UNIVERSIDAD DE MURCIA

FACULTAD DE VETERINARIA

Análisis Molecular, Proteómico y Filogenético de la Zona Pelúcida de Ovocitos de Coneja (*Oryctolagus cuniculus*) y Gata (*Felis catus*).

Dª. Irene Stetson

2015

UNIVERSIDAD DE
MURCIA

Murcia, España 18 de marzo de 2015

D. **Manuel Avilés Sánchez**, Profesor Titular en el departamento de Biología Celular e Histología de la Facultad de Medicina, AUTORIZA:

La presentación de la Tesis Doctoral titulada *"***Análisis molecular, proteómico y filogenético de la zona pelúcida de ovocitos de coneja (*Oryctolagus cuniculus*) y gata (*Felis catus*)***"*, realizada por D[a]. **Irene Stetson**, bajo mi inmediata dirección y supervisión, y que presenta para la obtención del grado de Doctor por la Universidad de Murcia

Dr. Manuel Avilés Sánchez

UNIVERSIDAD DE
MURCIA

Murcia, España 18 de marzo de 2015

Da. **María José Izquierdo-Rico**, Profesor en el departamento de Biología Celular e Histología de la Facultad de Medicina, AUTORIZA:

La presentación de la Tesis Doctoral titulada *"***Análisis molecular, proteómico y filogenético de la zona pelúcida de ovocitos de coneja (*Oryctolagus cuniculus*) y gata (*Felis catus*)***"*, realizada por D[a]. **Irene Stetson**, bajo mi inmediata dirección y supervisión, y que presenta para la obtención del grado de Doctor por la Universidad de Murcia

Dra. María José Izquierdo-Rico

Esta Tesis Doctoral ha sido realizada en el Departamento de Biología Celular e Histología de la Facultad de Medicina durante el periodo de disfrute de una Licencia por estudio predoctoral de la Universidad Nacional de Misiones Argentina y Ministerio de Salud Pública de Posada Misiones Argentina. Los proyectos de investigación han sido financiados por el Ministerio de Ciencias e Innovación (Proyecto AGL2009-12512-C02-02 y AGL2012-40180-C03-02). Así mismo, se ha desarrollado en el marco del Programa de Ayudas a Grupos de Excelencia de la Región de Murcia de la Fundación Séneca, 04542/GERM/09

AGRADECIMIENTOS

A Manuel Avilés Sánchez, mi estimado Director de tesis, a quien elegí como tutor siendo profesor del Máster y tuve la suerte de ser aceptada, gracias porque me alentaste siempre, me inspiraste, me has dado ánimo y me has trasmitido la pasión por tu trabajo, porque siempre felicitas y agradeces cada resultado positivo, y animas a seguir cuando así no lo fuera. Gracias por tu ayuda y por tu don de gente que han permitido que esta tesis sea una realidad.

A María José Izquierdo-Rico mi estimada Directora de tesis, quien entró al laboratorio a explicarme cada cosa diciendo esto es "fácil", haciendo sencillo lo complicado, demostrando tu amplia experiencia, tus respuestas eficiente y precisas ante las adversidades, gracias por tu super ayuda, tu dedicación y paciencia. Con vuestra ayuda y estímulo estimados tutores, fue posible finalizar esta tesis.

A todos los docentes del departamento de Biología Celular e Histología: Emma Martínez, Concha Ferrer, Esther Beltrán, José Ángel Martínez, Juan Francisco Madrid, José Ballesta, Luis Miguel Pastor, Teresa Castells, a Mari Carmen, gracias por vuestra ayuda, consejos, acompañamiento y charlas compartidos.

Agradezco a todos y cada uno de mis compañeros, Tere entrañable compañera y amiga, a Luis Miguel, a Willy y Lily, amigos, compañeros de posgrado y de piso, a mi querida Narci, a Sara y mi "colega" Vicente, a mi estimadísimo Salva, gracias a Ascen, y Blanca en los últimos meses de laboratorio, y un agradecimiento muy especial a tu predisposición, compañerismo y rectitud querida Carla, con quien compartimos experimentos, artículos y presentaciones en congresos.

A la querida gente del departamento de Fisiología Animal de Veterinaria, a los doctores Pilar Coy, Joaquín Gadea Mateos, y Fran García Vázquez, a Raquel

Román Andrés, Luis, Irene, Karen, a ti Juan y un agradecimiento especial a Sebastián Cánovas, con quien me contacte desde Argentina, me informo, animó y ayudó con los primeros trámites para realizar este postgrado, gracias por tu gestión y excelente actitud.

Mi agradecimiento al Servicio de Apoyo a la Investigación de la Universidad de Murcia, especialmente a Alejandro Torrecillas y a César Flores por su gran aporte a estas investigaciones.

Agradezco a las Clínicas Veterinarias que nos han aportado los ovarios de gatas (Clínica San Roque Plaza, Murcia) y a la Clínica Huellas (Vetpinos S.L.P., Pinoso, Alicante), mi gratitud al Matadero "Conejos Susi SL" Formentera del Segura, Alicante por las muestras de ovario de coneja.

A tí Conchitina increíble, única e irremplazable amiga, por el postgrado vital, experiencias entrañables, charlas, viajes, y caminos de 800 km a Santiago de Compostela, por tu amistad sincera, desapegada y generosa, gracias a tu presencia mi estancia tan lejos de mi familia y mi patria ha sido más liviana, agradable, y bella... gracias mil.

A mis compañeros de Taichí y a mis amigos de Biodanza, al apoyo de los meeting Ishas y el estado de amor consciencia, me trasmitieron fuerza y decisión.

Gracias a mis amigos de siempre Pepito, Ángel Carlos, Ana, Marta, Riqui, José, Tiki, Anita, Silvia, Cocho. Gracias Tito, amado y fiel compañero de mi vida.

A mis nuevos amigos de España, Pedro, Pablo, Tarsi, Toño, Sofi, a Piedi y Miguel en estos últimos meses de final de tesis y a tantos, tantos otros conocidos que sin saber ellos, han aportado a mi felicidad y aprendizaje, han pintado trazos de arcoíris en mi alma Gracias.

A mis compañeros de Cátedra y Autoridades de la Universidad Nacional de Misiones Argentina y a mis colegas y Autoridades del Hospital Materno Neonatal de Posadas Misiones y a las Autoridades del Ministerio de Salud Pública.

A mis hermanos admirables Robbie, y Charlie, por el aliento y acompañamiento constante.

A mis añorados y queridísimos padres por su amor incondicional, por su absoluta coherencia de pensamiento y obra, por inculcarme la constancia al estudio y el esfuerzo, la responsabilidad, y el respeto, el amor a la vida y el agradecimiento. A mamá por trasmitirme la Fe, Gracias a Dios y a la Virgen del Perpetuo Socorro.

Muy muy especial es el infinito agradecimiento a mis amados hijos, Sebastián, Daiana y Matías, quienes desde el primer momento me apoyaron, sostuvieron y felicitaron frente a este proyecto, para lo cual debieron afrontar y asumir compromisos y responsabilidades exitosamente. Gracias también a mis queridísimos hijos políticos Lu y Ele.

Y especialmente a mi amado nieto Simón, que cuando me enteré que venias al mundo decidí regresar a España, defender la tesis y concretar esta etapa de mi vida. Gracias mi amor bendito.

A mis padres

Harry y Nelly.

A mis hijos

Sebastián, Daiana y Matías,

y a mi nieto Simón.

ÍNDICE

1. RESUMEN

1. RESUMEN

El oocito de mamífero está rodeado por una matriz denominada zona pelúcida (ZP). Esta envoltura participa en procesos tales como la inducción de la reacción acrosómica, la unión de espermatozoides y también puede estar implicada en la especiación. Estudios recientes han demostrado que la matriz de ZP de ovocitos en varias especies se compone de cuatro glicoproteínas, denominadas ZP1, ZP2, ZP3 y ZP4, en lugar de las tres descritas en el ratón, cerdo y vaca. En la gata (*Felis catus*), esta matriz se compone de al menos tres glicoproteínas denominadas ZP2, ZP3 y ZP4. Sin embargo, la presencia de cuatro proteínas en varios mamíferos, sugiere que necesita una reevaluación de la ZP en la gata. Además, en esta tesis, se realizaron investigaciones para determinar la expresión de una cuarta glicoproteína en la ZP de conejo (*Oryctolagus cuniculis*). Los objetivos de esta investigación fueron analizar la composición de proteínas de ZP gato por medio de análisis proteómico y en el conejo se amplificó ZP1 mediante la reacción en cadena de la polimerasa (RT-PCR). En la ZP aislada de ovarios y ovocitos de gato, se detectaron varios péptidos correspondientes a cuatro proteínas, teniendo una cobertura del 33,17%, 71,50%, 50,23% y 49,64% para ZP1, ZP2, ZP3 y ZP4, respectivamente. Por otra parte, se confirmó por análisis molecular la expresión de los cuatro genes de ZP a partir de ARN total aislado de ovarios de gato, las ZPs de gato fueron parcialmente amplificadas por RT-PCR. En el conejo la ZP1 incluye un marco de lectura abierto de 1825 nucleótidos que codifican un polipéptido de 608 residuos de aminoácidos. La secuencia de aminoácidos deducida de ZP1 de conejo mostró una alta identidad con otras especies: 70% de identidad con ZP1 humana y caballo y 67% de identidad con la ZP1 de ratón y rata. A nivel proteómico, fueron detectados por espectrometría de masas péptidos correspondientes a las cuatro proteínas ZP. Además, mediante un análisis filogenético molecular de ZP1 mostró que este gen presenta una pseudogenización de por lo menos cuatro veces durante la evolución de los mamíferos. Los datos presentados en esta tesis proporcionan evidencia, por

primera vez, que la ZP de conejo y la ZP de gato se compone de cuatro glicoproteínas.

2. SUMMARY

2. SUMMARY

The mammalian oocyte is surrounded by a matrix called the zona pellucida (ZP). This envelope participates in processes such as acrosome reaction induction, sperm binding, and may be involved in speciation. Recent studies have shown that the ZP matrix of oocytes in several species is composed of four glycoproteins, designated ZP1, ZP2, ZP3 and ZP4, rather than the three described in mouse, pig and cow. In cat (*Felis catus*), this matrix is composed of at least three glycoproteins called ZP2, ZP3 and ZP4. However, the presence of a fourth protein in several mammals, meaning that a reevaluation of cat ZP is needed. Additionally, in this thesis, investigations were carried out to unveil a fourth glycoprotein in the rabbit (*Oryctolagus cuniculis*) ZP. The objectives of this research was to analyse of the protein composition of cat ZP by means of proteomic analysis and rabbit ZP1 was amplified by reverse transcribed polymerase chain reaction (RT-PCR). Using ZP from ovaries and oocytes of cat, several peptides corresponding to four proteins were detected, yielding a coverage of 33.17%, 71.50%, 50.23% and 49.64% for ZP1, ZP2, ZP3 and ZP4, respectively. Moreover, the expression of four genes was confirmed by molecular analysis. Using total RNA isolated from cat ovaries, the complementary deoxyribonucleic acids (cDNAs) encoding cat ZPs were partially amplified by RT-PCR. In the rabbit the ZP1 cDNA contains an open reading frame of 1825 nucleotides encoding a polypeptide of 608 amino acid residues. The deduced amino acid sequence of rabbit ZP1 showed high identity with other species: 70% identity with human and horse ZP1, and 67% identity with mouse and rat ZP1. At the proteomic level, peptides corresponding to the four proteins were detected by mass spectrometry. In addition, a molecular phylogenetic analysis of ZP1 showed that pseudogenization of this gene has occurred at least four times during the evolution of mammals. The data presented in this thesis provide evidence, for the first time, that the rabbit and cat ZP is composed of four glycoproteins.

3. INTRODUCCIÓN

3. INTRODUCCIÓN

En la fecundación, los gametos masculino y femenino interactúan para dar lugar a una célula huevo o cigoto. Durante este proceso, el espermatozoide se une al ovocito atravesando la zona pelúcida (ZP) para posteriormente atravesar la membrana plasmática (oolema) y alcanzar el ooplasma. La ZP es una matriz extracelular que rodea a ovocitos y embriones, demostrándose que juega un papel importante en el proceso de la fecundación y desarrollo embrionario temprano. Esta matriz interviene en la interacción especie-específica entre gametos, en la inducción de la Reacción Acrosómica (RA), en el bloqueo de la polispermia, así como en la protección del embrión preimplantado (Modlinski, 1970; Bleil y Wassarman, 1980; Florman y Storey, 1982; Berger *et al.*, 1989; Dumbar *et al.,* 1991; Liu *et al.*, 1996; Rankin *et al.*, 1996, 1999, 2001; Benoff, 1997; Fazeli *et al.*, 1997; Wassarm, 1998: Kölle *et al.*, 1998; Dean, 2004; Hoodbhoy y Dean, 2004; Cánovas *et al.*, 2007; Gupta *et al.*, 2011, 2012; Tanihara *et al.*, 2013).

Esta matriz extracelular es conocida como envoltura vitelina en anfibios, corion en los teleósteos y membrana perivitelina en las aves (Sasanami *et al.*, 2002, 2003; Barisone *et al.*, 2007; Wassarman y Litscher, 2008). Estas matrices extracelulares llevan a cabo funciones similares en las diferentes especies.

Referente a la composición de la ZP, es una matriz extracelular formada por unas pocas glicoproteínas. Su formación tiene lugar a lo largo de la ovogénesis y de la foliculogénesis (Wassarman, 1988; Jewgenow y Fickel, 1999; Lee *et al.*,1993; 2000; Blackmore *et al.*, 2004).

La ultraestructura y funciones de la ZP, así como su composición bioquímica han sido objeto de numerosos estudios en diferentes especies durante los últimos 35 años. No pocas investigaciones coinciden en definir que la composición de la ZP varía en las diferentes especies pudiendo estar constituída de 3 a 6 glicoproteínas (Bleil y Wassarman, 1980a; Hedrick y Wardrip, 1987;

Lefièvre *et al.,* 2004; Hoodbhoy *et al.*, 2005; Ganguly *et al.*, 2008; Goudet *et al.*, 2008). Sin embargo, la composición proteica así como la nomenclatura usada para clasificar las diferentes proteínas de la ZP continúan siendo bastante confusas. Un reciente estudio filogenético clarifica dicha nomenclatura relacionándola con la evolución de los genes codificantes para las diferentes proteínas de la ZP. Además, detecta la presencia de pseudogenes en diferentes especies. Así, estos autores proponen clasificar los genes de la ZP dentro de seis subfamilias: *ZPA/ZP2*, *ZPC/ZP3, ZPB/ZP4, ZP1, ZPAX* y *ZPD* (Goudet *et al.*, 2008).

Así, en especies no mamíferas han sido detectados más de cuatro genes. En el genoma de la gallina están presentes 6 genes (*ZP1, ZPA, ZPB, ZPC ZPAX y ZPD*) (Bausek *et al.*, 2000; Goudet *et al.*, 2008) y en el genoma de *Xenopus,* (rana africana) encontramos cinco genes codificantes para la ZP (*ZPA, ZPB, ZPC, ZPAX y ZPD*) (Goudet *et al.*, 2008). Para los mamíferos, la ZP tradicionalmente se ha considerado compuesta por tres glicoproteínas: ZP1, ZP2 y ZP3 tomando como modelo la ZP murina (Bleil y Wassarman, 1980). Ahora bien, la descripción del genoma completo en algunas especies, tales como el hombre, la rata o el caballo entre otras, ha dado lugar a la detección de nuevas proteínas asociadas a la ZP. Así, estudios recientes revelan que algunos mamíferos presentan cuatro glicoproteínas conformando esta matriz. Estas cuatro proteínas se designan como ZP1, ZP2 (ZPA), ZP3 (ZPC) y ZP4 (ZPB).

En la especie humana, ZP4 (ZPB) fue primeramente identificada como el gen ortólogo de ZP1 murina, pero estudios posteriores detectaron que el verdadero ortólogo de ZP1 humano era un gen distinto al llamado ZP1 en ratón (Hughes y Barrat, 1999). En cuanto a la composición proteica de la ZP se desprende que el ratón de laboratorio (*Mus musculus*), empleado como modelo de estudio de la ZP durante las últimas décadas, presenta una composición diferente al resto de mamíferos descritos hasta la fecha, con ZP1, ZP2 y ZP3, siendo ZP4 un pseudogén (Lefièvre *et al.*, 2004; Evsikov *et al.*, 2006; Goudet *et al.*, 2008).

Estudios recientes sobre la evolución de los genes codificantes de las proteínas de la ZP (Goudet *et al.*, 2008) revelan nuevos datos como la existencia de una pérdida de genes en determinadas especies tal y como ocurre con la ZP4 murina. Así, en la vaca (Noguchi *et al* 1994) y el perro (Goudet *et al.*, 2008) se describen sólo tres glicoproteínas, siendo estas: ZP2, ZP3 y ZP4, donde la proteína ZP1 ha sido identificada como un pseudogén en el genoma de estas especies (Goudet *et al.*, 2008). Así, teniendo en cuenta todo lo anteriormente expuesto podríamos agrupar a los mamíferos de acuerdo a la composición de su ZP en:

1. Los que presentan ZP1, ZP2 y ZP3 (con ZP4 como pseudogén), hasta la fecha solo el ratón (*Mus musculus*) (Bleil y Wassarman 1980a; Lefièvre, 2004; Evsikov *et al.*, 2006; Goudet *et al.*, 2008).

2. Los que presentan ZP2, ZP3 y ZP4 (con ZP1 como pseudogén) como en la perra (*Canis familiaris*) y la vaca (*Bos taurus*) (Goudet *et al.*, 2008).

3. Los que presentan ZP1, ZP2, ZP3 y ZP4: mujer, rata, macaco coronado y el hámster (Lefièvre *et al.*, 2004; Hoodbhoy *et al.*, 2005; Ganguly *et al., 2008*; Izquierdo-Rico *et al.*, 2009)

En la coneja (*Oryctolagus cuniculus*), la caracterización de la ZP por medio de electroforesis en geles de poliacrilamida (SDS-PAGE) sugirió la presencia de tres glicoproteínas, ZP2, ZP3 y ZP4, (Lee *et al.,* 1993; Harris *et al.*, 1994; Prasad, 1995) las cuales migran en el gel como una sola banda, con un peso molecular aparente de 85-95 kDa (Harris *et al.*, 1994). Los transcritos correspondientes a estas proteínas (ZP2, ZP3 y ZP4) fueron detectados por métodos de biología molecular y se depositaron en la base de datos GenBank con los números de acceso: L12167 (ZP2), NM_001195720.1 (ZP3) y NM_001082295 (ZP4).

En el gato (*Felis catus*), la caracterización de la ZP por medio de la electroforesis con gel de poliacrilamida (SDS-PAGE) sugirió la presencia de sólo dos glicoproteínas que cubren un amplio rango de masa molecular con bandas entre 50-110 kDa (Maresh y Dunbar, 1987). Más tarde, Harris *et al.*, clonaron y caracterizaron los ADNc de la ZP2, ZP3 y ZP4 de varias especies de mamíferos, incluido el gato. Las secuencias de nucleótidos y aminoácidos de estos genes y las correspondientes proteínas fueron depositadas en la base de datos GenBank con los siguientes números de acceso: ZP2 (NM_001009875), ZP3 (NM_001009330) y ZP4 (NM_001009260) (Harris *et al.*,1994)

En la presente Tesis Doctoral estudiamos mediante técnicas de biología molecular y proteómica, la composición de la ZP en estas dos especies de mamíferos, la coneja y la gata, a fin de esclarecer la estructura de su ZP y aportar nuevos datos respecto a otros modelos animales. Además se lleva a cabo un análisis filogenético de ZP1 en mamíferos.

4. OBJETIVOS

4. OBJETIVOS

El objetivo de esta Tesis Doctoral se centró en determinar la composición de la zona pelúcida en dos modelos animales. Se analiza la coneja (*Oryctolagus cuniculus*) por su interés como animal de experimentación alternativo al tratarse de una especie de consumo humano pudiendo obtenerse en matadero evitandose el sacrificio de animales cuya finalidad única sea la investigación. Se analiza la gata (*Felis catus*) como modelo para el desarrollo de vacunas anticonceptivas para el control de poblaciones de félidos domésticos y salvajes.

1. CONEJA (*Oryctolagus cuniculus*)

1.1. Análisis de la presencia del ARNm de ZP1 en el ovario de coneja.

1.2. Identificación de las proteínas de la ZP de coneja mediante espectrometría de masas.

2. GATA (*Felis catus*)

2.1 Análisis de la presencia de ARNm de ZP1 en el ovario de gata.

2.2. Identificación de las proteínas de la ZP de gata mediante espectrometría de masas.

3. ANÁLISIS FILOGENÉTICO DEL GEN *ZP1* EN MAMÍFEROS.

5. REVISIÓN BIBLIOGRÁFICA

5. REVISIÓN BIBLIOGRÁFICA

Durante la ovogénesis, los ovocitos de los mamíferos se vuelven más complejos en su organización, adquieren competencia para reactivar la meiosis y se preparan para la fecundación y la primera división embrionaria (Wassarman, 1988). Uno de los cambios más relevantes durante la ovogénesis es la producción de la ZP. Esta matriz extracelular rodea al ovocito externamente y desempeña papeles vitales durante la ovogénesis, la unión espermatozoide-ovocito, la inducción de la reacción del acrosoma (RA), bloqueo de la polispermia y el desarrollo temprano del embrión preimplantado (Wassarman, 1988; Tian *et a*l., 1997; Hyllner *et al.*, 2001; Sasanami *et al.*, 2002; Monné y Jovine, 2011).

5.1 COMPOSICIÓN DE LA ZONA PELÚCIDA

La ZP es una matriz extracelular traslúcida, compuesta por una matriz de glicoproteínas que rodea al ovocito y al embrión preimplantado de los mamíferos (Fig.1).

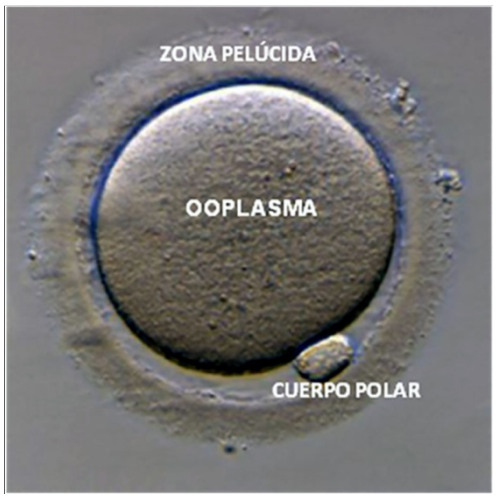


Figura 1. Imagen de un ovocito humano en metafase II. Se identifica el ooplasma, el primer corpúsculo polar y la ZP.

La ZP está constituida principalmente por proteínas y azúcares, que se asocian para formar varias glicoproteínas. Dependiendo de las especies éstas pueden variar en número; por ejemplo en mamíferos se describen 3 o 4 glicoproteínas diferentes y en otros vertebrados se describen hasta 6 proteínas.

La nomenclatura de las proteínas que conforman la ZP es algo confusa. En el ratón, modelo de estudio de la ZP desde hace más de 30 años, las glicoproteínas se clasificaron tomando en cuenta su peso molecular de acuerdo a la migración en gel SDS-PAGE. Éstas se denominaron ZP1, ZP2 y ZP3, siendo ZP1 la de mayor peso molecular (menor movilidad electroforética) con 200 kDa, ZP2 con 120 kDa y ZP3 la de menor peso molecular (mayor movilidad electroforética) con 83 kDa (Bleil y Wassarman, 1980a; Wassarman, 1988). Estas mismas glicoproteínas no siempre están presentes en la ZP de otras especies lo que llevó a revisar la nomenclatura.

En 2003, Spargo y Hope, tras realizar un estudio filogenético, proponen unificar el sistema de nomenclatura. Estableciendo que las 3 subfamilias *ZPA*, *ZPB* y *ZPC* serían identificadas numéricamente como *ZP2*, *ZP1* y *ZP3*. En este sistema de clasificación no consideraron la existencia de especies con más de tres genes codificantes para proteínas de la ZP lo que hace que la terminología usada todavía en nuestros días sea complicada.

Cinco años más tarde en 2008, Goudet y colaboradores, proponen clasificar los genes que codifican para las glicoproteínas de la ZP en seis subfamilias: *ZP1*, *ZP2/ZPA*, *ZP3/ZPC*, *ZP4/ZPB*, *ZPAX* y *ZPD*. La ZP varía considerablemente, no todos los genes están presentes en todas las especies, por ejemplo, en el genoma de mamíferos podemos encontrar 3 ó 4 genes, en el genoma de anfibios (*Xenopus*) 5 genes, y en el genoma de aves, como la gallina 6 genes (Goudet *et al.*, 2008).

La composición de la ZP se complica tras el descubrimiento de que algunos de estos genes como ZPAX y/ o ZP3 pueden estar duplicados en algunas especies. Por ejemplo, en algunos peces como la carpa espinosa (*Cyprinus carpio*), el pez globo moteado *(Tetraodon nigroviridis)*, el pez globo japonés (*Takifugu rubripes*)

entre otros (Conner y Hughes, 2003; Meslin *et al.*, 2012). También se describen duplicaciones de ZP3 en un marsupial como la zarigüeya de cola corta (*Monodelphis domestica*) (Meslin *et al.*, 2012), aves como el diamante mandarín (*Taeniopygia guttata*) y la gallina y en anfibios como la rana (Meslin *et al.*, 2012). Además, en el ornitorrinco (*Ornithorhynchus anatinus*), aparte de encontrarse ZP3 duplicada también lo está ZP2 (Meslin *et al.*, 2012).

Si nos centramos en los estudios realizados en mamíferos, se suponía que la ZP estaba formada por sólo tres glicoproteínas (ZP1, ZP2 y ZP3), debido a la extrapolación del modelo de ratón a otras especies (Bleil *et al.*, 1980; Bleil y Wassarman, 1980a).

También se demostró la presencia de tres proteínas en el cerdo (Hedrick *et al.*, 1987) y en la vaca (Noguchi, *et al.*, 1994), donde se describe que la ZP está conformada por tres glicoproteínas, identificadas como (ZP2, ZP3 y ZP4).

El análisis del genoma completo en varias especies, así como el desarrollo de las técnicas de espectrometría de masas, han proporcionado una importante alternativa para identificar las diferentes glicoproteínas presentes en la ZP de distintas especies. En 1999, el grupo del Dr. Barratt demostró la existencia de cuatro genes en la mujer (Hughes y Barratt, 1999) y en 2004 se identificaron por proteómica cuatro proteínas en la ZP de la mujer (Lefièvre *et al.*, 2004). Posteriores análisis filogenéticos detectan cuatro genes en el chimpancé, en el macaco de Rhesus (Goudet *et al.*, 2008) y en la yegua (Mugnier *et al.*, 2009). Mientras que, se demuestra la presencia de cuatro glicoproteínas en la rata (Hoodbhoy *et al.*, 2005), en el macaco coronado (*Macaca radiata*) (Ganguly *et al.*, 2008) y en el hámster (Izquierdo-Rico *et al.*, 2009a, Jiménez-Movilla *et al.*, 2009). Por tanto, podemos clasificar a los mamíferos en tres categorías:

1- Especies con **ZP1, ZP2 y ZP3**, hasta la fecha exclusivamente la ratona (Bleil y Wassarman, 1980a; Lefièvre *et al.*, 2004; Evsikov *et al.*, 2006; Goudet *et al.*, 2008).

2- Especies con **ZP2, ZP3 y ZP4**, donde ZP1 no está presente, como en la cerda, la vaca y la perra (Hedrick y Wardrip 1987, Noguchi *et al.*, 1994; Harris *et al.*, 1994, Goudet *et al.*, 2008)

3- Especies con 4 proteínas en su ZP: **ZP1, ZP2, ZP3 y ZP4**, como en la mujer, la rata, el macaco coronado o el hámster (Lefièvre *et al.*, 2004; Hoodbhoy *et al.*, 2005; Ganguly *et al.*, 2008; Izquierdo-Rico *et al.*, 2009a, Jiménez-Movilla *et al.*, 2009).

Los mamíferos cuya ZP puede estar conformada por 3 o 4 glicoproteínas desafían el modelo murino que es diferente al resto de los mamíferos, ya que cada día tenemos más evidencias de que la ZP con cuatro proteínas es la más común entre los mamíferos (Tabla 1).

Tabla.1. Resumen de las glicoproteínas de la ZP descritas en los diferentes mamíferos (Tomado de la Tesis Doctoral de Carla Moros Nicolás, 2015).

Especie	ZP1	ZP2	ZP3	ZP4	Referencias
Armiño (*Mustela erminea*)		**X**	**X**	**X**	Jackson y Beaton, 2004
Cabra (*Capra hircus*)		**X**	**X**		Chen *et al*., 2010
Cerda (*Sus scrofa domestica*)		**X**	**X**	**X**	Hedrick y Wardrip, 1984; Yurewicz et al., 1992, 1993; Harris et al., 1994; Hasegawa et al., 1994; Gupta et al., 1995; Taya et al., 1995; Kudo et al., 1998; Leièvre et al., 2003; Yonezawa et al., 2005
Coneja (*Oryctolagus cuniculus*)		**X**	**X**	**X**	Schoebel et al., 1991; Lee et al., 1993; Harris et al., 1994
Chimpancé (*Pan troglodytes*)	**X**	**X**	**X**	**X**	Goudet *et al*., 2008
Gata (*Felis catus*)		**X**	**X**	**X**	Harris *et al*., 1994, 1995; Okazaki y Sugimoto, 1995; Jewgenow y Fickel, 1999; Okazaki *et al*., 2007; Eade *et al*., 2009;
Hámster dorado (*Mesocricetus auratus*)	**X**	**X**	**X**	**X**	Kinloch *et al*., 1990; Koyama *et al*., 2005; Izquierdo-Rico *et al*., 2009
Hurona (*Mustela putorius furo*)			**X**		Jackson y Beaton, 2004
Macaco cangrejero (*Macaca fascicularis*)		**X**	**X**	**X**	Harris y Piersen, 2003
Macaco (*Macaca radiata*)	**X**	**X**	**X**	**X**	Gupta, 1994; Kolluri *et al*., 1995; Jethanandani *et al*., 1998; Ganguly *et al*., 2008
Mono tití (*Callithrix jacchus*)		**X**			Kerr *et al*., 1997

Macaco de Rhesus (Macaca mulata)	**X**	**X**	**X**	**X**	Goudet *et al*., 2008
Mujer (*Homo sapiens*)	**X**	**X**	**X**	**X**	Hughes y Barrat, 1999; Lefièvre *et al*., 2004
Oveja (*Ovis aries*)		**X**	**X**		Pariset *et al*., 2006; Chen *et al*., 2010
Papión amarillo (*Papio cynocephalus*)				**X**	Harris y Piersen, 2003
Perra (*Canis lupus familiaris*)		**X**	**X**	**X**	Harris *et al*., 1994; Okazaki *et al*., 1995; Okazaki y Sugimoto, 1995; Blackmore *et al*., 2004; McLaughlin *et al*., 2004
Rata (*Rattus norvegicus*)	**X**	**X**	**X**	**X**	Hoodbhoy *et al*., 2005
Ratona (*Mus musculus*)	**X**	**X**	**X**		Bleil y Wassarman, 1980a
Ratona de los Llanos (*Pseudomys australis*)			**X**		Swann *et al*., 2007
Ratón marsupial de cola gruesa (*Sminthopsis crassicaudata*)		**X**			Voyle *et al.*, 1999
Ratón marsupial de cara rayada *(Sminthopsis macroura)*		**X**			Au *et al*., 2008
Topillo de Brandt (*Microtus brandti*)			**X**		Li *et al*., 2000
Vaca (*Bos taurus*)		**X**	**X**	**X**	Harris *et al*., 1994; Noguchi *et al*., 1994; Yonezawa *et al*., 2001; Ikeda *et al*., 2002
Yegua (*Equus ferus caballus*)	**X**	**X**	**X**	**X**	Mugnier *et al*., 2009
Zarigüeya australiana (*Trichosurus vulpecula*)		**X**	**X**	**X**	Mate y McCartney, 1998; Voyle *et al*., 1999; McCartney y Mate, 1999; Haines *et al*., 1999; Mate *et al*., 2003
Zorra (*Vulpes vulpes*)		**X**	**X**		Reubel *et al*., 2005

En el conejo (*Oryctolagus cuniculus)* se realizó la caracterización de la ZP por medio de SDS-PAGE, evidenciándose la presencia de tres glicoproteínas: ZP2 (Lee *et al*., 1993; Von Witzendorff *et al.,* 2009); ZP3 (Shwoebel *et al.,* 1991) y ZP4 (Harris *et al.,* 1994; Prasad., 1995), las cuales migran como una banda única, con una masa molecular aparente de 85-95 kDa (Prasad *et al.,* 1995). (Fig.2) Las secuencias nucleotídicas y aminoacídicas se depositaron en el GenBank con los números de acceso: L12167 (ZP2), NM_001195720.1 (ZP3), NM_001082295 (ZP4).

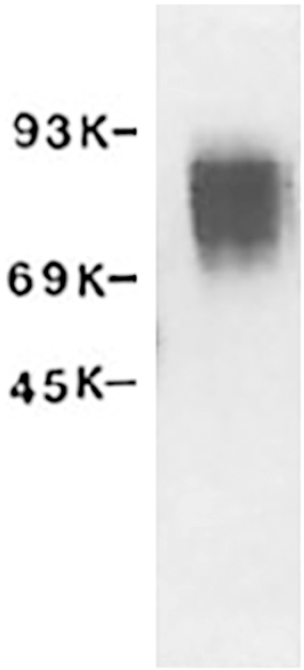


Fig 2. Zona pelúcida solubilizada de conejo analizada mediante electroforesis SDS-PAGE y tincíon de plata. Tomada de Prasad *et al*., 1995.

En el gato (*Felis catus*), la caracterización de la ZP por medio de la electroforesis con gel de poliacrilamida (SDS-PAGE) sugirió la presencia de sólo dos glicoproteínas que cubren un amplio rango de masa molecular con bandas entre 50-110 kDa (Maresh y Dunbar, 1987). Más tarde, Harris y colaboradores (1994), clonaron y caracterizaron los ADNc de la ZP2, ZP3 y ZP4 de varias especies de mamíferos, incluido el gato. Las secuencias de nucleótidos y aminoácidos de estos genes y las correspondientes proteínas fueron depositadas

en la base de datos GenBank con los siguientes números de acceso: ZP2 (NM_001009875), ZP3 (NM_001009330) y ZP4 (NM_001009260).

Sin embargo, ZP1 no se ha descrito ni en el conejo ni en el gato. Por otra parte, el gato es un miembro del orden de los mamíferos carnívoros, por lo cual se le ha incluído en la misma categoría que las especies relacionadas, como el perro en el que ZP1 es un pseudogén (Meslin, *et al* 2012). Sin embargo, las últimas actualizaciones de las bases de datos genómicas detectan en el genoma de estas especies un putativo gen ZP1 con un posible marco abierto de lectura (ENSOCUG00000015673 y NW_004065056 correspondientes a ZP1 de conejo y gato respectivamente).

5.2 Relevancia de la zona pelúcida

Entre las múltiples funciones atribuidas a la ZP, cabe destacar que interviene en la foliculogénesis, la organización y diferenciación de las células de la granulosa. Es la ZP quien se encarga de la unión específica de los gametos, ovocito y espermatozoide; la que induce la RA garantizando la fecundación y la estructura que evita la polispermia una vez fecundado el ovocito. Además la ZP se encarga de la protección del embrión oviductal hasta su implantación (Modlinski, 1970; Florman y Storey, 1982; Bleil y Wassarman, 1988; Berger *et al.*, 1989; Yanagimachi,1994; Liu *et al.*, 1996; Rankin *et al.*,1996,1999,2001; Benoff, 1997; Fazeli *et al.*, 1997; Kölle *et al.*, 1998; Wassarman y Litscher, 2001; Dean, 2004; Gupta *et al.*, 2011, 2012; Tanihara *et al.*, 2013).

Durante el desarrollo embrionario temprano, la ZP previene la disgregación de las blastómeras no compactadas, evitando la implantación prematura y ectópica; además protege al embrión contra virus, bacterias, toxinas y macrófagos y facilita la transmisión de señales entre el embrión y el útero (Herrler y Beier, 2000). En otros vertebrados como los peces, anfibios y aves, las capas que rodean al ovocito reciben diferente denominación. Así, son llamadas:

corion, membrana vitelina y membrana perivitelina respectivamente. Estas estructuras que rodean al ovocito son similares en tamaño y función a la ZP de los mamíferos, por lo que a veces pueden ser referidas colectivamente como ZP (Spargo *et al*, 2003).

Sin embargo, algunas de estas funciones están siendo puestas en evidencia en estudios recientes como ocurre con el papel de la ZP en la inducción de la RA. En 2011, Jim y colaboradores describen en la especie murina que la RA en su gran mayoría se inicia antes del contacto del espermatozoide con la ZP. Esta RA aparentemente comenzaría con el contacto de los espermatozoides con las células del *cumulus oophorus* (Jin et al., 2001; Yanagimachi 2011).

5.2.2 Funciones particulares de las glicoproteínas de la zona pelúcida

ZP1: se le atribuyó inicialmente una función estructural basada en el modelo de ratón (Wassarman *et al.*, 1988; Rankin, *et al.*, 1999), aportando estabilidad e integridad estructural a la matriz (Greve y Wassarman, 1985; Rankin *et al.*, 1999; Gupta *et al*, 2012). Recientemente, en humanos, nuevas funciones han sido atribuidas a ZP1, relacionándola con la unión y la inducción de la RA (Ganguly *et al.*, 2010a, 2010b)

Un estudio muy reciente, de Huang y colaboradores, reporta una mutación homocigótica de ZP1 en mujeres de la misma familia responsable de alteraciones ovocitarias. Fenotípicamente algunas de ellas carecen de ovocitos, mientras que en otras no se observó ningún ovocito, mientras que en otras, los ovocitos no tenían ZP y en ambos casos las mujeres son infértiles (Huang *et al.*, 2014). Este estudio muestra una diferencia entre la función de ZP1 en la ratona y la mujer; puesto que, en ratonas KO para ZP1 los ovocitos están rodeados por la ZP aunque esta es más delgada de lo normal y además estas ratonas no son estériles (Rankin *et al.*, 1999).

ZP2: esta proteína se ha relacionado tradicionalmente con la recepción secundaria del gameto masculino, es decir con la unión a espermatozoides reaccionados (Bleil *et al.*, 1988; Gupta *et al*, 2012). En ratón, ZP2 está involucrada en la unión secundaria como ligando a través de su interacción con componentes intraacrosomales (Miller *et al.*, 2002).

Sin embargo, estudios recientes empleando ratones transgénicos con ZP humanizada indican que el espermatozoide humano únicamente se une a la ZP, cuando ésta expresa ZP2, ya sea sola o coexpresada con otras glicoproteínas de la matriz. Esta unión se produce al dominio N-terminal de la proteína, tras la cual se induce el bloqueo a la polispermia (Baibakov *et al.*, 2012; Avella *et al.*, 2014).

Otros estudios recientes en ratón, también señalan a esta glicoproteína como el receptor primario; de modo que, el espermatozoide se uniría a un dominio de la glicoproteína ZP2 próximo a la región N-terminal, siendo el responsable del reconocimiento entre gametos tanto en la especie murina como en la humana. Tras la unión del espermatozoide al dominio N-terminal de esta glicoproteína; se produce la extrusión de los gránulos corticales, descargando una proteasa llamada ovastacina, que induce un cambio en la estructura del dominio N-terminal de ZP2, de modo que perdería su capacidad de unión al espermatozoide induciendo el bloqueo definitivo a la polispermia (Burkart *et al.*, 2012).

ZP3: en ratón se ha considerado como receptor primario, uniéndose al espermatozoide capacitado no reaccionado (Bleil y Wassarman, 1980b; Bleil y Wassarman, 1983; Wassarman, 1990; Gupta *et al.* 2012). Además es un agonista natural de la RA (Bleil y Wassarman, 1983; Wassarman, 1988, 2008). También es inductora de la RA en la especie humana (van Duin *et al.*, 1994; Dong *et al.*, 2001, Chakravarty *et al.*, 2005, 2008; Caballero-Campo *et al.*, 2006; José *et al.*, 2010).

ZP4: en la especie humana, al igual que ZP3 es capaz de inducir RA cuando es expresada en sistemas eucariotas, no así en procariotas (Chakravarty *et al.*, 2005, 2008; Caballero-Campo *et al.*, 2006), siendo su efecto mayor cuando se trata de proteína nativa purificada (Chiu *et al.*, 2008a). Estudios en ratones transgénicos expresando ZP4 humana indican que el espermatozoide humano no es capaz de unirse a esta ZP humanizada, por lo que son necesarias otras glicoproteínas; en particular ZP2, para que se produzca el reconocimiento y la unión por parte del espermatozoide humano (Yauger *et al.*, 2011; Baibakov *et al.*, 2012; Avella *et al.*, 2014).

En la vaca y en la cerda es esta proteína junto con ZP3 las que actúan como receptor primario, siendo ZP4 la que juega un papel más importante en la unión al espermatozoide (Yurewicz *et al.*, 1998; Yonezawa *et al.*, 1997; 2001; 2012).

5.2.3 Características de las proteínas de la zona pelúcida

Las proteínas de la familia ZP se caracterizan por presentar unos dominios estructurales semejantes. Los dominios comunes a esta familia de proteínas son:

Péptido señal: Es un domino hidrofóbico localizado en el extremo N-terminal de la proteína, que la dirige hacia la vía secretora y se elimina en la proteína madura (Boja *et al.*, 2003; Zhao *et al.*, 2003; Gupta *et al.*, 2012).

Dominio ZP: El dominio ZP, podemos encontrarlo en varias proteínas secretoras que juegan un papel importante en el desarrollo, en la audición, en la inmunidad y en el desarrollo de tumores (Bork y Sander, 1992; Jovine *et al.*, 2004, 2006). En el caso de la ZP, se ha visto que participa en la polimerización de la matriz extracelular de las proteínas (Jovine *et al.*, 2002, 2004, 2006). Por otro

lado, experimentos realizados con ZP1 recombinante humana, indican que este dominio es capaz de unirse al espermatozoide intacto e inducir RA (Ganguly *et al.*, 2010b).

Este dominio presenta unos 260 aminoácidos con de 8 (ZP3) a 10 (ZP1, ZP2 y ZP4) residuos de cisteína conservados (Jovine *et al.*, 2006; Wassarman, 2008; Monné y Jovine, 2011). A su vez está formado por dos subdominios, el subdominio C-terminal (ZP-C) y el subdominio N-terminal (ZP-N) (Jovine *et al.*, 2004), pudiendo haber varias copias del subdominio ZP-N, con la excepción de la glicoproteína ZP3 (Callebaut *et al.*, 2007) (Fig.3). Estas copias podrían participar en la especificidad de unión al espermatozoide de ZP1, ZP2 y ZP4 (Callebaut *et al.*, 2007).

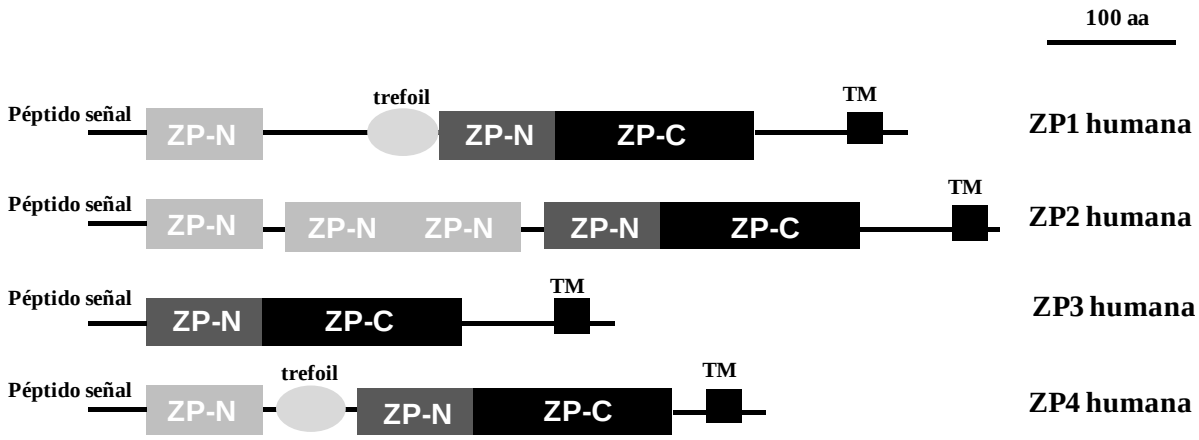


Figura 3. Subdominios ZP-C y ZP-N del dominio ZP en ZP1, ZP2, ZP3 y ZP4 humana. ZP3 no presenta copias de estos subdominios. ZP1 y ZP4 presentan una copia de ZP-N y ZP2 presenta tres copias de ZP-N (Modificado de Callebaut *et al.*, 2007).

La ZP-N debería ser considerado un dominio por sí solo, ya que se ha visto que puede inducir la polimerización individualmente (Jovine *et al.*, 2006). Estudios recientes, detectan dos péptidos en el dominio ZP-N de ZP1 humana con

propiedades amiloideas, induciendo el depósito de material fibroso insoluble, siendo los responsables de la polimerización de la matriz (Louros *et al.*, 2013).

Dominio trefoil: sólo se ha identificado en las glicoproteínas ZP1 y ZP4. Se trata de una región rica en cisteínas que forma una estructura de tres bucles unidos por puentes disulfuro (Braun *et al.*, 2009). Se le han atribuido funciones estructurales dada la estabilidad de los péptidos trefoil gastrointestinales a la digestión por proteasas (Braun *et al.*, 2009).

Sitio de corte de furina: secuencia de 4 aminoácidos RX(K/R)R, donde tendría lugar la escisión proteolítica de las distintas ZPs y su liberación a la vía secretora (Kiefer y Sailing, 2002; Zhao *et al.*, 2003; Tian *et al.*, 2011). Este dominio no está perfectamente conservado en todas las especies, como sucede en ZP4 de la mujer y la gata (Zhao *et al.*, 2003).

Dominios hidrofóbicos: Se han identificado dos motivos hidrofóbicos duplicados: el dominio hidrofóbico interno (IHP), localizado a nivel del dominio ZP y el dominio hidrofóbico externo (EHP), localizado entre el sitio consenso de corte de furina y el dominio transmembrana. Juntos evitan la polimerización de la proteína en el citoplasma celular e intervienen en la secreción e incorporación de las glicoproteínas a la ZP (Zhao *et al.*, 2003; Jovine *et al.*, 2004; Monné y Jovine, 2011).

Dominio transmembrana: se trata de un dominio hidrofóbico localizado en el extremo C-terminal de la proteína. El dominio transmembrana de ZP2 y ZP3 es requerido para la localización de estas glicoproteínas en la membrana del

ovocito, donde se procederá al procesamiento del extremo carboxilo terminal y la incorporación de las glicoproteínas a la ZP (Hoodbhoy *et al.*, 2006).

