PART A

ZONA PELLUCIDA DOMAIN PROTEINS

A.1 NATURE OF THE ZONA PELLUCIDA DOMAIN

In 1992, Peer Bork and Chris Sander coined the phrase "zona pellucida domain" (ZPD) to define a structural element present in proteins of the zona pellucida (ZP), an extracellular coat that surrounds all mammalian eggs and also present in transforming growth factor type-III receptor and some other receptor-like proteins. The location of the ZPD in these proteins suggested to Bork and Sander that the domain might play a common biological role. The new family of ZPD proteins was defined by pattern-based sequence analysis and it was suggested that this type of domain has a common tertiary structure.

A ZPD consists of ≃260 amino acids (aa) and has eight conserved Cys residues that participate in four intramolecular disulfides. The ZPD is composed of two sub-domains, referred to as ZP-N (≃120 aa) and ZP-C (≃130 aa), that are separated by a short protease sensitive region (Fig. A.1.1). Each sub-domain has four conserved Cys residues. However, the ZP-C sub-domain of some ZPD proteins may have additional Cys residues.

Since its identification more than 20 years ago, a ZPD or a ZP-N sub-domain has been found in hundreds of proteins of diverse functions in a wide variety of organs (e.g., ovary, ear, kidney, heart, liver, brain, pancreas, uterus, etc.; Table C.12.1) and organisms (e.g., jellyfish, sea urchins, worms, mollusks, fruit flies, tunicates, fish, amphibians, reptiles, birds, and mammals; Table E.3). ZPD proteins are frequently

A Guide to Zona Pellucida Domain Proteins, First Edition. Eveline S. Litscher and Paul M. Wassarman.

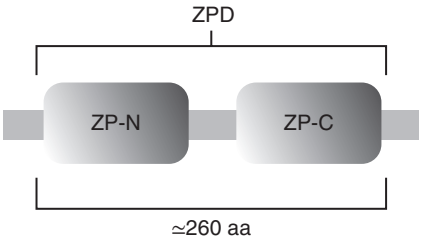

FIGURE A.1.1 Schematic representation of a ZPD. Each ZPD consists of ≃260 aa and the ZP-N and ZP-C sub-domains are connected by a short protease-sensitive linker region.

glycosylated and often display a mosaic architecture since they can consist of a combination of different structural and functional modules (Tables C.12.2 and E.3). ZPD proteins can be secreted into the extracellular space or sometimes be anchored to the cell's plasma membrane by a glycosylphosphatidylinositol (GPI) linkage. ZPD proteins function as structural components of egg coats and other tissues, and as receptors, mechanical transducers, and antimicrobials. They can also play vital roles during differentiation, morphogenesis, and signaling. ZPD proteins are present at the apical surface of many epithelia and participate in the functioning of the senses, including taste and smell. Mutations in genes encoding ZPD proteins can result in severe human pathologies, including deafness, vascular disease, renal disease, cancer, and possibly infertility (Table C.12.3).

FURTHER READING

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

Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona pellucida domain proteins. *Annu Rev Biochem* **74**, 83–114 (2005).

Monné M, Han L, Jovine L. Tracking down the ZP domain: from the mammalian zona pellucida to the molluscan vitelline envelope. *Semin Reprod Med* **24**, 204–216 (2006).

Plaza S, Chanut-Delalande H, Fernandes I, Wassarman PM, Payre F. From A to Z: apical structures and zona pellucida-domain proteins. *Trends Cell Biol* **20**, 524–532 (2010).

A.2 MOUSE ZP PROTEINS

Much of what is known today about ZPD proteins has its origins in early biochemical and molecular genetic studies of the mouse oocyte's ZP. A ZP surrounds all mammalian oocytes, ovulated eggs, and embryos up to the early blastocyst stage of development when embryos hatch from the ZP and implant in the uterus. In mice, the ZP first appears around growing oocytes during the final stages of oogenesis while

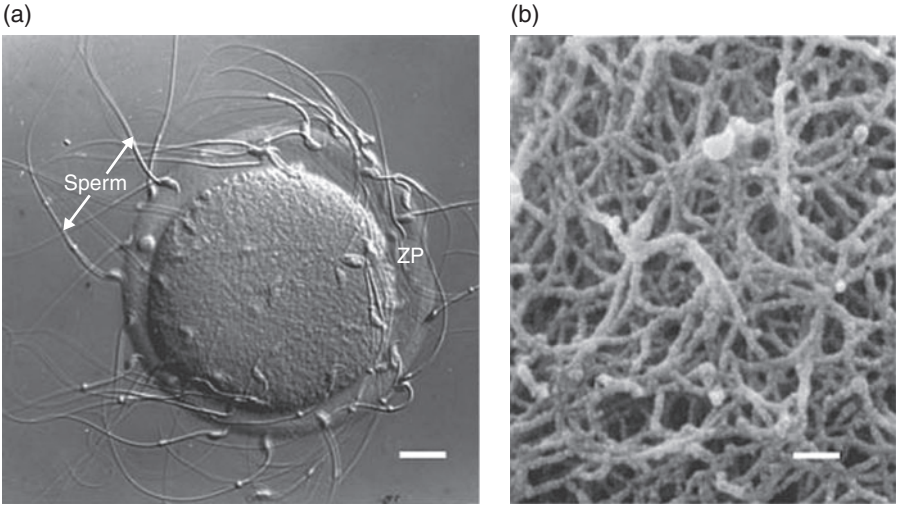

FIGURE A.2.1 Light and electron micrographs of the mouse ZP. (a) Light micrograph of sperm bound to the mouse egg's ZP. Bar ≃13 μm. (b) Scanning electron micrograph of the mouse egg's ZP. Bar ≃200 nm. Reproduced with permission from Wassarman (2008). © Journal of Biological Chemistry.

oocytes are arrested in first meiotic prophase. The ZP increases in thickness as oocytes increase in diameter from ≃12 to ≃80 μm. The ZP of fully-grown oocytes is ≃6 μm thick and contains ≃3.5 ng of protein. Overall, the ZP is a very porous (e.g., permeable to antibodies and viruses) and relatively elastic matrix that is composed of long, interconnected fibrils (Fig. A.2.1).

The mouse ZP is composed of three proteins, called mZP1–3. Together, mZP2 and 3 account for more than 80% of the mass of the ZP and are present in roughly equimolar amounts. mZP1 is the least abundant protein component of the ZP. A fourth ZP protein, mZP4, is missing from the ZP as it is encoded by a pseudogene (pseudogenes are dysfunctional relatives of genes that have lost their protein-coding ability or are no longer expressed). mZP1–3 are heterogeneously glycosylated with asparagine (N) and serine/threonine- (O-) linked oligosaccharides and the oligosaccharides are sialylated and sulfated making the proteins relatively acidic. mZP1, 2, and 3 possess 4, 6, and 5 N-linked oligosaccharides, respectively, and at least two O-linked oligosaccharides are present on mZP3. Under nonreducing conditions, mZP2 and 3 migrate on SDS-PAGE as ≃120 and ≃83 kD MW monomers, respectively, whereas mZP1 migrates as a ≃200 kD MW disulfide-linked homodimer. mZP1 crosslinks individual fibrils that consist of mZP2 and 3 and thereby ensures the structural integrity of the ZP matrix. mZP2 and 3 serve as building blocks of ZP fibrils and also as sperm receptors during fertilization. Modification of both mZP2 and 3 following fertilization renders the ZP refractory to sperm binding.

