Hamster zona pellucida is formed by four glycoproteins: ZP1, ZP2, ZP3, and ZP4

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · FEBRUARY 2009

Impact Factor: 4.25 · DOI: 10.1021/pr800568x · Source: PubMed

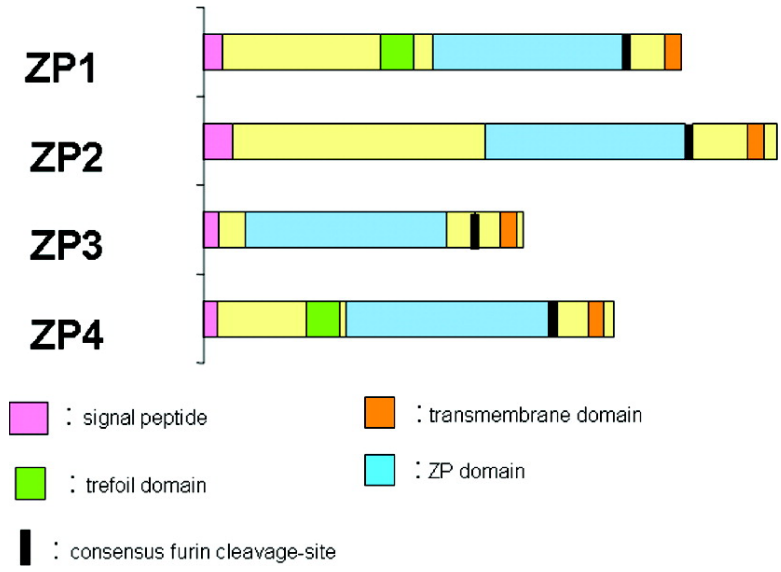
Journal of proteome research

Article

Hamster Zona Pellucida Is Formed by Four Glycoproteins: ZP1, ZP2, ZP3, and ZP4

M. J. Izquierdo-Rico, M. Jime#nez-Movilla, E. Llop, A. B. Pe#rez-Oliva, J. Ballesta, R. Gutie#rrez-Gallego, C. Jime#nez-Cervantes, and M. Avile#s

J. Proteome Res., **2009**, 8 (2), 926-941• DOI: 10.1021/pr800568x • Publication Date (Web): 21 January 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

research articles Journal of proteome research

Hamster Zona Pellucida Is Formed by Four Glycoproteins: ZP1, ZP2, ZP3, and ZP4

M. J. Izquierdo-Rico,[†,‡] M. Jiménez-Movilla,[†,‡] E. Llop,[§] A. B. Pérez-Oliva,[||] J. Ballesta,[‡] R. Gutiérrez-Gallego,[§] C. Jiménez-Cervantes,*[,||] and M. Avilés*[,‡]

Department of Cell Biology and Histology and Department of Biochemistry and Molecular Biology, School of Medicine, University of Murcia, 30100 Murcia, Spain, and Bio-analysis group, Neurophsycopharmacology Program, IMIM-Hospital del Mar & Department of Experimental and Health Sciences, University Pompeu Fabra (UPF), Barcelona Biomedical Research Park (PRBB), 08003-Barcelona, Spain

Received July 25, 2008

The zona pellucida (ZP) is an extracellular glycoprotein matrix that surrounds all mammalian oocytes. Recent data have shown the presence of four glycoproteins (ZP1, ZP2, ZP3, and ZP4) in the ZP of human and rat rather than the three glycoproteins proposed in the mouse model. In the hamster (*Mesocricetus auratus*), it was previously described that ZP was composed of three different glycoproteins, called ZP1, ZP2, and ZP3, even though only ZP2 and ZP3 have been cloned thus far. The aim of the study was to determine whether hamster might also express four, rather than three, ZP proteins. The full-length cDNAs encoding hamster ZP glycoproteins 1 and 4 were isolated using rapid amplification cDNA ends (RACE). The cDNA of ZP1 contains an open reading frame of 1851 nucleotides encoding a polypeptide of 616 amino acid residues. The amino acid sequence of ZP1 revealed a high homology with other mammalian species like human (66%), rat (80%), and mouse (80%). The cDNA of ZP4 contains an open reading frame of 1632 nucleotides encoding a polypeptide of 543 amino acid residues. The deduced amino acid sequence of ZP4 revealed high overall homology with rat (82%) and human (78%). Subsequent mass spectrometric analysis of the hamster ZP allowed identification of peptides from all four glycoproteins. The data presented in this study provide evidence, for the first time, that the hamster ZP matrix is composed of four glycoproteins.

Keywords: zona pellucida • *Mesocricetus auratus* • oocyte • sperm-binding protein

Introduction

Mammalian oocytes are surrounded by an extracellular coat called zona pellucida (ZP) which is involved in different processes during fertilization and early embryo development. This matrix is responsible for the species-specific recognition between gametes, inducing the acrosome reaction, preventing polyspermy and protecting the preimplantation embryo.[1] The composition of the ZP matrix has been elucidated for various species and shown to be composed of 3−6 or more glycoproteins depending on the species[2−7] (Table 1). However, the protein composition of the ZP and the nomenclature used to classify the different ZP proteins are quite confusing. A recent phylogenetic study clarified the nomenclature and the evolution of the ZP gene, especially after the presence of pseudogenes in different species was confirmed.[7] Thus, these authors proposed a classification for the ZP genes comprising six subfamilies: ZPA/ZP2, ZPC/ZP3, ZPB/ZP4, ZP1, ZPAX, and ZPD.[7]

Until very recently, mammalian ZP was believed to be composed of only three glycoproteins of the ZP family, ZP1, ZP2 and ZP3, as first described in the mouse species.[2] However, the description of the complete genome in some species, like human and rat, has resulted in the detection of new proteins expressed in the ZP. Recent studies have revealed that some mammals present a ZP formed by four glycoproteins, e.g., human,[4] rat[5] and bonnet monkey.[6] These four glycoproteins have been designated ZP1, ZP2, ZP3, and ZP4. In human, ZP4 was first identified as an orthologue of mouse ZP1, but later studies detected that the true orthologue of mouse ZP1 was a different gene called ZP1.[8] Further molecular and proteomic approaches identified the four genes and the corresponding proteins, respectively.[4]

The above studies contrast with other studies using a mouse model in which mass spectrometric analysis failed to identify ZP4. The orthologue of the human ZP4 gene is present in the mouse genome as a pseudogene[4] but a functional protein is not expressed. Thus, depending on the species analyzed, ZP is formed by three or four glycoproteins. In species like pig,[3] cow,[9] and dog,[7] the presence of three glycoproteins has also been

* To whom correspondence should be addressed. Department of Cell Biology and Histology, School of Medicine, University of Murcia, Espinardo, 30100 Murcia, Spain. E-mail: maviles@um.es.
† These authors contributed equally to this work.
‡ Department of Cell Biology, University of Murcia.
§ University Pompeu Fabra.
|| Department of Biochemistry and Molecular Biology, University of Murcia.

Published on Web 01/21/2009
10.1021/pr800568x CCC: $40.75

Table 1. Zona Pellucida Glycoproteins Described in the Different Species

scientific name	common name	zp protein	accession number Genbank
		Mammals	
Bos taurus	Cow	ZP2/ZPA	NM_173973
		ZP3/ZPC	NM_173974
		ZP4/ZPB	NM_173975
Callithrix jacchus	White-tufted-ear marmoset	ZP4/ZPB	Y10822
		ZP2/ZPA	Y10767
Canis familiaris	Dog	ZP2/ZPA	NM_001003304
		ZP3/ZPC	NM_001003224
		ZP4/ZPB (partial cds)	AY573930
Felis catus	Cat	ZP2/ZPA	NM_001009875
		ZP3/ZPC	NM_001009330
		ZP4/ZPB	NM_001009260
Homo sapiens	Human	ZP1	NM_207341
		ZP2/ZPA	NM_003460
		ZP3/ZPC	NM_007155
		ZP4/ZPB	NM_021186
Macaca fascicularis	Cynomolgus monkey	ZP2/ZPA	AY222645
		ZP3/ZPC	AY222644
		ZP4/ZPB	AY222647
Macaca radiata	Macaque, Bonnet monkey	ZP1	EF530200
		ZP2/ZPA	Y10690
		ZP3/ZPC	X82639
		ZP4/ZPB	
Macaca mulatta	Rhesus monkey	ZP1	XM_001084628
		ZP2/ZPA	XM_001093570
		ZP3/ZPC	XM_001114760
		ZP4/ZPB	XM_001096846
Mesocricetus auratus	Golden hamster	ZP2/ZPA	AY876920
		ZP3/ZPC	M63629
Microtus brandti	Brandt's vole	ZP3/ZPC	AF304487
Monodelphis domestica	Gray short-tailed opossum	ZP1	XM_001379208
		ZP2/ZPA	XM_001370665
		ZP3/ZPC	XM_001378889
		ZP4/ZPB	XM_001375250
Mus musculus	House mouse	ZP1	NM_009580
		ZP2/ZPA	NM_011775
		ZP3/ZPC	NM_011776
		ZP4/ZPB	XM_001481274
Mustela erminea	Ermine	ZP2/ZPA	AY779765 (partial cds)
		ZP3/ZPC	AY648050
		ZP4/ZPB	AY779766
Nottomys alexis	Hopping mouse	ZP3/ZPC	AY078054
Oryctolagus cuniculis	Rabbit	ZP2/ZPA	L12167
		ZP3/ZPC	U05782
		ZP4/ZPB	M58160
Pan troglodytes	Chimpanzee	ZP1	XM_522022
		ZP2/ZPA	XM_510869
		ZP3/ZPC	XM_519164
		ZP4/ZPB	XM_525105
Papio cynocephalus	Yellow baboon	ZP4/ZPB	AY222646
Pseudomys australis	Plains rat	ZPC/ZP3	AY078055
Rattus norvegicus	Rat	ZP1	XM_001074922
		ZP2/ZPA	NM_031150
		ZP3/ZPC	NM_053762
		ZP4/ZPB	NM_172330
Sus scrofa	Pig	ZP2/ZPA	NM_213848
		ZP3/ZPC	NM_213893
		ZP4/ZPB	NM_214045
Trichosurus vulpecula	Brush-tailed possum	ZP2/ZPA	AF079525
		ZP3/ZPC	AF079524
		ZP4/ZPB	AF263013
		Avian	
Coturnix japonica	Japonese quail	ZP1	AB061520
		ZP2/ZPA	AB295393
		ZP3/ZPC	AB081506
		ZPD	AB301422

Table 1. Continued

scientific name	common name	zp protein	accession number Genbank
Gallus gallus	chicken	ZP1	NM_204683
		ZP2/ZPA	BN000517
		ZP3/ZPC	NM-204389
		ZP4/ZPB	NM_204879
		ZPD	AB114441
		ZPAX	AJ698915
	Fish		
Carassius auratus	Goldfish	ZP2/ZPA	Z72495
		ZP3/ZPC	AF18045
Danio rerio	Zebrafish	ZP2/ZPA	AF331968
		ZP3/ZPC	NM_131331
Oncorhynchus mykiss	Rainbout trout	ZPBa	AF231706
		ZPBb	AF231707
		ZPC	AF271708
Oryzias latipes	Japanese medaka	ZPA	AF128807
		ZPB	AF128808
		ZPCe	AF128809
		ZPCd	AF128811
		ZPCb	AF128812
		ZPCa	AF128813
Pseudopleuronectes americanus	Winter flounder	ZPB	U03674
Salmo salar	Atlantic salmon	ZPC	X93306
		Ba	AY928800
		Bb	AY928798
		ZPB	AJ000665
	Amphibians		
Xenopus laevis	African clawed frog	ZP2/ZPA	AF038151
		ZP3/ZPC	U44952
		ZP4/ZPB	XLU44950
		ZPAX	AF225906
		ZPD	XLU44949
Xenopus tropicalis	Western clawed frog	ZP2/ZPA	NM_203524
		ZP3/ZPC	NM_203522
		ZP4/ZPB	NM_203523
		ZPAX	NM_203520
		ZPD	NM_203521
Bufo arenarum	Common toad	ZP2/ZPA	DQ394072
		ZP3/ZPC	AY185123
		ZP4/ZPB	DQ403815

described, but in these species the proteins are ZP2, ZP3 and ZP4. ZP1 has been identified as a pseudogene in the dog and bovine genome.[7] In non-mammalian species, more than four genes have been detected, for example, in chicken genome[7,10] six genes are present (ZP1, ZP2, ZP3, ZP4 ZPAX, ZPD) and in *Xenopus* genome there are five genes encoding ZP proteins (ZP2, ZP3, ZP4, ZPD, ZPAX).[7]

These observations suggest that the expression of both ZP1 and ZP4 genes represents an ancestral condition present before the mammalian and avian lineages diverged. Thus, ZP1 and ZP4, previously considered orthologs, are in fact paralogs. These two genes come from an ancestry gene through duplication.[7,8,10] Taking all these data together, it seems that the composition and, consequently, the structure of the mammalian ZP is more complicated than expected because, depending on the species: (1) it is formed by three or four glycoproteins; (2) in the three glycoprotein model it can be formed by ZP1, ZP2, and ZP3 or ZP2, ZP3, and ZP4; (3) the protein responsible for the sperm binding is different, for example, ZP3 in mice and ZP4-ZP3 in pig. It is therefore important to know the precise composition of the ZP in all species.

