Trends in Developmental Biology
Vol. 8, 2014

Mini-Review

Evolution, structure, and synthesis of vertebrate egg-coat proteins

Eveline S. Litscher[1,*] and Paul M. Wassarman[2,*]

Department of Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1020, New York, NY 10029-6574, USA.

ABSTRACT

All vertebrate eggs are surrounded by an extracellular coat that supports growth of oocytes, protects oocytes, eggs, and early embryos, and participates in the process of fertilization. In mammals (platypus to human beings) the coat is called a zona pellucida (ZP) and in non-mammals (molluscs to birds), a vitelline envelope (VE). The ZP and VE are composed of just a few proteins that are related to one another and possess a common motif, called the zona pellucida domain (ZPD). The ZPD arose more than ~600 million years ago, consists of ~260 amino acids, and has 8 conserved Cys residues that participate in 4 intramolecular disulfides. It is likely that egg-coat proteins are derived from a common ancestral gene. This gene duplicated several times during evolution and gave rise to 3-4 genes in fish, 5 genes in amphibians, 6 genes in birds, and 3-4 genes in mammals. Some highly divergent sequences, N- and C-terminal to the ZPD, have been identified in egg-coat proteins and some of these sequences may be under positive Darwinian selection that drives evolution of the proteins. These and other aspects of egg-coat proteins, including their structure and synthesis, are addressed in this review.

KEYWORDS: eggs, zona pellucida, vitelline envelope, *ZP* genes, ZP proteins, ZP domain

*Corresponding authors
[1]eveline.litscher@mssm.edu
[2]paul.wassarman@mssm.edu

INTRODUCTION

"*Evolutionary scenarios are an artform. They usefully exercise the brain, causing us to look at old data in new ways and stimulating us to collect new data. They do not have to be true*". W. Ford Doolittle, 2006 [1].

A relatively thick extracellular matrix or coat surrounds vertebrate eggs. In fish, amphibians, and birds the egg-coat is called a vitelline envelope (VE) and in mammals a zona pellucida (ZP) [2]. Vertebrate egg-coats differ considerably in size, but are quite uniform in composition since they consist of only a few proteins that possess a common motif, the zona pellucida domain (ZPD). ZP and VE proteins are a conserved group that share high sequence identity and have many structural features in common. The proteins assemble into fibrils and then into the matrix (egg-coat) that surrounds growing oocytes. Egg-coats protect oocytes, eggs, and early embryos, support oocyte growth, and play important roles during fertilization.

The mammalian ZP consists of 3-4 proteins, called ZP1-4 [3]. The amphibian and bird VE consists of 4-6 proteins, called ZP1-4, ZPd, and ZPax, and in many fish the VE consists of 2-4 proteins, called ZP1, ZP3, ZPax, and variants of ZP1 and ZP3 [4-6]. It has been proposed that ZP proteins are derived from a common ancestral gene, possibly *ZP3*, and that ancestral ZP3 protein functioned as an egg-coat protein. Perhaps a first duplication event of an ancestral *ZP3* gave rise to *ZP3* and a precursor to other *ZP* genes which

subsequently duplicated several times and evolved into all other *ZP* genes.

In recent years, the repertoire of egg-coat proteins has been expanded by whole-genome sequencing and provided amino acid (aa) sequences for ZP proteins from non-traditional vertebrate models. These models include platypus (monotremata) [7], lizard [8], turtle [9], snake (reptilia) [10], shark (cartilaginous fish) [11], lamprey (jawless fish) [12], as well as amphioxus (fish-like chordate) [13]. The lamprey is thought to belong to the oldest living group of vertebrates (> 350 million years old) and probably is a good candidate with which to address questions concerning the evolutionary origin and duplication events of ancestral *ZP* genes.

Here we discuss some characteristic traits of egg-coats, *ZP* genes, and ZP proteins from vertebrate species that represent fish (> 40,000 species), amphibians (> 15,000 species), birds (> 10,000 species), and mammals (> 5,000 species), as well as some non-traditional vertebrate models (e.g., shark, turtle, snake, and platypus). Emphasis is placed on shared features and functions that indicate a common origin of ZP proteins dating back more than 500 million years [2, 14]. We also point out some of the characteristic structural features of the ZPD and discuss aspects of the synthesis and assembly of ZP proteins into fibrils and matrix (egg-coats).

Vertebrate eggs and egg-coats

Vertebrate oocytes, eggs, and embryos, from fish to human beings, are surrounded by an egg-coat. Fish and amphibian embryos hatch from the VE as free-swimming larvae and tadpoles, respectively, whereas among birds a newly hatched chick emerges that is fully developed. Mammalian embryos are surrounded by a ZP until the early blastocyst stage of development when they hatch from the ZP and implant in the uterus (with the exception of monotremes, such as the platypus, that lay eggs) [2, 15].

Mammalian eggs are ~100 µm in diameter with not much more than a 2-to-1 variation in size for eggs from different mammals (with the exception of monotremes, such as platypus, whose eggs are 20-30 times larger) [16-19]. Similarly, the ZP of various mammalian eggs averages ~15 µm in width, from ~6 µm in mice to ~20-25 µm in human beings [20-23]. Platypus eggs are an exception since they have a very thin ZP, ~1 µm in width, and the eggs are protected by a leathery shell.

Bird eggs are very large compared to mammalian eggs due to the accumulation of vast amounts of yolk. However, the region of the egg that gives rise to the bird embryo lies on the surface of the yolk and is microscopic. The VE of bird (e.g., chicken) eggs consists of 3 layers, of which the innermost layer, called the pervitelline layer (PVL), is a relatively thin matrix. The PVL is a functional homolog of the mammalian ZP and the amphibian and fish VE [24].