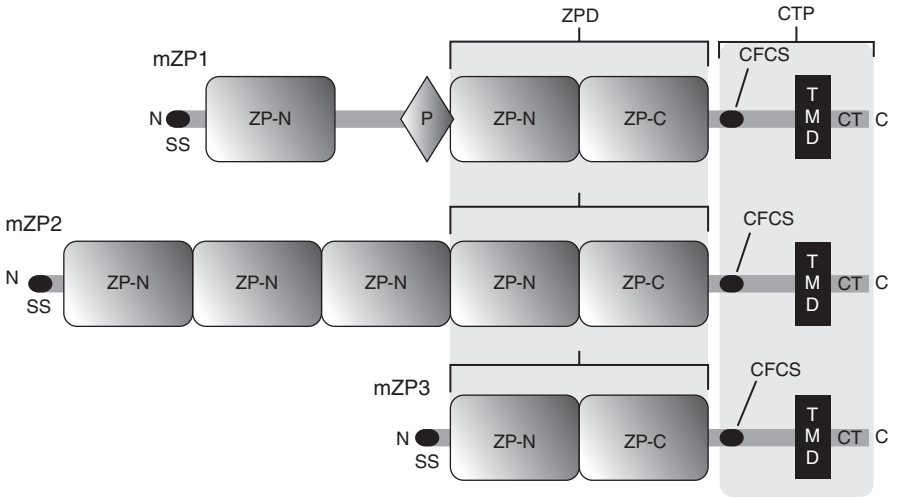

FIGURE A.2.2 Schematic representation of the organization of mZP1, 2 and 3. In each case, the polypeptide contains an SS at the N-terminus, a ZPD, and a CFCS, TMD, and CT at the CTP. mZP1 also has a trefoil (P) domain adjacent to the ZPD and an extra ZP-N sub-domain close to the N-terminus of the polypeptide. mZP2 has three extra ZP-N sub-domains between the ZPD and N-terminus of the polypeptide. mZP3, the smallest of the three proteins, consists primarily of a ZPD.

mZP1–3 are prototypical ZPD proteins. Their nascent precursor polypeptides consist of an N-terminal signal sequence (SS), a ZPD, a C-terminal propeptide (CTP) that has a consensus furin cleavage site (CFCS), a transmembrane domain (TMD), and a cytoplasmic tail (CT). The SS is a ≃25–30 aa peptide, almost always present at the N-terminus of the polypeptide, that directs proteins to the secretory pathway where the SS is removed. The CFCS is a short aa sequence, frequently R-X-X-R or R-X-R/K-R, that is recognized and cleaved by a member of the furin-like family of pro-protein convertases. The TMD is ≃20 aa in length, consists primarily of hydrophobic aa, and represents a stable structure, either an α- or β-helix when inserted in membrane.

The precursor polypeptide of mZP1 is ≃69 kD MW and has a ZPD that is preceded by a trefoil (P) domain (a 45 aa sequence characterized by six Cys residues that form three intramolecular disulfides linked 1,5, 2,4, and 3,6) and a single extra ZP-N sub-domain (Fig. A.2.2). The precursor polypeptide of mZP2 is ≃79 kD MW and has a ZPD that is preceded by three extra copies of the ZP-N sub-domain (Fig. A.2.2). The precursor polypeptide of mZP3 is ≃47 kD MW, the smallest of the three mouse ZP proteins, and consists primarily of a single ZPD (Fig. A.2.2).

mZP1 polypeptide aa sequence and domain organization:

The polypeptide of ZP1 consists of 623 aa residues and has an SS (aa 1–20; highlighted), a trefoil domain (aa 225–266; italicized) with six Cys residues (aa 228, 237, 247, 252, 253, 262; capitalized and underlined), a ZPD (aa 271–540; highlighted)

with 10 Cys residues (aa 272, 306, 325, 368, 449, 470, 522, 527, 535, 539; capitalized and underlined), followed by a CFCS (aa 545–548, RRRR; highlighted and underlined) and a TMD (aa 591–611; highlighted). Cys residues 272, 306, 325, and 368 are in the ZP-N sub-domain and Cys residues 449, 470, 522, 527, 535, and 539 are in the ZP-C sub-domain of the ZPD.

```
1   mawgcfvvll llaaaplrlg qrlhlepgfe ysydcgvrgm qllvfprpnq tvqfkvldef
61  gnrfevnncs icyhwvtsea qehtvfsady kgchvlekdg rfhlrvfiqa vlpngrvdia
121 qdvtlicpkp dhtvtpdpyl appttpepft phafalhpip dhtlagsght glttlypeqs
181 fihptpapps lgpgpagstv phsqwgtlep welteldsvg thlpqerCqv asghipCmvn
241 gssketCqqa gCCydstkee pCyygntvtl qCfksgyftl vmsqetalth gvlldnvhla
301 yapngCpptq ktsafvvfhv pltlCgtaiq vvgeqliyen qlvsdidvqk gpqgsitrds
361 afrlhvrCif nasdflpiqa sifspqppap vtqsgplrle lriatdktfs syyqgsdypl
421 vrllrepvyv evrllqrtdp slvlvlhqCw atpttspfeq pqwpilsdgC pfkgdnyrtq
481 vvaadrealp fwshyqrfti ttfmlldsss qnalrgqvyf fCsasaChpl gsdtCsttCd
541 sgiarrrrss ghhnitlral divsspgavg fedaakleps gssrnsssrm llllaitla
601 laagifvgli wawaqklweg iry
```

mZP2 polypeptide aa sequence and domain organization:

The polypeptide of ZP2 consists of 713 aa residues and has an SS (aa 1–34; highlighted), a ZPD (aa 364–628; highlighted) with 10 Cys residues (aa 272, 306, 325, 368, 449, 470, 522, 527, 535, 539; capitalized and underlined), followed by a CFCS (aa 632–635, RSKR; highlighted and underlined) and a TMD (aa 684–703; highlighted). Cys residues 365, 396, 417, and 458 are in the ZP-N sub-domain and Cys residues 538, 569, 608, 613, 623, and 627 are in the ZP-C sub-domain of the ZPD.

```
1   marwqrkasv sspcgrsiyr flsllftlvt svnsvslpqs enpafpgtli cdkdevrief
61  ssrfdmekwn psvvdtlgse ilnctyaldl erfvlkfpye tctikvvggy qvnirvgdtt
121 tdvrykddmy hffcpaiqae theiseivvc rrdlisfsfp qlfsrladen qnvsemgwiv
181 kignqtrahi lplkdaivqg fnllidsqkv tlhvpanatg ivhyvqessy lytvqlellf
241 sttgqkivfs shaicapdls vacnathmtl tipefpgkle svdfgqwsip edqwhangid
301 keatnglrln frksllktkp sekcpfyqfy lsslkltfyf qgnmlstvid pechcespvs
361 idelCaqdgf mdfevyshqt kpalnldtll vgnssCqpif kvqsvglarf hiplngCgtr
421 qkfegdkviy eneihalwen ppsnivfrns efrmtvrCyy irdsmllnah vkghpspeaf
481 vkpgplvlvl qtypdqsyqr pyrkdeyplv rylrqpiyme vkvlsrndpn iklvlddCwa
541 tssedpasap qwqivmdgCe yeldnyrttf hpagssaahs ghyqrfdvkt fafvseargl
601 ssliyfhCsa liCnqvslds plCsvtCpas lrskreanke dtmtvslpgp illlsdvsss
661 kgvdpsssei tkdiiakdia sktlgavaal vgsavilgfi cylykkrtir fnh
```

mZP3 polypeptide aa sequence and domain organization:

