HHS Public Access

Author manuscript

*Trends Dev Biol*. Author manuscript; available in PMC 2020 December 16.

Published in final edited form as:

*Trends Dev Biol*. 2020 ; 13: 21–33.

# Zona pellucida genes and proteins and human fertility

**Eveline S. Litscher**[*],

**Paul M. Wassarman**[*]

Department of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029-6574, USA.

## Abstract

The zona pellucida (ZP) is an extracellular matrix (ECM) that surrounds all mammalian oocytes, eggs, and embryos and plays vital roles during oogenesis, fertilization, and preimplantation development. The mouse and human ZP is composed of three or four unique proteins, respectively, called ZP1–4, that are synthesized, processed, and secreted by oocytes during their growth phase. All ZP proteins have a zona pellucida domain (ZPD) that consists of ≈270 amino acids and has 8 conserved Cys residues present as four intramolecular disulfides. Secreted ZP proteins assemble into long fibrils around growing oocytes with ZP2-ZP3 dimers located periodically along the fibrils. The fibrils are cross-linked by ZP1 to form a thick, transparent ECM to which sperm must first bind and then penetrate during fertilization of eggs. Inactivation of mouse *ZP1, ZP2,* or *ZP3* by gene targeting affects both ZP formation around oocytes and fertility. Female mice with eggs that lack a ZP due to inactivation of either *ZP2* or *ZP3* are completely infertile, whereas inactivation of *ZP1* results in construction of an abnormal ZP and reduced fertility. Results of a large number of studies of infertile female patients strongly suggest that gene sequence variations (GSV) in human *ZP1, ZP2,* or *ZP3* due to point, missense, or frameshift mutations have similar deleterious effects on ZP formation and female fertility. These findings are discussed in light of our current knowledge of ZP protein synthesis, processing, secretion, and assembly.

### Keywords

mouse; human; oogenesis; zona pellucida; ZP genes; ZP proteins; ZP domain; gene targeting; mouse fertility; gene sequence variations; human fertility

## 1. Introduction

The term “zona pellucida” was adopted by the Baltic German embryologist Karl Ernst von Baer (1792–1876) when describing human eggs for the first time in 1827 (Greek word *zone*, meaning belt or girdle; Latin word *pellucida*, meaning to shine or be transparent) [1]. By the 1840s the term zona pellucida (ZP) had gained widespread use among embryologists to describe the relatively thick transparent zone surrounding all mammalian eggs.

[*]Corresponding authors: eveline.litscher@mssm.edu, paul.wassarman@mssm.edu.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no disclosures.

The ZP is a relatively thick extracellular matrix (ECM) that surrounds all mammalian oocytes, eggs, and preimplantation embryos and plays various roles during oogenesis, fertilization, and preimplantation development [2–5]. It supports the health and growth of mammalian oocytes and follicles during the final stages of oogenesis, regulates species-restricted fertilization of eggs by sperm, and protects preimplantation embryos as they traverse the female reproductive tract toward the uterus. A ZP remains around preimplantation embryos until the expanded blastocyst stage when they hatch from the ZP and implant in the uterus. Shortly after eggs are fertilized by a single sperm, the ZP undergoes physical and biological changes that assist in both the prevention of polyspermy and protection of preimplantation embryos in the female reproductive tract.

The ZP is a viscoelastic, fibrillar ECM that is permeable to large macromolecules, such as antibodies, enzymes, and small viruses [3] (Figure 1). The width of the ZP varies relatively little, from ≈2 to ≈22 μm, whereas the diameter of eggs varies enormously among different animal species. ECM of most animal cells is constructed of proteoglycans, such as hyaluronic acid, heparin-, chondroitin-, and keratin-sulfate, and fibrous proteins, such as collagens, elastins, fibronectins, and laminins [6]. On the other hand, the ZP of mammalian eggs consists of either three (e.g., mouse) or four (e.g., human) glycosylated proteins, called ZP1–4, each with a unique polypeptide chain that includes a large ZP domain (ZPD) [3–5]. The ZPD consists of ≈270 aa, has 8 conserved cysteine (Cys) residues, and has two subdomains, called ZP-N and ZP-C, as well as a short linker region connecting the two subdomains [5, 7]. Eggs from fish, amphibia, reptiles, and birds are also surrounded by a ZP or vitelline envelope that consists of several ZPD proteins, all closely related to ZP1–4 [3, 5, 8].

Mouse ZP fibrils are polymers constructed of ZP2-ZP3 dimers and the long fibrils are interconnected by ZP1 to produce a thick ECM [7]. The ZP can be dissolved under conditions that do not result in breaking of covalent bonds, such as at low pH and elevated temperatures, suggesting that the structural integrity of the ZP is stabilized by non-covalent interactions between ZP proteins. Mechanical properties of the ZP, such as its viscosity and stiffness, change markedly after fertilization and is seen as an increased resistance of the ZP to dissolution by various agents (i.e., a hardening of the ZP) and an inability of free-swimming sperm to bind to the ZP of fertilized eggs [2, 4].

Mouse ZP proteins serve as structural proteins during assembly of the ZP around growing oocytes and as species-restricted receptors for sperm during fertilization [4, 9, 10]. Experiments involving gene targeting to produce *ZP1, ZP2,* and *ZP3* homozygous and heterozygous null mice suggest that inactivation or modification of ZP genes can have a significant effect on female fertility [4, 10]. In this context, results of a large number of relatively recent studies with infertile female patients suggest that gene sequence variations (GSV) in human *ZP1, ZP2,* or *ZP3* due to point, missense, or frameshift mutations can result in oocytes or eggs with either an abnormal or no ZP and in infertility. Here we provide some background information about ZP genes and proteins and discuss particular molecular aspects of GSV in human *ZP1, ZP2,* or *ZP3* that affect female fertility.

## 2. Mouse and human ZP genes and proteins

### 2.1. Mouse and human ZP genes

The mammalian egg ZP generally consists of either three or four proteins called ZP1–4. In mice and humans, ZP1–4 are encoded by single-copy genes located on different chromosomes (Table 1). Mouse *ZP1, ZP2,* and *ZP3* are located on chromosomes 19, 7, and 5, respectively, vary in length from 6.5 (*ZP1*) to 18.5 (*ZP2*) kb, and contain 12 (*ZP1*), 18 (*ZP2*), and 8 (*ZP3*) exons. Mouse *ZP4* is a pseudogene located on chromosome 13 [11–13]. Human *ZP1, ZP2, ZP3,* and *ZP4* are located on chromosomes 11, 16, 7, and 1, respectively, vary in length from 8.1 (*ZP1*) to 18.3 (*ZP3*) kb, and contain 12 (*ZP1*), 19 (*ZP2*), 8 (*ZP3*), and 12 (*ZP4*) exons [14].