In hamster, characterization of the ZP by SDS-PAGE suggested the presence of just three different glycoproteins called ZP1, ZP2, and ZP3.[11] However, only ZP2 and ZP3[12] have been cloned (GenBank accesion numbers: AY876920 (ZP2), M63629 (ZP3)). In the present study, molecular biology and proteomic analysis demonstrate the presence of mRNA codifying for the ZP1 and ZP4 glycoproteins and several peptides for each ZP1, ZP2, ZP3, and ZP4 glycoprotein are identified. These results indicate, for the first time, that the hamster ZP is formed by four different glycoproteins.

Materials and Methods

Molecular Analysis.

Purification of Hamster Ovarian RNA. Five 11 weeks-old female hamsters (*Mesocricetus auratus*) were injected with 25 IU of pregnant mare serum gonadotropin to stimulate folliculogenesis. The animals were sacrificed 48 h later by overdose of CO_2 and the ovaries were obtained and frozen in liquid nitrogen and kept at −80 °C until use. Total RNA was isolated using the RNeasy Mini Kit (Quiagen) according to the manufacturer's instructions.

Obtaining cDNA and Amplification of the Complete Open Reading Frame of Hamster ZP1 and ZP4 Genes. The first-strand cDNA was synthesized from total RNA with the SuperScript First-Strand Synthesis System kit for RT-PCR (Invitrogen-Life Technologies), according to the manufacturer's instructions.

Table 2. Primers used to amplify the ORF of ZP1 and ZP4 genes

	PCR primers	primers used in race
ZP1	F1: 5′-gaggctggagctgaggattg-3′ Rv1:5′-gcaggcagaggcactacagaag-3′	F1: 5′-caaatggtggccttggacagg-3′ Rv1:5′-gtgaagcgccggtagtgagacc-3′
ZP4	F1: 5′-acacagtgacctccc-3′ Rv1: 5′-gattggccattgtgg-3′ Nested: F2: 5′-acacagtgacctccc-3′ Rv2: 5′-caggaagtaagtgga-3′	F1: 5′-ttgctgtgtccagggatgtgacc-3′ Rv1: 5′-tggaaattggaacagggcaaagg-3′

Hamster ZP1 and ZP4 were partially amplified using the polymerase chain reaction (PCR) by means of specific primers. Two pairs of oligonucleotides were designed based on conserved sequences in mouse and rat ZP (see Table 2) for the specific detection of hZP1 and hZP4. In the case of mouse ZP4, we used a putative sequence deduced from the mouse pseudogene ZP4 (XM_001481274).

PCR amplifications were performed using 3 μL of target cDNA, 0.5 μg of each primer, 200 μM of each dNTP and 1 U of EcoTaq DNA polymerase. PCR for ZP1 was carried out using an initial denaturation cycle for 2 min, and then 30 cycles of 95 °C for 1 min, 67 °C for 1 min and 72 °C for 1 min. The final extension time was 10 min at 72 °C. PCR for ZP4 was carried out in similar conditions except of the annealing temperature (46 °C).

PCR products were analyzed by electrophoresis on 1.5% agarose gels. Four microliters of the PCR reaction mixture were mixed with loading buffer and separated for 90 min at 100 V, before being visualized under UV light using ethidium bromide.

Amplicons were carefully excised from the agarose gels and purified with a QIAquick Gel Extraction Kit Protocol (Quiagen), following the manufacter's protocol. After that, the amplicons were automatically sequenced.

To obtain the full-length hZP1 and hZP4 cDNAs, 5′ and 3′ rapid amplification of cDNA ends (RACE)[13,14] was performed with the BD SMART RACE cDNA Amplification Kit (Clontech), according to the manufacturer's protocol. The gene specific primers (GSP) were designed based on the initial RT-PCR nucleotide sequence. These primers are shown in Table 2. The amplicons obtained were purified and automatically sequenced.

Sequence data encoding the full length ZP1 and the full length ZP4 cDNA are deposited in the GenBank database under accession numbers EU003563 and DQ838550, respectively. The sequences were analyzed to determine the homology with other known sequences using the BLAST program (Basic Local Alignment Search Tool)[15](http://www.ncbi.nlm.nih.gov/blast/). The direct comparison between two sequences was made with the program ALIGN and the multiple alignment of the ZP1 and ZP4 sequences of different species with hamster sequences was carried out using Clustal W (http://www.ebi.ac.uk/clustalw/).

The amino acid sequences were analyzed with the following software packages: "SignalP"[16] to predict the putative signal sequence and cleavages sites, and "NetOGlyc"[17,18] and "NetNglyc"[19] to predict potential N-linked and O-linked glycosylation sites.

Accession numbers of the ZP genes evaluated here are: *Mus musculus* ZP1 (NM_009580); *Mus musculus* ZP4 (putative) (XM_001481274); *Rattus norvegicus* ZP1 (XM_001075428); *Rattus norvegicus* ZP4 (NM_172330); *Mesocricetus auratus* ZP1 (EU003563); *Mesocricetus auratus* ZP4 (DQ838550); *Homo sapiens* ZP1 (NM_207341); *Homo sapiens* ZP4 (NM_021186); *Macaca mulatta* ZP1 (XM_001084628); *Macaca mulatta* ZP4 (XM_001096846); *Macaca radiata* ZP1 (EF530200); *Macaca fascicularis* ZP4 (AY222647); *Pan troglodytes* ZP1 (XM_522022); *Pan troglodytes* ZP4 (XM_525105); *Equus caballus* ZP1 (XM_001493722); *Equus caballus* ZP4 (XM_001490753); *Monodelphis domestica* ZP1 (XM_001379208); *Trichosurus vulpecula* ZP4 (AF263013); *Papio cynocephalus* ZP4 (AY222646); *Callithrix jacchus* ZP4 (Y10822); *Oryctolagus cuniculus* ZP4 (M58160); *Bos taurus* ZP4 (NM_173975); *Sus scrofa* ZP4 (NM_214045); *Felis catus* ZP4 (NM_001009260); *Canis familiaris* ZP4 partial CDS (AY573930); *Mustela erminea* ZP4 (AY799766).

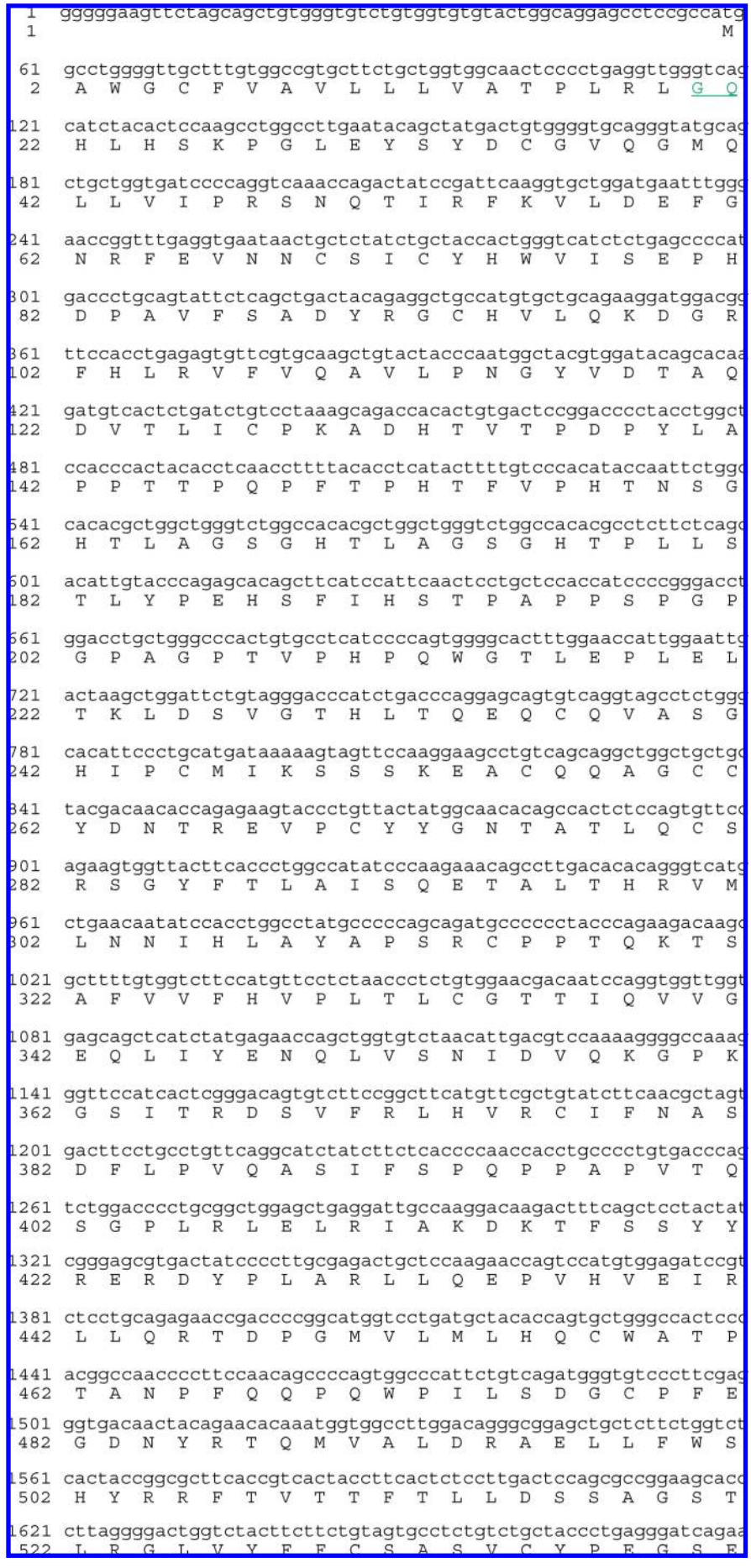
Proteomic Analysis.

Solubilization of Hamster ZP. The ovaries from hamster (n = 20) were trimmed using small scissors under a dissecting microscope to remove fat and connective tissue. The ovaries were homogenized with a Polytron at a setting of 4 for 5 s in 2 mL of homogenization buffer (25 mM triethanolamine-HCl, 150 mM NaCl, 1 mM $MgCl_2$, and 1 mM $CaCl_2$, pH 8.5) supplemented with 12 mg of soybean trypsin inhibitor, 4 mg of bovine testicular hyaluronidase, 1% NP-40 and 4 mg of DNase. The homogenate was placed on ice for 1 min and then homogenized twice as above. Deoxycholic acid solution (0.4 mL of 0.1 g/mL) was added to the homogenate and placed on ice for 10 min. The homogenate was further homogenized by 10 strokes in a glass homogenizer before centrifuging at 4 °C (5 min 10 000× g). The supernatant was discarded and the pellet was washed with, first, TEA buffer (25 mM triethanolamine-HCl, 150 mM NaCl, 1 mM $MgCl_2$, and 1 mM $CaCl_2$, pH 8.5) and then, with high salt PBS buffer (10 mM PBS plus 1 M NaCl, pH 7.0). The sample was finally resuspended in 1 mL of 10 mM phosphate buffer and heat-solubilized at 65 °C for 45 min. Finally, the sample was centrifuged again at room temperature (5 min, 10 000× g) and the supernatant which contained the heat solubilized ZP was stored at −20 °C until use.