The polypeptide of ZP3 consists of 424 aa residues and has an SS (aa 1–22; highlighted), a ZPD (aa 45–302; highlighted) with eight Cys residues (aa 46, 78, 98, 139, 216, 240, 283, 301; capitalized and underlined), followed by a CFCS (aa 350–353, RNRR; highlighted and underlined), and a TMD (aa 387–409; highlighted). Cys residues 46, 78, 98, and 139 are in the ZP-N sub-domain and Cys residues 216, 240, 283, and 301 are in the ZP-C sub-domain of the ZPD.

```
1    massyflflc lllcggpelc nsqtlwllpg gtptpvgsss pvkveCleae lvvtvsrdlf
61   gtgklvqpgd ltlgsegCqp rvsvdtdvvr fnaqlheCss rvqmtkdalv ystfllhdpr
121  pvsglsilrt nrvevpieCr ypraqnvssh piqptwvpfr atvsseekla fslrlmeenw
181  nteksaptfh lgevahlqae vqtgshlplq lfvdhCvatp splpdpnssp yhfivdfhgC
241  lvdglsesfs afqvprprpe tlqftvdvfh fanssrntly itChlkvapa nqipdklnka
301  Csfnktsqsw lpvegdadic dccshgncsn ssssqfqihg prqwsklvsr nrrhvtdead
361  vtvgpliflg kandqtvegw tasaqtsval glglatvafl tlaaivlavt rkchsssylv
421  slpq
```

Comparison of the aa sequences of the ZPDs of mZP1–3 (mZP1-aa 271–540; mZP2- aa 364–628; mZP3-aa 45–302) reveals that the only invariant residues are the Cys residues and 16 other aa residues; P, F, L between the second and third Cys residues; Y, R between the third and fourth Cys residues; four L residues between the fourth and fifth Cys residues; A, T, P, G between the fifth and sixth Cys residues; and two F and a Y residue between the sixth and seventh Cys residues. However, many aa positions in the sequences of other mammalian ZP proteins have conserved physiochemical character, such as always polar (≃15%), always small (≃15%), and always hydrophobic (≃7.5%).

FURTHER READING

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* **76**, 185–202 (1980).

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

Bork P. A trefoil domain in the major rabbit zona pellucida protein. *Protein Sci* **2**, 669–670 (1993).

Callebaut I, Mornon JP, Monget P. Isolated ZP-N domains constitute the N-terminal extensions of zona pellucida proteins. *Bioinformatics* **23**, 1871–1874 (2007).

Dietl J (ed.). *The Mammalian Egg Coat: Structure and Function*. 156 pp. Springer-Verlag, Berlin (1989).

Epifano O, Liang L, Dean J. Mouse *Zp1* encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish. *J Biol Chem* **270**, 27254–27258 (1995).

Kinloch RA, Roller RJ, Fimiani CM, Wassarman DA, Wassarman PM. Primary structure of the mouse sperm receptor's polypeptide chain determined by genomic cloning. *Proc Natl Acad Sci U S A* **85**, 6409–6413 (1988).

Liang L, Chamow SM, Dean J. Oocyte-specific expression of mouse *Zp-2*: developmental regulation of the zona pellucida genes. *Mol Cell Biol* **10**, 1507–1515 (1990).

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

Ringuette MJ, Chamberlin ME, Baur AW, Sobieski DA, Dean J. Molecular analysis of cDNA coding for ZP3, a sperm binding protein of the mouse zona pellucida. *Dev Biol* **127**, 287–295 (1988).

Wassarman PM. Zona pellucida glycoproteins. *Annu Rev Biochem* **57**, 415–442 (1988).

Wassarman PM. Mammalian fertilization: molecular aspects of gamete adhesion, exocytosis, and fusion. *Cell* **96**, 175–183 (1999).

Wassarman PM. Zona pellucida glycoproteins. *J Biol Chem* **283**, 24285–24289 (2008).

Wassarman PM, Litscher ES. Mammalian fertilization: the egg's multifunctional zona pellucida. *Int J Dev Biol* **52**, 665–676 (2008).

A.3 SYNTHESIS, SECRETION, AND ASSEMBLY OF ZP PROTEINS

ZP genes exhibit conserved organization with distinct domains defined by exon/intron boundaries. Mouse ZP genes share TATAA boxes ≃30 bp upstream of the transcription start sites and E-box sequences (CANNTG) that are involved in oocyte-specific expression of ZP genes. At least two ovary-specific DNA-binding proteins, ZAP-1 and OSP-1, bind to promoters of *mZP2* and *3*. As little as 153 nucleotides of the *mZP3* 5′-flanking sequence are sufficient to target the expression of a foreign protein (e.g., firefly luciferase) to growing oocytes. Messenger-RNA encoding mouse ZP proteins is undetectable in non-growing oocytes but appears in small oocytes that have entered the growth phase. For example, messenger-RNA encoding mZP3 is undetectable in non-growing oocytes but increases to ≃300,000 copies/oocyte in mid-stage growing oocytes, falls to ≃240,000 copies/oocyte in fully-grown oocytes, and decreases to undetectable levels in fertilized eggs (<1000 copies/egg). During oocyte growth, a period of ≃2–3 weeks in mice, ZP protein synthesis represents ≃5% of total protein synthesis by the oocyte.

As discussed previously, ZP proteins are synthesized as precursor polypeptides that have an SS and a CTP. The SS is removed during transit of the nascent proteins from the endoplasmic reticulum (ER) to the Golgi and the CTP is removed during secretion of the proteins into the extracellular space. In mice, all three ZP proteins are synthesized exclusively and coordinately by growing oocytes and are secreted independently. However, mZP2 and 3 are dependent on each other for incorporation into the ZP matrix. For example, female mice that are homozygous nulls for either *mZP2* or *3* fail to produce a ZP around growing oocytes and are infertile. Mice that are homozygous nulls for *mZP1* produce an abnormal ZP around growing oocytes but are fertile.

Several elements of nascent ZP proteins affect their secretion by oocytes and assembly into a ZP. For example, N-linked oligosaccharides are not required for secretion and assembly of mZP3 but are required for secretion and assembly of mZP2. mZP2 and 3 colocalize in unusually large secretory vesicles derived from Golgi but do not interact with each other inside the vesicles. Interaction between ZP

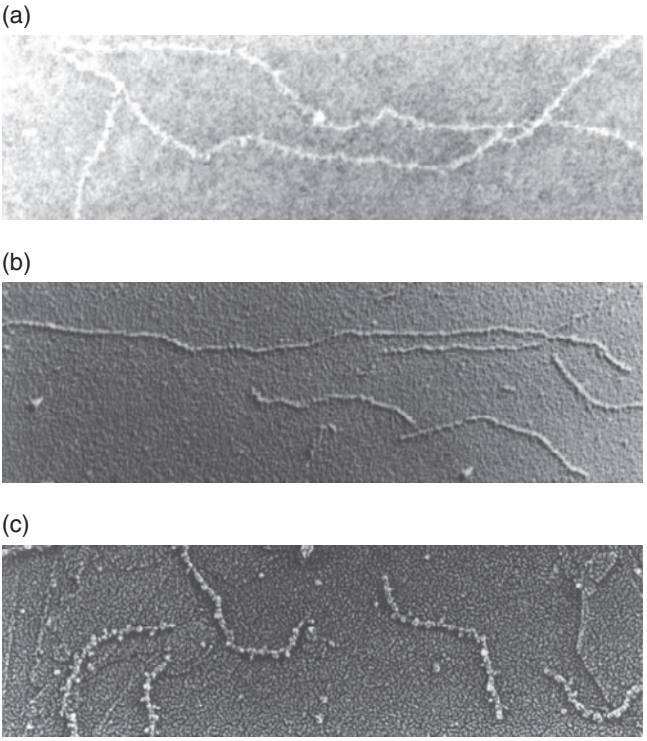

FIGURE A.3.1 Transmission electron micrographs of mouse ZP fibrils. Shown are (a) adsorbed, negatively stained, (b) sprayed, unidirectionally-shadowed, and (c) freeze-dried, unidirectionally-shadowed enzyme-solubilized preparations of ZP fibrils. Reproduced with permission from Wassarman (1991). © Springer.

proteins only occurs shortly before or just after the release of nascent mZP1–3 into the extracellular space. Nascent mZP2 and 3 are associated only with the innermost region of the thickening ZP, proximal to the oocyte's plasma membrane.

