Annu. Rev. Biochem. 2005. 74:83–114
doi: 10.1146/annurev.biochem.74.082803.133039

ZONA PELLUCIDA DOMAIN PROTEINS

Luca Jovine, Costel C. Darie, Eveline S. Litscher, and Paul M. Wassarman

Brookdale Department of Molecular, Cell, and Developmental Biology, Mount Sinai School of Medicine, New York, New York 10029-6574; email: luca.jovine@mssm.edu, costel.darie@mssm.edu, eveline.litscher@mssm.edu, paul.wassarman@mssm.edu

Key Words extracellular proteins, secretion, modular proteins, protein polymerization, fertilization, receptors, mechanical tension, mechanosensitivity

■ **Abstract** Many eukaryotic proteins share a sequence designated as the zona pellucida (ZP) domain. This structural element, present in extracellular proteins from a wide variety of organisms, from nematodes to mammals, consists of ∼260 amino acids with eight conserved cysteine (Cys) residues and is located close to the C terminus of the polypeptide. ZP domain proteins are often glycosylated, modular structures consisting of multiple types of domains. Predictions can be made about some of the structural features of the ZP domain and ZP domain proteins. The functions of ZP domain proteins vary tremendously, from serving as structural components of egg coats, appendicularian mucous houses, and nematode dauer larvae, to serving as mechanotransducers in flies and receptors in mammals and nonmammals. Generally, ZP domain proteins are present in filaments and/or matrices, which is consistent with the role of the domain in protein polymerization. A general mechanism for assembly of ZP domain proteins has been presented. It is likely that the ZP domain plays a common role despite its presence in proteins of widely diverse functions.

CONTENTS

INTRODUCTION

Since first recognized as a structural element by Bork & Sander in 1992 (1), the ZP domain has been found in hundreds of extracellular proteins of diverse functions from a wide variety of tissues and organisms (2). The latter include mammals, amphibia, birds, fish, flies, worms, molluscs, echinoderms, and tunicates (Table 1). In this context, it has been estimated that the *Drosophila melanogaster* and *Caenorhabditis elegans* genomes encode at least 18 and 40 ZP domain proteins, respectively (3, 4).

The designation ZP stands for zona pellucida, the thick extracellular coat that surrounds all mammalian eggs and preimplantation embryos. The mammalian egg coat is composed of only a few glycosylated proteins that all possess the archetypal ZP domain, which consists of ~260 amino acids (aa) and has 8 conserved cysteine (Cys) residues that participate in intramolecular disulfides (2, 5). It appears likely that the ZP domain is a bipartite structure divided by a protease-sensitive region. In most instances, the ZP domain is relatively close to the C terminus and is often only one of the recognizable motifs that make up the polypeptide (Table 1). For example, ZP domain proteins often contain other types of domains, such as proline-rich (P) or trefoil (6), epidermal growth factor (7), CUB or BMP (complement subcomponents Clr/s, Uegf protein, and bone morphogenetic protein) (8, 9), PAN (plasminogen N terminus) (10), SRCR (scavenger receptor cysteine rich) (11), von Willebrand factor (12), or other domains (13, 14). In addition, most ZP domain proteins are glycosylated and possess an N-terminal signal peptide and either a C-terminal putative transmembrane domain (TMD) or glycosyl phosphatidylinositol-(GPI-) anchor.

Here, we survey some of the extracellular proteins that possess a ZP domain, describe some of its structural features, and discuss a specific role ascribed to the ZP domain. Despite the disparate functions of ZP domain proteins, from sperm receptors and mechanotransducers to structural components, it is likely that the ZP domain plays an identical role in most, if not all, of these proteins. In this context,

TABLE 1 Representative ZP domain proteins

ZP domain proteins	Primary source	Other domains	References
Mammalian			
ZP1–3	Ovary	P/trefoil	5, 20, 23
Tectorins	Inner ear	G1, D	15, 25, 26, 33
Tamm-Horsfall protein	Kidney	EGF	17, 34–37
TGF-β type III receptor	Heart		41–43, 46–51
LZP	Liver	EGF	52
GP-2	Pancreas, kidney		53–55
Muclins	Intestine, pancreas	SRCR, CUB	63–65
Ebnerin	Tongue	SRCR, CUB	67
Vomeroglandin	Nose	SRCR, CUB	66
DMBT1	Brain, lung	SRCR, CUB	18, 68, 69
Hensin	Brain, epithelia	SRCR, CUB	62, 72
Itmap-1	Pancreas	CUB	73
UTCZP	Uterus	CUB	74
ERG-1	Uterus, oviduct	CUB	75
UO-44	Ovary	CUB	76, 77
PLAC1	Placenta		78–80
Oosp1	Ovary, spleen		81, 82
Nonmammalian			
H. rufescens			
VERL	Ovary		87, 89
D. melanogaster			
Dumpy/Piopio	Trachea, wings	EGF, DPY	4, 90–92
Miniature/Dusky	Cuticle, wings		93, 94
NompA	Dendritic cap	PAN	95, 96
C. elegans			
Cuticlin-1	Cuticle		97, 98
O. dioica			
Oikosins	Epithelium	CUB	99, 100

it should be noted that mutations in genes encoding ZP domain proteins can result in severe human pathologies, including nonsyndromic deafness (15), vascular (16) and renal (17) diseases, and cancer (18, 19).

MAMMALIAN ZP DOMAIN PROTEINS

ZP domain proteins are found in various organs from all mammals and play a wide variety of roles, from structural components, to receptors, to tumor suppressors. In each case, the ZP domain is close to the C terminus of a polypeptide, which usually contains other recognizable motifs.

ZP1–3

The ZP is a thick extracellular coat that surrounds all mammalian eggs and plays important roles during oogenesis, fertilization, and preimplantation development (2, 5, 20). All ZP proteins possess a ZP domain (Figure 1). For example, the mouse ZP is composed of three glycosylated proteins, called ZP1–3, each of which has a ZP domain relatively close to the C terminus of the polypeptide. In addition, each protein has an N-terminal signal peptide, a consensus furin cleavage site (CFCS), a C-terminal putative TMD, and a short cytoplasmic tail. ZP1 also has a P or trefoil domain (21) just upstream of its ZP domain. These proteins are concomitantly synthesized, secreted, and assembled into a ZP as mouse oocytes

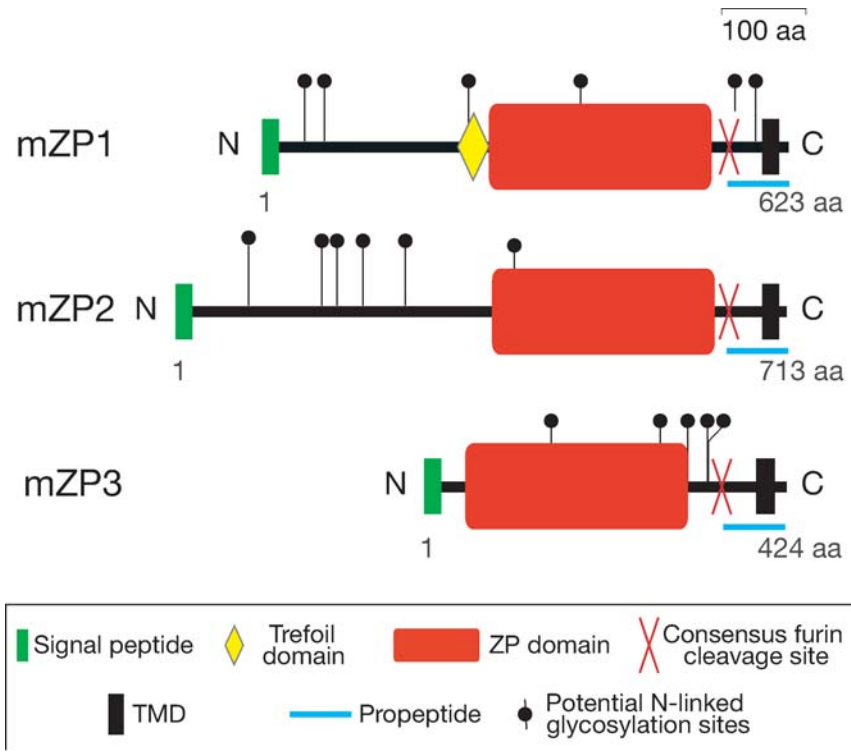


Figure 1 Schematic representation of the overall architecture of mouse ZP glycoproteins, ZP1, ZP2, and ZP3. The polypeptide of each ZP glycoprotein is drawn to scale, with the N and C termini indicated. Key features of the polypeptide, including the N-terminal signal peptide (*green*), P or trefoil domain (*yellow*), ZP domain (*red*), CFCS (X), TMD (*black*), and C-terminal propeptide region (*blue bar*) are depicted. Only putative N-linked glycosylation sites, conforming to the strict pattern Asn-X-Ser/Thr-X, where X can be any amino acid other than Pro, are shown.

grow. The secreted forms of ZP1, ZP2, and ZP3, lacking signal peptides and cleaved at the CFCS, have apparent M_rs of ~200 (dimer; ~59 kDa M_r polypeptides), ~120 (~68 kDa M_r polypeptide), and ~83 (~37 kDa M_r polypeptide) kDa, respectively. All three proteins possess N- (complex-type) and O-linked oligosaccharides, in addition to other posttranslational modifications (e.g., sulfation and sialylation). Mouse ZP proteins are encoded by single-copy genes (ZP1–12 exons, 623 aa, chromosome 19; ZP2–18 exons, 713 aa, chromosome 7; ZP3–8 exons, 424 aa, chromosome 5) (22).

ZP proteins have been characterized from a wide variety of mammalian eggs, including eggs from rodents, domesticated animals, marsupials, and primates (2, 23). Although the molecular weights differ for ZP glycoproteins from different species, due in large part to differential glycosylation and other modifications, it is apparent that all ZP consist of only a few glycoproteins whose polypeptides are related to those of mouse ZP1–3. The primary structures of ZP2- and ZP3-related ZP glycoproteins from different species are relatively well conserved (~65% to 98% identity), whereas ZP1-related glycoproteins are conserved to a lesser degree (~40% identity). It also is apparent that ZP1–3 have regions of polypeptide in common, suggesting that these regions may be derived from a common ancestral gene. It should be noted that there are additional forms of ZP glycoproteins (e.g., ZPB) (24), all containing a ZP domain and highly related to ZP1–3, that are present in egg coats from some mammals.

α- and β-Tectorin

The cochlear tectorial membrane is a specialized extracellular matrix (ECM) that contacts ciliated mechanosensory hair cells in the vertebrate inner ear and is important in the transduction of sound into neural potentials. The membrane is essentially a gel-like matrix positioned above hair cells such that stereocilia bundles of hair cells can bend against it in order to transmit sound (15, 25).

Three different collagens (types II, V, and IX) combine with three noncollagenous, glycosylated polypeptides, called α-tectorin, β-tectorin, and otogelin, to form the mammalian tectorial membrane (15, 25). Notably, no collagen components are detectable in the tectorial membrane of birds; however, it contains a set of glycoproteins homologous to mammals (26). The tectorins are associated with a striated matrix consisting of two types of fibrils that are connected to form flat sheets. The sheets are stacked on top of each other and wrap around bundles of collagen fibers (27, 28). Mice that are homozygous nulls for α-tectorin have a tectorial membrane that lacks a striated-sheet matrix and is detached from the Organ of Corti (28). Furthermore, the tectorial membrane of these mice lack β-tectorin, consistent with the idea that α- and β-tectorin interact with each other and polymerize into filaments that constitute the striated matrix. This could also explain the reported link between nonsyndromal hearing impairments and single-site mutations, deletions in human α-tectorin (15, 29–31), or deletion of the ZP domain of α-tectorin (32).