SDS-PAGE Electrophoresis and Silver Stain. Purified ZP obtained as described above was dissolved in sample buffer under reducing conditions (5% β-mercaptoethanol (vol/vol)). After boiling for 5 min, samples were separated in a 12% SDS-PAGE.[20] In brief, 4% stacking and 12% separating gels were used and run using 25 mM Tris/0.2 M glycine buffer, pH 8.6, containing 0.1% SDS for 1.5 h at 150 V and room temperature.

After electrophoresis, the gel was fixed in a 5% acetic acid/50% methanol solution for 30 min. The gel was then washed in a 50% methanol solution for 15 min followed by milliQ water for 15 min. Then, the gel was incubated in 0.01% sodium thiosulfate solution for 1 min and after two washes with milliQ water, the gel was incubated with 0.1% silver nitrate solution for 20 min at 4 °C. Finally, the gel was washed twice with milliQ water and incubated with 2% sodium carbonate solution with 250 μL of 35% formaldehyde solution, to visualize the protein bands. A 5 min incubation in 5% acetic acid solution was employed to immobilize proteins.

HPLC−MS Analysis. HPLC−MS/MS analysis was used to identify the hamster ZP proteins. The analysis was carried out on a HPLC-MS system consisting of an Agilent 1100 Series


1 gggggaagttctagcagctgtgggtgtctgtggtgtgtactggcaggagcctccgccatg
1 M
61 gcctggggttgctttgtggccgtgcttctgctggtggcaactcccctgaggttgggtcag
2 A W G C F V A V L L L V A T P L R L G Q
121 catctacactccaagcctggccttgaatacagctatgactgtggggtgcagggtatgcag
22 H L H S K P G L E Y S Y D C G V Q G M Q
181 ctgctggtgatccccaggtcaaaccagactatccgattcaaggtgctggatgaatttggg
42 L L V I P R S N Q T I R F K V L D E F G
241 aaccggtttgaggtgaataactgctctatctgctaccactgggtcatctctgagccccat
62 N R F E V N N C S I C Y H W V I S E P H
301 gaccctgcagtattctcagctgactacagaggctgccatgtgctgcagaaggatggacgg
82 D P A V F S A D Y R G C H V L Q K D G R
361 ttccacctgagagtgttcgtgcaagctgtactacccaatggctacgtggatacagcacaa
102 F H L R V F V Q A V L P N G Y V D T A Q
421 gatgtcactctgatctgtcctaaagcagaccacactgtgactccggacccctacctggct
122 D V T L I C P K A D H T V T P D P Y L A
481 ccacccactacacctcaacctttacacctcatacttttgtcccacataccaattctggc
142 P P T T P Q P F T P H T F V P H T N S G
541 cacacgctggctgggtctggccacacgctggctgggtctggccacacgcctcttctcagc
162 H T L A G S G H T L A G S G H T P L L S
601 acattgtacccagagcacagcttcatccattcaactcctgctccaccatccccgggacct
182 T L Y P E H S F I H S T P A P P S P G P
661 ggacctgctgggcccactgtgcctcatccccagtggggcactttggaaccattggaattg
202 G P A G P T V P H P Q W G T L E P L E L
721 actaagctggattctgtagggacccatctgacccaggagcagtgtcaggtagcctctggg
222 T K L D S V G T H L T Q E Q C Q V A S G
781 cacattccctgcatgataaaaagtagttccaaggaagcctgtcagcaggctggctgctgc
242 H I P C M I K S S S K E A C Q Q A G C C
841 tacgacaacaccagagaagtaccctgttactatggcaacacagccactctccagtgttcc
262 Y D N T R E V P C Y Y G N T A T L Q C S
901 agaagtggttacttcaccctggccatatcccaagaaacagccttgacacacagggtcatg
282 R S G Y F T L A I S Q E T A L T H R V M
961 ctgaacaatatccacctggcctatgcccccagcagatgccccctacccagaagacaagc
302 L N N I H L A Y A P S R C P P T Q K T S
1021 gcttttgtggtcttccatgttcctctaaccctctgtggaacgacaatccaggtggttggt
322 A F V V F H V P L T L C G T T I Q V V G
1081 gagcagctcatctatgagaaccagctggtgtctaacattgacgtccaaaaggggccaaag
342 E Q L I Y E N Q L V S N I D V Q K G P K
1141 ggttccatcactcgggacagtgtcttccggcttcatgttcgctgtatcttcaacgctagt
362 G S I T R D S V F R L H V R C I F N A S
1201 gacttcctgcctgttcaggcatctatcttctcaccccaaccacctgcccctgtgacccag
382 D F L P V Q A S I F S P Q P P A P V T Q
1261 tctggacccctgcggctggagctgaggattgccaaggacaagactttcagctcctactat
402 S G P L R L E L R I A K D K T F S S Y Y
1321 cgggagcgtgactatccccttgcgagactgctccaagaaccagtccatgtggagatccgt
422 R E R D Y P L A R L L Q E P V H V E I R
1381 ctcctgcagagaaccgaccccggcatggtcctgatgctacaccagtgctgggccactccc
442 L L Q R T D P G M V L M L H Q C W A T P
1441 acggccaacccccttccaacagccccagtggcccattctgtcagatgggtgtcccttcgag
462 T A N P F Q Q P Q W P I L S D G C P F E
1501 ggtgacaactacagaacacaaatggtggccttggacagggcggagctgctcttctggtct
482 G D N Y R T Q M V A L D R A E L L F W S
1561 cactaccggcgcttcaccgtcactaccttcactctccttgactccagcgccggaagcacc
502 H Y R R F T V T T F T L L D S S A G S T
1621 cttaggggactggtctacttcttctgtagtgcctctgtctgctaccctgagggatcagaa
522 L R G L V Y F F C S A S V C Y P E G S E

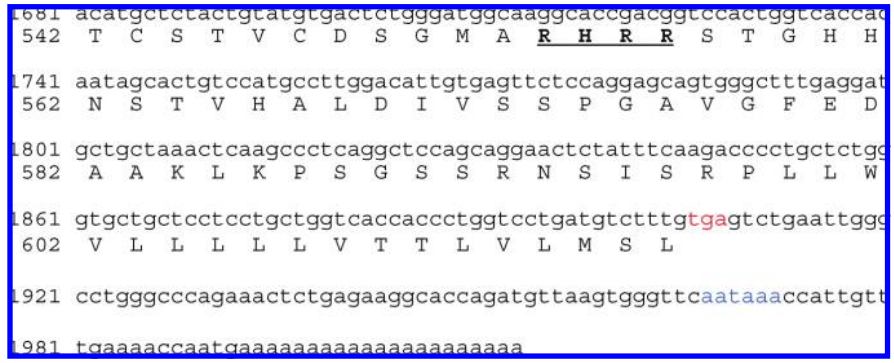


Figure 1. Nucleotide and deduced amino acid sequence of hamster ZP1. The signal peptidase cleavage site is between Gly20 and Gln21 and is marked in green color and underlined. The amino acids in bold and underlined indicate the C-terminal cleavage site. The stop codon is in red and the polyadenylation signal is in blue.

HPLC (Agilent Technologies, Santa Clara, CA) equipped with a μ-wellplate autosampler and a capillary pump, and connected to an Agilent Ion-Trap XCT Plus mass spectrometer (Agilent Technologies, Santa Clara, CA) equipped with an electrospray (ESI) interface.

Samples were separated using SDS-PAGE and bands were cut and washed twice with MilliQ distilled water and then twice with 25 mM ammonium bicarbonate buffer pH 8.5 in 50% acetonitrile for 30 min at 37 °C. The bands were dried by vacuum evaporator, and then incubated with 50 μL of 25 mM ammonium bicarbonate buffer pH 8.5 with 50 mM tris (2-carboxyethyl) phosphine at 60 °C for 10 min. After removing the supernatant, samples were alkylated by adding 25 mM ammonium bicarbonate buffer pH 8.5 containing 100 mM iodoacetamide and allowing to stand for 1 h at room temperature in the dark. The supernatant was removed and the bands were washed with 25 mM ammonium bicarbonate buffer pH 8.5 and then with 25 mM ammonium bicarbonate buffer pH 8.5 in 50% acetonitrile for 15 min at 37 °C each time. After washing, the bands were dried using a vacuum evaporator, and then incubated with 25 mM ammonium bicarbonate buffer pH 8.5 containing 0.3 μg of proteomics grade trypsin (Sigma-Aldrich) for 45 min at 4 °C and finally submitted to digestion for 16 h at 37 °C. The supernatant was collected in a new tube, and the bands were washed with 50 μL of a solution containing 50% acetonitrile and 0.5% TFA and then with 50 μL of acetonitrile for 30 min at 37 °C each time. These washes enhanced the extraction of digested fragments from the gel bands and, afterward, both supernatents were collected in the same tube and dried using a vacuum evaporator.

The separation and analysis of the tryptic digestions of the samples were performed with HPLC−MS. Dry samples (both from solution digestion and in-gel digestion) were resuspended in 10 μL of buffer A, consisting of water/acetonitrile/formic acid (94.9:5:0.1). Samples were injected into a Zorbax SB-C18 HPLC column (5 μm, 150 × 0.5 mm, Agilent Technologies, Santa Clara, CA), thermostatted at 40 °C, at a flow rate of 10 μL/min. After injection, the column was washed with buffer A and the digested peptides were eluted using a linear gradient of 0−80% B (buffer B: water/acetonitrile/formic acid, 10:89.9:0.1) for 120 min.

The mass spectrometer was operated in the positive mode with a capillary spray voltage of 3500 V and a scan speed of 8100 (m/z)/sec from 300 to 2200 m/z. The nebulizer gas (He) pressure was set at 15 psi, whereas the drying gas was set at a flow rate of 5 L/min at a temperature of 350 °C. MS/MS data were collected in an automated data-dependent mode. The most intense ions were sequentially fragmented using collision-induced dissociation (CID) using an isolation width of 2 Da. and a relative collision energy of 35%. Data processing was performed with DataAnalysis program for LC−MSD Trap Version 3.2 (Bruker Daltonik, GmbH, Germany) and Spectrum Mill MS Proteomics Workbench (Agilent Technologies, Santa Clara, CA).

MALDI-TOF MS. Peptide samples were dissolved in water, mixed with the corresponding matrix solution, <1 μL of these preparations applied to the MALDI target, and allowed to dry at room temperature. A solution of sinapinic acid (10 mg/ml) in ACN/water/TFA (50:50:0.1 v/v/v) was chosen for protein analyses, and a solution of α-cyano-4-hydroxycinnamic acid (20 mg/ml) in ACN/water/TFA (70:30:0.1 v/v/v) for peptide analyses. Experiments were carried out on a Voyager-DE STR Biospectrometry workstation (Applied Biosystems), equipped with a N_2 laser (337 nm). Samples were measured in reflectron mode to indentity molecular formulas based on precise mass measurements. External calibrations of the spectrometer were performed with standard peptides from the Sequazyme Peptide Mass Standards Kit (PerSeptive Biosystems). Recorded data were processed with Data Explorer Software (Applied Biosystems).

Results

Analysis of ZP1 and ZP4 cDNA Sequences. Using the 3′ and 5′ RACE technology, full-length hamster ZP1 and ZP4 cDNAs were obtained (Figure 1 and 2) from the total RNA prepared from hamster ovaries. The sequences were submitted to GenBank with the accession numbers EU003563 (ZP1) and DQ838550 (ZP4). The amplified sequence of ZP1 is 2013 nucleotides long and contains a single open reading frame (ORF) of 1851 nucleotides. The ATG initiation codon was predicted with Pedersen and Nielsen algorithm[21] and was found to be associated with vertebrate initiator codons.[22] This sequence contains a stop codon (TGA) in positions 1906−1908 and a polyadenylation signal (AATAAA) in positions 1967−1972.

The amplified sequence of ZP4 is 1767 nucleotides long and contains an ORF of 1632 nucleotides. The ATG initiation codon is associated with vertebrate initiator codons.[22] This sequence contains a stop codon (TGA) in positions 1663−1665 and a polyadenylation signal (AATAAA) in positions 1747−1752.