Tallo citoplasmático: El dominio transmembrana se continúa con un tallo citoplasmático hidrofílico que también evita la polimerización intracelular de las proteínas y asegura su correcta incorporación a la matriz (Jiménez-Movilla y Dean, 2011).

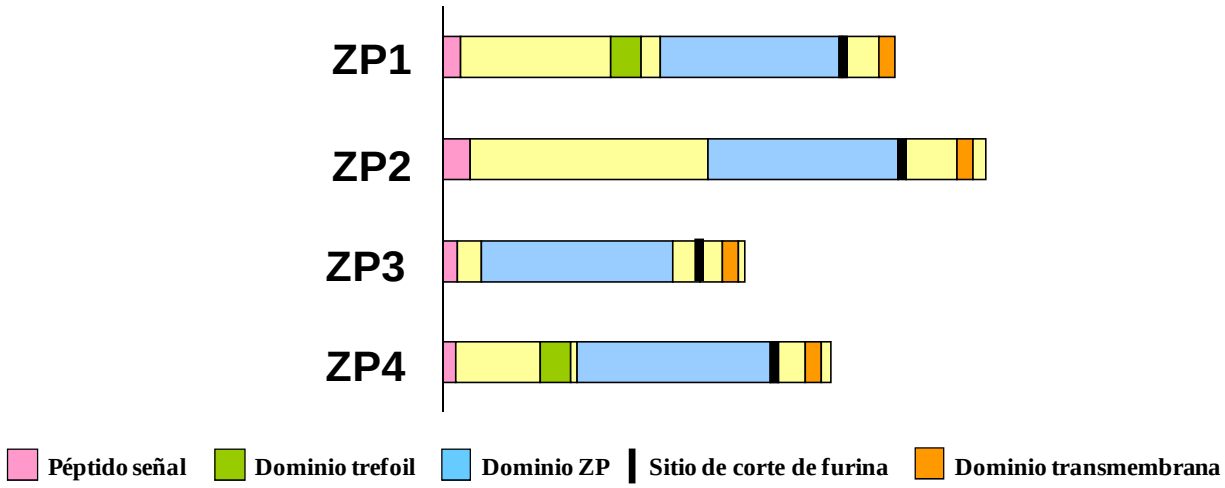


Figura 4 . Descripción esquemática de los diferentes dominios presentes en las glicoproteínas de la ZP (Modificado de Conner *et al.*, 2005).

5.3 CARACTERÍSTICAS MORFOLÓGICAS DE LA ZP

5.3.1. Grosor

El grosor de la ZP varía entre especies: 6 µm en la ratona, 15-20 µm en la mujer y en la cerda y 27 µm en la vaca (Wassarman, 1988; Dunbar *et al.*, 1994; Pelletier *et al.,* 2004), La diferencia en grosor implica que la cantidad proteica varía de una especie a otra. Así, el contenido proteico oscila entre 1 y 30 ng. El peso de la ZP de ratona es de aproximadamente 3,5 ng, mientras que en la mujer es 30 a 33 ng y en la vaca 30 ng (Wassarman, 1988; 2008; Topper *et al.*, 1997; Nakano y Yonezawa, 2001).

5.3.2 Estructura

En la mayoría de las especies la ZP es secretada morfológicamente en capas y encontramos distinta asimetría entre una capa interna y una capa externa (Philips y Shalgi, 1980a; Ahuja y Bolwell, 1983; Shalgi y Raz, 1997).

Se ha descrito una correlación entre el grado de madurez del ovocito y la estructura de la ZP tanto en la ratona (Calafell *et al.*, 1992), como en la mujer (Familiari *et al.*, 1988), describiéndose en esta última una ZP homogénea y compacta en ovocitos en Metafase I y mucho más porosa en ovocitos en metafase II (Tesarík *et al.*, 1988).

Los estudios mediante microscopía electrónica de barrido de Vanroose y su grupo en 2000 describen que en ovocitos y mórulas de vacas la superficie ZP mostró una apariencia áspera y esponjosa con numerosos poros; sin embargo, en cigotos, se observó que la superficie ZP tenía un aspecto más liso, con un menor número de poros (Vanroose *et al.*, 2000).

Cuando se utilizan técnicas más sofisticadas, con alta resolución, para estudiar la ZP en mamíferos, se observa una estructura de finos filamentos

interconectados, con un patrón regular de alternancia "mallas amplias y estrechas" (Familiari *et al.*, 2006). La superficie externa de la ZP presenta una apariencia de "queso suizo" mientras que la superficie interna muestra una apariencia regular y rugosa (Phillips y Shalgi, 1980; Familiari *et al.*, 1992; Keefe *et al.*,1997) (Fig.5).

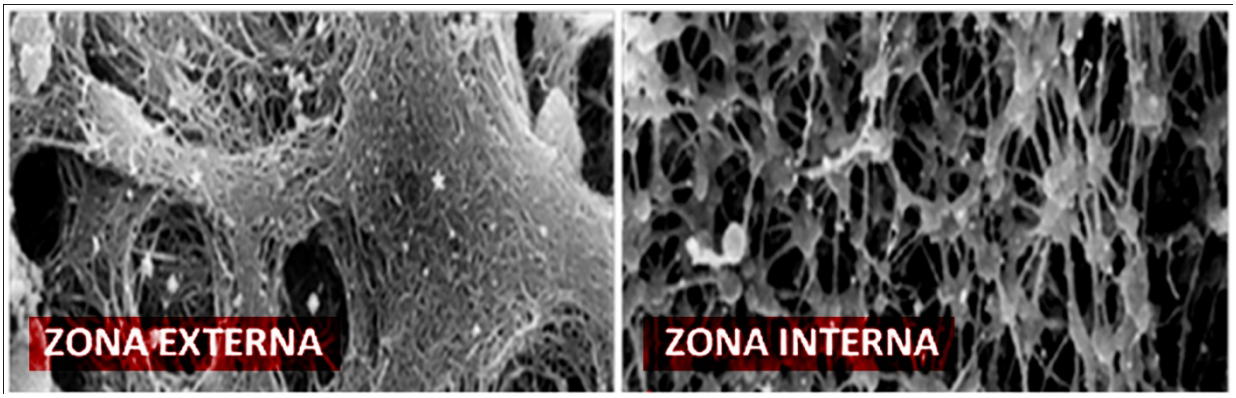


Figura 5. Microscopía electrónica de barrido de las zonas externa e interna de la ZP (modificado de Familiari *et al.,* 1992).

En la ZP se puede visualizar tres capas bien definidas al ser observada con PolScope: una interna, siendo la más birrefringente, brillante y gruesa; una media, la de menor grosor, que se observa oscura y con poca birrefringencia; y una externa con mayor grosor y menor birrefringencia que la anterior (Fig. 6). La apariencia de las mismas se emplea como marcador de calidad ovocitaria y desarrollo embrionario (Pelletier *et al.*, 2004; Shen *et al.*, 2005; Rama Raju *et al.*, 2007).

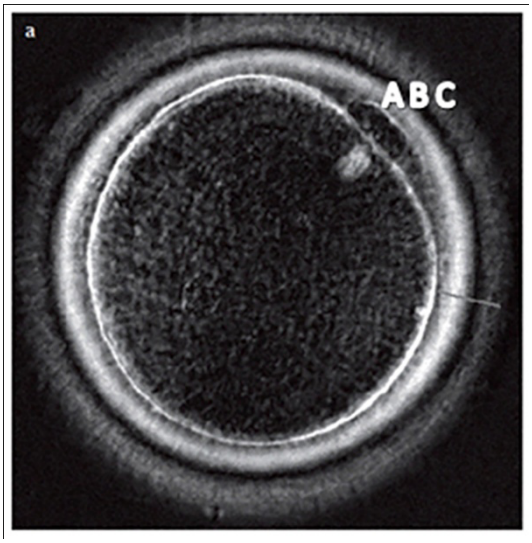


Figura 6 . Imagen de un ovocito observado con un microscopio Polscope. Se observan las tres capas que constituyen la ZP señaladas con A, B y C .Tomado de (Rama Raju *et al*., 2007).

5.3.3 Estructura molecular de zona pelúcida

En el modelo murino, es el mejor conocido hasta el momento, se describe que la ZP del ratón está formada por un entramado de fibras, las cuales están constituidas por las distintas glicoproteínas (Wassarman *et al.*, 1996). Las fibras se componen de dímeros de ZP2:ZP3 unidos entre sí por moléculas de ZP1. Sin embargo, Dean en 2004 propone que los filamentos que forman la ZP pueden estar compuestos por dímeros de ZP1:ZP3 y ZP2:ZP3 (Fig.7).

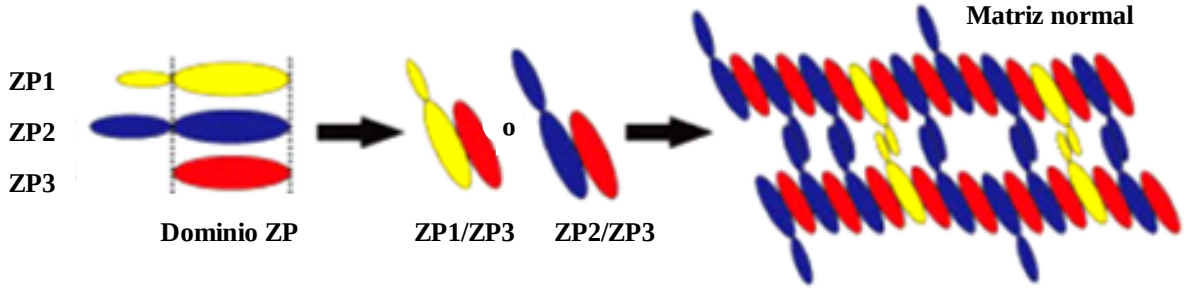


Figura 7. Modelo estructural de la ZP murina (Dean, 2004).

El modelo más reciente es una variación del modelo de Greve y Wassarman de 1985, de modo que, la glicoproteína ZP1 se incorpora a los largos filamentos a través de su dominio ZP. Por tanto en la ratona, la ZP estaría formada por un entramado fibrilar constituido por largos polímeros de ZP1-ZP2-ZP3 los cuales se unen entre sí mediante homodímeros de ZP1 a través de enlaces disulfuro formando una estructura tridimensional (Fig. 8) (Monné y Jovine, 2011).

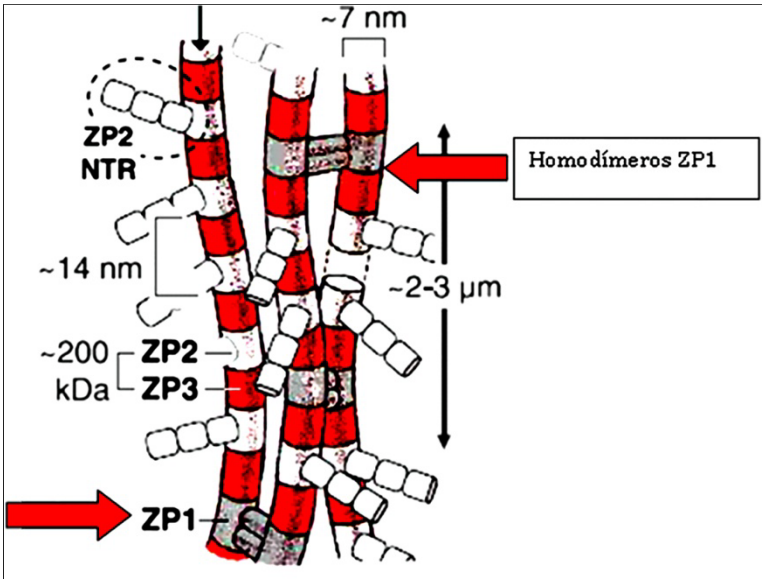


Figura 8. Modelo propuesto para ZP de ratón con 4 glicoproteínas. La ZP1 se incorpora a los filamentos de la ZP a partir de su domino ZP (Modificado de Monné y Jovine 2011).

En los casos de especies de mamíferos que presentan una composición glicoproteica diferente con 4 ZPs, o con 3 proteínas pero siendo éstas ZP2, ZP3 y ZP4 posiblemente tengan una estructura similar a la descrita anteriormente. Algunos autores proponen que las que presentan 4 proteínas estarían formadas por largos filamentos de ZP2 y ZP3 que estarían entrelazados por distintos tipos de filamentos: homodímeros de ZP1, homodímeros de ZP4 y heterodímeros de ZP1 y ZP4 (Florman y Ducibella, 2006) (Fig. 9).

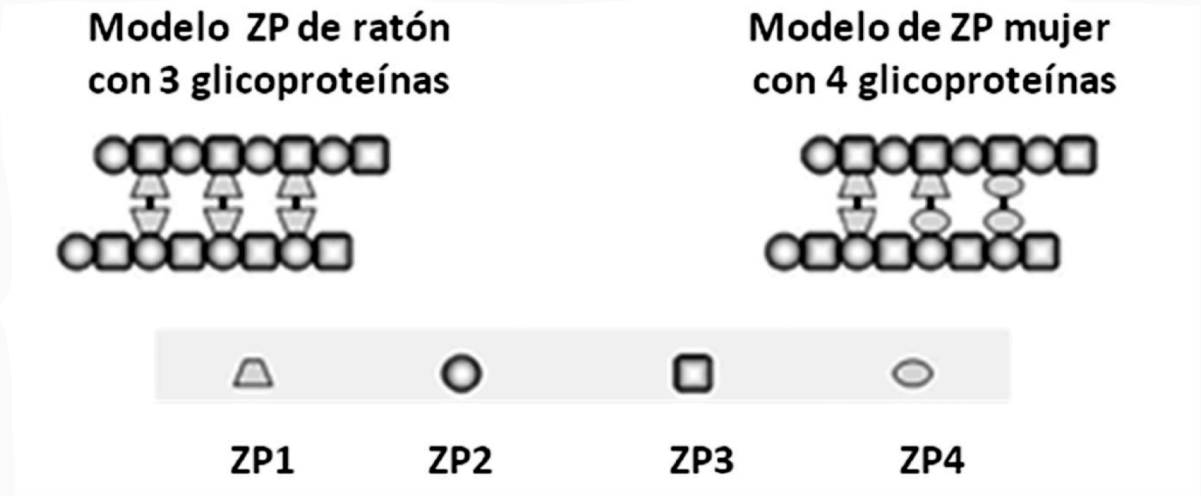


Figura 9. Esquema para 3 glicoproteínas (ratón) y modelo para 4 glicoproteínas (mujer) modificado de (Florman y Ducibella, 2006).

5.4 SÍNTESIS DE LAS GLICOPROTEÍNAS DE LA ZONA PELÚCIDA

El origen celular de las glicoproteínas de la ZP ha sido motivo de controversia durante largo tiempo. Las proteínas de la ZP de los mamíferos y las proteínas de la envoltura vitelina (EV) de los anfibios son sintetizadas en el ovario. Por otro lado las proteínas de la EV de aves y peces son producidas en el hígado, en el ovario o en ambas localizaciones según la especie (Litscher y Wassarman, 2007) (Tabla 2)

Tabla 2. Procedencia celular de las glicoproteínas de la ZP y de la VE en los diferentes taxones.

	Ovario	Hígado
Mamíferos (ZP)	+	-
Peces (EV)	+	+
Aves (EV)	+	+
Anfibios (EV)	+	-

Así, en los mamíferos, en la síntesis de la ZP que tiene lugar exclusivamente en el ovario pueden intervenir el ovocito o el ovocito junto a las células de la granulosa (Wassarman, 1988, 2008; Maresh *et al.*, 1990; Sinowatz *et al.*, 2001; Wassarman y Litscher, 2013).

Tradicionalmente, tomando como referencia el modelo del ratón (*Mus musculus*), se consideraba que la ZP de mamíferos era únicamente sintetizada por el ovocito (Haddad y Nagai, 1977; Bleil y Wassarman, 1980a; Wassarman, 1988; Epifano *et al.*, 1995). Sin embargo, en pocas especies: ratón (Haddad y Nagai, 1977; Epifano *et al.*, 1995; Eberspaecher *et al.*, 2001; Wassarman, 2008), rata (Akatsuka *et al.*, 1998; Scobie *et al*, 1999), hámster (Bousquet *et al.*, 1981; Izquierdo-Rico *et al.*, 2011), ratón marsupial de cara rayada (Au *et al.*, 2008) y zarigüeya australiana (Voyle *et al.*, 1999), se señala a esta célula como la única

responsable de su síntesis. En el resto de mamíferos estudiados habría una intervención conjunta del ovocito y las células de la granulosa.

Por tanto, en relación al lugar de la síntesis de las glicoproteínas de la ZP podemos encontrar dos tipos de modelo y dentro de cada especie una cinética de expresión propia (Maresh *et al.*, 1990; Kölle *et al*, 1996, 1998; Konrad *et al.*, 2012).

5.4.1 Síntesis de las glicoproteínas en coneja

Las proteínas de la ZP del conejo se sintetizan en los folículos durante las primeras etapas de la foliculogénesis (Dunbar *et al* 1994). En su síntesis intervienen el ovocito y las células de la granulosa (Dunbar *et al* 1994; Wolgemuth *et al.*, 1984; Schwoebel *et al.*, 1991; Grootenhuis *et al.*, 1996; Lee *et al.*, 1993; 2000).

En la coneja, ZP2 se detecta en ovocitos de folículos primordiales, primarios y secundarios tempranos, se describe además una mayor expresión en ovarios inmaduros de conejas (2-6 semanas de edad) que en ovarios maduros (8-12 semanas de edad) (Lee *et al.*, 1993). Por otra parte el ARNm de ZP3 también se expresa en mayor proporción (600 veces) en ovarios sexualmente inmaduros (animales de 6 semanas), es decir ovarios con una mayor proporción de folículos primordiales que en ovarios sexualmente maduros (animales de más de 6 meses) (Grootenhuis *et al.*, 1996) (Fig.10).

Por otro lado, ZP4 se expresa en folículos primordiales, primarios y secundarios tempranos (Lee, 2000).

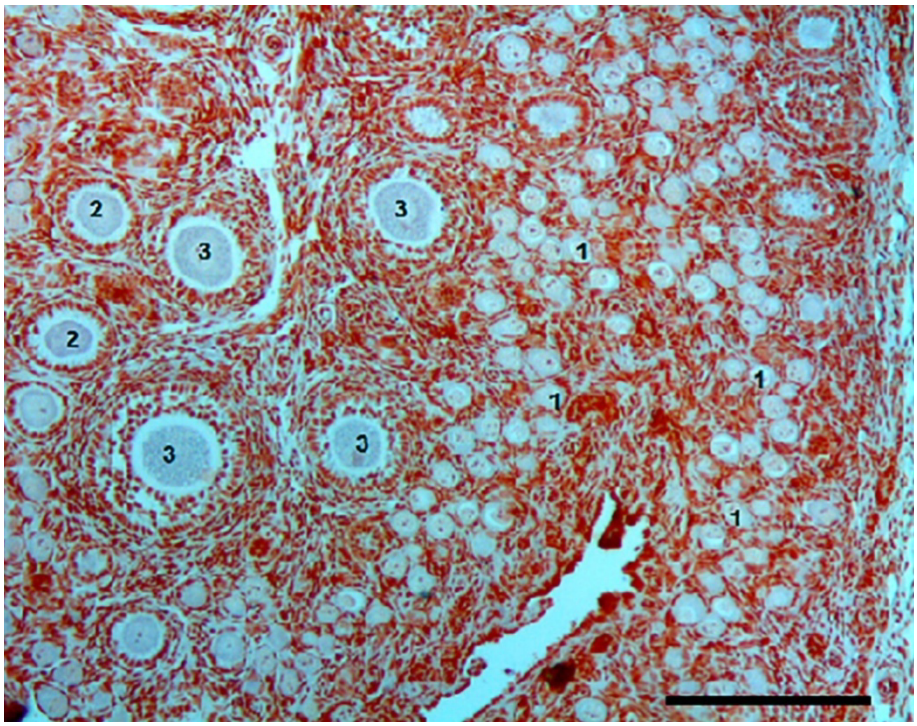


Figura10. Corte histológico de ovario de conejo. Se observan folículos en diferente estadio de crecimiento. Folículos primordiales (1), folículos primarios unilaminares (2) y folículos primarios multilaminares (3). Barra: 100 µm

5.4.2 Síntesis de las glicoproteínas de la zona pelúcida en la gata

En la gata (*Felis catus*), estudios realizados por Jewgenow y Rudolph en el año 2001 indicaban que en la síntesis de las diferentes glicoproteínas únicamente intervenían las células de la granulosa; sin embargo, estudios de ese mismo año realizados por Barber y colaboradores indican que en la síntesis interviene también el ovocito (Baber *et al.*, 2001). Se describe una síntesis secuencial, primero se detecta ZP4 en folículos primarios en crecimiento y secundarios y posteriormente se detectan ZP2 y ZP3 únicamente en folículos secundarios (Jewgenow y Fickel, 1999).

Ni los transcritos ni las proteínas de la ZP fueron detectables en los folículos primordiales.

El estudio de la expresión secuencial de proteínas de la ZP del gato es de relevancia práctica para el diseño de vacunas anticonceptivas (Fig.11).

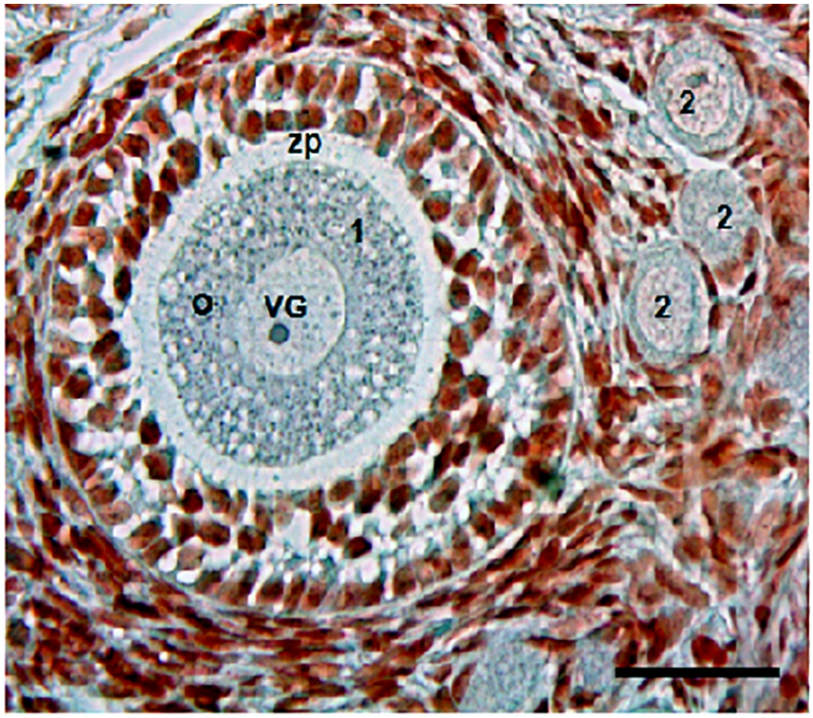


Figura 11. Corte histológico de ovario de gato. Se observa un folículo primario multilaminar (1) y varios folículos primordiales (2). Señalamos la zona pelúcida (zp) el ooplasma (O) y la vesícula germinal (VG). Barra: 50 µm

5.4.3. Traducción y procesamiento intracelular de las glicoproteínas de la zona pelúcida

Poco se conoce acerca del proceso de tráfico intracelular o de los mecanismos por los cuales las proteínas son secretadas para formar la ZP. Lo que si se admite es que estos mecanismos están bastante conservados en las diferentes especies, así aunque el anfibio *Xenopus laevis* (rana de uñas africana) y los mamíferos evolutivamente divergieron hace más de 300 millones de años, si microinyectamos ARNm murino que codifica para ZP1, ZP2 y ZP3 estos ARNm son traducidos y las proteínas incorporadas en la membrana vitelina que rodea a los ovocitos de *Xenopus* (Doren *et al.*, 1999).

Durante la síntesis y antes de la incorporación en la ZP extracelular, las proteínas de la ZP sufren importantes modificaciones. El péptido señal que se encuentra formado por aproximadamente los 20 primeros aminoácidos es cortado y resulta un extremo N-terminal de la proteína donde la glutamina es ciclada a piroglutamina (Ringuette *et al.*, 1988; Boja *et al.,* 2003). También son cortados o procesados el dominio transmembrana y el tallo citoplasmático puesto que anticuerpos que reconocen estas secuencias no reaccionan contra la ZP secretada (Boja *et al.,* 2003).

El tráfico intracelular de la ZP3 ha sido monitorizado mediante la expresión de ZP3 unida a EGFP (proteína fluorescente verde) en ratones transgénicos. La visualización de esta glicoproteína anclada al fluorocromo indica que la ZP3 es conducida por el péptido señal en la ruta de síntesis del ovocito desde el retículo endoplasmático al aparato de Golgi donde es glicosilada (Zhao *et al.*, 2003). En el citoplasma de ovocitos en crecimiento encontramos unas estructuras inusualmente grandes (0,5-3,0 μm) en las que encontramos anclada ZP3-EGFP (Qi *et al.,* 2002; Zhao *et al.,* 2003). La localización periférica de ZP3-EGFP en estas estructuras circulares ancladas a lípidos de membrana y la copresencia de synaptobrevina (VAMP), proteína de membrana asociada a vesículas (Lin y Scheller, 2000), sugirieron que estas estructuras surgirían del compartimento post-Golgi como vesículas secretoras (Qi *et al.,* 2002). Sin embargo, estas estructuras circulares surgen en etapas más tempranas en la ruta de síntesis puesto que se observó mediante microscopía electrónica en ovocitos de rata que estas formaciones concéntricas pertenecían a retículo endoplasmático (Kang, 1974) y mediante microscopía de fluorescencia, utilizando un anticuerpo anti-PDI (proteína disulfuro isomerasa) localizada en membranas de retículo endoplasmático, se pudo determinar que estas "mega-vesículas" en realidad eran retículo endoplasmático rugoso circular (Hoodbhoy *et al.*, 2006).

En estudios acerca del tráfico intracelular de ZP2 y ZP3 determinan que no hay interacciones físicas entre las dos proteínas durante el tráfico en el interior del ovocito. Las observaciones de que ZP3 es incorporada a una fina ZP en

ausencia de ZP2 (Rankin *et al.*, 2003) y que ZP2 está presente en la membrana plasmática en ausencia de ZP3 (Hoodbhoy *et al.*, 2006) están de acuerdo con la afirmación de que ZP2 y ZP3 progresan independientemente en el ovocito antes de ensamblarse para formar la matriz de la ZP (Hoodbhoy *et al.*, 2006).

Además, los análisis realizados mediante espectrometría de masas de las proteínas de la ZP demuestran el procesamiento de estas proteínas a estos dos niveles. A nivel del extremo carboxi-terminal el sitio de corte se realiza en un dominio bastante conservado RXK/RR reconocido por furina. El corte se sitúa efectivamente dentro del sitio consenso para corte de furina (justo en la mitad), entre los aminoácidos RX y los K/RR (Boja *et al.*, 2005, Hoodbhoy *et al.*, 2005). Sin embargo, el hecho de que la mutación del motivo RNRR a ANAA o RNGE no impida la secreción o incorporación de la ZP3 en la ZP sugiere que tiene que haber un sitio alternativo de procesamiento (Kiefer y Saling, 2002; Qi *et al.,* 2002; Zhao *et al.,* 2002).

Recientemente, también se ha observado la presencia de una región hidrofóbica que se encuentra entre el dominio furina y el dominio transmembrana. Esta región se encuentra bien conservada en los mamíferos y su posición es semejante en las tres glicoproteínas (ZP1, ZP2 y ZP3). La mutación de esta zona hidrofóbica (VTVGPLIFL) en ZP3 impide la incorporación de la misma en la ZP (Zhao *et al.,* 2003). Sin embargo, el papel de este pequeño dominio debe ser determinado. Se especula que podría servir como sitio de unión de proteínas tipo chaperonas importantes para el tráfico intracelular de las proteínas de la ZP.

En cuanto a los dominios transmembrana de ZP2 y ZP3 hay que destacar que pueden ser requeridos para la colocalización de las proteínas en la membrana plasmática del ovocito donde es procesado el extremo C-terminal para la incorporación de las proteínas a la ZP. Es interesante resaltar que este dominio transmembrana no está presente en las glicoproteínas de la ZP de peces que son sintetizadas en el hígado y que son transportadas vía sanguínea para su

incorporación a la ZP (Darie *et al.*, 2004, 2005; Sugiyama *et al.*, 1999; Hyllner *et al.*, 2001).

En cuanto a las regiones IHP y EHP se ha visto que van a permanecer unidas durante el transporte de las proteínas de la ZP hacia el exterior. Posteriormente, cuando se realice el corte de la proteína por el CFCS, se romperá dicha unión entre el IHP y el EHP, posibilitándose así que el péptido que atraviese la membrana plasmática y que alcance la ZP pueda adquirir su conformación tridimensional y unirse al resto de las proteínas de la ZP para la formación de fibrillas (Jovine et al., 2004).

El tallo citoplasmático hidrofílico también evita la polimerización intracelular de las proteínas y asegura su correcta incorporación a la matriz (Jiménez-Movilla y Dean, 2011).

5.5 EVOLUCIÓN MOLECULAR DE LAS GLICOPROTEÍNAS DE LA ZONA PELÚCIDA

Una evolución rápida de las proteínas involucradas en la reproducción ha sido documentada en una amplia variedad de taxones desde las diatomeas hasta los primates (Singh y Kulathinal, 2000; Swanson y Vacquier 2002). Así, se ha documentado en *Drosophila* que la evolución de las proteínas reproductoras es el doble de rápida que la que presentan las proteínas no relacionadas con la reproducción (Civetta y Singh, 1995). Además, se observa una evolución más rápida de las proteínas reproductoras masculinas que de las femeninas (Haerty *et al.*, 2007). Se ha sugerido que estas proteínas podrían evolucionar como resultado de una evolución adaptativa (selección positiva) (Swanson y Vacquier, 2002).

Estudios recientes sugieren que la coevolución entre pares de proteínas de los dos gametos que interaccionan en la fecundación es la principal fuerza que lidera la evolución adaptativa de las proteínas reproductoras (Swanson y Vacquier, 2002). Uno de los miembros del par espermatozoide-ovocito cambia primero y el otro miembro se adapta a este cambio para maximizar la interacción entre ambos. Por ejemplo se ha descrito la coevolución de la proteína CD9 del ovocito y la proteína IZUMO1 del espermatozoide en primates (Claw *et al.*, 2014).

En diversos organismos y a diferentes niveles del proceso reproductivo se han identificado proteínas que evolucionan de manera rápida. Podemos destacar, en la hembra, algunos ejemplos de proteínas involucradas en la reproducción bajo selección positiva. Una de ellas es la oviductina, ésta se ha relacionado con la regulación de la unión ovocito-espermatozoide contribuyendo al control de la polispermia (Coy *et al.*, 2008) estando, esta proteína, también sometida a una selección positiva (Swanson *et al.*, 2001).

También en la oreja de mar destacamos la proteína sp18. Esta proteína del espermatozoide ha sido relacionada con la fusión de gametos y quizás sea la proteína de metazoos que más rápidamente evoluciona de todas las descritas

hasta ahora (Swanson y Vacquier, 1995). Evoluciona 50 veces más rápido que la proteína de más rápida evolución de los mamíferos.

En relación a las proteínas de la ZP, decir que algunas de las glicoproteínas de la ZP han sido objeto de estudio. ZP3 y ZP2 son las proteínas constituyentes de la ZP que más han sido estudiadas desde el punto de vista evolutivo siendo consideradas como proteínas de evolución rápida. De hecho, ZP2 y ZP3 están entre el 10% de las proteínas que más difieren cuando comparamos los roedores y los humanos. Son muchos los autores que apoyan su evolución positiva (Makalowski y Boguski, 1998; Swanson *et al.*, 2001a; Swanson y Vacquier, 2002; Turner y Hoekstra, 2006, 2008). Regiones concretas de estas glicoproteínas han sido señaladas por estar sometidas a esta selección positiva, por ejemplo en ZP3 se detectaron múltiples regiones bajo selección positiva, la mayoría de las cuales se encontraban entre los aminoácidos 331 y 373 (Swanson *et al.*, 2001a).

En el ratón se ha descrito que los aminoácidos 324, 325 y 341 del exón 7 son sometidos a esta selección; región que se ha descrito como esencial en la interacción con el espermatozoide (Rosiere y Wassarman, 1992; Kinloch *et al.*, 1995; Wassarman y Litscher, 2009). Sin embargo otros autores descartan esta selección positiva (Berlin y Smith, 2005; Chen *et al.*, 2011; Meslin *et al.*, 2012), Berlin y Smith en el 2005 reanalizaron los datos de Swanson y no encontraron indicios de selección positiva para la glicoproteína ZP3 y atribuyen los resultados anteriores a falsos positivos originados por el método usado de análisis. Del mismo modo, los resultados propuestos por Jansa y colaboradores en 2003, donde señalan la presencia de selección positiva en ZP3 en varias especies del género *Mus*, son descartados cuando Turner y Hoekstra en 2006 reanalizan los resultados en ausencia del grupo externo empleado por los primeros autores.

En el año 2011, los análisis en 7 especies de bóvidos realizados por Chen y colaboradores, no encuentran indicios de selección positiva en ZP3, mientras que si detectan indicios, aunque débiles en ZP2 (aminoácidos 38, 117, 174 y 342). Lo que podría indicar que la ZP2 de mamíferos podría jugar un papel en la

interacción con el espermatozoide más importante que el que se había pensado hasta el momento (Chen *et al.*, 2011).

Estudios de Meslin *et al.* (2012) encuentran indicios de selección positiva para ZP2 y ZP4 en la rana y para ZP4 en el macaco de Rhesus. Sin embargo, al analizar otros genes sometidos a selección positiva concluyen que los aminoácidos bajo esta selección no se encuentran localizados en el dominio relacionado con la función principal de la proteína.

6. BIBLIOGRAFÍA

6. BIBLIOGRAFÍA

Ahuja KK, Bolwell GP(1983). Probable asymmetry in the organization of components of the hamster zona. J Reprod Fertil. 69:49-55.

Akatsuka K, Yoshida-Komiya H, Tulsiani DRP, Orgebin-Crist MC, Hiroi M, Araki Y (1998). Rat zona pellucida glycoproteins: Molecular cloning and characterization of the three major components. Mol Reprod Dev. 51:454-467.

Avella MA, Baibakov B, Dean J (2014). A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. J Cell Biol. 205:801-809.

Baibakov B, Boggs NA, Yauger B, Baibakov G, Dean J (2012). Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. J Cell Biol. 197:897-905.

Barber MR, Lee SM, Steffens WL, Ard M, Fayrer-Hosken RA (2001). Immunolocalization of zona pellucida antigens in the ovarian follicle of dogs, cats, horses and elephants. Theriogenology. 1;55:1705-1517.

Barisone GA, Krapf D, Correa-Fiz F,Arranz SE et al.(2007). Glycoproteins of the vitelline envelope of Amphibian oocyte: biological and molecular characterization of ZPC component (gp41) in Bufo arenarum. Mol. Reprod Dev.74, 629-640. Erratum in: Mol. Reprod. Dev 2009,76.689.

Barros C,Yanagimachi R.(1971). Indution of zona reaction in golden hamster egges by cortical granule material. Nature. 233: 268-269.

Bausek N, Waclawek M, Wolfgang JS, Wohlrab F (2000). The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. J Biol Chem. 275: 28866-28872.

Benoff S (1997). Carbohydrates and fertilization: an overview. Molecular Human Reproduction. 3:599-637.

Berger T, Turner KO, Meizel S, Hedrick JL (1989). Zona pellucida-induced acrosome reaction in boar sperm. Biology of Reproduction. 40:525-530.

Berlin S, Smith NG (2005). Testing for adaptative evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. BMC Evol Biol.5:65-74.

Blackmore DG, Baillie LR, Holt JE, Dierkx L, Aitken RJ, McLaughlin EA (2004). Biosynthesis of the canine zona pellucida requires the integrated participation of both oocytes and granulosa cells. Biol Reprod. 71:661-668.

Bleil JD, Greve JM, Wassarman PM (1988). Identification of a secondary sperm receptor in the mouse egg zona pellucida: role in maintenance of binding of acrosome-reacted sperm to eggs. Developmental Biology 128:376-385.

Bleil JD, Wassarman PM (1980a). Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. Dev Biol. 76:185-202.

Bleil JD, Wassarman PM (1980b). Mammalian sperm-egg Interaction: Identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. Cell. 20:873-882.

Bleil JD, Wassarman PM (1980c). Synthesis of zona pellucida proteins by denuded and follicle-enclosed mouse oocytes during culture in vitro. Proc Natl Acad Sci U S A. 77:1029-1033.

Bleil JD, Wassarman PM (1983). Sperm-egg interactions in the mouse: sequence of events and induction of the acrosome reaction by a zona pellucida glycoprotein. Dev Biol. 95:317-324.

Bodmer M, Ashburner M (1982). Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature. 309-425-430.

Berger T, Turner KO, Meizel S, Hedrick JL (1989). Zona pellucida-induced acrosome reaction in boar sperm. Biology of Reproduction. 40:525-530.

Boja ES, Hoodbhoy T, Fales HM, Dean J (2003). Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem. 278:34189-34202.

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

Bousquet D, Léveillé MC, Roberts KD, Chapdelaine A, Bleau G (1981). The cellular origin of the zona pellucida antigen in the human and hamster. J Exp Zool. 215:215-218.

Braun BC, Ringleb J, Waurich R, Viertel D, Jewgenow K (2009). Functional role of feline zona pellucida protein 4 trefoil domain: a sperm receptor or structural component of the domestic cat zona pellucida?. Reprod Domest Anim. 2:234-238.

Burkart AD, Xiong B, Baibakov B, Jiménez-Movilla M, Dean J (2012). Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. J Cell Biol. 197:37-44.

Caballero-Campo P, Chirinos M, Fan XJ, González-González ME, Galicia-Chavarría M, Larrea F, Gerton GL (2006). Biological effects of recombinant human zona pellucida proteins on sperm function. Biol Reprod. 74:760-768.

Calafell JM, Nogués C, Ponsà M, Santaló J, Egozcue J (1992). Zona pellucida surface of immature and in vitro matured mouse oocytes: analysis by scanning electron microscopy. J Assist. Reprod Genet. 9: 365-372

Callebaut I, Mornon JP, Monget P (2007). Isolated ZP-N domains constitute the N-terminal extensions of Zona Pellucida proteins. Bioinformatics. 23:1871-1874.

Cánovas S, Coy P Gomez E (2007). First steps in the development of a functional assay for human sperm using pig oocytes. J Androl. 28: 273-81.

Chakravarty S, Kadunganattil S, Bansal P, Sharma RK, Gupta SK (2008). Relevance of glycosylation of human zona pellucida glycoproteins for their binding to capacitated human spermatozoa and subsequent induction of acrosomal exocytosis. Mol Reprod Dev. 75:75-88.

Chakravarty S, Suraj K, Gupta SK (2005). Baculovirus-expressed recombinant human zona pellucida glycoprotein-B induces acrosomal exocytosis in capacitated spermatozoa in addition to zona pellucida glycoprotein-C. Mol Hum Reprod. 11: 365-372.

Chen S, Costa V, Beja-Pereira A (2011). Evolutionary patterns of two major reproduction candidate genes (ZP2 and ZP3) reveal no contribution to reproductive isolation between bovine species. BMC Evol Biol. 11:24-34

Chiu PC, Wong BS, Lee CL, Pang RT, Lee KF, Sumitro SB, Gupta SK, Yeung WS (2008a). Native human zona pellucida glycoproteins: purification and binding properties. Hum Reprod. 23:1385-1393.

Civetta A, Singh RS (1995). High divergence of reproductive tract proteins and their association with postzygotic reproductive isolation in Drosophila melanogaster and Drosophila virilis group species. J Mol Evol. 41:1085-1095.

Claw KG, George RD, Swanson WJ (2014). Detecting coevolution in mammalian sperm-egg fusion proteins. Mol Reprod Dev. 81:531-538.

Conner SJ, Hughes DC (2003). Analysis of fish ZP1/ZPB homologous genes evidence for both genome duplication and species-specific amplification models of evolution. Reproduction.. 126:347-352.

Coy P, Cánovas S, Mondéjar I, Saavedra MD, Romar R, Grullón L, Matás C, Avilés M (2008). Oviduct-specific glycoprotein and heparin modulate sperm-zona pellucida interaction during fertilization and contribute to the control of polyspermy. Proc Natl Acad Sci U S A. 105:15809-15814.

Darie CC, Biniossek ML, Jovine L, Litscher ES (2004). Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. Biochemistry. 43:7459-7478.

Darie CC, Biniossek ML, Gawinowicz MA, Milgrom Y, Thumfart JO, Jovine L, Litscher ES, Wassarman PM (2005). Mass spectrophotometric evidence that proteolytic processing of rainbow trout vitelline envelope proteins takes place on the egg. J. Biol. Chem. 280: 37585-37598.

Dean J (2002). Oocyte-specific genes regulate follicle formation, fertility and early mouse development. J Reprod Immunol. 53:171-180.

Dean J (2004). Reassesing the molecular biology of sperm-egg recognition with mouse genetics. BioEssays. 26:29-38.

Denker HW (200) Structural dynamics and función of early embryonic coats. Cells Tissues Organs 166:180-207.

Dong KW, Chi TF, Juan YW, Chen CW, Lin Z, Xiang XQ, Mahony M, Gibbons WE, Oehninger S (2001). Characterization of the biologic activities of a recombinant human zona pellucida protein 3 expressed in human ovarian teratocarcinoma (PA-1) cells. Am J Obstet Gynecol. 184:835-844.

Doren S1, Landsberger N, Dwyer N, Gold L, Blanchette-Mackie J, Dean J.(1999) Incorporation of mouse zona pellucida proteins into the envelope of Xenopus laevis oocytes. Dev Genes Evol.209:330-9.

Dunbar BS, Avery V, Lee S, Prasad D, Schwahn E, Schwoebel S, Skinner B, Wilkins(1994).The mammalian zona pellucida: its biochemistry, immunochemistry, molecular biology, and developmental expression. Reprod. Fert. Dev. 6:331-347.

Dunbar BS, Timmons TM, Skinner SM, Prasad SV(2001). Molecular analysis of a carbohydrate antigen involved in the structure and function of zona pellucida glycoproteins. Biol Reprod;65:951-60

Eberspaecher U, Becker A, Bringmann P, van der Merwe L, Donner P (2001). Immunohistochemical localization of zona pellucida proteins ZPA, ZPB and ZPC in human, cynomolgus monkey and mouse ovaries. Cell Tissue Res. 303:277-287.

Epifano O, Dean J (1994). Biology and structure of the zona pellucida: a target for immunocontraception. Reprod. Fertil. Dev. 6: 319-330.

Evsikov AV, Graber JH, Brockman JM, Hampl A, Holbrook AE, Singh P, Eppig JJ, Solter D, Knowles BB (2006). Cracking the egg: molecular dynamics and evolutionary aspects of the transition from the fully grown oocyte to embryo. Genes Dev. 20: 2713-2727.

Familiari G, Relucenti M, Heyn R, Micara G, Correr S (2006). Three-dimensional structure of the zona pellucida at ovulation. Microsc Res Tech. 69:415-426.

Familiari G, Nottola SA, Macchiarelli G, Micara G, Aragona C, Motta PM (1992). Human zona pellucida during in vitro fertilization: an ultrastructural study using saponin, ruthenium red, and osmium-thiocarbohydrazide. Mol Reprod Dev. 32:51-61.

Fazeli A, Hage WJ, Cheng FP, Voorhout WF, Marks A, Bevers MM, Colenbrander B (1997). Acrosome-intact boar spermatozoa initiate binding to the homologous zona pellucida in vitro. Biology of Reproduction. 56:430-438.

Florman HM, Ducibella T. (2006). Fertilization in mammals. In: Neill JD, editor. The physiology of reproduction 1. San Diego:Elsevier Press. pp: 55–122.

Florman HM, Storey BT (1982). Mouse gamete interactions: the zona pellucida is the site of the acrosome reaction leading to fertilization in vitro. Developmental Biology. 91:121-130.

Ganguly A, Bukovsky A, Sharma RK, Bansal P, Bhandari B, Gupta SK (2010a). In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. Hum Reprod. 25:1643-1656.

Ganguly A, Bansal P, Gupta T, Gupta SK (2010b). "ZP domain" of human zona pellucida glycoprotein-1 binds to human spermatozoa and induces acrosomal exocytosis. Reproductive Biology and Endocrinology. 8:110-123

Ganguly A, Sharma RK, Gupta SK (2008). Bonnet monkey (Macaca radiata) ovaries, like human oocytes, express four zona pellucida glycoproteins. Mol Reprod Dev. 75:156-166.

Goudet G, Mugnier S, Callebaut I, Monget P (2008). Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. Biol Reprod. 78:796-806.

Greve JM, Wassarman PM (1985). Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. J Mol Biol 181:253-264

Grootenhuis AJ, Philipsen HL, de Breet-Grijsbach JT, van Duin M (1996). Immunocytochemical localization of ZP3 in primordial follicles of rabbit, marmoset, rhesus monkey and human ovaries using antibodies against human ZP3. J Reprod Fertil. 50:43-54.

Gupta SK, Bhandari B (2011). Acrosome reaction: relevance of zona pellucida glycoproteins. Asian Journal of Andrology. 13:97-105.

Gupta SK, Bhandari B, Shrestha A, Biswal BK, Palaniappan C, Malhotra SS, Gupta N (2012). Mammalian zona pellucida glycoproteins: structure and function during fertilization. Cell Tissue Res. 349:665-678.

Haddad A, Nagai ME (1977). Radioautographic study of glycoprotein biosynthesis and renewal in the ovarian follicles of mice and the origin of the zona pellucida. Cell Tissue Res. 177:347-369.

Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS (2007). Evolution inthe fast lane: rapidly evolving sex-related genes in Drosophila. Genetics. 177:1321-1335.

Han L, Monné M, Okumura H, Schwend T, Cherry AL, Flot D, Matsuda T, Jovine L(2010). Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. Cell. Oct 29;143(3):404-15.

Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG (1994). Cloning and characterization of zona pellucida genes and cDNAS from a variety of mammalian species: the ZPA, ZPB and ZPC gene families. DNA Seq. 4:361-393.

Hedrick JL, Wardrip NJ (1987). On the macromolecular of the zona pellucida from porcine oocytes. Dev Biol. 121:478-488.

Hellberg ME, Vacquier VD (1999). Rapid evolution of fertilization selectivity and Lysin cDNA sequences in Teguline gastropods. Mol Biol Evol. 16-839-848.

Herrler A, BeierHM (2000). Early embryonic coats: morphology, function, practical applications. An overview. Cells Tissues Organs.;166:233-246.

Hoodbhoy T,Dean J.(2004). Insights into the molecular basis of sperm-egg recognition in mammals. Reprod. Fet. 127:417-422.

Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J (2005). Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. J Biol Chem. 280:12721-12731.