Mouse α- and β-tectorin are present as single-copy genes that are expressed by cells in and around mechanosensory epithelia. α-Tectorin and β-tectorin are synthesized as 239 (2150 aa) and 36 (320 aa) kDa M_r polypeptides, respectively (26, 33). Each contains potential N-linked glycosylation sites (α-tectorin, 33; β-tectorin, 4), an N-terminal signal peptide, and a hydrophobic C terminus characteristic of proteins that are membrane bound via a GPI-anchor. The latter are preceded a short distance upstream by a CFCS that is characteristic of endoproteinase cleavage sites that, when cleaved, result in the release of nascent tectorins into the extracellular compartment.

Mouse α-tectorin polypeptide (Figure 2) consists of three distinct modules. (*a*) An N-terminal region that is similar to a portion of the first globular domain (G1) of entactin/nidogen. (*b*) A long, central domain that is composed of three full repeats and two partial repeats homologous to the D domains of prepro-von Willebrand factor, zonadhesin, and intestinal mucin, muc2. (*c*) A ZP domain close to the C terminus that is similar to the ZP domains of THP and GP-2. Mouse β-tectorin polypeptide consists only of a ZP domain close to the C terminus.

Tamm-Horsfall Protein

Tamm-Horsfall protein (THP), also called uromodulin, is the most abundant protein in human urine and is excreted at the rate of ~50 mg/day (17, 34). THP has been studied extensively by electron microscopy (35–37). Single filaments of the protein originate from and merge into bundles at seemingly irregular intervals and generate a three-dimensional matrix with pores. Analyses of single fibrils reveal a so-called zig-zag course, and the structural features could be interpreted as two-dimensional projections of a helical superstructure (36, 108). Owing to their propensity to form a gel, THP filaments may ensure the water impermeability of the thick ascending limb of Henle's loop. In addition, there is evidence to suggest that THP may play a role in prevention of bacteria from adhering to urothelial receptors (17) and in kidney and systemic immunity (37).

The THP polypeptide (Figure 2) consists of an N-terminal signal peptide followed by an elastase-sensitive fragment containing one EGF-like domain, two Ca^{2+}-binding EGF-like domains, a ZP domain, and a GPI-anchor close to the C terminus. After processing of a GPI-linked precursor (38), THP is secreted into the urine as a highly glycosylated species that self-assembles into filaments with molecular weights of $\sim 10^4$ kDa (39).

Transforming Growth Factor-β Receptor Type III and Endoglin

Transforming growth factor-β TGF-β receptor type III (TGFR3), or betaglycan, is the most abundant TGF-β binding protein at the cell surface. Among its potential functions, TGFR3 appears to be essential for restructuring of blood vessels during angiogenesis in mammals (40). It has an N-terminal signal peptide, a ZP domain, a C-terminal putative TMD, and a cytoplasmic tail (41–43). TGFR3 binds all 3

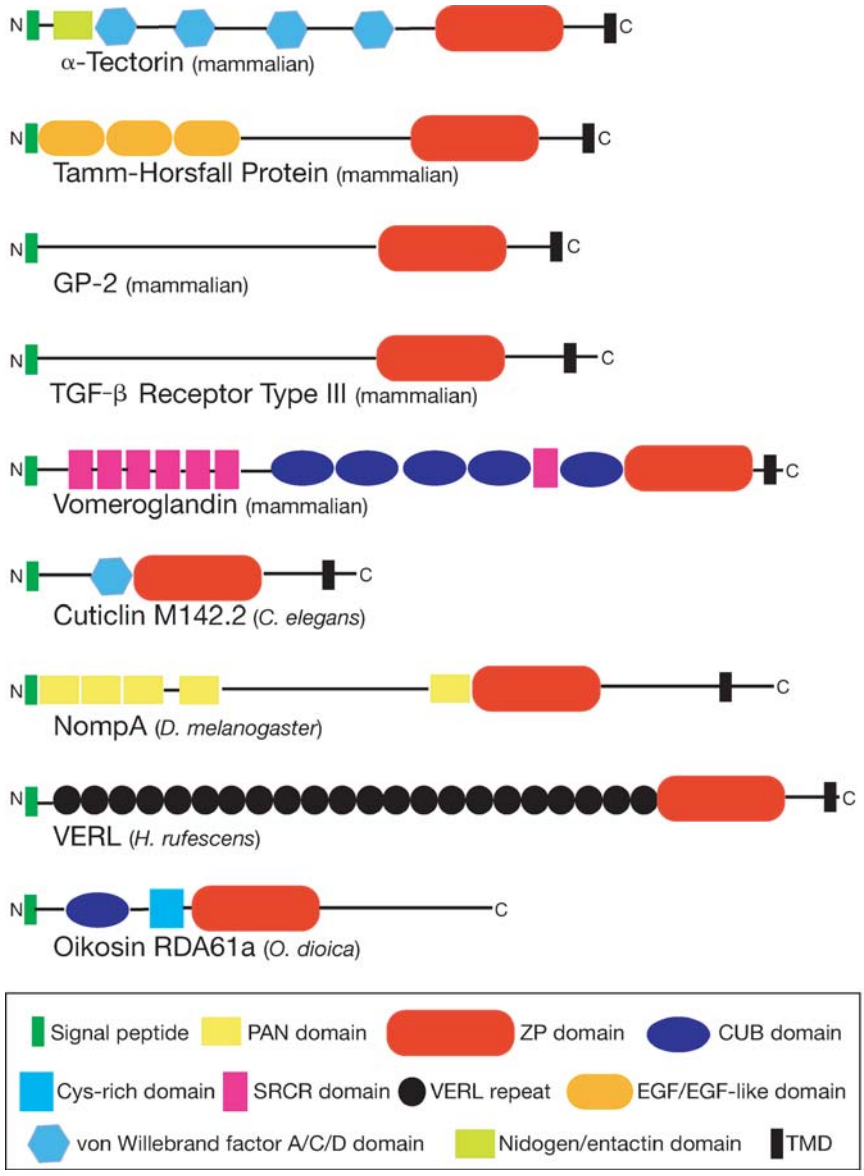


Figure 2 Schematic representation of the overall architecture of several mammalian and nonmammalian ZP domain proteins. The polypeptide of each ZP domain protein is not drawn to scale, but the N and C termini are indicated.

TGF-β isoforms with high affinity and is thought to facilitate binding of TGF-β to TGF-β type II receptor (41, 42, 44, 45). Endoglin ($\sim$180 kDa M_r; disulfide-linked homodimer), or CD105, is a membrane glycoprotein that is structurally related to TGFR3. Endoglin binds TGF-β isoforms 1 and 3 with high affinity through its association with the type II receptor (46, 47) and plays an important role in cardiovascular development and vascular remodeling (48). Mice lacking endoglin exhibit defective angiogenesis (49). Interactions between endoglin and TGFR3 may achieve a balance between the positive and negative regulation of TGF-β signaling pathways (50, 51), perhaps explaining the relationship between these proteins and cancer (19, 48).

Liver-Specific ZP Domain–Containing Protein

Human and mouse liver-specific ZP domain–containing proteins (LZPs) are secreted, liver-specific proteins that are involved in hepatocellular function and development (52). The polypeptide of LZP contains 546 aa organized as an N-terminal signal peptide, three EGF domains, and a C-terminal ZP domain; the EGF domains just precede the ZP domain, unlike their location in THP. No TMD or GPI-anchor has been identified, but a short polypeptide extension follows a CFCS C-terminal to the ZP domain.

GP-2 Protein

GP-2 is a GPI-linked glycosylated protein that is the most abundant protein in the pancreatic exocrine secretory granule, the zymogen granule (53–55). Coincident with exocrine secretion, GP-2 is released from the membrane and deposited into the pancreatic duct. It has been proposed that GP-2 may function in regulated membrane trafficking along the apical secretory pathway (56) and may be involved in diseases such as cystic fibrosis. GP-2 exhibits extensive sequence similarity to THP (57, 58) and, like THP, can form high molecular weight aggregates that obstruct the pancreatic duct in chronic pancreatitis (59). The GP-2 polypeptide ($\sim$58 kDa M_r) possesses an N-terminal signal peptide, a ZP domain, and a C-terminal GPI-anchor (Figure 2).

CRP-Ductin Gene Products

Transcripts from the mouse *CRP-ductin* gene (60) have been shown to undergo tissue-specific alternative splicing (61, 62), producing a group of proteins known as mouse pancreatic acinar cell glycoprotein, pro-Muclin/CRP-ductin(-α), mouse vomeronasal organ protein vomeroglandin, rat salivary gland ebnerin, rabbit kidney hensin, as well as human proteins DMBT1 and gp340. All of these proteins are highly glycosylated, possess an N-terminal signal peptide, and share a common modular organization, with different combinations of SRCR and CUB domains upstream of a ZP domain and a putative TMD close to the C terminus (note: DMBT1 and hensin do not have a TMD).

MUCLIN (GP300) Muclin is a major sulfated, mucin-like glycosylated protein product of the *CRP-ductin* gene (63) that is expressed in the mouse intestinal tract, pancreas, and gall bladder (64). It is derived by proteolytic cleavage of promuclin with the concomitant production of muclin (~300 kDa M_r; ~210 kDa M_r polypeptide), a glycosylated protein possessing eight SRCR, five CUB, and a ZP domain, and an ~80 kDa M_r membrane glycosylated protein (64). Muclin undergoes increased expression and altered posttranslational processing in CFTR (cystic fibrosis transmembrane conductance regulator) knockout mice (65) and is responsible in part for the cystic fibrosis phenotype. In general, CFTR mice are characterized by increased secretion of glycoconjugates with altered carbohydrate composition, and these may contribute to the pathology of cystic fibrosis by obstructing luminal spaces.

VOMEROGLANDIN Of the two chemoreceptive systems in vertebrates, the olfactory and vomeronasal systems, the latter plays an important role in perception of pheromones. The olfactory epithelium and vomeronasal organ of the nose are the sites of the olfactory and vomeronasal systems, respectively. Differential screening of mouse cDNA libraries resulted in the identification of a glycosylated protein, called vomeroglandin, that is localized specifically to the vomeronasal complex (66). Vomeroglandin is a subform of CRP-ductin (CRP-α) and bears homologies with ebnerin, DMBT1, and hensin, especially with respect to organization of its various structural domains. A full-length cDNA for vomeroglandin contains an open reading frame that encodes a protein of 1957 aa (~235 kDa M_r). Vomeroglandin is an N-glycosylated protein (21 potential sites) that has seven SRCR domains, five CUB domains, and a single ZP domain close to the C terminus of the protein (Figure 2). In addition, the protein has an N-terminal signal peptide, a C-terminal putative TMD, and a short cytoplasmic tail. Vomeroglandin is present in both membrane-bound and soluble (present in the lumen of the vomeronasal organ) forms, with the latter possibly involved in pheromone perception.

EBNERIN Von Ebner's glands are unique salivary glands, contained within the tongue, whose secretions modulate taste perception. Ebnerin is a protein found in ducts of von Ebner's glands and released into fluid bathing the taste buds in taste papillae (67). The ebnerin polypeptide contains 1290 aa and consists of four SRCR domains, three CUB domains, and a ZP domain close to the C terminus. There are a putative CFCS and TMD and a short cytoplasmic tail at the C terminus of the polypeptide, and there are 15 potential N-linked glycosylation sites. Despite the presence of a TMD, ebnerin exists mainly as a soluble protein, suggesting that the domain is removed by proteases.