Once outside the oocyte, ZP proteins polymerize into long, interconnected fibrils (Fig. A.3.1). An mZP2–mZP3 dimer is located every $\simeq$14 nm or so along the fibrils, imposing a structural periodicity that can be visualized in electron micrographs of solubilized mouse ZP. The fibrils in turn are crosslinked by mZP1, the least abundant of the ZP proteins, to create a 3-dimensional matrix. The propensity of purified mZP1, 2, and 3 to form higher order oligomers *in vitro* and the failure to construct a ZP around mouse oocytes that do not synthesize either mZP2 or 3 is consistent with this model for ZP fibrils.

Prior to incorporation into the oocyte's ZP, nascent mZP2 and 3 polypeptides are processed at their CFCS, located close to the C-terminus of the polypeptides, by a member of the furin family of serine proteases. All three nascent ZP proteins have a TMD downstream of the CFCS. The presence of a TMD is not required for

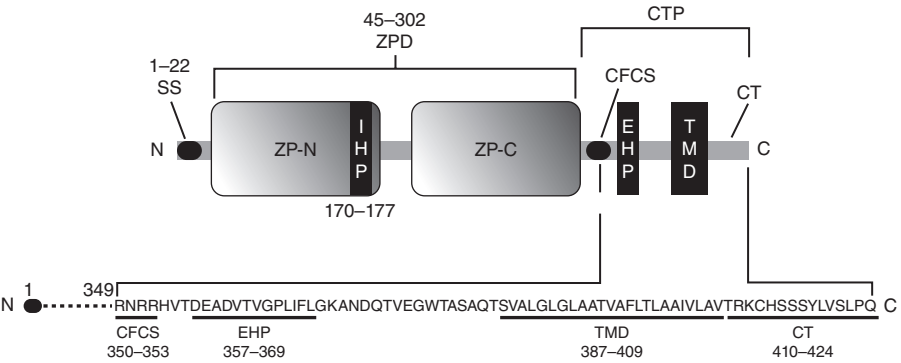

FIGURE A.3.2 Schematic representation of various features of mZP3. Shown are the positions of the SS (aa 1–22), ZPD (aa 45–302), IHP (aa 170–177), CFCS (aa 350–353), EHP (aa 357–369), TMD (aa 387–409), and CT (aa 410–424). The aa sequence of the CTP of mZP3, from aa 350–424, is shown together with the positions of the CFCS, EHP, TMD, and CT. Note that the IHP is located in the ZP-N sub-domain of the ZPD and that the EHP is located between the CFCS and TMD.

secretion but is required for cleavage at the CFCS and ensures proper localization and/or topological orientation of nascent ZP proteins so that assembly can take place. The carboxy-terminal propeptide (CTP) of nascent mZP1–3 possesses a short, hydrophilic cytoplasmic tail (CT) downstream of the TMD. Like the TMD, the CT is not required for secretion but is required for the incorporation of nascent ZP proteins into the thickening ZP.

Elements required for the secretion of ZP proteins are located in the CTP between the CFCS and TMD and include a hydrophobic peptide, referred to as an external hydrophobic patch (EHP) (Fig. A.3.2). ZP proteins must contain either an EHP or TMD to be secreted, and both the EHP and TMD must be present for incorporation of nascent ZP proteins into the thickening ZP. Another hydrophobic peptide, referred to as an internal hydrophobic patch (IHP), is present within the ZP-N sub-domain of the ZPD and also is required for the incorporation of ZP proteins into the ZP [Note: The IHP is not always present within the ZP-N sub-domain of the ZPD; for example, in chicken ZP3, the IHP is present within the ZP-C sub-domain (Part A.4). Accurate assignment of the position of an IHP depends upon the availability of a high-resolution, 3-dimensional structure.]. Secretion of nascent ZP proteins by oocytes is inhibited when either the EHP or IHP is mutated in mZP2 or 3 polypeptides truncated before the TMD (Fig. A.3.3). On the other hand, mutation of either the EHP or IHP in the presence of a TMD does not effect the secretion of mZP2 or 3 but prevents their assembly into fibrils (Fig. A.3.3).

The IHP and EHP are essential for the assembly of nascent ZP proteins into the ZP and possibly for the assembly of many other ZPD proteins, including all vertebrate egg envelopes. In this context, sequence alignments of ZP3 orthologues from

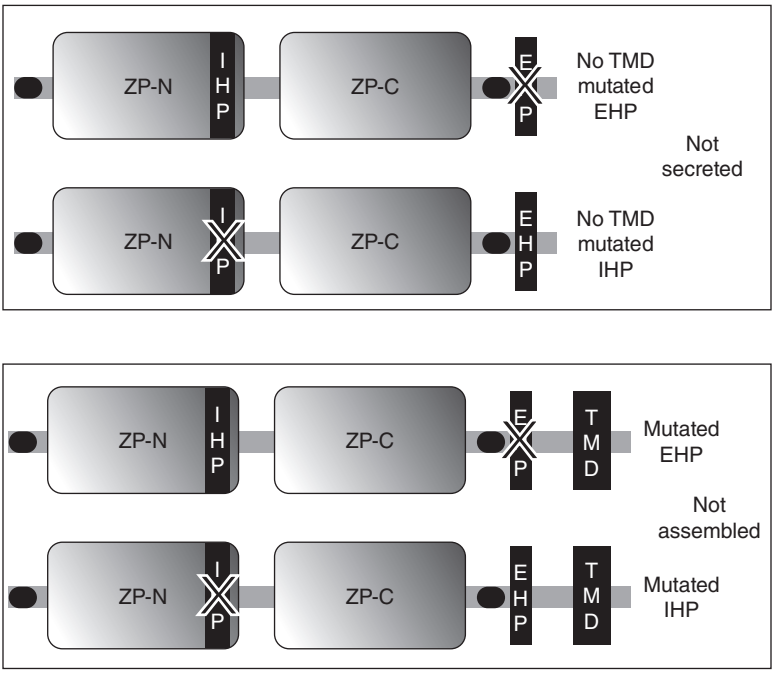

FIGURE A.3.3 Schematic representation of the effect of mutation of the EHP or IHP of mouse ZP proteins with and without a TMD. Top panel. In the absence of a TMD, mutation of either the EHP or IHP results in failure to secrete nascent ZP proteins. Bottom panel. In the presence of a TMD, mutation of either the EHP or IHP has no effect on the secretion of nascent ZP proteins but results in failure to assemble the proteins into a matrix.