Predicted Amino Acid Sequence of Hamster ZP1 and ZP4. The ORF of ZP1 encodes a polypeptide 616 amino acids long (Figure 1) with a theoretical molecular weight of 67.915 kDa. A signal peptide of 20 amino acids with a cleavage site between Gly20 and Gln21 was predicted with Bendtsen et al. algorithm.[23]

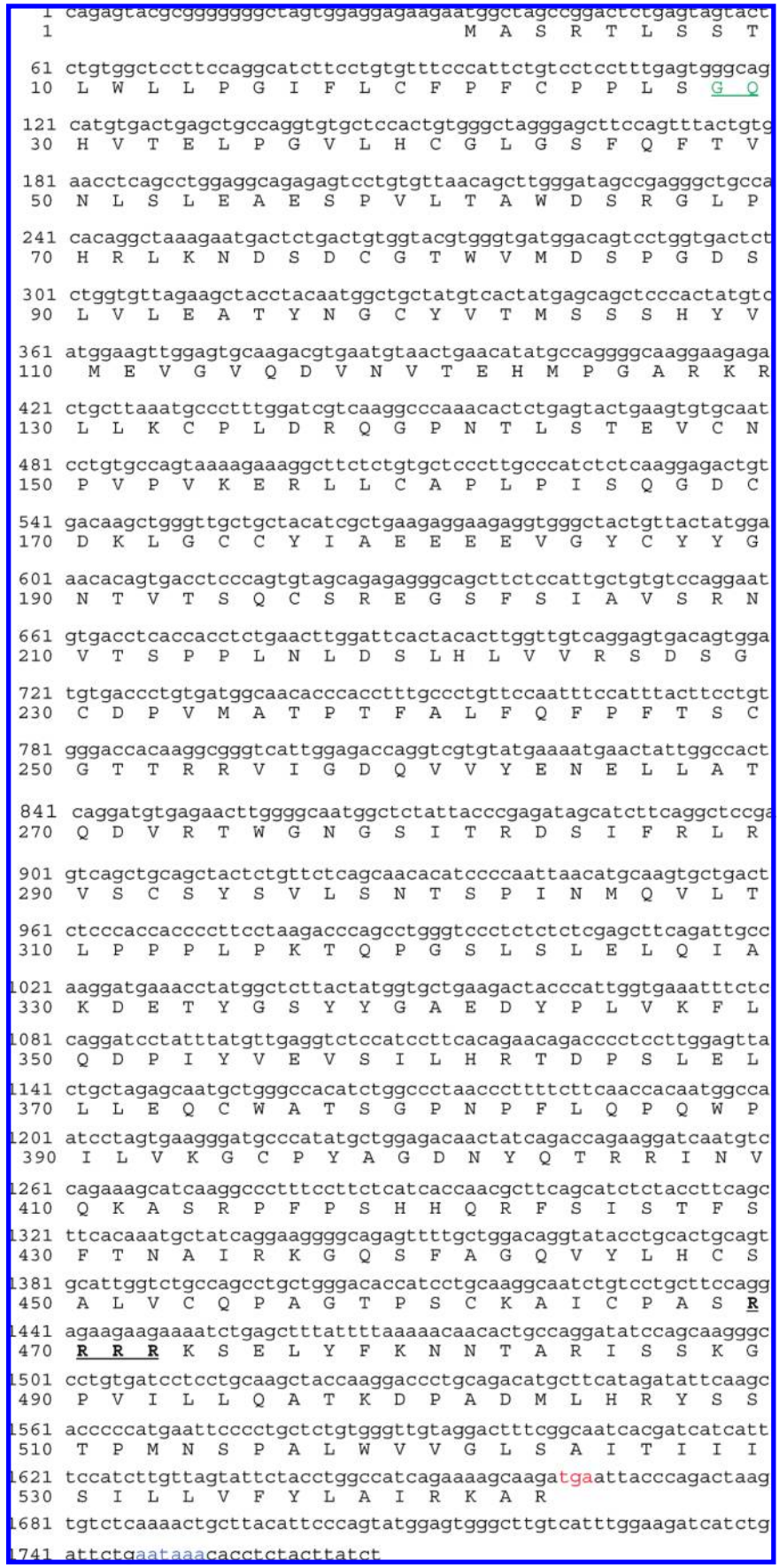


Figure 2. Nucleotide and deduced amino acid sequence of hamster ZP4. The signal peptidase cleavage site is between Gly28 and Gln29, is marked in green color and is underlined. The amino acids in bold and underlined indicate the C-terminal cleavage site. The stop codon is in red and the polyadenylation signal is in blue.

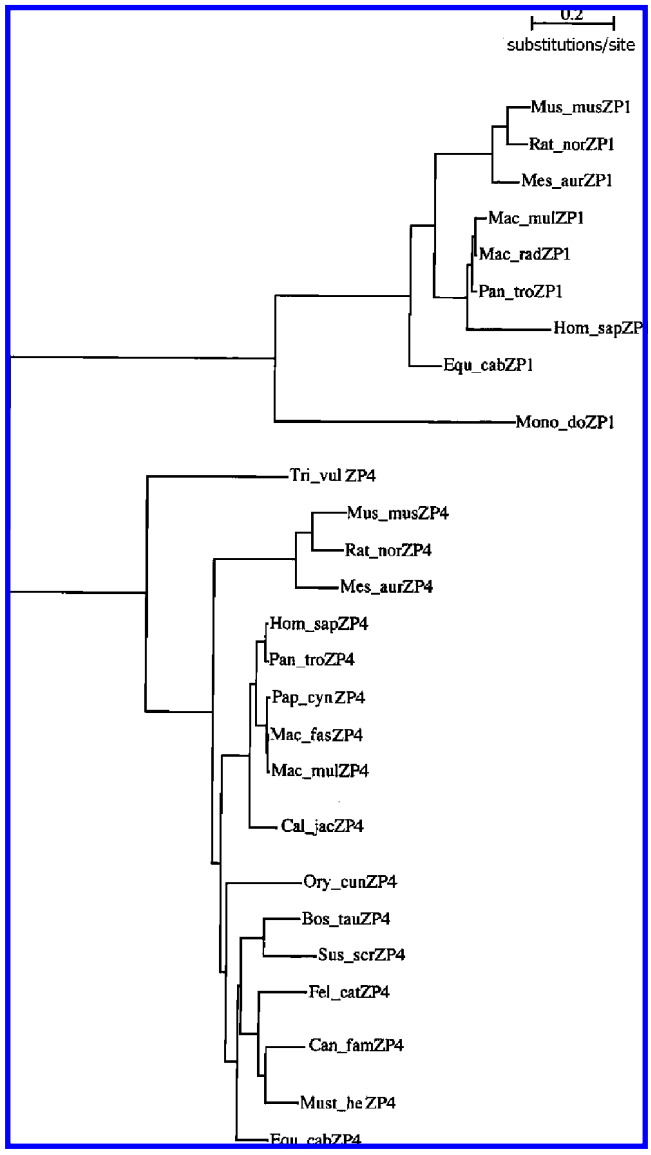


Figure 3. Phylogenetic relationship of ZP1 and ZP4 proteins. The tree shows phylogenetic relations between ZP1 and ZP4 of the different mammalian species. Alignments were performed with mega3 sofware and the tree was constructed using the maximum likelihood method.

The ORF of ZP4 encodes a polypeptide 543 amino acids long (Figure 2) with a theoretical molecular weight of 59.946 kDa. The putative signal peptide is cleaved between Gly28 and Gln29.

The sequences showed a high degree of hydrophobicity at the N-terminal (signal peptide) and C-terminal region. The latter correspond to the transmembrane domain (TMD) between Leu599 and Leu613 (ZP1) and between Leu517 and Ile539 (ZP4) which is followed by a cytoplasmic tail.[24] A basic amino acid domain (^{553}Arg-His-Arg-Arg^{556} in ZP1 and ^{469}Arg-Arg-Arg-Arg^{472} in ZP4) upstream of the TMD may serve as a consensus furin cleavage site (CFCS).[25]

These data could reflect the secretory pathway of ZP1 and ZP4 and their targeting to the extracellular matrix of hamster oocytes. The molecular mass of the processed peptide (with no signal peptide and cleavage at His554 in ZP1 and at Arg470 in ZP4) was calculated to be 59.200 kDa (ZP1) and 48.767 kDa (ZP4).

The molecules have a conserved ZP domain which is present in most sequences of envelope glycoproteins of many species. In the ZP1 sequence, this domain is 272 residues long (^{279}Gln-Gly^{550}) and in the ZP4 sequence 274 residues long (^{195}Gln-Ser^{468}) while in both the domain have 10 Cys residues, which form part of the signature of this domain.

Upstream of the ZP domain, a trefoil domain is present and contains 45 residues (^{234}Glu-Thr^{278}) in ZP1 and 49 residues (^{146}Glu-Ser^{194}) in ZP4. This domain is characteristic of the ZPB family and is a region rich in cystein amino acids.

A total of 102 potential O-glycosylation sites were predicted in the ZP1 sequence and 86 were predicted in the ZP4 sequence.[17,18] On the other hand, 3 potential N-glycosylation sites are present in mature hamster ZP1 (Asn49, Asn68, and Asn379) and 6 potential N-glycosylation sites are present in mature hamster ZP4 (Asn50, Asn74, Asn118, Asn209, Asn277, and Asn299).

Comparison of Hamster ZP1 and ZP4 with Other Oocyte ZP Glycoproteins. Hamster ZP1 and ZP4 were compared with ZP1 and ZP4 proteins from different mammalian species. The phylogenetic relationships obtained indicated that both new hamster ZPs are most closely related to mouse and rat ZP1 and ZP4, respectively (Figure 3).

Alignments of hamster ZP1 and ZP4 with ZP proteins of other species are shown in Figures 4 and 5, respectively. The nucleotide sequences show a high degree of homology with the ZP1 and ZP4 of other mammals. The nucleotide sequence of hamster ZP1 is 79% identical to human ZP1, 84% to rat ZP1, and 86% to mouse ZP1, whereas hamster ZP4 is 78% identical to human ZP4 and 82% to rat ZP4.

The amino acid sequence of hamster ZP1 is 66% identical to human ZP1, and 80% to rat and mouse ZP1. When analyzed, the homology of hamster ZP4 with other ZP4 proteins showed 62% identity with human ZP4 and 73% with rat ZP4.

Interestingly, the cysteine residues of the ZP domain are conserved in hamster, rat, mouse and human in the mature protein, and can be considered evidence of a similar tertiary structure between the proteins of different species.

The consensus furin cleavage site (CFCS) seems to be conserved in the different species both for ZP1 and ZP4. Although the sequences are not identical, a dibasic consensus motif is maintained.

Potential N-glycosylation sites for ZP1 (Asn49, Asn68, Asn379) are conserved in rat, mouse, and hamster and have been demonstrated by proteomic analysis[26] to be occupied, at least in rat and mouse. In human, only two sites (Asn68 and Asn379) are conserved albeit that no information on site occupancy is available. When analyzing the potential N-glycosylation sites of hamster ZP4, we observed that Asn50 and Asn74 are conserved in rat ZP4. Furthermore, these sites have been shown to be occupied in the latter case.[5] Both Asn74 and Asn209, but not Asn50, are conserved in human, cow and pig ZP4.

The O-glycosylated region demonstrated in rat ZP4 (Ser293, Ser295, Ser298, Ser301, Thr309) is conserved in hamster ZP4,[5] as seen in MS experiments.