DMBT1 AND GLYCOPROTEIN-340 DMBT1 (deleted in malignant brain tumors 1) is a mucin-like molecule, a candidate tumor suppressor gene for human brain, lung, and digestive tract cancer, and it may be a regulator of mucosal homeostasis (18). Homozygous deletions in the gene and lack of its expression have been observed

frequently in these cancers. Alternative splicing of *DMBT1* transcripts gives rise to a large number of isoforms, including gp340 (glycoprotein-340; salivary agglutinin), which is an isoform found in lung macrophages and binds to surfactant protein D (68, 69). Apparently, the expression pattern and subcellular distribution of DMBT1 in the human alimentary tract is reminiscent of epithelial mucins, especially MUC1, MUC5B, and MUC6. The DMBT1/gp340 prototype polypeptide contains 14 SRCR domains, 2 CUB domains, and a ZP domain.

HENSIN Hensin, an ECM protein, is expressed in virtually all epithelia and brain, is deleted in a large number of epithelial tumors, and is involved in apical secretion and endocytosis (70, 71). Initially, hensin assembles into dimers and tetramers and then, in the presence of gelactin-3, assembles into long fibers in the ECM (72). The hensin polypeptide (~230 kDa M_r) contains eight SRCR domains, two CUB domains, a ZP domain, followed by a CFCS and a short tail (62).

UTCZP, ERG-1, UO-44, and ITMAP-1

Mouse (uterine cub zona pelludica protein) *UTCZP* is expressed in the uterus 6 days prior to birth, but it is not expressed in the nonpregnant uterus or in a variety of adult and fetal tissues (74). *UTCZP* encodes a protein that contains 606 aa (~68 kDa M_r). The polypeptide consists of an N-terminal signal peptide followed by two contiguous CUB domains, a ZP domain, a putative TMD, and a short C-terminal tail. It is proposed that UTCZP may play an important role in events that transpire during late pregnancy. A highly related rat gene, *ERG-1* (estrogen-regulated gene-1), is regulated by estrogen, and its expression is restricted to the uterus and oviduct (75), whereas the identical rat gene, called *UO-44* (uterine-ovarian-specific gene 44), is also expressed in granulosa cells of the ovary (76, 77); its human homologue is highly expressed in the pancreas (77). ERG1 contains 607 aa and possesses an N-terminal signal peptide followed by two contiguous CUB domains, a ZP domain, a putative TMD, and a short C-terminal tail. Unlike mouse UTCZP, rat ERG1 and UO-44 are expressed in the nonpregnant uterus.

Mouse itmap-1 (integral membrane-associated protein-1) is a prominently expressed protein in pancreatic acinar cells where it is localized to zymogen granule membranes (73). The itmap-1 polypeptide (~69 kDa M_r; 110 kDa M_r glycosylated product), which consists of two CUB domains, a ZP domain, a TMD, and a short cytoplasmic tail, is located within the granules and may be identical to UTCZP. Interestingly, homozygous nulls of *Itmap-1* exhibit increased severity of pancreatitis, apparently attributable to impaired activation of trypsinogen, but these nulls have no effect on pregnancy.

PLAC1 and OOSP1

PLAC1 (placenta-specific 1; X-linked), one of several genes predominantly expressed in the placenta, encodes an extracellular protein that contains 212 (human)

or 173 (mouse) aa (78–80). The protein has an N-terminal signal peptide and bears significant homology to the N-terminal region of the ZP domain of ZP3. It is proposed that the protein may regulate specific interactions between the trophoblast and other placental or maternal components. On the other hand, mouse *Oosp1* (oocyte-secreted protein 1) is expressed at high levels in ovarian oocytes (81) and in the liver (82) but at much lower levels in the spleen (81). It encodes a 202-aa secreted protein that contains five potential N-linked glycosylation sites and shares amino acid identity with mouse PLAC1; four of six Cys residues in mouse PLAC1 and Oosp1 are conserved and correspond to Cys residues 1–4 of the ZP domain. Oosp1 is thought to play a role in signaling from the oocyte to surrounding somatic (follicle) cells.

NONMAMMALIAN ZP DOMAIN PROTEINS

Eggs from virtually all nonmammalian species are surrounded by a thin, filamentous, extracellular coat, often referred to as the vitelline envelope (VE). The VE performs some of the same functions as the ZP of mammalian eggs, and in this context, VE proteins closely resemble ZP proteins, are often glycosylated, and possess a single ZP domain close to their C terminus. In recent years, VE proteins from amphibian (*Xenopus laevis*), bird (chicken and quail), fish (e.g., zebrafish, rainbow trout, and medaka), mollusc (abalone), echinoderm (sea urchin), and tunicate (sea squirt) eggs have been characterized in some detail (2, 83–87). Interestingly, whereas all ZP glycoproteins are synthesized in the ovary, there is evidence that certain VE glycoproteins in birds and fish are synthesized by the liver and transported in the bloodstream to the ovary; synthesis of these glycoproteins is under estrogen control (2, 84, 85).

In addition to VE components, the ZP domain has been found in other kinds of proteins from flies (*D. melanogaster*), worms (*C. elegans*), and urochordates (*Oikopleura dioica*). Some of these ZP domain proteins, described below, have been shown to affect the morphology of organisms, organization of their tissues, as well as mechanosensation and fertilization.

Haliotis rufescens

VERL VERL (vitelline envelope receptor for lysin) is the major glycoprotein of the red abalone (*H. rufescens*) egg VE that is recognized in a species-specific manner by sperm lysin (87–89) during fertilization. The interaction between sperm lysin and egg VERL enables sperm to penetrate the VE. VERL consists of a ~3722-aa (~411 kDa M_r) polypeptide that is heavily glycosylated (112 potential N-linked glycosylation sites), resulting in a ~1000 kDa M_r glycoprotein. The polypeptide consists of an N-terminal signal peptide, 22 consecutive VERL repeats (~150 aa each), a C-terminal ZP domain, a potential CFCS, and a putative TMD (Figure 2). Therefore, vertebrate and invertebrate egg coat sequences share an homologous ZP domain.

Drosophila melanogaster

DUMPY AND PIOPIO The product of the *dumpy* gene of *D. melanogaster*, a gigantic ECM molecule, regulates mechanical forces (90). The gene is expressed within the invaginating ectoderm of the embryo, the developing trachea, muscle attachment sites (tendon cells that mediate muscle attachment to the cuticle body wall), and pupal wings. *dumpy* mutant phenotypes affect the size and shape of the limbs, thoracic cuticle, trachea, and mouthparts. It would appear that Dumpy is a membrane-anchored fiber ($\sim$1 μm in length) that is present within the cuticle and provides a strong anchor for the underlying tissue, thereby allowing it to maintain mechanical tension at sites under stress.

The *dumpy* locus is predicted to encode a 70-kb transcript that has an open reading frame encoding a $\sim$2.5 MDa M_r protein. Dumpy is a modular protein that consists of an N-terminal signal sequence, 308-epidermal growth factor modules interspersed with 185 copies of a novel 21-aa repeat (DPY module), 30 copies of a 101-aa, threonine/serine-rich repeat [PIGS-FEAST (P-F) repeat], a proline-rich region, a ZP domain, as well as a C-terminal putative TMD, and a short cytoplasmic tail (90). The TMD is thought to anchor the molecule in the epidermal cell membrane, and the ZP domain may cross-link the molecule to other ECM components. Structural information about the EGF-DPY-EGF modules, which represent the majority of the protein, suggests that these modules form a relatively rigid, extended structure, at least 5 nm in length. Consequently, the full length of the molecule would be $\sim$0.8 μm.

Recently, the products of two genes, *dumpy* and *piopio*, were shown to be essential for the generation of an interconnected tracheal system in *D. melanogaster* (4, 91, 92). Like *dumpy*, *piopio* encodes a secreted ZP domain protein but is considerably smaller than Dumpy. Piopio consists of 462 aa ($\sim$50 kDa M_r) and has an N-terminal signal sequence, followed by a ZP domain, a potential CFCS, a putative TMD, and a short cytoplasmic tail. It is proposed that Dumpy and Piopio interact with each other, possibly through their ZP domains, to form filaments that provide a structural network in the lumenal space of the trachea.

MINIATURE AND DUSKY The *miniature* and *dusky* genes of *D. melanogaster* are expressed in cuticle-secreting epithelia and are involved in generating the morphology of adult wing cells (93, 94). For both *miniature* and *dusky* mutants, wings have a normal number of correctly patterned epidermal cells, but the size of individual cells is significantly reduced. Consequently, wings of mutant flies are smaller than wild-type wings.

Miniature is a 682-aa protein ($\sim$75 kDa M_r) that resembles Dusky, a 699-aa protein ($\sim$77 kDa M_r) (93, 94), and both proteins resemble Cuticlin-1 in *C. elegans*. Both proteins have an N-terminal signal peptide, a ZP domain, potential CFCS, putative TMD, and short C-terminal cytoplasmic tail. The ZP domains of Miniature and Dusky exhibit 45% identity with one another, whereas a third related protein, named Dusky-like, has a ZP domain that is 70% identical with that of Dusky.

NOMPA For sensory organs that respond to mechanical stimuli, mechanosensitivity is attributable to extracellular structures linked directly to mechanically gated ion channels in sensory neurons. For example, this is the case for both tactile and auditory stimuli. *D. melanogaster* mechanosensory organs, such as bristles, campaniform sensilla, and chordotonal organs, include an extracellular structure, the dendritic cap. The cap is made by a specific support cell and contacts the neuronal ciliary sensory process. Contacts between sensory endings and support structures are disrupted in *nompA* (no-mechanoreceptor-potential A) mutants (95).

nompA is transcribed by a single support cell, probably the scolopale/sheath cell, in each type I larval sensory organ (96). In its pattern of expression and ZP domain architecture, *nompA* is similar to components of the tectorial membrane described above. In *nompA* mutants, gaps appear between the neuronal sensory process and cuticular sensory structures in bristles and campaniform sensilla. This results in loss of touch- and sound-evoked electrophysiological signals and, consequently, in defects in touch response, hearing, and sensing of body position (proprioception). The structure and location of NompA suggest that it forms part of a mechanical linkage required to transmit mechanical stimuli to the transduction apparatus.

The *nompA* gene was identified by positional cloning, and the NompA protein was shown to be localized specifically to the dendritic cap (96). *nompA* is predicted to encode a single-pass transmembrane protein (1557 aa; ~175 kDa M_r) that may be cleaved to release a large, multidomain extracellular fragment that includes a ZP domain (1254 aa; ~142 kDa M_r). The protein contains four consecutive PAN modules, a middle region (~500 aa), a fifth PAN module, a ZP domain, a tetrabasic motif that could serve as a CFCS, and a C-terminal cytoplasmic tail (Figure 2).

Caenorhabditis elegans

CUTICLINS The cuticle of the nematode *C. elegans* is an ECM that covers the entire organism, forming its exoskeleton and determining the shape of the animal. The cuticle consists for the most part of short collagens that are synthesized by the underlying hypodermis and become extensively cross-linked. However, an additional important component of the cuticle are the Cuticlins (97, 98), which include Cuticlin-1, -3, and -5 (CUT-1, -3, -5) and M142.2 (CUT-6), all of which are present at the dauer larva stage; a stage that worms enter when overcrowded or when food is scarce. CUT-1 and -6 are essential for assembly of the alae, two external lateral strips that run the length of the worm and are present at several stages of development, including the adult stage. Loss of function of CUT-6 by RNAi results in "dumpy" dauers that lack alae. All Cuticlins, except CUT-2, possess a ZP domain followed by a putative TMD and cytoplasmic tail. For example, CUT-6 (572 aa) has an N-terminal signal peptide followed in order by von Willebrand factor type-A, ZP, and putative TMDs, as well as a C-terminal cytoplasmic tail (Figure 2). There is a potential CFCS 14 aa downstream of the ZP domain.

