The zona pellucida is an extracellular matrix consisting of three glycoproteins that surrounds mammalian eggs and mediates fertilization. The primary structures of mouse ZP1, ZP2, and ZP3 have been deduced from cDNA. Each has a predicted signal peptide and a transmembrane domain from which an ectodomain must be released. All three zona proteins undergo extensive co- and post-translational modifications important for secretion and assembly of the zona matrix. In this report, native zonae pellucidae were isolated and structural features of individual zona proteins within the mixture were determined by high resolution electrospray mass spectrometry. Complete coverage of the primary structure of native ZP3, 96% of ZP2, and 56% of ZP1, the least abundant zona protein, was obtained. Partial disulfide bonds were made for each zona protein, and the size of the processed, native protein was determined. The N termini of ZP1 and ZP3, but not ZP2, were blocked by cyclization of glutamine to pyroglutamate. The C termini of ZP1, ZP2, and ZP3 lie upstream of a dibasic motif, which is part of, but distinct from, a proprotein convertase cleavage site. The zona proteins are highly glycosylated and 4/4 potential N-linkage sites on ZP1, 6/6 on ZP2, and 5/6 on ZP3 are occupied. Potential O-linked carbohydrate sites are more ubiquitous, but less utilized.

The zona pellucida is an extracellular matrix surrounding mammalian eggs that functions in taxon-specific gamete binding, provides a post-fertilization block to polyspermy, and protects the developing pre-implantation embryo. The mouse zona pellucida (ZP) is composed of three major glycoproteins (ZP1, ZP2, and ZP3) that are synthesized and secreted by oocytes during a 2–3 week growth period (4). The primary structures of ZP1 (623 amino acids), ZP2 (713 amino acids), and ZP3 (424 amino acids) have been deduced from cDNA (5–7). Each glycoprotein has a signal peptide directing it into a secretory pathway, a ~260 amino acid zona domain containing 8 conserved cysteine residues, and a transmembrane domain near the C terminus followed by a short cytoplasmic tail (8). The zona domain has been observed in multiple proteins (9) and has been implicated in the polymerization of extracellular matrices (10).

During oocyte growth, ZP1, ZP2, and ZP3 traffic through the growing oocyte, and their ectodomains are released from a transmembrane domain at the surface of the cell (11, 12). A conserved hydrophobic patch upstream of the transmembrane domain is required for progression to the cell surface2 and a consensus cleavage site (RX/K/R/R↓) for the proprotein convertase furin is present upstream of the transmembrane domain. Although this site has been implicated in the release of the zona ectodomain (13–15), mutations (RNRR→ANAA, or RNRR→ANGE), do not prevent incorporation of reporter-ZP3 proteins into the zona pellucida in growing oocytes (12, 16) or transgenic mice (12) and secretion of recombinant human ZP3 with a similar mutation (RNRR→ANAA) is not prevented (17).

The three zona proteins are extensively co- and post-translationally modified and a detailed structural analysis of mouse zona pellucida glycans has been reported (18). These observations are of particular interest because of the proposal that sperm bind to ZP3 O-glycans linked to Ser332 and Ser334, and the corollary that their removal by glycosidases released from egg cortical granules prevent sperm binding after fertilization (19). However, there has been controversy as to the nature of the glycans involved and the candidacy of individual terminal sugars as sperm receptors has not been supported by targeted null mutations in mice (8, 18). Moreover, recent genetic studies suggest that sperm binding to the zona pellucida is predicated on the three-dimensional structure of the zona pellucida matrix rather than a specific carbohydrate side chain. Cleavage of ZP2 by a protease released during cortical granule exocytosis that occurs upon fertilization may be sufficient to modify the supramolecular structure of the zona matrix and render it non-permissive to sperm binding (20).

Many of these controversies stem from the paucity of biological material that makes robust biochemical analysis difficult and has prompted reliance on recombinant zona proteins expressed in heterologous systems where processing and modifications may differ from those in mouse oocytes. This report takes advantage of microscale LC-MS to partially characterize mouse ZP1, ZP2, and ZP3 as a mixture in native zonae pellucidae. A hybrid QTOF instrument has the advantages of high mass accuracy, great sensitivity and resolution, and is well suited for detection of low levels of biological materials. Using these technologies we have determined both N and C termini, intramolecular disulfide linkages, and have identified N- and O-glycosylation sites on mouse ZP1, ZP2, and ZP3.
EXPERIMENTAL PROCEDURES

Mass Spectrometric Characterization of Mouse Zona Pellucida Proteins

Materials—Urea, dithiothreitol, iodacetamide (IAA), 4-vinylpyridine (4-VP), and ammonium bicarbonate were purchased from Sigma-Aldrich (St. Louis, MO). Tris (hydroxymethyl)phosphine hydrochloride (TCEP) 0.5 M was obtained from Pierce Biotechnology, Inc. (Rockford, IL). Sequencing grade porcine trypsin was from Promega, Inc. (Madison, WI) and Asp-N was from Roche Applied Science (Indianapolis, IN). All HPLC solvents were of the highest grade commercially available from J. T. Baker (Phillipsburg, NJ). Glycoprope Deglycosylation Kit was obtained from Invitrogen Corp. (San Diego, CA). An anti-rat secondary IgG–horseradish peroxidase was obtained from Jackson ImmunoResearch Laboratories, Inc. (West Grove, PA). All NOVEX gels were obtained from Invitrogen, Carlsbad, CA.

Disulfide Linkage Mapping—Zona pellucidae were isolated from an ovariian homogone using density gradient ultracentrifugation (21). Approximately 20 µg of zona proteins were lyophilized prior to denaturation in 4 ml of 8 M urea in 250 mM Tris-HCl, pH 8.0 at 37 °C for 1 h. Reduction with dithiothreitol (5 mM final concentration) and subsequent alkaliylation with IAA (80 mM final concentration) were performed in the same buffer at 37 °C for 1 h each. To this reaction mixture was added 100 µl of 50 mM ammonium bicarbonate, pH 7.8. The excess reagents including urea, dithiothreitol, and IAA were removed by buffer exchange (5×) using a YM-10 Amicon centrifugation filter device with a MW cutoff of 10 kDa (Millipore Corp., Bedford, MA).