### 2.2. Expression of ZP genes

ZP genes are expressed by oocytes during the latter stages of oogenesis when oocytes, arrested in first meiotic prophase, undergo tremendous growth (Figure 2). For example, mouse oocytes grow ≈300-fold, from ≈0.9 to ≈270 picoliters in volume, over 2–3 weeks [15]. Non-growing mouse oocytes, ≈12 μm in diameter, lack a ZP, whereas fully-grown oocytes, ≈80 μm in diameter, have a ZP that is ≈6.2 μm thick and contains ≈3.5 nanograms of protein. Consistent with the pattern of ZP1–3 synthesis, mRNA encoding mouse ZP proteins is undetectable in non-growing oocytes (i.e., <1,000 copies/cell), increases to hundreds-of-thousands of copies in mid-stage growing to fully-grown oocytes, and decreases to undetectable levels in fertilized eggs [16]. ZP genes are expressed only by growing oocytes, consequently, only by females, and is an example of gender-specific gene expression. Transgenic experiments, using a firefly luciferase reporter gene, strongly suggest that *cis*-acting sequences which restrict expression of ZP genes to growing oocytes are located very close to the transcription start-site [17, 18]. In addition, an E-box sequence (CANNTG) located about 200 bp upstream of the transcription start-site is apparently involved in oocyte-specific expression of ZP genes when E12/FIGα heterodimers bind to E-boxes [19, 20]. It should be noted that in fish and birds there are two sites of ZP protein synthesis, the ovary and the liver, and ZP precursor proteins synthesized in the liver are transported *via* circulating blood to the oocyte for uptake [5, 8].

### 2.3. Mouse and human ZP proteins

The mouse (m) ZP consists of three proteins, mZP1–3, and the human (h) ZP consists of four proteins, hZP1–4 (Figure 3). Nascent ZP proteins are processed by removal of an N-terminal signal sequence (SS) as proteins move from the endoplasmic reticulum to the Golgi and by removal of a C-terminal propeptide (CTP) at the oocyte's plasma membrane. The unprocessed polypeptides of mZP1, mZP2, and mZP3 are 623, 713, and 424 amino acids (aa) in length, respectively, and the processed polypeptides are 526 (≈120 kD MW), 599 (≈120 kD MW), and 329 (≈83 kD MW) aa in length, respectively [4, 15]. The unprocessed polypeptides of hZP1, hZP2, hZP3, and hZP4 are 638, 745, 424, and 540 aa in length, respectively, and the processed polypeptides are 528 (≈100 kD MW), 602 (≈75 kD MW), 328 (≈55 kD MW), and 445 (≈65 kD MW) aa in length, respectively [10, 21, 22] (Table 2). The polypeptides are heterogeneously glycosylated with asparagine-linked (N-linked) and serine/threonine-linked (O-linked) oligosaccharides that may be sialylated and sulfated [15].

As a consequence of these modifications, mouse and human ZP proteins are acidic (ave. pI <5.5) and migrate as very broad bands on SDS-PAGE.

Mouse and human ZP2 and ZP3 are monomers present in roughly equimolar amounts in the ZP and account for ≈80% of the mass of the ZP. ZP1 and ZP4 are homologous proteins and their genes are paralogous. ZP1 has a single intermolecular disulfide that supports the ZP1 dimer, whereas ZP4 does not have an intermolecular disulfide and in humans is a monomer [14, 23]. Unlike ZP2 and ZP3, ZP1 and ZP4 polypeptides possess a trefoil domain (TD), a three-loop compact structure with 6 Cys residues present as three intramolecular disulfides [24]. ZP1 serves as a cross-linker for the long fibrils that make up the ZP matrix. Mouse and human ZP1 have a proline (Pro)-rich N-terminus (100 aa, 17–21% Pro) that may provide the flexibility that contributes to the documented elasticity of the ZP prior to fertilization [4].

The primary structures of mouse and human ZP1–3 are well conserved having ≈64% identity and ≈84% similarity (Table 3) and human ZP1 and ZP4 are ≈43% identical and ≈72% similar. In this context, transgenic mouse lines have been established in which human ZP1–3 replaced the endogenous mouse ZP proteins [25, 26]. Mouse and human ZP polypeptides exhibit several regions characteristic of ZP proteins from jellyfish to humans, representing more than 600 million years of evolution [27]. These regions include an N-terminal SS (≈25 aa) that targets nascent ZP protein to the secretory pathway, followed by a long ZPD (≈270 aa) and a CTP required for secretion of nascent protein. The CTP includes a consensus furin cleavage-site (CFCS; ≈4 aa), an intervening sequence, a hydrophobic transmembrane domain (TMD; ≈20 aa), and a C-terminal tail (Figure 3; Table 2).

### 2.4. Structure of the ZPD

The ZPD, which has 8 conserved Cys residues present as four intramolecular disulfides, is found in all ZP proteins as well as in hundreds of other proteins of diverse functions, from receptors to mechanical transducers, from a wide variety of tissues and organs in all multicellular organisms [4, 5, 7]. The ZPD is a bipartite structure consisting of two subdomains, an N-terminal ZP-N (≈100 aa) and a C-terminal ZP-C (≈135–150 aa) subdomain, separated from each other by a short protease-sensitive region (≈25–30 aa). ZP1 and ZP4 have one extra ZP-N subdomain and ZP2 has three extra ZP-N subdomains as extensions at the N-terminus of their polypeptides [28]. ZP-N is used for polymerization of nascent ZP proteins and other extracellular ZPD proteins, such as tectorin and uromodulin, into long fibrils [29, 30].

The three-dimensional structure of the ZPD has been determined by X-ray diffraction and revealed that both the ZP-N and ZP-C subdomains adopt immunoglobulin (Ig)-like folds [4, 31, 32]. The ancestral gene for the Ig superfamily may have originated about 750 million years ago in sponges as a primitive sandwich-like fold used in vertebrate extracellular recognition systems [33]. The ZP-N fold consists of an antiparallel sandwich of two β-sheets made up of eight strands of polypeptide that enclose a hydrophobic core of buried residue side-chains, with two buried disulfides that clamp both sides of the sandwich. The ZP-C fold also consists of a β-sandwich comprising two stacked β-sheets, one with four strands and the other with six strands, that approximate a Greek-key like motif characteristic of Ig-like domains [34]. The strong structural similarity between the ZP-N and ZP-C subdomains

suggests that the ZPD may have arisen by duplication of an ancestral gene encoding a protein containing a single ZP-N subdomain.

### 2.5. Structure of ZP fibrils

In mice the ZP consists of cross-linked fibrils 7–8 nm in width and several μm in length with a ZP2-ZP3 dimer located every 14–15 nm along the fibrils [35–37]. The structural periodicity of fibrils can be visualized as protuberances along fibrils in electron micrographs of solubilized ZP, as well as in micrographs of solubilized ZP preparations decorated with monoclonal antibodies against either ZP2 or ZP3. Fibrils in the inner and outer layers of the ZP are oriented perpendicular and parallel, respectively, to the oolemma, whereas fibrils in the intervening layer are oriented randomly [4, 38–40]. It has been proposed that ZP fibrils have properties analogous to amyloid proteins that self-aggregate and form cross-β-sheet fibrillar structures [4, 41, 42]. However, ZP fibrils consist of heteromeric aggregates (i.e., ZP2-ZP3) rather than the homomeric aggregates typical of amyloids [43]. It has been suggested that the plasticity (rigid or flexible) of the linker between ZP-N and ZP-C subdomains of ZPD proteins determines whether the proteins form homo- or hetero-polymers [44].