Oikopleura dioica

OIKOSINS The *Appendicularia* are pelagic urochordates that live inside complex extracellular mucous houses composed of different compartments, valves, septa, and filter sets. The latter are used by the organism to filter food particles from seawater. In order to maintain sufficient filtration rates, the house is reconstructed repeatedly (every 3–4 h during the postmetamorphic phase of the life cycle) by secretion of at least 20 different related proteins, called oikosins, by the oikoplastic epithelium, a monolayer of cells that covers the trunk of the animal. Construction of a functional house requires synthesis of a particular oikosin, by the correct cell type, at a specific time. Recently, at least seven families of genes (*oikosins*), expressed by specific subregions of the oikoplastic epithelium, were identified and characterized for *O. dioica* (99, 100).

All oikosins are predicted to have an N-terminal signal peptide, both N- and O-linked oligosaccharides, and in some cases glycosaminoglycan attachment sites. Some, but not all, oikosins exhibit weak to modest similarities to known extracellular proteins, such as bone morphogenetic protein, tolloid, cubilin, and bovine mucins. These similarities are restricted to regions of CUB domains located close to the N terminus of the protein. In some cases, more than one CUB domain is present. Additionally, some of the oikosins possess a ZP domain downstream of both the CUB domain(s) and a Cys-rich region predicted to be present as disulfides (Figure 2). The ZP domain resembles that of THP.

ZP DOMAIN PROTEIN STRUCTURE

Despite their varied biological functions, a common set of features has emerged that is shared by most ZP domain proteins. Of particular interest are the molecular structure of these proteins, especially of the ZP domain itself, and the supramolecular organization of ZP domain proteins.

Organization of Genes Encoding ZP Domain Proteins

Analysis of ZP protein genes indicates that they share a conserved organization. For example, mouse and human ZP1 genes consist of 12 exons, whereas ZP2 genes consist of either 18 (mouse) or 19 (human) exons, and ZP3 genes consist of 8 exons. Although intron size varies extensively between species, the number and length of corresponding exons is highly conserved (101–104). This reflects the fact that exon/intron boundaries define the limits of distinct domains within ZP proteins, as recently confirmed by analysis of a cluster of genes encoding the major VE protein of zebrafish (105), identification of PLAC1 and Oosp1 proteins homologous to exons 1 and 2 of ZP3 (78, 79, 81, 82), and analysis of limited protease digests of ZP proteins (106, 107). Similarly, the homologous region of genes encoding THP and GP-2 contains identical exon/intron boundaries, with the first conserved exon encoding a polypeptide fragment that starts immediately before a protease-sensitive region of human THP, just N-terminal of the ZP domain (108). In this

context, it is noteworthy that alternative splicing of the human *GP-2* gene generates a truncated protein isoform whose mature sequence almost corresponds exactly to the protease-resistant ZP domain fragment of THP (109).

ZP genes also share TATAA boxes ~30 bp upstream of their transcription start sites, as well as E-box sequences (CANNTG) at ~−200 bp. Apparently, the latter are responsible for oocyte-specific expression that occurs in a coordinate manner upon binding of E12/FIG-α heterodimers to the E boxes (22, 110–112). E boxes have also been found in fish VE protein genes (105, 113), suggesting that the structure and regulation of egg coat protein genes have been conserved throughout vertebrate evolution.

General Features of ZP Domain Proteins

With two possible exceptions (67, 114), all ZP domain proteins share N-terminal signal peptides that target them to the secretory pathway through cotranslational import into the endoplasmic reticulum (ER) (Figure 2). In the ER, signal peptides are proteolytically removed, and proteins fold into their native structure while acquiring disulfide bonds. Additionally, in the ER, ZP domain proteins are generally modified with a variable number of high-mannose type, N-linked oligosaccharides. Upon transfer to the Golgi, the proteins can be modified further by addition of O-linked oligosaccharides and processing of high-mannose type, N-linked oligosaccharides to complex type. However, in some cases high-mannose type oligosaccharides remain unmodified in mature proteins, as shown for a single N-linked oligosaccharide that allows human THP to specifically bind type 1-fimbriated *Escherichia coli* (115). N-linked oligosaccharides have been shown to be required for secretion of individual ZP domain proteins, as in the case of ZP2 (2, 116), whereas O-linked oligosaccharides are thought to be responsible for the biological function of others, such as sperm receptor ZP3 (5). In general, the position of N-linked glycosylation sites is conserved between ZP domain proteins that are homologous but not necessarily between different families of ZP domain proteins. Furthermore, a few ZP domain proteins lack both N- and O-linked oligosaccharides (e.g., VEα and VEβ proteins of fish egg coats) (117). Therefore, although glycosylation is a modification common to the majority of ZP domain proteins, its extent and relevance in terms of protein structure and function can vary significantly.

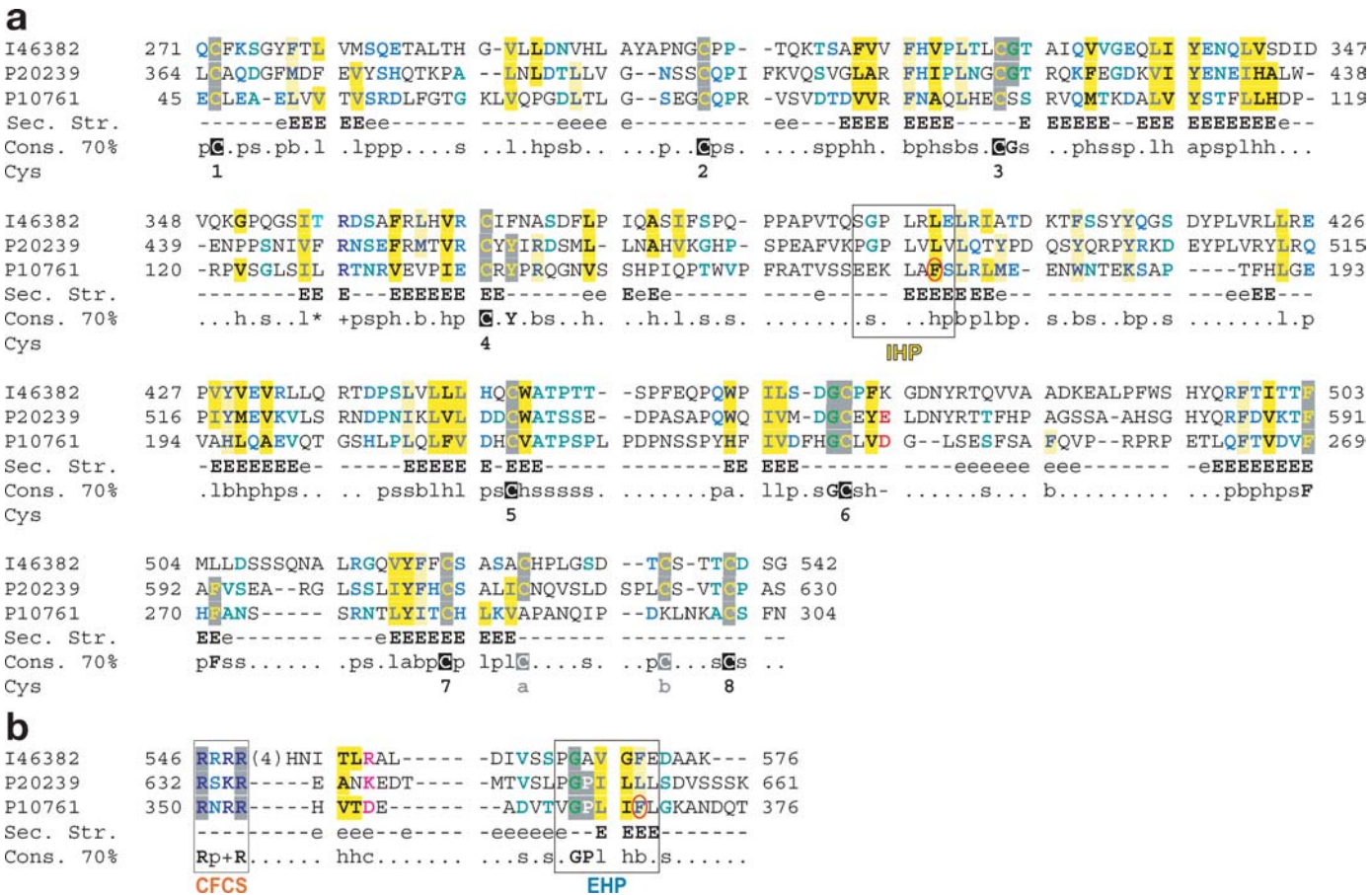
As discussed above, ZP domain proteins are characterized by a highly mosaic architecture, with a variable number and different combinations of structural domains between the signal peptide and a single ZP domain close to the C terminus (Figure 2). Following the latter, precursors of ZP domain proteins generally share a stretch of hydrophobic amino acids that either constitutes a single-spanning transmembrane helix or directs attachment of a GPI-anchor to nascent polypeptides in the ER. These features localize the proteins to secretory vesicles that transport them to the plasma membrane of the cell (118). Here, the majority of ZP domain proteins are released into the extracellular space upon proteolytic cleavage (occurring either within the *trans*-Golgi network or at the plasma membrane) at conserved basic sites between the ZP domain and the C-terminal membrane-anchoring elements (26, 38, 58, 118–126). However, certain ZP domain proteins, like endoglin, lack

such cleavage sites and apparently remain permanently associated with the plasma membrane (48). Although the C-terminal TMD of mammalian ZP proteins is not required for secretion (108, 127, 176), it indirectly plays a crucial role in polymerization of the proteins (107) (see below). Similarly, whereas deletion of the short cytoplasmic tail of TGFR3 does not affect protein targeting to the plasma membrane, it abrogates phosphorylation-dependent, β-arrestin 2-mediated endocytosis of the receptor (128). The cytoplasmic domain of the other TGF-β coreceptor, endoglin, is responsible for specific interactions with zyxin (129) and ZRP-1 (130) that interfere with their localization at focal adhesion sites. In other cases, however, the cytoplasmic tail is important for targeting (131).

Primary Structure of the ZP Domain

The ZP domain sequences of mouse ZP1–3, aligned and color coded according to the consensus for the current SMART ZP domain database (14), are shown in Figure 3*a*. Even using a relatively low 70% consensus threshold, the only invariant residues within the ~260 aa of the ZP domain are ten Cys residues (eight of which

Figure 3 ZP domain and C-terminal propeptide sequence features. (*a*) ZP domain sequences of mouse ZP1–3, aligned to the SMART ZP domain family [SM00241; (14)], are color coded (183) according to the consensus for the whole family (for clarity, no other sequences within the alignment are shown). Sequences are indicated by the accession numbers of the corresponding full-length proteins, to which the amino acid numbers also refer. Below the alignment, a *Jnet* secondary structure prediction (133) for ZP1–3 and a consensus sequence for the whole alignment, generated using a 70% threshold, are reported. Consensus keys are minus sign (−), negative (D, E); +, positive (H, K, R); p, polar (C, D, E, H, K, N, Q, R, S, T); *, (S, T); l, aliphatic (I, L, V); s, small (A, C, D, G, N, P, S, T, V); a, aromatic (F, H, W, Y); b, big (E, F, H, I, K, L, M, Q, R, W, Y); h, hydrophobic (A, C, F, G, H, I, L, M, T, V, W, Y); period (.), no consensus. Uppercase characters within the consensus sequence indicate the specific amino acids with the same one-letter code; conserved ZP domain Cys residues 1–8 and additional conserved Cys residues a, b of ZP1/ZP2-like proteins are numbered and highlighted in black and gray, respectively. Secondary structure keys are E, β-sheet (predicted for the corresponding amino acids of all mouse ZP proteins); e, β-sheet (predicted for the corresponding amino acids of only one or two of the mouse ZP proteins); minus sign (−), coil. Internal hydrophobic patch (IHP) is boxed; amino acid F171, whose mutation also impairs assembly but not secretion of mouse ZP3 (107), is circled in red. (*b*) C-terminal propeptide sequences of mouse ZP1–3 are aligned to the corresponding region of other ZP domain protein precursors (107) and annotated as above. Consensus furin-cleavage site (CFCS) and external hydrophobic patch (EHP) are boxed; amino acid F368 (107), whose mutation (deletion of EHP) impairs assembly but not secretion of mouse ZP3, is circled in red. TMD-containing propeptide sequences C-terminal to the conserved EHP are not shown.