Amphibian eggs typically are deposited in water in large groups or clusters and hatch within a relatively short time as free-swimming larvae. The eggs possess a VE that varies in thickness, from ~1 µm for *Xenopus* to ~8-15 µm for other amphibians. The eggs also have an additional gelatinous layer of protection, called the jelly coat [25].

Fish eggs are laid down in large numbers. They are surrounded by a VE that often consists of a thin outer layer and a thicker inner layer. The thin outer layer is possibly the result of crosslinking of VE proteins through their N-terminal proline-rich region (e.g., trout egg VE) [26-28]. Two kinds of fish eggs have been described; pelagic floating eggs that are highly hydrated and possess a "smooth" outer surface VE, and non-buoyant benthic eggs that have a "sticky" outer surface VE used to adhere to substrates [29]. Amphioxus (a fish-like marine chordate) eggs are ~150 µm in diameter and are surrounded by a VE that is ~6 µm thick [30].

Although the diameter of eggs and width of egg-coats can vary considerably among different organisms (Table 1), the relative elasticity and fibrillar nature of the egg-coat are characteristic features of all vertebrate VEs and ZPs.

Evolution of *ZP* genes

A better understanding of the evolution of *ZP* genes and proteins has recently emerged. For example, comparisons of gene structure and organization made it possible to locate, in different species, comparable *ZP* genes (Table 2) and their adjacent regions within portions of chromosomes

Table 1. Approximate egg-coat widths and egg diameters.

Organism	Egg-coat width (µm)	Egg diameter (µm)
Trout[a]	25-50	3,000-6,000
Frog[b]	1	1,000
Chicken	1-3.5	30,000
Platypus[c]	1	4,000
Possum[d]	6	230
Mouse	6	80
Rabbit, Sheep, Cat	15	120-130
Cow	16	120
Pig	16	130
Baboon[e]	13	-
Human	20-25	100-120

- not found.
[a]*Oncorhynchus mykiss*.
[b]*Xenopus laevis*.
[c]*Ornithorhynchus anatinus*.
[d]*Trichosurus vulpecula*.
[e]*Papio hamadryas*.

(synteny) [6]. ZP protein homologs are classified as either orthologs (e.g., mouse ZP3 vs chicken ZP3) or paralogs (e.g., mouse ZP3 vs mouse ZP2). It is likely that ZP proteins are derived from a common ancestral gene, possibly by an initial duplication event hundreds-of-millions of years ago. This event gave rise to *ZP3* in one branch and the *ZPd*, *ZPax*, *ZP2*, and *ZP1/ZP4* gene families, in this order, in other branches [4, 31]. *ZP1/ZP4* has been cited as an example of duplication of a single gene because adjacent regions of these two genes in the mouse are located on the same chromosome (number 19) [6].

ZP1-4 are found in mammals and other vertebrates, *ZPd* in amphibians and birds, and *ZPax* in fish, amphibians, and birds (Table 2) [6, 15]. The absence of *ZPd* and *ZPax* in mammals suggests that these genes were lost during evolution, as exemplified by non-functional *ZPax* (a pseudo-gene) found in some mammals (e.g., human beings, chimpanzees, macacas, and cows). In addition, *ZP1* and *ZP4* have been identified as pseudo-genes in several mammalian species (e.g., *ZP1* in dogs and cows, and *ZP4* in mice). This suggests that recent *ZP* gene evolution may have occurred by "gene death" [32].

Mammals, amphibians, and birds have 3-4, 5, and 6 *ZP* genes, respectively. Many fish (teleosts) have 2-4 *ZP* genes, whereas some fish (e.g., zebrafish) have multiple copies or tandem repeats of *ZP1* and *ZP3* genes.

ZP1-4 genes are also present in non-traditional vertebrate models, such as shark [11], turtle [9], snake [10], and platypus [7] (Table 3). The aa sequences of these ZP proteins have been derived by conceptual translation from whole-genome sequences and have not been characterized further. For amphioxus, 5 ZP proteins have been identified as components of its egg-coat (called BbZP1-5) and all have a type-II ZPD [30] (Figure 1b). Lamprey *ZP* genes have been predicted based on their similarity to fish *ZP* genes and identified as ZP protein homologs [30]. However, the predicted proteins are not full-length due to incomplete coverage of the genome.

Sequence identity of ZP proteins

The ZP/VE of vertebrate eggs, from fish to human beings, consists of only a few ZP proteins, called ZP1-4, ZPd, and ZPax. ZPd is only found in

Table 2. Location of *ZP* genes (Chromosome no.)*.

Organism	***ZP1***	***ZP2***	***ZP3***	***ZP4***	***ZPd***	***ZPax***
Trout	x		x			(-)[a]
Frog	-	x	x	x	x	x
Chicken	5	14	10	6	11	3
Possum	-	x	x	x		
Mouse	19	7	5	(13)[b]		
Rat	1	1	12	17		
Cow	(29)[c]	25	25	28		pseudo[d]
Pig	-	3	3	14		
Dog	(18)[c]	6	6	4		
Macaca[e]	14	20	3	1		pseudo[d]
Chimpanzee[f]	11	16	7	1		pseudo[d]
Human	11	16	7	1		pseudo[d]

*Taken from NCBI database.
x No information about location of *ZP* genes.
- Not found.
[a]In some fish, e.g. gilthead seabream, a ZP2 homolog (ZPx) is present.
[b]*ZP4* is a pseudo-gene in mice.
[c]*ZP1 i*s a pseudo-gene in cows and dogs.
[d]*ZPax* is a pseudo-gene in cows, macacas, chimpanzees, and humans.
[e]*Macaca mulatto*.
[f]*Pan troglodytes*.