TABLE A.3.1 Conservation of the IHP and EHP of ZP3

Organisms	IHP[a]	EHP[a]
Human	T **F S L R L M** E	E **E A D V T V G P L I F** L
Mouse	A **F S L R L M** E	D **E A D V T V G P L I F** L
Chicken	V **F S L R L M** S	V A **A D V** V I **G P** V L L S
Frog	A **F S L R L M** T	E H S L A **T** I **G P** I L V V
Trout	Y **F S** M **R L M** T	W E G **D V** Q L **G P** I F I S

[a]Shown are aa sequences for the IHP and EHP of ZP3 orthologues from mammals (human and mouse), birds (chicken), amphibians (frog), and fish (trout). Identical aa for different species are indicated in bold and nonidentical, hydrophobic aa are underlined.

human, mouse, chicken, frog, and fish egg envelopes reveal well-conserved IHPs and EHPs (Table A.3.1).

The presence of both an EHP and IHP within ZP protein precursors is thought to prevent premature assembly of nascent ZP proteins within oocytes. In fact, the high-resolution structure of full-length ZP3 (Part A.4) reveals that the EHP is

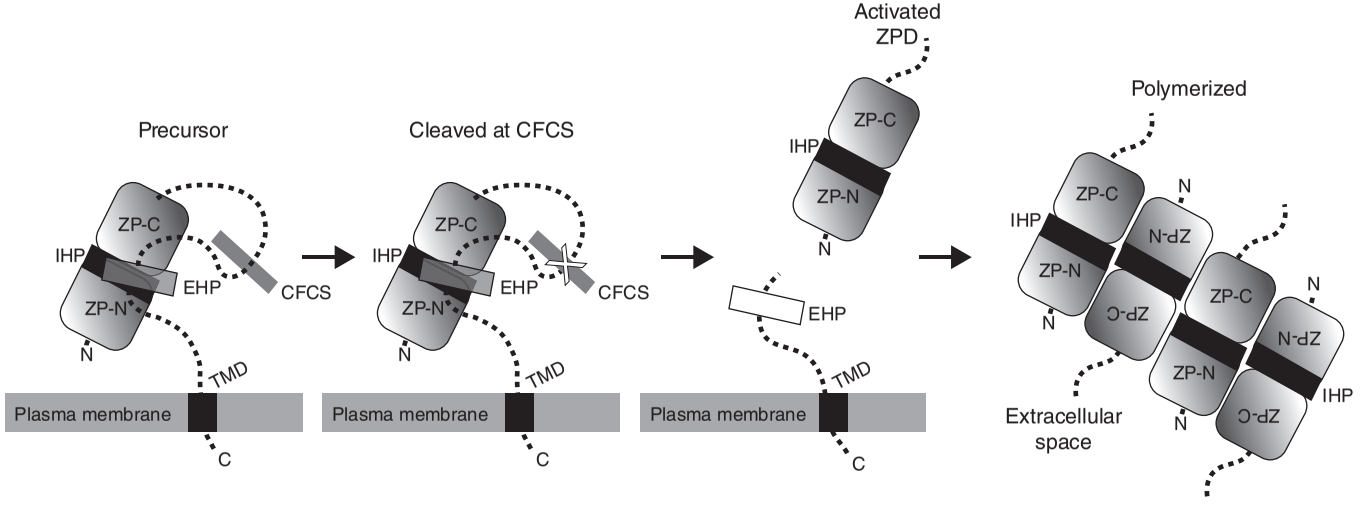

FIGURE A.3.4 Schematic representation of a general mechanism for assembly of nascent ZP proteins. In all ZPD precursor proteins (precursor) the ZPD is followed by a CTP that contains a basic cleavage site, such as a CFCS, an EHP, and, in most cases, a TMD or GPI-anchor site. Precursors do not polymerize within the cell either as a result of direct interaction between the EHP and IHP or because they adopt a conformation dependent on the presence of both hydrophobic patches. C-terminal processing at the CFCS by a pro-protein convertase (cleaved at CFCS) leads to dissociation of mature proteins from the EHP and activation of the ZPD (activated ZPD) for assembly (polymerized) into fibrils and matrices.

found very close to the IHP. When the CTP of ZP protein precursors is removed by proteolytic cleavage at the CFCS, the EHP no longer interacts with the IHP and ZP assembly ensues. This mechanism may apply to all ZPD proteins since it relies on sequence elements, the EHP and IHP, and certain events, like coupling between proteolytic processing and assembly, which are conserved in all ZPD proteins. The IHP and EHP apparently function as "control switches" in ZPD protein assembly by preventing premature assembly of nascent ZP proteins within oocytes or other cell types. ZP protein constructs that lack a TMD are neither secreted nor incorporated into the ZP if either the EHP or IHP is missing. However, if a TMD is present, such constructs are secreted but not incorporated into the ZP. It is likely that the EHP transiently masks the IHP that is required for interactions between secreted nascent ZP proteins. Cleavage at the CFCS releases the EHP and activates ZP proteins for assembly into fibrils and matrices in the extracellular space (Fig. A.3.4).

FURTHER READING

Bleil JD, Wassarman PM. 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 (1980).

Epifano O, Liang L, Familiari M, Moos MC, Dean J. Coordinate expression of the three zona pellucida genes during mouse oogenesis. *Development* **121**, 1947–1956 (1995).

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

Greve JM, Salzmann GS, Roller RJ, Wassarman PM. Biosynthesis of the major zona pellucida glycoprotein secreted by oocytes during mammalian oogenesis. *Cell* **31**, 749–759 (1982).

Hoodbhoy T, Aviles M, Baibakov B, Epifano O *et al*. ZP2 and ZP3 traffic independently within oocytes prior to assembly into the extracellular zona pellucida. *Mol Cell Biol* **26**, 7991–7998 (2006).

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

Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM. The ZP domain is a conserved module for protein polymerization. *Nat Cell Biol* **4**, 457–461 (2002).

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

Lira SA, Kinloch RA, Mortillo S, Wassarman PM. An upstream region of the mouse ZP3 gene directs expression of firefly luciferase specifically to growing oocytes in transgenic mice. *Proc Natl Acad Sci U S A* **87**, 7215–7219 (1990).

Lira SA, Schickler M, Wassarman PM. Cis-acting DNA elements involved in oocyte-specific expression of mouse sperm receptor gene mZP3 are located close to the gene's transcription start-site. *Mol Reprod Dev* **36**, 494–499 (1993).

Litscher ES, Qi H, Wassarman PM. Mouse zona pellucida glycoproteins mZP2 and mZP3 undergo carboxy-terminal proteolytic processing in growing oocytes. *Biochemistry* **38**, 12280–12287 (1999).

Litscher ES, Janssen WG, Darie CC, Wassarman PM. Purified mouse egg zona pellucida glycoproteins polymerize into homomeric fibrils under non-denaturing condition. *J Cell Physiol* **214**, 153–157 (2008).

Louros NN, Iconomidou VA, Giannelou P, Hamodrakas SJ. Structural analysis of peptide-analogues of human zona pellucida ZP1 protein with amyloidogenic properties: insights into mammalian zona pellucida formation. *PLoS One* **8**, 9 (2013).

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

Roller RJ, Wassarman PM. Role of asparagine-linked oligosaccharides in secretion of glycoproteins of the mouse egg's extracellular coat. *J Biol Chem* **258**, 13243–13249 (1983).

Salzmann GS, Greve JM, Roller RJ, Wassarman PM. Biosynthesis of the sperm receptor during oogenesis in the mouse. *EMBO J* **2**, 1451–1456 (1983).

Shimizu S, Tsuji M, Dean J. *In vitro* biosynthesis of three sulfated glycoproteins of murine zonae pellucidae by oocytes grown in follicle culture. *J Biol Chem* **258**, 5858–5863 (1983).