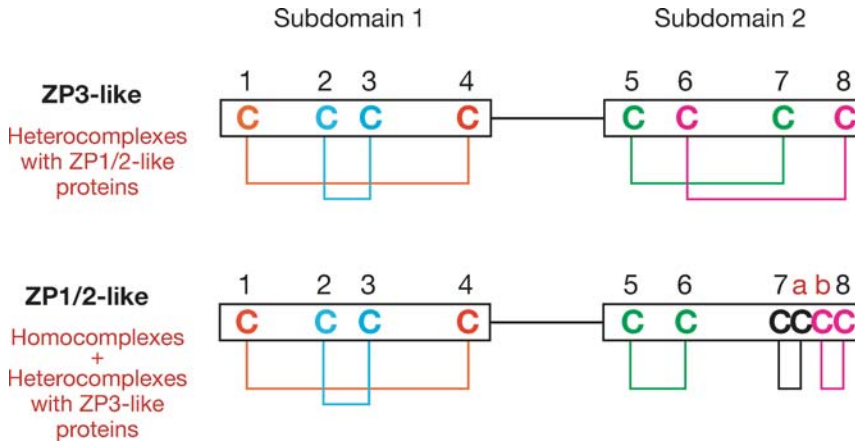


Figure 4 ZP domain disulfide bridges. Cys residues that are disulfide bonded within the ZP domain of mouse ZP (121), and fish VE (132) proteins are indicated with identical colors and connected by brackets. Whereas the connectivity of conserved Cys residues 1–4 within the N-terminal half of the ZP domain is identical in all ZP proteins, the presence of two additional Cys residues (a, b) within the ZP domain of ZP1/ZP2-like proteins (i.e., proteins with ten Cys residues in the ZP domain) makes their C-terminal connectivity different from that found in ZP3-like proteins (i.e., proteins with eight Cys residues in the ZP domain). Considering that ZP3-like proteins always form heterocomplexes with ZP1/ZP2-like proteins, whereas the latter can also form complexes in the absence of the former, it is possible that Cys residue connectivity plays an important role in specifying recognition between ZP domain proteins.

correspond to the first eight Cys residues of the ZP3 sequence; see below and Figure 4), two Gly residues immediately after and before the third and sixth Cys residues, respectively, and three aromatic residues. However, there are a relatively large number of positions with conserved physicochemical character, specifically, mainly polar (∼15%), small (∼15%), and hydrophobic (∼7.5%) amino acids. Consistent with these findings, recent studies strongly suggest that disulfide bonds involving the conserved Cys residues are crucial for the three-dimensional folding and the biological function of the ZP domain (29–31, 108, 121, 122, 132) (see below).

Secondary Structure of the ZP Domain

Predictions of the secondary structure of ZP domain proteins, based on analysis of both single sequences or aligned nonredundant sequence databases with the *Jnet* algorithm (133), were found to be essentially in agreement with previous analyses (101) (Figure 3*a*). Overall, ZP domain proteins are predicted to be relatively rich in β-structure (for example ∼21%, ∼32%, and ∼28% for mouse ZP1–3,

respectively), with very little α-helix content (<1.5% for all three ZP proteins). The percentage of predicted β-structure is even higher when only ZP domain sequences are considered (~34% to 46% for mouse ZP1–3 ZP domains), suggesting that the domain itself probably adopts an all-β fold.

Results of theoretical analyses are supported by a number of experimental reports on different systems. Attenuated total reflectance-infrared spectroscopy (ATR-IR) and Fourier Transform (FT)-Raman spectroscopy studies of intact VEs from teleostean fish detected a large fraction of β-sheets, together with a significant contribution of β-turns and a minor random coil component (134, 135). Regression analysis techniques, as well as spectral deconvolution data, suggest that the contents of β-sheets and β-turns are ~60% and 30% to 40%, respectively. Some features of the Raman spectra also indicate that the β-sheets of VE proteins are most likely antiparallel. Circular dichroism (CD) studies on human THP purified from native sources suggest that this ZP domain protein also adopts an overall β-fold (32% to 34%), with an α-helix content below 10% (39, 136). Similar conclusions were drawn from analysis of refolded monomeric THP, suggesting that the observed β-structure is primarily intramolecular and not strictly dependent on polymer assembly (L. Jovine and P.M. Wassarman, unpublished results). Finally, Fourier Transform-infrared (FT-IR) and CD spectroscopy studies on recombinant Cuticlin-1, both as inclusion bodies and in solubilized form, suggest that it contains ~50% β-structure, ~14% α-helix, and ~25% turns (137); furthermore, the structure of Cuticlin-1 was found to be unusually resistant to high temperature, as well as stabilized by detergents such as sodium dodecyl sulfate (SDS) (138).

Tertiary Structure of the ZP Domain

Because of their unique chemical properties, Cys residues play a fundamental role in the three-dimensional folding and biological activity of most secreted proteins (139–141). All ZP domain proteins are characterized by a number of conserved Cys residues, eight of which are an essentially invariant portion of the ZP domain signature. Formation of disulfide bonds was found to be the rate-limiting step for secretion of THP (142, 143) and endoglin (144); additional Cys linkages are apparently formed during ZP hardening (145). Mutations leading to either a reduction (29, 31) or an increase (29, 30) in the number of Cys residues within the ZP domain of human α-tectorin lead to deafness, most likely resulting from severe impairment of protein secretion leading to structural alterations of the tectorial membrane (108). From these findings, it is apparent that Cys residues play a major role in the biology of ZP domain proteins.

Numerous studies have increased our understanding of the structural and functional contribution of Cys residues to ZP domain proteins. Seminal work on human THP suggests that its 48 Cys residues are all involved in intrachain disulfide bonds (146). It was shown that, although Cys residues of mouse ZP2 and ZP3 are not responsible for association of the proteins within the ZP, at least some are involved in intramolecular disulfides. In addition, the apparent M_r of mouse ZP1 in the presence of reducing agents was found to be approximately one-half that in the

absence of reducing agents, suggesting that this protein forms homodimers linked by intermolecular disulfides (147). These results are consistent with the notion that all Cys residues of extracellular proteins generally exist in the oxidized state. In fact, mature THP, ZP2, and ZP3 all contain an even number of Cys residues, whereas mouse ZP1 possesses an odd number of Cys residues.

ATR-IR and FT-Raman spectroscopy have provided evidence for the presence of disulfide bonds within VE proteins of teleostean fish (135). Furthermore, CD spectra of purified human THP reveal features typical of disulfides (39, 136), and near-UV spectra of Cuticlin-1 suggest that at least a portion of its conserved Cys residues are involved in disulfides (137). Recently, advances in mass spectrometry have permitted determination of disulfide bond linkages within the ZP domain of native mouse (121) and pig (122) ZP proteins, as well as within fish VE proteins (132). Results with mouse and fish are identical for the ZP domain of homologous ZP/VE proteins (121, 132) (Figure 4). Whereas analysis of pig ZP1 suggests that it could share the same disulfide-bonding pattern as ZP3 homologues (122). Collectively, these results strongly support the contention that the first four conserved Cys residues of the ZP domain are engaged in identical intramolecular disulfide bonds in all ZP domain proteins. This is consistent with the state of PLAC1 and Oosp1 whose homology to other ZP domain proteins is restricted to the sequence encompassing Cys residues 1 to 4 (78, 79, 81, 82).

The presence of two extra Cys residues (a and b, Figure 4) within the C-terminal portion of the ZP domain of ZP1 and ZP2 makes the disulfide-bonding pattern of conserved Cys residues 5–8 different from that of ZP3. It is possible that the former proteins could also undergo disulfide exchange (132), as observed for other extracellular proteins (148). Such disulfide rearrangements could have important implications for specific recognition of individual ZP domain proteins during assembly into higher-order structures (discussed below). Furthermore, in view of other biochemical data and functional studies, the disulfide-bonding pattern of the ZP domain suggests that the domain itself consists of two independently folding subdomains (107, 122, 149) (see below).

Currently, no three-dimensional structure is available for any ZP domain protein. The limited resolution of published electron microscopy (EM) studies does not allow predictions to be made about the overall folding of this class of proteins. Similarly, the use of several established fold-recognition algorithms does not permit reliable predictions to be made (2). On the basis of recent studies (150–152), the availability of experimentally determined disulfide linkages for the ZP domain could have revealed structural similarities with other proteins. However, analyses using this approach have so far failed to detect any convincing matches, further supporting the idea that the ZP domain probably adopts a novel protein fold.

Supramolecular Structure of ZP Domain Proteins

Since the 1950s, a number of ultrastructural studies have been performed on the mammalian ZP, fish VE, human THP filaments, and mouse tectorial membrane.

The following is a brief review of what is currently known about the overall morphology of these ZP domain–containing structures.

MAMMALIAN ZP Results of light and scanning electron microscopic (SEM) analyses suggested that the mouse ZP has a sponge-like, porous structure, consistent with its permeability to relatively large macromolecules such as antibodies and small viruses (20, 153). Subsequent transmission electron microscopic (TEM) and biochemical studies of solubilized mouse ZP revealed that the ~3–4 ng of glycoprotein that give rise to the ~5–7 μm thick envelope are assembled into filaments (153). Although elastase treatment of the ZP generated mostly interconnected filaments containing all three ZP glycoproteins, digestion with chymotrypsin yielded unbranched filaments composed of only ZP2 and ZP3. Because the same result was obtained by solubilization of the ZP with reducing agents, it suggested a model in which ZP2 and ZP3 interact to form filaments that are, in turn, interconnected by disulfide-linked homodimers of ZP1 (20, 153). This model is supported by several lines of evidence. (*a*) The mouse ZP contains near equimolar amounts of ZP2 and ZP3, with ZP1 representing only ~10% of the total mass (20). (*b*) No ZP assembly is observed when synthesis of ZP3 is eliminated by using antisense oligonucleotides or homologous recombination (154–156). Furthermore, oocytes from $ZP3^{+/-}$ heterozygous mice have a ZP that is about half as thick as the ZP surrounding oocytes from wild-type mice (157). In $ZP2^{-/-}$ mice, the relatively less expressed ZP1 protein apparently interacts with ZP3 to form a very thin ZP that is not maintained in preovulatory follicles (158), whereas the ZP of ZP1-null mice appears as a loosely organized matrix with very large pores (159). (*c*) EM analysis of ZP filaments revealed the presence of a structural repeat of 140–150 Å in negatively stained or frozen/shadowed preparations. The periodicity collapsed into ~90 Å beads with ~170 Å spacing in samples that were sprayed, shadowed, and treated with glycerol (153). When ZP filaments were decorated with monoclonal antibodies against ZP2 or ZP3 prior to rotary shadowing, IgG molecules also displayed a periodicity of ~150 Å (160). Finally, a ladder of oligomers generated from a ~180-kDa species composed of both ZP2 and ZP3 could be visualized by SDS-polyacrylamide gel electrophoresis analysis of chemically cross-linked ZP filaments (160).