Hoodbhoy T, Avilés M, Baibakov B, Epifano O, Jiménez-Movilla M, Gauthier L, Dean J (2006). ZP2 and ZP3 traffic independently within oocytes prior to assembly into the extracellular zona pellucida. Mol Cell Biol. 26:7991-7998.

Huang HL, Lv C, Zhao YC, Li W, He XM, Li P, Sha AG, Tian X, Papasian CJ, Deng HW, Lu GX, Xiao HM (2014). Mutant ZP1 in familial infertility. N Engl J Med. 370:1220-1226.

Hughes DC, Barratt CL (1999). Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida. Biochim. Biophys Acta. 1447:303-6.

Hyllner SJ, Westerlund L, Olsson PE, Schopen A (2001).Cloning of rainbow trout egg envelope proteins: members of a unique group of structural proteins. Biol Reprod. Mar. 64:805-811.

Izquierdo-Rico MJ, Jimenez-Movilla M, Llop E, Pérez-Oliva AB, Ballesta J, Gutierrez-Gallego R, Jimenez-Cervantes C, Avilés M (2009a). Hamster zona pellucida is formed by four glycoproteins. J Proteome Res. 8:926-41.

Izquierdo-Rico MJ, Gimeno L, Jiménez-Cervantes C, Ballesta J, Avilés M (2011). Biosynthesis of hamster zona pellucida is restricted to the oocyte. Theriogenology. 75:463-72.

Jansa SA, Lundrigan BL, Tucker PK (2003). Tests for positive selection on immune and reproductive genes in closely related species of the murine genus mus. J Mol Evol. 56:294-307.

Jewgenow K, Fickel J (1999). Sequential expression of zona pellucida protein genes during the oogenesis of domestic cats. Biol Reprod. 60:522-526.

Jewgenow K, Rudolph M (2001). Timing and location of zona pellucida synthesis during oogenesis in domestic cats: an ultrastructural immunohistological investigation. J Reprod Fertil. 57:23-29

Jiménez-Movilla M, Martínez-Alonso E, Castells MT, Izquierdo-Rico MJ, Saavedra MD, Gutiérrez-Gallego R, Fayrer-Hosken R, Ballesta J, Avilés M (2009). Cytochemical and biochemical evidences for a complex tridimensional structure of the hamster zona pellucida. Histol Histopathol. 24:599-609

Jiménez-Movilla M, Dean J (2011). ZP2 and ZP3 cytoplasmic tails prevent premature interactions and ensure incorporation into the zona pellucida. J Cell Sci. 124:940-950.

Jin M, Fujiwara E, Kakiuchi Y, Okabe M, Satouh Y, Baba SA, Chiba K, Hirohashi N (2011). Proc Natl Acad Sci U S A. 108(12):4892-6.

José O, Hernández-Hernández O, Chirinos M, González-González ME, Larrea F, Almanza A, Felix R, Darszon A, Treviño CL (2010). Recombinant human ZP3-induced sperm acrosome reaction: evidence for the involvement of T- and L-type voltage-gated calcium channels. Biochem Biophys Res Commun. 395:530-534.

Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM (2002). The ZP domain is a conserved module for polymerization of extracellular proteins. Nature Cell Biol 4: 457-461.

Jovine L, Qi H, Williams Z, Litscher ES, Wassarman PM (2004). A duplicated motif controls assembly of zona pellucida domain proteins. Proc Natl Acad Sci USA. 101:5922-5927.

Jovine L.,DarieC.C.,Litscher, E.S., Wassarman, P.M., (2005). Zona pellucidadomain proteins. Annu. Rev. Biochem. 74, 83–114.

Jovine L, Janssen WG, Litscher ES, Wassarman PM (2006). The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. BMC Biochem.7:11-19

Kang YH, (1974). Development of the zona Pellucida in the rat oocyte. AM J. Anat 139:535-565.

Keefe, R.G., D.A. Ferrick, J.L. Scott (1997). Cytokine transcription in lymph nodes of cattle in diferent stages of bovine leukemia virus infection. Vet. Immunol. Immunpathol.59: 271-283.

Kiefer SM, Saling P (2002). Proteolytic processing of human zona pellucida proteins. Biol Reprod. 66:407-414.

Kinloch RA, Sakai Y, Wassarman PM (1995). Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. Proc Natl Acad Sci USA. 92:263-267.

Kölle S, Sinowatz F, Boie G, Palma G (1998). Differential expression of ZPC in the bovine ovary, oocyte, and embryo. Mol Reprod Dev. 49:435-443.

Konrad L, Kuhnert S, Nayudu PL, Einspanier R, Hinsch KD, Hinsch E (2012). Quantification of ZP1, ZP2 and ZP3 mRNA of marmoset monkey (Callithrix jacchus) oocytes from periantral and antral follicles. Andrologia. 1:349-353.

Kozak M (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem. 266:19867-19870.

Lee VH (2000). Expression of rabbit zona pellucida-1 messenger ribonucleic acid during early follicular development. Biol Reprod. 63:401-408.

Lee VH, Schwoebel E, Prasad S, Cheung P, Timmons TH, Cook R, Dunbar BS (1993). Identification and structural characterization of the 75-kDa rabbit zona pellucida protein. The Journal of Biological Chemistry. 268:12412-12417.

Lefièvre L, Conner S J, Salpekar A, Olufowobi O, Ashton P, Pavlovic B, Lenton W, Afnan M, Brewis IA, Monk M, Hughes DC, Barratt CL (2004). Four zona pellucida glycoproteins are expressed in the human. Hum Reprod.19:580-586.

Lin RC, Scheller RH. (2000).Mechanisms of synaptic vesicle exocytosis. Annu. Rev. Cell Dev. Biol. 16: 19-49

Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, Stewart CL, Wassarman PM (1996). Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. Proc Nati Acad Sci USA. 93:5431-5436.

Louros NN, Iconomidou VA, Giannelou P, Hamodrakas SJ (2013). Structural analysis of peptide-analogues of human Zona Pellucida ZP1 protein with amyloidogenic properties: insights into mammalian Zona Pellucida formation. PLoS One. 8:e73258.

Makalowski W, Boguski MS (1998). Evolutionary parameters of the transcribed mammalian genome: an analysis of 2820 orthologous rodent and human sequences. Proc Nati Acad Sci USA. 95:9407-9412.

Maresh GA, Dunbar BS (1987). Antigenic comparison of five species of mammalian zonae pellucidae. J Exp Zool.244:299-307.

Maresh GA, Timmons TM, Dunbar BS (1990). Effects of extracellular matrix on the expression of specific ovarian proteins. Biol Reprod. 43:965-976.

Meslin C, Mugnier S, Callebaut I, Laurin M, Pascal G, Poupon A, Goudet G, Monget P (2012). Evolution of genes involved in gamete interaction: evidence for positive selection, duplications and losses in vertebrates. PLoS One. 7: e44548.: 10.1371/journal.pone.0044548.

Miller DJ, Shi X, Burkon H (2002). Molecular basis of mammalian gamete bending. Recent Prog. Horm. Res. 57: 37-73.

Modlinsky J (1970). The role of the zona pellucida in the development of mouse eggs in vivo. Journal of embryology and Experimental Morphology. 23:539-547.

Monné M, Han L, Schwend T, Burendahl S, Jovine L, (2008). Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. Nature. 456:653-657.

Monné M, Jovine L (2011). A structural view of egg coat architecture and function in fertilization. Biol Reprod. 85:661-669.

Mugnier S, Dell'Aquila ME, Pelaez J, Douet C, Ambruosi B, De Santis T, Lacalandra GM, Lebos C, Sizaret PY,Delaleu B, Monget P, Mermillod P,Magistrini M, Meyers SA, Goudet G (2009). New Insights into the Mechanisms of Fertilization: Comparison of the Fertilization Steps, Composition, and Structure of the Zona Pellucida Between Horses and Pigs. Biology of Reproduction. 81:856-870.

Nakano M, Yonezawa N (2001). Localization of sperm ligand carbohydrate chains in pig zona pellucida glycoproteins. Cells Tissues Organs. 168:65-75.

Noguchi S, Yonezawa N, Katsumata T, Hashizume K, Kuwayama M, Hamano S, Watanabe S, Nakano M (1994). Characterization of the zona pellucida glycoproteins from bovine ovarian and fertilized eggs. Biochim Biophys Acta. 1201:7-14.

Pelletier C, Keefe DL, Trimarchi JR (2004). Noninvasive polarized light microscopy quantitatively distinguishes the multilaminar structure of the zona pellucida of living human eggs and embryos. Fertil Steril. 81 Suppl 1:850-856.

Philips DM, Shalgi RM (1980a) Surface architecture of the mouse and hamster zona pellucida and oocyte.J Ultrastruct Res. 72:1-12.

Prasad SV,Wilkins B,Skinner SM, Dumbar BS (1995). Isolation and characterization of a full-length cDNA 726 encoding the 55-kDa rabbit zona pellucida protein. Mol Reprod Dev 43(4):519-529.

Qi H, Williams Z, Wassarman PM (2002). Secretion and assembly of zona pellucida glycoproteins by growing mouse oocytes microinjected with epitope-tagged cDNAs for mZP2 and mZP3. Mol Biol Cell; 13:530-541

Rama Raju GA, Prakash GJ, Krishna KM, Madan K (2007). Meiotic spindle and zona pellucida characteristics as predictors of embryonic development: a preliminary study using PolScope imaging. Reprod Biomed Online. 14:166-174.

Rankin T, Talbot P, Lee E, Dean J (1999). Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development. 126:3847-3855.

Rankin TL, Coleman JS, Epifanio O, Hoodbhoy T, Tuener SG, Castle E, Lee E, Gore-Langton R, Dean J (2003). Fertility and taxxon specific sperm binding persist after replacement of mouse sperm receptors with human homologs. Dev. Cell. 5:33-43.

Rankin TL, O'Brien M, Lee E, Wigglesworth K, Eppig J, Dean J (2001). Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. Development. 128:1119-1126.

Rankin TL, Familari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, Drago J, Westphal H, Dean J (1996). Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. Development. 122:2903-2910

Ringuette, M.J., M.E. Chamberlin, A.W. Baur, D.A. Sobieski, J Dean (1988). Molecular analysis of cDNA coding for ZP3, a sperm binding protein of the mouse zona pellucida. Dev Biol. 127: 287-295.

Rosiere TK, Wassarman PM (1992). Identification of a region of mouse zona pellucida glycoprotein mZP3 that possesses sperm receptor activity. Dev Biol. 154:309-317.

Sasanami T, Pan J, Mori M (2003). Expression of perivitelline membrane glycoprotein ZP1 in the liver of Japanese quail (Coturnix japonica) after in vivo treatment with diethylstilbestrol. J Steroid Biochem Mol Biol. 84:109-16.

Sasanami T, Pan J, Doi Y, Hisada M, Kohsaka T, Toriyama M (2002). Secretion of egg envelope protein ZPC after C-terminal proteolytic processing in quail granulosa cells. Eur J Biochem. 269:2223-2231.

Schwoebel E, Prasad S, Timmons TM, Cook R, Kimura H, Niu 725 EM, Cheung P, Skinner S AveryS E., Wilkins B, and. Dunbar B S (1991) Isolation and characterization of a full-length cDNA 726 encoding the 55-kDa rabbit zona pellucida protein. J Biol Chemistry Inc. Vol. 266, No. Issue of April 15, pp. 72147219,1991 Printed in U. S.

Scobie GA, Kerr LE, MacDuff P, Aitken RJ (1999). Cloning, sequencing and site of origin of the rat sperm receptor protein, ZP3. Zygote. 7:27-35.

Shalgi R, Raz T (1997). The role of carbohydrate residues in mammalian fertilization. Histophatol.12:813-822

Shen Y, Stalf T, Mehnert C, Eichenlaub-Ritter U, Tinneberg HR (2005). High magnitude of light retardation by the zona pellucida is associated with conception cycles. Hum Reprod. 20:1596-1606.

Singh RS, Kulathinal RJ(2000). Sex gene pool evolution and speciation: a new paradigm. Genes Genet Syst. 75(3):119-30

Sinowatz F, Amselgruber W, Töpfer-Petersen E, Totzauer I, Calvete J, Plendl J (1995). Immunocytochemical characterization of porcine zona pellucida during follicular development. Anat Embryol. 191:41-46.

Sinowatz F, Kölle S, Töpfer-Petersen E (2001). Biosynthesis and expression of zona pellucida glycoproteins in mammals. Cells Tissues Organs. 168:24-35.

Spargo SC, Hope RM (2003). Evolution and nomenclature of pellucida gene family. Biol. Reprod.68:358-362.

Sugiyama H, Murata K, Iuchi I, Nomura K, Yamagami K. (1999)Formation of mature egg envelope subunit proteins from their precursors (choriogenins) in the fish, Oryzias latipes: loss of partial C-terminal sequences of the choriogenins. J Biochem. 125: 469-475.

Swanson WJ, Vacquier VD. (1995) Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc Natl Acad Sci U S A. 92:4957-4961.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001a). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proc Natl Acad Sci USA.98:2509-2514.

Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF (2001b). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc Natl Acad Sci U S A. 98:7375-7379.

Swanson WJ, Vacquier VD (2002). The rapid evolution of reproductive proteins. Nat RevGenet.3:137-44

Tanihara F, Nakai M, Kaneko H, Noguchi J, Otoi T, Kikuchi K (2013). Evaluation of zona pellucida function for sperm penetration during in vitro fertilization in pigs. J Reprod Dev. 59:385-392.

Tesarík J, Pilka L, Trávník P (1988). Zona pellucida resistance to sperm penetration before the completion of human oocyte maturation. J Reprod Fertil. 83:487-495.

Tian J, Gong H, Thomsen GH, Lennarz WJ (1997). Gamete interactions in Xenopus laevis: identification of sperm binding glycoproteins in the egg vitelline envelope. J Cell Biol. 136:1099-108.

Topper E K, Kruijt L, Calvete J, Mann E, Topfer-Petersen E, Woelders H. (1997) Identification of bovine zona pellucida glycoproteins. Mol Reprod. Dev 46: 344-350.

Totzauer I, Kölle S, Sinowatz F, Plendl J, Amselgruber W, Töpfer-Petersen E (1998). Localization of the zona glycoproteins ZPB (ZP3 alpha) and ZPC (ZP3 beta) in the bovine ovary during pre- and postnatal development. Ann Anat. 180:37-43.

Turner LM, Hoekstra HE (2006). Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol Biol Evol. 23: 1656-69.

Turner LM, Hoekstra HE (2008). Causes and consequences of the evolution of reproductive proteins. Int J Dev Biol. 52:769-780.

Van Duin M, Polman JE, De Breet IT, Van Ginneken K, Bunschoten H, Grootenhuis A, Brindle J, Aitken RJ (1994). Recombinant human zona pellucida protein ZP3 produced by Chinese hamster ovary cells induces the human sperm acrosome reaction and promotes sperm-egg fusion. Biol Reprod. 51:607-617.

Vanroose, G., H. Nauwynck, A. Van Soom, M.T. Ysebaert, G. Charlier, P. Van Oostveldt, A. de Kruif (2000). Structural aspects of the zona pellucida of in vitro produced bovine embryos: A scanning electron and confocal laser scanning microscopic study. Biol. Reprod. 62: 463-469.

von Witzendorff D, Maass K, Pich A, Ebeling S, Kölle S, Kochel C, (2009). Characterization of the acidic N-linked glycans of the zona pellucida of prepuberal pigs by a mass spectrometricapproach. Carbohydr Res.344:1541-1549.

Voyle RB, Haines BP, Loffler KA, Hope RM, Rathjen PD, Breed WG (1999). Isolation and characterisation of zona pellucida A (ZPA) cDNAs from two species of marsupial: regulated oocyte-specific expression of ZPA transcripts. Zygote. 7:239-248.

Wassarman PM(1988). Zona pellucida glycoproteins. Ann.Rev.Biochem 57: 415-442.

Wassarman PM, Liu C,Chen J, Qi H, Litscher ES (1998). Ovarian development in mice bearing homozygous or heterozygous null mutation in zona pellucida glycoprotein gene mZP3. Histol. Histopathol 13: 293-300.

Wassarman PM, Joviner L, Litsher ES,(2001). A profile of fertilization in mammals.Biol Nat Cell. 3: 59-64

Wassarman PM (2002). Sperm receptors and fertilization in mammals. Mt Sinai J Med;69:148-155.

Wassarman PM (2008). Zona pellucida glycoproteins. J Biol Chem. 283:24285-24289.

Wassarman PM, LitscherES (2008). Mammalian fertilization: the egg's multifunctional zona pellucida. Int J Dev Biol.;52(5-6):665-676.

Wassarman PM, Litscher ES (2009). The multifunctional zona pellucida and mammalian fertilization. J Reprod Immunol. 83:45-49.

Wassarman PM, Litscher ES (2013). Biogenesis of the mouse egg's extracellular coat, the zona pellucida. Curr Top Dev Biol. 102:243-266.

Wolgemuth DJ, Celenza J, Budman DS, Dumbar BS, 1984.Formation of the rabbit zona pellucida and its relationship to ovarian follicular development. Dev. Biol. 106: 71.90.

Yanagimachi R (1994). Mammalian fertilization. En physiology of reproduction, Knobil y Neil J (eds). 189-317.

Yanagimachi R (2011). Mammalian Sperm Acrosome Reaction: Where Does It Begin Before Fertilization?Biol Reprod 85:4-5.

Yauger B, Boggs NA, Dean J (2011). Human ZP4 is not sufficient for taxon-specific sperm recognition of the zona pellucida in transgenic mice. Reproduction. 141:313-319.

Yonezawa N, Fukui N, Kuno M, Shinoda M, Goko S, Mitsui S, Nakano M (2001). Molecular cloning of bovine zona pellucida glycoproteins ZPA and ZPB and analysis for sperm-binding component of the zona. European Journal of Biochemistry. 268:3587-3594.

Yonezawa N, Kanai-Kitayama S, Kitayama T, Hamano A, Nakano M (2012). Porcine zona pellucida glycoprotein ZP4 is responsible for the sperm-binding activity of the ZP3/ZP4 complex. Zygote. 20:389-397.

Yonezawa N, Mitsui S, Kudo K, Nakano M (1997). Identification of an N-glycosylated region of pig zona pellucida glycoprotein ZPB that is involved in sperm binding. Eur J Biochem. 248:86-92.

Yurewicz EC, Sacco AG, Gupta SK, Xu N, Gage DA (1998). Hetero-oligomerization-dependent binding of pig oocyte zona pellucida glycoproteins ZPB and ZPC to boar sperm membrane vesicles. J Biol Chem. 273:7488-7494.

Zhao M, Gold L, Dorward H, Liang LF, Hoodbhoy T, Boja E, Fales HM, Dean J (2003). Mutation of a conserved hydrophobic patch prevents incorporation of ZP3 into the zona pellucida surrounding mouse eggs. Mol Cell Biol. 23:8982-8991.

7. CHAPTERS

CHAPTER I

Rabbit zona pellucida composition: a molecular, proteomic and phylogenetic approach.

INTRODUCTION

During the *in vivo* fertilization process, sperm interact with the extracellular coats that surround the mammalian oocyte. These coats are the cumulus oophorus and the zona pellucida (ZP). The ZP has been related with species-specific gamete recognition, the sperm acrosome reaction, the control of polyspermy and protection of the oviductal embryo[1,2,3].

It has been shown that the ZP genes that encode ZP proteins can be classified into six subfamilies: ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, ZP1, ZPAX and ZPD[4]. However, not all these genes are present in all species. The ZP or equivalent extracellular coat in vertebrates is formed by several proteins ranging from 3 to 6[5,4]. Phylogenetic studies and the finding of different pseudogenes suggest that the evolution of ZP genes is mainly produced by duplications and death of genes[6,4].

Early studies in mouse, demonstrated that the ZP is formed of only three glycoproteins: ZP1, ZP2 and ZP3[7]. Later, the presence of three glycoproteins was demonstrated in other species like pig[8] and cow[9]. However, in these species the proteins are ZP2, ZP3 and ZP4.

Moreover, analysis of the complete genome in different species suggests the existence of additional genes coding for ZP proteins and shows that mammalian ZP could be formed of four proteins. Some studies have reported the existence of four proteins in the ZP of species like human[10,11], rat[12] and hamster[13,14] and phylogenetic analysis have detected four genes in other species like chimpanzee and macaque[15] .

Therefore, mammals could be classified into three categories according to their ZP composition. 1) species with a ZP formed by ZP1, ZP2 and ZP3 (to date, include only the mouse); 2) species showing three proteins, where ZP1 is not

present (e.g. cow, dog and pig). 3) species with four proteins (ZP1, ZP2, ZP3 and ZP4) as, for example, in human, rat and hamster.

The confusing results obtained in different studies on the ZP composition in some species is mainly due to the scarce amount of ZP available and, especially, to the heterogeneous glycosylation of the ZP proteins, resulting in broad, partially overlapping, bands in SDS-PAGE[16,14]. These facts make the purification of these proteins very difficult and, subsequently, accurate analysis is also difficult. Moreover, the general acceptation of mouse zona pellucida model (with 3 proteins) made difficult to take into consideration the analysis of the ZP composition in other species.

The development of mass spectrometric techniques has provided an important opportunity to identify the different proteins and glycoforms present in a complex mixture. Thus, proteomic analysis clarified ZP protein composition in human, rat, and hamster[11,12,13] and provided detailed information on the carbohydrate composition of the ZP proteins in some species[17,18,19].

In rabbit, characterization of the ZP by SDS-PAGE suggested the presence of three glycoproteins, ZP2[20,21], ZP3[22] and ZP4[23,24] , which migrate as one band with an apparent molecular mass of 85-95 kDa[23]. These proteins (ZP2[20], ZP3[22] and ZP4[25)] were detected by molecular biology approaches and are deposited in the GenBank database (GenBank accession numbers: L12167 (ZP2), NM_001195720.1 (ZP3), NM_001082295 (ZP4).

In addition, ZP1 cDNA sequence has been deposited after *in silico* analysis in GenBank with the accession number: XM_002709016.

The aim of this study was to demonstrate the presence of ZP1 mRNA in rabbit ovaries and ZP1 protein in the ovary and ZP from isolated oocytes.

MATERIAL AND METHODS

Collection of rabbit (Oryctolagus cuniculis) ovaries

12 week-old New Zealand California white rabbits were used to obtain ovarian RNA (n=3). The rabbits were injected with 25 IU of pregnant mare serum gonadotropin (PMSG) to stimulate folliculogenesis and sacrificed 48 hours later by overdose of CO_2. Ovaries were obtained and frozen in liquid nitrogen and kept at -80 ºC until use.

In addition, 20 female 8 week-old rabbits, obtained from a slaughterhouse (Conejos Susi, S.L., Alicante, Spain), were used for proteomic analysis.

Collection of rabbit (Oryctolagus cuniculis) oocytes and obtention of the zona pellucida.

Ovaries were obtained from 18 week-old animals (n=12) killed with an overdose of CO_2 and subjected to laparotomy. Cumulus-oocyte complexes (COCs) were obtained by aspiration with a 2 ml syringe and a 25 gauge needle from ovarian follicles, <1 mm in size, as previously described[26]. The COCs were placed in PBS 4-well dishes and the cumulus cells were removed by gentle pipetting using 2 mM hyaluronidase.

The oocyte ZPs were obtained after vigorous pipetting of each oocyte by using a narrow-bore micropipette, following four washes in PBS to eliminate the oocyte debris.

Purification of rabbit ovarian RNA, obtention of cDNA and amplification of the complete open reading frame of rabbit ZP1 gene

Total RNA was isolated using RNAqueous® kit (Ambion, Austin, Texas, USA) according to the manufacturer's instructions. The first-strand cDNA was synthesized from total RNA with the SuperScript First-Strand Synthesis System kit for RT-PCR (Invitrogen-Life Technologies, Carlsbad, CA, USA), according to the manufacturer's protocol.

Rabbit ZP1 was amplified using polymerase chain reaction (PCR) by means of specific primers (Table I) designed according to the predicted cDNA sequence obtained from genomic sequences in the ENSEMBL server (ensemble accession number: ENSOCUG00000015673).

PCR amplifications were performed using 3 µl of target cDNA, 0.5 µg of each primer, 200 µM of each dNTP and 1 IU of Advantage GC 2 Polymerase (Clontech Laboratories, CA, USA). PCR was carried out using an initial denaturation cycle of 2 min, and then 30 cycles of 1 min at 95ºC, 1 min at annealing temperature (depending on the primers) and then 1 min at 72 ºC. The final extension time was 10 min at 72 ºC. 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 visualising under UV light using ethidium bromide.

Amplicons were carefully excised from the agarose gels and purified with the QIAquick Gel Extraction Kit Protocol (Quiagen, Hilden, Germany) according to the manufacter's protocol. After that, the amplicons were automatically sequenced. The sequences were analyzed to determine the degree of homology with other known sequences using the BLAST program (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/blast/)[27]. Direct comparison between two sequences was made with the ALIGN program, and the multiple alignment of the ZP1 sequences of different species with the rabbit ZP1 sequence was carried out using Clustal W (http://www.ebi.ac.uk/clustalw/).

The amino acid sequences were analyzed with the following software packages: "SignalP"[28] to predict putative signal sequence and cleavage sites, and "NetOGlyc"[29] and "NetNglyc"[30] to predict potential O-linked and N-linked glycosylation sites, respectively.

In addition, amplicons corresponding to ZP2, ZP3 and ZP4 were amplified in the same conditions as ZP1, while the primers were designed based on cDNA sequences obtained from the GenBank database (ZP2: XM_002711834, ZP3: NM_001195720, ZP4: NM_001082295) (supplementary file 1).

Phylogenetic analysis of ZP1

Sequences of ZP1 for different mammals were retrieved from GenBank (when mRNA sequences were available) and from ENSEMBL gene predictions (supplementary file 2). All these predictions were checked manually to detect annotation errors especially close to splicing sites. Similarity searches were performed using BLAST and BLAT against assembled genomes (http://ensembl.org), and TRACE data (deposited in the trace archive of GenBank) followed by manual compilation of data to predict further genes or exons missing from the ENSEMBL predictions. We also checked that the new sequences corresponded to a syntenic region of the corresponding chromosome or contig. Sequences were aligned using Muscle in Seaview[31] and the alignment was refined visually. Only the exonic portions were kept for the phylogenetic analysis. Phylogenetic trees were reconstructed using maximum likelihood with PhymL[32] and the robustness of the nodes was estimated with bootstrap percentage (n=1000). The appropriate model of evolution was determined using corrected Akaike information criterion (AICc) and Modeltest software[33].

PROTEOMICS ANALYSIS

Solubilization of rabbit ZP, SDS-PAGE and silver staining

The rabbit ovaries (three different experiments: n=14; n=22 and n=37) were trimmed using small scissors and dissected to remove fat and connective tissue. Solubilized ZP was obtained according to the p*rotocol previously described* by our group[13,14].

In addition, solubilized ZP was also obtained from oocytes. For that, oocyte ZP (n=200) was solubilized at 65ºC for 30 min. in PBS buffer; the sample was then centrifuged and the supernatant was recovered.

Partially purified ZP (ovary) and isolated ZP (oocytes) were dissolved in sample buffer in reducing conditions (5%-mercaptoethanol). After boiling for 5 min, samples were separated by 12% SDS-PAGE. In brief, 4% stacking and 12% separating gels were used with 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. Next, the gel was incubated in a 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. The proteins were immobilised by incubating for 5 min in a 5% acetic acid solution.

HPLC-MS Analysis

HPLC-MS/MS analysis was used to identify the rabbit ZP proteins. The analysis was carried out on an HPLC-MS system consisting of an Agilent 1100 Series 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 interface (ESI). Details of the mass LC-MS conditions are described below.

Samples were separated using SDS-PAGE and the 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 allowed 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 supernatants were combined and dried using a vacuum evaporator.

In the case of soluble samples, these were diluted up to a final volume of 100 μL of 25 mM ammonium bicarbonate buffer pH 8.5 and then incubated with 50 mM tris (2-carboxyethyl) phosphine at 60°C for 10 min. After that, samples were alkylated by adding 100 mM iodoacetamide and left to stand for 1 h at room temperature in the dark. Finally, 0.3 µg of proteomics grade trypsin (Sigma-Aldrich) were added to each sample for a digestion time of 16 h at 37°C. After this incubation, the tryptic digestion was stopped with 0.5% TFA and the samples were dried using a vacuum evaporator.

The tryptic digestions were separated and analysed by HPLC-MS. Dried 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) in 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 2200m/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) with an isolation width of 2 Da and a relative collision energy of 35%. Data processing was performed with the Data Analysis program for LC-MSD Trap Version 3.2 (Bruker Daltonik, GmbH, Germany) and Spectrum Mill MS Proteomics Workbench (Agilent Technologies, Santa Clara, CA).

Data processing was performed with DataAnalysis program for LC/MSD Trap Version 3.3 (Bruker Daltonik, GmbH, Germany) and Spectrum Mill MS Proteomics Workbench (Rev A.03.02.060B, Agilent Technologies, Santa Clara,

CA, USA). Briefly, raw data were extracted under default conditions as follows: unmodified or carbamidomethylated cysteines; sequence tag length >1; [MH]+ 50–7000 m/z; maximum charge +7; minimum signal-to-noise (S/N) 25; finding ^{12}C signals. The MS/MS search against mammalian sequences in the NCBInr database was performed with the following criteria: identity search mode; tryptic digestion with 2 maximum missed cleavages; carbamidomethylated cysteines; peptide charge +1, +2, +3; monoisotopic masses; peptide precursor mass tolerance 2.5 Da; product ion mass tolerance 0.7 amu; ESI ion trap instrument; minimum matched peak intensity 50%; oxidized methionine, N-terminal glutamine conversion to pyroglutamic acid and STY phosphorylation as variable modifications. Two o more validated peptides were considered to demonstrate the existence of the protein. Peptides were considered valid with a score threshold of 5, and a percentage-scored peak intensity of 60%. All database matches above the threshold score of 3 were reported and used for discussion purposes.

RESULTS AND DISCUSSION

Analysis of rabbit ZP1 cDNA and amino acid sequences

The open reading frame (ORF) of rabbit ZP1 was completely amplified and characterized for the first time in this work.

The ORFs corresponding to rabbit ZP2, ZP3 and ZP4 have been characterized in previous studies[20,22,25]. Moreover, amplifications of a fragment corresponding with each gene (ZP2, ZP3 and ZP4) were made to confirm that results (Fig. 1) demonstrating the presence of the four transcripts in the rabbit ovary.

Full-length rabbit ZP1 cDNA was obtained from the total RNA prepared from rabbit ovaries and the sequence was submitted to GenBank with the accession number HQ702467. The amplified sequence of ZP1 contains a single ORF of 1884 nucleotides (Fig. 2) and 100% similarity with the predicted rabbit ZP1 (XM_002709016). The ATG iniation codon which was predicted with the Pedersen and Nielsen algorithm, was found to be associated with vertebrate initiator codons. This sequence contains a stop codon (TAA) in position 1882-1884.

The ORF of ZP1 encodes a polypeptide 627 amino acids long (Fig. 2) with a theoretical molecular weight of 68.7 kDa. A signal peptide of 20 amino acids with a cleavage between Gly20 and Gln21 was predicted with Bendtsen *et al.* Algorithm[34].

The ZP protein possesses the archetypal 'ZP domain', a signature domain comprised of 272 amino acid residues (^{279}Gln-Gly^{550}) rich in Cys residues (ten). Upstream of the ZP domain, a trefoil domain contains 45 residues (^{217}Glu-Thr^{260}). This domain is characteristic of ZP1 and ZP4 and is a region rich in Cys amino acids (six).

The sequence showed high hydrophobicity in the N-terminal (signal peptide) and C-terminal regions, the latter corresponding to the transmembrane domain (TMD) between Leu586 and Leu608, which is followed by a cytoplasmic tail. A basic amino acid domain (^{537}Arg-Arg-Arg-Ser^{540}) upstream of the TMD may serve as a consensus furin cleavage site[35,36,37].

Taking into account the presence of a signal peptide and that the putative cutting site is Arg538, as observed for mouse and rat ZP1[36,37], the molecular mass of the putative mature protein is estimated to be 57.3 kDa.

The amino acid sequence showed high similarity with the ZP1 of other mammals (Fig. 3). The protein sequence of rabbit ZP1 is 70% identical to human and horse ZP1, and 67% identical to mouse ZP1. The similarity with mammalian ZP1, the presence of the same domains and the conservation of the Cys (Fig. 3) strongly suggest that the amplified ORF correspond to rabbit ZP1.

A total of 12 potential O-glycosylation sites were predicted in the mature protein and two potential N-glycosylation sites (Asn-X-S/T) are present in mature rabbit ZP1 in position Asn71 and Asn362. These equivalent positions are conserved in horse, human, mouse and rat ZP1; however, an additional N-glycosylation site, Asn49, is present in the mouse and rat species[36,37] and is lacking in horse, human and rabbit ZP1 (Fig. 4 and supplementary file 3).

Mass spectrometry of rabbit ZP glycoproteins

Following amplification of ZP1 ORF, the next step was to confirm expression of the four proteins in rabbit ZP. For this purpose, the rabbit ZP extracted from ovaries or oocytes, as described in the Material and Methods section, was analyzed directly by MS/MS or after separation of the proteins by SDS-PAGE electrophoresis followed by silver staining of the gel. In this last situation, gel segments were reduced and alkylated, trypsinized and analyzed by

LC-ESI-MS-MS. A summary of the peptides identified is included in Table II and Figure 4.

Several peptides corresponding to the immature version of the proteins were detected, indicating that ZP expression is continuous and elevated or, alternatively, that signal peptides and the carboxyl terminal region are not efficiently removed. These regions of the proteins could probably be identified because the protocol used for ZP isolation differ from those of previous studies performed in other species.

Taking into account that HPLC-MS analysis can be considered as a semiquantitative technique, the fact that the coverage of ZP2 and ZP3 was similar (≈ 50%) might indicate the similar abundance of these two proteins. However, ZP4 seems less abundant (≈43%) and ZP1 may be the least abundant protein in ZP matrix (≈33%).. Taking into account these data, it seems that the levels of ZP2 and ZP3 in the ZP matrix may be higher than the levels of ZP1 and ZP4, even though one should realise that is indicative only given the sequence differences and ionisation efficiency differences of the different peptides. Nevertheless, a similar situation is observed in mouse, in which the levels of ZP1 mRNA are four times lower than those of ZP2 and ZP3[38] and so only 56% of the ZP1 polypeptide chain can be identified by direct MS/MS, compared with the 96 and 100% of ZP2 and ZP3 respectively[36]. Other species show a similar pattern to human or rat, in both of which proteomic analysis has revealed a coverage percentage that is similar between ZP2 and ZP3, being ZP1 the least abundant protein[11,12]. Future quantitative proteomic analysis will be performed to ascertain the ZP glycoprotein stoichiometry of the mature ZP.

ZP1

A total of 21 different peptides were identified in the different analyses yielding a sequence coverage of 33.9%. None of the identified peptides contained an N-glycosylation site, suggesting that both the described consensus sequences may be occupied in the mature glycoprotein. An N-glycosylation site present in the N-terminal region of the mouse and rat ZP1 protein was not conserved in the rabbit and human ZP1[36,37] (supplementary file 3). On the other hand, 35 out of the predicted O-glycosylation sites were contained in the identified peptides, from which it can be deduced that these residues are either not glycosylated at all or are, at most, partially glycosylated. A similar result was previously reported for ZP1 in mouse and rat where proteomic analysis did not detect O-glycosylation sites[36,37].

ZP2

Forty peptides corresponding to rabbit ZP2 were detected. These peptides correspond to 55.7% of the protein. 65 potential sites of O-glycosylation and five sites of N-glycosylation (Asn99, Asn134, Asn278 and Asn302) can be localized in the detected peptides (Fig. 4). These residues might not be glycosylated at all or be partially glycosylated. In contrast, the conserved Asn corresponding to Asn99 and Asn278 were detected as glycosylated peptides in mouse and rat[36,37]. Other differences in the degree of N-glycosylation in this protein between cow, human, pig and rabbit can also be detected (supplementary file 3).

ZP3

A total of eighteen peptides from rabbit ZP3 could be convincingly identified in the different experiments. This corresponds to 50.3% of the sequence (Fig. 4). Two sites of N-glycosylation (Asn139 and Asn264) were detected in the analysis (Fig. 4). These two N-glycosylation sites may not be glycosylated in the native protein or, at most, be partially glycosylated. In contrast, these conserved

sites are glycosylated in human, mouse and rat[39,36,37] (supplementary file 3). Different studies have suggested that carbohydrates play a key role during the sperm-ZP interaction in different species. N-glycans of human ZP3 have important roles in the induction of the AR[40]. However, more attention has been paid to O-glycans. In particular, O-linked oligosaccharides in mice ZP3 have been analyzed in depth, althought there is a controversy about the exact role played by O-glycans in this process[41,42,43,44]. Thirty four potential O-glycosylation sites were detected in this analysis, suggesting that they are not occupied or only partially occupied. In recombinant human, mouse and rat ZP3, two clusters of O-glycans have been reported, some of them are similar among species; however, they are not identical and this probably contributes to the species specificity of the gamete interaction. The first cluster corresponds to the amino acid residues 156-173[39]. The Thr155 and Thr156 in mouse and human, respectively, are glycosylated and have been suggested to play a role in sperm binding[42]; however, the corresponding amino acid in rabbit ZP3, Thr154, is at least partially unoccupied, indicating that differences exist between these species. A similar result was observed with the second cluster observed in the human ZP3, which corresponds to Thr260, Thr264 and Thr281. Identification in this study of the peptide (Aa 251-296) in rabbit ZP3 indicates that there is no a clustering because the first two Thr are not conserved. The Thr279 is conserved but is not totally glycosylated. The role of the different oligosaccharide chains in fertilization in rabbit remains unresolved thus far.

ZP4

Twenty three peptides were detected in ZP4 (43.5% coverage). Forty one O-glycosylation and two N-glycosylation (Asn478, Asn482) sites were observed, which might mean that these sites are not glycosylated in mature protein. Asn75 and Asn206 are not detected in the peptides probably because they are occupied. The corresponding amino acid in the rat[12] and pig are also glycosylated (supplementary file 3).

Different O-glycosylation sites were detected in the pig and rat ZP4[12]. In rat ZP4, the precise amino acid involved was not determined; however, in the pig ZP4 the amino acid residues (Ser293 and Thr303) are glycosylated and are conserved in the rabbit ZP4. A peptide including this region is not detected suggesting that probably it is glycosylated as observed in pig ZP4. Future glycomics studies are necessary to obtain more precise information about the ZP4 glycosylation.

Thus, although the mouse ZP composition might originally have supported the hypothesis that mammalian ZP has three proteins, the mouse model has been revealed to be an exception within mammals[4]. This study demonstrates that the rabbit ZP is formed by four proteins, as is the human ZP, making this species a good animal model for understanding the role played by the ZP for several reasons. First, the similarity between rabbit and human ZP proteins is generally greater than in other species (except primates and horse, supplementary file 4). Second, ZP biogenesis in rabbit is similar to primates including, human[45-50]. In human and rabbit ovarian follicles, oocytes and granulosa cells contribute to the formation of the ZP[45,46]; however, in hamster, mice and rat the ZP is only formed by the oocytes[51-57]. Third, these animals are used for human consumption in numerous countries, and so a large amounts of rabbit ovaries are available in the slaughterhouses and could contribute to reducing the number of animal sacrifices necessary for research purposes.

ZP1 phylogeny

It was previously reported using *in silico* approaches that ZP genes have been gradually lost during the evolution of vertebrates[4]. Thus, in several mammals the ZP1, ZPB/ZP4, ZPD, and/or ZPAX gene is lacking. These data suggested the presence of pseudogenes in the genome of these species. In particular, the same authors did not find ZP1 in rabbit and other species[4]. Our results demonstrate, however, that rabbit ZP is composed of four glycoproteins,

including ZP1. The difference observed between both studies is probably due to the incompleteness of the rabbit genome sequence available at that time.

Additionally, four ZPs sequences are annotated in the genome of pika (*Ochotona princeps*), which belongs to the same order as the rabbit (the Lagomorpha), suggesting that this species also have four proteins. Future proteomic analyses are necessary to confirm these *in silico* findings. Our similarity search also found sequences corresponding to ZP1 in the genome of *Canis* (chromosome 18) and *Bos* (chromosome 29), as already mentioned by Goudet et al (2008); and new ones in the genome of *Sus* (chromosome.2), *Callithrix* (chromosome11),and *Spermophilus* (scaffold_129656) (Supplementary file 2). Our analysis of the sequences indicates that at least 4 sequences are probably pseudogenes due to the presence of deletions and/or stop codons in addition to *Canis* and *Bos* (Goudet et al., 2008): *Callithrix*, *Tarsius* and:twoothercetartiodactyls *Sus* and *Tursiops.* (Fig.5)

The phylogenetic tree reconstructed with phyml with the GTR + I +G model of sequence evolution is presented in Figure 6. The topology is congruent with the classic phylogeny of mammals. Pseudogenization of the ZP1 occurred at least four times during the evolution of mammals (indicate by red branches in Fig. 6).

ZP1 and ZP4, two paralogous genes from the ZPB subfamily, were formed by gene duplication. Previous phylogenetic studies indicated that they share a common ancestral gene[4,5,58].

After a gene duplication is possible three situations: a) the ancestral function is partitioned and shared by the two members of the duplicated pair (subfunctionalization), b) one duplicate acquires a new function while the other retains the original function (neofunctionalization) c) one gene degenerate to a pseudogen by accumulation of mutations and the other mantain the original gene function. The last situation corresponds with the species with three glycoproteins in which ZP1 or ZP4 is lost.

Therefore, the common origin of ZP1 and ZP4 is suggested by the observation that both are involved in identical molecular ***mechanisms***. Studies have shown that ZP3-induced acrosome reaction (AR) involves activation of the G(i)-coupled receptor pathway, whereas ZP1- and ZP4-mediated ARs are independent of this pathway. The ZP3-induced AR involves the activation of T-type voltage-operated calcium channels (VOCCs), whereas ZP1-and ZP4-induced ARs involve both T-and L-type VOCCs[59].

Thus, in species with four glycoproteins, the fact that ZP1 and ZP4 participate in AR through similar pathways may indicate a possible subfunctionalization

The function played by each ZP protein is not totally clear and differs in the species. A structural function was initially attributed to the ZP1 based on the mouse model[60,61]. In human, ZP1 has been related with the induction of the AR[62,63]. However, ZP4 seems to be implicated more directly in the fertilization process in numerous mammal species[64-67]. Thus, in human, it was reported that ZP4 is involved in sperm binding and the induction of the AR[68]. In other species like pig and cow, it was reported that both ZP3 and ZP4 act as receptors of the spermatozoa[69,70].

Evidence suggests that when a gene suffers duplication, the functional divergence of gene copies is a major factor promoting their retention in the genome. So, species with four ZP glycoproteins, like rabbit, the two copies might play a different role and the two genes are necessary for the ZP to play its role correctly.

CONCLUDING REMARKS

In summary, in this study, the cDNA encoding ZP1 has been identified in rabbit (*Oryctolagus cuniculus*) ovaries. The nucleotide sequence shows a high similarity with the ZP1 of other mammals. Mass spectrometric analysis confirmed the presence of ZP1, ZP2, ZP3 and ZP4 proteins in rabbit ovaries and oocytes. Phylogenetic analysis indicates that the pseudogenization of the ZP1 has occurred at least four times during the evolution of mammals. Finally, due to the similar composition and expression pattern, rabbit ZP could be proposed as a suitable experimental model for studying the human ZP and its role during fertilization.

REFERENCES

(1) Yanagimachi, R. Mammalian fertilization. Physiology of reproduction; Knobil, E., Neill, J. D., Eds.; Raven Press: New York, 1994; pp 189-317.

(2) Dean J. The enigma of sperm-egg recognition in mice. Soc Reprod Fertil Suppl 2007;63:359-65.

(3) Wassarman PM, Litscher ES. The multifunctional zona pellucida and mammalian fertilization. J Reprod Immunol 2009;83:45-9.

(4) 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.

(5) Bausek N, Waclawek M, Wolfgang JS, Wohlrab F. The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. J Biol Chem 2000; 275:28866–72.

(6) Spargo SC, Hope RM. Evolution and nomenclature of zona pellucida gene family. Biol Reprod 2003;68:358-62.

(7) Bleil JD, Wassarman PM. 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.

(8) Hedrick JL, Wardrip NJ. On the macromolecular composition of the zona pellucida from porcine oocytes. Dev Biol 1987;121:478–88.

(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) Hughes DC, Barratt CL. Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida. Biochim. Biophys Acta 1999;1447:303–6.

(11) Lefièvre L, Conner SJ, Salpekar A, Olufowobi O, Ashton P, Pavlovic B, Lenton W, Afnan M, Brewis IA, Monk M, Hughes DC, Barratt CLR. Four zona pellucida glycoproteins are expressed in the human. Hum Reprod 2004;19:580–6.

(12) Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. J Biol Chem. 2005;280:12721–31.

(13) Izquierdo-Rico MJ, Jimenez-Movilla M, Llop E, Pérez-Oliva AB, Ballesta J, Gutierrez-Gallego R, Jimenez-Cervantes C, Aviles M. Hamster zona pellucida is formed by four glycoproteins. J Proteome Res 2009;8:926-41.

(14) Jiménez-Movilla M, Martínez-Alonso E, Castells MT, Izquierdo-Rico MJ, Saavedra MD, Gutiérrez-Gallego R, Fayrer-Hosken R, Ballesta J, Avilés M. Cytochemical and biochemical evidences for a complex tridimensional structure of the hamster zona pellucida. Histol Histopathol 2009;24:599-609.

(15) Goudet G, Mugnier S, Callebaut I, Monget P. Phylogenetic analysis and identification of pseudogenes reveal a progressive loss ofzona pellucida genes during evolution of vertebrates. Biol Reprod 2008;78:796-806.

(16) Dunbar BS, Timmons TM, Skinner SM, Prasad SV. Molecular analysis of a carbohydrate antigen involved in the structure and function of zona pellucida glycoproteins. Biol Reprod 2001;65:951-60.

(17) Easton RL, Patankar MS, Lattanzio FA, Leaven TH, Morris HR, Clark GF, 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;275:7731-42.

(18) von Witzendorff D, Ekhlasi-Hundrieser M, Dostalova Z, Resch M, Rath D, Michelmann HW, Töpfer-Petersen E. Analysis of N-linked glycans of porcine zona pellucida glycoprotein ZPA by MALDI-TOF MS: a contribution to understanding zona pellucida structure. Glycobiology 2005;15:475-88.

(19) von Witzendorff D, Maass K, Pich A, Ebeling S, Kölle S, Kochel C, Ekhlasi-Hundrieser M, Geyer H, Geyer R, Töpfer-Petersen E. Characterization of the acidic N-linked glycans of the zona pellucida of prepuberal pigs by a mass spectrometric approach. Carbohydr Res 2009;344:1541-9.

(20) Lee VH, Schwoebel E, Prasad S, Cheung P, Timmons TM, Cook R, Dunbar BS. Identification and structural characterization of the 75-kDa rabbit zona pellucida protein. J Biol Chem 1993;268:12412-7.