The proteins were re-dissolved in 50 µl of 50 mM ammonium bicarbonate, pH 7.8 and deglycosylated using a Prozyme Glycoprope Deglycosylation Kit. N-glycans were removed using 1 µl of PNGase F (5000 units/ml) for 26 h at 37 °C. After N-deglycosylation, the sample was divided into three equal aliquots and lyophilized. Half of the material was digested in 50 mM ammonium bicarbonate buffer, pH 6.1, prior to O-glycan removal. O-Deglycosylation was performed using 1 µl of the following exoglycosidases: sialidase A (5 units/ml), β(1→4)-galactosidase (3 units/ml), and β-N-acetylglucosaminidase (45 units/ml) ± 1 µl of endo-O-glycosidase (1.25 units/ml) at 37 °C for 36 h. The pH of this sample was raised to 6.5 in the middle of the reaction. The O-deglycosylation samples were subsequently lyophilized, and re-dissolved in 50 mM ammonium bicarbonate buffer, pH 7.8, to give ~10 pmol/ml final concentration of ZP3 in the MIP mix, 1 µl of ZP mix containing 10 pmol of ZP3) was digested in a 10-µl volume consisting of 1 µl of acetonitrile, 7 µl of 50 mM ammonium bicarbonate buffer, pH 7.8, and either 1 µl of trypsin (1 pmol) for 18 h, Asp-N (0.5 pmol) for 18 h, or trypsin (1 pmol) for 48 h followed by Asp-N (0.5 pmol) for an additional 18 h. Trypsin cleaves C-terminal to lysine and arginine; Asp-N cleaves N-terminal to aspartic acid, although infrequent cleavage N-terminal to glutamic acid also has been reported (22).

Disulfide Linkage Mapping—A non-reduced zona protein mixture (20 µl) was denatured in 8 M urea, pH 7.2 at 37 °C for 1 h. Free thiol s of cysteine residues were blocked with 1 M (final concentration) of 4-VP for 36 h. Reduction with dithiothreitol (5 mM final concentration) and subsequent alkylolation with IAA (50 mM final concentration) were performed in the same buffer at 37 °C for 1 h each. To this reaction mixture was added 100 µl of 50 mM ammonium bicarbonate, pH 7.8, and deglycosylated in a 25-µl reaction mixture prepared in an ammonium bicarbonate buffer, pH 7.8 to give ~10 pmol/ml final concentration of ZP3 in the ZP mix. 1 µl of ZP mix containing 10 pmol of ZP3 was digested in a 10-µl volume consisting of 1 µl of acetonitrile, 7 µl of 50 mM ammonium bicarbonate buffer, pH 7.8, and either 1 µl of trypsin (1 pmol) for 18 h, Asp-N (0.5 pmol) for 18 h, or trypsin (1 pmol) for 48 h followed by Asp-N (0.5 pmol) for an additional 18 h. Trypsin cleaves C-terminal to lysine and arginine; Asp-N cleaves N-terminal to aspartic acid, although infrequent cleavage N-terminal to glutamic acid also has been reported (22).

LC-MS Analysis of Protein Digests—Trypsin, Asp-N, and trypsin/Asp-N double digests of ZP mix were analyzed on a Micromass QTOF Ultima Global (Micromass, Manchester, UK) in electrospray mode interfaced with an Agilent HP1100 CapLC (Agilent Technologies, Palo Alto, CA) prior to the mass spectrometer. 2 µl (~2 pmol) of each digest was loaded onto a Vydac C4, MS column (100 × 15 mm; Grace Vydac, Hesperia, CA) and chromatographic separation was performed at 1 µl/min using the following gradient: 0–10% B over 5 min; gradient from 10%–40% B over 60 min; 40–95% B over 5 min; 95% B held over 5 min (solvent A: 0.2% formic acid in water; solvent B: 0.2% formic acid in acetonitrile). A data-dependent analysis (DDA) method collected CID data for the three most abundant peptide ions observed in the preceding survey scan (m/z 300–1990) above a threshold of 10 counts/sec. Collision energy for CID experiments was optimized using peptide standards with a wide mass range (m/z 400–1600) and charge state (+1 to +4) and was typically between 20–65 eV. Data was processed using the MassLynx software package (version 3.5) to generate peak list files before submitting them to in-house licensed Mascot search (24) (bispec.ncbi.nlm, London, UK). Error tolerant searches were performed to consider irregular cleavages and post-translational modifications. In addition, manual data analysis in search of specific ions of interest was carried out. All MS/MS fragment ions were within 50 ppm of their theoretical values determined by the BioLynx Protein/Pepptide Editor and most were within 10 ppm.

Determination of the N Termini of ZP1, ZP2, and ZP3—Virtually all extracellular proteins have N-terminal signal peptides that direct them into secretory pathways and are removed in the endoplasmic reticulum by signal peptidases. A predictive algorithm (33) predicts cleavage of ZP1, ZP2 and ZP3 immedi-
Mass Spectrometric Characterization of Mouse Zona Pellucida Proteins

Fig. 1. Determination of the N termini of native mouse zona proteins.

Asp-N, trypsin, and both were used to map peptides at the N termini using microscale LC-MS analysis. A, the N terminus of ZP1 defined by the Asp-N peptide \( ^{21}\text{qRLHLEPGFEYSY}^{33} \) with pyroglutamate \((q)\) in place of Gln\(^{21}\) exhibits the +2 charged ion at \( m/z \) 811.37 Da. CID spectrum confirms the sequence. B, CID spectrum of the +2 charged precursor ion at \( m/z \) 915.45 Da corresponding to the N-terminal peptide \( ^{35}\text{VSLPQSENPAFPQGLTLIC}^{51} \) of mouse ZP2 derived from sequential trypsin and Asp-N cleavage. Many internal fragment ions near prolines were observed together with partial sequence ions from the peptide. C, the observed masses at \( m/z \) 527.06 (+4 charged) and 702.42 (+3 charged) from tryptic cleavage match the expected value of the N-terminal peptide \( ^{23}\text{qTLWLPQPGTPTPVGSSPVPK}^{43} \) with a pyroglutamate in place of a glutamine.

ately upstream of Gln\(^{21}\), Val\(^{35}\), and Gln\(^{23}\), respectively. Edman degradation sequence confirmed the N terminus of ZP2 (6), but was either imprecise for ZP1 (7) or uninformative for ZP3 (5).