Tong ZB, Nelson LM, Dean J. Inhibition of zona pellucida gene expression by antisense oligonucleotides injected into mouse oocytes. *J Biol Chem* **270**, 849–853 (1995).

Wassarman PM. Fertilization in the mouse. I. The egg. In: *A Comparative Overview of Mammalian Fertilization* (Dunbar BS, O'Rand MG eds.), Plenum Press, New York, pp. 151–165 (1991).

Wassarman PM, Litscher ES. Influence of the zona pellucida of the mouse egg on folliculogenesis and fertility. *Int J Dev Biol* **56**, 833–839 (2012).

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

Wassarman PM, Mortillo S. Structure of the mouse egg extracellular coat, the zona pellucida. *Int Rev Cytol* **130**, 85–110 (1991).

Williams Z, Wassarman PM. Secretion of mouse ZP3, the sperm receptor, requires cleavage of its polypeptide at a consensus furin cleavage-site. *Biochemistry* **40**, 929–937 (2001).

Zhao M, Gold L, Dorward H, Liang L *et al*. Mutation of a conserved hydrophobic patch prevents incorporation of ZP3 into the zona pellucida surrounding mouse oocytes. *Mol Cell Biol* **23**, 8982–8991 (2003).

A.4 STRUCTURE OF THE ZPD

As mentioned previously, the ZPD is a bipartite structure consisting of two sub-domains, ZP-N and ZP-C, linked by a protease-sensitive region. However, ZP-N can serve as an independent structural domain since it is present in ZPD proteins Plac1, Oosp1, and papillote in the absence of the ZP-C sub-domain. In addition, divergent copies of ZP-N are found in single or multiple copies in the N-terminal extensions of ZP1, 2, 4, and ax (Fig. A.4.1). It has been shown that the ZP-N

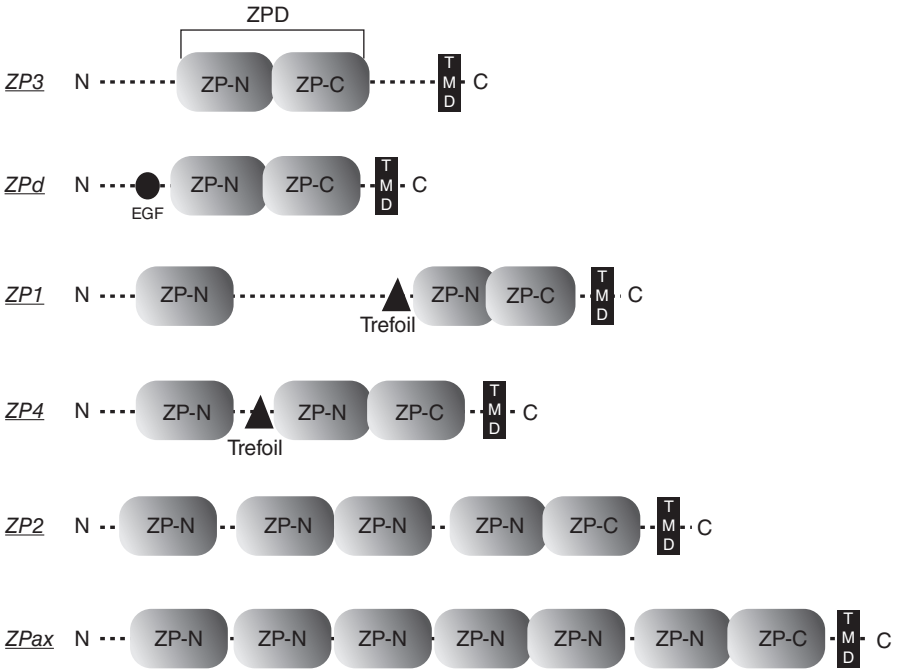

FIGURE A.4.1 In each representation, the positions of the ZPD and TMD are indicated. For ZP1 and 4, the postion of the trefoil domain (P) is indicated and for ZPd the position of the EGF domain is indicated. Note the extra copy or copies of the ZP-N sub-domain in ZP1, 2, 4 and ax.

sub-domain alone is a biologically active folding unit that can assemble into fibrils. This suggests that the ZP-C sub-domain may play a regulatory role in the assembly of ZPD protein complexes.

ZP-N and ZP-C sub-domains share a common basic topology with β-sheet arrangements symmetrical to each other, although the sub-domains have significantly different primary structures and intramolecular disulfide bonds. There are two types of ZPDs. Type-I (ZP3-like) with eight Cys residues and type-II (ZP1/2-like) with 10 Cys residues. The type-I ZPD has a ZP-N sub-domain with four Cys residues, linked 1,4 and 2,3, and a ZP-C sub-domain with four Cys residues, linked 5,7 and 6,8. The type-II ZPD has a ZP-N sub-domain with four Cys residues, linked 1,4 and 2,3, and a ZP-C sub-domain with six Cys residues, linked 5,6, 7,a, and b,8 (Fig. A.4.2).

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 ZP1-like proteins (α and β) and are located between the ZP-N and ZP-C sub-domains (Table A.4.1).

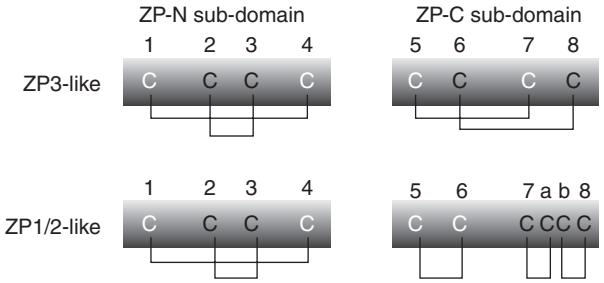

FIGURE A.4.2 Schematic representation of intramolecular disulfides in ZP3-like (type-I) and ZP1/2-like (type-II) ZPD proteins. Top: ZP3-like ZP-N sub-domain with four Cys residues linked 1,4 and 2,3 and ZP-C sub-domain with four Cys residues linked 5,7 and 6,8. Bottom: ZP1/2-like ZP-N sub-domain with four Cys residues linked 1,4 and 2,3 and ZP-C sub-domain with six Cys residues linked 5,6, 7,a, and b,8.

TABLE A.4.1 Alignment of Additional Conserved Cys Residues in Trout ZP1[a]

		C_x	C_y	
Fish ZP1-α	· AVLHVELRLA	NGRCLSKGCD	EMQEAYTSYY	
ZP1-β	· PLI.VELRLG	SGGCLTKGCN	EEEVAYTSYY	
Mouse ZP1	· PLR.LELRIA	T.....	..DKTFSSYY	
ZP2	· PLV.LVLQTY	P.....	..DQSYQRPY	

[a]Shown are partial sequences from trout (*Oncorhynchus mykiss*) ZP1α/ZP1β with Cx and Cy and partial sequences of mZP1/mZP2 for comparison.

There are a few other ZPD proteins with 11, 12, or more Cys residues within their ZPD, for example, piopio (*Drosophila melanogaster*) (Part D.5.d), RAM-5 (*Caenorhabditis elegans*) (Part D.3.c), and tectorin-α (mammals) (Part C.9). However, no disulfide assignments have been made for the extra Cys residues of these ZP1/2-like proteins. Cys residue clustering in ZP-C sub-domains is variable in chicken and pig ZP3; there is a disulfide linkage to a Cys residue C-terminal to the ZPD, 6,11 and 8,9, whereas in mZP3 it is 6,8 and 9,11. It is likely that these two different disulfide bonding patterns cause the polypeptides to adopt different conformations that may determine the specificity of egg coat assembly in different species.