(21) Schwoebel ED, Vandevoort CA, Lee VH, Lo YK, Dunbar BS. Molecular analysis of the antigenicity and immunogenicity of recombinant zona pellucida antigens in a primate model. Biol Reprod 1992;47:857-65.

(22) Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG. Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families. DNA Seq 1994;4:361-93.

(23) Prasad SV, Mujtaba S, Lee VH, Dunbar BS. Immunogenicity enhancement of recombinant rabbit 55-kilodalton zona pellucida protein expressed using the baculovirus expression system. Biol Reprod 1995;52:1167-78.

(24) Lee VH. Expression of rabbit zona pellucida-1 messenger ribonucleic acid during early follicular development. Biol Reprod 2000;63:401-8.

(25) Schwoebel E, Prasad S, Timmons TM, Cook R, Kimura H, Niu EM, Cheung P, Skinner S, Avery SE, Wilkins B, Dunbar BS. Isolation and characterization of a full-length cDNA encoding the 55-kDa rabbit zona pellucida protein. J Biol Chem 1991; 266:7214-9.

(26) Arias-Álvarez M, García-García RM, Torres-Rovira L, González-Bulnes A, Rebollar PG, Lorenzo PL. Influence of leptin on in vitro maturation and steroidogenic secretion of cumulus–oocyte complexes through JAK2/STAT3 and MEK 1/2 pathways in the rabbit model. Reproduction 2010;139:523–32.

(27) Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

(28) 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.

(29) Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accesibility. Glycoconj J 1998;15:115–30.

(30) Gupta R, Brunak S. Prediction of glycosilation across the human proteome and the correlation to protein function. Pac. Symp. Biocomput. 2002, 310-322.

(31) Gouy M, Guindon S, Gascuel O. SeaView Version 4: A Multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 2010;27:221-4.

(32) Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 2003;52:696-704.

(33) Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998;14:817-8.

(34) Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783-95.

(35) Ducker P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel 2004;17:107–12.

(36) Boja ES, Hoodbhoy T, Fales HM, Dean J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 2003;278:34189-202.

(37) Boja ES, Hoodbhoy T, Garfield M, Fales HM. Structural conservation of mouse and rat zona pellucida glycoproteins. Probing the native rat zona pellucida proteome by mass spectrometry. Biochemistry 2005;44:16445–60.

(38) Epifano O, Liang LF, Familiari M, Moos MC, Dean J. Coordinate expression of the three zona pellucida genes during mouse oogenesis. Development 1995;121:1947-56.

(39) Zhao M, Boja ES, Hoodbhoy T, Nawrocki J, Kaufman JB, Kresge N, Ghirlando R, Shiloach J, Pannell L, Levine RL, Fales HM, Dean J. Mass spectrometry analysis of recombinant human ZP3 expressed in glycosylation-deficient CHO cells. Biochemistry 2004;43:12090-104.

(40) Chakravarty S, Kadunganattil S, Bansal P, Sharma RK, Gupta SK. Relevance of glycosylation of human zona pellucida glycoproteins for their binding to capacitated human spermatozoa and subsequent induction of acrosomal exocytosis. Mol Reprod Dev 2008;75:75-88.

(41) Dean J. The enigma of sperm-egg recognition in mice. Soc Reprod Fertil Suppl. 2007;63:359-65.

(42) Visconti PE, Florman HM. Mechanisms of sperm-egg interactions: between sugars and broken bonds. Sci Signal 2010;3:35.

(43) Clark GF. Molecular models for mouse sperm-oocyte binding. Glycobiology 2011;21:3-5.

(44) Gahlay G, Gauthier L, Baibakov B, Epifano O, Dean J. Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. Science 2010;329:216-9.

(45) Lee VH, Dunbar BS. Developmental expression of the rabbit 55-kDa zona pellucida protein and messenger RNA in ovarian follicles. Dev Biol 1993;155:371-82.

(46) Dunbar BS, Avery S, Lee V, Prasad S, Schwahn D, Schwoebel E, Skinner S,Wilkins B. The mammalian zona pellucida: its biochemistry, immunochemistry, molecular biology, and developmental expression. Reprod Fertil Dev 1994;6:331-47.

(47) Hinsch KD, Hinsch E, Meinecke B, Töpfer-Petersen E, Pfisterer S, Schill WB. Identification of mouse ZP3 protein in mammalian oocytes with antisera against synthetic ZP3 peptides. Biol Reprod 1994;51:193-204.

(48) Grootenhuis AJ, Philipsen HL, de Breet-Grijsbach JT, van Duin M. Immunocytochemical localization of ZP3 in primordial follicles of rabbit, marmoset, rhesus monkey and human ovaries using antibodies against human ZP3. J. Reprod. Fertil Suppl 1996;50:43-54.

(49) Martínez ML, Fontenot GK, Harris JD. The expression and localization of zona pellucida glycoproteins and mRNA in cynomolgus monkeys (Macaca fascicularis). J. Reprod Fertil Suppl 1996;50:35-41.

(50) Gook DA, Edgar DH, Borg J, Martic M. Detection of zona pellucida proteins during human folliculogenesis. Hum Reprod 2008;23:394-402.

(51) Bleil JD, Wassarman PM. Synthesis of zona pellucida proteins by denuded and follicle-enclosed mouse oocytes during culture in vitro. Proc Natl Acad Sci USA 1980; 77:1029-33.

(52) Fléchon JE, Pavlok A, Kopecný V. Dynamics of zona pellucida formation by the mouse oocyte. An autoradiographic study. Biol Cell 1984;51:403-6.

(53) Epifano O, Liang LF, Familiari M, Moos MC, Dean J. Coordinate expression of the three zona pellucida genes during mouse oogenesis. Development 1995;121:1947-56.

(54) Sinowatz F, Kölle S, Töpfer-Petersen E. Biosynthesis and expression of zona pellucida glycoproteins in mammals. Cells Tissues Organs 2001;168:24-35.

(55) El Mestrah M, Castle PE, Borrosa G, Kan FW. Subcellular distribution of ZP1, ZP2, and ZP3 glycoproteins during folliculogenesis and demonstration of their topographical disposition within the zona matrix of mouse ovarian oocytes. Biol Reprod 2002;66:866-76.

(56) Scobie GA, Kerr LE, MacDuff P, Aitken RJ. Cloning, sequencing and site of origin of the rat sperm receptor protein, ZP3. Zygote 1999;7:27-35.

(57) Izquierdo-Rico MJ, Gimeno L, Jiménez-Cervantes C, Ballesta J, Avilés M. Biosynthesis of hamster zona pellucida is restricted to the oocyte. Theriogenology 2011;75:463-72.

(58) Conner SJ, Hughes DC. Analysis of fish ZP1/ZPB homologous genes--evidence for both genome duplication and species-specific amplification models of evolution. Reproduction 2003;126:347-52.

(59) Gupta SK, Bhandari B. Acrosome reaction: relevance of zona pellucida glycoproteins. Asian J Androl 2011;13:97-105.

(60) Wassarman PM. Zona pellucida glycoproteins. Ann Rev Biochem 1988;57:415-42.

(61) Rankin T, Talbot P, Lee E, Dean J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Develop 1999;126:3847-55.

(62) Ganguly A; Bukovsky A, Sharma RK, Bansal P, Bhandari B, Gupta SK. In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. Hum Reprod 2010;25:1643-56.

(63) Ganguly A, Bansal P, Gupta T, Gupta SK. 'ZP domain' of human zona pellucida glycoprotein-1 binds to human spermatozoa and induces acrosomal exocytosis. Reprod Biol Endocrinol 2010;11;8:110.

(64) Topper EK, Kruijt L, Calvete J, Mann K, Töpfer-Petersen E, Woelders H. Identification of bovine zona pellucida glycoproteins. Mol Reprod Dev 1997;46:344-50.

(65) Govind CK, Hasegawa A, Koyama K, Gupta SK. Delineation of a conserved B cell epitope on bonnet monkey (Macaca radiata) and human zona pellucida glycoprotein-B by monoclonal antibodies demonstrating inhibition of sperm-egg binding. Biol Reprod 2000;62: 67-75.

(66) Chakravarty S, Suraj K, Gupta SK. Baculovirus-expressed recombinant human zona pellucida glycoprotein-B induces acrosomal exocytosis in capacitated spermatozoa in addition to zona pellucida glycoprotein-C. Mol Hum Reprod 2005;11:365-72.

(67) Prasad SV, Wilkins B, Skinner SM, Dunbar BS. Evaluating zona pellucida structure and function using antibodies to rabbit 55 kDa ZP protein expressed in baculovirus expression system. Mol Reprod Dev 1996;43:519-29.

(68) Chiu PC, Wong BS, Chung MK, Lam KK, Pang RT, Lee KF, Sumitro SB, Gupta SK, Yeung WS. Effects of native human zona pellucida glycoproteins 3 and 4 on acrosome reaction and zona pellucida binding of human spermatozoa. Biol Reprod 2008;79:869-77.

(69) Yurewicz EC, Sacco AG, Gupta SK, Xu N, Gage DA. Hetero-oligomerization-dependent binding of pig oocyte zona pellucida glycoproteins ZPB and ZPC to boar sperm membrane vesicles. J Biol Chem. 1998;273:7488-94.

(70) Kanai S, Yonezawa N, Ishii Y, Tanakura M, Nakano M. Recombinant bovine zona pellucida glycoproteins ZP3 and ZP4 coexpressed in Sf9 cells form a sperm-binding active hetero-complex. FEBS J 2007;274:5390-405.

FIGURES

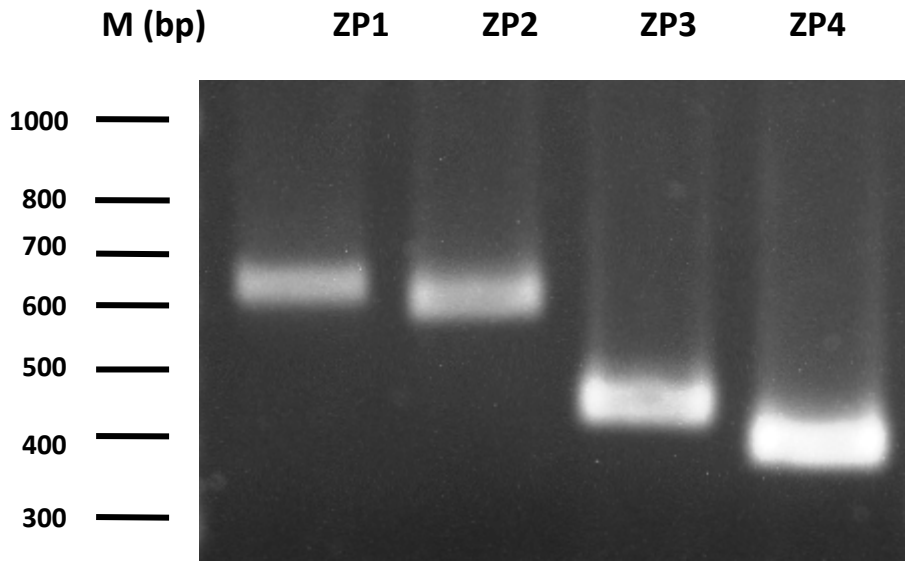


Figure 1: Analysis of ZP1, ZP2, ZP3 and ZP4 gene expression in rabbit ovary by RT-PCR.

```
   1 atgactgggggtcgcctggtggccctactactgctggtggcggcctccctggggctgggt
   1  M  T  G  G  R  L  V  A  L  L  L  L  V  A  A  S  L  G  L  G
  61 cagcagccacaccctgagcccggcctcccaggcctccagtacagctatgactgtgggatg
  21  Q  Q  P  H  P  E  P  G  L  P  G  L  Q  Y  S  Y  D  C  G  M
 121 cggggcatgcaactggtggtgctccccaggccgggccggactatccgtttcaaggtggtg
  41  R  G  M  Q  L  V  V  L  P  R  P  G  R  T  I  R  F  K  V  V
 181 gatgaattcgggaaccggtttgaggtgaacaactgttccatctgcttccactgggtcagc
  61  D  E  F  G  N  R  F  E  V  N  N  C  S  I  C  F  H  W  V  S
 241 gccgagccccaggcgcccgccgtcttctctgctgattacaaaggctgccacgtgctggag
  81  A  E  P  Q  A  P  A  V  F  S  A  D  Y  K  G  C  H  V  L  E
 301 aaggaggggcattcccacctgacggtgttcatagaagcaatgctgcctgatggtcacgtg
 101  K  E  G  H  S  H  L  T  V  F  I  E  A  M  L  P  D  G  H  V
 361 gaggtcgcacaggaggcggttctgatctgtcccaaacctggccacacctgggccgtgggt
 121  E  V  A  Q  E  A  V  L  I  C  P  K  P  G  H  T  W  A  V  G
 421 tcccaccaggtgcccccccaccacgccctcgcctaccacccccccatgctctccccttccac
 141  S  H  Q  V  P  P  T  T  P  S  P  T  T  P  H  A  L  P  F  H
 481 ctctcctcagcccacaccttccccatccctctgtacctggagcacagcctcatgctccca
 161  L  S  S  A  H  T  F  P  I  P  L  Y  L  E  H  S  L  M  L  P
 541 acccctgctgggccctccctgggacctggccccaccccccgccgtgctggctcaggtggaa
 181  T  P  A  G  P  S  L  G  P  G  P  T  P  A  V  L  A  Q  V  E
 601 cgctgggaagtggacaagccggatgccgtaggtagccatctgccccaggagtggtgccag
 201  R  W  E  V  D  K  P  D  A  V  G  S  H  L  P  Q  E  W  C  Q
 661 gtggcctctgggcacatcccctgcatagtgcaaagcagctccaaggaggcctgtgagcag
 221  V  A  S  G  H  I  P  C  I  V  Q  S  S  S  K  E  A  C  E  Q
 721 gccggctgttgctatgacagtgccagggaggtgccctgctactatggcaacacagccact
 241  A  G  C  C  Y  D  S  A  R  E  V  P  C  Y  Y  G  N  T  A  T
 781 gtccagtgcttccgaaacggctacttcatcttggttgttgcccaagaaatggccttggca
 261  V  Q  C  F  R  N  G  Y  F  I  L  V  V  A  Q  E  M  A  L  A
 841 cacagaatcacgctggccaacgtccacctggcctatgcccccacgcgctgccccccggcc
 281  H  R  I  T  L  A  N  V  H  L  A  Y  A  P  T  R  C  P  P  A
 901 cagaagaccagtgcttttgtcatcttccacgtccccctcacccactgcggcaccacagtt
 301  Q  K  T  S  A  F  V  I  F  H  V  P  L  T  H  C  G  T  T  V
 961 caggtgctggggagccagctcttctacgagaaccagctggtgtctgacatcgatgtccgg
 321  Q  V  L  G  S  Q  L  F  Y  E  N  Q  L  V  S  D  I  D  V  R
1021 gaggggccgcagggttccatcacacgggacagctccttccggcttctcgtccgctgtatc
 341  E  G  P  Q  G  S  I  T  R  D  S  S  F  R  L  L  V  R  C  I
1081 ttcaatgccagcgacttcctgcccatccaggcgtccatcttctcacctccactgcctgcc
 361  F  N  A  S  D  F  L  P  I  Q  A  S  I  F  S  P  P  L  P  A
1141 cctgtaactcaggctggccccctgcgcctggagctacggattgccagggatgagacttttt
 381  P  V  T  Q  A  G  P  L  R  L  E  L  R  I  A  R  D  E  T  F
1201 agctccttctatgaggaggaggactacccccctcgtgaggctgctccgagaaccggtacac
 401  S  S  F  Y  E  E  E  D  Y  P  L  V  R  L  L  R  E  P  V  H
1261 gtggaggtccggctgctgcagaggacagaccccagtctggtgctggagctgcaccagtgc
 421  V  E  V  R  L  L  Q  R  T  D  P  S  L  V  L  E  L  H  Q  C
1321 tgggccactcccagtgccaaccccgtccagcagccccagtggcccctcctgtcagacggg
 441  W  A  T  P  S  A  N  P  V  Q  Q  P  Q  W  P  L  L  S  D  G
1381 tgtcctttcaagggcgacagctacagaacccgagtgctagccttggaccgggcagagctg
 461  C  P  F  K  G  D  S  Y  R  T  R  V  L  A  L  D  R  A  E  L
1441 cccttccggtctcattaccagcgtttcacggttgccaccttcaccttcctggactcgggc
 481  P  F  R  S  H  Y  Q  R  F  T  V  A  T  F  T  F  L  D  S  G
1501 gctcagcgagccctcaggggactggtttacttcttctgcagcgcctcagcctgccaccct
 501  A  Q  R  A  L  R  G  L  V  Y  F  F  C  S  A  S  A  C  H  P
1561 tcagggccagagacttgctcatctacgtgtagctccaggactgccaaacgccgacgatcc
 521  S  G  P  E  T  C  S  S  T  C  S  S  R  T  A  K  R  R  R  S
1621 tcaggttaccatgacggcacccccagggccctggacatcgtgagttctccagggccagtg
 541  S  G  Y  H  D  G  T  P  R  A  L  D  I  V  S  S  P  G  P  V
1681 ggcttccaggattctcacaggcaggagcccacactggagtccacaggctccggcaggaac
 561  G  F  Q  D  S  H  R  Q  E  P  T  L  E  S  T  G  S  G  R  N
1741 tccaacccgaagcctctgctctgggtggtccttctgctgctggccattgctcttgtcctg
 581  S  N  P  K  P  L  L  W  V  V  L  L  L  L  A  I  A  L  V  L
1801 gggattggtgtctttgtgggcctgagccaggcctgggcccacaagctccgggaaggccac
 601  G  I  G  V  F  V  G  L  S  Q  A  W  A  H  K  L  R  E  G  H
1861 aggctcaccgaccaggctcaataa
 621  R  L  T  D  Q  A  Q  *
```

Figure 2: Nucleotide and deduced amino acid sequence of rabbit ZP1. The initial and final codons are in pink. The signal peptidase cleavage site is between Gly20 and Gln21 and is marked in green. The underlined amino acids indicate the C-terminal cleavage site. The zona domain is shown in red. The trefoil domain is shown in blue. The consensus furin cleavage-site is underlined. The transmembrane domain is in orange.

```
HUMAN           ------------------------------------------------------------
CHIMPANZEE      ------------------------------------------------------------
HORSE           MGSSAFCPRLRGEISAACWSLSSPVALILLQALGAGAVSLQAPKAPPEPEPSPSGGRWGL 60
RABBIT          ------------------------------------------------------------
MOUSE           ------------------------------------------------------------
MACAQUE         ------------------------------------------------------------


HUMAN           --------MAGGSATTWGYPVALLLLVAT-LGLGRWLQPDPGLPGLRHSYDCGIKGMQLL 51
CHIMPANZEE      --------MAGGSATTWGYRVALLLLVAT-LGLGRRLQPNPGLPGLRHSYDCGIKGMQLL 51
HORSE           LWGLWVYLVAGTSAMVWGLSVA-LLLVAT-LGLGQQPRLEPGLLGLRHTYDCGMKGLQLL 118
RABBIT          -------------MTGGRLVALLLLVAASLGLGQQPHPEPGLPGLQYSYDCGMRGMQLV 46
MOUSE           -------------MAWGCFVVLLLLAAAPLRLGQRLHLEP---GFEYSYDCGVRGMQLL 43
MACAQUE         ---------------MRLLQCVLLCVSLSLVLSGQHKPET--PGYSSVLHCGLWSFQFA 42
                              ** .:  * *.   : :.   *     .**: .:*:


HUMAN           VFPRPGQTLR--FKVVDEFGNRFDVNNCSICYHWVTSRPQEPAVFSADYRGCHVLEKDGR 109
CHIMPANZEE      VFPRPGQTLR--FKVVDEFGNRFDVNNCSICYHWVTSRPQEPAVFSADYRGCHVLEKDGR 109
HORSE           VFPPPGWTVR--FKVVDEFGNTFEVNNCSICYHWVTAKPKGPVVFSADYKGCHVLEKDGR 176
RABBIT          VLPRPGRTIR--FKVVDEFGNRFEVNNCSICFHWVSAEPQAPAVFSADYKGCHVLEKEGH 104
MOUSE           VFPRPNQTVQ--FKVLDEFGNRFEVNNCSICYHWVTSEAQEHTVFSADYKGCHVLEKDGR 101
MACAQUE         VNLSQEATSPPVLITWDNQGLLHKLQNDSDCGTWIRKRPGSSVVLEATYSSCYVTEWDSH 102
                *      *    : . *: *  ..:: * *  *:  ..   .*:.* * .*:* * :.:


HUMAN           FHLRVFMEAVLPNGRVDVAQDATLICPKPDPSRTLDSQLAPPAMFSVSTPQTLSFLPTSG 169
CHIMPANZEE      FHLRVFMEAVLPNGRVDVAQDATLICPKPDPSRTPDSQLAPPAMFSVSTPQTLSFLPTSG 169
HORSE           SQLRVLIEAVLPEGRVDVARDVTLICPKPDHTWPLDSYLVPPSAP--STPHTRPLRPT-- 232
RABBIT          SHLTVFIEAMLPDGHVEVAQEAVLICPKPGHTWAVGSHQVPPTTPSPTTPHALPFHLSSA 164
MOUSE           FHLRVFIQAVLPNGRVDIAQDVTLICPKPDHTVTPDPYLAPPTTPEPFTPHAFALHPIPD 161
MACAQUE         YIMPVGVEGVGVAEHKMVPERKLLKCP------------------------------- 129
                  : * ::.:    :  :..   * **


HUMAN           HTSQGSGHAFPSPLDPGHSSVHPTPALPSPGPGPTLATLAQPHWGTLEHWDVNKRDYIGT 229
CHIMPANZEE      HTSQGSGHAFPSPLDPGHSSVHPTPALPSPGPGPTLATLAQPHWGTLEHWDVNKRDYIGT 229
HORSE           --------------PEHSFVHPTPALPSLGPGPTHPTLAQPQWGTLGPWEVDKPASIGT 277
RABBIT          HT-------FPIPLYLEHSLMLPTPAGPSLGPGPTPAVLAQ-----VERWEVDKPDAVGS 212
MOUSE           HTLAGSGHTGLTTLYPEQSFIHPTPAPPSLGPGPAGSTVPHSQWGTLEPWELTELDSVGT 221
MACAQUE         -------------------------------------------------MDLLARDAPDT 140
                                                                 ::      .:


HUMAN           HLSQEQCQVASGHLPCIVRRTSKEACQQAGCCYDNTREVPCYYGNTATVQCFRDGYFVLV 289
CHIMPANZEE      HLSQERCQVASGHLPCIVRRTSKEACQQAGCCYDNTREVPCYYGNTATVQCFRDGYFVLV 289
HORSE           HLAQEQCQVASGRIPCRVSGSSREACQQAGCCYDNTREVPCYYGNTATVQCFRNGHFILV 337
RABBIT          HLPQEWCQVASGHIPCIVQSSSKEACEQAGCCYDSAREVPCYYGNTATVQCFRNGYFILV 272
MOUSE           HLPQERCQVASGHIPCMVNGSSKETCQQAGCCYDSTKEEPCYYGNTVTLQCFKSGYFTLV 281
MACAQUE         DWCDS--IPARDRLPCAPSPISRGDCEGLGCCYSS--ENSCYYGNTVTLRCTREGHFSIA 196
                .  :.    * .::**     *:  *:  ****..  * .******.*::* :.*:* :.


HUMAN           VSQEMALTHRITLANIHLAYAP-TSCSPTQHTEAFVVFYFPLTHCGTTMQVAGDQLIYEN 348
CHIMPANZEE      VSQEMALTHRITLANIHLAYAP-TSCSPTQHMEAFVVFYFPLTHCGTTMQVAGDQLIYEN 348
HORSE           VSQETASAHSFTLANVRLAYAP-TGCSPTQETGSFALFQFPLTHCGTTVQVVGNRLIYEN 396
RABBIT          VAQEMALAHRITLANVHLAYAP-TRCPPAQKTSAFVIFHVPLTHCGTTVQVLGSQLFYEN 331
MOUSE           MSQETALTHGVLLDNVHLAYAP-NGCPPTQKTSAFVVFHVPLTLCGTAIQVVGEQLIYEN 340
MACAQUE         VSRNVVSPP-LLLDSVRLALRNDSACNPVMATQAFVLFHFPFTSCGTTRRITGDRAVYEN 255
                :::: . .  . * .::**    . * *.    :*.:* .*:* ***: :: *.: .***


HUMAN           WLVSGIHIQKGPQGSITRDSTFQLHVRCVFNAS-DFLPIQASIFPPPSPAPMTQPGPLRL 407
CHIMPANZEE      WLVSGIHIQKGPQGSITRDSTFQLHVRCVFNAS-DFLPIQASIFPPPSPAPMTQPGPLRL 407
HORSE           QLVSDMDVRRGPQGSITREGTFRLHMHCIFNAS-DFLPLQASIFPPPSPAAVTQSGPLRL 455
RABBIT          QLVSDIDVREGPQGSITRDSSFRLLVRCIFNAS-DFLPIQASIFSPPLPAPVTQAGPLRL 390
MOUSE           QLVSDIDVQKGPQGSITRDSAFRLHVRCIFNAS-DFLPIQASIFSPQPPAPVTQSGPLRL 399
MACAQUE         ELVATRDVKNGSRGSVTRDSIFRLHVSCSYSVSSNSLPIKVQVFTLPPPFPETQPGPLTL 315
                 **:  .::.*.:**:**:. *:* : * :..* : **::..:*.   * . **.*** *


HUMAN           ELRIAKDETFSSYYGEDDYPIVRLLREPVHVEVRLLQRTDPNLVLLLHQCWGAPSANPFQ 467
CHIMPANZEE      ELRIAKDETFSSYYGEDDYPIVRLLREPVHVEVRLLQRTDPNLVLLLHQCWGAPSANPFQ 467
HORSE           ELRIAKDVTYGSYYGEGDYPIVRLLREPVPVEVRLLQRTDPSLVLVLHQCWATPSANPVQ 515
RABBIT          ELRIARDETFSSFYEEEDYPLVRLLREPVHVEVRLLQRTDPSLVLELHQCWATPSANPVQ 450
MOUSE           ELRIATDKTFSSYYQGSDYPLVRLLREPVYVEVRLLQRTDPSLVLVLHQCWATPTTSPFE 459
MACAQUE         ELQIAKDKNYGSYYGVGDYPVVKLLRDPIYVEVSILHRTDPSLGLLLHQCWATPSTDPLS 375
                **:** * .:.*:*   ***:*:***:*: *** :*:****.* * *****.:*::.*..


HUMAN           QPQWPILSDGCPFKGDSYRTQMVALDGAT--PFQSHYQRFTVATFALLD-SGSQRALRGL 524
CHIMPANZEE      QPQWPILSDGCPFKGDSYRTQMVALDGAT--PFQSHYQRFTVATFALLD-SGSQRALRGL 524
```

```
HORSE           QPQWPILWDGCPFDGDSYRTRLVALDGAEL-PFPSHYQRFTVATFVLLD-SGSQRALRGP 573
RABBIT          QPQWPLLSDGCPFKGDSYRTRVLALDRAEL-PFRSHYQRFTVATFTFLD-SGAQRALRGL 508
MOUSE           QPQWPILSDGCPFKGDNYRTQVVAADREAL-PFWSHYQRFTITTFMLLD-SSSQNALRGQ 517
MACAQUE         QPQWPILVKGCPYIGDNYQTQLIPVQKALDLPFPSHYQRFSIFTFSFVDPTVEKQALRGP 435
                *****:* .***: **.*:*:::. :     ** ******:: ** ::* :  :.****

HUMAN           VYLFCSTSACHTSGLETCSTACSTGTTRQRRSSGHRNDTARPQDIVSSPGPVGFEDSYGQ 584
CHIMPANZEE      VYLFCSTSACHTSGLETCSTACSTGTTRQRRSSGHRNDTARPQDIVSSPGPVGFEDSYGQ 584
HORSE           VYFFCSASACAPSGLETCATACSSRTARQRRSQSHRSDTAEPQNIVSSPGPVHFEGTHGQ 633
RABBIT          VYFFCSASACHPSGPETCSSTCSSRTAKRRRSSGYHDGTPRALDIVSSPGPVGFQDSHRQ 568
MOUSE           VYFFCSASACHPLGSDTCSTTCDSGIARRRRSSGHHNITLRALDIVSSPGAVGFEDAAKL 577
MACAQUE         VHLHCSVSVCQPAETPSSVRTCPDLSRRRKFSTIFQNTTAS----VSSKGPMILLQATKD 491
                *::.**.*.* .    :.  :*     ::: *  .:. *      *** *.: :  :

HUMAN           EPTLGPTDSNGNSSLRPLLWAVLLLPAVALVLGFGVFVGLSQTWAQKLWESNRQ----- 638
CHIMPANZEE      EPTLGPTDSNGNSSLRPLLWVVLSLPAVALVLGFGVFVGLRQTWAQKLWESNRQ----- 638
HORSE           EPTLRPTGSTRNSKPRPLLWMVLLLVAIALVLGVGIFVGLRQARAQKLQEGNRG----- 687
RABBIT          EPTLESTGSGRNSNPKPLLWVVLLLLAIALVLGIGVFVGLSQAWAHKLREGHRLTDQAQ 627
MOUSE           EP----SGSSRNSSSR----MLLLLLAITLALAAGIFVGLIWAWAQKLWEGIRY----- 623
MACAQUE         PP--EKLRAPVDS-------KVLWVAGLSGTLILGGLVVSYLAIKQLNCPDQTCQ---- 537
                 *      :  :*        :* : .:: .*  * :*    :  :    .
```

Figure 3. Comparison of amino acid sequences of ZP1 from human, chimpanzee, horse, rabbit, mouse and macaque. The deduced amino acid sequence of rabbit ZP1 was aligned with the ZP1 sequences of the other species using the Clustal W program. The accession numbers of the sequences used are as follows: horse ZP1 (XP_001493772), human ZP1 (NP_997224), mouse ZP1 (NP_033606) and rabbit ZP1. Identical amino acids are marked by an asterisk. The colon (:) represents conserved residues and the dot (.) represents semi-conserved residues. The potential signal peptidase cleavage is between Gly20 and Gln21. The zona domain is shown in red. The trefoil domain is shown in blue. The consensus furin cleavage-site is underlined. The transmembrane domain is in orange. The cystein residues are in green. The potential N-glycosylation sites are in pink.

Rabbit ZP1

```
  1 MTGGRLVALL LLVAASLGLG QQPHPEPGLP GLQYSYDCGM RGMQLVVLPR PGRTIRFKVV
 61 DEFGNRFEVN NCSICFHWVS AEPQAPAVFS ADYKGCHVLE KEGHSHLTVF IEAMLPDGHV
121 EVAQEAVLIC PKPGHTWAVG SHQVPPTTPS PTTPHALPFH LSSAHTFPIP LYLEHSLMLP
181 TPAGPSLGPG PTPAVLAQVE RWEVDKPDAV GSHLPQEWCQ VASGHIPCIV QSSSKEACEQ
241 AGCCYDSARE VPCYYGNTAT VQCFRNGYFI LVVAQEMALA HRITLANVHL AYAPTRCPPA
301 QKTSAFVIFH VPLTHCGTTV QVLGSQLFYE NQLVSDIDVR EGPQGSITRD SSFRLLVRCI
361 FNASDFLPIQ ASIFSPPLPA PVTQAGPLRL ELRIARDETF SSFYEEEDYP LVRLLREPVH
421 VEVRLLQRTD PSLVLELHQC WATPSANPVQ QPQWPLLSDG CPFKGDSYRT RVLALDRAEL
481 PFRSHYQRFT VATFTFLDSG AQRALRGLVY FFCSASACHP SGPETCSSTC SSRTAKRRRS
541 SGYHDGTPRA LDIVSSPGPV GFQDSHRQEP TLESTGSGRN SNPKPLLWVV LLLLAIALVL
601 GIGVFVGLSQ AWAHKLREGH RLTDQAQ
```

Rabbit ZP2

```
  1 MQVSNSGSRG KRLPWPSLTK FTFPYLSPPS SSSACTWLFL FFTLVTSVNS IYFLQLSDPA
 61 FPGTVTCNEN EIMVEFPSYV GTKTLHASVV DPLGVEMLNC TYILDPEKLT LRVPYKACTR
121 AVHGGHQMSI RVMNNSAAL RHTDVEYQFFC PVEQTLEFSK SAACTKDFMS LSFPRIPTGL
181 GDSTMVNESQ MGWMVQAGHG PGAQTLSLEE AKGQGFGVLI DDNKMTLSVL LNATGVTHYV
241 EGTSHLHTMF LKLSLVSPGQ KMTFPSRAIC LSGPVTCNAT HMTLTIPEFP GKLESVSIEN
301 RNITVSQLHD QGIDVEAING LRLHFSKTVL KTKFSEKCLH DQLYISSLKL TFNLELDTVS
361 TVINPECPCD SPASIVSGEL CTQDGFMDFE VYTHQTKPAL NLDTLRVGSS SCQPVFKAQS
421 QGLVRFRIPL NGCGTRHKFE DEKVIYENEV HALWENLPPS KISRDSEFRM TVQCYYTRDD
481 MLLNANIKSL PPPVASVKPG PLALSLQTYP DESYQQPYRV NEYPIVKYLR QPIYMEVRVL
541 NRNDPNIKLA LDDCWATSSM DPASLPKWSI VMDGCEYSLD NYQTNFHPVG SSVTYPEHYQ
601 RFDVKTFAFV SEAQARSSLV YFHCSALICN QHYPDSPLCS VTCPGSSRHR RATGNTEEER
661 VTASLPGPIL LLPNGSSFRG VGDSKEHGMA GDVTSKTMAA VAAVAGVVAT LGFISYLCKK
721 RTMMLSH
```

Rabbit ZP3

```
  1 MGLSYGLFVC LLLWGGSELC CPQPLWFWQG GTRQPAPSVT PVVVECLEAR LVVTVSRDLF
 61 GTGKLIQEAD LSLGPEGCEP QASTDAVVRF EVGLHECGNS VQVTDDSLVY SSFLAGNLSI
121 LRTNRAEVPI ECRYPRQGNV SSRAILPTWV PFWTTVLSEE RLVFSLRLME ENWSREKMSP
181 TFHLGDTAHL QAEVRTGSHP PLLLFVDRCV ATPTRDQSGSP YHTIVDLHG CLVDGLSDGA
241 SKFKAPRPKP DVLQFMVAVF HFANDSRHTV YITCHLRVIPA QQAPDRLNK ACSFNQSSSS
301 WAPVEGSADI CECCGNGDCD LIAGSPMNQN HAARSSLRSRR HVTEEADVT VGPLIFLGKA
361 GDPAGTEGLA SAAQATLVLG LRMATIVFLA VAAVVLGLTRG RHAASHPRS ASQ
```

Rabbit ZP4

```
  1  MAPGSTMWLL GYIFLCFPVS PKPFALIKQE TPTDPGVLHC RPWNFKFTIN FQNQETGSSP
 61  VLVTWDNQGR LHRLQNDTDC GTRVGEGPGP SVVLEANYSS CYVTESEPYY VMLVGVEEVD
121  AAGQNLVTKQ QLLKCPMHLP APDAGLCDSV PVQDRLPCAT APISQEDCEE LGCCHSSEEV
181  NACYYGNTVT SHCTQEGHFS IAVSRNVSSP PLHLDSVHLV FGNDSECQPV VATRAFVLFL
241  FPFTACGTTR QITGDRAIYE NELLATREVR TWSRGSITRD SIFRLRVSCS YSISSSALPV
301  DMHVLTLPPP LPETQPGPLT VVLQIAKDKD YHSYYTMDDY PVVKLLRDPI YVDVSILYRT
361  DPYLGLRLHQ CWATPRTNPL YQPQWPILVK GCPYTGDNYQ TQLIPVQEAF DLPFPSHHQR
421  FSISTFSFLD SSVAKEALKG PIYLHCSVSV CQPTGTQSCT VTCPIDSRRR NSDINFQNST
481  ANISSKGPMI LLQATEDPSE KLHKHSGVPV HPGALWVAGL SGIFIIGALL VSYVAIRTRR
```

Figure 4. Rabbit ZP1, ZP2 (XP_002711880), ZP3 (NP_001182649), ZP4 (NP_001075764) amino acid sequences. Underlined sequences are the tryptic peptides obtained by MS/MS. The detected putative N-glycosylation sites are in red and the O-glycosylation sites are in blue.

A) Tursiops

```
                D  E  F  G  N  R  F  D  V  N  N  C  S  I  C  Y  16
Homo          G ATGAATTTGGGAACCGATTTGATGTCAACAACTGCTCCATCTGCTAC 48
                |||||||||||||||  | ||| || |||| ||||||||||||  |
Tursiops      ? ATGAATTTGGGAACCCGTCTGAGGTGAACAGCTGCTCCATCTGTCAT 47
                  E  F  G  N  P  S  E  V  N  S  C  S  I  C  H   15
```

```
                H  W  V  T  S  R  P  Q  E  P  A  V  F  S  A  D  32
Homo          CACTGGGTCACCTCCAGGCCGCAGGAGCCTGCAGTCTTCTCGGCCGAT 96
              |||  |    |  | || ||   ||  ||       || || ||| ||
Tursiops      CACGGGTGACCGCCGAGCCCCAGGGGCCCGCGGTCTTTTTCTGCCAAT 95
                H  G  *  P  P  S  P  R  G  P  R  S  F  S  A  N  31
```

```
                Y  R  G  C  H  V  L  E  K    41
Homo          TACAGAGGCTGCCACGTGCTGGAGAAG  123
              ||||  ||  |||  ||  ||||||||||||
Tursiops      TACAAAGACTGTCATGTGCTGGAGAAG  122
                Y  K  D  C  H  V  L  E  K    40
```

B) Callithrix

```
                D  E  F  G  N  R  F  D  V  N  N  C  S  I  C  Y  16
Homo          G ATGAATTTGGGAACCGATTTGATGTCAACAACTGCTCCATCTGCTAC 48
                |||||||||||||| |||||||||| ||||||||||||||||||||
Callithrix    ? ATGAATTTGGGAACTGATTTGATGTGAACAACTGCCCCATCTGCTAC 47
                   E  F  G  N  *  F  D  V  N  N  C  P  I  C  Y  15
```

```
                H  W  V  T  S  R  P  Q  E  P  A  V  F  S  A  D  32
Homo          CACTGGGTCACCTCCAGGCCGCAGGAGCCTGCAGTCTTCTCGGCCGAT 96
              || |||||||||||||||| |||||||||||||||||||||| |||
Callithrix    CAGTGGGTCACCTCCAGGCCTCAGGAGCCTGCAGTCTTCTCGGCTGAT 95
                Q  W  V  T  S  R  P  Q  E  P  A  V  F  S  A  D  31
```

```
                Y  R  G  C  H  V  L  E  K   41
Homo          TACAGAGGCTGCCACGTGCTGGAGAAG 123
              |||||||||||||||||||||||||||
Callithrix    TACAGAGGCTGCCACGTGCTGGAGAAG 122
                Y  R  G  C  H  V  L  E  K   40
```

C) Tarsius

```
             D  G  R  F  H  L  R  V  F  M  E  A  V  L  P  N  G  R  V  D  20
Homo       GATGGGCGTTTCCACCTGAGGGTGTTCATGGAGGCTGTGCTGCCCAATGGTCGTGTGGAT 60
           | |||||||||||||||||| ||||| |||||    ||| |||| |   ||  ||||||
Tarsius    GGTGGGCGTTTCCACCTGAGGATGTTCGTGGAGAGCATGCAGCCCGACCATCAGGTGGAT 60
             G  G  R  F  H  L  R  M  F  V  E  S  M  Q  P  D  H  Q  V  D  20

             V  A  Q  D  A  T  L  I  C  P  K  P  D  P  S  R  T  L  D  S   40
Homo       CAGGCACAAGACGCTACTCTGATCTGTCCCAAACCTGACCCCTCCCGGACTCTGGACTCC 120
            ||||||| ||| |  |||||| || |||||||||||||||| || ||   | |||||||
Tarsius    TAGGCACAGGACACCGCTCTGACCTATCCCAAACCTGACCCCACCTGGGTCCCGGACTCC 120
             V  A  Q  D  T  A  L  T  Y  P  K  P  D  P  T  W  V  P  D  S   40
```

```
           Q  L  A  P  P  A  M  F  S  V  S  T  P  Q  T  L  S  F  L  P   60
Homo       CAGCTGGCACCACCCGCCATGTTCTCTGTCTCAACCCCACAAACCCTTTCCTTCCTCCCC 180
           | | ||||||||||| ||  ||||||  ||||  ||||  | | |   || |||  ||||
Tarsius    TACCCGGCACCACCCACCGAGTTCTCACTCTCTGCCCCTAATAGC---TCTTTCTCCCCC 177
           Y  P  A  P  P  T  E  F  S  L  S  A  P  N  S     S  F  S  P   59

           T  S  G  H  T  S  Q  G  S  G  H  A  F  P  S  P  L  D  P  G   80
Homo       ACCTCTGGCCATACCTCCCAAGGCTCTGGCCATGCCTTTCCCAGCCCACTGGACCCAGGG 240
           |||| |||||  || ||   ||  |  |||| ||| | | |||||| ||||||  || |
Tarsius    TCCTCCGGCCACGCCCCCG--GGGCCCAGCCACGCCCTGCTCAGCCCTCTGGACAAAGAG 235
           S  S  G  H  A  P    G  A  Q  P  R  P  A  Q  P  S  G  Q  R   78

           H  S  S  V  H  P  T  P  A  L  P  S  P  G  P  G  P  T  L  A  100
Homo       CACAGCTCTGTCCACCCAACCCCTGCTTTACCATCCCCTGGACCTGGACCTACCCTCGCC 300
           ||||||||| ||||||||| | || | |   | ||| |  ||| ||| ||| ||| | ||
Tarsius    CACAGCTCTATCCACCCAAGCGCTTCCTCTTCCTCCTCCAGACTTGGGCCTGCCCACCCC 295
           A  Q  L  Y  P  P  K  R  F  L  F  L  L  Q  T  W  A  C  P  P   98

           T  L  A  Q  P  H  W  G  T  L  E  H  W  D  V  N  K  R  D  Y  120
Homo       ACCCTGGCTCAACCCCACTGGGGCACCTTGGAACACTGGGATGTGAACAAACGAGATTAC 360
           ||||||||||||| || | |||||||||||| || |||||| ||| || ||| || ||
Tarsius    ACCCTGGCTCAACTCCTCGGGGGCACCTTGGGACCCTGGGAAGTGGACGAACCAGGTTCT 355
           H  P  G  S  T  P  R  G  H  L  G  T  L  G  S  G  R  T  R  F  118

           I  G  T  H  L  S  Q  E  Q  C  Q  V  A  S  G  H  L  P  C  I  140
Homo       ATAGGTACCCACCTGAGCCAGGAGCAGTGCCAGGTGGCCTCAGGGCACCTCCCCTGCATC 420
           |||||||||| |||| |||||| ||||||| ||||||||| |||| | ||||||||||||
Tarsius    TTAGGTACCCATCTGACCCAGGAACAGTGCCGGGTGGCCTCCGGGCCCATCCCCTGCATC 375
           F  R  Y  P  S  D  P  G  T  V  P  G  G  L  R  A  H  P  L  H  138

           V  R  R  T  S  K  E  A  C  Q  Q  A  G  C  C  Y  D  N  T  R  160
Homo       GTGAGAAGAACTTCAAAAGAAGCCTGTCAGCAGGCTGGCTGCTGCTATGACAACACCAGA 480
           ||||  |    |||| ||  |||||||||||||| |||||||||||||||| ||||
Tarsius    ATGAGTGG---CCCAAAGGAGTCCTGTCAGCAGGCTGACTGCTGCTATGACAACATCAGA 432
           H  E  W     P  K  G  V  L  S  A  G  *  L  L  L  *  Q  H  Q  157

           E  V  P  C  Y  Y  G  N  T  169
Homo       GAGGTTCCCTGTTACTATGGCAACACAG 508
           |||||||||| || ||||||||||||||
Tarsius    GAGGTTCCCTGCTATTATGGCAACACAG 460
           R  G  S  L  L  L  W  Q  H    166
```

D)Sus

```
          D  G  R  F  H  L  R  V  F  M  E  A  V  L  P  N  G  R  V  D    20
Homo      GATGGGCGTTTCCACCTGAGGGTGTTCATGGAGGCTGTGCTGCCCAATGGTCGTGTGGAT    60
          ||  ||       || |  |     || ||| |   ||| |     || | | |
Sus       AATCAGCAGCCGAGCCAGCAGAAAGCCA-GGAAG---TGGAGTTCC-TGTT-GCGGCTCA    55
          N  Q  Q  P  S  Q  Q  K  A   R  K    W  S  S   C   C  G  S    18

                                                                       }

          V  A  Q  D  A  T  L  I  C  P  K  P  D  P  S  R  T  L  D  S    40
Homo      GTGGCACAAGACGCTACTCTGATCTGTCCCAAACCTGACCCCTCCCGGACTCTGGACTCC   120
          |      ||||  ||     ||||||  |||||||||||| | || ||||     |||  |
Sus       GCAGATGAAGAACCT-----CATCTGTGGCAAACCTGACCACACCTGGACCTGACACTAC   110
           A  D  E  E  P       H  L  W  Q  T  *  P  H  L  D  L  T  L    36

          Q  L  A  P  P  A  M  F  S  V  S  T  P  Q  T  L  S  F  L  P    60
Homo      CAGCTGGCACCACCCGCCATGTTCTCTGTCTCAACCCCACAAACCCTTTCCTTCCTCCCC   180
          |  |||               |||||  |  |  |||| ||  |   | || ||| ||||
Sus       CCACTG-------------CATTCTCACTTCCTGCCCCGCAGCCTTGTCCCCTCCACCCC   157
          P  T              A  F  S  L  P  A  P Q  P  C  P  L  H  P    52
```

Figure 5. Ilustration of the presence of stop codon in the different sequences of the putative pseudogens. All the sequences are aligned to the ZP1 sequence of Homo. A) exon 2 of *Tursiops. B*) exon 2 of *Callithrix.* Exon 1 is missing in these two species, but the last nucleotide of the first exon of Homo was added for the translation. C) exon 3 and exon 4 (underlined) of *Tarsius.* D) first 180 bp of the exon 3 of *Sus* (the sequence of *Sus* is highly incomplete with numerous indels).