Because all ZP proteins possess a ZP domain essential for polymerization (108) (see below) and because ZP domains of ZP2 and ZP1, but not ZP3, have identical disulfide linkages (121, 132), it is likely that ZP3 can bind with comparable affinity to either ZP2 or ZP1. In fact, the ZP domains of ZP1 and ZP2 are more homologous to each other (~34% identity) than to ZP3 (~19%). Consequently, ZP filaments probably grow by random addition of either ZP3-ZP2 or ZP3-ZP1 complexes, with the latter being incorporated much less than the former owing to relatively low expression of the *ZP1* gene. The amount of stochastically distributed cross-links between ZP filaments is ultimately controlled by the extent of transcription of the *ZP1* gene. These proposals are consistent with the lack of ZP2 homologues in fish, which instead contain, in addition to ZP3 homologues, multiple ZP1-like

proteins that assemble into a VE having substantial mechanical stability (161). Furthermore, these proposals are compatible with results of reconstitution studies of ZP3-ZP1 complexes (162), as well as with a report showing that high molecular weight components of the fish VE consist of covalently linked heterodimers of ZP3 and ZP1 homologues (132).

FISH VE The general ultrastructure of the envelopes of teleostean fish eggs has been investigated extensively (163). Unlike the ZP, the structural complexity and macromolecular composition of the fish VE vary considerably from species to species. With the exception of salmonids, which are characterized by a monolaminar envelope (117), the VE is a multilayered structure of variable thickness; nevertheless, in all cases, a wave-shaped fibrillar component, which is embedded in an amorphous matrix, can been recognized (163, 164).

MAMMALIAN THP FILAMENTS Because of its availability in large amounts, THP has been the subject of several EM studies since its identification in 1950 (34, 36, 39, 108, 136, 165). Single filaments of the protein, with diameters between 10 and 40 Å, originate from and merge into bundles (av diameter 120 Å) at seemingly irregular intervals, generating a three-dimensional matrix with 0.1–1 μm pores. Analysis of single fibrils revealed a zig-zag course with a periodicity of ~100–140 Å, the single branch of each zig-zag measuring ~60 Å in length and 20–40 Å in width. It was noted that these features could be interpreted as two-dimensional projections of a helical superstructure (36). This interpretation was confirmed by EM analysis of a proteolytic fragment of THP essentially corresponding to the protein's ZP domain. The analysis revealed a double-helical core structure with a pitch of ~120 Å and a diameter of 90–140 Å (108). Similar features were also noted in electron micrographs of rotary shadowed mouse ZP filaments, suggesting that polymers assembled by different ZP domains share a similar overall three-dimensional structure (108).

MAMMALIAN TECTORIAL MEMBRANE Within the mammalian tectorial membrane, tectorins are associated with a striated matrix composed of two types of 7–9 nm diameter fibrils, alternatively light- and dark-staining, connected by staggered cross bridges to form flat sheets. These sheets are, in turn, stacked on top of each other and wrapped around bundles of collagen fibers (27, 28). EM studies of the tectorial membrane of mice homozygous for a targeted deletion of α-tectorin revealed that they lack a striated-sheet matrix. Because β-tectorin could not be detected in the tectorial membrane of these animals, it was hypothesized that α- and β-tectorins interact with each other and polymerize into filaments that constitute the striated matrix (28). This model would explain the hearing impairments that have been linked to either single-site mutations (29–31) or a complete deletion (32) of the ZP domain of human α-tectorin or that are caused by a reduction of β-tectorin levels prior to the onset of hearing in mice (166).

ZP DOMAIN FUNCTION

The ZP domain has been shown to be a conserved module for polymerization of extracellular proteins. The recent identification of sequence elements regulating assembly of ZP domain proteins, in the context of posttranslational processing of the proteins, supports a general mechanism for ZP domain–mediated polymerization.

Role of the ZP Domain in Protein Polymerization

It was first proposed in 1995 that the ZP domain might play a role in polymerization of ZP domain–containing proteins into filaments and/or matrices (26, 33). Subsequently, development of an assay to follow assembly of epitope-tagged recombinant ZP glycoproteins into the ZP of mouse oocytes permitted an experimental approach to investigate the role of the ZP domain (118). Using such an assay, it was shown that the ZP domain together with the N-terminal signal peptide and C-terminal propeptide (Figure 1) are both necessary and sufficient for assembly of nascent protein into the ZP (108). Furthermore, EM analysis of human THP demonstrated that proteolytic digestion of non-ZP domain sequences did not disrupt the organization of THP filaments (108). Together with other relevant findings (167), the results strongly suggest that the ZP domain of extracellular proteins functions as a "polymerization module." Other functions of ZP domain proteins can be ascribed to sequences N- and C-terminal to the ZP domain. For example, sequences involved in the sperm-binding functions of mammalian ZP2 and ZP3, as well as those of ascidian sperm receptor HrVC70, map to regions N- or C-terminal to the ZP domain of all three proteins (5, 86, 168).

Aspects of ZP Domain Protein Polymerization

Extracellular proteins that assemble into filaments or matrices have evolved a variety of strategies to regulate their assembly both spatially and temporally (169–173). For example, these proteins should not polymerize prematurely inside cells. Recent evidence suggests that regulation of ZP domain protein polymerization is achieved through a relatively complex, mutually dependent set of events.

Nearly all precursors of ZP domain proteins are characterized by either a C-terminal TMD or GPI-anchor (Figures 1 and 2). Some ZP domain proteins, such as fish VE proteins synthesized by the liver (161), do not possess these membrane-anchoring elements but do have relatively short extensions C-terminal to the ZP domain. However, in all cases the C-terminal propeptides are lost, either before or during protein polymerization, by proteolytic cleavage at conserved basic sites located immediately after the ZP domain (26, 58, 118, 120–126, 132, 174) (Figure 5). For the majority of ZP domain proteins, the cleavage sites conform to the consensus for furin cleavage (Arg-X-Lys/Arg-Arg) (Figure 3*b*) and are considered to be targets of proprotein convertase enzymes (119, 125, 175) in the Golgi or at the plasma membrane (118, 174, 175). Some ZP domain proteins have a modified cleavage site as, for example, Arg-Lys-X-Arg for fish VE proteins (123, 132). Following

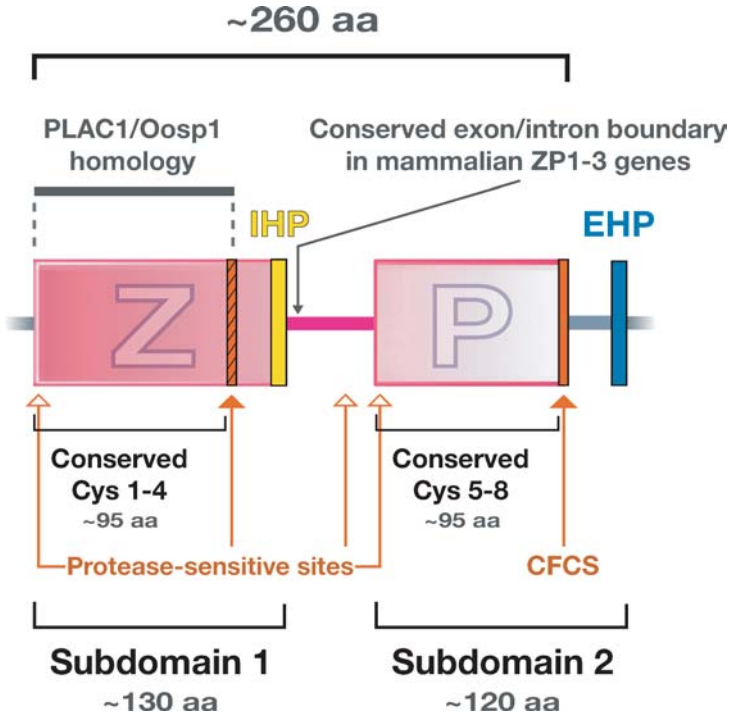


Figure 5 The ZP domain consists of two subdomains. Disulfide bond connectivity (121, 122, 132) (Figure 4), limited proteolytic digestion (106, 107), homology to PLAC1/Oosp1 proteins (78, 81), conservation of exon-intron boundaries (2), identification of glycosylation sites (121, 122), and the relative location of IHP and EHP sequences (107) are all consistent with the view that two subdomains interact to form the entire ZP domain. Although not apparent from their primary structure and disulfide connectivity, it is possible that the two subdomains share a similar three-dimensional structure.

cleavage at basic sites, carboxypeptidase H-like enzymes can trim the C terminus of mature ZP domain proteins (121, 124, 132). Although the specific details of C-terminal processing can vary, this event relies on the polypeptide between the ZP domain and the membrane-anchoring elements being conformationally flexible, thereby accessible to proteases. This may account for the observation that mutation of the CFCS does not impair secretion and assembly of ZP3 in transgenic mice (167); it is likely that a protease cleaved the precursor somewhere else within this flexible region of the polypeptide.

At least two observations suggest that the propeptide of ZP proteins plays a role in both their secretion and assembly. First, recombinant ZP proteins truncated at the CFCS are retained in the ER (119), and second, the propeptide is invariably removed prior to assembly of the mature proteins (26, 58, 118, 120–126, 132, 174).

In addition, although the TMD is not required for secretion of ZP proteins (108, 176), it ensures cleavage of precursors at the CFCS and is essential for assembly in mammals (107, 108). Because a native TMD can be replaced by an unrelated one without affecting secretion and assembly (107), its presence is needed only to anchor precursors in the plasma membrane for proteolytic processing.

Two short hydrophobic motifs are conserved in ZP domain proteins. One, termed an external hydrophobic patch (EHP), is present in the C-terminal propeptide between the CFCS and TMD (Figures 3*b* and 5) (107, 177), and the other, called an internal hydrophobic patch (IHP), is present in the ZP domain itself, following conserved Cys residues 1–4 (Figures 3*a* and 5) (107). Both motifs are predicted to form β-strands. The relative locations of the EHP and IHP are consistent with the proposal that the ZP domain consists of two subdomains, with each subdomain connected to a C-terminal hydrophobic patch by a short, protease-sensitive linker. Although, under some conditions, modification of the EHP apparently can impair secretion of ZP proteins (177), deletion of the EHP does not affect secretion of ZP proteins but rather prevents their assembly (107). Mutation of the IHP also prevents assembly without affecting secretion (107). Secretion is inhibited when either the IHP or EHP is mutated in the context of a ZP protein truncated before the TMD (107). These results strongly suggest that the EHP and IHP are functionally related to each other and, together with the CFCS and TMD, control assembly of ZP proteins.

The observations just described have led to the proposal of a mechanism for activation of polymerization of ZP domain proteins (Figure 6) (107). It is based on the loss of the EHP when the propeptide of ZP protein precursors is cleaved prior to assembly of mature ZP proteins. Presumably, the EHP and IHP of ZP protein precursor interact with each other, and in this manner, the ZP domain is prevented from participating in assembly within the cell. This mechanism probably applies to ZP domain proteins in general because it relies on sequence elements (EHP and IHP) and events (coupling between proteolytic processing and polymerization) that are conserved in different ZP domain proteins. In this context, cleavage of inhibitory sequences from protein precursors with concomitant exposure of polymerization elements has been shown to regulate assembly of several types of proteins (178–182).