Proteomic Analysis of the Hamster Zona Pellucida Glycoproteins. The expression of the two new hamster ZP proteins was studied by means of proteomic analysis. For this purpose, hamster ovaries were isolated and dissected as described. After homogenization, the hamster ZP was heat-solubilized and separated by SDS-PAGE followed by silver staining of the gel (Figure 6). Subsequently, different protocols were employed to further analyze the individual segments of the gel. Samples were either analyzed directly from the gel or electro-eluted prior to further processing. Gel segments were reduced and alkylated, trypsinized and analyzed by MALDI-

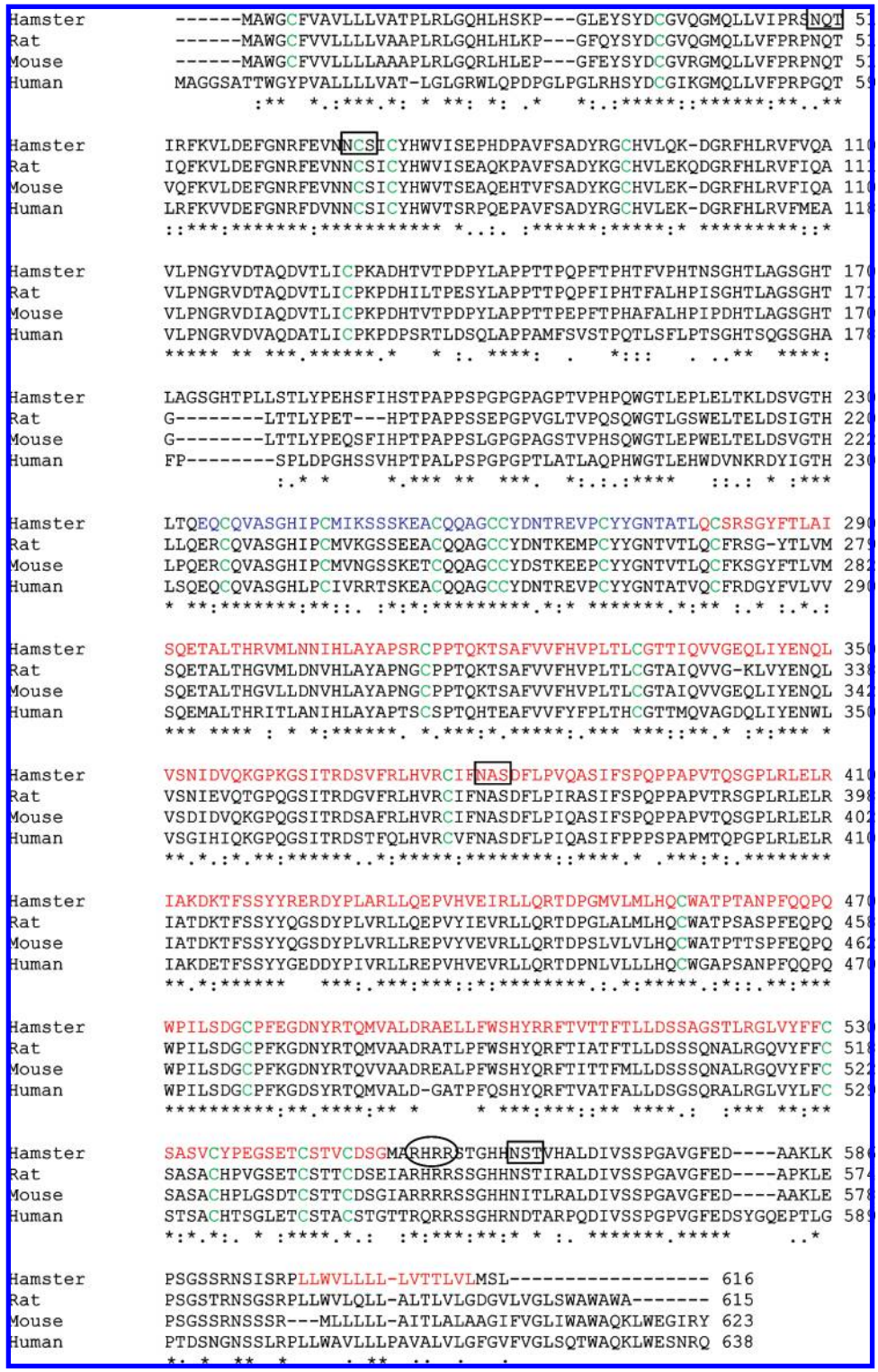


Figure 4. Comparison of amino acid sequences of ZP1 from hamster, mouse, rat and human. The deduced amino acid sequence of hamster ZP1 was aligned with the ZP1 sequences of the other species using the Clustal W program. The accession numbers of the sequences used are the following: hamster ZP1 (ABS86997), rat ZP1 (O54766), mouse ZP1 (NP_033606), and human ZP1 (P60852). Identical amino acids are marked by an asterisk. Colon (:) represents conserved residues and dot (.) represents semiconserved residues. The potential signal peptidase cleavage is between G20 and Q21. The ZP domain is shown in red. The trefoil domain is shown in blue. The consensus furin cleavage-
site is marked with a circle. Potential N-glycosylation sites are marked by a square. The transmembrane domain is in orange. The cysteine residues are in green.

TOF peptide mass finger printing and LC−ESI−MS/MS. Figure 6 shows an example of the different protein bands detected by silver staining and a summary of the peptides identified is included in Table 3. ESI−MS/MS spectra of some peptides corresponding to hamster ZP1 and ZP4 are shown in Figure 7. Peptides corresponding to hamster ZP1 were detected in bands

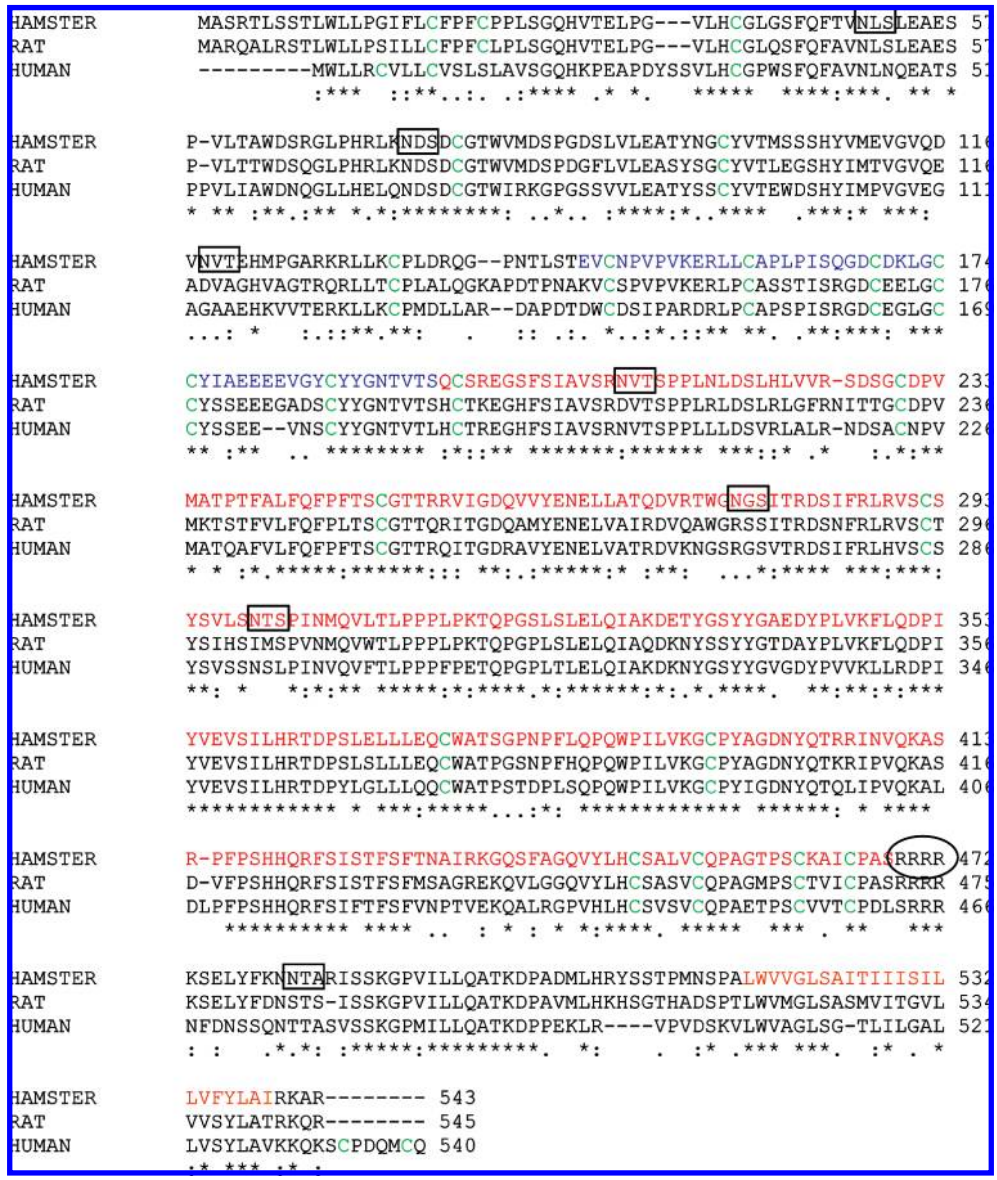


Figure 5. Comparison of amino acid sequences of the ZP4 from hamster, rat, and human. Identical amino acids are marked by an asterisk. Colon (:) represents conserved residues, and dot (.) represents semiconserved residues. The potential signal peptidase cleavage is between G29 and Q30.The ZP domain is shown in red. The trefoil domain is shown in blue. The consensus furin cleavage-site is marked with a circle. Potential N-glycosylation sites are marked by a square. The transmembrane domain is in orange. The cysteine residues are in green.

4 (~90 kDa) and 7 (~65 kDa). A total of seven different peptides were identified in the different analyses yielding a sequence coverage of 12.6% with respect to the primary sequence as derived from the gene (ABS86997). From a random search in over 5.5×10^6 sequences, only ZP1 from *Mesocricetus auratus* was identified with a significant ($P < 0.05$), degree of probability, based on the Mowse score. None of the identified peptides contained an N-glycosylation site, suggesting that all three described consensus sequences may be occupied in the mature glycoprotein. On the other hand, 13 out of the 102 predicted O-glycosylation sites were contained in the identified peptides, from which it can be deduced that these residues are either not glycosylated or, alternatively, suffer from incomplete site-occupancy. The peptide sequence 593−616 of hamster ZP1 NSISRPLLWVLLLLLVTTLVLMSL was also encountered. However, this peptide corresponds to the immature version of ZP1 and was excluded from the identifier peptides.

Three peptides corresponding to hamster ZP2 were detected in bands 1 and 2 (MW between 182−115 kDa, Figure 6). The fact that only 5.1% of the nonprocessed sequence could be identified from the largest glycoprotein was attributed to the fact that it contains 10 potential N-glycans and 120 potential O-glycans (6 of which are at least partially unoccupied), which affects the effectivity of the proteolytic digestion and alters the analytical properties of the resulting peptides. A total of eight peptides from hamster ZP3 could be convincingly identified in bands 7, 8 and 9 (MW between 64.2−48 kDa, Figure 6). This corresponds to 19.2% of the sequence as derived directly from the gene (P23491). Again, none of the indentified peptides contained an N-glycosylation site whereas 15 of the 77 potential O-glycosylation sites were contained in these sequences. Finally, peptides corresponding to hamster ZP4, as described by the gene-derived sequence (ABH06548), were detected in band 3 (from ~182 to ~115 kDa) as well as band 9 (from ~64.2

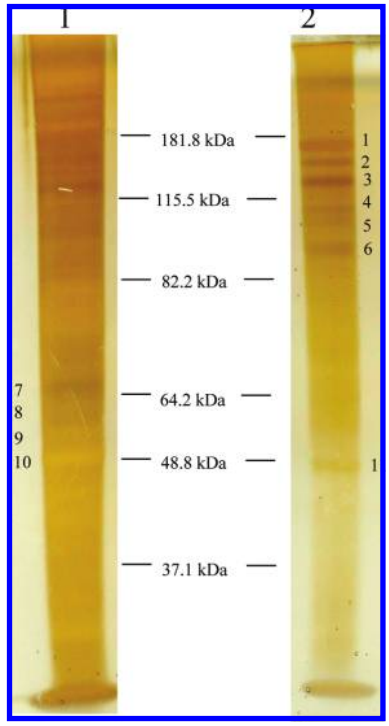


Figure 6. Silver-stained SDS-PAGE gel containing purified proteins from ZP of hamster ovaries. Proteins detected by silver staining were numbered, excised, and submitted for peptide analysis and protein identification as described in Materials and Methods.