## 3. Mouse ZP genes and female fertility

Results of experiments in which antisense oligonucleotides directed against either ZP2 or ZP3 mRNAs were injected into growing mouse oocytes suggest that ZP2 and ZP3 are dependent upon each other for incorporation into the ZP [45]. To extend these observations, gene targeting was used to establish mouse lines in which ZP genes were inactivated by either homologous recombination or insertional mutagenesis and the fertility of the mice was assessed. Results of the gene targeting experiments in mice are summarized below and in Table 4.

### 3.1. Wild-type male mice

Male mice that are homozygous nulls for *ZP1, ZP2,* or *ZP3* are as fertile as wild-type males. ZP genes are only expressed in female mice.

### 3.2. ZP2 and ZP3 homozygous null female mice

Female mice that are homozygous nulls for either *ZP2* ($ZP2^{-/-}$) or *ZP3* ($ZP3^{-/-}$) produce eggs that lack a ZP and the females are infertile [46–48]. The infertility is due to a scarcity of growing oocytes in ovaries and ovulated eggs in oviducts of homozygous null mice. This suggests that the presence of both ZP2 and ZP3 is absolutely required for assembly of a ZP around growing oocytes and these findings are consistent with results of antisense experiments mentioned above. The paucity of growing oocytes and follicles in ovaries of $ZP3^{-/-}$ mice is reflected in the weight difference of ovaries from wild-type females (20 days-of-age, 1.0 ± 0.17 mg/ovary) and ovaries from $ZP3^{-/-}$ females (20 days-of-age, 0.26 ± 0.1 mg/ovary) [49]. In the absence of a ZP, formation of gap junctions between oocytes and surrounding follicle cells is severely reduced, thereby compromising transfer of nutrients, metabolites, and other molecules essential for oocyte and follicle growth. The latter finding is consistent with the phenotype of female mice that are homozygous nulls for gap junction

proteins, such as connexin-37 and −43, that are deficient in multi-layered follicles and growing oocytes and are infertile [50, 51].

### 3.3. ZP3 heterozygous null female mice

Female mice that are heterozygous nulls for *ZP3* (*ZP3*$^{+/-}$) are as fertile as wild-type females, but their eggs have a thin ZP (ave. width 2.7 ± 1.2 μm) compared to the ZP of eggs from wild-type females (ave. width 6.2 ± 1.9 μm) [52]. The thin ZP contains about one-half the amount of ZP2 and ZP3 found in ZP of eggs from wild-type mice. These observations suggest that the width of the ZP is not a critical parameter for either binding of free-swimming sperm to the ZP or fertilization of eggs.

### 3.4. ZP1 homozygous null female mice

Female mice that are homozygous nulls for *ZP1* (*ZP1*$^{-/-}$) are fertile, but exhibit reduced fertility compared to wild-type mice due to early loss of preimplantation embryos in the oviduct [53]. This loss is attributable to a ZP that is insufficiently cross-linked and, consequently, extremely fragile as cleavage-stage embryos traverse the reproductive tract on their way to the uterus. The presence of ZP2 and ZP3 in growing oocytes of *ZP1*$^{-/-}$ mice supports formation of heterodimers that can assemble into long fibrils. However, in the absence of ZP1 the fibrils are not properly cross-linked, creating an unusually porous ZP matrix that even permits follicle cells to enter the perivitelline space between the oocyte's ZP and plasma membrane. New insights into the structural basis of ZP1/ZP4 crosslinking of human ZP have recently been obtained [54].

## 4. Human ZP genes and female fertility

Some early evidence suggested that there might be a causal relationship between gene sequence variations (GSV) in human *ZP* genes and female fertility [55–57]. For example, it was found that there was ≈1.5-times more GSV in *ZP1* and *ZP3* of women who were unsuccessful in IVF trials compared to women with proven fertility [55]. This finding has now been extended by a large number of studies carried out to assess whether GSV in human *ZP1–3* have an effect on female fertility. Results of these experiments with human patients are summarized below and in Table 5.

### 4.1. GSV in human ZP1

One study revealed a homozygous frameshift deletion of eight bp in *ZP1* of women who were infertile and whose eggs lacked a ZP [58]. The deletion was predicted to result in a premature stop-codon (SC) in *ZP1* and synthesis of a truncated form of ZP1; Ile390*fs*404X, a 404 aa polypeptide versus a 638 aa polypeptide in wild-type *ZP1*. Truncated ZP1 had the N-terminal SS, TD, and first half of the ZPD, but was missing the CTP essential for protein secretion [59–62]. Since oocytes from *ZP1* homozygous null mice have a ZP, albeit a very loose and porous ZP, it was surprising that oocytes from these women lacked a ZP. However, subsequently it was reported that accumulation of truncated ZP1 in the oocyte's cytoplasm apparently interfered with secretion of nascent ZP3 and ZP4 and thereby prevented assembly of a ZP around growing ocytes [63]. An alternative explanation for the observation has

recently been put forward that does not involve interference with secretion of nascent ZP proteins by truncated ZP1, but rather by affecting the cross-linking function of ZP1 [54].

Other studies also have attributed female infertility to GSV in *ZP1*. A heterozygous missense mutation in *exon-3* of *ZP1* was identified in an infertile patient whose oocytes lacked a ZP [64]. The mutation resulted in His replacing Arg109 at the N-terminus of ZP1. Similarly, a compound heterozygous mutation consisting of a point mutation and deletion in *ZP1* was identified in an infertile woman whose oocytes lacked a ZP [65]. The mutation in *exon-5* resulted in synthesis of ZP1 stopping at Gln292 and a two basepair deletion in *exon-7* also resulted in a premature SC and ZP1 synthesis stopping at Ile386.

Several additional studies also led to identification of GSV in *ZP1* in infertile females who had abnormal oocytes. For example, a missense mutation in exon-2, Trp83Arg, was found in a patient with degenerated oocytes and an abnormal or no ZP, and in another patient a nonsense mutation with a premature SC in exon-8, Trp471→X, had a similar phenotype [66]. A compound heterozygous mutation, Arg61Cys and Ile390Thr*fs**16, was found to be associated with abnormal oocytes and no ZP since replacement of Arg61 with Cys was predicted to be deleterious to ZP1 and the frameshift mutation introducing an SC in *exon-7* (Ile390Thr*fs*404X) resulted in a 234 aa deletion at the C-terminus of ZP1 [67, 68]. In another case, two frameshift mutations in *ZP1* resulted in premature SCs in *exon-1* (Gly57Asp*fs**9) and *exon-7* (Ile390Thr*fs**16) and apparently disrupted interactions between ZP proteins and caused degeneration of oocytes [68]. Missense mutations, Val570Met and Arg410Trp, were identified in two infertile females that had no oocytes or oocytes lacking a ZP [69]. Similarly, a compound heterozygous mutation with premature SCs in *exon-9* (Cys478→X) and *exon-12* (Asp592Gly*fs**29) and a frameshift mutation in *exon-3,* His170Ilefs*52, were identified that possibly resulted in a truncated ZP1 that interfered with ZP formation [69].

In many instances GSV in *ZP1* affected its ZPD (Figure 4), a region of all ZP proteins considered critical for proper secretion of nascent ZP proteins and proper assembly of a ZP around growing oocytes [4, 7, 29, 30, 61].