Peptide mapping of ZP1 from Asp-N digestion followed by LC-MS indicated that the N terminus starts at Gln\(^{21}\), which had been converted to pyroglutamate. The CID spectrum (Fig. 1A) of the precursor ion at \( m/z \) 811.37 (inset, calc. 811.39\(^{+}\)) corresponding to the mass of the N-terminal peptide \( ^{21}\text{qRLHLEPGFEYSY}^{33} \) (q = pyroglutamate) indicated the presence of both y and b ion series including \( y_{1-2}, y_3-\text{H}_2\text{O}, y_7, b_{2-6}, b_{6-10}, b_{2-}\text{NH}_3, b_{7-}\text{NH}_3, b_{9-}\text{NH}_3 \). In addition, an ion series \( a_{5-6}, a_8, a_{11} \) as well as immonium ions of tyrosine and phenylalanine were observed. MS data from the combined trypsin/Asp-N digestion revealed the presence of the \( [M+2\text{H}]^{2+} \) ion at \( m/z \) 915.45 (inset, calc. 915.46\(^{+}\)) corresponding to the N-terminal carbamidomethylated peptide \( ^{35}\text{VSLPQSENPAFPQGLTLIC}^{51} \) of ZP2 (Fig. 1B). The CID spectrum of this ion generated many internal fragment ions (PG, PQ, PGT, PGTLI, PQSENPAF, etc.) near proline residues and, together with sequence ions \( y_1, y_2, y_6, a_4, b_{2-}\text{H}_2\text{O}, b_{7-}\text{NH}_3, b_{11}, \) confirmed its identity.

For mouse ZP3, tryptic digestion revealed \( [M+3\text{H}]^{3+} \) and \( [M+4\text{H}]^{4+} \) at \( m/z \) 702.42 and 527.06 that match the N-terminal peptide \( ^{23}\text{qTLWLPQPGTPTPVGSSPVPK}^{43} \), again with a pyroglutamate in place of a glutamine (Fig. 1C). Unfortunately, the low abundance of these multiply charged ions prevented them from being selected for fragmentation (CID). Furthermore, the highly charged state of this peptide is unusual since there is only one basic lysine residue. However, gas phase basicity can promote proton trapping by proline, tryptophan, and glutamine (34, 35) and may account for these observations.

Determination of the C Termini of ZP1, ZP2, ZP3—A potential proprotein convertase (furin) cleavage site (RX\((R/K)\text{R} \)) that lies 35–40 amino acids N-terminal of the transmembrane domain is conserved among the mouse zona proteins and has been implicated in the release of the mature zona ectodomain (13). Because trypsin cuts within the furin site and could have
provided ambiguous results, samples were digested with Asp-N. MS data was obtained from both N-deglycosylated and N/O-deglycosylated zonae pellucidae. For mouse ZP1, we observed a peptide of MH\(^{+}\)11001774.42 Da corresponding to the sequence of 540DSGIARR546 both as a +1 (calc. 774.421 Da) and +2 charged ion at m/z 387.72 (Fig. 2A). This indicates that the C terminus of mouse ZP1 (Arg546) lies two amino acids upstream of the furin cleavage site. Due to the low abundance of these ions, CID data were not obtained.

For ZP2, Asp-N digestion and LC-MS data revealed the presence of a precursor ion of MH\(^{+}\)1649.76 representing the C-terminal peptide 619DSPLCSTCPASLRS633 where Cys 623 and Cys 627 were both carbamidomethylated (calc. MH\(^{+}\)1649.76). The CID spectrum of the +2 charged ion at m/z 825.38 corresponds to the C-terminal peptide 619DSPLCSTCPASLRS633 of ZP2; the C-terminal peptide of ZP3 \(^{351}\)DSSSSQFQHGPRQWSKLVS\(^{351}\) was detected at m/z 636.80 (+4 charged) as well as 848.75 (+3 charged), 0.96 Da higher than expected, demonstrating that Asn\(^{350}\) was replaced by Asp. The CID spectrum confirmed the sequence identity of this peptide (see text).

Disulfide Linkage Mapping—Blocking with 4-VP at pH 7.2 revealed no S-pyridylethylated cysteine-containing peptides in the mixture, suggesting that all cysteines (at least those de-
detected in the digest) participate in disulfide bonding. In the following discussion, the two disulfide bonded peptide chains have been arbitrarily designated as P1 and P2, priming fragmentations that arise from the latter, e.g. \( y_y^+/y_{110.3}^+ \). Because the disulfide bridge is sometimes "reductively" cleaved either between or on each side of sulfur, peptide fragment ions will appear carrying either an SH or SSH at the cysteine site, and these are referred to as \( y_r \) (or \( y_{110.3}^r \)) and \( y_d \) (or \( y_{110.3}^d \)), respectively.

ZP1 forms a homodimer in the native zona pellucida. It has 21 cysteine residues and the potential to form 10 intramolecular disulfide bonds with the remaining cysteine residue available for intermolecular ZP1-ZP1 linkage. However, due to the low abundance of ZP1 in the zona protein mixture only one disulfide-bonded peptide was detected. The low abundances of the +3 and +4 charged ions at \( m/z \) 1013.50 and 765.37 (precursor ion of MH+ 3058.45) CID spectra were obtained, and as expected, both ions disappeared after treatment with tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 1 h. Unfortunately, the reduced ion 2 Da higher was not available to corroborate the reduction. ZP2 has 20 cysteine residues capable of 10 disulfide bonds. Within the zona domain (containing ten cysteines, eight of which
A disulfide bond between Cys102 and Cys104, near the N terminus of ZP2, outside the zona domain was also identified. The +3, +4, and +5 charged ions at m/z 1269.95, 592.71, and 762.35 (MH+ 3807.87) with strong ion intensities correspond to the accurate mass of the S-S-linked peptide 659PNPSV-DVDGSEILNCTYALKDL286 (P1; Asp393 → Asp83 conversion) and 573FPYETCTIK105 (P2). Moreover, the CID spectra of both 952.95+ and 1270.30+ showed a similar fragmentation pattern corresponding to the sequence of both peptides linked via disulfide bonds (Fig. 3B). The presence of y5 ions of P1 from the ion at m/z 952.95+ showed fragmentation up to Cys844 where the disulfide linkage was located. The linkage was confirmed by analysis of the combined trypsin and Asp-N cleavage (data not shown). The ions at m/z 892.44+ and 595.29+ correspond to the mass of the disulfide-linked peptide 680NCTYALKDL286 (Asp83 → Asp63) and 573FPYETCTIK105 (MH+ 1783.81). These ions disappeared upon reduction with TCEP and additional ions corresponding to their reduced forms at m/z 658.36+ [659PNPSV-DVDGSEILNCTYALKDL286 (Asp83 → Asp83) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.