FINAL COMMENTS

In recent years the ZP domain has been identified in numerous proteins from both mammalian and nonmammalian sources. It is a frequently found module in extracellular proteins that polymerize into higher-order structures, such as filaments and matrices (e.g., egg ZP and VE, inner ear tectorial membrane, kidney THP filaments, nematode cuticle, and fly tracheal system). ZP domains of different proteins have a relatively large number of amino acids with conserved physicochemical character, and folding of the domain is dependent on disulfides formed between conserved Cys residues. The ZP domain probably adopts a novel

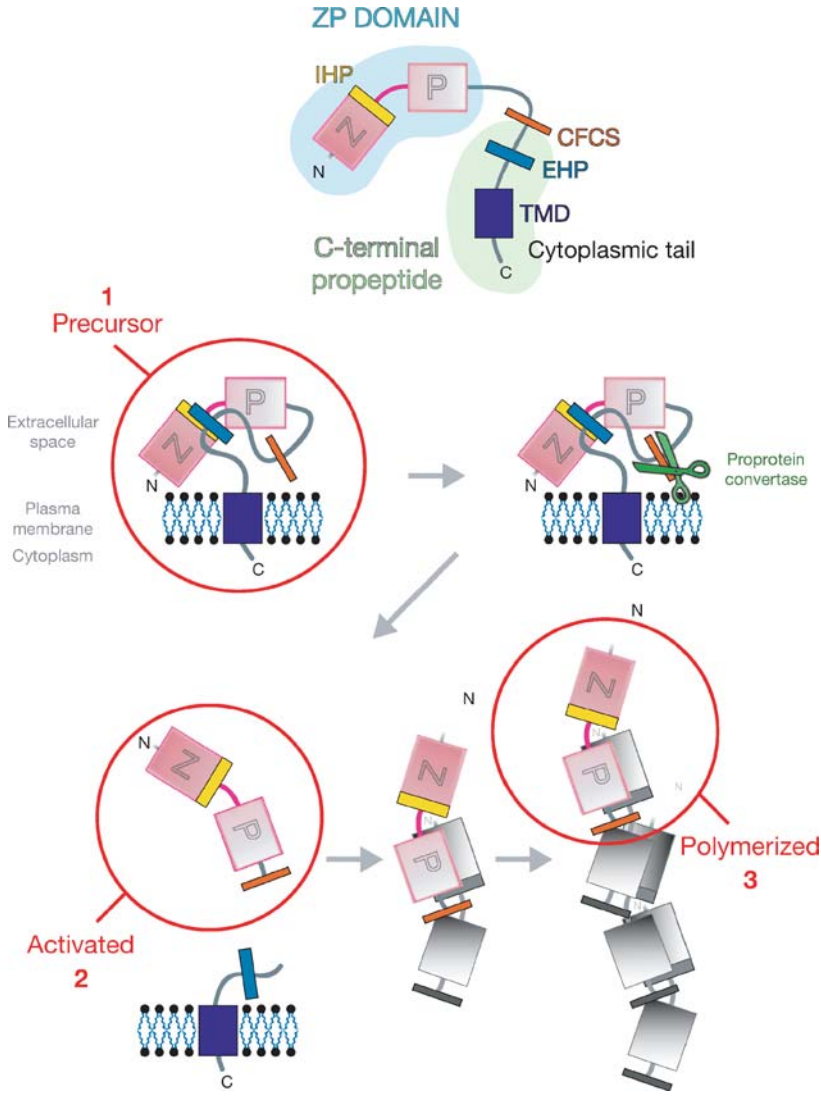


Figure 6 A general mechanism for assembly of ZP domain proteins. In all ZP domain precursors, the ZP domain is followed by C-terminal propeptide that contains a basic cleavage site (such as a CFCS), an EHP, and, in most cases, a TMD or a GPI-anchor (*top panel*). Precursors do not polymerize within the cell either as a result of direct interaction between EHP and IHP or because they adopt an inactive conformation dependent on the presence of both patches (*middle left panel*). C-terminal processing at the CFCS by a proprotein convertase (*middle right*) would lead to dissociation of mature proteins from the EHP (*bottom left*), activating them for assembly into filaments and matrices (*bottom right*).

protein fold. Polymerization of these proteins is attributable to the ZP domain and is controlled by short, hydrophobic sequences in the C-terminal propeptide of protein precursors (EHP) and in the ZP domain itself (IHP). In some cases, ZP domain mutations have been shown to result in loss of protein function and the onset of disease. ZP domain proteins often are glycosylated, have mucin-like properties, and, in some instances, interact with collagens. Frequently, these proteins possess other types of modules (e.g., CUB, EGF, and PAN domains) and perform functions distinct from the structural role played by the ZP domain. For example, ZP domain proteins may serve as receptors or mechanotransducers, owing to sequences that lie outside the ZP domain. Clearly, this is a large and important family of proteins that will continue to grow in size and be of great interest to investigators in several areas of research for many years to come.

ACKNOWLEDGMENTS

Because of editorial restrictions, we used reviews whenever possible and apologize to those investigators whose contributions were not cited directly. We thank Drs. Huayu Qi and Zev Williams for their contributions to our research on the structure and function of ZP domain proteins. We are grateful to previous members of our laboratory for providing a strong foundation on which to build. L.J. was supported by a Human Frontier Science Program long-term fellowship. Research from the authors' laboratory was supported most recently by National Institutes of Health grant HD35105.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Bork P, Sander C. 1992. *FEBS Lett.* 300: 237–40
2. Jovine L, Litscher E, Wassarman PM. 2002. *Adv. Dev. Biol. Biochem.* 12:31–54
3. Muriel JM, Brannan M, Taylor K, Johnstone IL, Lithgow GJ, Tuckwell D. 2003. *Dev. Biol.* 260:339–51
4. Jazwinska A, Affolter M. 2004. *Gene Expr. Patterns* 4:413–21
5. Wassarman PM, Jovine L, Litscher ES. 2001. *Nat. Cell Biol.* 3:E59–64
6. Carr MD. 1992. *Biochemistry* 31:1998–2004
7. Appella E, Weber IT, Blasi F. 1988. *FEBS Lett.* 231:1–4
8. Bork P, Beckmann G. 1993. *J. Mol. Biol.* 231:539–45
9. Fukagawa M, Suzuki N, Hogan BL, Jones CM. 1994. *Dev. Biol.* 163:175–83
10. Tordai H, Banyai L, Patthy L. 1999. *FEBS Lett.* 461:63–67
11. Resnick D, Pearson A, Krieger M. 1994. *Trends Biochem. Sci.* 19:5–8
12. Ruggeri ZM. 2003. *Curr. Opin. Hematol.* 10:142–49
13. Bork P, Downing AK, Kieffer B, Campbell ID. 1996. *Q. Rev. Biophys.* 29:119–67
14. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, et al. 2004. *Nucleic Acids Res.* 32: D142–44
15. Steel KP, Kros CJ. 2001. *Nat. Genet.* 27:143–49

16. Marchuk DA, Srinivasan S, Squire TL, Zawistowski JS. 2003. *Hum. Mol. Genet.* 12(Spec. Issue 1):R97–112
17. Serafini-Cessi F, Malagolini N, Cavallone D. 2003. *Am. J. Kidney Dis.* 42:658–76
18. Kang W, Reid KB. 2003. *FEBS Lett.* 540:21–25
19. Copland JA, Luxon BA, Ajani L, Maity T, Campagnaro E, et al. 2003. *Oncogene* 22:8053–62
20. Wassarman PM. 1988. *Annu. Rev. Biochem.* 57:415–42
21. Bork P. 1993. *Protein Sci.* 2:669–70
22. Epifano O, Liang LF, Dean J. 1995. *J. Biol. Chem.* 270:27254–58
23. Spargo SC, Hope RM. 2002. *Biol. Reprod.* 68:358–62
24. Lefievre L, Conner SJ, Salpekar A, Olufowobi O, Ashton P, et al. 2004. *Hum. Reprod.* 19:1580–86
25. Goodyear RJ, Richardson GP. 2002. *J. Neurobiol.* 53:212–27
26. Killick R, Legan PK, Malenczak C, Richardson GP. 1995. *J. Cell Biol.* 129: 535–47
27. Hasko JA, Richardson GP. 1988. *Hear. Res.* 35:21–38
28. Legan PK, Lukashkina VA, Goodyear RJ, Kossi M, Russell IJ, Richardson GP. 2000. *Neuron* 28:273–85
29. Verhoeven K, Van Laer L, Kirschhofer K, Legan PK, Hughes DC, et al. 1998. *Nat. Genet.* 19:60–62
30. Kirschhofer K, Kenyon JB, Hoover DM, Franz P, Weipoltshammer K, et al. 1998. *Cytogenet. Cell Genet.* 82:126–30
31. Moreno-Pelayo MA, del Castillo I, Villamar M, Romero L, Hernandez-Calvin FJ, et al. 2001. *J. Med. Genet.* 38:E13–16
32. Mustapha M, Weil D, Chardenoux S, Elias S, El-Zir E, et al. 1999. *Hum. Mol. Genet.* 8:409–12
33. Legan PK, Rau A, Keen JN, Richardson GP. 1997. *J. Biol. Chem.* 272:8791–801
34. Porter KR, Tamm I. 1954. *J. Biol. Chem.* 212:135–39
35. Maxfield M, Davis MS. 1963. *Ann. NY Acad. Sci.* 106:288–97
36. Bayer ME. 1964. *J. Cell Biol.* 21:265–74
37. Kokot F, Dulawa J. 2000. *Nephron* 85:97–102
38. Cavallone D, Malagolini N, Serafini-Cessi F. 2001. *Biochem. Biophys. Res. Commun.* 280:110–14
39. Robinson JP, Puett D. 1973. *Arch. Biochem. Biophys.* 159:615–21
40. Bandyopadhyay A, Zhu Y, Malik SN, Kreisberg J, Brattain MG, et al. 2002. *Oncogene* 21:3541–51
41. Lopez-Casillas F, Cheifetz S, Doody J, Andres JL, Lane WS, Massague J. 1991. *Cell* 67:785–95
42. Wang XF, Lin HY, Ng-Eaton E, Downward J, Lodish HF, Weinberg RA. 1991. *Cell* 67:797–805
43. Moren A, Ichijo H, Miyazono K. 1992. *Biochem. Biophys. Res. Commun.* 189:356–62
44. Lopez-Casillas F, Wrana JL, Massague J. 1993. *Cell* 73:1435–44
45. Massagué J. 2000. *Nat. Rev. Mol. Cell Biol.* 1:169–78
46. Fonsatti E, del Vecchio L, Altomonte M, Sigalotti L, Nicotra MR, et al. 2001. *J. Cell Physiol.* 188:1–7
47. Sorensen LK, Brooke BS, Li DY, Urness LD. 2003. *Dev. Biol.* 261:235–50
48. Fonsatti E, Altomonte M, Nicotra MR, Natali PG, Maio M. 2003. *Oncogene* 22:6557–63
49. Li DY, Sorensen LK, Brooke BS, Urness LD, Davis EC, et al. 1999. *Science* 284:1534–37
50. Wong SH, Hamel L, Chevalier S, Philip A. 2000. *Eur. J. Biochem.* 267:5550–60
51. Parker WL, Goldring MB, Philip A. 2003. *J. Bone Miner. Res.* 18:289–302
52. Xu ZG, Du JJ, Zhang X, Cheng ZH, Ma ZZ, et al. 2003. *Hepatology* 38:735–44
53. Scheele GA, Fukuoka SI, Kern HF, Freedman SD. 1996. *Pancreas* 12:1–9
54. Fritz BA, Poppel CS, Fei MW, Lowe AW. 2002. *Pancreas* 24:336–43
55. Rindler MJ, Hoops TC. 1990. *Eur. J. Cell Biol.* 53:154–63