### 4.2. GSV in human ZP2 and ZP3

GSV in human *ZP2* and *ZP3* of infertile women can also result in synthesis of ZP proteins that are unable to undergo normal secretion and assembly during oocyte growth. An infertile woman was found to have a heterozygous missense mutation in *ZP2* (*exon-19* Arg698→X) with insertion of a SC at aa 698 and a heterozygous frameshift mutation in *ZP3* (*exon-8* Arg349Leu→X) followed by a SC [70]. Both mutations resulted in the synthesis of truncated ZP proteins, ZP2 lacking a TMD and ZP3 lacking a CTP. Three other cases of GSV in *ZP2* have been described in which Cys372 was changed to Ser, Arg533 to Ser, and Cys566 to Arg [66, 69]. All these changes occurred at conserved aa residues in the ZPD of ZP2, aa 371–637. A heterozygous missense variant in *exon-2* of *ZP3* also was identified as a change of Ala134 to Thr; a change proposed to cause empty follicle syndrome and female infertility [64, 71]. A similar missense mutation in *exon-5* of *ZP3*, Arg255 to Gly, was found in a female with primary infertility [69]. In both cases the mutations occurred in the ZPD of ZP3, aa 45–304.

## 5. Final comments

ZP1–4 are structural proteins essential for constructing a proper ZP around mammalian oocytes during their growth phase [3, 15, 37]. ZP2-ZP3 dimers polymerize to form the long fibrils cross-linked by ZP1 that constitute the thick ECM. ZP genes in mice and humans are single-copy genes located on different chromosomes and are expressed only in growing oocytes. Gene targeting of *ZP1–3* in mice [46–48, 52] and analysis of GSV in *ZP1–3* in human patients [58, 63–71] have revealed that female fertility can be severely affected by the absence of ZP1, ZP2, or ZP3 or by the presence of mutated forms of ZP1, ZP2, or ZP3 during oocyte growth. Failure to assemble a proper ZP around growing oocytes can result in infertility.

Inactivation of mouse *ZP1, ZP2, or ZP3* by gene targeting results in infertility ($ZP2^{-/-}$ or $ZP3^{-/-}$) or reduced fertility ($ZP1^{-/-}$) in mice and strongly suggests that both ZP2 and ZP3 must be present to construct a proper ZP. The presence of mouse ZP1, a cross-linker protein, is not required since a loose ZP can be constructed from ZP2 and ZP3. $ZP1^{-/-}$ female mice are fertile, but a reduced number of preimplantation embryos survive in their reproductive tract due to a labile ZP. The thickness of the ZP does not appear to affect fertility since heterozygous null mice ($ZP3^{+/-}$) with a thin ZP, about one-half the width of a wild-type ZP, are as fertile as wild-type mice.

GSV in human *ZP1, ZP2,* or *ZP3* as point, missense, or frameshift mutations also can result in infertility in humans. In many instances the mutations affect secretion of nascent ZP1–3 by growing oocytes due to insertion of premature SCs in ZP genes and thereby prevent construction of a proper ZP and causes infertility. The extent of female infertility has increased dramatically (≈15%) over the past 25 years, such that today it is estimated that ≈10% of married women worldwide are infertile. It is likely that many more mutations in human *ZP1–3* and *ZP4* will be identified in the future that account for female infertility. Correction of these mutations by genetic engineering, perhaps by using CRISPR/Cas9 technology, remains an interesting and tempting possibility.

## ACKNOWLEDGEMENTS

We are grateful to all members of our laboratory over the years who contributed to our research on the zona pellucida. We respectfully acknowledge the late Jeffrey David Bleil (1952-2014) who initiated our research on the zona pellucida as a graduate student at Harvard Medical School more than forty years ago. We thank Giuseppe Familiari at the University of Rome, Italy, for kindly providing images from his publications for our mini-review. Our research was supported in part by the NIH (NICHD) and F. Hoffmann-La Roche AG.

## ABBREVIATIONS

| | |
|---|---|
| **aa** | amino acid |
| **N-terminus** | amino-terminus |
| **bp** | basepairs |
| **CTP** | carboxy-terminal propeptid |
| **C-terminus** | carboxy-terminus |

| | |
|---|---|
| **CFCS** | concensus furin cleavage-site |
| **Cys** | cysteine |
| **ECM** | extracellular matrix |
| **GSV** | gene sequence variations |
| **hZP1–4** | human ZP1–4 |
| **pI** | isoelectric point |
| **kD** | kilodaltons |
| **MW** | molecular weight |
| **mZP1–3** | mouse ZP1–3 |
| **SC** | stop-codon |
| **SS** | signal sequence |
| **TD** | trefoil domain |
| **TMD** | transmembrane domain |
| **ZP** | zona pellucida |
| **ZP-C** | ZPD carboxy-terminal subdomain |
| **ZPD** | zona pellucida domain |
| **ZP-N** | ZPD amino-terminal subdomain |

## REFERENCES


1. von Baer KE 1827, De ovi mammalium et hominis genesi [On the Genesis of the Ovum of Mammals and of Man]. Leopold Voss, Leipzig, Germany.
2. Wassarman PM 1988, Annu. Rev. Biochem, 57, 415–442. [PubMed: 3052278]
3. Litscher ES and Wassarman PM (Eds.) 2018, Extracellular Matrix and Egg Coats. Academic Press/Elsevier, Oxford, UK.
4. Litscher ES and Wassarman PM 2020, Annu. Rev. Biochem, 89, 695–715. [PubMed: 32569527]
5. Litscher ES and Wassarman PM 2015, A Guide to Zona Pellucida Domain Proteins. John Wiley and Sons, Hoboken.
6. Franz C, Stewart KM and Weaver VM 2010, J. Cell Sci, 123, 4195–4200. [PubMed: 21123617]
7. Jovine L, Darie CC, Litscher ES and Wassarman PM 2005, Annu. Rev. Biochem, 74, 83–114. [PubMed: 15952882]
8. Litscher ES and Wassarman PM 2014, Trends Dev. Biol, 8, 65–76. [PubMed: 26504367]
9. Wassarman PM 1987, Science, 235, 553–560. [PubMed: 3027891]
10. Avella MA, Xiong B and Dean J 2013, Mol. Hum. Reprod, 19, 279–289. [PubMed: 23335731]
11. Liang L, Chamow SM and Dean J 1990, Mol. Cell. Biol, 10, 1507–1515. [PubMed: 1690843]
12. Liang L-F and Dean J 1993, Dev. Biol, 156, 399–408. [PubMed: 8385033]
13. Epifano O, Liang L-F and Dean J 1995, J. Biol. Chem, 270, 27254–27258. [PubMed: 7592984]
14. Hughes DC and Barratt CL 1999, Biochem. Biophys. Acta, 1447, 303–306. [PubMed: 10542331]