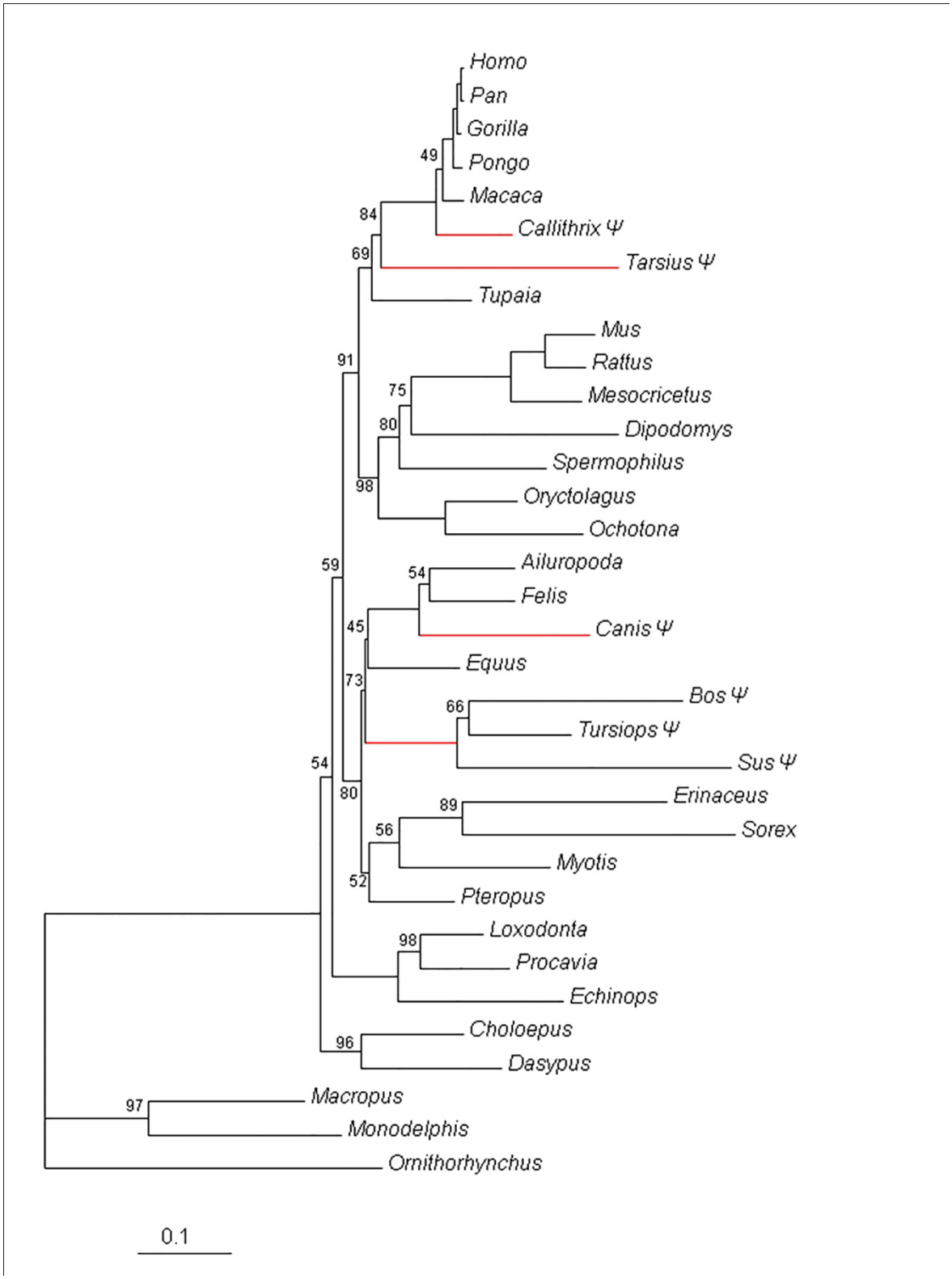


Figure 6. Phylogenetic relationship of ZP1 proteins reconstructed with PhyML. The numbers above each node are the bootstrap supports (only the values < 99 are shown). The symbol Ψ indicates possible pseudogenes. The branches in red indicate the position of the probable pseudogenization event.

TABLES

Table I: primers used in amplification of rabbit ZP1.

Primers	Forward (position in sequence)	Reverse	Amplified region (pb)
Fw1 and Rv1	at gac tgg ggg tcg cct ggt (1)	ctc ctg ggg cag atg gct acc tac	651
Fw2 and Rv2	ggt gga acg ctg gga agt gg (594)	gaa gat gga cgc ctg gat gg	522
Fw3 and Rv3	tct tca atg cca gcg act tc (1079)	ctca ggc cca caa aga cac ca	746
Fw4 and Rv4	aga ctt gct cat cta cgt gt (1571)	tta ttg agc ctg gtc ggt ga	314

Supplementary file 1: Primers used in amplification of rabbit ZP2, ZP3 and ZP4.

Primers	Forward (position in sequence)	Reverse	Amplified region (pb)
ZP2			
Fw1 and Rv1	aca tgg agt gag agt cct a (1601)	tta ttt caa cca ttt tat ttg caa g	614
ZP3			
Fw1 and Rv1	ggc tgg tgt tct ccc tgc g (500)	aga tgt ctg cac tgc ctt cca	450
ZP4			
Fw1 and Rv1	gtc ctg ttt ctg ttc cca ttt a (709)	ggc gta gac cta ggt atg ga	394

TABLE II: Peptides identified by proteomics analysis. Peptides with a score higher than 5, and percentage-scored peak intensity of 60%, which are the threshold criteria for a positive identification, are shown in red. The * indicates that the peptide has been also detected in ZP from oocytes.

ZP1							
Peptides	**Theoretical $[M+H]^+$**	**Sequence**	**z**	**m/z**	**Number of detections**	**Score**	**SPI**
GMQLVVLPRPGRTIR*	1693.0059	42-56	3	565.16	2	3.36	75
GMQLVVLPRPGRTIRFKVVDEFGNR	2884.6096	42-66	3	967.62	7	5.71	53.2
EVPCYYGNTATVQCFR	1850.8206	250-265	2	983.42	2	10.57	83.8
ITLANVHLAYAPTRCPPAQK*	2164.1701	283-302	3	741.80	1	3.86	62.6
EGPQGSITRDSSFR	1536.7406	341-354	3	565.89	1	3.91	51.7
EGPQGSITRDSSFRLLVR	2018.0783	341-358	3	674.09	1	5.21	60.5
IARDETFSSFYEEEDYPLVRLLR	2848.4157	394-416	3	950.07	1	3.06	57.9
LLREPVHVEVR*	1346.7908	414-424	2	674.17	1	4.10	87.7
GDSYRTRVLALDRAELPFR	2235.1998	465-483	2	1028.34	1	3.61	54.8
GDSYRTRVLALDRAELPFRSHYQR	2906.5138	465-488	3	996.28	1	4.21	61.4
TRVLALDR	943.5689	470-477	1	943.04	1	5.20	54.3
SHYQRFTVATFTFLDSGAQRALR*	2672.3697	484-506	3	918.66	9	5.34	86.4
FTVATFTFLDSGAQR	1660.8335	489-503	2	831.27	1	6.60	77.3
ALRGLVYFFCSASACHPSGPETCSSTCSSR	3124.3749	504-533	3	1095.04	1	5.51	71.5
ALRGLVYFFCSASACHPSGPETCSSTCSSRTAK	3424.5546	504-536	3	1188.22	1	6.31	54.5
GLVYFFCSASACHPSGPETCSSTCSSR	2784.1526	507-533	3	982.78	8	5.63	59.6
GLVYFFCSASACHPSGPETCSSTCSSRTAKR	3240.4335	507-537	3	1137.15	1	5.11	59.6
GLVYFFCSASACHPSGPETCSSTCSSRTAKRR	3396.5346	507-538	3	1133.40	1	4.58	55.6
RSSGYHDGTPRALDIVSSPGPVGFQDSHR	3095.5047	539-567	3	1059.79	3	6.22	79.0
SSGYHDGTPR*	1076.4761	540-549	2	578.23	3	3.78	61.8
SSGYHDGTPRALDIVSSPGPVGFQDSHR*	2399.4036	540-567	3	980.36	1	3.59	86.9

ZP2							
Peptides	**Theoretical $[M+H]^+$**	**Sequence**	**z**	**m/z**	**Number of detections**	**Score**	**SPI**
GKRLPWPSLTK	1282.7636	10-20	2	641.80	1	5.20	57.6
TLHASVVDPLGVEMLNCTYILDPEK*	2814.4058	84-108	3	973.15	17	10.24	63.2
TLHASVVDPLGVEMLNCTYILDPEKLTLR	2814.4058	84-112	3	1134.69	1	3.45	50.6
LTLRVPYKACTR	1420.8098	109-120	3	520.18	8	5.95	59.8
LTLRVPYKACTRAVHGGHQMSIR*	2594.3924	109-131	3	924.72	7	5.62	74.1
VPYKACTRAVHGGHQMSIR	2194.1014	113-131	3	750.36	1	5.62	54.8
VPYKACTRAVHGGHQMSIRVMNNSAALR	3067.5617	113-141	3	1074.28	2	5.13	54.6
ACTRAVHGGHQMSIR*	1623.7960	117-131	2	821.41	2	5.58	53.8
ACTRAVHGGHQMSIRVMNNSAALR*	2580.2822	117-141	3	919.09	8	4.97	54.7
AVHGGHQMSLRVMNNSAALR	2149.0871	121-141	3	743.70	2	5.11	65.2
HTDVEYQFFCPVEQTLEFSK	2504.1444	142-160	2	1252.16	1	16.57	94.0
SAACTKDFMSLSFPR	1660.7827	161-175	2	899.35	1	4.09	50.8
LSLVSPGQKMTFPSR	1647.8892	253-267	2	872.86	1	3.08	71.0
MTFPSRAICLSGPVTCNATHMTLTIPEFPGK	3320.6304	262-292	3	1145.56	1	5.13	53.5
AICLSGPVTCNATHMTLTIPEFPGK*	2601.8879	268-292	3	918.12	14	8.95	95.6
NITVSQLHDQGIDVEAINGLRLHFSK	2904.5332	302-327	3	995.05	1	3.90	52.5
LHFSKTVLKTKFSEK*	1793.0325	323-337	2	897.05	1	3.75	59.9
TVLKTKFSEKCLHDQLYISSLK	2581.4064	328-349	3	887.71	4	4.71	64.9
FSEKCLHDQLYISSLK	1910.9686	334-349	3	637.09	1	3.95	60.4
VGSSSCQPVFKAQSQGLVRFR	2281.1875	407-427	3	814.33	1	4.27	61.6

ZP2							
Peptides	**Theoretical $[M+H]^+$**	**Sequence**	**z**	**m/z**	**Number of detections**	**Score**	**SPI**
VGSSSCQPVFKAQSQGLVRFRIPLNGCGTRHK	3457.8061	407-438	6	599.62	1	4.40	54.8
AQSQGLVRFRIPLNGCGtRHKFEDEK	2986.5433	418-443	3	1022.77	1	3.16	57.3
FRIPLNGCGTRHKFEDEK	2147.0820	426-443	3	762.16	1	5.39	67.9
DSEFRMTVQCYYTR	1798.7892	465-478	3	654.00	1	3.51	70.0
DSEFRMTVQCYYTRDDMLLNANIK	2926.3538	465-488	3	999.96	1	4.97	67.7
MTVQCYYTRDDMLLNANIK	2292.0826	470-488	3	820.47	1	5.87	51.8
VNEYPIVK	961.5358	520-527	1	961.32	1	3.97	81.0
VNEYPIVKYLRQPIYMEVR	2410.2957	520-538	3	836.52	1	4.48	80.0
YLRQPIYMEVRVLNRNDPNIK	2631.4193	528-548	3	930.48	2	3.27	69.3
QPIYMEVRVLNRNDPNIK*	2199.1708	531-548	3	733.72	2	5.83	57.5
NDPNIKLALDDCWATSSMDPASLPK*	2702.2806	543-567	3	933.83	13	6.56	52.4
LALDDCWATSSMDPASLPK	2020.9360	549-567	2	1039.91	1	4.28	51.2
SSLVYFHCSALICNQHYPDSPLCSVTCPGSSRHRR	3920.8205	617-651	6	677.37	1	4.78	60.7
VTASLPGPILLLPNGSSFRGVGDSK	2482.3669	661-685	3	827.92	1	3.55	75.8
GVGDSKEHGMAGDVTSK*	1674.7757	680-696	2	917.15	3	6.01	71.5
EHGMAGDVTSKTMAAVAAVAGVVATLGFISYLCKK	3496.8006	686-720	3	1199.02	3	5.12	56.5
MAGDVTSKTMAAVAAVAGVVATLGFISYLCKKRTMMLSH	4353.2054	686-727	5	906.90	4	6.47	60.7
EHGMAGDVTSK*	1131.5104	686-696	2	606.24	1	4.19	63.3
TMAAVAAVAGVVATLGFISYLCKK*	2384.3086	697-720	3	827.53	4	6.31	72.1
TMAAVAAVAGVVATLGFISYLCKKR	2540.4097	697-721	3	893.13	1	3.46	85.1

ZP3							
Peptides	**Theoretical $[M+H]^+$**	**Sequence**	**z**	**m/z**	**Number of detections**	**Score**	**SPI**
QPAPSVTPVVVECLEARLVVTVSR	2549.4125	34-57	3	877.52	1	3.44	61.3
QPAPSVTPVVVECLEARLVVTVSRDLFGTGK	3267.7775	34-64	3	1104.02	1	3.94	94.1
LVVTVSR	773.4885	51-57	1	773.55	5	4.90	53.7
DLFGTGK*	737.3834	58-64	1	737.60	12	9.53	74.9
TNRAEVPIECRYPRQGNVSSR*	2432.2217	123-143	3	856.27	1	3.16	50.5
QGNVSSRAILPTWVPFWTTVLSEER*	2837.4950	137-161	3	984.42	3	4.62	60.5
AILPTWVPFWTTVLSEERLVFSLR	2860.5765	144-167	3	980.71	1	6.18	78.2
MSPTFHLGDTAHLQAEVRTGSHPPLLLFVDR	3442.7694	178-208	3	1175.04	1	4.69	50.3
DQSGSPYHTIVDLHGCLVDGLSDGASK*	2771.2946	216-242	3	943.90	18	13.38	74.1
FKAPRPKPDVLQFMVAVFHFANDSR*	2917.5299	243-267	3	973.10	1	3.01	59.4
APRPKPDVLQFMVAVFHFANDSR*	2642.3666	245-267	3	881.35	6	5.21	55.5
HTVYITCHLR*	1242.6417	268-277	2	650.33	1	4.11	72.7
HTVYITCHLRVIPAQQAPDRLNK*	2673.4411	268-290	3	918.57	1	6.11	75.1
SSLRSRRHVTEEADVTVGPLIFLGK*	2767.5219	335-359	3	923.12	4	5.42	68.5
SRRHVTEEADVTVGPLIFLGK	2324.2726	339-359	3	827.93	1	3.29	57.9
RHVTEEADVTVGPLIFLGK*	2081.1395	340-359	3	721.25	1	4.03	71.6
MATIVFLAVAAVVLGLTR*	1845.1036	383-399	2	970.90	2	4.80	68.5
MATIVFLAVAAVVLGLTRGRHAASHPR*	2814.6041	383-408	3	970.87	4	5.51	57.1

ZP4							
Peptides	**Theoretical $[M+H]^+$**	**Sequence**	**z**	**m/z**	**Number of detections**	**Score**	**SPI**
FTINFQNQETGSSPVLVTWDNQGR*	2738.3174	47-70	3	966.51	2	5.59	58.1
QQLLKCPMHLPAPDAGLCDSVPVQDR	2831.4006	130-155	3	984.75	5	5.31	84.5
CPMHLPAPDAGLCDSVPVQDR	2221.0204	135-155	3	778.71	1	4.65	73.9
AFVLFLFPFTACGTTR	1847.9518	235-250	2	924.46	4	16.01	91.8
QITGDRAIYENELLATR*	1963.0249	251-264	3	648.94	2	6.95	65.7
QITGDRAIYENELLATREVR	2347.2370	251-270	3	836.85	3	4.86	75.9
QITGDRAIYENELLATREVRTWSR*	2877.4971	251-274	3	987.25	2	3.49	61.2
AIYENELLATR	1292.6850	257-267	2	648.04	6	18.30	97.5
AIYENELLATREVRTWSRGSITR*	2721.4436	257-279	3	960.92	3	4.00	55.6
EVRTWSRGSITRDSIFR*	2066.0895	268-284	3	741.91	4	4.96	78.3
TWSRGSITRDSIFR*	1681.8774	271-284	2	921.26	1	3.49	67.7
DSIFR	637.3309	280-284	1	637.93	5	6.85	89.2
DKDYHSYYTMDDYPVVK	2138.9381	328-344	3	739.56	2	4.85	87.0
DKDYHSYYTMDDYPVVKLLR*	2521.2073	328-347	3	841.06	1	3.30	51.6
DYHSYYTMDDYPVVK	1895.8162	330-344	3	632.41	1	3.68	79.7
LLRDPIYVDVSILYR*	1835.0431	345-359	2	918.96	3	14.09	90.6
LLRDPIYVDVSILYRTDPYLGLR*	2750.5245	345-367	3	970.47	5	6.12	56.7
TNPLYQPQWPILVK	1696.9427	377-390	2	848.13	2	11.44	74.5
FSISTFSFLDSSVAK	1635.8270	421-435	2	818.34	3	12.47	79.6
EALKGPIYLHCSVSVCQPTGTQSCTVTCPIDSR	3493.6588	436-468	3	1191.29	1	3.82	84.7
RNSDINFQNSTANISSK	1895.9211	470-486	2	1028.34	1	3.40	58.5
NSDINFQNSTANISSK*	1739.8200	471-486	2	950.03	2	4.33	85.6
GPMILLQATEDPSEK*	1628.8205	487-501	2	895.71	1	3.68	93.7

Supplementary file 2: List of taxa used in the phylogenetic analysis with their origin. The * indicates that the sequences have been corrected after manual checking of the genomic sequences.

Scientific name	Common name	Accession numbers/ Origin-additional informations
Ailuropoda melanoleuca	Panda	ENSAMET00000008075*
Bos taurus	Cow	Predicted from genome/pseudogene, chr 29
Callithrix jacchus	Marmoset	Predicted from genome /pseudogene, chr 11
Canis familiaris	Dog	Predicted from genome /pseudogene, chr 18
Choloepus hoffmanni	Sloth	ENSCHOT00000008192*
Dasypus novemcinctus	Armadillo	ENSDNOT00000005786
Dipodomys ordii	Kangaroo rat	ENSDORT00000006903*
Echinops telfairi	Lesser hedgedog tenrec	ENSETET00000001351*
Equus caballus	Horse	XM_001493722
Erinaceus europaeus	Hedgehog	ENSEEUT00000006283*
Felis catus	Cat	ENSFCAT00000013671*
Gorilla gorilla	Gorilla	ENSGGOT00000013730
Homo sapiens	Human	NM_207341
Loxodonta africana	Elephant	ENSLAFT00000009456
Macaca mulatta	Macaque	XM_001084628
Macropus eugenii	Wallaby	ENSMEUT00000000742*
Mesocricetus auratus	Hamster	EU003563
Monodelphis domestica	Opossum	ENSMODT00000035038
Mus musculus	Mouse	U20448

Scientific name	Common name	Accession numbers/ Origin-additional informations
Myotis lucifugus	Microbat	ENSMLUT00000005035*
Ochotona princeps	Pika	ENSOPRT00000013207*
Ornithorhynchus anatinus	Platypus	ENSOANT00000023958*
Oryctolagus cuniculus	Rabbit	XM_002709016
Pan troglodytes	Chimpanzee	XM_522022
Pongo pygmaeus	Orangutan	ENSPPYT00000003825
Procavia capensis	Hyrax	ENSPCAT00000014254*
Pteropus vampyrus	Megabat	ENSPVAT00000014667*
Rattus norvegicus	Rat	NM_053509
Sorex araneus	Shrew	ENSSART00000003987*
Spermophilus tridecemlineatus	Squirrel	Predicted from genome /scaffold_129656
Sus scrofa	Pig	Predicted from genome /pseudogene, chr2
Tarsius syrichta	Tarsier	ENSTSYT00000001755*/ Pseudogene
Tupaia belangeri	Tree shrew	ENSTBET00000015513
Tursiops truncatus	Dolphin	ENSTTRT00000012462*/ Pseudogene

Supplementary file 3

Comparison of the different glycosylation sites of the four zona pellucida glycoproteins from different species.

ZP1

```
human           MAGGSATTWGYPVALLLLVAT-LGLGRWLQPDPGLPGLRHSYDCGIKGMQLLVFPRPGQT 59
rabbit          MTG------GRLVALLLLVAASLGLGQQPHPEPGLPGLQYSYDCGMRGMQLVVLPRPGRT 54
mouse           MAW------GCFVVLLLLAAAFLRLGQRLHLEP---GFEYSYDCGVRGMQLLVFPRPNQT 51
rat             MAW------GCFVVLLLLVAAFLRLCQHLHLKP---GFQYSYDCCVQCMQLLVFPRPNQT 51
                *:       *  *.****.*: * **:  : .*   *:.:*****::****:*:***.:*

human           LRFKVVDEFGNRFDVNNCSICYHWVTSRPQEPAVFSADYRGCHVLEK-DGRFHLRVFMEA 118
rabbit          IRFKVVDEFGNRFEVNNCSICFHWVSAEPQAPAVFSADYKGCHVLEK-EGHSHLTVFIEA 113
mouse           VQFKVLDEFGNRFEVNNCSICYHWVTSEAQEHTVFSADYKGCHVLEK-DGRFHLRVFIQA 110
rat             IQFKVLDEFGNRFEVNNCSICYHWVISEAQKPAVFSADYKGCHVLEKQDGRFHLRVFIQA 111
                ::***:*******:*******:*** :..*  :******:******* :*: ** **::*

human           VLPNGRVDVAQDATLICPKPDPSRTLDSQLAPPAMFSVSTPQTLSFLPTSGHTSQGSGHA 178
rabbit          MLPDGHVEVAQEAVLICPKPGHTWAVGSHQVPPTTPSPTTPHALPFHLSSAHT------- 166
mouse           VLPNGRVDIAQDVTLICPKPDHTVTPDPYLAPPTTPEPFTPHAFALHPIPDHTLAGSGHT 170
rat             VLPNGRVDTAQDVTLICPKPDHILTPESYLAPPTTPQPFIPHTFALHPISGHTLAGSGHT 171
                :**:*:*: **:..******.   :  .  .**:  .   *:::.:   . **

human           FPSPLDPGHSSVHPTPALPSPGPGPTLATLAQPHWGTLEHWDVNKRDYIGTHLSQEQCQV 238
rabbit          FPIPLYLEHSLMLPTPAGPSLGPGPTPAVLAQ-----VERWEVDKPDAVGSHLPQEWCQV 221
mouse           GLTTLYPEQSFIHPTPAPPSLGPGPAGSTVPHSQWGTLEPWELTELDSVGTHLPQERCQV 230
rat             GLTTLYPET---HPTPAPPSSEPGPVGLTVPQSQWGTLGSWELTELDSIGTHLLQERCQV 228
                   .*        **** **  ***.  .:.:     :  *:: : * :*:** ** ***

human           ASGHLPCIVRRTSKEACQQAGCCYDNTREVPCYYGNTATVQCFRDGYFVLVVSQEMALTH 298
rabbit          ASGHIPCIVQSSSKEACEQAGCCYDSAREVPCYYGNTATVQCFRNGYFILVVAQEMALAH 281
mouse           ASGHIPCMVNGSSKETCQQAGCCYDSTKEEPCYYGNTVTLQCFKSGYFTLVMSQETALTH 290
rat             ASGHIPCMVKGSSEEACQQAGCCYDNTKEMPCYYGNTVTLQCFRSG YTLVMSQETALTH 287
                ****:**:*. :*:*:*:*******.::* *******.*:***:.* : **::** **:*

human           RITLANIHLAYAPTSCSPTQHIEAFVVFYFPLTHCGTTMQVAGDQLIYENWLVSGIHIQK 358
rabbit          RITLANVHLAYAPTRCPPAQKISAFVIFHVPLTHCGTTVQVLGSQLFYENQLVSDIDVRE 341
mouse           CVLLDNVHLAYAPNGCPPTQKISAFVVFHVPLTLCGTAIQVVCEQLIYENQLVSDIDVQK 350
rat             GVMLDNVHLAYAPNGCPPTQKISAFVVFHVPLTLCGTAIQVVG-KLVYENQLVSNIEVQT 346
                 : * *:******. *.*:*:*.***:*:.*** ***::** * :*.*** ***.*.::

human           GPQGSITRDSTFQLHVRCVFNASDFLPIQASIFPPPSPAPMTQPGPLRLELRLAKDETFS 418
rabbit          CPQCSITRDSSFRLLVRCIFNASDFLPIQASIFSPDLPAPVTQACPLRLELRIARDETFS 401
mouse           GPQGSITRDSAFRLHVRCIFNASDFLPIQASIFSPQPPAPVTQSGPLRLELRIATDKTFS 410
rat             GPQGSITRDGVFRLHVRCIFNASDFLPIRASIFSPQPPAPVTRSGPLRLELRIATDKTFS 406
                *********. *:* ***:*********:****.*  ***:*:.********** *:***

human           SYYGEDDYPIVRLLREPVHVEVRLLQRTDPNLVLLLHQCWGAPSANPFQQPQWPILSDGC 478
rabbit          SFYEEEDYPLVRLLREPVHVEVRLLQRTDPSLVLELHQCWATPSANPVQQPQWPLLSDGC 461
mouse           SYYQGSDYPLVRLLREPVYVEVRLLQRTDPSLVLVLHQCWATPTTSPFEQPQWPILSDGC 470
rat             SYYQGSDYPLVRLLQEPVYIEVRLLQRTDPGLALMLHQCWATPSASPFEQPQWPILSDGC 466
                *:*  .***:****:***::**********.*.* *****.:*::.*.:******:*****

human           PFKGDSYRTQMVALDGAT-PFQSHYQRFTVATFALLDSGSQRALRGLVYLFCSTSACHTS 537
rabbit          PFKGDSYRTRVLALDRAELPFRSHYQRFTVATFTFLDSGAQRALRGLVYFFCSASACHPS 521
mouse           PFKGDNYRTQVVAADREALPFWSHYQRFTITTFMLLDSSSQNALRGQVYFFCSASACHPL 530
rat             PFKGDNYRTQMVAADRATLPFWSHYQRFTIATFTLLDSSSQNALRGQVYFFCSASACHPV 526
                *****.***:::* *    ** *******::** :***.:*.**** **:***:****.

human           GLETCSTACSTGTTRQRRSSGHRNDTARPQDIVSSPGPVGFEDSYGQEPTLGPTDSNGNS 597
rabbit          GPETCSSTCSSRTAKRRRSSGYHDGTPRALDIVSSPGPVGFQDSHRQEPTLESTGSGRNS 581
mouse           GSDTCSTTCDSGIARRRRSSGHHNITLRALDIVSSPGAVGFED----AAKLEPSGSSRNS 586
rat             GSETCSTTCDSEIARHRRSSGHHNSTIRALDIVSSPGAVGFED----APKLEPSGSTRNS 582
                * :***::*.:  :::*****::: * *. *******.***:*     ..* .:.*  **

human           SLRPLLWAVLLLPAVALVLGFGVFVGLSQTWAQKLWESNRQ----- 638
rabbit          NPKPLLWVVLLLLAIALVLGIGVFVGLSQAWAHKLREGHRLTDQAQ 627
mouse           SSR---MLLLLL-ATTLALAAGTFVGLTWAWAQKLWEGTRY----- 623
rat             GSRPLLWVLQLL-ALILVLGDGVLVGLSWAWAWA------------ 615
                . :     : ** *::*.*. *::***  :**
```

ZP2

```
cow             MACRQRGDSGRPSSWFRADW---------------RSFFLSFTLLTSVNSIDVNQLD-PA 44
pig             MACRHRGDSGRPLSWLSASW---------------RSLLLFFPLVTSVNSIGVNQLVNTA 45
human           MACRQRGGSWSPSGWFNAGWSTY------------RSISLFFALVTSVNSIDVSQLVNPA 48
rabbit          MOVSNSGSRGKRLPWPSLTKFTFPYLSPPSSSSACTWLFLFFTLVTSVNSIYFLOLSDPA 60
mouse           MARWQRKASVSSPCGRSIYRFLS---------------LLFTLVTSVNSVSLPQSENPA 44
rat             MARWQR-----------VYWLRS---------------LFFALVTSVNSLSLPQSENPA 33
                *   :                                 * *.*:*****: . *   .*

cow             FPGTVTCYENRMVVEFKRTLGNKIQHASVVDSLGLKMLNCTYVLDPEKLTLKAPYESCTK 104
pig             FPGIVTCHENRMVVEFPRILGTKIQYTSVVDPLGLEMMNCTYVLDPENLTLKAPYEACTK 105
human           FPGTVTCDEREITVEFPSSPGTKKWHASVVDPLGLDMPNCTYILDPEKLTLRATYDNCTR 108
rabbit          FPGTVTCNENEIMVEFPSYVGTKTLHASVVDPLGVEMLNCTYILDPEKLTLRVPYKACTR 120
mouse           FPGTLICDKDEVRIEFSSRFDMEKWNPSVVDTLGSEILNCTYALDLERFVLKFPYETCTI 104
rat             FPGTLICDKDEVRVEFSSRFDMEKWNPSLVDTFGNEISNCTYALDLEKFILKFPYETCTI 93
                *** : * : .: :**    . :   .*:**.:* .: **** ** *.: *: .*. **

cow             RVLGQHQMTITFMNDNTAHRQKTVLYHVSCPVMQAGRHDQHSGSTICSKDFMSFTFH-FF 163
pig             RVRGHHQMTIRLIDDNAALRQEALMYHISCPVMGAEGPDQHSGSTICMKDFMSFTFN-FF 164
human           RVHGGHQMTIRVMNNSAALRHGAVMYQFFCPAMQVEETQGLSASTICQKDFMSFSLPRVF 168
rabbit          AVHGGHQMSIRVMNNSAALRHTDVEYQFFCPVEQT---LEFSKSAACTKDFMSLSFPRIP 177
mouse           KVVGGYQVNIRVGDTTTDVRYKDDMYHFFCPAIQAET-HEISEIVVCRRDLISFSFPQLF 163
rat             KVIGGYQVNIRVQDTNADVSYKDDVHHFFCPAIQAEI-HEVSEIVVCMEDLISFSFPQLF 152
                 * * :*:.* . : .:        ::. **.  .      *  . * .*::*:::  .

cow             PGLADDTAG---PKPQMGWTVTVGDGERAQNLTLQEALTQGYNLLIENQKMSIQVLFHAT 220
pig             PGMADENVKREDSKQRMGWSLVVGDGERARTLTFQEAMTQGYNFLIENQKMNIQVSFHAT 224
human           SGLADDSKG---TKVQMGWSIEVGDGARAKTLTLPEAMKEGFSLLIDNHRMTFHVPFNAT 225
rabbit          TGLGDSTMVN---ESQMGWMVQAGHGPGAQTLSLEEAKGQGFGVLIDDNKMTLSVLLNAT 234
mouse           SRLADENQN----VSEMGWIVKIGNGTRAHILSLKDAIVQGFNLLIDSQKVTLHVPANAT 219
rat             SRLADENQN----VSEMGWIIKIGNGTRVHTLPLKDAIVQGFNLLIDSQKITLHVPANAT 208
                . :.*..        .*** :  *.*  .: *.: :*  :*:..**:.:::.: *  :**

cow             GVTHYSQGNSHLYMVPLKLTHVSPGQTIILSSRLICASDP-VTCNATHMTLTIPEFPGKL 279
pig             GVTRYSQGNSHLYMVPLKLKHVSHGQSLILASQLICVADP-VTCNATHVTLAIPEFPGKL 283
human           GVTHYVQGNSHLYMVSLKLTFISPGQKVIFSSQAICAPDP-VTCNATHMTLTIPEFPGKL 284
rabbit          GVTHYVEGTSHLHTMFLKLSLVSPGQKMTFPSRAICLSGP-VTCNATHMTLTIPEFPGKL 293
mouse           GIVHYVQESSYLYTVQLELLFSTTGQKIVFSSHAICAPDLSVACNATHMTLTIPEFPGKL 279
rat             GVAHYVQESSYLYTVQLKLLFSSPGQKITFSSQAICAPDLSVACNVTHMSLTIPEFPGKL 268
                *:.:* : .*:*: : *:*   : **.: :.*: ** ..  *:**.**::*:********

cow             KSVSFENKNIAVNQLHNSGIVMEIANGLRLHFSKTLLKTKFSEKCLPYQFYLSSLKLTFY 339
pig             KSVNLGSGNIAVSQLHKHGIEMETTNGLRLHFNQTLLKTNVSEKCQPHQLYLSSLRLTFH 343
human           KSVSFENQNIDVSQLHDNGIDLEATNGMKLHFSKTLLKTKLSEKCLLHQFYLASLKLTFL 344
rabbit          ESVSIENRNITVSQLHDQGIDVEAINGLRLHFSKTVLKTKFSEKCLHDQLYISSLKLTFN 353
mouse           ESVDFGQWSIPEDQWHANGIDKEATNGLRLNFRKSLLKTKPSEKCPFYQFYLSSLKLTFY 339
rat             KSVGFGQRNIPEDQWHANGIDKEATNGLRLHFRKSLLKTKPSEKCPFYQFYFSSLELTFN 328
                :**.: . .*  .* *  **  *  **::*:* :::***: ****   *:*::**.***

cow             TQLETVSMVIYPECVCESTVSIVSGELCTQDGFMDVEVYRHQTKPALNLDTLRVGDSSCQ 399
pig             SQLEAVSMVIYPECLCESTVSLVSEELCTQDGFMDVKVHSHQTKPALNLDTLRVGDSSCQ 403
human           LRPETVSMVIYPECLCESPVSIVTGELCTQDGFMDVEVYSYQTQPALDLGTLRVGNSSCQ 404
rabbit          LELDTVSTVINPECPCDSPASIVSGELCTQDGFMDFEVYTHQTKPALNLDTLRVGSSSCQ 413
mouse           FQGNMLSTVIDPECHCESPVSID--ELCAQDGFMDFEVYSHQTKPALNLDTLLVGNSSCQ 397
rat             FQGDMLSTVIDPECHCESPVSID--ELCTRDGFMDFEVYSHQTKPALNLESLLVGNSSCQ 386
                 . : :* ** *** *:*..*:   ***::*****.:*: :**:***:* :* **.****

cow             PTIKAPFQGLVKFHIPLNGCGTRHKFENGKVIYENEIHALWADLPPSTISRDSEFRMTVR 459
pig             PTFKAPAQGLVQFRIPLNGCGTRHKFKNDKVIYENEIHALWAD-PPSAVSRDSEFRMTVR 462
human           PVFEAQSQGLVRFHIPLNGCGTRYKFEDDKVVYENEIHALWTDFPPSKISRDSEFRMTVK 464
rabbit          PVFKAQSQGLVRFRIPLNGCGTRHKFEDEKVIYENEVHALWENLPPSKISRDSEFRMTVQ 473
mouse           PIFKVQSVGLARFHIPLNGCGTRQKFEGDKVIYENEIHALWENPPSNIVFRNSEFRMTVR 457
rat             PIFKVQSLGLARFHIPLNGCGTRQKFEGDKVIYENEIHALWENPPSNIIFRNSEFRMTVR 446
                * ::.   **.:*:********* **:. **:****:**** : *.. : *:*******:

cow             CYYSSSNMLINTNVESLPPPVASVKPGPLALTLQTYPDNSYLQPYGDKDYPVVRYLRQPI 519
pig             CSYSSSNMLINTNVESLPSPEASVKPGPLTLTLQTYPDNAYLQPYGDKEYPVVKYLRQPI 522
human           CSYSRNDMLLNINVESLTPPVASVKLGPFTLILQSYPDNSYQQPYGENEYPLVRFLRQPI 524
rabbit          CYYTRDDMLLNANIKSLPPPVASVKPGPLALSLQTYPDESYQQPYRVNEYPIVKYLRQPI 533
mouse           CYYIRDSMLLNAHVKGHPSPEAFVKPGPLVLVLQTYPDQSYQRPYRKDEYPLVRYLRQPI 517
rat             CHYIRDSMLLRAHIKSHSSPVASVKPGPLALVLQTYPDISYQRPYRKNEYPLVRYLRQPI 506
                * *  ..**:. :::. ..* * ** **:.* **:*** :* :**  .:**:*::*****

cow             YLEVRVLNRTDPNIKLVLDDCWATSTMDPASLPQWNIIVDGCEYNLDNHRTTFHPVGSSV 579
pig             YLEVRILNRTDPNIKLVLDDCWATSTEDPASLPQWNVVMDGCEYNLDNHRTTFHPVGSSV 582
```

ZP3

```
Rabbit          MGLSYGLFVCLLLWGGSELCCPQPLWFWQGGTRQPAPSVTPVVVECLEARLVVTVSRDLF 60
Human           MELSYRLFICLLLWGSTELCYPQPLWLLQGGASHPETSVQPVLVECQEATLMVMVSKDLF 60
Mouse           MASSYFLFLCLLLCGGPELCNSQTLWLLPGGTPTPVGSSSPVKVECLEAELVVTVSRDLF 60
Rat             MGPSCLLFLCLLLCGGPELCYPQTQWLLPGGTPTPAGSSSPVEVECKEAELVVTARRDLF 60
                *  *  **:**** *..*** .*. *:  **:  *  *  ** *** ** *:* . :***

Rabbit          GTGKLIQEADLSLGPEGCEPQAS--TDAVVRFEVGLHECGNSVQVTDDSLVYSSFLLHDP 118
Human           GTGKLIRAADLTLGPEACEPLVSMDTEDVVRFEVGLHECGNSMQVTDDALVYSTFLLHDP 120
Mouse           GTGKLVQPGDLTLGSEGCQPRVS-VDTDVVRFNAQLHECSSRVQMTKDALVYSTFLLHDP 119
Rat             GTGKLVQPGDLTLGSEGCQPLVA-VDTDVVRLNAQLHECSSGVQVTEDALVYSTFLLHDP 119
                *****:: .**:**.*.*:* .:     ***::. ****.. :*:*.*:****:******

Rabbit          RPAGNLSILRTNRAEVPIECRYPRQGNVSSRAILPTWVPFWTTVLSEERLVFSLRLMEEN 178
Human           RPVGNLSIVRTNRAEIPIECRYPRQGNVSSQAILPTWLPFRTTVFSEEKLTFSLRLMEEN 180
Mouse           RPVSGLSILRTNRVEVPIECRYPRQGNVSSHPIQPTWVPFRATVSSEEKLAFSLRLMEEN 179
Rat             RPVNGLSILRTNRVEVPIECRYPRQGNVSSHPIQPTWVPFSATVSSEEKLAFSLRLMEED 179
                **...***:****.*:**************:.* ***:** :** ***:*.********:

Rabbit          WSREKMSPTFHLGDTAHLQAEVRTGSHPPLLLFVDRCVATPT--RDQSGSPYHTIVDLHG 236
Human           WNAEKRSPTFHLGDAAHLQAEIHTGSHVPLRLFVDHCVATPT--PDQNASPYHTIVDFHG 238
Mouse           WNTEKSAPTFHLGEVAHLQAEVQTGSHLPLQLFVDHCVATPSPLPDPNSSPYHFIVDFHG 239
Rat             WNTEKSSPTFHLGEVAHLQAEVQTGSHLPLQLFVDHCVATPSPLPGQNSSPHHFIVDSHG 239
                *. ** :******:.******::**** ** ****:*****:   . ..**:* *** **

Rabbit          CLVDGLSDGASKFKAPRPKPDVLQFMVAVFHFANDSRHTVYITCHLRVIPAQQAPDRLNK 296
Human           CLVDGLTDASSAFKVPRPGPDTLQFTVDVFHFANDSRNMIYITCHLKVTLAEQDPDELNK 298
Mouse           CLVDGLSESFSAFQVPRPRPETLQFTVDVFHFANSSRNTLYITCHLKVAPANQIPDKLNK 299
Rat             CLVDGLSESFSAFQVPRPRPETLQFTVDVFHFANSSRNTVYITCHLKVAPANQIPDKLNK 299
                ******::. * *:.*** *:.*** * ******.**: :*******:*  *:* **.***

Rabbit          ACSFNQSSSSWAPVEGSADICECCGNGDCDLIAGS---PMNQNHAARSSLRSRRHVTEEA 353
Human           ACSFSKPSNSWFPVEGSADICQCCNKGDCGTPSHSRRQPHVMSQWSRSASRNRRHVTEEA 358
Mouse           ACSFNKTSQSWLPVEGDADICDCCSHGNCSNSSSSQFQIHGPRQWSKLVSRNRRHVTDEA 359
Rat             ACSFNKTSQSWLPVEGDADICDCCSNGNCSNSSSSEFETHEPAQWSTLVSRNRRHVTDEA 359
                ****.:.*.** ****.****:**.:*:*.  : *       : :    *.*****:**

Rabbit          DVTVGPLIFLGKAGDPAGTEGLASAAQATLVLGFGMATIVFLAVAAVVLGLTRGRHAASH 413
Human           DVTVGPLIFLDRRGDHEVEQWALPSDTSVVLLGVGLAVVVSLTLTAVILVLTRRCRTASH 418
Mouse           DVTVGPLIFLGKANDQTVEGWTASAQTS-VALGLGLATVAFLTLAAIVLAVTRKCHSSSY 418
Rat             DVTVGPLIFLGKANDQAVEGWTSSAQTS-VALGLGLATVAFLTLAAIVLGVTRMCHTSSY 418
                **********.: .*        .:  : : **.*:*.:. *:::*::* :**  :::*:

Rabbit          PRSASQ 419
Human           PVSASE 424
Mouse           LVSLPQ 424
Rat             LVSLPQ 424
                  * .:
```

```
human           YMEVRVLNRDDPNIKLVLDDCWATSTMDPDSFPQWNVVVDGCAYDLDNYQTTFHPVGSSV 584
rabbit          YMEVRVLNRNDPNIKLALDDCWATSSMDPASLPKWSIVMDGCEYSLDNYQTNFHPVGSSV 593
mouse           YMEVKVLSRNDPNIKLVLDDCWATSSEDPASAPQWQIVMDGCEYELDNYRTTFHPAGSSA 577
rat             YMEVTVLNRNDPNIKLVLDDCWATTFEDPASVPQWQIIMDGCEYELDNYRTTFHAANSSA 566
                *:** :*.* ******.*******:  ** * *:*.:::*** *.***::*.**...**.

cow             AYPNHYQRFAVKTFAFVSEDPAFSHLVYFHCSALICDQLSSNFPLCSASCLVSSRSRRAT 639
pig             TYPNHHQRFDVKTFAFVSGAQGVSQLVYFHCSVFICNQLSPTFSLCSVTCHGPSRSRRAT 642
human           THPDHYQRFDMKAFAFVSEAHVLSSLVYFHCSALICNRLSPDSPLCSVTCPVSSRHRRAT 644
rabbit          TYPEHYQRFDVKTFAFVSEAQARSSLVYFHCSALICNQHYPDSPLCSVTCPGSSRHRRAT 653
mouse           AHSGHYQRFDVKTFAFVSEARGLSSLIYFHCGALICNQVSLDSPLCSVTCPASLRSKRE- 636
rat             AHSGHYQRFDVKTFAFVSESRGLSSLIYFHCSALICNQA---SPLCSVTCPAPLRNKRE- 622
                ::. *:*** :*:*****     * *:****..:**::     .***.:*  . * :*

cow             GATEEEKMIVSLPGPILLLSDGSSFR----------------------------DAVDSK 671
pig             GTTEEEKMIVSLPGPILLLSDGSSLR----------------------------DAVNSK 674
human           GATEAEKMTVSLPGPILLLSDDSSFRGVGSSDLKASGSSGEKSRSETGEEVGSRGAMDTK 704
rabbit          GNTEEERVTASLPGPILLLPNGSSFR-----------------------------GVGDSK 685
mouse           -ANKEDTMTVSLPGPILLLSDVSSSKGVD-------------------------PSSSEIT 671
rat             -ASKEGTMTVSLPGPIILLSDDSSSKGVMN-----------------------PDSYEIT 658
                  .:   : .******:**.: ** :                              .  : .

cow             GHGTSGYAAFKTMVAVVALAGVVATLSLISYLRKKRITVLNH 713
pig             GSRTNGYVAFKTMVAMVASAGIVATLGLISYLHKKRIMMLNH 716
human           GHKTAGDVGSKAVAAVAAFAGVVATLGFIYYLYEKR-TVSNH 745
rabbit          EHGMAGDVTSKTMAAVAAVAGVVATLGFISYLCKKRTMMLSH 727
mouse           KDIIAKDIASKTLGAVAALVGSAVILGFICYLYKKRTIRFNH 713
rat             -----KDIASKTLGAVAALVGSAVIIGFICYLHKKRIVRFNS 695
                     *:: *:.* .* .. :.:* ** :**    .
```

ZP4

```
Pig             ---------MWLRPS-IWLCFPLCLALPGQSQPKAADDLGGLYCGPSSFHFSINLLSQDT 50
Rabbit          ---MAPGSTMWLLGY-IFLCFPVSFALIKQPKPETPTDPGVLHCRPWNFKFTINFQNQET 56
Human           ---------MWLLRC-VLLCVSLSLAVSGQHKPEAPDYSSVLHCGPWSFQFAVN-LNQEA 49
Rat             MARQALRSTLWLLPS-ILLCFPFCLPLSGQHVTELP---GVLHCGLQSFQFAVN-LSLEA 55
                         :**    : **. ..: :     .: .   . * *   .*:*::*  . ::

Pig             ATPPALVVWDRRGRLHKLQNDSGCGTWVHKGPGSSMGVEASYRGCYVTEWDSHYLMPIGL 110
Rabbit          GSSPVLVTWDNQGRLHRLQNDTDCGTRVGEGPGPSVVLEANYSSCYVTESEPYYVMLVGV 116
Human           TSPPVLIAWDNQGLLHELQNDSDCGTWIRKGPGSSVVLEATYSSCYVTEWDSHYIMPVGV 109
Rat             ES-PVLTTWDSQGLPHRLKNDSDCGTWVMDSPDGFLVLEASYSGCYVTLEGSHYIMTVGV 114
                 : *.* .** :*  * *:**:.*** : ..*.  :  **.* .****   .:*:* :*:

Pig             EEADAGGHRTVTETKLFKCPVDFLA--LDVPTIGLCDAVPVWDRLPCAPPPITQGECKQL 168
Rabbit          EEVDAAGQNLVTKQQLLKCPMHLP-----APDAGLCDSVPVQDRLPCATAPISQEDCEEL 171
Human           EGAGAAEHKVVTERKLLKCPMDLLA--RDAPDTDWCDSIPARDRLPCAPSPISRGDCEGL 167
Rat             QEADVAGHVAGTRQRLLTCPLALQGKAPDTPNAKVCSPVPVKERLPCASSTISRGDCEEL 174
                : ....      . :*:.**: :      .*    *..:*. :*****...*:: :*: *

Pig             GCCYNSEE--VPSCYYGNTVTSRCTQDGHFSIAVSRNVTSPPLLWDSVHLAFRNDSE-CK 225
Rabbit          GCCHSSEE--VNACYYGNTVTSHCTQEGHFSIAVSRNVSSPPLHLDSVHLVFGNDSE-CQ 228
Human           GCCYSSEE--VNSCYYGNTVTLHCTREGHFSIAVSRNVTSPPLLLDSVRLALRNDSA-CN 224
Rat             GCCYSSEEEGADSCYYGNTVTSHCTKEGHFSIAVSRDVTSPPLRLDSLRLGFRNITTGCD 234
                ***:.***  . :******** :**::**********:*:****  :*::* : * :  *.

Pig             PVMETHTFVLFRFPFSSCGTAKRVTGNQAVYENELVAARDVRTWSHGSITRDSIFRLRVS 285
Rabbit          PVVATRAFVLFLFPFTACGTTRQITGDRAIYENELLATREVRTWSRGSITRDSIFRLRVS 288
Human           PVMATQAFVLFQFPFTSCGTTRQITGDRAVYENELVATRDVKNGSRGSVTRDSIFRLHVS 284
Rat             PVMKTSTFVLFQFPLTSCGTTQRITGDQAMYENELVAIRDVQAWGRSSITRDSNFRLRVS 294
                **: * :**** **:::***::::**.:*:*****:* *:*:  .:.*:**** ***:**

Pig             CIYSVSSSALPVNIQVFTLPPPLPETHPGPLTLELQIAKDERYGSYYNASDYPVVKLLRE 345
Rabbit          CSYSISSSALPVDMHVLTLPPPLPETQPGPLTVVLQIAKDKDYHSYYTMDDYPVVKLLRD 348
Human           CSYSVSSNSLPINVQVFTLPPPFPETQPGPLTLELQIAKDKNYGSYYGVGDYPVVKLLRD 344
Rat             CTYSIHSIMSPVNMQVWTLPPPLPKTQPGPLSLELQIAQDKNYSSYYGTDAYPLVKFLQD 354
                * **  *   *::::* *****:*:*:** *:: *:**:*: * ***  . **:**:*::

Pig             PIYVEVSIRHRTDPSLGLHLHQCWATPGMSPLLQPQWPMLVNGCPYTGDNYQTKLIPVQK 405
Rabbit          PIYVDVSILYRTDPYLGLRLHQCWATPRTNPLYQPQWPILVKGCPYTGDNYQTQLIPVQE 408
Human           PIYVEVSILHRTDPYLGLLLQQCWATPSTDPLSQPQWPILVKGCPYIGDNYQTQLIPVQK 404
Rat             PIYVEVSILHRTDPSLSLLLEQCWATPGSNPFHQPQWPILVKGCPYAGDNYQTKRIPVQK 414
                ****:***  **** * * *.******  ..: *****:**:**** ******: *** :

Pig             ASNLLFPSHYQRFSVSTFSFVDS-VAKQALKGPVYLHCTASVCKPAGAPICVTTCPAARR 464
Rabbit          AFDLPFPSHHQRFSISTFSFLDSSVAKEALKGPIYLHCSVSVCQPTGTQSCTVTCPIDSR 468
Human           ALDLPFPSHHQRFSIFTFSFVNPTVEKQALRGPVHLHCSVSVCQPAETPSCVVTCPDLSR 464
Rat             ASDV-FPSHHQRFSISTFSFMSAGREKQVLGGQVYLHCSASVCQPAGMPSCTVICPASRR 473
                * :: ****:****: ****:..   *..* * ::***:.***:*:    *.. **   *

Pig             RRSSDIHFQNGTASISSKGPMILLQATRDSSERLHKYSRPPVDSHALWVAGLLG-SLIIG 523
Rabbit          RRNSDINFQNSTANISSKGPMILLQATEDPSEKLHKHSGVPVHPGALWVAGLSG-IFIIG 527
Human           RRNFDNSSQNTTASVSSKGPMILLQATKDPPEKLR----VPVDSKVLWVAGLSG-TLILG 519
Rat             RRKSELYFDNSTS-ISSKGPVILLQATKDPAVMLHKHSGTHADSPTLWVMGLSASMVITG 532
                **. :   :* *: :*****:***** .*..  *:      ... .*** ** .  .* *

Pig             ALLVSYLVFRKWR-------- 536
Rabbit          ALLVSYVAIRTRR-------- 540
Human           ALLVSYLAVKKQKSCPDQMCQ 540
Rat             VLVVSYLATRKQR-------- 545
                .*.:**:. :. :
```

Supplementary file 4

Percentage of similarities of the ZP proteins of different species compared to the corresponding human ZP proteins.