Table 3. *ZP* genes of non-traditional vertebrate models.

Organism	***ZP* genes**
Amphioxus[c]	(5 *ZP* genes)[a]
Lamprey[d]	(*ZP2, ZP4, ZPax*)[b]
Shark[e]	*ZP1-4, ZPd*
Turtle[f]	*ZP1-4*
Snake[g]	*ZP1-3*
Platypus	*ZP1-4*

[a]Named 5 ZP proteins based on mass spectrometric analyses.
[b]ZP proteins not full-length and tentatively named.
[c]*Branchiostoma belcheri.*
[d]*Petromyzon marinus.*
[e]*Callorhynchus milii.*
[f]*Chelonia mydas.*
[g]*Ophiophagus hannah.*

Xenopus and chicken, and ZPax is found in some fish, *Xenopus*, and chicken. ZP3 is always the smallest of the four ZP proteins and ZP4 is always smaller than ZP1 (Table 4).

Sequences of vertebrate ZP proteins (ZP1-4) have been compared with orthologs from human beings and, although aa sequences between species vary in length, there is a high degree of sequence identity in overlapping sequences (Table 5). The average % identity suggests that the twelve organisms examined can be divided into four groups, I-IV: I - trout, 33% ave. identity; II - frog, chicken, and possum, 43-51% ave. identity; III - mouse, rat, cow, pig, and dog, 64-69% ave. identity; IV - macaca and chimpanzee, 93-99% ave. identity. It has been suggested that proteins with sequence identities of 40% or more perform the same function (ZP proteins in groups II, III, and IV) and those with identities of 25-40% perform similar functions

Table 4. Sizes of polypeptide precursors (No. aa residues)*.

Organism	ZP1	ZP2	ZP3	ZP4	ZPd	ZPax
Trout	563/524[a]		441			(698)[b]
Frog		699	460	544	376	905
Chicken	934	695	446	543	418	837
Possum		712	422	527		
Mouse	623	713	424			
Rat	617	695	424	545		
Cow		713	421	534		
Pig		716	421	536		
Dog		715	426	531		
Macaca	640	745	424	539		
Chimpanzee	638	745	424	540		
Human	638	745	424	540		

*Taken from NCBI database.
[a]ZP1α/ZP1β.
[b]Gilthead seabream ZPx.

Table 5. Comparison of vertebrate and human ZP1-4.

	Average % identity[a]			
Organism	**hZP1[b]**	**hZP2**	**hZP3**	**hZP4**
Trout	33/31		36 (40)[c]	
Frog		42	44 (49)	42
Chicken	45	44	48 (58)	52
Possum		54	46 (54)	54
Mouse	67	58	68 (75)	
Rat	68	58	67 (73)	64
Cow		67	72 (82)	69
Pig		63	73 (84)	67
Dog		67	71 (80)	69
Macaca	94	93	93 (97)	93
Chimpanzee	99	99	99 (100)	99

[a]Whole sequences of ZP proteins (ZP1, 2, 3 and 4) are compared with sequences of human ZP1-4 (see Table 4). The numbers represent % identity of each of the ZP proteins with the corresponding human ortholog. Note that aa sequences vary in length.
[b]Human ZP proteins are designated with an h.
[c]Comparison of ZPD sequences of ZP3. Note that the % identity of the ZPDs is 1-11% higher than the % identity of whole sequences.

(ZP proteins in group I) [33]. In groups II-IV, with sequence identities of 40% or more, ZP proteins are structural and sperm-binding proteins, whereas in group I, with sequence identities of 25-40%, ZP proteins play only a structural role (e.g., fish sperm do not bind to the VE, but enter the egg via an opening in the VE, called the micropyle).

Structure of the ZPD

A structural element present in ZP proteins and in a variety of other proteins was defined by pattern-based sequence analysis and called the "zona pellucida domain" (ZPD) [34]. It was suggested that the ZPD has a common tertiary structure and plays a common biological role. Proteins with the same domain combination tend to share an ancestor and have functional features in common [35]. The ZPD arose more than ~600 million years ago and is found in a variety of animal species, in vertebrates and invertebrates, and in a variety of tissues and organs. It is a component of all ZP proteins and is present in hundreds of other proteins [36]. The latter include secreted proteins that function as receptors or mechanical transducers between adjacent cells, in intermediators during differentiation, morphogenesis and signaling, or in proteins of extracellular matrices other than egg-coats.

A ZPD consists of ~260 aa and has 8 conserved Cys residues that participate in 4 intramolecular disulfides. It is a bi-partite structure consisting of 2 sub-domains, ZP-N (~120 aa) and ZP-C (~130 aa), linked by a protease-sensitive region (Figure 1a). Each sub-domain has 4 conserved Cys residues, however, the ZP-C sub-domain can have 2 additional Cys residues. ZP-N and ZP-C share a common immunoglobulin (Ig)-like topology with β-sheet arrangements symmetrical to each other, although the sub-domains have significantly different primary structures and intramolecular disulfide bonds [37]. There are two types of ZPDs: Type-I (ZP3-like) with 8 Cys residues and type-II (ZP1/2-like) with 10 Cys residues. The type-I ZPD has a ZP-N sub-domain with 4 Cys residues, linked 1,4 and 2,3, and a ZP-C sub-domain with 4 Cys residues, linked 5,7 and 6,8 (Figure 1b). The type-II ZPD has a ZP-N sub-domain with 4 Cys residues, linked 1,4 and 2,3, and a ZP-C sub-domain with 6 Cys residues, linked 5,6, 7,a and b,8 (Figure 1b).