15. Wassarman PM and Litscher ES 2018, Curr. Top. Dev. Biol, 130, 331–356. [PubMed: 29853182]
16. Roller RJ, Kinloch RA, Hiraoka BY, Li SS and Wassarman PM 1989, Development, 106, 251–261. [PubMed: 2591314]
17. Lira SA, Kinloch RA, Mortillo S and Wassarman PM 1990, Proc. Natl. Acad. Sci. USA, 87, 7215–7221. [PubMed: 2402504]
18. Lira SA, Schickler M and Wassarman PM 1993, Mol. Reprod. Dev, 36, 494–499. [PubMed: 8305213]
19. Liang L-F, Soyal SM and Dean J 1997, Development, 124, 4939–4947. [PubMed: 9362457]
20. Soyal SM, Amleh A and Dean J 2000, Development, 127, 4645–4654. [PubMed: 11023867]
21. Lefievre L, Conner SJ, Salpekar A, Olufowobi O, Ashton P, Pavlovic B, Lenton W, Afnan M, Brewis IA, Monk M, Hughes DC and Barratt CL 2004, Hum. Reprod, 19, 1580–1586. [PubMed: 15142998]
22. Gupta SK 2018, Curr. Top. Dev. Biol, 130, 379–411. [PubMed: 29853184]
23. Connor SJ, Lefievre L, Hughes DC and Barratt CL 2005, Hum. Reprod, 20, 1148–1152. [PubMed: 15760956]
24. Thim L 1997, Cell. Mol. Life Sci, 53, 888–903. [PubMed: 9447240]
25. Rankin TL, Coleman JS, Epifano O, Hoodbhoy T, Turner SG, Castle PE, Lee E, Gore-Langton R and Dean J 2003, Dev. Cell, 5, 33–43. [PubMed: 12852850]
26. Baibakov B, Boggs NA, Yauger B, Baibakov G and Dean J 2012, J. Cell Biol, 197, 897–905. [PubMed: 22734000]
27. Killingbeck EE and Swanson WJ 2018, Curr. Top. Dev. Biol, 130, 443–488. [PubMed: 29853187]
28. Callebaut I, Mornon JP and Monget P 2007, Bioinformatics, 23, 1871–1874. [PubMed: 17510169]
29. Jovine L, Qi H, Williams Z, Litscher E and Wassarman PM 2002, Nature Cell Biol, 4, 457–461. [PubMed: 12021773]
30. Jovine L, Janssen WG, Litscher ES and Wassarman PM 2006, BMC Biochem, 7, 11. [PubMed: 16600035]
31. Raj I, Sadat Al Hosseini H, Dioguardi E, Nishimura K, Han L, Villa A, de Sanctis D and Jovine L 2017, Cell, 169, 1315–1326. [PubMed: 28622512]
32. Bokhove M and Jovine L 2018, Curr. Top. Dev. Biol, 130, 413–442. [PubMed: 29853186]
33. Williams AF and Barclay AN 1988, Annu. Rev. Immunol, 6, 381–405. [PubMed: 3289571]
34. Hutchinson EG and Thornton JM 1993, Protein Eng, 6, 233–245. [PubMed: 8506258]
35. Greve JM and Wassarman PM 1985, J. Mol. Biol, 181, 253–264. [PubMed: 3845123]
36. Wassarman PM and Mortillo S 1991, Intl. Rev. Cytol, 130, 85–110.
37. Wassarman PM, Liu C and Litscher ES 1996, J. Cell Sci, 109, 2001–2004. [PubMed: 8856495]
38. Keefe D, Tran P, Pellegrini C and Oldenbourg R 1997, Hum. Reprod, 12, 1250–1252. [PubMed: 9222011]
39. Pelletier C, Keefe DL and Trimarchi JR 2004, Fertil. Steril, 81, 850–856. [PubMed: 15019819]
40. El-Mestrah M, Castle PE, Borossa G and Kan FWK 2002, Biol. Reprod, 66, 866–876. [PubMed: 11906903]
41. Egge N, Muthusubramanian A and Cornwall GA 2015, PLoS One, 10, e129907.
42. Louros NN, Iconomidou VA, Giannelou P and Hamodrakas SJ 2013, PLoS One, 8, e73258. [PubMed: 24069181]
43. Eisenberg DS and Sawaya MR 2017, Annu. Rev. Biochem, 86, 69–95. [PubMed: 28125289]
44. Bokhove M, Nishimura K, Brunati M, Han L, de Sanctis D, Rampoldi L and Jovine L 2016, Proc. Natl. Acad. Sci. USA, 113, 1552–1557. [PubMed: 26811476]
45. Tong Z-B, Nelson LM and Dean J 1995, J. Biol. Chem, 270, 849–853. [PubMed: 7822321]
46. Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, Stewart CL and Wassarman PM 1996, Proc. Natl. Acad. Sci. USA, 93, 5431–5436. [PubMed: 8643592]
47. Rankin T, Familiari M, Lee E, Ginsberg A, Dwyer N, Blanchette-Mackie J, Drago J, Westphal H and Dean J 1996, Development, 122, 2903–2910. [PubMed: 8787763]

48. Rankin T, O'Brien M, Lee E, Wigglesworth K, Eppig J and Dean J 2001, Development, 128, 1119–1126. [PubMed: 11245577]
49. Wassarman PM and Litscher ES 2012, Intl. J. Dev. Biol, 56, 833–839.
50. Simon AM, Goodenough DA, Li E and Paul DL 1997, Nature, 385, 525–529. [PubMed: 9020357]
51. Ackert CL, Gittens JE, O'Brien MJ, Eppig JJ and Kidder GM 2001, Dev. Biol, 233, 258–270. [PubMed: 11336494]
52. Wassarman PM, Qi H and Litscher ES 1997, Proc. Roy. Soc. London, Biol. Sci, 264, 323–328.
53. Rankin TL, Talbot P, Lee E and Dean J 1999, Development, 126, 3847–3855. [PubMed: 10433913]
54. Nishimura K, Dioguardi E, Nishio S, Villa A, Han L, Matsuda T and Jovine L 2019, Nature Comm, 10, 3086.
55. Männikkö M, Törmälä RM, Tuuri T, Haltia A, Martikainen H, Ala-Kokko L, Tapanainen JS and Lakkakorpi JT 2005, Hum. Reprod, 20, 1578–1585. [PubMed: 15860499]
56. Pökkylä RM, Lakkakorpi JT, Noujua-Huttunen SH and Tapanainen JS 2011, Fertil. Steril, 95, 2669–2672. [PubMed: 21334611]
57. Margalit M, Paz G, Yavetz H, Yogev L, Amit A, Hevlin-Schwartz T, Gupta SK and Kleiman SE 2012, Eur. J. Obstet. Gynecol. Reprod. Biol, 165, 70–76. [PubMed: 22889493]
58. Huang H-L, Lv C, Zhao Y-C, Li W, He X-M, Li P, Sha A-G, Tian X, Papasian CJ, Deng H-W, Lu G-X and Xiao H-M 2014, New Engl. J. Med, 370, 1220–1226. [PubMed: 24670168]
59. Williams Z and Wassarman PM 2001, Biochemistry, 40, 929–937. [PubMed: 11170414]
60. Qi H, Williams Z and Wassarman PM 2002, Mol. Biol. Cell, 13, 530–541. [PubMed: 11854410]
61. Jovine L, Qi H, Williams Z, Litscher ES and Wassarman PM 2004, Proc. Natl. Acad. Sci. USA, 101, 5922–5927. [PubMed: 15079052]
62. Jimenez-Movilla M and Dean J 2011, J. Cell Sci, 124, 940–950. [PubMed: 21378311]
63. Lv C, Huang H-L, Wang Y, Peng T-L, Tan H-J, Zeng MH, Quan R-P, Deng H-W and Xiao H-M 2019, BioRxiv, 825018.
64. Cao Q, Zhao C, Zhang X, Zhang H, Lu Q, Wang C, Hu Y, Ling X, Zhang J and Huo R 2020, J. Cell. Mol. Med, 2020, 1–10.
65. Zhang Z, Shangguan T, Li Y-Y and He W 2020, Reprod. Dev. Med, 2, 183–186.
66. Yang P, Luan X, Peng Y, Chen T, Su S, Zhang C, Wang Z, Cheng L, Zhang X, Wang Y, Chen Z-J and Zhao H 2017, Fertil. Steril, 107, 1364–1369. [PubMed: 28577617]
67. Yuan P, Li R, Li D, Zheng L, Ou S, Zhao H, Zhang Q and Wang W 2019, J. Assist. Reprod. Genet, 36, 741–747. [PubMed: 30778819]
68. Sun L, Fang X, Chen Z, Zhang H, Zhang Z, Zhou P, Xue T, Peng X, Zhu Q, Yin M, Liu C, Deng Y, Hu H and Li N 2019, Hum. Mut, 40, 2001–2006. [PubMed: 31292994]
69. Zhou Z, Ni C, Wu L, Chen B, Xu Y, Zhang S, Mu J, Li B, Yan Z, Fu J, Wang W, Zhao L, Dong J, Sun X, Kuang Y, Sang Q and Wang L 2019, Hum. Genet, 138, 327–337. [PubMed: 30810869]
70. Liu W, Li K, Bai D, Yin J, Tang Y, Chi F, Zhang L, Wang Y, Pan J, Liang S, Guo Y, Ruan J, Kou X, Zhao Y, Wang H, Chen J, Teng X and Gao S 2017, Hum. Genet, 136, 1822–1827.
71. Chen T, Bian Y, Liu X, Zhao S, Wu K, Yan L, Li M, Yang Z, Liu H, Zhao H and Chen Z-J 2017, Amer. J. Hum. Genet, 101, 459–465. [PubMed: 28886344]
72. Familiari G, Relucenti M, Heyn R, Micara G and Correr S 2006, Microsc. Res. Tech, 69, 415–426. [PubMed: 16703610]