	Human			
	ZP1	ZP2	ZP3	ZP4
Cow	Pseudogene	67%	74%	69%
Hamster	68%	58%	70%	62%
Horse	72%	71%	79%	75%
Mouse	68%	58%	69%	Pseudogene
Pig	Pseudogene	63%	75%	67%
Rabbit	71%	69%	69%	70%
Rat	69%	58%	67%	64%

Cow (NM_173973, NM_173974, NM_173975), hamster (EU003563, AY876920, M63629, DQ838550), horse (XP_001493772, XP_001494819, XP_001493094, XP_001490803), human (P60852, AAH96304, NP_001103824, NM_021186.), mouse (AAI25614, NM_011775.6 and NM_011776.1), pig (NM_213848, NM_213893, NM_214045), rabbit (HQ702467, XM_002711834.1, NM_001195720.1, NM_001082295.1) and rat (NP_445961, NM_031150.1, NM_053762.1 and NM_172330.1)

CHAPTER II

Four glycoproteins are expressed in the cat zona pellucida.

INTRODUCTION

The zona pellucida (ZP) is an extracellular matrix that envelopes the mammalian oocyte and early embryos. A similar matrix is found in other vertebrates, including birds, amphibians and fish [1,2,3,4,5]. The specific functions of ZP are highly determined by its morphological structure, which is directly related to its composition [6]. However, instead of being an evolutionarily-conserved structure, this matrix is not constituted by the same proteins in all species. In birds and amphibians (where the matrix is called vitelline envelope) and fish (vitelline envelope or chorion) as many as six genes may be present: ZP1, ZP2, ZP3, ZP4, ZPD and ZPAX. Moreover, some species present duplications in some genes, for example, ZPAX in teleosts [7] and ZP3, which presents multiple copies in fish, amphibians (Xenopus) and birds (chicken, zebra finch) [8]. However, in euterian the number is more reduced and could vary between 3 and 4 genes codifying 3 or 4 proteins: ZP1, ZP2, ZP3 and ZP4, with ZP2 and ZP3 always being present [9,10,11,12,13,14]. However, in some mammals, duplications of ZP2 and ZP3 genes, have recently been detected. This is the case of platypus (*Ornithorhynchus anatinus*) and marsupials such opossum [8]. The meaning of these duplications remains to be investigated.

Mus musculus, the house mouse, has three proteins (ZP1, ZP2 and ZP3). This species has been used as a model to study the ZP since more than 30 years [9]. However, such ZP composition is not frequent in the mammalian oocytes described to date. Species such as pig, dog or cow also have three ZP proteins but in these cases the proteins are: ZP2, ZP3 and ZP4 [10,14,15]. And more recently, four proteins (ZP1, ZP2, ZP3 and ZP4) have been described in rat, human, macaque, hamster or rabbit [11,12,13,16,17]. The development of new techniques of genome sequencing and the apparition of –IOMICs techniques like transcriptomics or proteomics that has made an accurate analysis of the ZP genes and protein composition of ZP possible. Proteomic analysis has allowed the

detection of four proteins in rat, human, hamster and rabbit ZP [11,12,16,17]. Moreover, when the mRNA/protein has not been detected by transcriptomic/proteomic analysis in a given species, phylogenetic and bioinformatic analyses have detected the presence of pseudogenes, as reported in the mouse [11], cow and dog [14] and pig [17].

In the cat (*Felis catus*), characterization of the ZP by two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) suggested the presence of only two glycoproteins, covering a wide molecular mass range as bands between 50-110 kDa [18]. Later, Harris *et al.* [19] cloned and characterized the full length cDNAs of ZP2, ZP3 and ZP4 from a variety of mammalian species, including cat. The nucleotide and amino-acid sequences of these genes and corresponding proteins were deposited in Genbank database with these accession numbers: ZP2 (NM_001009875), ZP3 (NM_001009330) and ZP4 (NM_001009260). However, ZP1 has not been described in cat.

Furthermore, cat is a member of the mammalian order Carnivora and has sometimes been included in the same category as its related species, the dog, in which ZP1 is a pseudogen [8]. However, the lastest updates show that cat genome has a ZP1 gene with a possible ORF encoding a complete protein (GenBank accession number: NM_001284435). This record is derived from a genomic sequence (NW_004065056) annotated using the gene prediction method GNOMON.

The aim of this study was to determine the precise composition of cat ZP by means of proteomic and molecular biology techniques.

MATERIALS AND METHODS

2.1. Collection of cat (Felis catus) ovaries

Ovaries were obtained from domestic cats castrated at local veterinary clinics (24-32 week-old) (n=7). Depending on the experiment, they were frozen in liquid nitrogen and kept at −80 °C (for molecular and proteomic analysis) or transported to the lab in saline solution (0.9% NaCl, w/v) containing 100 μg/ml of kanamycin sulphate (Sigma-Aldrich Quimica SA, Spain) at 37 ºC within 1h of the ovariectomy (to obtain the ZPs).

2.2. Isolation of the zona pellucida

Cumulus-oocyte complexes (COCs) were excised from ovarian follicles using a 25 gauge needle. The COCs were placed in PBS four-well multidish (Nunc®, Thermo-Scientific Nunc, Roskilde, Denmark) and the cumulus cells were removed by gentle pipetting using 0.5% hyaluronidase (Sigma-Aldrich Quimica SA, Spain) in PBS. The oocyte ZPs were obtained after vigorous pipetting of each oocyte using a narrow-bore micropipette, following four washes in PBS to eliminate the oocyte debris.

2.3. Purification of cat ovarian RNA, cDNA synthesis and amplification by PCR of the open reading frame of cat ZP genes

Total RNA was isolated using the RNAqueous® kit (Ambion, Austin, Texas, USA) according to the manufacturer's instructions. The first-strand cDNA was synthesized from total RNA with the SuperScript First-Strand Synthesis System kit for RT-PCR (Invitrogen-Life Technologies, Carlsbad, CA, USA), according to the manufacturer's protocol.

Cat ZP1 was totally amplified by polymerase chain reaction (PCR) using specific primers (Table 1) designed according to the predicted cDNA sequence obtained from GenBank database (accession numberXM_003993490).

PCR amplifications were performed using 2 µl of target cDNA, 0.5 µg of each primer, 200 µM of each dNTP and 1 IU of Taq DNA polymerase (Thermo Scientific, Massachusetts, USA). PCR was carried out using an initial denaturation cycle of 2 min, and then 30 cycles of 1 min at 95 ºC, 1 min at annealing temperature (depending on the primers) and then 1 min at 72 ºC. The final extension time was 10 min at 72 ºC. 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 visualising under UV light using ethidium bromide.

Amplicons were carefully excised from the agarose gels and purified with the QIAquick Gel Extraction Kit Protocol (Quiagen, Hilden, Germany) according to the manufacter's protocol. After that, the amplicons were automatically sequenced (using 3500 Genetic Analyzer of Applied Biosystem). The sequences were analyzed to determine the degree of homology with other known sequences using the BLAST program (Basic Local Alignment Search Tool) (http://www.ncbi.nlm.nih.gov/blast/) [20]. Direct comparison between two sequences was made with the ALIGN program, and the multiple alignment of the ZP1 sequences of different species with the cat ZP1 sequence was carried out using Clustal W (http://www.ebi.ac.uk/clustalw/) [21].

The amino acid sequence was analyzed with the following software packages: "NetOGlyc" [22] and "NetNglyc" [23] to predict potential O-linked and N-linked glycosylation sites, respectively and SMART to predict the localization of the protein domains [24,25].

In addition, amplicons corresponding to ZP2, ZP3 and ZP4 were amplified in the same conditions as ZP1, while the primers were designed based on cDNA

sequences obtained from the GenBank database (ZP2: NM_001009875, ZP3: NM_001009330, ZP4: NM_001009260) (Table 1).

2.4. Proteomics analysis

The cat ovaries were trimmed using small scissors and dissected to remove fat and connective tissue. The solubilized ZP was obtained according to the protocol previously described by our group [16,26].

Solubilized ZP was obtained from oocytes, for which oocyte ZP (n=100) was solubilized at 65 ºC in PBS buffer for 30 min.

The analysis was carried out on a HPLC-MS system consisting of an Agilent 1100 Series 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.

The samples were diluted up to a final volume of 100 µL of 25 mM ammonium bicarbonate buffer pH 8.5 and then incubated with 50 mM tris (2-carboxyethyl) phosphine at 60°C for 10 min. The samples were then alkylated by adding 100 mM iodoacetamide and left to stand for 1 h at room temperature in the dark. Finally, 0.3 µg of proteomic grade trypsin (Sigma-Aldrich) was added to each sample for a digestion time of 16 h at 37°C. After this incubation, the tryptic digestion was stopped with 0.5% trifluoroacetic acid (TFA) and the samples were dried using a vacuum evaporator. The tryptic digestions were separated and analyzed by HPLC–MS. Dried samples 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), thermostated 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)/s 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 automated data-dependent mode. The most intense ions were sequentially fragmented using collision-induced dissociation (CID) with an isolation width of 2 Da and a relative collision energy of 35%. Data processing was performed with Data Analysis program for LC/MSD Trap Version 3.3 (Bruker Daltonik, GmbH, Germany) and Spectrum Mill MS Proteomics Workbench (Rev A.03.02.060B, Agilent Technologies, Santa Clara, CA, USA). Briefly, raw data were extracted under default conditions as follows: unmodified or carbamidomethylated cysteines; sequence tag length>1; [MH]+ 50–7000 m/z; maximum charge +7; minimum signal-to-noise (S/N) ratio 25; finding 12C signals. The MS/MS search against mammalian sequences in the NCBInr database was performed with the following criteria: identity search mode; tryptic digestion with 2 maximum missed cleavages; carbamidomethylated cysteines; peptide charge +1, +2, +3; monoisotopic masses; peptide precursor mass tolerance 2.5 Da; product ion mass tolerance 0.7 amu; ESI ion trap instrument; minimum matched peak intensity 50%; oxidized methionine, N-terminal glutamine conversion to pyroglutamic acid and STY phosphorylation as variable modifications. Two or more validated peptides were considered to demonstrate the existence of the protein. Peptides were considered valid with a score threshold of 5, and a percentage-scored peak intensity of 60%. All database matches above the threshold score of 3 were reported and used for discussion purposes.

RESULTS

3.1. MS/MS analysis

To confirm the existence of four proteins in cat ZP, the ZP extracted from ovaries and the ZP isolated from oocytes, as described in the Material and Methods section, were analyzed directly by MS/MS. A summary of the peptides identified is included in Table 2 and Figure 1.

For ZP1, a total of 20 different peptides were identified in the different analyses, yielding a sequence coverage of 33.17%. *In silico* analysis of the protein sequence revealed the existence of two N-glycosylation sites (Asn84 and Asn370). None of them was detected by proteomic analysis, suggesting that both the described consensus sequences may be occupied in the mature glycoprotein. On the other hand, 32 out of the predicted O-glycosylation sites part of the identified peptides, from which it can be deduced that these residues are either not glycosylated at all or are, at most, partially glycosylated (Fig. 1).

Thirty four peptides corresponding to cat ZP2 were detected. These peptides correspond to 71.50% of the protein. Ninety-one potential O-glycosylation sites and five N-glycosylation (Asn87, Asn96, Asn222, Asn268, Asn531 Asn710) sites were localized in the detected peptides (Fig. 1). The last one (Asn710) is localized in a peptide that is not present in mature protein. These residues might not be glycosylated or only be partially glycosylated in the native protein.

A total of 22 peptides from cat ZP3 were convincingly identified in different experiments. This corresponds to 50.23% of the sequence (Fig. 1). Only one N-glycosylation site (Asn145) was detected in the analysis and 31 O-glycosylation sites were detected in the experiments (Fig. 1). These sites may not be glycosylated in the native protein or, at most, be partially glycosylated.

Twenty four peptides were detected in ZP4 (49.64% coverage). Forty two O-glycosylation and one N-glycosylation (Asn237) sites were observed, which might mean that these sites are not glycosylated in mature protein. The rest of the N-glysosylation sites in the mature protein (Asn44, Asn68 and Asn477) were not detected in the peptides probably because they are occupied.

3.2. Transcript analysis

First, the cDNA of cat ZP1 was totally amplified, and the existence of a complete open reading frame (ORF) that could encode a protein was confirmed. Although the ORFs corresponding to cat ZP2, ZP3 and ZP4 were previously characterized [19], partial amplifications of each transcript (ZP2, ZP3 and ZP4) were made to confirm the above results (Fig. 2) and to confirm the presence of all four transcripts in the cat ovary.

Full-length rabbit ZP1 cDNA contains an ORF of 1956 nucleotides (Fig. 3) and shows 99% similarity with the predicted cat ZP1 deposited in Genbank database with accession number NM_001284435. The ATG initiation codon predicted with the Pedersen and Nielsen algorithm was found to be associated with vertebrate initiator codons. The ZP1 sequence contains a stop codon (TAG) in positions 1957-1959. The ORF of ZP1 codifies for a polypeptide 633 amino acids long (Fig. 3) with a theoretical molecular weight of 71.5 kDa. Two single nucleotide polymorphisms were detected in position 279 (c/t) and in position 1803 (c/g) but, in no case were changes detected in the amino acid sequence. This ZP protein share domains with other proteins of the same family. The protein possesses the archetypal 'ZP domain', a signature domain comprising 271 amino acid residues (271Gly-Asn541) rich in cysteine amino acid (nine Cys residues). The trefoil domain is also present. This part contains 44 residues (Glu225-Leu268) and is characteristic of ZP1 and ZP4. The last domain is the transmembrane domain (TMD) between Phe594 and Ile616, which is followed by

a cytoplasmic tail. A basic amino acid domain (^{525}Arg-Gln-Arg-Arg^{528}) upstream of the TMD may serve as a consensus furin cleavage site [27,28,29].

The comparative analysis of ZP1 nucleotide and amino acid sequence revealed high similarity with ZP1 from other species. The nucleotide sequence of cat ZP1 is 78% identical to human ZP1, 76% to rabbit ZP1, 74% to hamster ZP1 and 72% to rat ZP1. The analysis of amino acid sequence of cat ZP1 revealed that it is 69% identical to human ZP1, 67% to rabbit ZP1, 62% to rat ZP1 and 60% to hamster ZP1 (Fig. 4).

3.3. Identification of syntenic regions between human and cat genomes

Using the tool Ensembl Genome Browser *(*www.ensembl.org/*) w*e analyze the localization of cat ZP genes in their chromosomes and their alignment on the human genome. ZP1 is mapped in chromosome D1 in cat (106,541,040-106,548,667), corresponding to a syntenic region of chromosome 11 in human genome (60,867,562-60,875,693). ZP2 is present in cat genome in chromosome E3 (26,347,688-26,359,772 and align with chromosome 16 in human (21,197,450-21,214,510).

ZP3 is in chromosome E3 (8,878,565-8,885,605) and align with a region in chromosome 7 in human (76,397,518-76,442,071). ZP4 is in chromosome D2 in cat (10,881,111-10,888,573) and corresponds with the region situated in chromosome 1 in human (237,877,864-237,890,922).

DISCUSSION

The ZP, an extracellular matrix surrounding the oocyte, plays a key role during folliculogenesis, fertilization and early embryo development. Knowledge of the precise composition of ZP is necessary not only for understanding processes like oocyte-sperm interaction since it is also involved in acrosome reaction induction, the blocking of polyspermy or protection of the preimplantation embryo. Until now, mammalian ZP has been considered to be formed by 3 to 4 highly glycosylated proteins [9,10,11,12.13,14]. Different models have been proposed for mammalian ZP composition: 1) three proteins in cases where ZP1 is present as in the lab mouse (*Mus musculus*) (ZP1, ZP2 and ZP3) [9], 2) three proteins when ZP1 is absent; e.g. in ungulates (cow and pig) [14,17] and dog (ZP2, ZP3 and ZP4) [14] and 3) four proteins (ZP1, ZP2, ZP3 and ZP4), an increasingly numerous group. The rat, human, horse, macaque, hamster and rabbit are included in this last group [11,12,13,30,16,17].

ZP composition becomes more complicated if we look at other non-mammalian species like birds, amphibians or fishes, where more genes are involved (ZPAX, ZPD), and some of which have multiple copies of some ZP genes: four ZPAX copies in carp [7,31], five ZPAX copies in zebrafish [8,32,33], and six ZPAX copies in medaka [8,34,35]. Moreover, in these species ZP3 presents multiple copies: in fish (carp, medaka), birds (chicken, zebra finch) and amphibians (Xenopus) [8].

Nevertheless, the mouse model has been used for more than thirty years to study the ZP and its role in fertilization. However, more and more species show a different protein ZP composition, with mouse being the exception. The question arises, therefore, whether or not this model accurately reflects the mechanisms of ZP behaviour in other species.

The cat ZP has long been considered to be constituted by three glycoproteins based on description of the cDNA of ZP2, ZP3 and ZP4 in cat ovary [19]. Moreover, cat was included in the 3 ZP-glycoprotein group due to its closeness to the dog, a species where a ZP1 pseudogene has been described [14]. The situation in cat could be similar to other species containing ZP2, ZP3 and ZP4. Such is the case with *Canis* and *Bos,* where a pseudogene was found in the genome corresponding to ZP1 [14], while, more recently, ZP1 has been described as a pseudogene in *Sus*, *Tursiops*, *Callithrix* or *Tarsius* [17]. However, in *Felis* a complete ZP1 sequence has been annotated in the genome.

For that reason, the aim of this study was to determine of the precise composition of cat ZP using proteomic and molecular biology techniques, paying special attention to the detection of ZP1 protein. As a result, we present the first evidence supporting the existence of four proteins in cat ZP: ZP1, ZP2, ZP3 and ZP4. For this purpose, the cat ZP extracted from ovaries or oocytes were analyzed directly by MS/MS. A summary of the peptides identified is included in Table II and Figure 1. ESI-MS/MS spectra of some peptides corresponding to cat ZP1, ZP2, ZP3 and ZP4 are shown in supplementary file 1.

Peptides for the four proteins (ZP1, ZP2, ZP3 and ZP4) were detected by means of proteomic analysis in the ZP matrix, providing information about the putative glycosylation degree of these proteins. These glycans may be related with sperm binding and induction of the sperm acrosome reaction [36,37,38]. In ZP1, several O-glycosylation sites are present in the detected peptides, which suggests that these sites are unoccupied in the native protein and suggested that ZP1 is poor in O-glycans as previously reported for ZP1 in mouse and rat, in which proteomic analysis did not detect O-glycosylation sites [23,25]. On the other hand, neither of two putative N-glycosylation sites (Asn84 and Asn370) were identified by MS/MS analysis suggesting that both the described consensus sequences are occupied in the mature glycoprotein. These two potential N-glycosylation sites are conserved in the human, mouse, rat, hamster and rabbit

(supplementary file 2) and direct evidence of glycosylation is provided for mouse and rat [27,29]

As regards ZP2, several potential O-glycosylation sites and four N-glycosylation sites in mature protein (Asn87, Asn96, Asn222, Asn268) can be identified in the detected peptides (Fig. 3), meaning that they are not occupied in the native protein. In contrast, the conserved Asn corresponding to Asn87 and Asn222 was detected as glycosylated peptides in mouse and rat [27,29]. The Asn87 is also conserved in human, rabbit, cow and pig; however, Asn222 there is not present in cow and pig ZP2 (supplementary file 2) demonstrating the differences in glycosylation that exist between species.

In ZP3, 25 potential O-glycosylation sites and only one N-glycosylation site (Asn145) were identified in the detected peptides. The absence of glycosylation in the Asn145 site agrees with the proteomic analysis data obtained in mouse and rat ZP3, where no glycosylation was detected [27,29]. The fact that Asn123 was not detected suggests possible glycosylation. This result is similar to finding in pig and human where this Asn (Asn124-Asn125 in pig and human, respectively) is glycosylated [39,40]. As regards O-glycosylation, the presence of two clusters was described in mouse, rat and human ZP3 [27,29,40]. Only two Ser present in the first cluster of mouse and rat ZP3 sites seem to be glycosylated because no peptides were identified in this region, suggesting a glycosylation pattern similar to mouse and rat ZP3. However, the peptides containing the Ser and/or Thr in the position between Ser147 and Ser171 were detected in our proteomic analysis, indicating that these residues are not always glycosylated. It is important to take in consideration that the Thr155 previously identified in mouse was detected in cat ZP3 in our study, suggesting that this Thr is not always glycosylated; however, this Thr is considered to play a important role in sperm binding to the ZP [41].

In cat ZP4, the fact that Asn68 has not been identified suggests that this aminoacid is glycosylated, as found in rat [12] and in the pig ZP4 [42]. However,

the N-glycosylation site in position Asn237 was detected, suggesting that this amino acid is not occupied, whereas the equivalent site in the pig ZP (Asn203) is occupied. Another N-glycosylation site that is conserved in pig and rat has not been conserved in the cat (Asn220 in pig and Asn228 in rat). A substitution of N by K was produced. There is a N-glycosylation site in cat ZP4 in position Asn477 that it is probably glycosylated because the peptide containing this amino acid was not found by MS/MS.

Information on the O-glycosylation sites present in the ZP4 of different species is only available for rat and porcine [12,42], although a large number of potential O-linked glycosylation sites exists in both these species and the cat ZP4; however, only one glycosylated site is observed in the rat ZP4 and two sites in the pig ZP4. It was observed that the glycosylated Thr conserved in both the rat and and porcine ZP4 is also conserved in the cat ZP4 (Thr337), while the corresponding peptide was not detected using MS/MS analysis, suggesting that it is also occupied in the cat ZP4.

Future glycomic studies are necessary to obtain more precise information about ZP glycosylation in this species, which will allow a comparative analysis of the glycosylation patterns.

Several peptides corresponding to the immature version of the proteins were detected in the ZP obtained from ovaries. However, the peptides detected in mature ZP (ZP obtained from oocytes) are always present in the sequence corresponding to the mature protein (without the signal peptide and the carboxyl terminal region) (supplementary file 3) indicating that mature ZP corresponds to the polypeptide located between the signal peptide and the furin cleavage site.

Taking into account that HPLC–MS analysis can be considered as a semiquantitative technique, the fact that the coverage of ZP3 and ZP4 was similar (≈50%) might indicate the similar abundance of these two proteins. However, ZP2 seems to be the most abundant (71.50%), while ZP1 may be the least abundant protein in the ZP matrix (≈33%). Nevertheless, a similar situation

is observed in mouse, in which the levels of ZP1 mRNA are four times lower than those of ZP2 and ZP3 [43] so that only 56% of the ZP1 polypeptide chain can be identified by direct MS/MS, compared with the 96 and 100% of ZP2 and ZP3, respectively. The same occurs in rabbit [17], where ZP1 is the least detectable protein with this technique. However, these results do not agree with data obtained from other species. For example, in hamster [16], the coverage obtained in the different proteins were: 12.6% (ZP1), 5.1% (ZP2), 19.2% (ZP3), 11.2% (ZP4). In interpreting these data, it is important to know that the detection of peptides depends on the glycosylation. If the peptide is glycosylated it cannot usually be detected by this technique. For this reason, HPLC–MS analysis can be considered as a semiquantitative rather than a quantitative technique.

The second objective in our work was to confirm that the four genes are expressed by detecting their transcripts using molecular analysis. We also obtain a full-length cDNA for cat ZP1. Analysis of the sequence indicated that it has a complete coding region: an ORF, an initiation codon and a termination codon. A computer homology search in the GenBank database revealed significant homology with the ZP1 glycoprotein reported in other mammalian species, including human, mouse, rat, and rabbit.

From a phylogenetic point of view, the presence of ZP1 in cat suggests that the pseudogenization event that happened in dog (*Canis familiaris*) could be an isolated event, affecting only to the dog branch. Further transcriptomic and proteomic studies are necessary to clarify whether there are more members of Carnivora order with this four glycoprotein composition.

The function of ZP1 protein inside the whole ZP in species with four proteins remains to be investigated. A structural function was initially attributed to ZP1 based on the mouse model [44,45]. But in human, new functions have been attributed to ZP1, which has been related with binding and induction of the AR [46]. Better understanding of the ZP composition is basic to knowing the molecular bases of the sperm-ZP interaction. Because of its major role in the

fertilization process (interaction with sperm, induction of acrosome reaction, control of polyspermy), the ZP has been used as an attractive target for contraceptive vaccination [47,48]. This study provides a complete information about the proteins present in the cat ZP, contributing the development of new contraceptive studies. Vaccination with native porcine ZP (pZP) has been used in different species successfully [49,50,51]. However, in cats the effect of immunization with pZP is not clear, and contradictory results have been obtained [52,53,54,55,56]. Moreover, the different ZP composition could explain the lack of efficiency of some heterospecific antigens when they are used as anticonception vaccines in cat [53,57]. Pig presents a three glycoprotein composition compared with the four proteins present in cat. A previous study reported the effect of a panel of native ZP antigens isolated from several mammalian species (cows, cats, ferrets, dog and mink) [57], of which the most immunogenic antigen was from mink followed by ferret, both of them of the family Mustelidae. *In silico* analysis showed that the Ferret genome presents four ZP genes, as does mink probably, making them more similar to the cat ZP.

It was previously reported that the use of two proteins in the cat is more efficient than the use of only one ZP protein [56]. Probably, the development of a contraceptive vaccine including the ZP1 as antigen would be beneficial because the antibodies produced against ZP1 reduce sperm binding to the ZP [58,59,60].

CONCLUDING REMARKS

In summary, this study confirms the four ZP protein composition of cat ZP. Using a double approach, MS/MS analysis confirmed the presence of ZP1, ZP2, ZP3 and ZP4 proteins and molecular biology analysis detected the transcripts corresponding to the four proteins. This finding not only provides insight into the cat ZP composition but also promises benefits in the potential development of new contraceptive strategies. Some intriguing questions regarding the pseudogenization event in the Carnivore group, the glycosylation role of the ZP proteins, and the functional implication of ZP1 inside the matrix remain to be investigated.

REFERENCES

[1] Darie CC, Biniossek ML, Jovine L, Litscher ES, Wassarman PM. Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. Biochemistry 2004;43:7459-78.

[2] Barisone GA, Krapf D, Correa-Fiz F, Arranz SE, Cabada MO. Glycoproteins of the vitelline envelope of Amphibian oocyte: biological and molecular characterization of ZPC component (gp41) in Bufo arenarum. Mol Reprod Dev 2007;74:629-40. Erratum in: Mol Reprod Dev 2009;76:689.

[3] Modig C, Raldúa D, Cerdà J, Olsson PE. Analysis of vitelline envelope synthesis and composition during early oocyte development in gilthead seabream (*Sparus aurata*). Mol Reprod Dev 2008;75:1351-60.

[4] Kinoshita M, Rodler D, Sugiera K, Matsushima K, Kansaku N, Tahara K, Tsukada A, Ono H, Yoshimura T, Yoshizaki N, Tanaka R, Kohsaka T et al. Zona pellucida protein ZP2 is expressed in the oocyte of Japanese quail (*Coturnix japonica*). Reproduction 2010;139:359-71.

[5] Serizawa M, Kinoshita M, Rodler D, Tsukada A, Ono H, Yoshimura T, Kansaku N, Sasanami T. Oocytic expression of zona pellucida protein ZP4 in Japanese quail (*Coturnix japonica*). Anim Sci J 2011;82:227-35.

[6] Monné M, Jovine L. A structural view of egg coat architecture and function in fertilization. Biol Reprod 2011;85:661-9.

[7] Chang YS, Hsu CC, Wang SC, Tsao CC, Huang FL. Molecular cloning, structural analysis, and expression of carp ZP2 gene. Mol Reprod Dev 1997;46:258-67.

[8] Meslin C, Mugnier S, Callebaut I, Laurin M, Pascal G, Poupon A, Goudet G, Monget P. Evolution of genes involved in gamete interaction: evidence

for positive selection, duplications and losses in vertebrates. PLoS One 2012 *7*:e44548. doi: 10.1371/journal.pone.0044548.

[9] Bleil JD, Wassarman PM. 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.

[10] Hedrick JL, Wardrip NJ. On the macromolecular composition of the zona pellucida from porcine oocytes. Dev Biol 1987;121:478-88.

[11] Lefièvre L, Conner SJ, Salpekar A, Olufowobi O, Ashton P, Pavlovic B, Lenton W, Afnan M, Brewis IA, Monk M, Hughes DC, Barratt CL. Four zona pellucida glycoproteins are expressed in the human. Hum Reprod 2004;19:580-6.

[12] Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. J Biol Chem 2005;280:12721-31.

[13] Ganguly A, Sharma RK, Gupta SK. Bonnet monkey (*Macaca radiata*) ovaries, like human oocytes, express four zona pellucida glycoproteins. Mol Reprod Dev 2008; *75:* 156-66.

[14] 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.

[15] 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.

[16] Izquierdo-Rico MJ, Jimenez-Movilla M, Llop E, Perez-Oliva AB, Ballesta J, Gutierrez-Gallego R, Jimenez-Cervantes C, Aviles M. Hamster zona pellucida is formed by four glycoproteins. J Proteome Res 2009;8:926-41.

[17] Stetson I, Izquierdo-Rico MJ, Moros C, Chevret P, Lorenzo PL, Ballesta J, Rebollar PG, Gutiérrez-Gallego R, Avilés M. Rabbit zona pellucida composition: a molecular, proteomic and phylogenetic approach. J Proteomics 2012; 75: 5920-35.

[18] Maresh GA, Dunbar BS. Antigenic comparison of five species of mammalian zonae pellucidae. J Exp Zool 1987; 244: 299-307.

[19] Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG. Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB and ZPC gene families. DNA Seq 1994; 4:361-93.

[20] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol 1990;215:403-10.

[21] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994; 22:4673-80.

[22] Julenius K, Mølgaard A, Gupta R, Brunak S. Prediction, conservation analysis and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 2005;15:153-64.

[23] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput 2002;310-22.

[24] Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc Natl Acad Sci 1998;*95:5857-64.

[25] Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 2012;40:302-5.

[26] Jiménez-Movilla M, Martínez-Alonso E, Castells MT, Izquierdo-Rico MJ, Saavedra MD, Gutiérrez-Gallego R, Fayrer-Hosken R, Ballesta J, Avilés M. Cytochemical and biochemical evidences for a complex tridimensional structure of the hamster zona pellucida. Histol Histopathol 2009;24:599-609.

[27] Boja ES, Hoodbhoy T, Fales HM, Dean J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 2003;278:34189-202.

[28] Ducker P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel 2004;17:107-12.

[29] Boja ES, Hoodbhoy T, Garfield M, Fales HM. Structural conservation of mouse and rat zona pellucid glycoproteins. Probing the native rat zona pellucid proteome by mass spectrometry. Biochemistry 2005;44:16445-60.

[30] Mugnier S, Dell'Aquila ME, Pelaez J, Douet C, Ambruosi B, De Santis T, Lacalandra GM, Lebos C, Sizaret PY, Delaleu B, Monget P, Mermillod P et al. New insights into the mechanisms of fertilization: comparison of the fertilization steps, composition, and structure of the zona pellucida between horses and pigs. Biol Reprod 2009;81:856-70.

[31] Chang YS, Wang SC, Tsao CC, Huang FL. Molecular cloning, structural analysis, and expression of carp ZP3 gene. Mol Reprod Dev 1996;44:295-304.

[32] Wang H, Gong Z. Characterization of two zebrafish cDNA clones encoding egg envelope proteins ZP2 and ZP3. Biochim Biophys Acta 1999;1446:156-60.

[33] Mold DE, Kim IF, Tsai CM, Lee D, Chang CY, Huang RC. Cluster of genes encoding the major egg envelope protein of zebrafish. Mol Reprod Dev 2001;58:4-14.

[34] Kanamori A. Systematic identification of genes expressed during early oogenesis in medaka. Mol Reprod Dev 2000;55:31-6.

[35] Conner SJ, Hughes DC. Analysis of fish ZP1/ZPB homologous genes-evidence for both genome duplication and species-specific amplification models of evolution. Reproduction 2003;16:347-52.

[36] Velásquez JG, Canovas S, Barajas P, Marcos J, Jiménez-Movilla M, Gallego RG, Ballesta J, Avilés M, Coy P. Role of sialic acid in bovine sperm-zona pellucida binding. Mol Reprod Dev 2007;74:617-28.

[37] Chakravarty S, Kadunganattil S, Bansal P, Sharma RK, Gupta SK. Relevance of glycosylation of human zona pellucid glycoproteins for their binding to capacitated human spermatozoa and subsequent induction of acrosomal exocytosis. Mol Reprod Dev 2008;75:75-88.

[38] Pang PC, Chiu PC, Lee CL, Chang LY, Panico M, Morris HR, Haslam SM, Khoo KH, Clark GF, Yeung WS, Dell A. Human sperm binding is mediated by the sialyl-Lewis(x) oligosaccharide on the zona pellucida. Science 2011;333:1761-4.

[39] Nakano M, Yonezawa N. Localization of sperm ligand carbohydrate chains in pig zona pellucida glycoproteins. Cells Tissues Organs 2001;168:65-75.

[40] Zhao M, Boja ES, Hoodbhoy T, Nawrocki J, Kaufman JB, Kresge N, Ghirlando R, Shiloach J, Pannell L, Levine RL, Fales HM, Dean J. Mass spectrometry analysis of recombinant human ZP3 expressed in glycosylation-deficient CHO cells. Biochemistry 2004;43:12090-104.

[41] Visconti PE, Florman HM. Mechanisms of sperm-egg interactions: between sugars and broken bonds. Sci Signal 2010;3:35.

[42] Kudo K, Yonezawa N, Katsumata T, Aoki H, Nakano M. Localization of carbohydrate chains of pig sperm ligand in the glycoprotein ZPB of egg zona pellucida. Eur J Biochem 1998;252:492-9.

[43] Epifano O, Liang LF, Familari M, Moos MC, Dean J. Coordinate expression of the three zona pellucida genes during mouse oogenesis. Development 1995;121:1947-1956.

[44] Wassarman PM. Zona pellucida glycoproteins. Annu Rev Biochem 1988;57:415-42.

[45] Rankin T, Talbot P, Lee E, Dean J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development 1999;126:3847-55.

[46] Ganguly A, Bukovsky A, Sharma RK, Bansal P, Bhandari B, Gupta SK. In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. Hum Reprod 2010;25:1643-56.

[47] Fayrer-Hosken R. Controlling animal populations using anti-fertility vaccines. Reprod Domest Anim 2008;2:179-85.

[48] Levy JK. Contraceptive vaccines for the humane control of community cat populations. Am J Reprod Immunol 2011;66:63-70.

[49] Delves PJ, Roitt IM. Vaccines for the control of reproduction--status in mammals, and aspects of comparative interest. Dev Biol 2005;121:265-73.

[50] Jewgenow K, Dehnhard M, Hildebrandt TB, Göritz F. Contraception for population control in exotic carnivores. Theriogenology 2006;66:1525-9.

[51] Kirkpatrick JF, Rowan A, Lamberski N, Wallace R, Frank K, Lyda R. The practical side of immunocontraception: zona proteins and wildlife. J Reprod Immunol 2009;83:151-7.

[52] Ivanova M, Petrov M, Klissourska D, Mollova M. Contraceptive potential of porcine zona pellucida in cats. Theriogenology 1995;43:969-81.

[53] Jewgenow K, Rohleder M, Wegner I. Differences between antigenic determinants of pig and cat zona pellucida proteins. J Reprod Fertil 2000;119:15-23.

[54] Barber MR, Lee SM, Steffens WL, Ard M, Fayrer-Hosken RA. Immunolocalization of zona pellucida antigens in the ovarian follicle of dogs, cats, horses and elephants. Theriogenology 2001;55:1705-17.

[55] Gorman SP, Levy, JK, Hampton AL, Collante WR, Harris AL, Brown RG. Evaluation of a porcine zona pellucida vaccine for the immunocontraception of domestic kittens (*Felis catus*). Theriogenology 2002;58:135-49.

[56] Eade JA, Roberston ID, James CM. Contraceptive potential of porcine and feline zona pellucida A, B and C subunits in domestic cats. Reproduction 2009;137:913-22.

[57] Levy JK, Mansour M, Crawford PC, Pohajdak B, Brown RG. Survey of zona pellucida antigens for immunocontraception of cats. Theriogenology 2005;63:1334-41.

[58] Sivapurapu N, Upadhyay A, Hasegawa A, Koyama K, Gupta SK. Native zona pellucida reactivity and in-vitro effect on human sperm-egg binding with antisera against bonnet monkey ZP1 and ZP3 synthetic peptides. J Reprod Immunol 2002;56:77-91.

[59] Hardy CM, Clydesdale G, Mobbs KJ. Development of mouse-specific contraceptive vaccines: infertility in mice immunized with peptide and polyepitope antigens. Reproduction 2004;128:395-407.

[60] Sivapurapu N, Hasegawa A, Gahlay GK, Koyama K, Gupta SK. Efficacy of antibodies against a chimeric synthetic peptide encompassing epitopes of bonnet monkey (*Macaca radiata*) zona pellucida-1 and zona pellucida-3 glycoproteins to inhibit in vitro human sperm-egg binding. Mol Reprod Dev 2005;70: 247-54.

FIGURES

ZP1

MVAVVYLMTRASAMVWDGCVALLLLLVAALGLGQRPHPEPGLRGLRHSSDCGIKGMQLLVFPRPGQTVRFKVVDE
FGNQFEVHNCSVCYHWVTARPLGPAVFSADYRGCHVLEKGGRFHLRVFVEAVLRDGRVDAAGEVTLICPKPGHTW
TPESHLASRTGFSLPTPHTQPLRPTQEHSFTRPTPALLPLRPGATRPTLTLPQWDILEHWGVDEPLHPGAPLTWE
QCQVPSGHIPCVVRRGSKEACQKAGCCYDNSRAVPCYYGNTATVQCFRNGHFVLVVSQETALAHGITLANIHVAY
APTSCSPTQDTGSFVVFQFPLTHCGTTVQVVGNQLLYENQLVSDIDVRMGPQGSITRDGAFRLHVRCTVNASDFL
PLQASIFSPPSPVPVIQSGPLRFQLRIATDETFRSFYEEGDYPIVRLLREPVSVEVRLLDRTDPGLVLLLHQCWA
TPGVSPFQQPQWPILSEGCPFDGDSYRTRMVASDGAGLSFPSHHQRFTVTTFALLDPDSQRALRGQVYFFCHSSA
CSPSGLETCSTTCSSRPARQRRSYNPHGEATRPQNLVSSPGPVDFEDSSGQEPPLGPTGSPRNANQRPLLWVVLL
LVAVALVLGVGVFEGLSQAKPRSSRRATEGDWAQ*

ZP2

MASRQKGDSGSPSSWFNADWSTYRSLFLLFILVTSVNSIGVLQLVNPVFPGTVTCYETRMAVEFPSDFGTKKWHT
SVVDPFSFELLNCTYILDPENLTLKAPYETCTRRTLGQHRMIIRLKDHNAASRHNSLMYQINCPVMQAEETHEHA
GSTICTKDSMSFTFNVIPGLADENTDIKNPMGWSIEVGDGTKAKTLTLQDVLRQGYNILFDNHKITFQVSFNATG
VTHYMQGNSHLYMVPLKLIHESLGQKIILTTRVLCMSDAVTCNATHVTLTIPEFPGKLKSVSSENRNFAVSQLHN
NGIDKEESSGLTLHFSKTLLKMEFSEKCLPYQFYLASLKLTFAFNQETISTVLYPECVCESPVSIVTGDLCTQDG
FMDIKVYSHQTKPALNLETLRGGDSSCQPTFQAASQGLILFHIPLNGCGTRHKFKEGKVIYENEIHAVWADLPPS
TISRDSEFRMTVQCHYSKGDLLINTRVQSLPPPEASVRPGPLALILQTYPDKSYLQPYGEKEYPVVRYLRQPIYL
EVRVLNRSDPNIKLVLDDCWATPTMDPASVPQWNIIMDGCEYNLDNHRTTFHPVGSSVTYPTHYRRFDVKTFAFV
SEAQVLSSLVYFHCSVLICSRLSADSPLCSVTCPVSSRHRRATGTTEEEKMIVSLPGPILLLSDSSSLRDVVDSK
GYGAAGYVAFKTVVAVAALAGLVATLGFITYLRKNRTMINH*

ZP3

MGLSYGLFICFLLWAGTGLCYPPTTTEDKTHPSLPSSPSVVVECRHAWLVVNVSKNLFGTGRLVRPADLTLWPEN
CEPLISGDSDDTVRFEVELHKCGNSVQVTEDALVYSTFLLHNPRPMGNLSILRTNRAEVPIECRYPRHSNVSSEA
ILPTWVPFRTTMLSEEKLAFSLRLMEEDWGSEKQSPTFQLGDLAHLQAEVHTGRHIPLRLFVDYCVATLTPDQNA
SPHHTIVDFHGCLVDGLSDASSAFKAPRPRPETLQFTVYTFHFANDPRNMIYITCHLKVTPASRVPDQLNKACSF
IKSSNRWFPVEGPADICNCCNKGSCGLQGRSWRLSHLDRPWHKMASRNRRHVTEEADITVGPLIFLGKAADRGVE
GSTSPHTSVMVGIGLATVLSLTLATIVLGLARRHHTASRPMICPVSASQ*

ZP4

MWLLQPLLLCVPLSLAVHGQQKPQVPDYPGELHCGLQSLQFAINPSPGK**ATPALIVWDNR**GLPHKLQNNSGCGTW
VRESPGGSVLLDASYSSCYVNEWVSTTQSPGTSRPPTPASR**VTPQDSHYVMIVGVEGTDAAGRR**VTNTKVLRCPR
NPPDQALVSSLSPSPLQNVALEAPNADLCDSVPK**WDRLPCASSPITQGDCNKLGCCYKSEANSCYYGNTVTSRCT
QDGHFSIAVSRNVTSPPLLLNSLR**LAFGKDRECNPVKATR**AFALFFFPFNSCGTTRWVTGDQAVYENELVAARDV
RTWSHGSITRDSIFR**LRVSCSYSVRSNAFPLSVQVFTIPPPHLK**TQHGPLTLELKIAKDKHYGSYYTIGDYPVVK
LLRDPIYVEVSIRHRTDPSLGLLLHNCWATPGK**NSQSLSQWPILVK**GCPYVGDNYQTQLIPVQK**ALDTPFPSYYK
RFSIFTFSFVDTMAKWALR**GPVYLHCNVSICQPAGTSSCRITCPVARR**RRHSDLHHHSSTASISSKGPMILLQAT
MDSAEK**LHKNSSSPIDSQALWMAGLSGTLIFGFLLVSYLAIRKRR*

Figure 1. Cat ZP1 (XM_003993490), ZP2 (NM_001009875), ZP3 (NM_001009330) and ZP4 (NM_001009260) amino acid sequences. Bold underlined sequences are the tryptic peptides obtained by MS/MS. The putative N-glycosylation sites present in the detected peptides are in red and the putative O-glycosyilation sites are in blue. The furin cleavage site (Arg-X-(Lys/Arg)-Arg) is showed in yellow.