The mature mouse 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 (Table I). In the first pair, masses corresponding to the peptide 448VECLEAELVVTYTSR57 disulfide-linked to 1383VEPYIECR410 (with loss of 2 Da) were observed at m/z 622.83+ and 830.13+ from both the trypsin only and Asp-N/trypsin double digest (data not shown). These ions, however, were not selected for fragmentation by the software. After reduction, these ions vanished while ions at m/z 773.91+ and 472.75+ corresponding to both reduced peptides respectively were detected. In the second pair, a precursor ion of MH+ 3058.45 as detected by its +3 and +4 charged ions at m/z 1020.16+ and 765.37+ corresponds to 659PNPSV-DVDGSEILNCTYALKDL286 (P1) and 551.25+ [573FPYETCTIK105 were generated, further confirming the original disulfide linkage between the two peptides.
Since six out of a total of eight cysteines in mZP3 had been accounted for in disulfide bonding, it seemed reasonable that the last linkage would be between 386DFHGLL242 and 308ACSF303. However, ions corresponding to this linkage calculated as MH⁺ 1214.50 Da were not detected in the double digest sample. A +1 charged ion at m/z 427.16 corresponding to the mass of 300ACSF303 was detected only after TCEP reduction, but not present in the non-reduced digest. Similarly, the +1 and +2 charged ions at m/z 790.31 and 395.66, which correspond to 238DFHGLL242 were only detected after TCEP treatment. This observation implies that Cys301 was originally internally disulfide-bonded twice, as demonstrated by the presence of a +1 charged ion at m/z 988.41 (MH⁺ 3950.68). The conversion of two asparagines to aspartic acids resulted in a mass increase of 1.97 Da, however, the loss of 4.03 Da from the formation of two disulfide bridges caused a net decrease of 2.06 Da. 

Two additional proteolytic cleavages were subsequently carried out. One was the conversion of Asn to Asp at position 371 upon PNGase F treatment. Asn371 was converted to Asp. In Fig. 4B, glycosylation site identification by CID is illustrated for the +3 charged ion of a glycopeptide 69WPSVDTLGSEILNCYDALER292 at m/z 922.74 derived from trypsin digestion. Again, the experimental precursor ion MH⁺ 2766.20 was 0.86 Da higher than the calculated value of this peptide (MH⁺ 2765.34). The y$_{10-15}$ ions unequivocally confirm the conversion of Asn to Asp at position 83 within the N-glycosylation motif [83NCT85]. The presence of y$_7$, y$_8$-H$_2$O, y$_9$-NH$_3$, y$_{10-12}$, b$_{13-14}$ and a$_{14}$ ions further confirm the sequence identity and demonstrate that Asn70 preceding a proline was not N-glycosylated, as predicted. In addition, the +2 charged ion of this peptide at m/z 1383.65 was observed (0.95 Da higher than the calculated mass) and its CID spectrum showed a very similar fragmentation pattern to that of the +3 charged ion (data not shown).

As a result of Asp-N as well as trypsin/Asp-N sequential digestion, Asn217 and Asn264 were also identified as N-glycosylation sites in ZP2. In the case of Asn266, a peptide 260NATHMTLTIPEFPGK278 resulting from the double digest was detected at m/z 829.41 (+2 charged) and 553.27 (+3 charged), a mass increase of 0.98 Da which was further confirmed by the CID spectrum of 829.41 (data not shown). This conversion led to Asp-N cleavage at position 264 which allowed detection of this peptide. The same observation was made with a peptide 217NATGNIHVYQESSLYTVQLEFLSTTGK246 at m/z 1130.89 (+3 charged) derived from the sequential digest that resulted from the Asn-Asp conversion at position 217 for Asp-N cleavage. In both cases, a mass increase of 0.93–0.98 Da was noted from the conversion.

Similarly, trypsin digestion of ZP3 generated five out of six Asp-containing peptides after PNGase F deglycosylation (Table IV). A +4 charged ion at m/z 1046.43 indicates that the C-terminal peptide 306TSQSWLPVEGDADCDCCHGCNSSSQQFQHGP242 was N-glycosylated at both Asn227 and Asn330. Interestingly the observation of a second co-elyuting ion at m/z 1046.16+ implies the presence of another population of the same peptide N-glycosylated at either Asn227 or Asn330 (data not shown). However, no CID data was available to locate the precise glycosylation site on the second ion. The very large tryptic peptide fragment from residue 185 to 256 encompassing the predicted N-glycosylation site Asn227 was not detected. To obtain additional information on this middle region of the ZP3 sequence, Asp-N as well as trypsin/Asp-N sequential digestion was performed. Two Asp-N fragments, 214DHCATPSLPDP-PNSSYPHYTFV235 and 225DPNSSYPHYTFV235 at m/z 817.37+ and 638.29+, the masses of which match the calculated values of these peptides (MH⁺ 2450.14 and 1275.60) clearly indicated the presence of another population of the same peptide N-glycosylated at both Asn227 or Asn330 (data not shown). However, no CID data was available to locate the precise glycosylation site on the second ion. The very large tryptic peptide fragment from residue 185 to 256 encompassing the predicted N-glycosylation site Asn227 was not detected. To obtain additional information on this middle region of the ZP3 sequence, Asp-N as well as trypsin/Asp-N sequential digestion was performed. Two Asp-N fragments, 214DHCATPSLPDP-PNSSYPHYTFV235 and 225DPNSSYPHYTFV235 at m/z 817.37+ and 638.29+, the masses of which match the calculated values of these peptides (MH⁺ 2450.14 and 1275.60) clearly indicated the absence of a predicted N-linked Asn227 residue. Asn330 was found to be N-glycosylated from the tryptic peptide 300ACS-FNK260 showing a mass shift of +0.98 Da and confirmed by the CID spectrum of its +2 charged ion at m/z 364.16 (data not shown). Further confirmation for this N-linked asparagine site came from Asp-N digestion where a peptide 285DKLNKACSF303 was observed at m/z 541.76+ due to the generation of a new cleavage site at position 304 after PNGase F deglycosylation.
The same observation was made with a trypsin/Asp-N fragment 330NSSSSQFQIHGPR342 at m/z 723.34 and 482.56, where Asp-N cleavage occurred at Asn 330 indicated a mass shift of 0.95 Da. As shown in Fig. 4C, low energy CID generated the sequential y ion series y1-y10 and y6-NH3 ions, as well as b2–3 and b2-H2O ions. Hence, the N-terminal Asn 330 of this ZP3 peptide was unambiguously assigned as an N-glycosylation site.

O-linked Glycosylation Sites—Although O-glycans attach to threonines and serines, there is no specific consensus sequence to readily predict potential linkage sites. Instead, monosaccharides must be removed by a series of exoglycosidases (sialidase, -(1–4)-galactosidase, -N-acetylglucosaminidase) until only the Galβ1–3GalNAc core remains attached to the serine/threonine residues. This results in a mass increase of 365.13 Da/core glycan over the basic peptide. Further O-deglycosylation with endo-O-glycosidase removes the core sugar leaving serine and threonine residues unmodified. Shifts in mobility on SDS-PAGE after deglycosylation suggest that ZP1 contains considerably more O-linked carbohydrate side chains than either ZP2 or ZP3 (data not shown) (29), although estimates of glycosylation based on SDS-PAGE are inexact (30). However, due to its low abundance, no mass spectrometric data was obtained on ZP1 O-linked glycosylation.