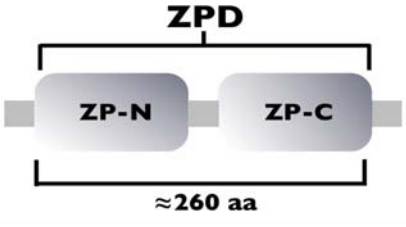

Figure 1a. Schematic representation of the ZPD composed of a ZP-N and a ZP-C sub-domain.

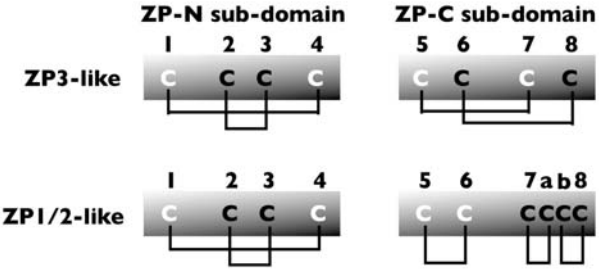

Figure 1b. Schematic representation of intramolecular disulfides of ZP3-like (type I) and ZP1/2-like (type II) ZPD proteins.

There are exceptions to the two types of ZPD described above, including ZPDs with 12 Cys residues. The two additional Cys residues, referred to as Cx and Cy and linked x,y, are present in fish (e.g., trout) ZP1-like proteins (ZP1α and ZP1β) and are located between the ZP-N and ZP-C sub-domains [27]. In chicken and pig ZP3, Cys residue clustering in ZP-C sub-domains is variable [37]; there is a disulfide linkage to a Cys residue C-terminal to the ZPD, linked 6,11 and 8,9, whereas in mouse ZP3 it is 6,8 and 9,11 (Cys residues 9 and 11 are located downstream of the ZPD). It is likely that the additional disulfide linkage in trout and the different disulfide bonding patterns in chicken and pig cause the polypeptides to adopt different conformations, as compared to ZP1/2-like and ZP3-like proteins, which in turn could alter the specificity of egg coat assembly in these species.

Function of the ZPD and adjoining regions

Protein domains like the ZPD are evolutionary units that can be duplicated and recombined, and typically, pairs of domains are found in one sequential order (A→B or B→A), almost never in

both [33]. Considering the two ZP sub-domains, ZP-N and ZP-C, as separate units, the sequential order is always ZP-N→ZP-C and never ZP-C→ ZP-N, although the ZP-N sub-domain can be present in the absence of ZP-C.

The ZPD is common to all vertebrate egg-coat proteins and the domain conservation is reflected in the high percentage of its sequence identity within vertebrate ZP proteins (Table 5). It has been proposed that the ZPD is a conserved module for polymerization of extracellular proteins (e.g., to form ZP fibrils) [38, 39] and it is likely that the bi-partite structure of the ZPD (ZP-N and ZP-C sub-domains) endows it with dual functions. For example, the ZP-N sub-domain can serve as an independent structural domain in the absence of ZP-C (e.g., in proteins Oosp1 [40], Plac1 [41], and papillote [42]), and the ZP-N sub-domain alone has been shown to be an active folding unit that can polymerize into fibrils [43]. In addition, divergent copies of ZP-N sub-domains are found in single or multiple copies in the N-terminal regions of ZP1, ZP2, ZP4, and ZPax [44]. On the other hand, no protein has been found to consist only of the ZP-C sub-domain which suggests that its role is dependent on its partner ZP-N. Perhaps ZP-C plays a regulatory role during assembly of ZPD protein complexes.

N- and C-terminal to the ZPD are highly divergent aa sequences that have been identified in mouse and chicken ZP3 [45, 46]. For example, in mouse ZP3 there are 2 clusters of sites N-terminal (aa 25-50) and C-terminal (aa 331-373) (Table 6) to the ZPD and, similarly, in chicken ZP3 (aa 2-30, aa 381-436), that apparently are under positive Darwinian selection (including a few mutations within the ZPD). Regions that are N- and C-terminal to the polymerization module are more likely to be involved in species-specific functions (e.g., sperm-egg interactions) and it is possible that selective pressure drives the mutations in these regions and thereby promotes divergence and evolution of ZP3 proteins. Mutations within the ZPD module and adjoining regions can also affect post-translational modifications, such as N- (Asn-X-Ser/Thr) or O- (Ser,Thr) linked glycosylation. For example, 2 potential O-linked glycosylation sites (Ser331 and Ser333) are within the highly divergent C-terminal cluster of mouse ZP3 and might promote divergence between mouse ZP3 and other closely related mammalian species due to differences in glycosylation pattern.

In general, the ZPD module is a conserved evolutionary unit essential for polymerization of proteins [38], whereas adjoining regions contribute functional diversification that may be caused by selective pressure related to species-specific functions.

ZP protein synthesis, secretion, and assembly

ZP proteins are secreted proteins and as such are synthesized as precursor polypeptides with

Table 6. Divergent ZP3 sequences in mammals.

	↓ ↓ ↓↓ ↓ ↓ ↓
Mouse	$_{331}$SSS<u>S</u>QFQIHGPRQWSKLV
Rat	$_{331}$SSS<u>S</u>EFETHEPAQWSTLV
Cow	$_{329}$SGR<u>S</u>MRLSH--RE-GRPV
Pig	$_{328}$PSL<u>S</u>R-KLSMPKRQS--A
Dog	$_{325}$PGR<u>S</u>RRLSHLERGWRRSV
Macaca	$_{330}$PSH<u>S</u>RRQPHVVSQWSRSA
Chimpanzee	$_{330}$PSH<u>S</u>RRQPHVVSQWSRSA
Human	$_{330}$PSH<u>S</u>RRQPHVMSQWSRSA

Highly divergent aa's (↓) in mouse ZP3 as compared to homologous C-terminal sequences of rat, cow, pig, dog, macaca, chimpanzee, and human. Note that mouse S_{334} (underlined) is conserved in all eight species.