Table 3. Peptides Identified by Proteomic Analysis

glycoprotein	peptides	sequence	mass $[M+]^+$
ZP1	EVPCYYGNTATLQCSR	267–282	1641.7359
	SGYFTLAISQETALTHR	283–299	1894.9657
	VMLNNIHLAYAPSR	300–313	1598.8471
	DKTFSSYYR	414–422	1166.5476
	LLQEPVHVEIR	431–441	1332.7634
	AELLFWSHYRR	495–505	1477.7699
	AELLFWSHYR	495–504	1321.6687
ZP2	HPSFPVTVSCDENEVR	42–57	1815.8330
	TFGGYQVTIR	106–115	1141.6000
	QGFNFLIDTR	196–205	1210.6215
ZP3	LTSSVEVECLEAELVVTVSR	38–57	2173.1200
	DLFGTGK	58–64	737.3828
	LVFSLRLMEENWNTEK	169–184	2009.0160
	LMEENWNTEK	175–184	1293.5780
	NTIYITCHLK	275–284	1205.6347
	VTPANQTPDELNK	285–298	1426.7172
	YQAHGVSQWPKSASR	334–348	1701.8455
	YQAHGVSQWPK	334–344	1300.6433
ZP4	EGSFSIAVSR	199–208	1052.5371
	DETYGSYYGAEDYPLVK	331–347	1969.8701
	RVIGDQVVYENELLATQDVR	254–273	2317.2146
	VIGDQVVYENELLATQDVR	255–273	2161.1135
	TQPGSLSLELQIAK	317–330	1484.8318

to ~48 kDa). The fact that ZP4 coelutes with ZP3 and ZP2 has potentially hampered its premature disclosure. A total of five different peptides could be sequenced by MS-MS analysis (Figure 8) covering 11.2% of the nonprocessed sequence, and permitted the protein to be identified. As with ZP1 and ZP3, also for ZP4 none of the identified peptides contained an N-glycosylation site for which full occupation of the seven identified Asn residues could be anticipated. Yet, 9 out of the 86 (from a total of 92) potential O-glycosylation sites were contained in the sequenced peptides, indicating either the absence of glycosylation in these serine and threonine residues, or incomplete site-occupancy. The appearance of ZP1 and ZP4 in the high molecular weight fraction might well correspond to the fully glycosylated structures. In contrast, the appearance of ZP1 and ZP4 in the lower-molecular weight fraction may be explained by a nonglycosylated variant or a truncated variant. Much more material will be required before an unequivocal explanation can be offered.

Discussion

Mammalian fertilization is a complex process, which involves interactions among various proteins of the oocyte with other proteins in the spermatozoa. The binding site in the oocyte is the ZP. So the determination of the composition of the zona pellucida and precise functions of each glycoprotein are topics of great interest. Using the mouse model system, the structural and functional significance of individual ZP glycoproteins has been extensively investigated.[2,27−29] Recent studies have described the presence of four glycoproteins in some species, including human, rat, or bonnet monkey,[4−6] and suggest the need for a reinterpretation of the numerous electrophoretic studies on ZP in other species. Such revision is seriously complicated by the variable patterns of glycosylation observed in ZP in different species. Furthermore, additional functional aspects may interfere with the expression of particular proteins. For example, bioinformatic analysis has indicated that, as in other species, the mouse has four genes. However, comparative sequence analysis revealed that the mouse ZPB gene has acquired a number of mutations in its sequence that avoids the synthesis of a functional protein from DNA. These hypotheses were backed-up by extensive MS analysis that failed to identify ZP4.[30]

In this paper, we have studied the hamster ZP using molecular and proteomic approaches to determine the precise composition of this extra-cellular matrix of the oocyte. The presence of three different proteins in hamster ZP (hZP1, hZP2, and hZP3) has been previously described by SDS-PAGE.[31,32,11] These three proteins appear as two fairly diffuse bands that cover a wide molecular mass range (i.e., from ~50 kDa to ~185 kDa). Wassarman et al. (1990) also described the mRNA that encodes hZP3 glycoprotein and results of *in vitro* competition assays strongly suggested that hZP3 was the hamster sperm receptor and, thus, was functionally analogous to mouse ZP3. Still, a description of the cDNA encoding the full-length hZP2 and hZP3 was not made until a decade later (Koyama et al., 2005, Kinloch et al., 2002). However, no solid evidence was produced on the cDNA for hZP1, not to mention the presence of a fourth member of the ZP glycoprotein family. In our studies we investigated the genetic background of the hZP family and present here the first evidence for the existence of four glycoproteins in hamster ZP: ZP1, ZP2, ZP3 and ZP4. We obtained a full-length cDNA for hamster ZP1 and ZP4. The analysis of the sequences indicated that these are complete coding regions: they have an ORF, an initiation codon and a termination codon. The 3′ UTR regions includes a polyadenylation signal.

A computer homology search with the GenBank database revealed significant homology of the ZP1 and ZP4 sequences with the ZP glycoproteins reported in other mammalian species, including human. The basic structure of the new proteins is similar to the other ZP glycoproteins previously

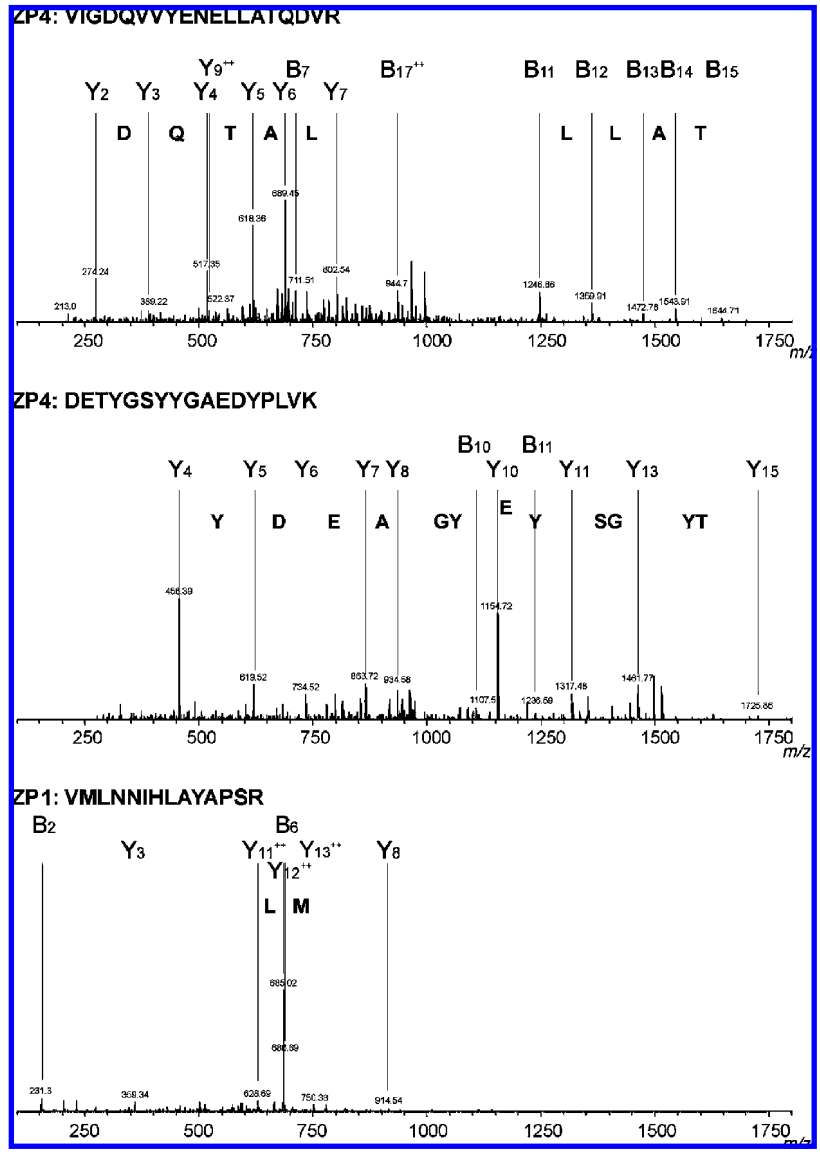


Figure 7. ESI-MS/MS spectra of ZP4 peptide 255–273 (VIGDQVVYENELLATQDVR) in the top panel, the ZP4 peptide 331−347 (DETYGSYYGAEDYPLVK) in the center panel, and the ZP1 peptide 300–313 (VMLNNIHLAYAPSR) in the lower panel.

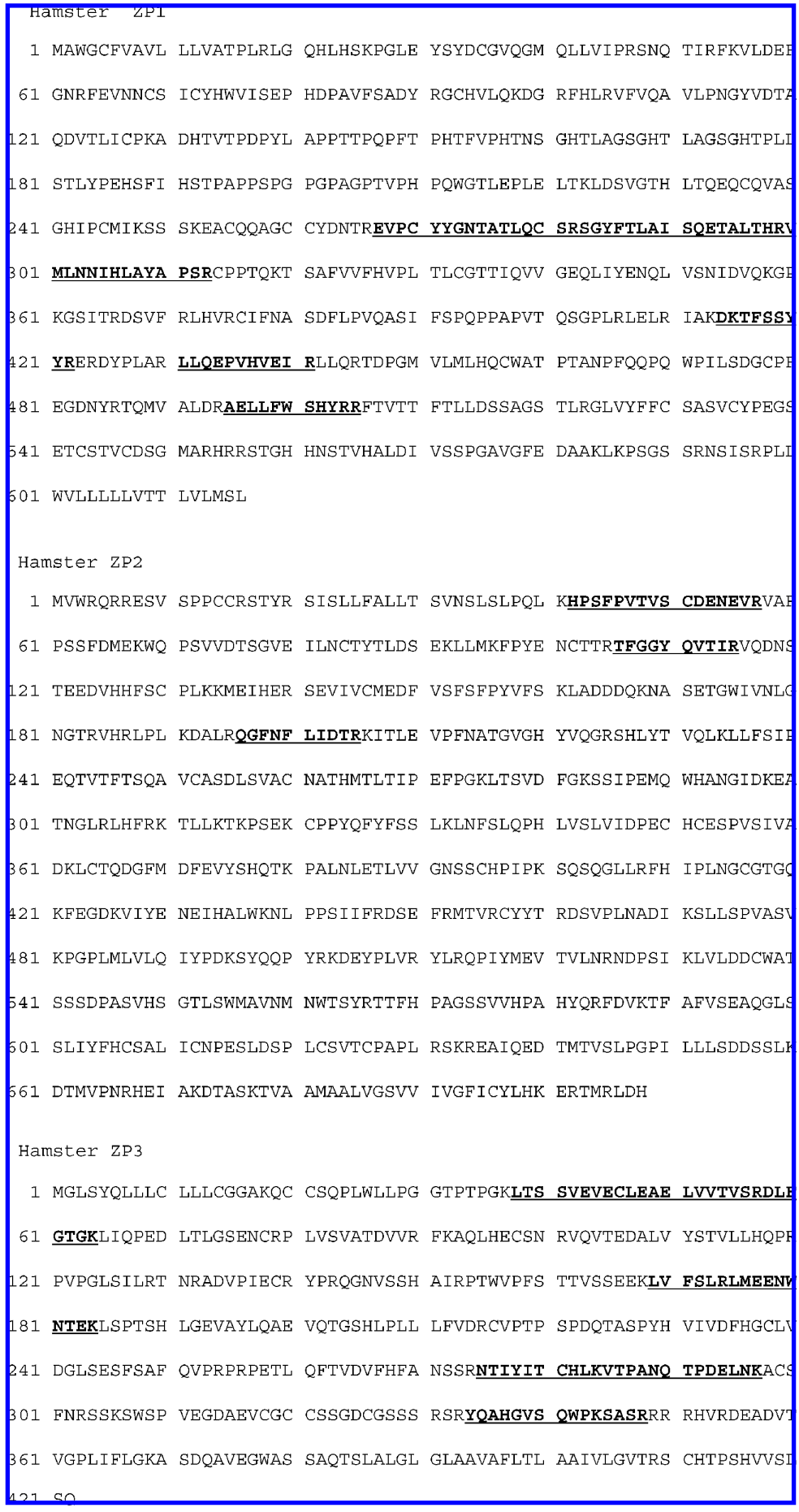
described, with the highest degree of homology in the so-called ZP domains. ZP proteins from different species share these conserved motifs and domains, showing a similar predicted folding with a central ZP domain.[33] The presence of a signal peptide, trefoil domain and transmembrane domain is also ubiquitous. The conservation of C-terminal CFCS and the TMD in different mammalian ZP proteins may reflect the constitutive secretory pathway involved in their biosynthesis.[34]

A comparison with other mammalian ZP demonstrates that the putative sites of glycosylation are conserved in hamster. In hamster ZP1 the potential N-glycosylation sites (Asn49, Asn68, Asn379) are conserved as in rat and mouse. In hamster ZP4, the potential N-glycosylation sites Asn50 and Asn74 are conserved as in rat ZP4, while Asn74 and Asn209 are conserved as in human, cow, and pig ZP4. On the other hand the O-glycosylated region that contains a O-glycan described by MS-MS in rat ZP4 is conserved in hamster ZP4.