Two ZP3 domains were identified that contain one or more O-linked oligosaccharide side chains: one at the N terminus (residues 23–43 with 5 potential sites) and the other within the zona domain (residues 144–168 with six potential sites). The concomitant identification of peptides from these domains prior to deglycosylation implies a mixture of ZP3 molecules, some with O-glycans and others without. Multiply charged ions (+3 and +4) at m/z 702.42 and 527.06 of the N-terminal ZP3 peptide 330NSSSSQFQIHGPR342 resulting from trypsin/Asp-N sequential digestion clearly shows that the Asn-Asp conversion at position 330 results in a mass increase of 0.95 Da and a new cleavage site at Asp330. The observation of ions at m/z 723.34 and 482.56 pertaining to this peptide in the N-deglycosylated sample indicated that Ser332–334 were not O-glycosylated.
Thr34, Ser39 are predicted to be glycosylated (probabilities of Ser39, Ser40) appear to have Asn330. H11001 converted to aspartate after PNGase F treatment (\(N^*\)tamate) were detected in the corresponding to this N-terminal peptide with the attachment Glyc/) (36). WVPFR160 were detected at position 146). The CID spectrum of m/z 976.02+ confirmed the sequence by the presence of Y10–s, Y12–11 ions, as well as b10–s, b9 ions. The b8–s, b7 ions demonstrated that Asn146 was converted to an aspartic acid upon PNGase F deglycosylation (Fig. 5A). This suggests that a population of this peptide was either not O-glycosylated, or the labile sugar core structure was lost during MS analysis. The presence of a Gal[\(\beta1\)–3]GalNAc core was detected by the ions at m/z 1046.16+ that represents the charged N-deglycosylated species of ZP3 peptide 306TSQSWLPVEGDADICDCCSHGNCSNSSSQQFQIHGPR of at either Asn327 or m/z 1067.47+ and 772.71+ (corresponding to a mass increase of 365.13 Da). Selected ion chromatograms for ions at m/z 976.02+ (unglycosylated) and m/z 772.71+ (glycosylated) co-eluted during chromatography, which is uncommon for differentially glycosylated species and is more consistent with the labile sugar core structure being lost during MS analysis. The CID spectrum of 773.03+ showed the presence of not only sequence ions resulting from the peptide backbone including b3, b2, H2O, b1 ions, as well as Y1, Y3–4, Y7–8, Y10–13, Y10–12, H2O, Y11–NH2 ions, but also low mass carbohydrate marker ions at m/z 204.09 (GalNAc+H)+, 168.08 [GalNAc2H2O+H]+, 144.08 [GalNAc-HAC+H]+, and 366.14 [Gal[\(\beta1\)–3]GalNAc+H]+ (Fig. 5B). Moreover, these ions were no longer present upon deglycosylation with endoglycosidases, again supporting that this peptide was previously O-glycosylated (data not shown). Unfortunately, even with CID data, we could not determine the exact site of the sugar linkage among the three potential sites (Ser148, Ser149, Thr155) due to the loss of the sugar moiety prior to the peptide backbone cleavage. However, Thr155 is a pre-

### Table III

**LC-MS analysis of mouse ZP3 N-linked glycosylation sites**

The following definitions are in the table. C° is carboxamidomethylated cysteine, M° is methionine sulfoxide, N°XS/T is N-linked asparagine converted to aspartate after PNGase F treatment (+0.984 Da), and N is the non-N-linked asparagine site.

| Residue # | Sequence | Enzymes | m/z exp. | m/z calc. |
|-----------|----------|---------|----------|-----------|
| 69–92     | WPNVVDTLGSEILN°C°TYALDLER | Trypsin | 1383.65+ | 1383.18+ |
| 166–181   | LADENQV°SEM*GWIVK | Trypsin | 922.73+ | 922.45+ |
| 168–181   | DENQNV°SEMGWIVK | Trypsin | 925.42+ | 924.94+ |
| 172–183   | NV°SEM*GWIVKIG | Asp-N | 825.34+ | 824.89+ |
| 182–187   | IGN°GTR | Trypsin | 675.34+ | 674.85+ |
| 184–194   | (G/°G/°TRAHLPLKD) | Asp-N | 675.34+ | 674.85+ |
| 217–246   | N°AGTVHIVQESSLYL | Trypsin | 1130.89+ | 1130.58+ |
| 217–236   | N°AGTVHIVQESSLYTVQL | Trypsin + Asp-N | 1143.56+ | 1143.08+ |
| 264–282   | N°ATHTMTLTFEPGBKLESV(D) | Asp-N | 986.91+ | 985.89+ |
| 264–278   | N°ATHTMTLTFEPGBK | Trypsin + Asp-N | 989.91+ | 988.92+ |
| 382–401   | PALNLDTLLVGN°SSC°QPIFK | Trypsin | 1094.58+ | 1094.08+ |
| 393–401   | N°SSC°QPIFK | Trypsin + Asp-N | 541.25+ | 540.76+ |

### Table IV

**LC-MS analysis of mouse ZP3 N-linked glycosylation sites**

The following definitions are in the table. C° is carboxamidomethylated cysteine, N°XS/T is N-linked asparagine converted to aspartate after PNGase F treatment (+0.984 Da), and N is the non-N-linked asparagine site. 1° indicates the presence of an additional ion at m/z 1046.16+ that represents the charged N-deglycosylated species of ZP3 peptide 306TSQSWLPVEGDADICDCCSHGNCSNSSSQQFQIHGPR of at either Asn327 or Asn330.

| Residue # | Sequence | Enzymes | m/z exp. | m/z calc. |
|-----------|----------|---------|----------|-----------|
| 144–160   | QGN°VSSHIPQPTWVPFR | Trypsin | 978.00+ | 975.50+ |
| 225–235   | DPNSSSPHFIV | Asp-N | 650.98+ | 650.67+ |
| 214–235   | DHCVATPSPLDPNSSPHFIV | Asp-N | 638.29+ | 638.30+ |
| 257–276   | PRPETLQTVDFVFIHAN°SSR | Trypsin | 817.37+ | 817.39+ |
| 259–276   | PETLQFTVDVFIFAN°SSR | Trypsin | 783.71+ | 783.40+ |
| 295–303   | DKLHAC°SF(N°KT) | Asp-N | 588.03+ | 587.80+ |
| 300–305   | AC°SFN°R(T) | Trypsin | 611.33+ | 610.92+ |
| 306–342   | TSQSWLPVEGDADIC°DC°C°SHG | Trypsin | 948.52+ | 948.02+ |
| 330–342   | N°SSSQQFQIHGPR | Trypsin + Asp-N | 699.33+ | 699.01+ |

Table II:

**LC-MS analysis of mouse ZP2 N-linked glycosylation sites**

LC-MS analysis of mouse ZP2 N-linked glycosylation sites was lost during MS analysis. The presence of a Gal[\(\beta1\)–3]GalNAc core was detected by the ions at m/z 1046.16+ that represents the charged N-deglycosylated species of ZP3 peptide 306TSQSWLPVEGDADICDCCSHGNCSNSSSQQFQIHGPR of at either Asn327 or Asn330.
dicted O-linked glycosylation site in mouse ZP3 with a probability of 98% (www.cbs.dtu.dk/services/NetOGlyc/).