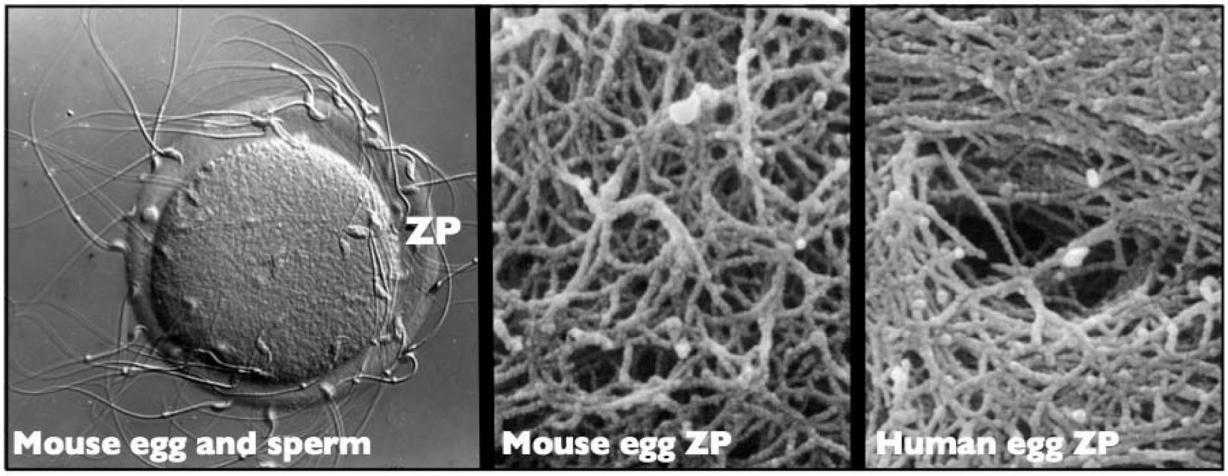


**Figure 1.**

Photographic images of the mouse and human egg zona pellucida (ZP). *Left:* A light micrograph (Nomarski differential interference contrast) of an unfertilized mouse egg incubated in the presence of free-swimming sperm. Sperm are shown bound to the ZP (≈550 × magnification). *Middle:* Scanning electron micrograph of the outer surface of the ZP of a mouse oocyte showing its fibrillar organization (≈50,000 × magnification). *Right:* Scanning electron micrograph of the outer surface of the ZP of a human oocyte showing its fibrillar organization and the presence of a pore (≈50,000 × magnification). The fibrils are 0.1–0.4 μm long and 10–14 nm wide. The latter two micrographs (*Middle* and *Right*) were obtained with samples treated with saponin-ruthenium red-osmium thiocarbohydrazide to reveal mouse and human ZP fibrils and are included here with the permission of G. Familiari at the University of Rome, Italy [72].

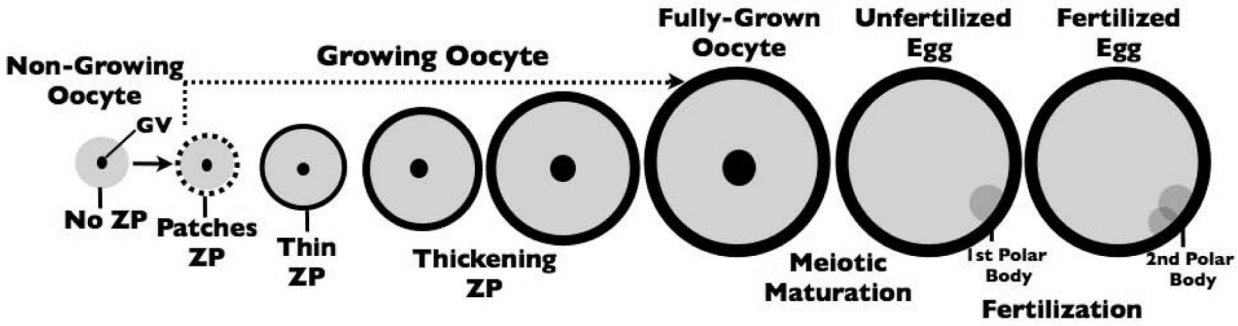


**Figure 2.**

Schematic representation of zona pellucida (ZP) production during oocyte growth in mice (2–3 weeks). Non-growing oocytes lack a ZP, but as soon as they begin to grow they lay down patches of ZP fibrils (*black dashes*). The fibrils coalesce into a ZP matrix (*black*) relatively early in oocyte growth and the ZP (*black*) continues to thicken throughout oocyte growth. The ZP remains around fully-grown oocytes [germinal vesicle (GV) present, *black*], unfertilized eggs (dissolution of the GV and emission of the first polar body, *dark gray*), and fertilized eggs (emission of second polar body, *dark gray*). The ZP remains around the early embryo until the expanded blastocyst stage when the embryo hatches from the ZP and implants in the uterus.