Table 7. Site(s) of ZP protein synthesis.

Source	Ovary[a]	Liver
Fish[b]	+	+
Amphibians[c]	+	-
Birds[d]	+	+
Mammals[c]	+	-

[a]Oocytes and/or granulosa cells.
[b]Rainbow trout - liver; zebrafish - ovary; medaka - liver and oocytes.
[c]*Xenopus*, mice, humans - oocytes.
[d]Chicken - liver and ovary.

a signal sequence (SS) at the N-terminus and a C-terminal propeptide (CTP) at the C-terminus. They have a ZPD with 8 or more conserved Cys residues present as intramolecular disulfides, a consensus furin-cleavage site (CFCS) located close to the C-terminus, and a transmembrane domain (TMD) downstream of the CFCS (some fish lack a TMD or a hydrophobic C-terminus). ZP1 and ZP4 also have a trefoil domain adjacent to the ZPD.

For mammals and amphibians the ovary is the only site of synthesis of ZP proteins [47, 48] (Table 7). However, in fish that release several tens-of-thousands of relatively small eggs at a time, and in birds that have very large, yolk-laden eggs, there are two sites of ZP protein synthesis, the ovary and the liver [49-52] (Table 7). Perhaps the additional sites of ZP protein synthesis in fish and birds reflect the necessity of synthesizing large amounts of ZP proteins in a relatively short time. For some fish, it might therefore be advantageous to synthesize ZP proteins in the liver. On the other hand, small fish like zebrafish release relatively small clutches of eggs (~hundreds) and do not express *ZP* genes in the liver. In birds, ZP proteins are synthesized in the liver and in granulosa cells that surround growing oocytes. A fully-grown chicken oocyte is ~200-300 times larger in diameter than a mammalian egg and thus the ZP (although thinner than a mammalian ZP) covers an area that is tens-of-thousands times larger than the mammalian counterpart.

All vertebrate ZP proteins are synthesized as precursor polypeptides that have a SS and a CTP downstream of the ZPD, and it has been suggested that sequences within the CTP are required for secretion and assembly of ZP proteins [48, 53]. The CTP includes a CFCS, a hydrophobic peptide, called an external hydrophobic patch (EHP), a TMD, followed by a short hydrophilic C-terminus. Another hydrophobic peptide, called an internal hydrophobic patch (IHP), is present within the ZPD (ZP-N or ZP-C sub-domain). The IHP and EHP are well conserved units within vertebrate ZP proteins as revealed in sequence alignments of ZP1-4 orthologs from trout, frog, chicken, possum, mouse, and human beings (with the possible exception of trout ZP1α) (Table 8). Based on experiments with mouse ZP3 it has been suggested that IHP-EHP elements regulate assembly of ZP proteins [54]. In ZP protein precursors the IHP interacts with the EHP thereby preventing premature polymerization and assembly of ZP fibrils and matrix within oocytes. When the CTP is removed by proteolysis at the CFCS during secretion of ZP precursor proteins, the EHP dissociates from the IHP, “activated” ZP protein is released from the oocyte, and ZP assembly (i.e., formation of fibrils and matrix) can take place [36].

In fish and birds some ZP precursor proteins are synthesized in the liver and are transported via circulating blood to the oocyte for uptake (similar to the synthesis of egg-yolk precursor, vitellogenin, in the liver of amphibians and birds [55]) [56]. It is not known how these proteins are transported to their site of assembly or how they are assembled into an egg-coat. Synthesis of individual ZP proteins must be coordinated in a timely manner to ensure that they are present simultaneously so that they can assemble first into fibrils and then into matrix. It is possible that ZP precursors are transported in the bloodstream in small vesicles, perhaps bound to specific receptors, that dock on to the oocyte’s plasma membrane for uptake into the oocytes. The precursor proteins would then follow the biosynthetic pathway described above.

Final comments

During the past 30 years or so, identification and characterization of vertebrate egg-coat genes and proteins, from organisms as different as fish and human beings, has come into its own. ZP proteins

Table 8. Conservation of IHP and EHP[a].