Figure 2. Analysis of ZP1, ZP2, ZP3 and ZP4 gene expression in cat ovary by RT-PCR. The primers used are shown in Table 1.

```
   1 atggtggccgtggtgtatctgatgacgagagcctcggccatggtctgggatggctgcgtg
   1  M  V  A  V  V  Y  L  M  T  R  A  S  A  M  V  W  D  G  C  V
  61 gcattgctgctgctgctggtggccgctctggggctggggcagcggccgcaccctgagcct
  21  A  L  L  L  L  L  V  A  A  L  G  L  G  Q  R  P  H  P  E  P
 121 ggtctccggggcctgcggcacagctctgactgtgggatcaagggcatgcagctgctggtg
  22  G  L  R  G  L  R  H  S  S  D  C  G  I  K  G  M  Q  L  L  V
 181 ttcccccggccgggccagaccgtccgcttcaaggtcgtagatgaatttgggaaccaattc
  42  F  P  R  P  G  Q  T  V  R  F  K  V  V  D  E  F  G  N  Q  F
 241 gaggtccacaactgctctgtctgctaccactgggtcacYgccaggcccctgggacccgcc
  62  E  V  H  N  C  S  V  C  Y  H  W  V  T  A  R  P  L  G  P  A
 301 gtcttctctgctgactacagaggttgccacgtgctggagaagggcgggcgcttccacctg
  82  V  F  S  A  D  Y  R  G  C  H  V  L  E  K  G  G  R  F  H  L
 361 agggtgttcgtagaagccgtgctgcgcgacggtcgagttgacgcggcaggagaggtcacc
 102  R  V  F  V  E  A  V  L  R  D  G  R  V  D  A  A  G  E  V  T
 421 ctgatctgtcctaaacctggccacacctggactccggagtcccacctggcctcacgcaca
 122  L  I  C  P  K  P  G  H  T  W  T  P  E  S  H  L  A  S  R  T
 481 ggcttctccctgcctacccctcatacccggcccctccgccccacccaggagcacagcttc
 142  G  F  S  L  P  T  P  H  T  R  P  L  R  P  T  Q  E  H  S  F
 541 acccgtccgacccctgccctgctgcccccttagacctggagccacccatcccaccctgact
 162  T  R  P  T  P  A  L  L  P  L  R  P  G  A  T  H  P  T  L  T
 601 ctgccccagtgggacatcttggaacactgggggttgacgagccccttcacccaggggca
 182  L  P  Q  W  D  I  L  E  H  W  G  V  D  E  P  L  H  P  G  A
 661 cctctgacctgggagcagtgccaggtgccctctgggcacatcccctgtgtggtgagaaga
 202  P  L  T  W  E  Q  C  Q  V  P  S  G  H  I  P  C  V  V  R  R
 721 ggttcgaaggaagcctgtcagaaggctggctgctgctatgacaatagcagagcagttccc
 222  G  S  K  E  A  C  Q  K  A  G  C  C  Y  D  N  S  R  A  V  P
 781 tgttactatggcaacacagcaacagtccagtgcttcagaaatggccatttcgtcctggtg
 242  C  Y  Y  G  N  T  A  T  V  Q  C  F  R  N  G  H  F  V  L  V
 841 gtgtcccaagaaacagccttggcacatgggatcacgctggccaacatccacgtggcctat
 262  V  S  Q  E  T  A  L  A  H  G  I  T  L  A  N  I  H  V  A  Y
 901 gccccccaccagctgctccccgactcaggacacggggtccttcgtggtcttccaattccct
 282  A  P  T  S  C  S  P  T  Q  D  T  G  S  F  V  V  F  Q  F  P
 961 ctgacccactgtgggaccaccgtccaggtggttggcaaccagctcctctatgagaaccag
 302  L  T  H  C  G  T  T  V  Q  V  V  G  N  Q  L  L  Y  E  N  Q
1021 ctggtgtcagacatcgacgtccgaatggggccacagggctccatcacacgggatggcgcc
 322  L  V  S  D  I  D  V  R  M  G  P  Q  G  S  I  T  R  D  G  A
1081 ttccggcttcacgtgcgctgtaccgtcaacgccagtgacttcctgccactccaggcgtct
 342  F  R  L  H  V  R  C  T  V  N  A  S  D  F  L  P  L  Q  A  S
1141 atcttctcacccccctcgccggtccccgtgatccagtccggccccctgcggttccagctg
 362  I  F  S  P  P  S  P  V  P  V  I  Q  S  G  P  L  R  F  Q  L
1201 cggatcgccacagatgagactttccgctccttctacgaggaggggactaccccatcgtg
 382   R  I  A  T  D  E  T  F  R  S  F  Y  E  E  G  D  Y  P  I  V
1261 aggctgctccgcgagcctgtgtccgtggaggtccggctcctggacaggacagaccccggt
 402   R  L  L  R  E  P  V  S  V  E  V  R  L  L  D  R  T  D  P  G
1321 ctggtcctgcttctgcaccagtgctgggccactcccggtgtcagccccttccagcagcct
 422   L  V  L  L  L  H  Q  C  W  A  T  P  G  V  S  P  F  Q  Q  P
1381 cagtggcccatcctgtcagaggggtgtcctttcgacggtgacagctacaggacccgaatg
 442   Q  W  P  I  L  S  E  G  C  P  F  D  G  D  S  Y  R  T  R  M
1441 gtagcctcagacggggcagggctgtccttcccgtcacaccaccagcgctttaccgtcacc
 462   V  A  S  D  G  A  G  L  S  F  P  S  H  H  Q  R  F  T  V  T
1501 accttcgccctcctggaccctgactcccagagggccctcaggggacaggtttacttcttc
 482   T  F  A  L  L  D  P  D  S  Q  R  A  L  R  G  Q  V  Y  F  F
1561 tgccactcctctgcctgctccccctcggggctggagacgtgctcgaccacgtgcagctct
 502   C  H  S  S  A  C  S  P  S  G  L  E  T  C  S  T  T  C  S  S
1621 aggcccgctagacagcgaagatcctacaatccgcacggcgaggccaccaggccgcagaac
 522   R  P  A  R  Q  R  R  S  Y  N  P  H  G  E  A  T  R  P  Q  N
1681 ctcgtgagctctccagggccagtggactttgaggattcctccgggcaggagcctccgctg
 542   L  V  S  S  P  G  P  V  D  F  E  D  S  S  G  Q  E  P  P  L
1741 gggcccacaggctcccccaggaacgctaatcagaggcctctcctctgggtggtccttctg
```

```
 562  G  P  T  G  S  P  R  N  A  N  Q  R  P  L  L  W  V  V  L  L
1801 ctSgtggctgtggccctggtccttggggtcggggttttcgagggcctgagccaagccaaa
 582  L  V  A  V  A  L  V  L  G  V  G  V  F  E  G  L  S  Q  A  K
1861 gcccagaagctccaggagggcgacagagggcgactgggctcaataaagcaccgtgtccag
 602   A  Q  K  L  Q  E  G  D  R  G  R  L  G  S  I  K  H  R  V  Q
1921 cctgcccagcgggtgtggaaagctattcgtggcgactag
 622   P  A  Q  R  V  W  K  A  I  R  G  D -
```

Figure 3. Nucleotide and deduced amino acid sequence of cat ZP1. The initial (atg) and final codon (tag) are in pink. The trefoil domain is shown in blue. The zona domain is shown in red. The consensus furin cleavage-site is underlined. The transmembrane domain is in orange. In red: two polymorphisms (positions: 279 (Y) and 1803 (S)).

```
MOUSE           ------------MAWGCFVVLLLLAAAPLRLGQRLHLEP---GFEYSYDCGVRGMQLLV 44
RAT             ------------MAWGCFVVLLLLVAAPLRLGQHLHLKP---GFQYSYDCGVQGMQLLV 44
HAMSTER         ------------MAWGCFVAVLLLVATPLRLGQHLHSKP---GLEYSYDCGVQGMQLLV 44
RABBIT          ------------MTGGRLVALLLLVAASLGLGQQPHPEPGLPGLQYSYDCGMRGMQLVV 47
HUMAN           -------MAGGSATTWGYPVALLLLVAT-LGLGRWLQPDPGLPGLRHSYDCGIKGMQLLV 52
CAT             MVAVVYLMTRASAMVWDGCVALLLLLVAALGLGQRPHPEPGLRGLRHSSDCGIKGMQLLV 60
                             . .  *.:*** .: * **:  : .*   *:.:* ***::****:*

MOUSE           FPRPNQTVQFKVLDEFGNRFEVNNCSICYHWVTSEAQEHTVFSADYKGCHVLEK-DGRFH 103
RAT             FPRPNQTIQFKVLDEFGNRFEVNNCSICYHWVISEAQKPAVFSADYKGCHVLEKQDGRFH 104
HAMSTER         IPRSNQTIRFKVLDEFGNRFEVNNCSICYHWVISEPHDPAVFSADYRGCHVLQK-DGRFH 103
RABBIT          LPRPGRTIRFKVVDEFGNRFEVNNCSICFHWVSAEPQAPAVFSADYKGCHVLEK-EGHSH 106
HUMAN           FPRPGQTLRFKVVDEFGNRFDVNNCSICYHWVTSRPQEPAVFSADYRGCHVLEK-DGRFH 111
CAT             FPRPGQTVRFKVVDEFGNQFEVHNCSVCYHWVTARPLGPAVFSADYRGCHVLEK-GGRFH 119
                :**..:*::***:*****:*:*:***:*:*** :..   :******:*****:*  *: *

MOUSE           LRVFIQAVLPNGRVDIAQDVTLICPKPDHTVTPDPYLAPPTTPEPFTPHAFALHPIPDHT 163
RAT             LRVFIQAVLPNGRVDTAQDVTLICPKPDHILTPESYLAPPTTPQPFIPHTFALHPISGHT 164
HAMSTER         LRVFVQAVLPNGYVDTAQDVTLICPKADHTVTPDPYLAPPTTPQPFTPHTFVPHTNSGHT 163
RABBIT          LTVFIEAMLPDGHVEVAQEAVLICPKPGHTWAVGSHQVPPTTPSPTTPHALPFHLSSAHT 166
HUMAN           LRVFMEAVLPNGRVDVAQDATLICPKPDPSRTLDSQLAPPAMFSVSTPQTLSFLPTSGHT 171
CAT             LRVFVEAVLRDGRVDAAGEVTLICPKPGHTWTPESHLASRTGFSLPTPHTQPLRPTQEHS 179
                * **::*:* :* *: * :..*****..   :  .  .. :  .   *::       *:
```

```
MOUSE           LAGSGHTG--------LTTLYPEQSFIHPTPAPPSLGPGPAGSTVPHSQWGTLEPWELTE 215
RAT             LAGSGHTG--------LTTLYPET---HPTPAPPSSEPGPVGPTVPQSQWGTLGSWELTE 213
HAMSTER         LAGSGHTLAGSGHTPLLSTLYPEHSFIHSTPAPPSPGPGPAGPTVPHPQWGTLEPLELTK 223
RABBIT          FP---------------IPLYLEHSLMLPTPAGPSLGPGPTPAVLAQ-----VERWEVDK 206
HUMAN           SQGSGHAFP--------SPLDPGHSSVHPTPALPSPGPGPTLATLAQPHWGTLEHWDVNK 223
CAT             FT------------------------RPTPALLPLRPGATRPTLTLPQWDILEHWGVDE 214
                                          .***  .  **.. ..:.     :    : :

MOUSE           LDSVGTHLPQERCQVASGHIPCMVNGSSKEACQQAGCCYDSTKEEPCYYGNTVTLQCFKS 275
RAT             LDSIGTHLLQERCQVASGHIPCMVKGSSEEACQQAGCCYDNTKEMPCYYGNTVTLQCFRS 273
HAMSTER         LDSVGTHLTQEQCQVASGHIPCMIKSSSKEACQQAGCCYDNTREVPCYYGNTATLQCSRS 283
RABBIT          PDAVGSHLPQEWCQVASGHIPCIVQSSSKEACEQAGCCYDSAREVPCYYGNTATVQCFRN 266
HUMAN           RDYIGTHLSQEQCQVASGHLPCIVRRTSKEACQQAGCCYDNTREVPCYYGNTATVQCFRD 283
CAT             PLHPGAPLTWEQCQVPSGHIPCVVRRGSKEACQKAGCCYDNSRAVPCYYGNTATVQCFRN 274
                    *: *  * ***.***:**::.  *:***::******.::  *******.*:** :.

MOUSE           GYFTLVMSQETALTHGVLLDNVHLAYAPNGCPPTQKTSAFVVFHVPLTLCGTAIQVVGEQ 335
RAT             GYFTLVMSQETALTHGVMLDNVHLAYAPNGCPPTQKTSAFVVFHVPLTLCGTAIQVVGKQ 333
HAMSTER         GYFTLAISQETALTHRVMLNNIHLAYAPSRCPPTQKTSAFVVFHVPLTLCGTTIQVVGEQ 343
RABBIT          GYFILVVAQEMALAHRITLANVHLAYAPTRCPPAQKTSAFVIFHVPLTHCGTTVQVLGSQ 326
HUMAN           GYFVLVVSQEMALTHRITLANIHLAYAPTSCSPTQHTEAFVVFYFPLTHCGTTMQVAGDQ 343
CAT             GHFVLVVSQETALAHGITLANIHVAYAPTSCSPTQDTGSFVVFQFPLTHCGTTVQVVGNQ 334
                *:* *.::** **:* : * *:*:****. *.*:*.* :**:* .*** ***::** *.*

MOUSE           LIYENQLVSDIDVQKGPQGSITRDSAFRLHVRCIFNASDFLPIQASIFSPQPPAPVTQSG 395
RAT             LVYENQLVSNIEVQTGPQGSITRDGVFRLHVRCIFNASDFLPIRASIFSPQPPAPVTRSG 393
HAMSTER         LIYENQLVSNIDVQKGPKGSITRDSVFRLHVRCIFNASDFLPVQASIFSPQPPAPVTQSG 403
RABBIT          LFYENQLVSDIDVREGPQGSITRDSSFRLLVRCIFNASDFLPIQASIFSPPLPAPVTQAG 386
HUMAN           LIYENWLVSGIHIQKGPQGSITRDSTFQLHVRCVFNASDFLPIQASIFPPPSPAPMTQPG 403
CAT             LLYENQLVSDIDVRMGPQGSITRDGAFRLHVRCTVNASDFLPLQASIFSPPSPVPVIQSG 394
                *.*** ***.*.:: **:******. *:* *** .*******::****.*  *.*: :.*
```

```
MOUSE          PLRLELRIATDKTFSSYYQGSDYPLVRLLREPVYVEVRLLQRTDPSLVLLLHQCWATPTT 455
RAT            PLRLELRIATDKTFSSYYQGSDYPLVRLLQEPVYIEVRLLQRTDPGLALMLHQCWATPSA 453
HAMSTER        PLRLELRIAKDKTFSSYYRERDYPLARLLQEPVHVEIRLLQRTDPGMVLMLHQCWATPTA 463
RABBIT         PLRLELRIARDETFSSFYEEEDYPLVRLLREPVHVEVRLLQRTDPSLVLELHQCWATPSA 446
HUMAN          PLRLELRIAKDETFSSYYGEDDYPIVRLLREPVHVEVRLLQRTDPNLVLLLHQCWGAPSA 463
CAT            PLRFQLRIATDETFRSFYEEGDYPIVRLLREPVSVEVRLLDRTDPGLVLLLHQCWATPGV 454
               ***::**** *:** *:*   ***:.***:*** :*:***:****.::.* *****.:* .

MOUSE          SPFEQPQWPILSDGCPFKGDNYRTQVVAADKEALPFWSHYQRFTITTFMLLDSSSQNALR 515
RAT            SPFEQPQWPILSDGCPFKGDNYRTQMVAADRATLPFWSHYQRFTIATFTLLDSSSQNALR 513
HAMSTER        NPFQQPQWPILSDGCPFEGDNYRTQMVALDRAELLFWSHYRRFTVTTFTLLDSSAGSTLR 523
RABBIT         NPVQQPQWPLLSDGCPFKGDSYRTRVLALDRAELPFRSHYQRFTVATFTFLDSGAQRALR 506
HUMAN          NPFQQPQWPILSDGCPFKGDSYRTQMVALDGAT-PFQSHYQRFTVATFALLDSGSQRALR 522
CAT            SPFQQPQWPILSEGCPFDGDSYRTRMVASDGAGLSFPSHHQRFTVTTFALLDPDSQRALR 514
               .*.:*****:**:****.**.***::* *     * **::***::** :**..:  :**

MOUSE          GQVYFFCSASACHPLGSDTCSTTCDSGIARRRRSSGHHNITLRALDIVSSPGAVGFEDAA 575
RAT            GQVYFFCSASACHPVGSETCSTTCDSEIARHRRSSGHHNSTIRALDIVSSPGAVGFEDAP 573
HAMSTER        GLVYFFCSASVCYPEGSETCSTVCDSGMARHRRSTGHHNSTVHALDIVSSPGAVGFEDAA 583
RABBIT         GLVYFFCSASACHPSGPETCSSTCSSRTAKRRRSSGYHDGTPRALDIVSSPGPVGFQDSH 566
HUMAN          GLVYLFCSTSACHTSGLETCSTACSTGTTRQRRSSGHRNDTARPQDIVSSPGPVGFEDSY 582
CAT            GQVYFFCHSSACSPSGLETCSTTCSSRPARQRRSYNPHGEATRPQNLVSSPGPVDFEDSS 574
               * **:** :*.* . * :***:.*.:  :::*** . :. : :. ::*****.*.*:*:

MOUSE          KLEP----SGSSRNSSSR---MLLLLLA-ITLALAAGIFVGLIWAWAQKLWEGIRY*--- 623
RAT            KLEP----SGSTRNSGSRPLLWVLQLLA-LTLVLGDGVLVGLSWAWAWA*---------- 617
HAMSTER        KLKP----SGSSRNSISRPLLWVLLLLL-VTTLVLMSL*--------------------- 616
RABBIT         RQEPTLESTGSGRNSNPKPLLWVVLLLLAIALVLGIGVFVGLSQAWAHKLREGHRLTDQA 626
HUMAN          GQEPTLGPTDSNGNSSLRPLLWAVLLLPAVALVLGFGVFVGLSQTWAQKLWESNRQ*--- 638
CAT            GQEPPLGPTGSPRNANQRPLLWVVLLLVAVALVLGVGVFEGLSQAWAQKLQEGDRGRLGS 634
                 :*    :.*  *:  :     : **  ::  :  .:
```

```
MOUSE           ------------------
RAT             ------------------
HAMSTER         ------------------
RABBIT          Q*----------------  627
HUMAN           ------------------
CAT             IKHRVQPAQRVWKAIRGD* 652
```

Figure 4. Comparison of amino acid sequences of ZP1 from mouse, rat, hamster, rabbit, human and cat. The deduced amino acid sequence of cat ZP1 was aligned with the ZP1 sequences of the other species using the Clustal W program. The accession number of the sequences used are as follows: mouse ZP1 (U20448), rat ZP1 (NM_053509), hamster ZP1 (EU003563), rabbit ZP1 (HQ702467), human ZP1 (NP_997224), and cat (XM_003993490). Identical aminoacids are marked by an asterisk. Colon (:) represents conserved residues and dot (.) represents semi-conserved residues. The signal peptide is in purple. The zona domain is shown in red. The trefoil domain is shown in blue. The consensus furin cleavage-site is underlined. The transmembrane domain is in orange. The cysteines are shown in green. The final of each protein is marked with *.

TABLES

Table I: primers used to amplify ZP genes.

Primers	Sequence	Position in sequence
ZP1		
Fw1	atggtggccgtggtgtatc	1
Fw2	actgctctgtctgctaccac	251
Rv3	acgtggcaacctctgtagtc	313
Rv1	atgggtggctccaggtc	572
Rv2	tgctcccaggtcagaggtg	659
Fw3	ctatgagaaccagctggtg	1008
Rv3	cactgaggctgctggaag	1368
Fw4	ctccagggccagtggact	1691
Rv4	ctagtcgccacgaatagct	1941
ZP2		
FwZP2	gagagtcctaaataggtctg	1631
RvZP2	ccttagtgatttatcatggttc	2182
ZP3		
FwZP3	ccaggccggagactctc	785
RvZP3	cttttattgggaagcagacac	1282
ZP4		
FwZP4	gctcctccataactgttgg	1232
RvZP4	gaataattcacctccgtttcc	1741
Primers	**Sequence**	**Position in sequence**
ZP1		
Fw1	atggtggccgtggtgtatc	1
Fw2	actgctctgtctgctaccac	251
Rv3	acgtggcaacctctgtagtc	313
Rv1	atgggtggctccaggtc	572
Rv2	tgctcccaggtcagaggtg	659
Fw3	ctatgagaaccagctggtg	1008
Rv3	cactgaggctgctggaag	1368
Fw4	ctccagggccagtggact	1691
Rv4	ctagtcgccacgaatagct	1941
ZP2		
FwZP2	gagagtcctaaataggtctg	1631
RvZP2	ccttagtgatttatcatggttc	2182
ZP3		
FwZP3	ccaggccggagactctc	785
RvZP3	cttttattgggaagcagacac	1282
ZP4		
FwZP4	gctcctccataactgttgg	1232
RvZP4	gaataattcacctccgtttcc	1741

Table II: Peptides identified by proteomics analysis. Peptides with a score higher than 5, and percentage-scored peak intensity of 60%, which are the threshold criteria for a positive identification, are shown in red. The peptides that have been detected in ZP from oocytes are in bold. n: number of times that the peptide has been detected.

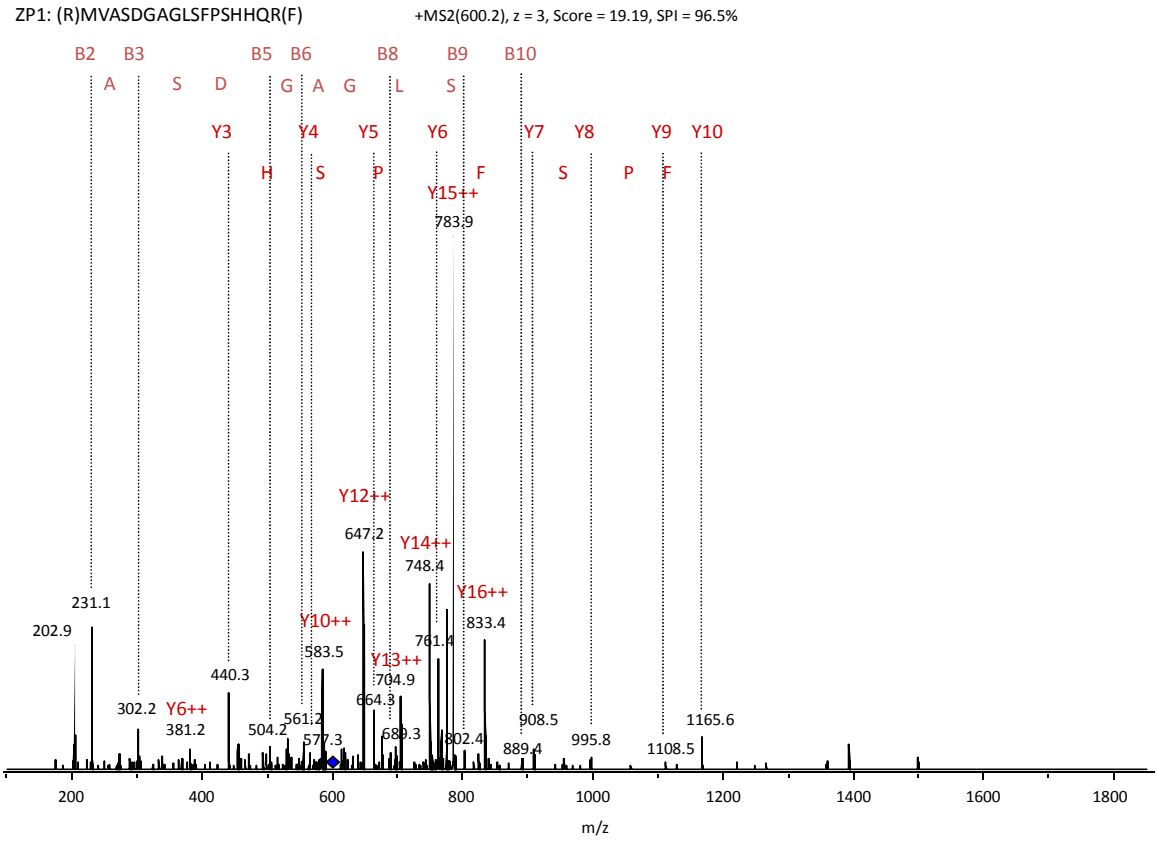
ZP1						
Peptides	z	m/z	Score	SPI	Sequence	n
GLRHSSDCGIK	2	665.77	4.36	62.5	37-47	2
GLRHSSDCGIKGMQLLVFPRPGQTVR	3	996.41	6.19	86.1	37-62	11
GLRHSSDCGIKGMQLLVFPRPGQTVRFK	3	1062.82	4.12	74.3	37-64	2
HSSDCGIKGMQLLVFPRPGQTVR	3	893.59	7.15	76.8	40-62	4
HSSDCGIKGMQLLVFPRPGQTVRFK	3	975.16	7.07	76.2	40-64	8
GCHVLEKGGRFHLRVFVEAVLRDGR	3	969.49	6.05	57.8	99-125	2
VDAAGEVTLICPKPGHTWTPESHLASR	3	999.60	11.15	82.7	126-152	3
RGSKEACQKAGCCYDNSR	3	742.19	3.85	87.3	233-250	2
GSKEACQKAGCCYDNSR	2	966.23	5.12	53.7	234-250	6
EACQKAGCCYDNSRAVPCYYGNTATVQCFR	3	1161.05	3.78	60.8	237-266	1
AGCCYDNSRAVPCYYGNTATVQCFR	3	978.90	5.77	62.9	242-266	5
AVPCYYGNTATVQCFR	2	955.06	9.37	72.6	251-266	4
FQLRIATDETFRSFYEEGDYPIVR	3	1038.53	3.91	67.4	391-414	1
IATDETFRSFYEEGDYPIVRLLR	3	984.08	3.87	60.3	395-417	2
IATDETFRSFYEEGDYPIVRLLREPVSVEVR	3	1129.83	5.40	91.4	395-425	1
SFYEEGDYPIVRLLREPVSVEVR	3	944.99	3.82	54.9	403-425	1
LLREPVSVEVR	2	649.91	11.20	85.0	415-425	1
MVASDGAGLSFPSHHQR	3	600.10	16.19	96.5	473-489	7
GQVYFFCHSSACSPSGLETCSTTCSSRPAR	3	1076.02	4.20	76.0	508-537	2
SSRRATEGDWAQ	2	763.29	7.90	75.7	616-627	3
ZP2						
Peptides	z	m/z	Score	SPI	Sequence	n
MASRQKGDSGSPSSWFNADWSTYR	3	967.08	8.37	89.2	1-24	6
SLFLLFILVTSVNSIGVLQLVNPVFPGTVTCYETRMA VEFPSDFGTKK	6	919.17	6.38	63.3	25-72	1
MAVEFPSDFGTKK	3	486.87	7.70	92.4	60-72	
WHTSVVDPFSFELLNCTYILDPENLTLKAPYETCTRR	5	913.54	5.99	69.3	73-109	1
APYETCTRRTLGQHRMIIRLK	3	925.80	6.15	52.1	101-121	7
MIIRLKDHNAASR	2	762.32	3.91	92.2	116-128	2
DSMSFTFNVIPGLADENTDIK	3	830.77	3.64	89.9	158-178	4
NPMGWSIEVGDGTKAK	2	893.59	4.33	70.3	179-194	1
NPMGWSIEVGDGTKAKTLTLQDVLR	3	956.60	6.68	56.1	179-203	4
TLTLQDVLRQGYNILFDNHK	3	836.72	3.07	77.6	193-214	1
ITFQVSFNATGVTHYMQGNSHLYMVPLK	3	1099.93	6.20	78.0	215-242	1
LIHESLGQKIILTTR	2	922.76	5.06	84.3	243-257	2
VLCMSDAVTCNATHVTLTIPEFPGK	3	952.97	8.90	56.9	258-282	6

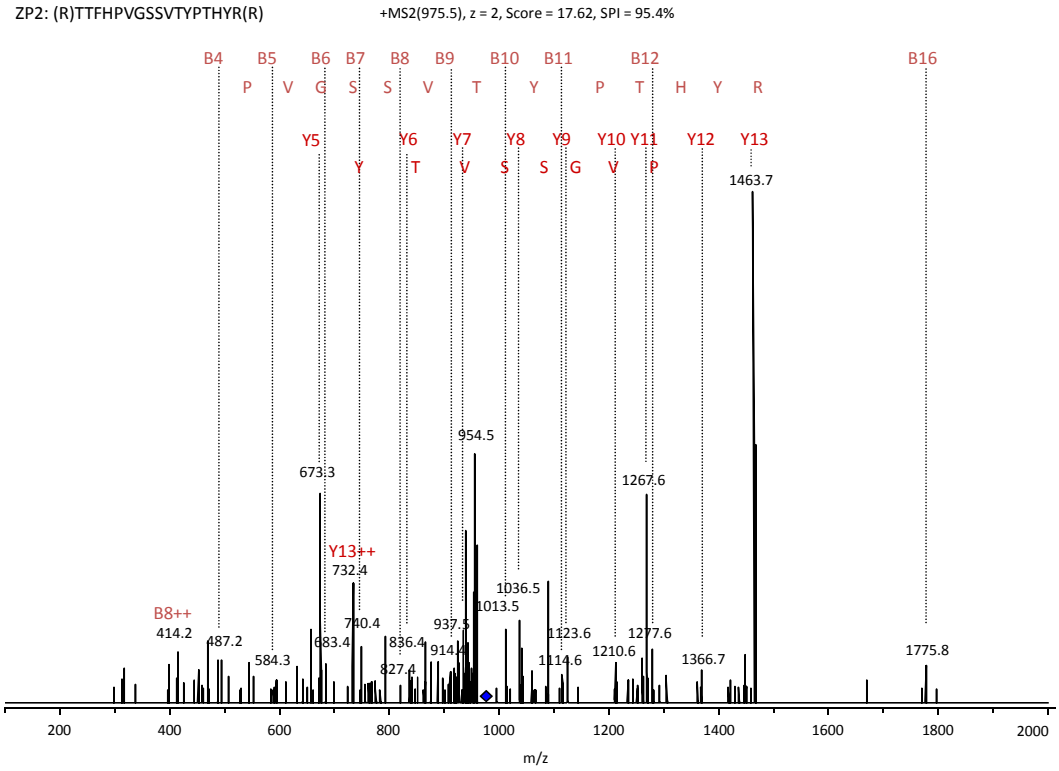
ZP2						
LKSVSSENRNFAVSQLHNNGIDK	3	893.18	3.39	51.8	283-305	1
EESSGLTLHFSK	3	472.68	3.37	55.3	306-317	1
EESSGLTLHFSKTLLKMEFSEK	3	887.75	4.14	75.4	306-327	1
TLLKMEFSEKCLPYQFYLASLK	3	956.34	4.71	53.0	318-339	3
MEFSEKCLPYQFYLASLK	3	783.66	3.42	64.0	322-339	1
VYSHQTKPALNLETLR	2	936.06	14.34	92.7	381-396	2
VQSLPPPEASVRPGPLALILQTYPDK	3	982.35	7.80	57.4	431-454	2
EGKVIYENEIHAVWADLPPSTISR	3	962.12	7.07	66.3	431-454	4
VIYENEIHAVWADLPPSTISR	2	1204.95	17.04	91.3	434-454	2
DSEFRMTVQCHYSK	2	983.36	3.13	54.6	455-468	1
DSEFRMTVQCHYSKGDLLINTR	3	949.86	4.61	64.5	455-476	2
GDLLINTR	2	451.98	7.88	86.6	469-476	2
QPIYLEVR	2	509.43	9.33	85.6	521-528	6
TTFHPVGSSVTYPTHYR	2	975.45	17.62	95.4	574-590	4
TTFHPVGSSVTYPTHYRR	3	755.62	4.26	63.1	574-591	1
LSADSPLCSVTCPVSSR	2	999.60	3.21	61.9	622-638	1
RATGTTEEEKMIVSLPGPILLLSDSSSLR	3	1087.94	3.88	54.4	641-669	1
ATGTTEEEKMIVSLPGPILLLSDSSSLRDVVDSK	3	1210.35	3.42	63.6	642-675	1
MIVSLPGPILLLSDSSSLRDVVDSK	3	934.84	4.78	54.0	651-675	1
DVVDSKGYGAAGYVAFK	2	953.90	3.79	56.7	670-686	1
TVVAVAALAGLVATLGFITYLRKNR	3	926.56	4.21	54.2	687-711	1
ZP3						
Peptides	**z**	**m/z**	**Score**	**SPI**	**Sequence**	**n**
AEVPIECR	1	974.80	5.21	83.7	132-139	2
HSNVSSEAILPTWVPFRTTTMLSEEK	3	979.75	7.25	51.4	143-167	4
TTMLSEEKLAFSLRLMEEDWGSEK	3	976.43	5.08	86.1	160-183	7
LAFSLR	2	354.31	8.20	97.2	168-173	3
LAFSLRLMEEDWGSEK	2	995.84	4.33	55.7	168-183	3
LMEEDWGSEK	2	611.83	3.90	52.2	174-183	1
QSPTFQLGDLAHLQAEVHTGR	2	1152.79	16.04	87.3	184-204	2
QSPTFQLGDLAHLQAEVHTGRHIPLR	3	968.80	4.83	50.9	184-209	1
APRPRPETLQFTVYTFHFANDPR	3	947.65	4.68	63.2	251-273	1
NMIYITCHLKVTPASRVPDQLNK	3	919.56	3.52	69.2	274-296	1
VTPASRVPDQLNKACSFIKSSNR	3	885.21	3.28	56.2	284-306	1
VPDQLNKACSFIKSSNR	2	1062.82	6.99	82.4	290-306	2
SSNRWFPVEGPADICNCCNKGSCGLQGR	3	999.98	3.68	50.8	303-330	1
WFPVEGPADICNCCNK	2	984.18	13.69	89.1	307-322	1
GSCGLQGRSWRLSHLDRPWHK	3	871.95	3.55	64.7	323-342	1
GSCGLQGRSWRLSHLDRPWHKMASR	3	998.02	3.17	56.5	323-347	1
SWRLSHLDRPWHKMASR	3	762.40	5.86	88.3	331-347	6
SWRLSHLDRPWHKMASRNR	3	838.87	3.17	65.4	331-349	1
SWRLSHLDRPWHKMASRNRR	3	904.00	7.58	73.4	331-350	2
MASRNRRHVTEEADITVGPLIFLGK	3	943.57	4.31	70.9	344-368	4
RHVTEEADITVGPLIFLGKAADR	3	837.44	4.17	59.1	350-372	2
HHTASRPMICPVSASQ	2	978.20	4.36	59.4	409-424	2

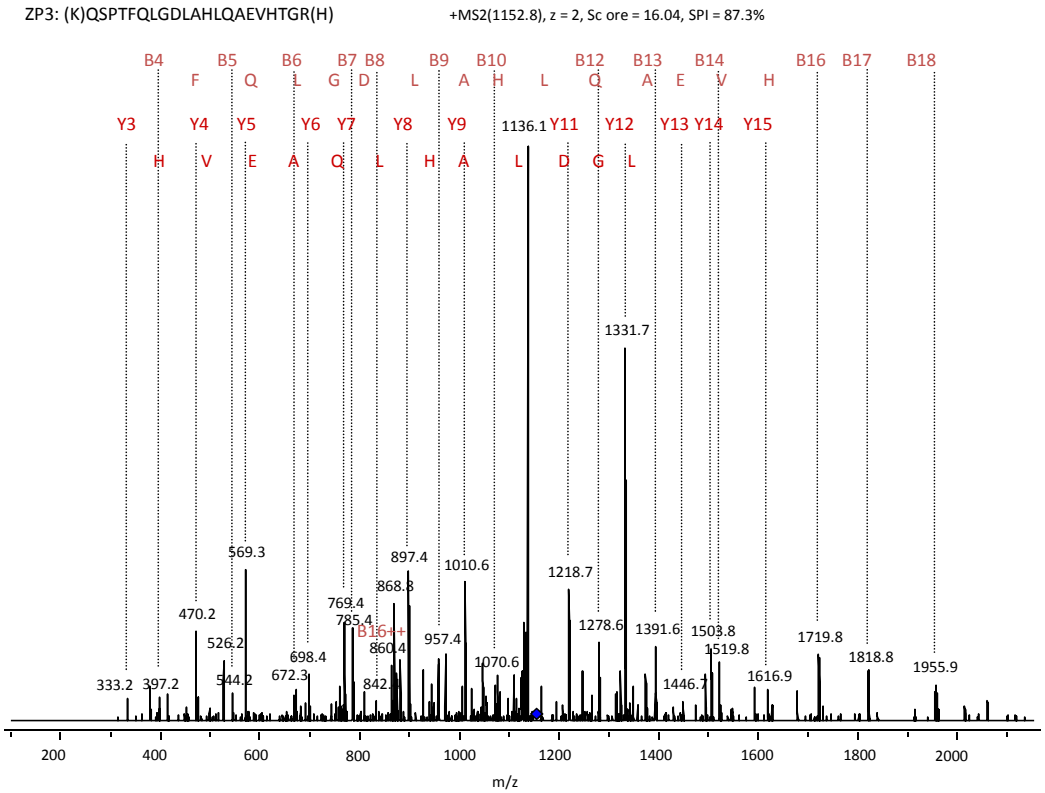
ZP4						
Peptides	z	m/z	Score	SPI	Sequence	n
ATPALIVWDNR	2	627.22	11.84	82.5	50-60	1
VTPQDSHYVMIVGVEGTDAAGRR	3	846.30	5.06	69.9	117-139	1
WDRLPCASSPITQGDCNKLGCCYK	3	919.79	7.01	77.3	183-208	2
WDRLPCASSPITQGDCNK	3	742.75	6.89	70.7	185-201	3
LGCCYKSEANSCYYGNTVTSR	3	830.06	3.68	77.9	203-223	1
CTQDGHFSIAVSR	3	493.63	9.91	72.4	224-236	1
CTQDGHFSIAVSRNVTSPPLLLNSLR	3	988.98	6.46	67.1	224-249	3
AFALFFFPFNSCGTTR	2	943.06	9.37	84.6	266-281	1
WVTGDQAVYENELVAAR	2	961.79	19.97	99.0	282-298	1
DVRTWSHGSITR	2	787.30	3.14	53.1	299-310	1
DSIFR	1	637.89	6.78	84.8	311-315	2
TQHGPLTLELKIAKDKHYGSYYTIGDYPVVK	3	1205.25	3.15	90.6	345-365	1
IAKDKHYGSYYTIGDYPVVKLLR	3	927.59	7.00	61.7	356-378	3
DKHYGSYYTIGDYPVVK	2	1003.68	16.68	86.4	359-375	1
HYGSYYTIGDYPVVK	2	942.45	3.07	70.2	361-375	1
ZP4						
LLRDPIYVEVSIR	3	525.70	14.34	84.8	376-388	1
DPIYVEVSIRHR	2	783.16	3.69	77.5	379-390	1
HRTDPSLGLLLHNCWATPGK	3	811.35	4.47	54.7	389-410	1
GCPYVGDNYQTQLIPVQK	2	1039.65	14.58	88.1	422-439	3
RFSIFTFSFVDTMAK	2	927.74	3.15	64.2	451-465	1
RFSIFTFSFVDTMAKWALR	3	833.85	3.20	60.7	451-469	1
RRHSDLHHHSSTASISSKGPMILLQATMDSAEK	4	911.76	7.83	86.8	499-531	1
RHSDLHHHSSTASISSKGPMILLQATMDSAEK	3	1190.92	6.45	58.5	500-531	1
GPMILLQATMDSAEK	2	810.15	3.47	65.9	515-531	1

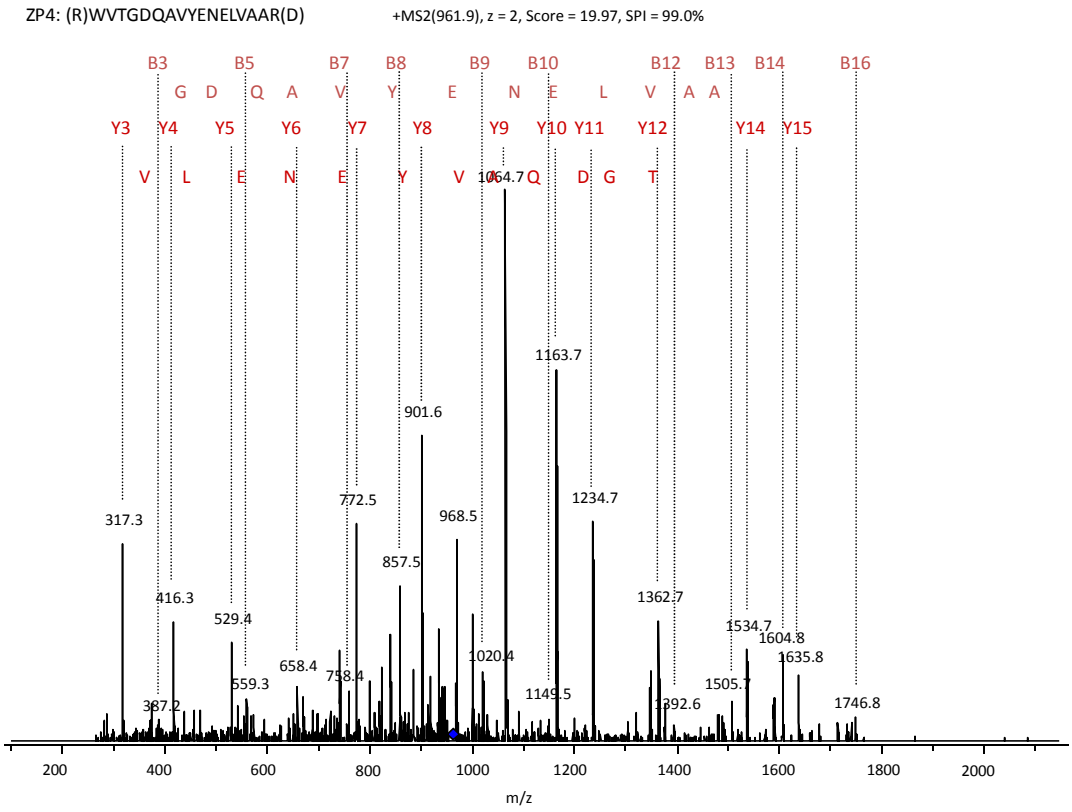
Supplementary file 1:

ESI-MS/MS spectra of ZP1 peptide 473-489 (MVASDGAGLSFPSHHQR), ZP2 peptide 574-590 (TTFHPVGSSVTYPTHYR), ZP3 peptide 184-204 (QSPTFQLGDLAHLQAEVHTGR) and ZP4 peptide 282-298 (WVTGDQAVYENELVAAR), showing the detected b- and y- ions from each sequence.


ZP1: (R)MVASDGAGLSFPSHHQR(F)
+MS2(600.2), z = 3, Score = 19.19, SPI = 96.5%
B2 B3 B5 B6 B8 B9 B10
A S D G A G L S
Y3 Y4 Y5 Y6 Y7 Y8 Y9 Y10
H S P F S P F
Y15++
783.9
Y12++
647.2
Y14++
748.4
Y16++
833.4
231.1
202.9
Y10++
583.5
761.4
Y13++
704.9
440.3
664.3
302.2
Y6++
381.2
561.2
504.2
577.3
689.3
802.4
908.5
889.4
995.8
1165.6
1108.5
200
400
600
800
1000
1200
1400
1600
1800
m/z


ZP2: (R)TTFHPVGSSVTYPTHYR(R)
+MS2(975.5), z = 2, Score = 17.62, SPI = 95.4%
B4 B5 B6 B7 B8 B9 B10 B11 B12 B16
P V G S S V T Y P T H Y R
Y5 Y6 Y7 Y8 Y9 Y10 Y11 Y12 Y13
Y T V S S G V P
1463.7
954.5
673.3
1267.6
Y13++
732.4
1036.5
1013.5
B8++
414.2
740.4
937.5
1123.6
1277.6
487.2
683.4
836.4
914.4
1210.6
584.3
827.4
1114.6
1366.7
1775.8
200 400 600 800 1000 1200 1400 1600 1800 2000
m/z


ZP3: (K)QSPTFQLGDLAHLQAEVHTGR(H)
+MS2(1152.8), z = 2, Sc ore = 16.04, SPI = 87.3%
B4 B5 B6 B7 B8 B9 B10 B12 B13 B14 B16 B17 B18
F Q L G D L A H L Q A E V H
Y3 Y4 Y5 Y6 Y7 Y8 Y9 1136.1 Y11 Y12 Y13 Y14 Y15
H V E A Q L H A L D G L
1331.7
569.3
897.4
1010.6
1218.7
868.8
769.4
785.4
B16++
470.2
526.2
860.4
957.4
1278.6
1391.6
1503.8
1519.8
1719.8
1818.8
698.4
1070.6
1955.9
672.3
1616.9
333.2
397.2
544.2
842
1446.7
200 400 600 800 1000 1200 1400 1600 1800 2000
m/z



Supplementary file 2:

Comparison of the different glycosylation sites of the four zona pellucida glycoproteins from different species. Letters in red colour indicates putative N-glycosylation site. Yellow colour indicates that N-glycosylation site has been confirmed by proteomic analyses. In green, N-glycosylation sites not occupied (confirmed by proteomic analyses). The O-glycosylation sites in blue indicate that this sites are occupied (demonstrated by proteomic analysis).