Assuming a furin type processing similar to that reported for other species, secreted hamster ZP1 and ZP4 would correspond to the region between ^{21}Gln-Arg^{554} and ^{29}Gln-Arg^{470}, respectively, as a result of the elimination of the signal peptide and TMD.

Peptides of the four hamster mature proteins, ZP1, ZP2, ZP3, and ZP4 were detected by means of peptide mass finger printing and MS-MS analyses, confirming for the first time that the four distinct genes are effectively expressed and follows the discovery of four proteins in the human and rat ZP,[4,26] also using proteomics approaches. The similar composition (four proteins) of hamster and human ZP make the hamster a good model for analyzing the structure and function of human ZP.

Zona Pellucida Glycoproteins Heterogeneity (Glycoforms). Many types of studies have been previously carried out in an attempt to assess the biochemical constitution of the ZP of the mammalian oocyte.[11,35,36] In these studies the ZP has been described as a highly porous envelope constructed of glycoprotein units. Its histochemical staining properties,[37] and ability to bind lectins[38,39] and incorporate radio-labeled sugars[40,41] strongly suggest that the zona pellucida is a typical glycocalyx



Hamster ZP1

1 MAWGCFVAVL LLVATPLRLG QHLHSKPGLE YSYDCGVQGM QLLVIPRSNQ TIRFKVLDEF
61 GNRFEVNNCS ICYHWVISEP HDPAVFSADY RGCHVLQKDG RFHLRVFVQA VLPNGYVDTA
121 QDVTLICPKA DHTVTPDPYL APPTTPQPFT PHTFVPHTNS GHTLAGSGHT LAGSGHTPLI
181 STLYPEHSFI HSTPAPPSPG PGPAGPTVPH PQWGTLEPLE LTKLDSVGTH LTQEQCQVAS
241 GHIPCMIKSS SKEACQQAGC CYDNTR**EVPC YYGNTATLQC SRSGYFTLAI SQETALTHRV**
301 **MLNNIHLAYA PSR**CPPTQKT SAFVVFHVPL TLCGTTIQVV GEQLIYENQL VSNIDVQKGE
361 KGSITRDSVF RLHVRCIFNA SDFLPVQASI FSPQPPAPVT QSGPLRLELR IAK**DKTFSSY**
421 **YR**ERDYPLAR **LLQEPVHVEI R**LLQRTDPGM VLMLHQCWAT PTANPFQQPQ WPILSDGCPE
481 EGDNYRTQMV ALDR**AELLFW SHYRR**FTVTT FTLLDSSAGS TLRGLVYFFC SASVCYPEGS
541 ETCSTVCDSG MARHRRSTGH HNSTVHALDI VSSPGAVGFE DAAKLKPSGS SRNSISRPLI
601 WVLLLLLVTT LVLMSL

Hamster ZP2

1 MVWRQRRESV SPPCCRSTYR SISLLFALLT SVNSLSLPQL K**HPSFPVTVS CDENEVR**VAE
61 PSSFDMEKWQ PSVVDTSGVE ILNCTYTLDS EKLLMKFPYE NCTTR**TFGGY QVTIR**VQDNS
121 TEEDVHHFSC PLKKMEIHER SEVIVCMEDF VSFSFPYVFS KLADDDQKNA SETGWIVNLG
181 NGTRVHRLPL KDALR**QGFNF LIDTR**KITLE VPFNATGVGH YVQGRSHLYT VQLKLLFSIE
241 EQTVTFTSQA VCASDLSVAC NATHMTLTIP EFPGKLTSVD FGKSSIPEMQ WHANGIDKEA
301 TNGLRLHFRK TLLKTKPSEK CPPYQFYFSS LKLNFSLQPH LVSLVIDPEC HCESPVSIVA
361 DKLCTQDGFM DFEVYSHQTK PALNLETLVV GNSSCHPIPK SQSQGLLRFH IPLNGCGTGC
421 KFEGDKVIYE NEIHALWKNL PPSIIFRDSE FRMTVRCYYT RDSVPLNADI KSLLSPVASV
481 KPGPLMLVLQ IYPDKSYQQP YRKDEYPLVR YLRQPIYMEV TVLNRNDPSI KLVLDDCWAT
541 SSSDPASVHS GTLSWMAVNM NWTSYRTTFH PAGSSVVHPA HYQRFDVKTF AFVSEAQGLS
601 SLIYFHCSAL ICNPESLDSP LCSVTCPAPL RSKREAIQED TMTVSLPGPI LLLSDDSSLK
661 DTMVPNRHEI AKDTASKTVA AMAALVGSVV IVGFICYLHK ERTMRLDH

Hamster ZP3

1 MGLSYQLLLC LLLCGGAKQC CSQPLWLLPG GTPTPGK**LTS SVEVECLEAE LVVTVSRDLE**
61 **GTGK**LIQPED LTLGSENCRP LVSVATDVVR FKAQLHECSN RVQVTEDALV YSTVLLHQPR
121 PVPGLSILRT NRADVPIECR YPRQGNVSSH AIRPTWVPFS TTVSSEEK**LV FSLRLMEENW**
181 **NTEK**LSPTSH LGEVAYLQAE VQTGSHLPLL LFVDRCVPTP SPDQTASPYH VIVDFHGCLV
241 DGLSESFSAF QVPRPRPETL QFTVDVFHFA NSSR**NTIYIT CHLKVTPANQ TPDELNK**ACS
301 FNRSSKSWSP VEGDAEVCGC CSSGDCGSSS RSR**YQAHGVS QWPKSASR**RR RHVRDEADVT
361 VGPLIFLGKA SDQAVEGWAS SAQTSLALGL GLAAVAFLTL AAIVLGVTRS CHTPSHVVSI
421 SQ

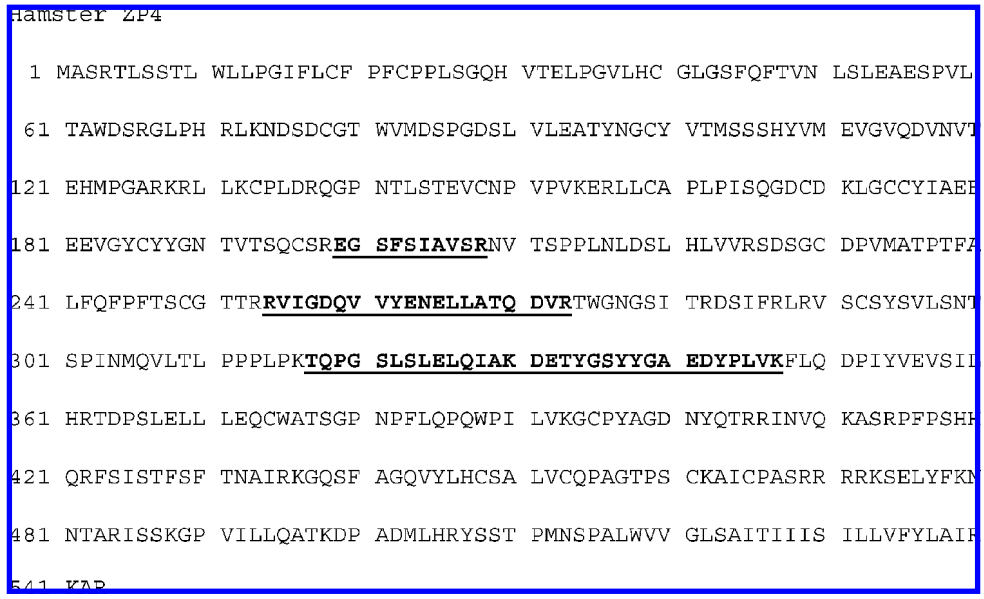


Figure 8. Hamster ZP1 (ABS86997), ZP2 (AAW66610), ZP3 (P23491), ZP4 (ABH06548) amino acid sequences. Bold underlined sequences are the tryptic peptides obtained by MS/MS.

composed of protein and saccharide. Such oligosaccharide moieties are known to play a key role in the interaction with spermatozoa.[42−44] The difficulty involved in obtaining definitive protein identification in mammalian ZP is largely due to the presence of these chains of carbohydrates. The presence of different glycoforms for each ZP protein is responsible for the appearance as a broad band when the ZP glycoproteins are separated by SDS-PAGE. Thus, identification of the different ZP pig glycoproteins could only be performed when the ZP was deglycosylated.[45]

Confirmation of the existence of four hamster ZP proteins requires a reinterpretation of the numerous electrophoretic studies on the ZP. Although only two diffuse bands were observed in one-dimensional (1D) electrophoresis, we were able to detect four proteins using biochemical and molecular approaches.

Processed ZP1 polypeptide has a molecular weight without glycosylation of 59.2 kDa. ZP1 peptides were detected in bands 4 (90 kDa) and 7 (65 kDa). ZP2 has a molecular weight of 67.4 kDa and its peptides were detected in bands 1 and 2 (MW between 182 and 115 kDa). ZP3 polypeptide backbone has a molecular weight of 36.4 kDa and its peptides were detected in bands 7, 8, and 9 (MW between 64.2 and 48 kDa). ZP4 has a molecular weight of 48.767 kDa and its peptides were detected in band 3 (181−115 kDa) and 9 (64.2−48 kDa) (described in Materials and Methods). So, the different proteins appear in different positions and some of the proteins appear in upper bands and in lower bands (ZP1 and ZP4). The presence of different glycoforms for each protein could partially explain this phenomenon. So, the different glycoforms probably have a different migration behavior in SDS-PAGE and it is also possible that different ZP proteins comigrate in the electrophoresis process.[31,32,11] However, it has to be to taken into consideration that the ZP mixture obtained from the ovaries could contribute or increase the heterogeneity of the ZP.

The presence of distinct glycoforms has been described previously by several authors in different species: in bovine ZP,[36] mouse ZP[46,47,39,30,48] hamster ZP,[39] rat ZP,[39] pig ZP,[49] and human ZP.[50] So, it is very likely that the dense and heterogeneous glycosylation observed is the reason for the fourth glycoprotein in hamster not being detected earlier. The number of the different possible carbohydrates chains is very high, even though a limited number of glycosylation sites are available. So, for example, for 5 O-glycosylation sites in mouse a total of 28 different carbohydrates chains have been described.[51] Furthermore, the different glycoforms are heterogeneously distributed throughout the ZP.[51−56] Thus, the inner region, close to the oocyte, is more densely packed than the outer region, which is in contact with the cumulus matrix or the cumulus cells. The different degrees of compaction in these areas may be responsible for the higher intensity of the lectin binding observed in the inner region of ZP.[51−53] The results of different experiments confirm that sugar residues are heterogeneously distributed in the ZP, some being restricted to the outer zone, and some having mainly, been detected in the inner zone.[53] Future structural analyses at the glycomic level will be required to unravel whether the differential glycosylation detected at the biochemical level holds true for all four hamster ZP proteins.

In summary, two cDNAs encoding ZP1 and ZP4 have been demonstrated in hamster (*Mesocricetus auratus*) ovaries. The nucleotide sequences show a high degree of homology with ZP1 and ZP4 of other mammalians. Employing mass spectrometry analysis we have probed the presence of the proteins of ZP1, ZP2, ZP3, and ZP4 in hamster ovaries. Future analyzes of isolated mature ZP are necessary to confirm the expression of four hamster ZP proteins in the ovulated oocyte.