		IHP	EHP
ZP1/4			
Trout	zp1α	$_{370}$DAVLHVE$_{376}$	$_{551}$SQKVIMI$_{557}$
	zp1β	$_{330}$PGPLIVE$_{336}$	$_{511}$SGELILT$_{517}$
Frog	zp1	-	-
	zp4	$_{308}$PGPLMLE$_{314}$	$_{480}$DGPVDFI$_{486}$
Chicken	zp1	$_{746}$PGPLQLQ$_{752}$	$_{920}$RGRIVLP$_{926}$
	zp4	$_{321}$PGPLSLE$_{327}$	$_{491}$KGPVIFL$_{497}$
Possum	zp1	-	-
	zp4	$_{309}$PGPLALE$_{315}$	$_{479}$QGPIFFL$_{485}$
Mouse	zp1	$_{394}$SGPLRLE$_{400}$	$_{566}$PGAVGFE$_{572}$
	zp4	-	-
Human	zp1	$_{402}$PGPLRLE$_{408}$	$_{573}$PGPVGFE$_{579}$
	zp4	$_{312}$PGPLTLE$_{318}$	$_{482}$KGPMILL$_{488}$
ZP2			
Gilthead seabream[b]		$_{521}$RGELQIT$_{527}$	$_{691}$SGPILVN$_{698}$
Frog		$_{468}$DGPLTLV$_{474}$	$_{634}$SGPILIV$_{640}$
Chicken		$_{467}$QGPLSLI$_{473}$	$_{633}$QGPVLLV$_{639}$
Possum		$_{486}$PGPLSLV$_{492}$	$_{651}$PGPVFLV$_{657}$
Mouse		$_{483}$PGPLVLV$_{489}$	$_{648}$PGPILLL$_{654}$
Human		$_{490}$LGPFTLI$_{496}$	$_{657}$PGPILLL$_{663}$
ZP3			
Trout		$_{231}$YFSMRLMT$_{238}$	$_{424}$WEGDVQLGPIFIS$_{436}$
Frog		$_{198}$AFSLRLMT$_{205}$	$_{390}$EHSLATIGPILVV$_{402}$
Chicken		$_{192}$VFSLRLMS$_{199}$	$_{377}$VAADVVIGPVLLS$_{389}$
Possum		$_{175}$KFSLRLMA$_{182}$	$_{354}$FEADLMLGPLVLS$_{366}$
Mouse		$_{170}$AFSLRLME$_{177}$	$_{357}$DEADVTVGPLIFL$_{369}$
Human		$_{171}$TFSLRLME$_{178}$	$_{356}$EEADVTVGPLIFL$_{368}$

[a]ZP1 and ZP2 were compared and aligned with mouse ZP1 and mouse ZP2 as in [36].
ZP4 was compared and aligned with vertebrate ZP1, and ZP3 sequences were taken from [37].
[b]Gilthead seabream has a ZP2 homolog called ZPx. Rainbow trout has no ZP2 or ZP2 homolog.
Conserved residues are underlined.

apparently are derived from a common ancestral gene that gave rise to *ZP3* and 4 other *ZP* gene families (*ZPd*, *ZPax*, *ZP2*, and *ZP1/ZP4*). One of the gene families (*ZPax*) is found in fish, amphibians, and birds, and another (*ZPd*) in amphibians and birds. *ZPax* is found in a few mammals, but is always present as a pseudo-gene. *ZP1-4* are found in mammals and other vertebrates.

All of the corresponding ZP proteins possess a ZPD that participates in polymerization of processed ZP/VE precursor proteins into higher-order structures, such as fibrils and matrix (egg-coats). The presence of a ZPD in proteins from jellyfish (cnidarians), a member of the oldest of the true metazoan phyla, suggests that it arose more than ~600 million years ago. Furthermore, a ZPD is present in a host of organisms, including echinoderms, worms, and flies, in hundreds of different proteins [2, 36], including TGF-β receptor type III (betaglycan), hensin, vomeroglandin, tectorin-α and -β, uromodulin, mesoglein, DMBT-1, cuticlins, oikosins, DYF-7 and RAM-5, and two-to-three dozen proteins in *Drosophila* [57]. It is very likely that as more and more organisms are subjected to whole-genome sequencing the number and variety of ZPD proteins will continue to increase and evolutionary relationships between the relevant genes/proteins will be clarified further.

ACKNOWLEDGEMENTS

We are grateful to the Bodleian Library, Oxford University, for providing a stimulating and comfortable environment in which to write this review. Our research was supported in part by the NIH (NICHD).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no disclosures.

REFERENCES

1. Koonin, E. V., Senkevich, T. G. and Dolja, V. V. 2006, Biol. Direct., 1, 29.
2. Litscher, E. S. and Wassarman, P. M. 2015, A Guide to Zona Pellucida Domain Proteins. Wiley and Sons, Hoboken.
3. Wassarman, P. M. 2008, J. Biol. Chem., 283, 24285.
4. Spargo, S. C. and Hope, R. M. 2003, Biol. Reprod., 68, 358.
5. Okumura, H., Kohno, Y., Iwata, Y., Mori, H., Aoki, N., Sato, C., Kitajima, K., Nadano, D. and Matsuda, T. 2004, Biochem. J., 384, 191.
6. Smith, J., Paton, I. R., Hughes, D. C. and Burt, D. W. 2005 Mol. Reprod. Dev., 70, 133.
7. Warren, W. C., Hillier, L. W., Marshall Graves, J. A., Birney, E., Ponting, C. P., Grützner, F., Belov, K., Miller, W., Clarke, L., Chinwalla, A. T., Yang, S. P., Heger, A., Locke, D. P., Miethke, P., Waters, P. D., Veyrunes, F., Fulton, L., Fulton, B., Graves, T., Wallis, J., Puente, X. S., López-Otín, C., Ordóñez, G. R., Eichler, E. E., Chen, L., Cheng, Z., Deakin, J. E., Alsop, A., Thompson, K., Kirby, P., Papenfuss, A. T., Wakefield, M. J., Olender, T., Lancet, D., Huttley, G. A., Smit, A. F., Pask, A., Temple-Smith, P., Batzer, M. A., Walker, J. A., Konkel, M. K., Harris, R. S., Whittington, C. M., Wong, E. S., Gemmell, N. J., Buschiazzo, E., Vargas Jentzsch, I. M., Merkel, A., Schmitz, J., Zemann, A., Churakov, G., Kriegs, J. O., Brosius, J., Murchison, E. P., Sachidanandam, R., Smith, C., Hannon, G. J., Tsend-Ayush, E., McMillan, D., Attenborough, R., Rens, W., Ferguson-Smith, M., Lefèvre, C. M., Sharp, J. A., Nicholas, K. R., Ray, D. A., Kube, M., Reinhardt, R., Pringle, T. H., Taylor, J., Jones, R. C., Nixon, B., Dacheux, J. L., Niwa, H., Sekita, Y., Huang, X., Stark, A., Kheradpour, P., Kellis, M., Flicek, P., Chen, Y., Webber, C., Hardison, R., Nelson, J., Hallsworth-Pepin, K., Delehaunty, K., Markovic, C., Minx, P., Feng, Y., Kremitzki, C., Mitreva, M., Glasscock, J., Wylie, T., Wohldmann, P., Thiru, P., Nhan, M. N., Pohl, C. S., Smith, S. M., Hou, S., Nefedov, M., de Jong, P. J., Renfree, M. B., Mardis, E. R. and Wilson, R. K. 2008, Nature, 453, 175.
8. Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C. B., Glor, R. E., Jaffe, J. D., Ray, D. A., Boissinot, S., Shedlock, A. M., Botka, C., Castoe, T. A., Colbourne, J. K., Fujita, M. K., Moreno, R. G., Ten Hallers, B. F., Haussler, D., Heger, A., Heiman, D., Janes, D. E., Johnson, J., de Jong, P. J., Koriabine, M. Y., Lara, M., Novick, P. A., Organ, C. L., Peach, S. E., Poe, S., Pollock, D. D., de Queiroz, K., Sanger, T., Searle, S., Smith, J. D., Smith, Z., Swofford, R., Turner-Maier, J., Wade, J., Young, S., Zadissa, A., Edwards, S. V., Glenn, T. C., Schneider, C. J., Losos, J. B., Lander, E. S., Breen, M., Ponting, C. P. and Lindblad-Toh, K. 2011, Nature, 477, 587.

9. Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., Li, C., White, S., Xiong, Z., Fang, D., Wang, B., Ming, Y., Chen, Y., Zheng, Y., Kuraku, S., Pignatelli, M., Herrero, J., Beal, K., Nozawa, M., Li, Q., Wang, J., Zhang, H., Yu, L., Shigenobu, S., Wang, J., Liu, J., Flicek, P., Searle, S., Wang, J., Kuratani, S., Yin, Y., Aken, B., Zhang, G. and Irie, N. 2013, Nature Genet., 45, 701.
10. Vonk, F. J., Casewell, N. R., Henkel, C. V., Heimberg, A. M., Jansen, H. J., McCleary, R. J., Kerkkamp, H. M., Vos, R. A., Guerreiro, I., Calvete, J. J., Wüster, W., Woods, A. E., Logan, J. M., Harrison, R. A., Castoe, T. A., de Koning, A. P., Pollock, D. D., Yandell, M., Calderon, D., Renjifo, C., Currier, R. B., Salgado, D., Pla, D., Sanz, L., Hyder, A. S., Ribeiro, J. M., Arntzen, J. W., van den Thillart, G. E., Boetzer, M., Pirovano, W., Dirks, R. P., Spaink, H. P., Duboule, D., McGlinn, E., Kini, R. M. and Richardson, M. K. 2013, Proc. Natl. Acad. Sci. USA, 110, 20651.
11. Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., Hoon, S., Gangu, V., Roy, S. W., Irimia, M., Korzh, V., Kondrychyn, I., Lim, Z. W., Tay, B. H., Tohari, S., Kong, K. W., Ho, S., Lorente-Galdos, B., Quilez, J., Marques-Bonet, T., Raney, B. J., Ingham, P. W., Tay, A., Hillier, L. W., Minx, P., Boehm, T., Wilson, R. K., Brenner, S. and Warren, W. C. 2014, Nature, 505, 174.
12. Smith, J. J., Kuraku, S., Holt, C., Sauka-Spengler, T., Jiang, N., Campbell, M. S., Yandell, M. D., Manousaki, T., Meyer, A., Bloom, O. E., Morgan, J. R., Buxbaum, J. D., Sachidanandam, R., Sims, C., Garruss, A. S., Cook, M., Krumlauf, R., Wiedemann, L. M., Sower, S. A., Decatur, W. A., Hall, J. A., Amemiya, C. T., Saha, N. R., Buckley, K. M., Rast, J. P., Das, S., Hirano, M., McCurley, N., Guo, P., Rohner, N., Tabin, C. J., Piccinelli, P., Elgar, G., Ruffier, M., Aken, B. L., Searle, S. M., Muffato, M., Pignatelli, M., Herrero, J., Jones, M., Brown, C. T., Chung-Davidson, Y. W., Nanlohy, K. G., Libants, S. V., Yeh, C. Y., McCauley, D. W., Langeland, J. A., Pancer, Z., Fritzsch, B., de Jong, P. J., Zhu, B., Fulton, L. L., Theising, B., Flicek, P., Bronner, M. E., Warren, W. C., Clifton, S. W., Wilson, R. K. and Li, W. 2013, Nature Genet., 45, 415.
13. Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J. K., Benito-Gutiérrez, E. L., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J. J., Grigoriev, I. V., Horton, A. C., de Jong, P. J., Jurka, J., Kapitonov, V. V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L. A., Salamov, A. A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L. Z., Holland, P. W., Satoh, N. and Rokhsar, D. S. 2008, Nature, 453, 1064.
14. Benton, M. J. and Ayala, F. J. 2003, Science, 300, 1698.
15. Litscher, E. S. and Wassarman, P. M. 2007, Histol. Histopathol., 22, 337.
16. Hartman, C. G. 1929, Quart. Rev. Biol., 4, 373.
17. Hamilton, W. J. and Laing, J. A. 1946, J. Anat., 80, 194.
18. Selwood, L. 2000, Cell Tissue Organs, 166, 208.
19. Gruetzner, F., Nixon, B. and Jones, R. C. 2008, Sex. Dev., 2, 115.
20. Dunbar, B. S., Prasad, S. V. and Timmons, T. M. 1991, A Comparative Overview of Mammalian Fertilization, B. S. Dunbar and M. G. O'Rand (Eds.), Plenum, NY, 97.
21. Selwood, L., Robinson, E. S., Pedersen, R. A. and Vandeberg, J. I. 1997, Intl. J. Dev. Biol., 41, 397.
22. Wassarman, P. M., Qi, H. and Litscher, E. S. 1997, Proc. Roy. Soc. London, Biol. Sci., 264, 323.
23. Mate, K. E. and McCartney, C. A. 1998, Mol. Reprod. Dev., 51, 322.
24. Bellairs, R., Harkness, M. and Harkness, R. D. 1963, J. Ultrastr. Res., 8, 339.
25. Hedrick, J. L. 2008, Intl. J. Dev. Biol., 52, 683.
26. Brivio, M. F., Bassi, R. and Cotelli, F. 1991, Mol. Reprod. Dev., 28, 85.