ZP1

```
human           -------MAGGSATTWG-YPVALLLLVATLGLGRWLQPDPGLPGLRHSYDCGIKGMQLLV 52
cat             MVAVVYLMTRASAMVWDGCVALLLLLVAALGLGQRPHPEPGLRGLRHSSDCGIKGMQLLV 60
rabbit          ------------MTGGRLVALLLLVAASLGLGQQPHPEPGLPGLQYSYDCGMRGMQLVV 47
mouse           ------------MAWGCFVVLLLLAAAPLRLGQRLHLEP---GFEYSYDCGVRGMQLLV 44
rat             ------------MAWGCFVVLLLLVAAPLRLGQHLHLKP---GFQYSYDCGVQGMQLLV 44
                             . .   . *** .*.* **:  : .*   *:.:* ***::****:*

human           FPRPGQTLRFKVVDEFGNRFDVNNCSICYHWVTSRPQEPAVFSADYRGCHVLEK-DGRFH 111
cat             FPRPGQTVRFKVVDEFGNQFEVHNCSVCYHWVTARPLGPAVFSADYRGCHVLEK-GGRFH 119
rabbit          LPRPGRTIRFKVVDEFGNRFEVNNCSICFHWVSAEPQAPAVFSADYKGCHVLEK-EGHSH 106
mouse           FPRPNQTVQFKVLDEFGNRFEVNNCSICYHWVTSEAQEHTVFSADYKGCHVLEK-DGRFH 103
rat             FPRPNQTIQFKVLDEFGNRFEVNNCSICYHWVISEAQKPAVFSADYKGCHVLEKQDGRFH 104
                :***.:*::***:*****:*:*:***:*:*** :..   :******:*******  *: *

human           LRVFMEAVLPNGRVDVAQDATLICPKPDPSRTLDSQLAPPAMFSVSTPQTLSFLPTSGHT 171
cat             LRVFVEAVLRDGRVDAAGEVTLICPKPGHTWTPESHLASRTGFSLPTPHTQPLRPTQ--- 176
rabbit          LTVFIEAMLPDGHVEVAQEAVLICPKPGHTWAVGSHQVPPTTPSPTTPHALPFHLSS--- 163
mouse           LRVFIQAVLPNGRVDIAQDVTLICPKPDHTVTPDPYLAPPTTPEPFTPHAFALHPIPDHT 163
rat             LRVFIQAVLPNGRVDTAQDVTLICPKPDHILTPESYLAPPTTPQPFIPHTFALHPISGHT 164
                * **::*:* :*:*: * :..******.   :  .  .. :  .   *:: .:

human           SQGSGHAFPSPLDPGHSSVHPTPALPSPGPGPTLATLAQPHWGTLEHWDVNKRDYIGTHL 231
cat             --------------EHSFTRPTPALLPLRPGATRPTLTLPQWDILEHWGVDEPLHPGAPL 222
rabbit          ----AHTFPIPLYLEHSLMLPTPAGPSLGPGPTPAVLAQ-----VERWEVDKPDAVGSHL 214
mouse           LAGSGHTGLTTLYPEQSFIHPTPAPPSLGPGPAGSTVPHSQWGTLEPWELTELDSVGTHL 223
rat             LAGSGHTGLTTLYPET---HPTPAPPSSEPGPVGPTVPQSQWGTLGSWELTELDSIGTHL 221
                            ****  .  **.. ..:.      :  * : :    *: *

human           SQEQCQVASGHLPCIVRRTSKEACQQAGCCYDNTREVPCYYGNTATVQCFRDGYFVLVVS 291
cat             TWEQCQVPSGHIPCVVRRGSKEACQKAGCCYDNSRAVPCYYGNTATVQCFRNGHFVLVVS 282
rabbit          PQEWCQVASGHIPCIVQSSSKEACEQAGCCYDSAREVPCYYGNTATVQCFRNGYFILVVA 274
mouse           PQERCQVASGHIPCMVNGSSKETCQQAGCCYDSTKEEPCYYGNTVTLQCFKSGYFTLVMS 283
```

```
rat             LQERCQVASGHIPCMVKGSSEEACQQAGCCYDNTKEMPCYYGNTVTLQCFRSGYFTLVMS 281
                  * ***.***:**:*.  *:*:*::******.::  *******.*:***:.*:* **::

human           QEMALTHRITLANIHLAYAPTSCSPTQHTEAFVVFYFPLTHCGTTMQVAGDQLIYENWLV 351
cat             QETALAHGITLANIHVAYAPTSCSPTQDTGSFVVFQFPLTHCGTTVQVVGNQLLYENQLV 342
rabbit          QEMALAHRITLANVHLAYAPTRCPPAQKTSAFVIFHVPLTHCGTTVQVLGSQLFYENQLV 334
mouse           QETALTHGVLLDNVHLAYAPNGCPPTQKTSAFVVFHVPLTLCGTAIQVVGEQLIYENQLV 343
rat             QETALTHGVMLDNVHLAYAPNGCPPTQKTSAFVVFHVPLTLCGTAIQVVGKQLVYENQLV 341
                ** **:* : * *:*:****. *.*:*.* :**:* .*** ***::** *.**.*** **

human           SGIHIQKGPQGSITRDSTFQLHVRCVFNASDFLPIQASIFPPPSPAPMTQPGPLRLELRI 411
cat             SDIDVRMGPQGSITRDGAFRLHVRCTVNASDFLPLQASIFSPPSPVPVIQSGPLRFQLRI 402
rabbit          SDIDVREGPQGSITRDSSFRLLVRCIFNASDFLPIQASIFSPPLPAPVTQAGPLRLELRI 394
mouse           SDIDVQKGPQGSITRDSAFRLHVRCIFNASDFLPIQASIFSPQPPAPVTQSGPLRLELRI 403
rat             SNIEVQTGPQGSITRDGVFRLHVRCIFNASDFLPIRASIFSPQPPAPVTRSGPLRLELRI 401
                *.*.:: *********. *:* *** .*******::****.*  *.*: :.****::***

human           AKDETFSSYYGEDDYPIVRLLREPVHVEVRLLQRTDPNLVLLLHQCWGAPSANPFQQPQW 471
cat             ATDETFRSFYEEGDYPIVRLLREPVSVEVRLLDRTDPGLVLLLHQCWATPGVSPFQQPQW 462
rabbit          ARDETFSSFYEEEDYPLVRLLREPVHVEVRLLQRTDPSLVLELHQCWATPSANPVQQPQW 454
mouse           ATDKTFSSYYQGSDYPLVRLLREPVYVEVRLLQRTDPSLVLVLHQCWATPTTSPFEQPQW 463
rat             ATDKTFSSYYQGSDYPLVRLLQEPVYIEVRLLQRTDPGLALMLHQCWATPSASPFEQPQW 461
                * *:** *:*   ***:****:*** :*****:****.*.* *****.:* ..*.:****

human           PILSDGCPFKGDSYRTQMVALDGA-TPFQSHYQRFTVATFALLDSGSQRALRGLVYLFCS 530
cat             PILSEGCPFDGDSYRTRMVASDGAGLSFPSHHQRFTVTTFALLDPDSQRALRGQVYFFCH 522
rabbit          PLLSDGCPFKGDSYRTRVLALDRAELPFRSHYQRFTVATFTFLDSGAQRALRGLVYFFCS 514
mouse           PILSDGCPFKGDNYRTQVVAADREALPFWSHYQRFTITTFMLLDSSSQNALRGQVYFFCS 523
rat             PILSDGCPFKGDNYRTQMVAADRATLPFWSHYQRFTIATFTLLDSSSQNALRGQVYFFCS 521
                *:**:****.**.***:::* *    .* **:****::** :**..:*.**** **:**

human           TSACHTSGLETCSTACSTGTTRQRRSSGHRNDTARPQDIVSSPGPVGFEDSYGQEPTLGP 590
cat             SSACSPSGLETCSTTCSSRPARQRRSYNPHGEATRPQNLVSSPGPVDFEDSSGQEPPLGP 582
```

```
rabbit          ASACHPSGPETCSSTCSSRTAKRRRSSGYHDGTPRALDIVSSPGPVGFQDSHRQEPTLES 574
mouse           ASACHPLGSDTCSTTCDSGIARRRRSSGHHNITLRALDIVSSPGAVGFEDAAKLEP---- 579
rat             ASACHPVGSETCSTTCDSEIARHRRSSGHHNSTIRALDIVSSPGAVGFEDAPKLEP---- 577
                :*** . * :***::*.:  :::*** . :. : *. ::*****.*.*:*:   **

human           TDSNGNSSLRPLLWAVLLLPAVALVLGFGVFVGLSQTWAQKLWESNRQ------------ 638
cat             TGSPRNANQRPLLWVVLLLVAVALVLGVGVFEGLSQAXAQKLQEGDRGRLGSIKHRVQPA 642
rabbit          TGSGRNSNPKPLLWVVLLLLAIALVLGIGVFVGLSQAWAHKLREGHR-----LTDQAQ-- 627
mouse           SGSSRNSSSR----MLLLLLAITLALAAGIFVGLIWAWAQKLWEGIRY------------ 623
rat             SGSTRNSGSRPLL-WVLQLLALTLVLGDGVLVGLSWAWAWA------------------- 617
                :.*  *:. :     :* * *::*.*. *:: **  : *

human           ----------
cat             QRVWKAIRGD 652
rabbit          ----------
mouse           ----------
rat             ----------
```

ZP2

```
cow             ------MACRQRGDSGRPSSWFRADW--------------RSFFLSFTLLTSVNSIDVN 39
pig             ------MACRHRGDSGRPLSWLSASW--------------RSLLLFFPLVTSVNSIGVN 39
cat             ------MASRQKGDSGSPSSWFNADWSTY-----------RSLFLLFILVTSVNSIGVL 42
human           ------MACRQRGGSWSPSGWFNAGWSTY-----------RSISLFFALVTSGNSIDVS 42
rabbit          ------MQVSNSGSRGKRLPWPSLTKFTFPYLSPPSSSSACTWLFLFFTLVTSVNSIYFL 54
mouse           ------MARWQRKASVSSPCGRSIYRFLS----------------LLFTLVTSVNSVSLP 38
rat             ------MARWQR-----------VYWLRS----------------LFFALVTSVNSLSLP 27
                      *   :                                  * * *:** **: .

cow              QLD-PAFPGTVTCYENRMVVEFKRTLGNKIQHASVVDSLGLKMLNCTYVLDPEKLTLKAP 98
pig              QLVNTAFPGIVTCHENRMVVEFPRILGTKIQYTSVVDPLGLEMMNCTYVLDPENLTLKAP 99
cat             QLVNPVFPGTVTCYETRMAVEFPSDFGTKKWHTSVVDPFSFELLNCTYILDPENLTLKAP 102
human           QLVNPAFPGTVTCDEREITVEFPSSPGTKKWHASVVDPLGLDMPNCTYILDPEKLTLRAT 102
rabbit          QLSDPAFPGTVTCNENEIMVEFPSYVGTKTLHASVVDPLGVEMLNCTYILDPEKLTLRVP 114
mouse            QSENPAFPGTLICDKDEVRIEFSSRFDMEKWNPSVVDTLGSEILNCTYALDLERFVLKFP 98
rat              QSENPAFPGTLICDKDEVRVEFSSRFDMEKWNPSLVDTFGNEISNCTYALDLEKFILKFP 87
                 *   ..*** : * : .: :**    . :   .*:**.:. .: **** ** *.: *: .

cow              YESCTKRVLGQHQMTITFMNDNTAHRQKTVLYHVSCPVMQAGRHDQHSGSTICSKDFMSF 158
pig              YEACTKRVRGHHQMTIRLIDDNAALRQEALMYHISCPVMGAEGPDQHSGSTICMKDFMSF 159
cat              YETCTRRTLGQHRMIIRLKDHNAASRHNSLMYQINCPVMQAEETHEHAGSTICTKDSMSF 162
human            YDNCTRRVHGGHQMTIRVMNNSAALRHGAVMYQFFCPAMQVEETQGLSASTICQKDFMSF 162
rabbit           YKACTRAVHGGHQMSIRVMNNSAALRHTDVEYQFFCPVEQT---LEFSKSAACTKDFMSL 171
mouse            YETCTIKVVGGYQVNIRVGDTTTDVRYKDDMYHFFCPAIQA-ETHEISEIVVCRRDLISF 157
rat              YETCTIKVIGGYQVNIRVQDTNADVSYKDDVHHFFCPAIQA-EIHEVSEIVVCMEDLISF 146
                 *. **  . * ::: * . : .:        ::. **.  .      :  . * .* :*:

cow              TFH-FFPGLADDTAG---PKPQMGWTVTVGDGERAQNLTLQEALTQGYNLLIENQKMSIQ 214
pig              TFN-FFPGMADENVKREDSKQRMGWSLVVGDGERARTLTFQEAMTQGYNFLIENQKMNIQ 218
cat              TFN-VIPGLADENTD---IKNPMGWSIEVGDGTKAKTLTLQDVLRQGYNILFDNHKITFQ 218
human            SLPRVFSGLADDSKG---TKVQMGWSIEVGDGARAKTLTLPEAMKEGFSLLIDNHRMTFH 219
rabbit           SFPRIPTGLGDSTMVN---ESQMGWMVQAGHGPGAQTLSLEEAKGQGFGVLIDDNKMTLS 228
```

```
mouse           SFPQLFSRLADENQN----VSEMGWIVKIGNGTRAHILPLKDAIVQGFNLLIDSQKVTLH 213
rat             SFPQLFSRLADENQN----VSEMGWIIKIGNGTRVHTLPLKDAIVQGFNLLIDSQKITLH 202
                ::  . . :.*..        *** :  *.*  .: *.: :.  :*:..*::.:::.:

cow             VLFHATGVTHYSQGNSHLYMVPLKLTHVSPGQTIILSSRLICASDP-VTCNATHMTLTIP 273
pig             VSFHATGVTRYSQGNSHLYMVPLKLKHVSHGQSLILASQLICVADP-VTCNATHVTLAIP 277
cat             VSFNATGVTHYMQGNSHLYMVPLKLIHESLGQKIILTTRVLCMSDA-VTCNATHVTLTIP 277
human           VPFNATGVTHYVQGNSHLYMVSLKLTFISPGQKVIFSSQAICAPDP-VTCNATHMTLTIP 278
rabbit          VLLNATGVTHYVEGTSHLHTMFLKLSLVSPGQKMTFPSRAICLSGP-VTCNATHMTLTIP 287
mouse           VPANATGIVHYVQESSYLYTVQLELLFSTTGQKIVFSSHAICAPDLSVACNATHMTLTIP 273
rat             VPANATGVAHYVQESSYLYTVQLKLLFSSPGQKITFSSQAICAPDLSVACNVTHMSLTIP 262
                *  :***:.:* : .*:*: : *:*   : **.: :.:: :* ..  *:**.**::*:**

cow             EFPGKLKSVSFENKNIAVNQLHNSGIVMEIANGLRLHFSKTLLKTKFSEKCLPYQFYLSS 333
pig             EFPGKLKSVNLGSGNIAVSQLHKHGIEMETTNGLRLHFNQTLLKTNVSEKCQPHQLYLSS 337
cat             EFPGKLKSVSSENRNFAVSQLHNNGIDKEESSGLTLHFSKTLLKMEFSEKCLPYQFYLAS 337
human           EFPGKLKSVSFENQNIDVSQLHDNGIDLEATNGMKLHFSKTLLKTKLSEKCLLHQFYLAS 338
rabbit          EFPGKLESVSIENRNITVSQLHDQGIDVEAINGLRLHFSKTVLKTKFSEKCLHDQLYISS 347
mouse           EFPGKLESVDFGQWSIPEDQWHANGIDKEATNGLRLNFRKSLLKTKPSEKCPFYQFYLSS 333
rat             EFPGKLKSVGFGQRNIPEDQWHANGIDKEATNGLRLHFRKSLLKTKPSEKCPFYQFYFSS 322
                ******:**.  . .:  .* *  **  *  .*: *:* :::** : ****   *:*::*

cow             LKLTFYTQLETVSMVIYPECVCESTVSIVSGELCTQDGFMDVEVYRHQTKPALNLDTLRV 393
pig             LRLTFHSQLEAVSMVIYPECLCESTVSLVSEELCTQDGFMDVKVHSHQTKPALNLDTLRV 397
cat             LKLTFAFNQETISTVLYPECVCESPVSIVTGDLCTQDGFMDIKVYSHQTKPALNLETLRG 397
human           LKLTFLLRPETVSMVIYPECLCESPVSIVTGELCTQDGFMDVEVYSYQTQPALDLGTLRV 398
rabbit          LKLTFNLELDTVSTVINPECPCDSPASIVSGELCTQDGFMDFEVYTHQTKPALNLDTLRV 407
mouse           LKLTFYFQGNMLSTVIDPECHCESPVSID--ELCAQDGFMDFEVYSHQTKPALNLDTLLV 391
rat             LELTFNFQGDMLSTVIDPECHCESPVSID--ELCTRDGFMDFEVYSHQTKPALNLESLLV 380
                *.***  . : :* *: *** *:*..*:   :**::*****.:*: :**:***:* :*

cow             GDSSCQPTIKAPFQGLVKFHIPLNGCGTRHKFENGKVIYENEIHALWADLPPSTISRDSE 453
pig             GDSSCQPTFKAPAQGLVQFRIPLNGCGTRHKFKNDKVIYENEIHALWAD-PPSAVSRDSE 456
```

```
cat             GDSSCQPTFQAASQGLILFHIPLNGCGTRHKFKEGKVIYENEIHAVWADLPPSTISRDSE 457
human           GNSSCQPVFEAQSQGLVRFHIPLNGCGTRYKFEDDKVVYENEIHALWTDFPPSKISRDSE 458
rabbit          GSSSCQPVFKAQSQGLVRFRIPLNGCGTRHKFEDEKVIYENEVHALWENLPPSKISRDSE 467
mouse           GNSSCQPIFKVQSVGLARFHIPLNGCGTRQKFEGDKVIYENEIHALWENPPSNIVFRNSE 451
rat             GNSSCQPIFKVQSLGLARFHIPLNGCGTRQKFEGDKVIYENEIHALWENPPSNIIFRNSE 440
                *.***** ::.   **  *:********* **:  **:****:**:* : *.. : *:**

cow             FRMTVRCYYSSSNMLINTNVESLPPPVASVKPGPLALTLQTYPDNSYLQPYGDKDYPVVR 513
pig             FRMTVRCSYSSSNMLINTNVESLPSPEASVKPGPLTLTLQTYPDNAYLQPYGDKEYPVVK 516
cat             FRMTVQCHYSKGDLLINTRVQSLPPPEASVRPGPLALILQTYPDKSYLQPYGEKEYPVVR 517
human           FRMTVKCSYSRNDMLLNINVESLTPPVASVKLGPFTLILQSYPDNSYQQPYGENEYPLVR 518
rabbit          FRMTVQCYYTRDDMLLNANIKSLPPPVASVKPGPLALSLQTYPDESYQQPYRVNEYPIVK 527
mouse           FRMTVRCYYIRDSMLLNAHVKGHPSPEAFVKPGPLVLVLQTYPDQSYQRPYRKDEYPLVR 511
rat             FRMTVRCHYIRDSMLLRAHIKSHSSPVASVKPGPLALVLQTYPDISYQRPYRKNEYPLVR 500
                *****:* *  ..:*:. .::. ..* * *: **:.* **:*** :* :**  .:**:*:

cow             YLRQPIYLEVRVLNRTDPNIKLVLDDCWATSTMDPASLPQWNIIVDGCEYNLDNHRTTFH 573
pig             YLRQPIYLEVRILNRTDPNIKLVLDDCWATSTEDPASLPQWNVVMDGCEYNLDNHRTTFH 576
cat             YLRQPIYLEVRVLNRSDPNIKLVLDDCWATPTMDPASVPQWNIIMDGCEYNLDNHRTTFH 577
human           FLRQPIYMEVRVLNRDDPNIKLVLDDCWATSTMDPDSFPQWNVVVDGCAYDLDNYQTTFH 578
rabbit          YLRQPIYMEVRVLNRNDPNIKLALDDCWATSSMDPASLPKWSIVMDGCEYSLDNYQTNFH 587
mouse           YLRQPIYMEVKVLSRNDPNIKLVLDDCWATSSEDPASAPQWQIVMDGCEYELDNYRTTFH 571
rat             YLRQPIYMEVTVLNRNDPNIKLVLDDCWATTFEDPASVPQWQIIMDGCEYELDNYRTTFH 560
                :******:** :*.* ******.*******.  ** * *:*.:::*** *.***::*.**

cow             PVGSSVAYPNHYQRFAVKTFAFVSEDPAFSHLVYFHCSALICDQLSSNFPLCSASCLVSS 633
pig             PVGSSVTYPNHHQRFDVKTFAFVSGAQGVSQLVYFHCSVFICNQLSPTFSLCSVTCHGPS 636
cat             PVGSSVTYPTHYRRFDVKTFAFVSEAQVLSSLVYFHCSVLICSRLSADSPLCSVTCPVSS 637
human           PVGSSVTHPDHYQRFDMKAFAFVSEAHVLSSLVYFHCSALICNRLSPDSPLCSVTCPVSS 638
rabbit          PVGSSVTYPEHYQRFDVKTFAFVSEAQARSSLVYFHCSALICNQHYPDSPLCSVTCPGSS 647
mouse           PAGSSAAHSGHYQRFDVKTFAFVSEARGLSSLIYFHCSALICNQVSLDSPLCSVTCPASL 631
rat             AANSSAAHSGHYQRFDVKTFAFVSESRGLSSLIYFHCSALICNQA---SPLCSVTCPAPL 617
                ...**.::. *::** :*:*****     * *:*****.:**.:     .***.:*  .
```

```
cow             RSRRATGATEEEKMIVSLPGPILLLSDGSSFR---------------------------- 665
pig             RSRRATGTTEEEKMIVSLPGPILLLSDGSSLR---------------------------- 668
cat             RHRRATGTTEEEKMIVSLPGPILLLSDSSSLR---------------------------- 669
human           RHRRATGATEAEKMTVSLPGPILLLSDDSSFRGVGSSDLKASGSSGEKSRSETGEEVGSR 698
rabbit          RHRRATGNTEEERVTASLPGPILLLPNGSSFR---------------------------- 679
mouse           RSKRE--ANKEDTMTVSLPGPILLLSDVSSSKGVD-----------------------P 665
rat             RNKRE--ASKEGTMTVSLPGPIILLSDDSSSKGVMN----------------------P 652
                * :*    .:   : .******:**.: ** :

cow             DAVDSKGHGTSGYAAFKTMVAVVALAGVVATLSLISYLRKKRITVLNH 713
pig             DAVNSKGSRTNGYVAFKTMVAMVASAGIVATLGLISYLHKKRIMMLNH 716
cat             DVVDSKGYGAAGYVAFKTVVAVAALAGLVATLGFITYLRKNR-TMINH 716
human           GAMDTKGHKTAGDVGSKAVAAVAAFAGVVATLGFIYYLYEKR-TVSNH 745
rabbit          GVGDSKEHGMAGDVTSKTMAAVAAVAGVVATLGFISYLCKKRTMMLSH 727
mouse           SSSEITKDIIAKDIASKTLGAVAALVGSAVILGFICYLYKKRTIRFNH 713
rat             DSYEIT-----KDIASKTLGAVAALVGSAVIIGFICYLHKKRIVRFNS 695
                .  : .          *:: *:.* .* .. :.:* ** ::*    .
```

ZP3

```
cat             MGLSYGLFICFLLWAGTGLCYPPTTTEDKTHPSLPS--SPSVVVECRHAWLVVNVSKNLF 58
pig             MAPSWRFFVCFLLWGGTELCSPQPVWQDEGQRLRPSK-PPTVMVECQEAQLVVIVSKDLF 59
human           MELSYRLFICLLLWGSTELCYPQPLWLLQGGASHPETSVQPVLVECQEATLMVMVSKDLF 60
rabbit          MGLSYGLFVCLLLWGGSELCCPQPLWFWQGGTRQPAPSVTPVVVECLEARLVVTVSRDLF 60
mouse           MASSYFLFLCLLLCGGPELCNSQTLWLLPGGTPTPVGSSSPVKVECLEAELVVTVSRDLF 60
rat             MGPSCLLFLCLLLCGGPELCYPQTQWLLPGGTPTPAGSSSPVEVECKEAELVVTARRDLF 60
                *  *  :*:*:** ... ** . .          *     .* *** .* *:* . ::**

cat             GTGRLVRPADLTLWPENCEPLISGDSDDTVRFEVELHKCGNSVQVTEDALVYSTFLLHNP 118
pig             GTGKLIRPADLSLGPAKCEPLVSQDTDAVVRFEVGLHECGSSLQVTDDALVYSTFLRHDP 119
human           GTGKLIRAADLTLGPEACEPLVSMDTEDVVRFEVGLHECGNSMQVTDDALVYSTFLLHDP 120
rabbit          GTGKLIQEADLSLGPEGCEPQAS--TDAVVRFEVGLHECGNSVQVTDDSLVYSSFLLHDP 118
mouse           GTGKLVQPGDLTLGSEGCQPRVSVDTD-VVRFNAQLHECSSRVQMTKDALVYSTFLLHDP 119
```

```
rat             GTGKLVQPGDLTLGSEGCQPLVAVDTD-VVRLNAQLHECSSGVQVTEDALVYSTFLLHDP 119
                ***:*:: .**:* .  *:*  :  :: .**::. **:*.. :*:*.*:****:** *:*

cat             RPMGNLSILRTNRAEVPIECRYPRHSNVSSEAILPTWVPFRTTMLSEEKLAFSLRLMEED 178
pig             RPAGNLSILRTNRAEVPIECHYPRQGNVSSWAILPTWVPFRTTVFSEEKLVFSLRLMEEN 179
human           RPVGNLSIVRTNRAEIPIECRYPRQGNVSSQAILPTWLPFRTTVFSEEKLTFSLRLMEEN 180
rabbit          RPAGNLSILRTNRAEVPIECRYPRQGNVSSRAILPTWVPFWTTVLSEERLVFSLRLMEEN 178
mouse           RPVSGLSILRTNRVEVPIECRYPRQGNVSSHPIQPTWVPFRATVSSEEKLAFSLRLMEEN 179
rat             RPVNGLSILRTNRVEVPIECRYPRQGNVSSHPIQPTWVPFSATVSSEEKLAFSLRLMEED 179
                ** ..***:****.*:****:***:.**** .* ***:** :*: ***:*.********:

cat             WGSEKQSPTFQLGDLAHLQAEVHTGRHIPLRLFVDYCVATLT--PDQNASPHHTIVDFHG 236
pig             WSAEKMTPTFQLGDRAHLQAQVHTGSHVPLRLFVDHCVATLT--PDWNTSPSHTIVDFHG 237
human           WNAEKRSPTFHLGDAAHLQAEIHTGSHVPLRLFVDHCVATPT--PDQNASPYHTIVDFHG 238
rabbit          WSREKMSPTFHLGDTAHLQAEVRTGSHPPLLLFVDRCVATPT--RDQSGSPYHTIVDLHG 236
mouse           WNTEKSAPTFHLGEVAHLQAEVQTGSHLPLQLFVDHCVATPSPLPDPNSSPYHFIVDFHG 239
rat             WNTEKSSPTFHLGEVAHLQAEVQTGSHLPLQLFVDHCVATPSPLPGQNSSPHHFIVDSHG 239
                *. ** :***:**: *****:::** * ** **** **** :   . . ** * *** **

cat             CLVDGLSDASSAFKAPRPRPETLQFTVYTFHFANDPRNMIYITCHLKVTPASRVPDQLNK 296
pig             CLVDGLTEASSAFKAPRPGPETLQFTVDVFHFANDSRNTIYITCHLKVTPADRVPDQLNK 297
human           CLVDGLTDASSAFKVPRPGPDTLQFTVDVFHFANDSRNMIYITCHLKVTLAEQDPDELNK 298
rabbit          CLVDGLSDGASKFKAPRPKPDVLQFMVAVFHFANDSRHTVYITCHLRVIPAQQAPDRLNK 296
mouse           CLVDGLSESFSAFQVPRPRPETLQFTVDVFHFANSSRNTLYITCHLKVAPANQIPDKLNK 299
rat             CLVDGLSESFSAFQVPRPRPETLQFTVDVFHFANSSRNTVYITCHLKVAPANQIPDKLNK 299
                ******::. * *:.*** *:.*** * .*****..*: :******:*  *.: **.***

cat             ACSFIKSSNRWFPVEGPADICNCCNKGSCGLQGRSWRLSHLDRPWHKMASRNRRHVTEEA 356
pig             ACSFSKSSNRWSPVEGPAVICRCCHKGQCGTPSLSRKLSMPKR---QSAPRSRRHVTDEA 354
human           ACSFSKPSNSWFPVEGPADICQCCNKGDCGTPSHSRRQPHVMSQWSTSASRNRRHVTEEA 358
rabbit          ACSFNQSSSSWAPVEGSADICECCGNGDCDLIAGS---PMNQNHAARSSLRSRRHVTEEA 353
mouse           ACSFNKTSQSWLPVEGDADICDCCSHGNCSNSSSSQFQIHGPRQWSKLVSRNRRHVTDEA 359
rat             ACSFNKTSQSWLPVEGDADICDCCSNGNCSNSSSSEFETHEPAQWSTLVSRNRRHVTDEA 359
```

```
                **** :.*. * **** * ** ** :*.*.  . *              *.*****:**

cat             DITVGPLIFLGKAAD-RGVEG-STSPHTSVMVGIGLATVLSLTLATIVLGLARRHHTASR 414
pig             DVTVGPLIFLGKTSD-HGVEG-STSSPTSVMVGLGLATVVTLTLATIVLGVPRRRRAAAH 412
human           DVTVGATDLPGQEW------------------------------------------- 372
rabbit          DVTVGPLIFLGKAGDPAGTEGLASAAQATLVLGFGMATIVFLAVAAVVLGLTRGRHAASH 413
mouse           DVTVGPLIFLGKAND-QTVEGWTASAQTSVALGLGLATVAFLTLAAIVLAVTRKCHSSSY 418
rat             DVTVGPLIFLGKAND-QAVEGWTSSAQTSVALGLGLATVAFLTLAAIVLGVTRMCHTSSY 418
                *:***.  : *:

cat             PMICPVSASQ 424
pig             -LVCPVSASQ 421
human           ----------
rabbit          ----PRSASQ 419
mouse           ----LVSLPQ 424
rat             ----LVSLPQ 424
```

ZP4

```
pig             ---------MWLRPSIWLCFPLCLALPGQSQPKAADDLGGLYCGPSSFHFSINLLSQDTA 51
cat             ---------MWLLQPLLLCVPLSLAVHGQQKPQVPDYPGELHCGLQSLQFAINP-SPGKA 50
rabbit          ---MAPGSTMWLLGYIFLCFPVSFALIKQPKPETPTDPGVLHCRPWNFKFTINFQNQETG 57
human           ---------MWLLRCVLLCVSLSLAVSGQHKPEAPDYSSVLHCGPWSFQFAVNL-NQEAT 50
rat             MARQALRSTLWLLPSILLCFPFCLPLSGQHVTELP---GVLHCGLQSFQFAVNL-SLEAE 56
                         :**   : **....:.:  *  .: .   . *:*   .::*::*  .

pig             TPPALVVWDRRGRLHKLQNDSGCGTWVHKGPGSSMGVEASYRGCYVTEW---------- 100
cat             TP-ALIVWDNRGLPHKLQNNSGCGTWVRESPGGSVLLDASYSSCYVNEWVSTTQSPGTSR 109
rabbit          SSPVLVTWDNQGRLHRLQNDTDCGTRVGEGPGPSVVLEANYSSCYVTES---------- 106
human           SPPVLIAWDNQGLLHELQNDSDCGTWIRKGPGSSVVLEATYSSCYVTEW---------- 99
rat             SP-VLTTWDSQGLPHRLKNDSDCGTWVMDSPDGFLVLEASYSGCYVTLEG--------- 105
                :. .* .** :*  *.*:*::.*** : ..*.  : ::*.* .***.
```

```
pig             -----------DSHYLMPIGLEEADAGGHRTVTETKLFKCP----------------VDF 133
cat             PPTPASRVTPQDSHYVMIVGVEGTDAAGRR-VTNTKVLRCPRNPPDQALVSSLSPSPLQN 168
rabbit          -----------EPYYVMLVGVEEVDAAGQNLVTKQQLLKCP----------------MHL 139
human           -----------DSHYIMPVGVEGAGAAEHKVVTERKLLKCP----------------MDL 132
rat             ------------SHYIMTVGVQEADVAGHVAGTRQRLLTCP--------------LALQG 139
                            .:*:* :*:: .... :   *. ::: **               :.

pig             LALDVPTIGLCDAVPVWDRLPCAPPPITQGECKQLGCCYNSE--EVPSCYYGNTVTSRCT 191
cat             VALEAPNADLCDSVPKWDRLPCASSPITQGDCNKLGCCYKS---EANSCYYGNTVTSRCT 225
rabbit          PA---PDAGLCDSVPVQDRLPCATAPISQEDCEELGCCHSSE--EVNACYYGNTVTSHCT 194
human           LARDAPDTDWCDSIPARDRLPCAPSPISRGDCEGLGCCYSSE--EVNSCYYGNTVTLHCT 190
rat             KAPDTPSAKVCSPVPVKERLPCASSTISRGDCEELGCCYSSEEEGADSCYYGNTVTSHCT 199
                 *   *    *..:*  :*****...*:: :*: ****:.*    . :******** :**

pig             QDGHFSIAVSRNVTSPPLLWDSVHLAFRNDSE-CKPVMETHTFVLFRFPFSSCGTAKRVT 250
cat             QDGHFSIAVSRNVTSPPLLLNSLRLAFGKDRE-CNPVKATRAFALFFFPFNSCGTTRWVT 284
rabbit          QEGHFSIAVSRNVSSPPLHLDSVHLVFGNDSE-CQPVVATRAFVLFLFPFTACGTTRQIT 253
human           REGHFSIAVSRNVTSPPLLLDSVRLALRNDSA-CNPVMATQAFVLFQFPFTSCGTTRQIT 249
rat             KEGHFSIAVSRDVTSPPLRLDSLRLGFRNITTGCDPVMKTSTFVLFQFPLTSCGTTQRIT 259
                ::*********:*:****  :*::* : :    *.**  * :*.** **:.:***:: :*

pig             GNQAVYENELVAARDVRTWSHGSITRDSIFRLRVSCIYSVSSSALPVNIQVFTLPPPLPE 310
cat             GDQAVYENELVAARDVRTWSHGSITRDSIFRLRVSCSYSVRSNAFPLSVQVFTIPPPHLK 344
rabbit          GDRAIYENELLATREVRTWSRGSITRDSIFRLRVSCSYSISSSALPVDMHVLTLPPPLPE 313
human           GDRAVYENELVATRDVKNGSRGSVTRDSIFRLHVSCSYSVSSNSLPINVQVFTLPPPFPE 309
rat             GDQAMYENELVAIRDVQAWGRSSITRDSNFRLRVSCTYSIHSIMSPVNMQVWTLPPPLPK 319
                *:.*:*****:* *:*:  .:.*:**** ***:*** **: *   *:.::* *:***  :

pig             THPGPLTLELQIAKDERYGSYYNASDYPVVKLLREPIYVEVSIRHRTDPSLGLHLHQCWA 370
cat             TQHGPLTLELKIAKDKHYGSYYTIGDYPVVKLLRDPIYVEVSIRHRTDPSLGLLLHNCWA 404
rabbit          TQPGPLTVVLQIAKDKDYHSYYTMDDYPVVKLLRDPIYVDVSILYRTDPYLGLRLHQCWA 373
human           TQPGPLTLELQIAKDKNYGSYYGVGDYPVVKLLRDPIYVEVSILHRTDPYLGLLLQQCWA 369
rat             TQPGPLSLELQIAQDKNYSSYYGTDAYPLVKFLQDPIYVEVSILHRTDPSLSLLLEQCWA 379
```

```
                *: ***:: *:**:*: * ***  . **:**:*::****:*** :**** *.* *.:***

pig             TPGMSPLLQPQWPMLVNGCPYTGDNYQTKLIPVQKASNLLFPSHYQRFSVSTFSFVD-SV 429
cat             TPGKNSQSLSQWPILVKGCPYVGDNYQTQLIPVQKALDTPFPSYYKRFSIFTFSFVD-TM 463
rabbit          TPRTNPLYQPQWPILVKGCPYTGDNYQTQLIPVQEAFDLPFPSHHQRFSISTFSFLDSSV 433
human           TPSTDPLSQPQWPILVKGCPYIGDNYQTQLIPVQKALDLPFPSHHQRFSIFTFSFVNPTV 429
rat             TPGSNPFHQPQWPILVKGCPYAGDNYQTKRIPVQKASDV-FPSHHQRFSISTFSFMSAGR 438
                **  ..   .***:**:**** ******: ****:* :  ***:::***: ****:.

pig             AKQALKGPVYLHCTASVCKPAGAPICVTTCPAARRRRSSDIHFQNGTASISSKGPMILLQ 489
cat             AKWALRGPVYLHCNVSICQPAGTSSCRITCPVARRRRHSDLHHHSSTASISSKGPMILLQ 523
rabbit          AKEALKGPIYLHCSVSVCQPTGTQSCTVTCPIDSRRRNSDINFQNSTANISSKGPMILLQ 493
human           EKQALRGPVHLHCSVSVCQPAETPSCVVTCPDLSRRRNFDNSSQNTTASVSSKGPMILLQ 489
rat             EKQVLGGQVYLHCSASVCQPAGMPSCTVICPASRRRRKSELYFDNSTS-ISSKGPVILLQ 497
                 * .* * ::***..*:*:*:    *   **   ***  :   .. *: :*****:****

pig             ATRDSSERLHKYSRPPVDSHALWVAGLLG-SLIIGALLVSYLVFRKWR-------- 536
cat             ATMDSAEKLHKNSSSPIDSQALWMAGLSG-TLIFGFLLVSYLAIRKRR-------- 570
rabbit          ATEDPSEKLHKHSGVPVHPGALWVAGLSG-IFIIGALLVSYVAIRTRR-------- 540
human           ATKDPPEKLR----VPVDSKVLWVAGLSG-TLILGALLVSYLAVKKQKSCPDQMCQ 540
rat             ATKDPAVMLHKHSGTHADSPTLWVMGLSASMVITGVLVVSYLATRKQR-------- 545
                ** *..  *:       .. .**: ** .  .* * *:***:. :. :
```

Supplementary file 3:

Cat ZP1 (XM_003993490), ZP2 (NM_001009875), ZP3 (NM_001009330) and ZP4 (NM_001009260) amino acid sequences. Bold underlined sequences are the tryptic peptides obtained by MS/MS in ZP obtained from oocytes. The putative N-glycosilation sites present in the detected peptides are in red. The putative O-glycosilation sites present in the detected peptides are in blue. The furin cleavage site (Arg-X-(Lys/Arg)-Arg) is showed in yellow.

ZP1

MVAVVYLMTRASAMVWDGCVALLLLLVAALGLGQRPHPEPGLRGLRHSSDCGIKGMQLLVFPRPGQTVRFKVVDE
FGNQFEVHNCSVCYHWVTARPLGPAVFSADYRGCHVLEKGGRFHLRVFVEAVLRDGRVDAAGEVTLICPKPGHTW
TPESHLASRTGFSLPTPHTQPLRPTQEHSFTRPTPALLPLRPGATRPTLTLPQWDILEHWGVDEPLHPGAPLTWE
QCQVPSGHIPCVVRRGSKEACQKAGCCYDNSR**AVPCYYGNTATVQCFR**NGHFVLVVSQETALAHGITLANIHVAY
APTSCSPTQDTGSFVVFQFPLTHCGTTVQVVGNQLLYENQLVSDIDVRMGPQGSITRDGAFRLHVRCTVNASDFL
PLQASIFSPPSPVPVIQSGPLRFQLRIATDETFRSFYEEGDYPIVR**LLREPVSVEVR**LLDRTDPGLVLLLHQCWA
TPGVSPFQQPQWPILSEGCPFDGDSYRTR**MVASDGAGLSFPSHHQR**FTVTTFALLDPDSQRALRGQVYFFCHSSA
CSPSGLETCSTTCSSRPARQRRSYNPHGEATRPQNLVSSPGPVDFEDSSGQEPPLGPTGSPRNANQRPLLWVVLL
LVAVALVLGVGVFEGLSQAXAQKLQEGDRGRLGSIKHRVQPAQRVWKAIRGD

ZP2

MASRQKGDSGSPSSWFNADWSTYRSLFLLFILVTSVNSIGVLQLVNPVFPGTVTCYETR**MAVEFPSDFGTKK**WHT
SVVDPFSFELLNCTYILDPENLTLKAPYETCTRRTLGQHRMIIRLKDHNAASRHNSLMYQINCPVMQAEETHEHA
GSTICTKDSMSFTFNVIPGLADENTDIKNPMGWSIEVGDGTKAKTLTLQDVLRQGYNILFDNHKITFQVSFNATG
VTHYMQGNSHLYMVPLKLIHESLGQKIILTTRVLCMSDAVTCNATHVTLTIPEFPGKLKSVSSENRNFAVSQLHN
NGIDKEESSGLTLHFSKTLLKMEFSEKCLPYQFYLASLKLTFAFNQETISTVLYPECVCESPVSIVTGDLCTQDG
FMDIK**VYSHQTKPALNLETLR**GGDSSCQPTFQAASQGLILFHIPLNGCGTRHKFKEGK**VIYENEIHAVWADLPPS
TISR**DSEFRMTVQCHYSK**GDLLINTR**VQSLPPPEASVRPGPLALILQTYPDKSYLQPYGEKEYPVVRYLR**QPIYL
EVR**VLNRSDPNIKLVLDDCWATPTMDPASVPQWNIIMDGCEYNLDNHR**TTFHPVGSSVTYPTHYR**RFDVKTFAFV
SEAQVLSSLVYFHCSVLICSRLSADSPLCSVTCPVSSRHRRATGTTEEEKMIVSLPGPILLLSDSSSLRDVVDSK
GYGAAGYVAFKTVVAVAALAGLVATLGFITYLRKNRTMINH*

ZP3

MGLSYGLFICFLLWAGTGLCYPPTTTEDKTHPSLPSSPSVVVECRHAWLVVNVSKNLFGTGRLVRPADLTLWPEN
CEPLISGDSDDTVRFEVELHKCGNSVQVTEDALVYSTFLLHNPRPMGNLSILRTNR**AEVPIECR**YPRHSNVSSEA
ILPTWVPFRTTMLSEEK**LAFSLRLMEEDWGSEKQSPTFQLGDLAHLQAEVHTGR**HIPLRLFVDYCVATLTPDQNA
SPHHTIVDFHGCLVDGLSDASSAFKAPRPRPETLQFTVYTFHFANDPRNMIYITCHLKVTPASRVPDQLNKACSF
IK**SSNRWFPVEGPADICNCCNKGSCGLQGR**SWRLSHLDRPWHKMASRNRRHVTEEADITVGPLIFLGKAADRGVE
GSTSPHTSVMVGIGLATVLSLTLATIVLGLARRHHTASRPMICPVSASQ*

ZP4

MWLLQPLLLCVPLSLAVHGQQKPQVPDYPGELHCGLQSLQFAINPSPGK**ATPALIVWDNR**GLPHKLQNNSGCGTW
VRESPGGSVLLDASYSSCYVNEWVSTTQSPGTSRPPTPASRVTPQDSHYVMIVGVEGTDAAGRRVTNTKVLRCPR
NPPDQALVSSLSPSPLQNVALEAPNADLCDSVPK**WDRLPCASSPITQGDCNKLGCCYK**SEANSCYYGNTVTSR**CT
QDGHFSIAVSR**NVTSPPLLLNSLRLAFGKDRECNPVKATR**AFALFFFPFNSCGTTRWVTGDQAVYENELVAAR**DV
RTWSHGSITR**DSIFR**LRVSCSYSVRSNAFPLSVQVFTIPPPHLKTQHGPLTLELK**IAKDKHYGSYYTIGDYPVVK
LLR**DPIYVEVSIRHRTDPSLGLLLHNCWATPGKNSQSLSQWPILVK**GCPYVGDNYQTQLIPVQK**ALDTPFPSYYK
RFSIFTFSFVDTMAKWALRGPVYLHCNVSICQPAGTSSCRITCPVARRRRHSDLHHHSSTASISSKGPMILLQAT
MDSAEKLHKNSSSPIDSQALWMAGLSGTLIFGFLLVSYLAIRKRR*

8. CONCLUSIONES

8. CONCLUSIONES

Utilizando un enfoque doble, análisis molecular y proteómico, describimos la composición de cuatro proteínas en la ZP de gata y coneja permitiendo futuros ensayos para el desarrollo potencial de vacunas anticonceptivas así como estudios acerca de la interacción ovocito-espermatozide en estas especies. Además, nuestro análisis filogenético describe los eventos de pseudogenización de ZP1 que han tenido lugar a lo largo de la evolución de los mamíferos. Por todo ello, podemos concluir lo siguiente:

1. CONEJO

1.1 El ARNm de ZP1 al igual que el de ZP2, ZP3 y ZP4 se encuentran presentes en el ovario de la coneja. El ADNc correspondiente a ZP1 ha sido amplificado, secuenciado y caracterizado.

1.2 El ARNm de ZP1 al igual que el de ZP2, ZP3 y ZP4 se expresa en el ovario de la coneja, dando lugar a una proteína. Estos resultados se apoyan en la detección de distintos péptidos pertenecientes a las cuatro glicoproteínas (ZP1, ZP2, ZP3 y ZP4) mediante espectrometría de masas.

1.3 La ZP de coneja formada por 4 proteínas puede constituir un nuevo y buen modelo experimental para el estudio de la función de las proteínas de la ZP en un contexto de cuatro proteínas.

2. GATO

2.1 El ARNm de ZP1 al igual que el de ZP2, ZP3 y ZP4 se encuentran presentes en el ovario del gato. El ADNc correspondiente a ZP1 ha sido amplificado, secuenciado y caracterizado.

2.2 El ARNm de ZP1 al igual que el de ZP2, ZP3 y ZP4 se expresa en el ovario del gato, dando lugar a una proteína. Estos resultados se apoyan en la detección de distintos péptidos pertenecientes a las cuatro glicoproteínas (ZP1, ZP2, ZP3 y ZP4) mediante espectrometría de masas.

2.3 La ZP de los diferentes felinos está probablemente formada por 4 proteínas lo que implica la existencia en los carnívoros de una ZP formada por 3 o 4 proteínas.

3. ANÁLISIS FILOGENÉTICO DEL GEN *ZP1* EN MAMÍFEROS

3.1 Un análisis filogenético indica que la pseudogenización de la ZP1 ha ocurrido al menos cuatro veces durante la evolución de los mamíferos afectando a las siguientes especies: el titi (Género *Callithrix*), el tarsero (Género *Tarsius*), el perro (Género Canis), el delfín (Género *Tursiops*) y el cerdo (Género *Sus*).

9. ABREVIATURAS

9. ABREVIATURAS

ADNc: ácido desoxirribonucleótido complementario

ADNg: ADN genómico

ARN: ácido ribonucleico

BSA: albúmina sérica bovina

CCO: complejo cúmulo-ovocito

dNTPs: desoxinucleótidos trifosfato

EDTA: ácido etildiaminotetraacético

DTT: ditiotreitol

FIV: fecundación *in vitro*

GC: gránulos corticales

VG: vesícula germinal

hCG: gonadotropina coriónica humana

HIS: hibridación *in situ*

HTF: human tubal fluid

HPLC: cromatografía líquida de alta presión

kDa: kilodalton

KO: *knock out*

m/z: relación masa/carga

MS: espectrometría de masas

Ma: millones de años

PB: tampón fosfato

PBS: tampón fosfato salino

PCR: reacción en cadena de la polimerasa

PMSG: gonadotropina sérica de yegua gestante

RA: reacción acrosómica

rpm: revoluciones por minuto

RT-PCR: reacción en cadena de la polimerasa a partir de ADNc

RT-qPCR: PCR en tiempo real

SDS-PAGE: electroforesis en gel de poliacrilamida con dodecilsulfato sódico

SDS: dodecilsulfato sódico

TAE: tris acetato EDTA

TFA: ácido trifluoroacético

Tm: temperatura de fusión o de melting

UI: unidades internacionales

ZP: zona pelúcida