Similarly, a /H110012 charged ion at m/z 608.27 eluting early in the chromatogram from the tryptic digest of the N/O-deglycosylated sample corresponds to the mass of the ZP3 peptide /H1100161ATVSSEEK/ with the Gal/1–3GalNAc core attached, presumably to either Thr 162 or one of the two serines at positions 164 and 165. The CID spectrum of this peptide up by 965.13 Da (i.e. the mass of O-linked Galβ1–3GalNAc), were detected in the same sample. CID spectrum of 772.713... showed the presence of sequence ions as well as carbohydrates marker ions. This observation implies that one of the three potential sites (Ser/148, Ser/149, Thr/155) was O-glycosylated. Upon endo-O-glycosidase treatment, which removes the Galβ1–3GalNAc core, these two ions disappeared.

![Diagram](image)

**Fig. 5. O-glycosylation of mouse ZP3.** A, a peptide 144QGDVSSHPQPTWPFR160 was present in the N/O-deglycosylated sample of ZP3. The ions at m/z 976.050 and 650.982 corresponding to the above peptide without any O-sugars were present with an Asn-Asp conversion at position 146. The CID spectrum of m/z 976.050 confirmed the sequence of this peptide: B, additional ions at m/z 1158.568 and 772.713, which shifted the mass of this peptide up by 365.13 Da (i.e. the mass of O-linked Galβ1–3GalNAc), were detected in the same sample. CID spectrum of 772.713 showed the presence of sequence ions as well as carbohydrates marker ions. This observation implies that one of the three potential sites (Ser/148, Ser/149, Thr/155) was O-glycosylated. Upon endo-O-glycosidase treatment, which removes the Galβ1–3GalNAc core, these two ions disappeared.

Similarly, a +2 charged ion at m/z 608.27 eluting early in the chromatogram from the tryptic digest of the N/O-deglycosylated sample corresponds to the mass of the ZP3 peptide 161ATVSSEEK168 with the Galβ1–3GalNAc core attached, presumably to either Thr162 or one of the two serines at positions 164 and 165. The CID spectrum of this ion produced only MH+–(Galβ1–3GalNAc) at m/z 850.41, perhaps due to being subjected to CID late in peak elution when less precursor ion signal is available. However, its low mass carbohydrate marker ions including GalNAc+H+, (GalNAc-H2O)+H+, (GalNAc-2H2O)+H+, and (GalNAc-HAc)+H+, at m/z 204.09, 186.08, 168.08, and 144.07 resembled that of the O-glycosylated peptide described above, indicating that this peptide is clearly O-glycosylated. The lack of peptide ions with sugar moieties attached made it impossible to assign the site of the O-glycan linkage, but based on the predictive algorithm, Thr162 has a 70% probability of being glycosylated.

Earlier studies have described mouse ZP3 as the primary sperm receptor, an activity ascribed to O-glycans attached at Ser332 and Ser334 (37, 38). However, the trypsin/Asp-N digest of
DISCUSSION

The mammalian zona pellucida is a unique biological structure that surrounds growing oocytes, ovulated eggs, and the pre-implantation embryo (39). Although essential for in vivo fertilization and early development, its biochemical characterization has been impeded by the difficulty of purifying adequate quantities of native material. Earlier studies had determined the presence of three major glycoproteins (ZP1, ZP2, ZP3) and their primary structures have been deduced from cDNA (8). More recent genetic studies using null mutations and replacement with human homologues have provided information about the conservation of these sequences (8). More recent genetic studies using null mutations and replacement with human homologues have provided information about the conservation of these sequences (8).

The native protein mixture generated the masses at m/z 723.342+ and 482.563+ as described above (Fig. 4). These masses correspond to the peptide DSSSSQFQHGPR214, where Asp-N cleavage took place at Asn213 due to the Asn-Asp conversion (i.e. a mass shift of -0.95 Da). Since these masses match the calculated masses of this peptide with the replacement of Asn with Asp (MH+ 1445.68) without any prior O-deglycosylation treatment (N-deglycosylated sample), and since the peptide identity was confirmed by CID sequence data, it indicates that neither Ser232 nor Ser338 are O-glycosylated at a measurable level. Because glycosylation at these sites was inferred from previous microsequencing studies (37), we looked specifically for the masses corresponding to various combinations of glycosylation sites using extracted ion chromatograms in the N/O-deglycosylated samples, but did not find them. Thus, to the extent of our mass spectrometric detection (low femtomole levels), we did not observe glycosylation of any potential O-glycosylation sites except an N-terminal cluster (predicted to be Thr32, Thr34, Ser39) and a second cluster in the zona domain (predicted to be Thr155, Thr162).

Peptide Identification—The three zona proteins are distinct from one another with ZP1 and ZP2 more evolutionarily conserved than ZP3 (40). However, as a cohort they share certain common features. Each has a signal peptide to direct it into a secretory pathway and each has an ectodomain that must be released from a transmembrane domain prior to incorporation into the extracellular zona matrix. The native N terminus of each zona protein was determined by mass spectrometry. Both ZP1 (Fig. 6) and ZP3 (Fig. 8) are blocked by a pyroglutamate (pyroGlu21 and pyroGlu23, respectively) and the N-terminal Val235 of ZP2 (Fig. 7) confirms an earlier determination by Edman degradation (6). Thus, the signal peptides of ZP1, ZP2, and ZP3 are 20, 34, and 22 amino acids long, respectively, and the experimentally determined cleavage sites correspond to those of von Heijne’s predictive algorithm (33).