27. Darie, C. C., Biniossek, M. L., Jovine, L., Litscher, E. S. and Wassarman, P. M. 2004, Biochemistry, 43, 7459.
28. Darie, C. C., Janssen, W. G., Litscher, E. S. and Wassarman, P. M. 2008, Biochim. Biophys. Acta, 1784, 385.
29. Berois, N., Arezo, M. J. and Papa, N. G. 2011, Biol. Res., 44, 119.
30. Xu, Q., Li, G., Cao, L., Wang, Z., Ye, H., Chen, X., Yang, X., Wang, Y. and Chen, L. 2012, BMC Evol. Biol., 12, 239.
31. Claw, K. G. and Swanson, W. J. 2012, Annu. Rev. Genomics Hum. Genet., 13, 109.
32. Goudet, G., Mugnier, S., Callebaut, I. and Monget, P. 2008, Biol. Reprod., 78, 796.
33. Chothia, C., Gough, J., Vogel, C. and Teichmann, S. A. 2003, Science, 300, 1701.
34. Bork, P. and Sander, C. 1992, FEBS Letts., 300, 237.
35. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. and Teichmann, S. A. 2004, Curr. Opin. Struct. Biol., 14, 208.
36. Jovine, L., Darie, C. C., Litscher, E. S. and Wassarman, P. M. 2005, Annu. Rev. Biochem., 74, 83.
37. Han, L., Monné, M., Okumura, H., Schwend, T., Cherry, A. L., Flot, D., Matsuda, T. and Jovine, L. 2010, Cell, 143, 404.
38. Jovine, L., Qi, H., Williams, Z., Litscher, E. and Wassarman, P. M. 2002, Nature Cell Biol., 4, 457.
39. Litscher, E. S., Janssen, W. G., Darie, C. C. and Wassarman, P. M. 2008, J. Cell. Physiol. 214, 153.
40. Yan, C., Pendola, F. L., Jacob, R., Lau, A. L., Eppig, J. J. and Matzuk, M. M. 2001, Genesis, 31, 105.
41. Cocchia, M., Huber, R., Pantano, S., Chen, E. Y., Ma, P., Forabosco, A., Ko, M. S. and Schlessinger, D. 2000, Genomics, 68, 305.
42. Bökel, C., Prokop, A. and Brown, N. H. 2005, J. Cell Sci., 118, 633.
43. Jovine, L., Janssen, W. G., Litscher, E. S. and Wassarman, P. M. 2006, BMC Biochem., 7, 11.
44. Callebaut, I., Mornon, J. P. and Monget, P. 2007, Bioinformatics, 23, 1871.
45. Swanson, W. J., Yang, Z., Wolfner, M. F. and Aquadro, C. F. 2001, Proc. Natl. Acad. Sci. USA, 98, 2509.
46. Calkins, J. D., El-Hinn, D. and Swanson, W. J. 2007, J. Mol. Evol., 65, 555.
47. Yamaguchi, S., Hedrick, J. L. and Katagiri, C. 1989, Dev. Growth Differ., 31, 85.
48. Wassarman, P. M. and Litscher, E. S. 2013, Curr. Top. Dev. Biol., 102, 243.
49. Tesoriero, J. V. 1978, J. Ultrastr. Res., 64, 315.
50. Bausek, N., Waclawek, M., Schneider, W. J. and Wohlrab, F. 2000, J. Biol. Chem., 275, 28866.
51. Arukwe, A. and Goksøyr, A. 2003, Comp. Hepatol., 2, 4.
52. Sano, K., Kawaguchi, M., Watanabe, S., Nagakura, Y., Hiraki, T. and Yasumasu, S. 2013, J. Exp. Zool. (Mol. Dev. Evol.), 320B, 332.
53. Qi, H., Williams, Z. and Wassarman, P. M. 2002, Mol. Biol. Cell, 13, 530.
54. Jovine, L., Qi, H., Williams, Z., Litscher, E. S. and Wassarman, P. M. 2004, Proc. Natl. Acad. Sci. USA, 101, 5922.
55. Wallace, R. A. 1985, Developmental Biology, A Comprehensive Synthesis. Oogenesis, L. W. Browder (Ed.), Plenum, NY, 127.
56. Darie, C. C., Biniossek, M. L., Gawinowicz, M. A., Milgrom, Y., Thumfart, J. O., Jovine, L., Litscher, E. S. and Wassarman, P. M. 2005, J. Biol. Chem., 280, 37585.
57. Plaza, S., Chanut-Delalande, H., Fernandes, I., Wassarman, P. M. and Payre, F. 2010, Trends Cell Biol., 20, 524.