The 3-dimensional structure of a ZPD has been determined by X-ray crystallographic analysis at 2.0 Å resolution using crystals of chicken ZP3. Chicken ZP3 was engineered as a shortened polypeptide (358 aa residues) with a deleted TMD (V403–I425), a mutated N-glycosylation site (N159VS → Q159VS), and no CFCS (R359FRR → A359FAA) [Note: There are two chicken ZP3 sequences; one is nine aa longer than the other at the N-terminus (446 vs. 437 aa). The numbering shown here is for the 437 aa ZP3

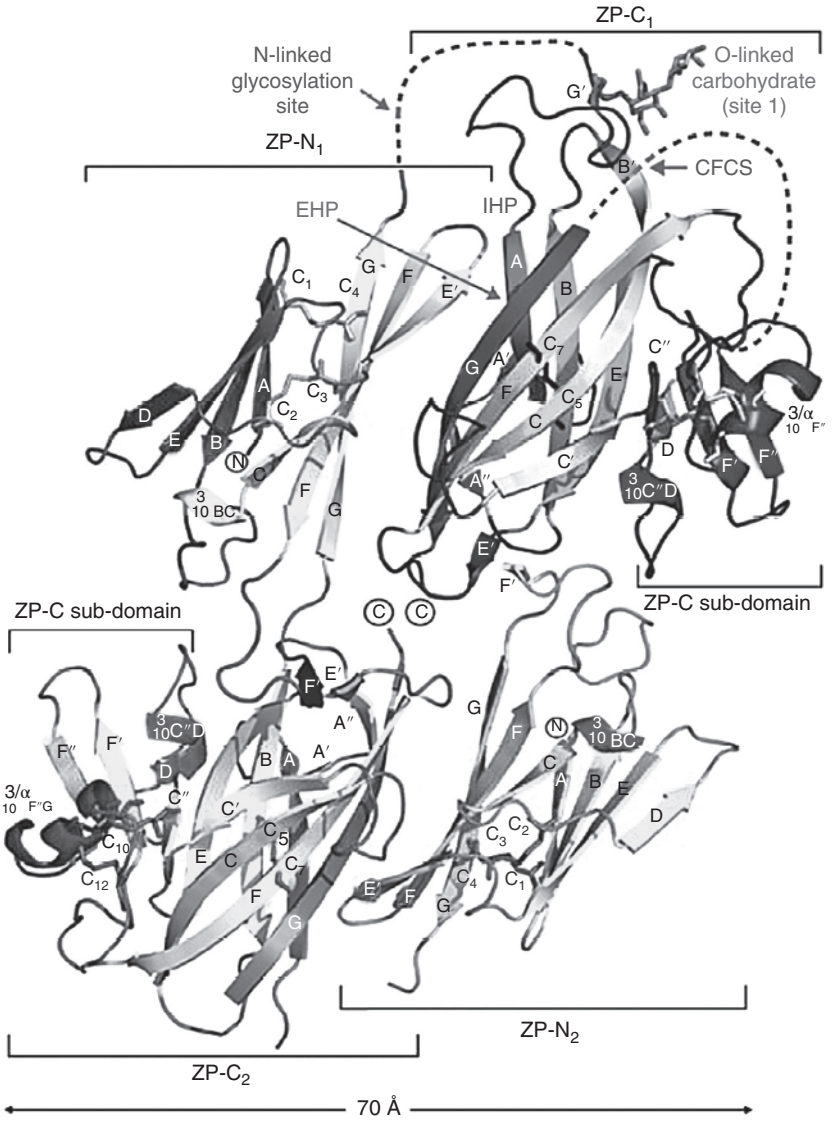

FIGURE A.4.3 Chicken ZP3 homodimer structure formed by two ZP modules each consisting of a ZP-N and ZP-C sub-domain. Dashed lines represent disordered loops. Reproduced with permission from Han *et al*. (2010). © Elsevier. For color detail, please see color plate section.

sequence.] The protein was expressed in Chinese hamster ovary cells and secreted as an ≃80 kD MW soluble homodimer.

The crystallographic analysis revealed that two chicken ZP3 molecules are arranged in antiparallel orientation in the homodimer to form a flat,

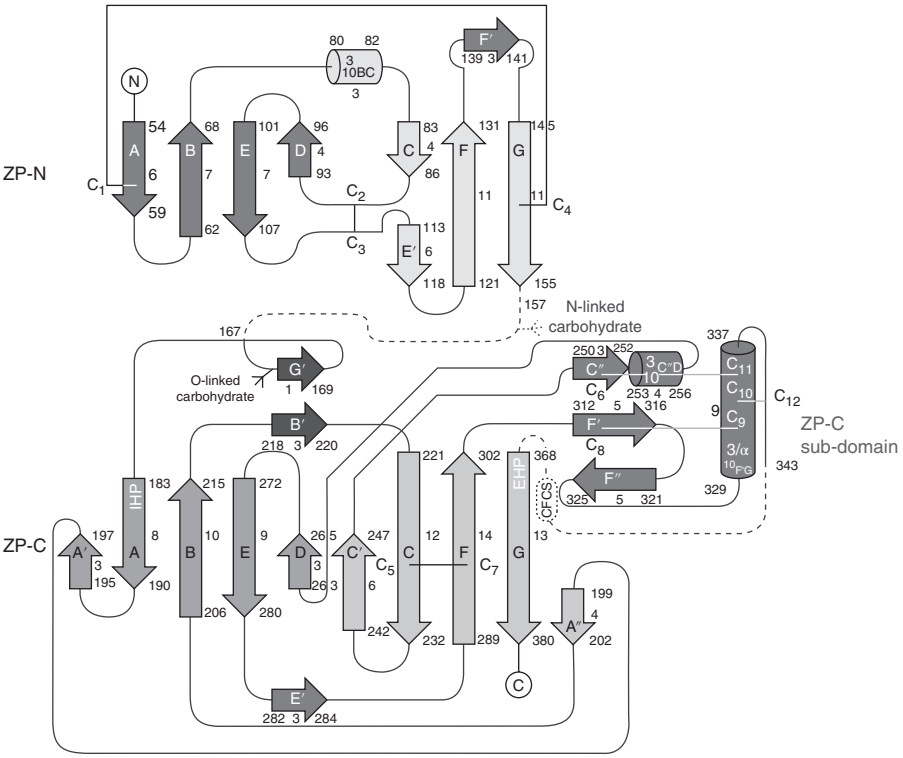

FIGURE A.4.4 Topology scheme of chicken ZP3 with secondary structure and disulfide connectivity. Reproduced with permission from Han *et al*. (2010). © Elsevier. [Note: In this figure, V54 corresponds to V63 and N316 corresponds to N325 in the chicken ZP3 sequence.] For color detail, please see color plate section.

Yin-Yang-shaped, asymmetric structure. The two ZPDs are held together by electrostatic interactions between ZP-N and ZP-C sub-domains of opposing molecules. The β-strands of the ZP-N and ZP-C sub-domains share a common immunoglobulin (Ig)-like topology that give the ZPD an internal symmetry (Fig. A.4.3).

The ZPD of chicken ZP3 consists of 258 aa residues, from Q66 (ZP-N) to S323 (ZP-C) and does not include β-strands F″ and G and helix F″G (Fig. A.4.4). In this overall topography, the C-terminal strand, including the EHP and mutated CFCS, is an integral part of the chicken ZP3-fold (Fig. A.4.3). The EHP is part of β-strand G and faces directly the IHP on β-strand A of the ZP-C sub-domain. Apparently this arrangement is stable in the uncleaved protein precursor but dissociates slowly after cleavage at the CFCS. The structure is consistent with the idea that the EHP blocks premature protein assembly by acting as "molecular glue" that keeps the ZPD module in a conformation that is essential for secretion but is incompatible with assembly.