Once directed into the secretory pathway, the zona proteins remain associated with the endomembrane system until they are released at the surface of the oocyte. There has been controversy as to the cleavage site required for release of the ectodomain from the predicted transmembrane domain near the C terminus (12, 14, 15, 17). The mass spectrometric data indicates that the C terminus of ZP1 (Arg364, ZP2 (Ser383), and ZP3 (Asn531) in native zona pellucidae are N-terminal to a dibasic motif (ZP1, Arg547-Arg548; ZP2, Lys634-Arg635; ZP3, Arg572-Arg573). These presumed cleavage sites are part of, but distinct from, a proprotein convertase (furin) site (13) that is imperfectly conserved among zona proteins. The ZP1, ZP2, and ZP3 dibasic motif lies 43, 50, 37 amino acids, respectively, upstream of the mouse protein transmembrane domain and is conserved in all mammalian species examined to date. It has been suggested that similarly positioned C termini in the quail and Xenopus homologues of ZP3 result from cleavage at the proprotein convertase followed by carboxypeptidase trimming of two basic residues (41, 42). The observation that mutation of
the dibasic motif does not preclude secretion and incorporation of mouse ZP3 into the zona pellucida suggests that alternative cleavage sites are available as has been reported for other secreted proteins (43, 44).

Thus, after N- and C-terminal processing, the polypeptide chains of ZP1 (Fig. 6), ZP2 (Fig. 7), and ZP3 (Fig. 8) will have molecular masses of 58, 68, and 36 kDa, respectively. These predictions are in good agreement with the apparent molecular masses observed after N/O-deglycosylation of ZP1 (63 kDa), ZP2 (68 kDa) and ZP3 (39 kDa) in native zonae pellucidae by immunoblot (data not shown) and autoradiography (29). The minor discrepancies may reflect residual O-linked sugars predicted after enzymatic deglycosylation or aberrant migration and are well within estimation errors associated with SDS-PAGE.

Formation of Intramolecular Disulfide Bonds within the Zona Domain—Disulfide linkages are thought to be one of the major factors in stabilizing native conformations of secreted proteins (45, 46). No free cysteine residues were detected in the native zona pellucida proteins and intermolecular disulfide bonds have been observed only in ZP1 (31). A 260 amino acid zona domain with eight conserved cysteine residues is present in ZP1 (amino acids 288–542), ZP2 (amino acids 363–630), and ZP3 (amino acids 45–308) (9). The mass spectrometric data is most complete for the mouse ZP3 zona domain in which four disulfide bonds are defined (Fig. 8). The two N-terminal bonds (Cys46/Cys139; Cys78/Cys98) form 1–4 and 2–3 linkages (loop-within-loop) and the two C-terminal disulfide bonds (Cys216/Cys283; Cys240/Cys301) form 1–3 and 2–4 crossover linkages. The four additional cysteine residues in ZP3 (Cys320, Cys322, Cys323, Cys328) lie C-terminal to the zona domain and form two disulfide bonds, the linkage of which is indeterminate due to their tight clustering within nine amino acid residues.

Although incompletely determined, the formation of disulfide bonds in the zona domain of ZP1 (Fig. 6) and ZP2 (Fig. 7) appear to differ from that of ZP3. The two, N-terminal bonds (Cys365/Cys457; Cys396/Cys417) in the ZP2 zona domain conform with the loop within a loop motif observed in ZP3, but the two disulfide bonds at the C terminus of the ZP2 zona domain (Cys608/Cys613; Cys623/Cys627) do not share the ZP3 crossover motif. Disulfide linkage between the remaining cysteine residues (Cys538, Cys559) in ZP2 zona domain was not determined, but the corresponding residues (Cys449, Cys470) in ZP1 form a disulfide bond. Thus, there appear to be two additional residues (beyond the 8 conserved cysteines) in the zona domain of ZP1 and ZP2 that are not present in ZP3 and disulfide bond formations in the C-terminal half of the ZP2 (and perhaps ZP1) zona domain differ from those of ZP3.

The zona domain has been implicated in forming protein polymers not only in the zona pellucida matrix, but between constituents of the extracellular tectorin membrane found in the inner ear (10, 47). Genetically altered mice lacking ZP1 form a zona matrix composed of ZP2 and ZP3 (48); mice lacking ZP2 form a thinner, more fragile matrix composed of ZP1 and ZP3 (49); but mice lacking ZP3 do not form a zona pellucida (11, 50). Thus, a zona matrix can be formed by either ZP1/ZP3 or ZP2/ZP3 consistent with the necessity of two types of zona domains: one from ZP3 and the other either from ZP1 or ZP2.

Taken together these data suggest that the structure of ZP1 and ZP2 zona domains may be similar to each other and different from that of ZP3.

Glycosylation of Zona Proteins—N-glycosylation plays an essential role in the folding/trafficking of glycoproteins (51, 52), and can only occur at asparagines that have a consensus...
NX(S/T) motif (where X cannot be a proline). O-glycosylation derivatizes the hydroxyl groups of threonine and serine residues and, although there is no particular sequence motif dictating whether glycosylation can take place, flanking amino acids are thought to exert an influence (53, 54). Each of the proteolytically processed mouse zona proteins contains a limited number of potential N-linkage glycosylation sites (ZP1, 4 sites; ZP2, 6 sites; ZP3, 6 sites), but considerably more potential O-linkage sites (ZP1, 82 sites; ZP2, 84 sites; ZP3, 58 sites). Zona glycoproteins were either N- or O-glycosylated as described above to identify glycosylated asparagine, serine, and threonine residues.

Deglycosylation with PNGase F releases the entire N-glycan bound to asparagine residues and by converting the residue to aspartic acid provides an unequivocal mass spectrometric signature of the glycosylation site. All four potential N-linked sites on ZP1 (Asn184, Asn217, Asn264, Asn393) are also occupied (Fig. 7) in accord with early estimates (55, 56). Five of the six potential N-linked sites on ZP3 (Asn146, Asn273, Asn304, Asn327, Asn330) have carbohydrate side chains (Fig. 8), which is somewhat more extensive than earlier reports (57). Only Asn227 on ZP3 was experimentally determined by mass spectrometry and CID not to be glycosylated, perhaps due to inaccessibility or the presence of proline residues immediately upstream and downstream of the consensus motif. Taken together, these data show that all but one asparagine residue within the NX(S/T) consensus motif is N-glycosylated in mature, native ZP1, ZP2, and ZP3. The molecular masses of N-glycans attached to the mouse zona pellucida ranges from 1.6–3.8 kDa (18), and based on the number of side chains it appears that 15–30% of the mass of individual mouse zona proteins is N-linked carbohydrate side chains.