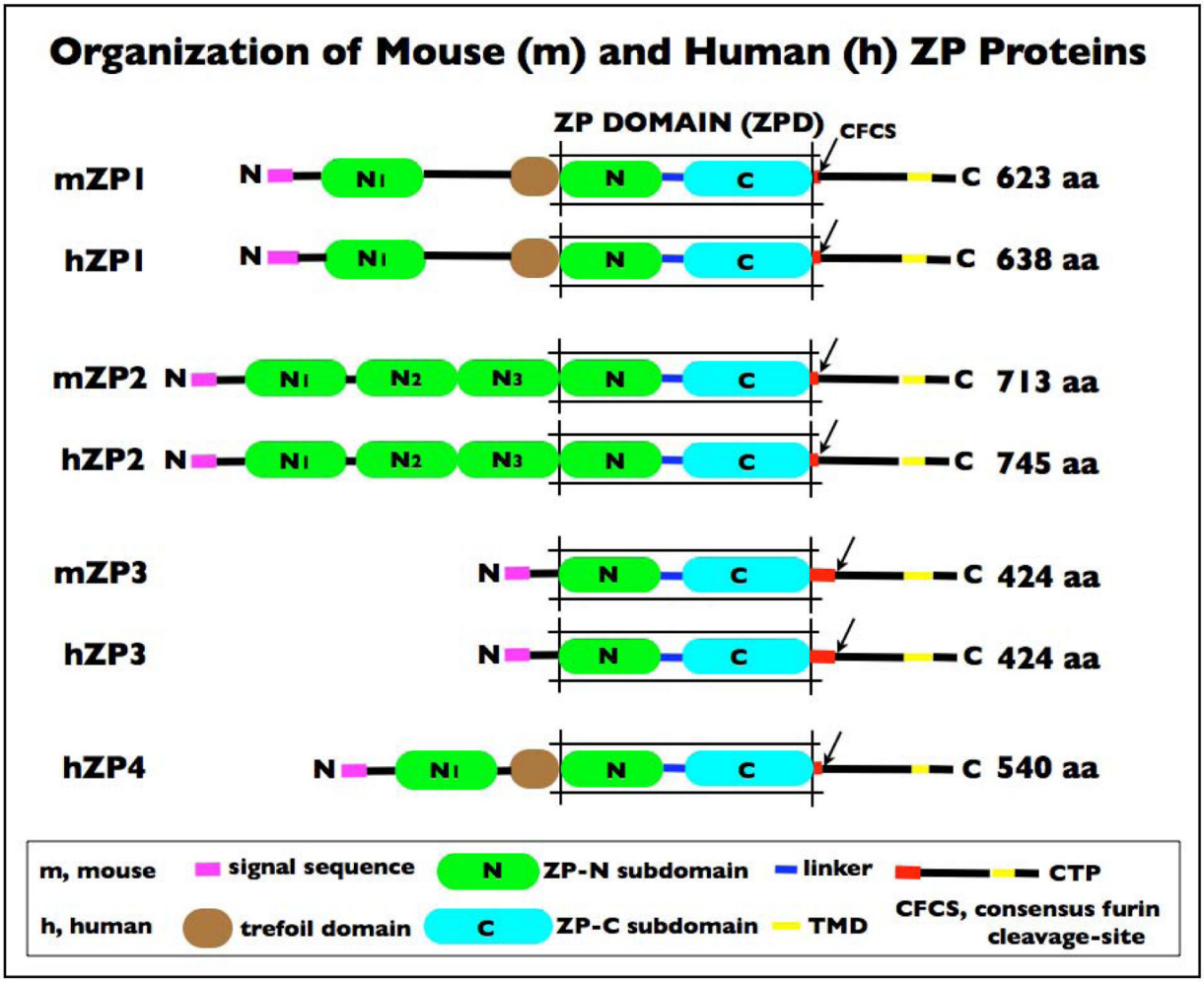


**Figure 3.**

Schematic representation of the organization of mouse (m) zona pellucida (ZP) proteins mZP1–3 and human (h) ZP proteins hZP1–4. Linker region (*blue*); concensus furin cleavage-site (CFCS; *arrow*); transmembrane domain (TMD; *yellow*); and cytoplasmic tail (*black*) in the C-terminus. In each case, the polypeptide has a signal sequence (SS; *magenta*) at the N-terminus; a ZP domain (ZPD) consisting of two subdomains, ZP-N (*green*) and ZP-C (*cyan*), and a C-terminal propeptide (CTP). mZP1 (623 aa), hZP1 (638 aa), and hZP4 (540 aa) also have a trefoil domain (TD; *brown*) adjacent to the ZPD. mZP2 (713 aa) and hZP2 (745 aa) have 3 additional ZP-N subdomains (N1-N3; *green*) between the N-terminus of the polypeptide and the ZPD. mZP1, hZP1, and hZP4 have one additional ZP-N subdomain (N1; *green*) between the N-terminus of the polypeptide and the TD. mZP3 (424 aa) and hZP3 (424 aa) are the smallest of the ZP proteins and consist primarily of a ZPD (mZP3 260 aa; hZP3 259 aa). Notably, mZP1 and 2 and hZP1, 2, and 4 have only 3 to 5 aa between the ZPD and the CFCS (*red*), whereas mZP3 has 47 aa and hZP3 has 45 aa between the ZPD and the CFCS which includes the sperm combining-site [9, 59] and represents a region of positive Darwinian selection during evolution [15, 27]. This region of the ZP3 polypeptide also includes 4 conserved Cys residues in close proximity to one

another (mZP3, Cys-320, −322, −323, and −328; hZP3, Cys-319, −321, −322, and −327) that could form two intramolecular disulfides.