FURTHER READING

Callebaut IU, Mornon AP, Monget P. Isolated ZP-N domains constitute the N-terminal extensions of zona pellucida proteins. *Bioinformatics* **23**, 1871–1874 (2007).

Han L, Monné M, Okumura H, Schwend T *et al*. Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3. *Cell* **143**, 404–415, (2010).

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

Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona pellucida domain proteins. *Annu Rev Biochem* **74**, 83–114 (2005).

Jovine L, Janssen WG, Litscher ES, Wassarman PM. The PLAC-1 homology region of the ZP domain is sufficient for protein polymerization. *BMC Biochem* **7**, 11–19 (2006).

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

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

A.5 EVOLUTION OF ZPD PROTEINS

The ZPD arose more than ≃600 million years ago and is a component of proteins from a wide range of animal species, from jellyfish to humans (Fig. A.5.1; Parts B, C, and D). Protein domains like the ZPD are evolutionary units that can be duplicated and recombined. Pairs of domains are usually found in one sequential order (A → B or B → A) but almost never in both. So it is with the ZPD. ZP-N and ZP-C are always present in one order (ZP-N → ZP-C) and not the other (ZP-C → ZP-N), although the ZP-N sub-domain can be found by itself. Proteins with sequence identities of 40% or more usually have the same function (e.g., mammalian egg ZP proteins from platypus to human), those with identities of 25–40% have similar functions (e.g., fish and human ZP proteins), and those with identities below 25% have different functions (e.g., mammalian ZP proteins, tectorin-α, and uromodulin).

It is likely that ZP proteins are derived from a common ancestral gene. In this context, it has been proposed that a first duplication event in evolution gave rise to ZP3 and an ancestral ZP gene subsequently duplicated several times and evolved into all the other ZP genes (Fig. A.5.2). *ZPd* genes are found in amphibians (Part D.8) and birds (Part D.10) and *ZPax* genes are found in fish (Part D.7), amphibians, and birds. The absence of *ZPd* and *ZPax* in mammals suggests that these genes have been lost during evolution. Accordingly, mammals have three to four ZP genes, amphibians have five ZP genes, and birds have six ZP genes.

Since more and more DNA sequencing data has become available, a better understanding of the genetics of the ZP has emerged. For example, ZP homologs are classified as orthologues (e.g., mZP3 vs. human ZP3) or paralogues (e.g., mZP3 vs.

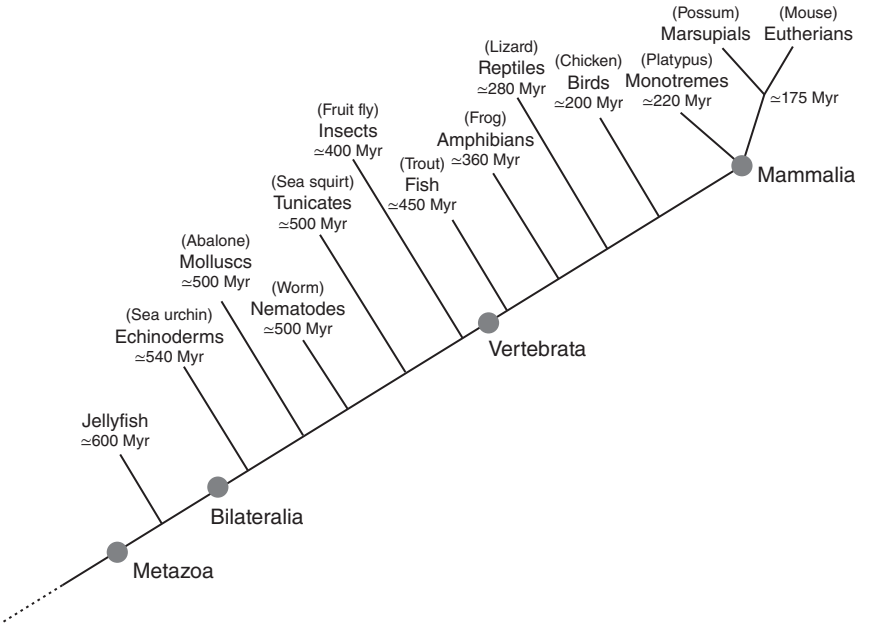

FIGURE A.5.1 Phylogenetic relationships of ZPD proteins as depicted in the "tree of life." A ZPD is present as early as ≃600 million years ago in jellyfish (Part D.1). ZPD proteins that have diverse functions are found in every major animal group and in a wide variety of tissues and organs.

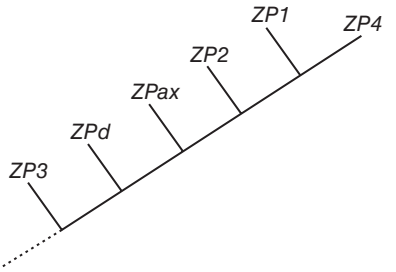

FIGURE A.5.2 Evolutionary scheme of the organization of *ZP* genes in mammals, fish, amphibians, reptiles, and birds. *ZP1–4* are found in mammals and other vertebrates, *ZPd* in amphibians and birds, and *ZPax* in fish, amphibians, and birds.

mZP2) and ZP locations are compared to portions of chromosomes in different species (e.g., comparable locations are referred to as synteny). In addition, several ZP pseudogenes have been identified in mammals, for example, a *ZP4* pseudogene in mice and *ZP1* pseudogene in dogs and cows.

Highly divergent sequences have been identified in ZP2 and 3. In mZP3, there are two clusters of sites—N-terminal (aa 25–50) and C-terminal (aa 331–373)—to the

ZPD that are under positive selection. mZP2 has several single positive selection sites. It is possible that the high divergence of these regions is the result of positive Darwinian selection and drives the evolution of the proteins. Similar analyses of abalone and avian egg coat proteins, Vitelline Envelope Receptor for Lysin (VERL) (Part D.4) and ZP3 (Part D.10), respectively, have led to the same conclusion.

FURTHER READING

Benton MJ, Ayala FJ. Dating the tree of life. *Science* **300**, 1698–1700 (2003).

Chothia C, Gough J, Vogel C, Teichmann SA. Evolution of the protein repertoire. *Science* **300**, 1701–1703 (2003).

Claw KG, Swanson WJ. Evolution of the egg: new findings and challenges. *Annu Rev Genomics Hum Genet* **13**, 109–125 (2012).

Findlay GD, Swanson WJ. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *Bioessays* **32**, 26–36 (2010).

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

Litscher E, Wassarman PM. Egg extracellular coat proteins: from fish to mammals. *Histol Histopathol* **22**, 337–347 (2007).

Pennisi E. Drafting a tree. *Science* **300**, 1694 (2003).

Spargo SC, Hope RM. Evolution and nomenclature of the zona pellucida gene family. *Biol Reprod* **68**, 358–362 (2003).

Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nat Rev Genet* **31**, 137–144 (2002).

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A* **98**, 2509–2514 (2001).

Swanson WJ, Aagaard JE, Vacquier VD, Monné M *et al*. The molecular basis of sex: linking yeast to human. *Mol Biol Evol* **28**, 1963–1966 (2011).

Turner LM, Hoekstra HE. Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol Biol Evol* **23**, 1656–1669 (2006).