The composition of O-glycans isolated from native mouse zona pellucida has been determined by chromatography and mass spectrometry (18). Although association with individual zona proteins was not reported, O-linked sugars ranged in size from three to six residues, did not include fucose, and the great majority had core-2 type structures, Gal(β1-3)GalNAc, which provides a useful identification tag. We have reasoned that if a peptide is detected prior to deglycosylation or in a N-deglycosylated sample, then it is not O-glycosylated. Conversely, O-glycosylated peptides would only be found after removal of its O-glycans. Exo-O-glycosidases remove O-linked sugars from zona proteins leaving a Gal(β1-3)GalNAc core attached to serine/threonine residues. Endo-O-glycosidase can be used in addition to exo-O-glycosidases to remove the core sugars with no modification of the serine/threonine residues. Thus, in addition to CID data detecting the attached sugar, the presence of the Gal(β1-3)GalNAc tag (365.13 Da), on the serine/threonine residues before, but not after, treatment with Endo-O-glycosidase is useful in identifying O-glycan sites. However, in view of the fact that evidence has been found for loss of at least one type of O-linked sugar (mannose) upon collision in a triple stage quadrupole (58), one must consider the possibility that similar losses of the closely related O-linked GalNAc residue may arise from collisional processes in the source region.

Experimental determination by mass spectrometry of O-linked sites on ZP1 and ZP2 was not successful either due to incomplete coverage (ZP1) or a paucity of O-linked sugars (ZP2). Greater success was obtained with ZP3. Two clusters of O-linked glycosylation were detected on native ZP3 (Fig. 8). One, at the N terminus appears to contain three occupied amino acid residues (predicted to be Thr32, Thr34, Ser39) and there appear to be two clusters of O-linked glycans at the N terminus (predicted at T32, T34, S39) and within the zona domain (predicted at T155, T162). Clusters are indicated by bracket, potential sites by asterisks, and number of glycans by arabic numbers.

Fig. 8. Summary of mouse ZP3. The primary amino acid sequence (single letter code) of ZP3 obtained from the native mouse zona pellucida extends from an N-terminal pyroglutamate (q23) to a C-terminal asparagine (N351) immediately upstream of a dibasic cleavage site. There are eight conserved cysteine (yellow on blue background) residues in the zona domain yellow background) that are disulfide-linked, C46/C139, C78/C98, C216/C253, C240/C301 (solid line) as well four cysteines (C320, C322, C325, C328) that are C-terminal to the zona domain. The linkage of the latter (dotted lines) was indeterminate due to clustering of cysteine residues and the absence of appropriate cleavage sites. Five of the six potential N-linked sites (white on green background) are glycosylated (N146, N273, N304, N327, N330, but not N227) and there appear to be two clusters of O-linked glycans at the N terminus (predicted at T32, T34, S39) and within the zona domain (predicted at T155, T162). Clusters are indicated by bracket, potential sites by asterisks, and number of glycans by arabic numbers.
conditions (confirmed by MS and CID data), which was detected without any prior O-deglycosylation. Additionally, transgenic mice expressing mutant ZP3 (Ser332 → Gly332; Ser334 → Ala334) have normal fertility (61), although the more definitive assessment of their reproductive fitness in the Zp3-null background has not been reported. Whether the N-terminal or zona domain cluster of O-glycans plays a role in sperm binding remains to be determined, but it seems unlikely that they act as the sole sperm receptor given the genetically altered mice in which sperm continue to bind to the zona pellucida despite the cortical granule reaction and the release of putative glycosidases (20).

Acknowledgments—We thank the members of our laboratories for the many useful discussions, the initial help in the project by Stephanie Gill and the critical reading of the manuscript by Dr. Douglas Sheeley.

REFERENCES

1. Talbot, P., Shur, B. D., and Myles, D. G. (2003) Biol. Reprod. 68, 1–9
2. Evans, J. P., and Florman, H. M. (2002) Nat. Cell Biol. 4, Suppl. 1, 857–863
3. Herrler, A., and Becker, H. M. (2000) Cells Tissues Organs 166, 233–246
4. Wassarman, P. M. (1988) Annu. Rev. Biochem. 57, 415–442
5. Ringuette, M. J., Chamberlin, M. E., Baur, A. W., Sobieski, D. A., and Dean, J. (1988) Dev. Biol. 127, 287–295
6. Liang, L.-F., Chamow, S. M., and Dean, J. (1990) Mol. Cell. Biol. 10, 1507–1515
7. Epifano, O., Liang, L.-F., Familiari, M., Moos, M. C., Jr., and Dean, J. (1995) Development 121, 1947–1956
8. Rankin, T., and Dean, J. (2000) Rev. Reprod. 5, 114–121
9. Bork, P., and Sander, C. (1992) FEBS Lett. 300, 237–240
10. Jovine, L., Qu, H., Williams, Z., Litscher, E., and Wassarman, P. M. (2002) Nat. Cell Biol. 4, 457–461
11. Rankin, T., Familiari, M., Lee, E., Ginsberg, A. M., Dwyer, N., Blanchette-Mackie, J., Dragu, J., Westphal, H., and Dean, J. (1996) Development 122, 2903–2910
12. Zhao, M., Gell, L., Ginsberg, A. M., Liang, L.-F., and Dean, J. (2002) Mol. Cell. Biol. 22, 3111–3120
13. Yurewicz, E. C., Hidler, D., Fontenot, G. K., Sacco, A. G., and Harris, J. (1993) Biochim. Biophys. Acta 1174, 211–214
14. Litscher, E., Qu, H., and Wassarman, P. M. (1999) Biochemistry 38, 12280–12287
15. Williams, Z., and Wassarman, P. M. (2001) Biochemistry 40, 929–937
16. Qu, H., Williams, Z., and Wassarman, P. M. (2002) Mol. Biol. Cell 13, 530–541
17. Kiesser, S. M., and Saling, P. (2002) Biol. Reprod. 66, 407–414
18. Easton, R. L., Patankar, M. S., Lattanzio, F. A., Leaven, T. H., Morris, H. R., Clark, G. F., and Dell, A. (2000) J. Biol. Chem. 275, 7731–7742
19. Wassarman, P. M. (2002) Mt. Sinai J. Med. 69, 148–155
20. Rankin, T. L., Coleman, J. S., Epifano, O., Hoofdduy, T., Turner, S. G., Epifano, O., Lee, E., Gore-Langton, R., and Dean, J. (2003) Dev. Cell 5, 33–43
21. Blei, J. D., and Wassarman, P. M. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 6778–6782
22. Lopaticki, S., Morrow, C. J., and Gorman, J. J. (1998) J. Mass Spectrom. 33, 950–960
23. Sechi, S., and Chait, B. T. (1998) Anal. Chem. 70, 5150–5158
24. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Electrophyseres 20, 3551–3567