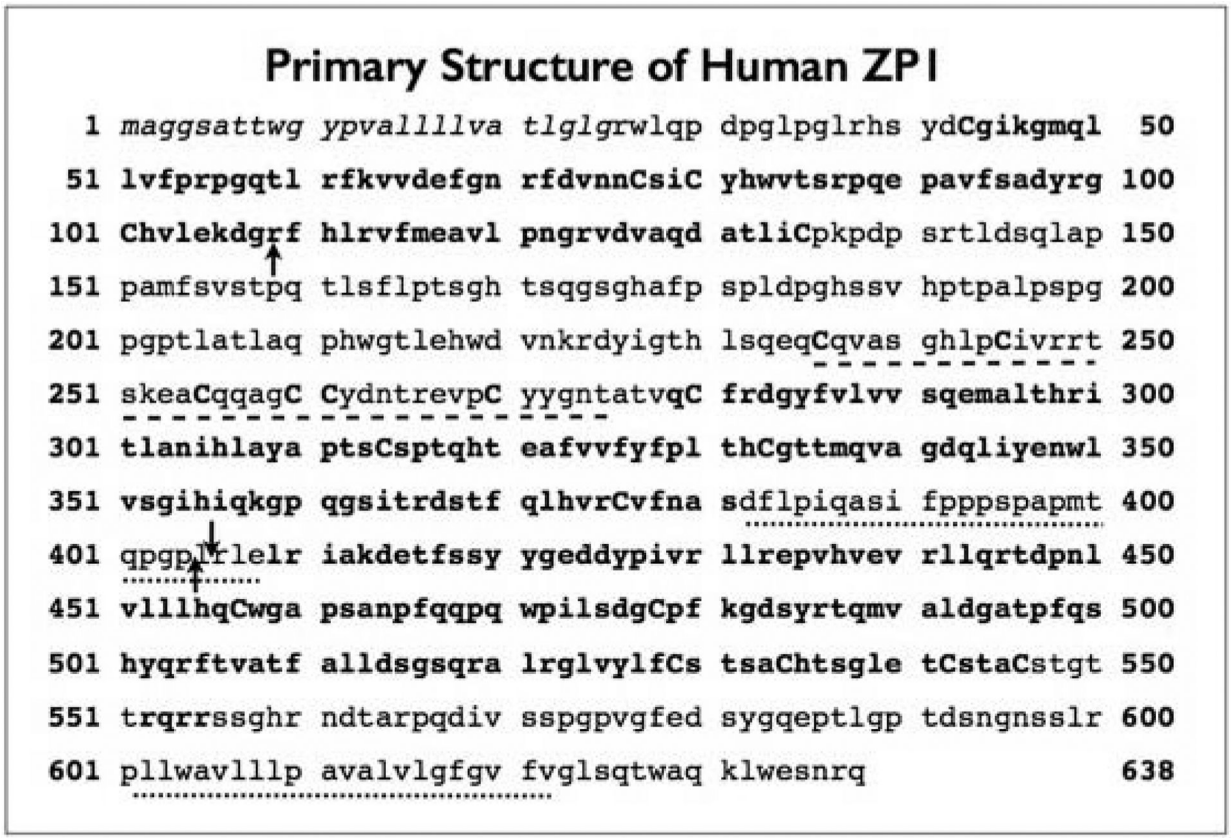


**Figure 4.**

Primary structure of human (h) ZP1 (NCBI Reference Sequence: NP_997224.2) using the single letter amino acid code. The polypeptide consists of 638 aa and has a signal sequence (SS; aa 1–25), a ZP-N1 (aa 43–135) with 5 Cys residues (*bold*), a proline-rich region (aa 136–235), a trefoil domain (TD; aa 236–275) with 6 Cys residues (*bold*), a ZPD (aa 279–546) consisting of a ZP-N (aa 279–381) with 4 Cys residues (*bold*), a ZP-C (aa 409–546) with 6 Cys residues (*bold*), and a linker region (aa 382–408; internal hydrophobic patch, aa 402–408), a concensus furin cleavage-site (CFCS; aa 552–555), an d a transmembrane domain (TMD; aa 602–622). The SS is *italicized*, ZP-N1 is in *bold*, the TD is indicated by a *dashed line*, ZP-N and ZP-C are in *bold*, the linker region is indicated by a *dotted line*, the CFCS (*bold;* aa 552–555), and the TMD is indicated by a *dotted line*. Three *black arrows* indicate the positions of gene sequence variations (GSV) that result in changes in ZP1 that are proposed to cause female infertility. One arrow indicates the replacement of Arg-109 with His (R109H; *exon-3,* N1 before TD) and the two other arrow*s* indicate the insertion of an SC that results in the synthesis of truncated ZP1 that lacks the ZP-C (aa 409–546) and CTP (aa 547–638).

**Table 1.**

Mouse (m) and human (h) ZP genes.

| ZP gene | Chromosome location (No.) | Gene length (kb) | Number exons |
|---|---|---|---|
| mZP1 | 19 | 6.5 | 12 |
| hZP1 | 11 | 8.1 | 12 |
| mZP2 | 7 | 18.5 | 18 |
| hZP2 | 16 | 14 | 19 |
| mZP3 | 5 | 8.6 | 8 |
| hZP3 | 7 | 18.3 | 8 |
| mZP4 | 13 | - | - |
| hZP4 | 1 | 17 | 12 |

**Table 2.**

Primary structure of mouse (m) and human (h) ZP proteins.

| ZP protein | Polypeptide length (aa) | Signal sequence (aa) | ZP domain (aa) | Consensus furin cleavage-site (aa) | Transmembrane domain (aa) | Trefoll domain (aa) |
|---|---|---|---|---|---|---|
| mZP1 | 623 | 1–20 | 271–542 | 545–548 | 591–611 | 225–266 |
| hZP1 | 638 | 1–25 | 279–549 | 552–555 | 602–622 | 234–274 |
| mZP2 | 713 | 1–34 | 364–630 | 632–635 | 684–703 | - |
| hZP2 | 745 | 1–38 | 372–637 | 639–642 | 717–736 | - |
| mZP3 | 424 | 1–22 | 45–304 | 350–353 | 388–408 | - |
| hZP3 | 424 | 1–22 | 45–303 | 349–352 | 387–409 | - |
| hZP4 | 540 | 1–19 | 188–462 | 463–466 | 505–526 | 141–183 |

**Table 3.**

Mouse (m) and human (h) ZP proteins.

| ZP protein | % Similarity | % Identity |
|---|---|---|
| mZP1/hZP1 | 84 | 67 |
| mZP2/hZP2 | 80 | 58 |
| mZP3/hZP3 | 87 | 68 |

**Table 4.**

Phenotypes of *ZP1, 2, 3* null female mice.

| Genotype | Fertility | Zona pellucida |
|---|---|---|
| Wild-type | Fertile | Normal |
| ZP1$^{-/-}$ | Reduced fertility | Abnormal |
| ZP2$^{-/-}$ | Infertile | None |
| ZP3$^{-/-}$ | Infertile | None |
| ZP3$^{+/-}$ | Fertile | Thin |

**Table 5.**

ZP1, 2, and 3 mutations in infertile human patients.

| Designation: hZP1 mutations | Location of mutation | Status of zona pellucida | Reference |
|---|---|---|---|
| G57Dfs*9 | *exon-1,* SC in N1 before TD | none | [68] |
| R61C | *exon-1,* N1 before TD | none (?) | [67] |
| W83R | *exon-2,* N1 before TD | abnormal/none | [66] |
| R109H | *exon-3,* N1 before TD | none | [64] |
| H170Ifs*52 | *exon-3,* SC between N1 and TD | none | [69] |
| Q292→X | *exon-5,* SC in ZP-N | none | [65] |
| I386→X | *exon-7,* SC between ZP-N and ZP-C (linker) | none | [65] |
| I390fs404X | *exon-7,* SC between ZP-N and ZP-C (linker) | none | [58, 63] |
| I390Tfs*16 | *exon-7,* SC between ZP-N and ZP-C (linker) | none | [67, 68] |
| R410W | *exon-7,* SC between ZP-N and ZP-C (linker) | none | [69] |
| W471→X | *exon-8,* SC in ZP-C | abnormal/none | [66] |
| C478→X | *exon-9,* SC in ZP-C | none | [69] |
| V570M | *exon-11,* between CFCS and EHP | none | [69] |
| D592Gfs*29 | *exon-12,* SC between CFCS and TMD | none | [69] |
| **hZP2 mutations** | | | |
| C372S | *exon-11,* ZP-N | thin/none | [69] |
| R533S | *exon-15,* ZP-C | abnormal/none | [66] |
| C566R | *exon-16,* ZP-C | abnormal/none | [66] |
| R698→X | *exon-19,* SC between CFCS and TMD | very thin/none | [70] |
| **hZP3 mutations** | | | |
| A134T | *exon-2,* ZP-N | none | [64, 71] |
| R255G | *exon-5,* ZP-C | none | [69] |
| R349L→X | *exon-8,* SC at CFCS | very thin/none | [70] |

**Abbreviations:** CFCS, concensus furin cleavage-site; EHP, external hydrophobic patch; h, human; SC, stop-codon; TD, trefoil domain; TMD, transmembrane domain; ZPD, zona pellucida domain; ZP-C, ZPD carboxy-terminal subdomain; ZP-N, ZPD amino-terminal subdomain.