Acknowledgment. This work was supported by Spanish MEC grant (BFU2004-05568/BFI) and Fundación Seneca (04542/GERM/06). We thank Dr. Anne Dell and all the members of her laboratory for the proteomic analysis of the hamster ZP3 performed during the visit of María Jiménez-Movilla to her laboratory. We thank Mr. Alejandro Torrecillas Sánchez of the Proteomic Service of the University of Murcia. We thank Dr. Pascale Chevret for the phylogenetic analysis.

References

(1) Yanagimachi, R. Mammalian fertilization. Physiology of reproduction; Knobil, E., Neill, J. D., Eds.; Raven Press: New York, 1994; pp 189−317.
(2) Bleil, J. D.; Wassarman, P. M. Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Dev. Biol.* **1980**, *76*, 185–202.
(3) Hedrick, J. L.; Wardrip, N. J. On the macromolecular composition of the zona pellucida from porcine oocytes. *Dev. Biol.* **1987**, *121*, 478–488.
(4) 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.; Barratt, C. L. R. Four zona pellucida glycoproteins are expressed in the human. *Hum. Reprod.* **2004**, *19*, 1580–1586.
(5) Hoodbhoy, T.; Joshi, S.; Boja, E. S.; Williams, S. A.; Stanley, P.; Dean, J. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. *J. Biol. Chem.* **2005**, *280*, 12721–12731.
(6) Ganguly, A.; Sharma, R. K.; Gupta, S. K. Bonnet Monkey (*Macaca radiata*) Ovaries, like human oocytes, express four zona pellucida glycoproteins. *Mol. Reprod. Dev.* **2008**, *75*, 156–166.
(7) Goudet, G.; Mugnier, S.; Callebaut, I.; Monget, P. Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biol. Reprod.* **2008**, *78*, 796–806.
(8) Hughes, D. C.; Barratt, C. L. Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida. *Biochim. Biophys. Acta* **1999**, *1447*, 303–306.
(9) Noguchi, S.; Yonezawa, N.; Katsumata, T.; Hashizume, K.; Kuwayama, M.; Hamano, S.; Watanabe, S.; Nakano, M. Characterization of the zona pellucida glycoproteins from bovine ovarian and fertilized eggs. *Biochim. Biophys. Acta* **1994**, *1201*, 7–14.
(10) Bausek, N.; Waclawek, M.; Wolfgang, J. S.; Wohlrab, F. The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. *J. Biol. Chem.* **2000**, *275*, 28866–28872.
(11) Moller, C. C.; Bleil, J. D.; Kinloch, R. A.; Wassarman, P. M. Structural and functional relationships between mouse and hamster zona pellucida glycoproteins. *Dev. Biol.* **1990**, *137*, 276–286.
(12) Kinloch, R. A.; Ruiz-Seiler, B.; Wassarman, P. M. Genomic organization and polypeptide primary structure of zona pellucida glycoprotein hZP3, the hamster sperm receptor. *Dev. Biol.* **1991**, *145*, 203–204.
(13) Frohman, M. A.; Dusk, M. K.; Martin, G. R. Rapid production of full-length cDNAs from rare transcript: amplification using a single gene specific oligonucleotide primer. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8998–9002.
(14) Belyavsky, A.; Vinogradova, T.; Rajewsky, K. PCR-based cDNA library construction: general cDNA libraries at the level of a few cells. *Nucleic Acids Res.* **1989**, *17*, 2919–2932.
(15) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
(16) Nielsen, H.; Engelbrecht, J; Brunak, S.; von Heijne, G. Identification of prokaryotic and eukariotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **1997**, *10*, 1–6.
(17) Hansen, J. E.; Lund, O.; Rapacki, K.; Brunak, S. O-GLYBASE version 2.0: A revised database of O-glycosylated proteins. *Nucleic Acids Res.* **1997**, *25*, 278–282.
(18) Hansen, J. E.; Lund, O.; Tolstrup, N.; Gooley, A. A.; Williams, K. L.; Brunak, S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accesibility. *Glycoconj. J.* **1998**, *15*, 115–130.
(19) Gupta, R.; Brunak, S. Prediction of glycosilation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, 310–322.
(20) Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970**, *227*, 680–685.
(21) Pedersen, A. G.; Nielsen, H. Neural network prediction of translation initiation sites in eukaryotes: perspectives for EST and genome analysis. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1997**, *5*, 226–233.
(22) Kozak, M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **1991**, *266*, 19867–19870.
(23) Bendtsen, J. D.; Nielsen, H.; von, H. G.; Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **2004**, *340*, 783–795.
(24) Krogh, A.; Larsson, B.; von, H. G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.
(25) Ducker, P.; Brunak, S.; Blom, N. Prediction of proprotein convertase cleavage sites. *Protein Eng. Des. Sel.* **2004**, *17*, 107–112.
(26) Boja, E. S.; Hoodbhoy, T.; Garfield, M.; Fales, H. M. Structural conservation of mouse and rat zona pellucida glycoproteins. Probing the native rat zona pellucida proteome by mass spectrometry. *Biochemistry* **2005**, *44*, 16445–16460.
(27) Greve, J. M.; Wassarman, P. M. Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *J. Mol. Biol.* **1985**, *181*, 253–264.
(28) Rankin, T. L.; Coleman, J. S.; Epifano, O.; Hoodbhoy, T.; Turner, S. G.; Castle, P. E.; Lee, E.; Gore-Langton, R.; Dean, J. Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. *Dev. Cell* **2003**, *5*, 33–43.
(29) Dean, J. Reassessing the molecular biology of sperm-egg recognition with mouse genetics. *Bioessays* **2004**, *26*, 29–38.
(30) Boja, E. S.; Hoodbhoy, T.; Fales, H. M.; Dean, J. Structural characterization of native mouse zona pellucida 'proteins using mass spectrometry. *J. Biol. Chem.* **2003**, *278*, 34189–34202.
(31) Ahuja, K. K.; Bolwell, G. P. Probable asymmetry in the organization of components of the hamster zona. *J. Reprod. Fertil.* **1983**, *69*, 49–55.
(32) Oikawa, T.; Sendai, Y.; Kurata, S.; Yanagimachi, R. A glycoprotein of oviductal origin alters biochemical properties of the zona pellucida of hamster egg. *Gamete Res.* **1988**, *19*, 113–122.
(33) Bork, P.; Sander, C. A large domain common to sperm receptors (ZP2 and ZP3) and TGF-β type III receptor. *Fed. Eur. Biochem. Soc.* **1992**, *300*, 237–240.
(34) Sasanami, T.; Pan, J.; Doi, Y.; Hisada, M.; Koshaka, T.; Toriyama, M. Secretion of egg envelope protein ZPc after C-terminal proteolytic processing in quail granulosa cells. *Eur. J. Biochem.* **2002**, *269*, 2223–2231.
(35) Akatsuka, K.; Yoshida-Komiya, H.; Tulsiani, D. R.; Orgebin-Crist, M. C.; Hiroi, M.; Araki, Y. Rat zona pellucida glycoproteins: molecular cloning and characterization of the three major components. *Mol. Reprod. Dev.* **1998**, *51*, 454–467.
(36) Ikeda, K.; Yonezawa, N.; Naoi, K.; Katsumata, T.; Hamano, S.; Nakano, M. Localization of N-linked carbohydrate cahins in glycoprotein ZPA of the bovine egg zona pellucida. *Eur. J. Biochem.* **2002**, *269*, 4257–4266.
(37) Kang, Y. Development of the zona pellucida in the rat oocyte. *Am. J. Anat.* **1974**, *139*, 535–566.
(38) Nicolson, G. L.; Yanagimachi, R.; Yanagimachi, H. Ultrastructural localization of lectin binding sites on the zona pellucida and plasma membranes of mammalian eggs. *J. Cell Biol.* **1975**, *66*, 263–274.
(39) Avilés, M.; Okinaga, T.; Shur, B. D.; Ballesta, J. Diferential expression of glycoside residues in the mammalian zona pellucida. *Mol. Reprod. Dev.* **2000**, *57*, 296–308.
(40) Oakberg, E. F.; Tyrell, P. D. Labeling of the zona pellucida of mouse oocytes. *Biol. Reprod.* **1975**, *12*, 477–482.
(41) Haddad, A.; Nagai, E. T. Radioautographic study of glycoprotein biosynthesis and renewal in the ovarian follicles of mice and the origin of the zona pellucida. *Cell Tissue Res.* **1977**, *177*, 347–369.
(42) Benoff, S. Carbohydrates and fertilization: an overview. *Mol. Hum. Reprod.* **1997**, *3*, 599–637.
(43) Tulsiani, D. R.; Yoshida-Komiya, H.; Araki, Y. Mammalian fertilization: a carbohydrate-mediate event. *Biol. Reprod.* **1997**, *57*, 487–494.
(44) Ling, R.; Shur, B. D. Sperm-egg binding requires a multiplicity of receptor-ligand interactions: new insights into the nature of gamete receptors derived from reproductive tract secretions. *Soc. Reprod. Fertil. Suppl.* **2007**, *65*, 335–351.
(45) Yonezawa, N.; Nakano, M. Identification of the carboxyl termini of porcine zona pellucida glycoproteins ZPB and ZPC. *Biochem. Biophys. Res. Commun.* **2003**, *307*, 877–882.
(46) Easton, R. L.; Patankar, M. S.; Lattanzio, F. A.; Leaven, T. H.; Morris, H. R.; Clark, G. F.; Dell, A. 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.* **2000**, *275*, 7731–7742.
(47) Avilés, M.; Castells, M. T.; Abascal, I.; Martínez-Menárguez, J. A.; Draber, P.; Kan, F. W. K.; Ballesta, J. Cytochemical localization of GalNAc and GalNAcβ1,4Galβ1,4 disaccharide in mouse zona pellucida. *Cell Tissue Res.* **1999**, *295*, 269–277.
(48) Dell, A.; Chalabi, S.; Easton, R. L.; Haslam, S. M.; Sutton-Smith, M.; Patankar, M. S.; Lattanzio, F.; Panico, M.; Morris, H. R.; Clark, G. F. Murine and human zona pellucida 3 derived from mouse eggs express identical O-glycans. *PNAS Dev. Biol.* **2003**, *100*, 15631–15636.
(49) Yonezawa, N.; Fukui, K.; Kudo, K.; Nakano, M. Localization of neutral N-linked carbohydrate chains in pig zona pellucida glycoprotein ZPC. *Eur. J. Biochem.* **1999**, *260*, 57–63.

(50) Chalabi, S.; Panico, M.; Sutton-Smith, M.; Haslam, S. M.; Patankar, M. S.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A. Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. *Biochemistry* **2006**, *45*, 637–647.
(51) Shalgi, R.; Maymon, R.; Bar-Shira, B.; Amihai, D.; Skutelsky, E. Distribution of lectin receptors sites in the zona pellucida of follicular and ovulated rat oocytes. *Mol. Reprod. Dev.* **1991**, *29*, 365–372.
(52) Avilés, M.; Martínez-Menárguez, J. A.; Castells, M. T.; Madrid, J. F.; Ballesta, J. Cytochemical characterization of oligosaccharide side chains of the glycoproteins of rat zona pellucida: an ultrastructural study. *Anat. Rec.* **1994**, *239*, 137–149.
(53) Avilés, M.; Jaber, L.; Castells, M. T.; Kan, F. W. K.; Ballesta, J. Modifications of the lectin binding pattern in the zona pellucida of rat after fertilization. *Mol. Reprod. Dev.* **1996**, *44*, 370–381.
(54) Avilés, M.; Jaber, L.; Castells, M. T.; Ballesta, J.; Kan, F. W. K. Modifications of carbohydrate residues in ZP2 and ZP3 in the mouse zona pellucida glycoproteins after fertilization. *Biol. Reprod.* **1997**, *57*, 1155–1163.
(55) Avilés, M.; Castells, M. T.; Martínez-Menárguez, J. A.; Abascal, I.; Ballesta, J. Localization of penultimate carbohydrate residues in zona pellucida and acrosomes by means of lectin cytochemistry and enzymatic treatments. *Histochem. J.* **1997**, *29*, 583–592.
(56) Skutelsky, E.; Ranen, E.; Shalgi, R. Variations in the distribution of sugar residues in the zona pellucida as possible species-specific determinants of mammalian oocytes. *J. Reprod. Fert.* **1994**, *100*, 35–41.

PR800568X