Web 11/24/05 (ASAP) and in the December 20, 2005, issue (Vol. 44, No. 50, pp 16445−16460). The correct legend is now associated with the correct graphic for each figure in the electronic version published 01/11/06, and an Addition and Correction appears in the February 7, 2006, issue (Vol. 45, No. 5).

ACKNOWLEDGMENT

We thank Dr. Douglas Sheeley and Dr. Jurrien Dean for their critical reading of the manuscript.

REFERENCES

1. Bleil, J. D., and Wassarman, P. M. (1980) Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida, *Dev. Biol. 76*, 185−202.
2. Ringuette, M. J., Chamberlin, M. E., Baur, A. W., Sobieski, D. A., and Dean, J. (1988) Molecular analysis of cDNA coding for ZP3, a sperm binding protein of the mouse zona pellucida, *Dev. Biol. 127*, 287−295.
3. Liang, L.-F., Chamow, S. M., and Dean, J. (1990) Oocyte-specific expression of mouse Zp-2: Developmental regulation of the zona pellucida genes, *Mol. Cell. Biol. 10*, 1507−1515.
4. Epifano, O., Liang, L.-F., and Dean, J. (1995) Mouse Zp1 encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish, *J. Biol. Chem. 270*, 27254−27258.
5. Easton, R. L., Patankar, M. S., Lattanzio, F. A., Leaven, T. H., Morris, H. R., Clark, G. F., and Dell, A. (2000) Structural analysis of murine zona pellucida glycans. Evidence for the expression of core 2-type O-glycans and the Sd(a) antigen, *J. Biol. Chem. 275*, 7731−7742.
6. 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.
7. Zhao, M., Boja, E., Hoodbhoy, T., Nawrocki, J., Kaufman, J. B., Kresge, N., Ghirlando, R., Shiloach, J., Pannell, L., Levine, R., Fales, H., and Dean, J. (2004) Mass spectrometry analysis of recombinant human ZP3 expressed in glycosylation-deficient CHO cells, *Biochemistry 43*, 12090−12104.
8. Akatsuka, K., Yoshida-Komiya, H., Tulsiani, D. R., Orgebin-Crist, M. C., Hiroi, M., and Araki, Y. (1998) Rat zona pellucida glycoproteins: Molecular cloning and characterization of the three major components, *Mol. Reprod. Dev. 51*, 454−467.
9. Lefièvre, L., Conner, S. J., Salpekar, A., Olufowobi, O., Ashton, P., Pavlovic, B., Lenton, W., Afnan, M., Brewis, I. A., Monk, M., Hughes, D. C., and Barratt, C. L. R. (2004) Four zona pellucida glycoproteins are expressed in the human, *Hum. Reprod. 19*, 1580−1586.
10. Hoodbhoy, T., Joshi, S., Boja, E. S., Williams, S. A., Stanley, P., and Dean, J. (2005) Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins, *J. Biol. Chem. 280*, 12721−12731.
11. Hoodbhoy, T., and Dean, J. (2004) Insights into the molecular basis of sperm-egg recognition in mammals, *Reproduction 127*, 417−422.
12. Bedford, J. M. (1977) Sperm/egg interaction: The specificity of human spermatozoa, *Anat. Rec. 188*, 477−487.
13. Rankin, T., Talbot, P., Lee, E., and Dean, J. (1999) Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss, *Development 126*, 3847−3855.
14. Von Heijne, G. (1986) A new method for predicting signal sequence cleavage sites, *Nucleic Acids Res. 14*, 4683−4690.
15. Wassarman, P. M. (2002) Sperm receptors and fertilization in mammals, *Mt. Sinai J. Med. 69*, 148−155.
16. Tulsiani, D. R., Yoshida-Komiya, H., and Araki, Y. (1997) Mammalian fertilization: A carbohydrate-mediated event, *Biol. Reprod. 57*, 487−494.
17. Ellies, L. G., Tsuboi, S., Petryniak, B., Lowe, J. B., Fukuda, M., and Marth, J. D. (1998) Core 2 oligosaccharide biosynthesis distinguishes between selectin ligands essential for leukocyte homing and inflammation, *Immunity 9*, 881−890.
18. Shi, S., Williams, S. A., Seppo, A., Kurniawan, H., Chen, W., Zhengyi, Y., Marth, J. D., and Stanley, P. (2004) Inactivation of the Mgat1 gene in oocytes impairs oogenesis, but embryos lacking complex and hybrid N-glycans develop and implant, *Mol. Cell. Biol. 24*, 9920−9929.
19. Lu, Q., and Shur, B. D. (1997) Sperm from β-1,4-galactosyltransferase-null mice are refractory to ZP3-induced acrosome reactions and penetrate the zona pellucida poorly, *Development 124*, 4121−4131.
20. Thall, A. D., Maly, P., and Lowe, J. B. (1995) Oocyte Galα1,3Gal epitopes implicated in sperm adhesion to the zona pellucida glycoprotein ZP3 are not required for fertilization in the mouse, *J. Biol. Chem. 270*, 21437−21440.
21. Moller, C. C., and Wassarman, P. M. (1989) Characterization of a proteinase that cleaves zona pellucida glycoprotein ZP2 following activation of mouse eggs, *Dev. Biol. 132*, 103−112.
22. Bauskin, A. R., Franken, D. R., Eberspaecher, U., and Donner, P. (1999) Characterization of human zona pellucida glycoproteins, *Mol. Reprod. Dev. 5*, 534−540.
23. Rankin, T. L., Coleman, J. S., Epifano, O., Hoodbhoy, T., Turner, S. G., Castle, P. E., Lee, E., Gore-Langton, R., and Dean, J. (2003) Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs, *Dev. Cell 5*, 33−43.
24. Zapun, A., Jakob, C. A., Thomas, D. Y., and Bergeron, J. J. (1999) Protein folding in a specialized compartment: The endoplasmic reticulum, *Struct. Folding Des. 7*, R173−R182.
25. Fassio, A., and Sitia, R. (2002) Formation, isomerisation and reduction of disulphide bonds during protein quality control in the endoplasmic reticulum, *Histochem. Cell Biol. 117*, 151−157.
26. 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.
27. 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.
28. Rankin, T., Familari, M., Lee, E., Ginsberg, A. M., 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.
29. Rankin, T. L., O'Brien, M., Lee, E., Wigglesworth, K., Eppig, J. J., and Dean, J. (2001) Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development, *Development 128*, 1119−1126.
30. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum, *Annu. Rev. Biochem. 73*, 1019−1049.
31. Hanisch, F. G. (2001) O-Glycosylation of the mucin type, *Biol. Chem. 382*, 143−149.

BI051883F