56. Scheele GA, Fukuoka S, Freedman SD. 1994. *Pancreas* 9:139–49
57. Hoops TC, Rindler MJ. 1991. *J. Biol. Chem.* 266:4257–63
58. Fukuoka S, Freedman SD, Yu H, Sukhatme VP, Scheele GA. 1992. *Proc. Natl. Acad. Sci. USA* 89:1189–93
59. Freedman SD, Sakamoto K, Venu RP. 1993. *J. Clin. Invest.* 92:83–90
60. Cheng H, Bjerknes M, Chen H. 1996. *Anat. Record* 244:327–43
61. Mollenhauer J, Herbertz S, Helmke B, Kollender G, Krebs I, et al. 2001. *Cancer Res.* 61:8880–86
62. Takito J, Yan L, Ma J, Hikita C, Vijayakumar S, et al. 1999. *Am. J. Physiol. Renal Physiol.* 277:F277–89
63. De Lisle RC, Petitt M, Isom KS, Ziemer D. 1998. *Am. J. Physiol. Gastrointest. Liver Physiol.* 275:G219–27
64. De Lisle RC. 1994. *J. Cell. Biochem.* 56: 385–96
65. De Lisle RC, Petitt M, Huff J, Isom KS, Agbas A. 1997. *Gastroenterology* 113: 521–32
66. Matsushita F, Miyawaki A, Mikoshiba K. 2000. *Biochem. Biophys. Res. Commun.* 268:275–81
67. Li XJ, Snyder SH. 1995. *J. Biol. Chem.* 270:17674–79
68. Holmskov U, Lawson P, Teisner B, Tornoe I, Willis AC, et al. 1997. *J. Biol. Chem.* 272:13743–49
69. Holmskov U, Mollenhauer J, Madsen J, Vitved L, Gronlund J, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:10794–99
70. van Adelsberg J, Edwards JC, Takito J, Kiss B, Al-Awqati Q. 1994. *Cell* 76:1053–61
71. Al-Awqati Q. 2003. *Annu. Rev. Physiol.* 65:567–83
72. Hikita C, Takito J, Vijayakumar S, Al-Awqati Q. 1999. *J. Biol. Chem.* 274: 17671–76
73. Imamura T, Asada M, Vogt SK, Rudnick DA, Lowe ME, Muglia LJ. 2002. *J. Biol. Chem.* 277:50725–33
74. Kasik JW. 1998. *Biochem. J.* 330:947–50
75. Chen DH, Xu XP, Zhu LJ, Angervo M, Li QX, et al. 1999. *J. Biol. Chem.* 274:32215–24
76. Huynh H, Ng CY, Lim KB, Ong CK, Ong CS, et al. 2001. *Endocrinology* 142:2985–95
77. Leong CT, Ng CY, Ng CP, Ma ZS, Nguyen TH, et al. 2004. *Oncogene* 23: 5707–18
78. Cocchia M, Huber R, Pantano S, Chen EY, Ma P, et al. 2000. *Genomics* 68:305–12
79. Hemberger M, Himmelbauer H, Ruschmann J, Zeitz C, Fundele R. 2000. *Dev. Biol.* 222:158–69
80. Fant M, Weisoly DL, Cocchia M, Huber R, Khan S, et al. 2002. *Mol. Reprod. Dev.* 63:430–36
81. Yan C, Pendola FL, Jacob R, Lau AL, Eppig JJ, Matzuk MM. 2001. *Genesis* 31: 105–10
82. Mano H, Nakatani S, Aoyagi R, Ishii R, Iwai Y, et al. 2002. *Biochem. Biophys. Res. Commun.* 297:323–28
83. Tian J, Gong H, Lennarz WJ. 1999. *Proc. Natl. Acad. Sci. USA* 96:829–34
84. Hyllner SJ, Westerlund L, Olsson PE, Schopen A. 2001. *Biol. Reprod.* 64:805–11
85. Bausek N, Waclawek M, Schneider WJ, Wohlrab F. 2000. *J. Biol. Chem.* 275: 28866–72
86. Sawada H, Sakai N, Abe Y, Tanaka E, Takahashi Y, et al. 2002. *Proc. Natl. Acad. Sci. USA* 99:1223–28
87. Swanson WJ, Vacquier VD. 1997. *Proc. Natl. Acad. Sci. USA* 94:6724–29
88. Kresge N, Vacquier VD, Stout CD. 2000. *J. Mol. Biol.* 296:1225–34
89. Galindo BE, Moy GW, Swanson WJ, Vacquier VD. 2002. *Gene* 288:111–17
90. Wilkin MB, Becker MN, Mulvey D, Phan I, Chao A, et al. 2000. *Curr. Biol.* 10:559–67
91. Jazwinska A, Ribeiro C, Affolter M. 2003. *Nat. Cell Biol.* 5:895–901
92. Denholm B, Skaer H. 2003. *Curr. Biol.* 13:R909–11

93. DiBartolomeis SM, Akten B, Genova G, Roberts MA, Jackson FR. 2002. *Mol. Genet. Genomics* 267:564–76
94. Roch F, Alonso CR, Akam M. 2003. *J. Cell Sci.* 116:1199–207
95. Kernan M, Cowan D, Zuker C. 1994. *Neuron* 12:1195–206
96. Chung YD, Zhu J, Han Y, Kernan MJ. 2001. *Neuron* 29:415–28
97. Fujimoto D, Kanaya S. 1973. *Arch. Biochem. Biophys.* 157:1–6
98. Sebastiano M, Lassandro F, Bazzicalupo P. 1991. *Dev. Biol.* 146:519–30
99. Thompson EM, Kallesoe T, Spada F. 2001. *Dev. Biol.* 238:260–73
100. Spada F, Steen H, Troedsson C, Kallesoe T, Spriet E, et al. 2001. *J. Biol. Chem.* 276:20624–32
101. Kinloch RA, Ruiz-Seiler B, Wassarman PM. 1990. *Dev. Biol.* 142:414–21
102. Chamberlin ME, Dean J. 1990. *Proc. Natl. Acad. Sci. USA* 87:6014–18
103. Liang LF, Dean J. 1993. *Dev. Biol.* 156: 399–408
104. McLeskey SB, Dowds C, Carballada R, White RR, Saling PM. 1998. *Int. Rev. Cytol.* 177:57–113
105. Mold DE, Kim IF, Tsai CM, Lee D, Chang CY, Huang RC. 2001. *Mol. Reprod. Dev.* 58:4–14
106. Rosiere TK, Wassarman PM. 1992. *Dev. Biol.* 154:309–17
107. Jovine L, Qi H, Williams Z, Litscher ES, Wassarman PM. 2004. *Proc. Natl. Acad. Sci. USA* 101:5922–27
108. Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM. 2002. *Nat. Cell Biol.* 4:457–61
109. Fukuoka S. 2000. *Biochim. Biophys. Acta* 1491:376–80
110. Epifano O, Liang LF, Familari M, Moos MC Jr, Dean J. 1995. *Development* 121: 1947–56
111. Liang L, Soyal SM, Dean J. 1997. *Development* 124:4939–47
112. Huntriss J, Gosden R, Hinkins M, Oliver B, Miller D, et al. 2002. *Mol. Hum. Reprod.* 8:1087–95
113. Kanamori A, Naruse K, Mitani H, Shima A, Hori H. 2003. *Gene* 305:35–45
114. Shibuya K, Nagamine K, Okui M, Ohsawa Y, Asakawa S, et al. 2004. *Biochem. Biophys. Res. Commun.* 319:1181–89
115. Cavallone D, Malagolini N, Monti A, Wu XR, Serafini-Cessi F. 2004. *J. Biol. Chem.* 279:216–22
116. Roller RJ, Wassarman PM. 1983. *J. Biol. Chem.* 258:13243–49
117. Brivio MF, Bassi R, Cotelli F. 1991. *Mol. Reprod. Dev.* 28:85–93
118. Qi H, Williams Z, Wassarman PM. 2002. *Mol. Biol. Cell* 13:530–41
119. Williams Z, Wassarman PM. 2001. *Biochemistry* 40:929–37
120. Litscher ES, Qi H, Wassarman PM. 1999. *Biochemistry* 38:12280–87
121. Boja ES, Hoodbhoy T, Fales HM, Dean J. 2003. *J. Biol. Chem.* 278:34189–202
122. Yonezawa N, Nakano M. 2003. *Biochem. Biophys. Res. Commun.* 307:877–82
123. Sugiyama H, Murata K, Iuchi I, Nomura K, Yamagami K. 1999. *J. Biochem.* 125:469–75
124. Kubo H, Matsushita M, Kotani M, Kawasaki H, Saido TC, et al. 1999. *Dev. Genet.* 25:123–29
125. Sasanami T, Toriyama M, Mori M. 2003. *Biol. Reprod.* 68:1613–19
126. Sasaki K, Sato K, Akiyama Y, Yanagihara K, Oka M, Yamaguchi K. 2002. *Cancer Res.* 62:4894–98
127. Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG. 1994. *DNA Seq.* 4:361–93
128. Chen W, Kirkbride KC, How T, Nelson CD, Mo JY, et al. 2003. *Science* 301:1394–97
129. Conley BA, Koleva R, Smith JD, Kacer D, Zhang DW, et al. 2004. *J. Biol. Chem.* 279:27440–49
130. Sanz-Rodriguez F, Guerrero-Esteo M, Botella LM, Banville D, Vary CP, Bernabeu C. 2004. *J. Biol. Chem.* 279:32858–68
131. De Lisle RC, Ziemer D. 2000. *Eur. J. Cell Biol.* 79:892–904

132. Darie CC, Biniossek ML, Jovine L, Litscher ES, Wassarman PM. 2004. *Biochemistry* 43:7459–78
133. Cuff JA, Barton GJ. 2000. *Proteins* 40: 502–11
134. Hamodrakas SJ. 1992. *Results Probl. Cell Differ.* 19:115–86
135. Iconomidou VA, Chryssikos DG, Gionis V, Pavlidis MA, Paipetis A, Hamodrakas SJ. 2000. *J. Struct. Biol.* 132:112–22
136. Puett D, Holladay LA, Robinson JP. 1977. *Mol. Cell. Biochem.* 15:109–16
137. D'Auria S, Rossi M, Tanfani F, Bertoli E, Parise G, Bazzicalupo P. 1998. *Eur. J. Biochem.* 255:588–94
138. D'Auria S, Bazzicalupo P, Rossi M, Gryczynski I, Lakowicz JR. 2000. *J. Fluoresc.* 10:27–33
139. Betz SF. 1993. *Protein Sci.* 2:1551–58
140. Thornton JM. 1981. *J. Mol. Biol.* 151: 261–87
141. Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. 2000. *Biochemistry* 39:4207–16
142. Serafini-Cessi F, Malagolini N, Hoops TC, Rindler MJ. 1993. *Biochem. Biophys. Res. Commun.* 194:784–90
143. Malagolini N, Cavallone D, Serafini-Cessi F. 1997. *Kidney Int.* 52:1340–50
144. Sullivan DC, Huminiecki L, Moore JW, Boyle JJ, Poulsom R, et al. 2003. *J. Biol. Chem.* 278:47079–88
145. Iwamoto K, Ikeda K, Yonezawa N, Noguchi S, Kudo K, et al. 1999. *J. Reprod. Fertil.* 117:395–402
146. Fletcher AP, Neuberger A, Ratcliffe WA. 1970. *Biochem. J.* 120:425–32
147. Bleil JD, Wassarman PM. 1980. *Dev. Biol.* 76:185–202
148. Hogg PJ. 2003. *Trends Biochem. Sci.* 28: 210–14
149. Patra AK, Gahlay GK, Reddy BV, Gupta SK, Panda AK. 2000. *Eur. J. Biochem.* 267:7075–81
150. Chuang CC, Chen CY, Yang JM, Lyu PC, Hwang JK. 2003. *Proteins* 53:1–5
151. van Vlijmen HW, Gupta A, Narasimhan LS, Singh J. 2004. *J. Mol. Biol.* 335:1083–92
152. Vinayagam A, Pugalenthi G, Rajesh R, Sowdhamini R. 2004. *Nucleic Acids Res.* 32:D200–2
153. Greve JM, Wassarman PM. 1985. *J. Mol. Biol.* 181:253–64
154. Tong ZB, Nelson LM, Dean J. 1995. *J. Biol. Chem.* 270:849–53
155. Rankin T, Familari M, Lee E, Ginsberg A, Dwyer N, et al. 1996. *Development* 122:2903–10
156. Liu C, Litscher ES, Mortillo S, Sakai Y, Kinloch RA, et al. 1996. *Proc. Natl. Acad. Sci. USA* 93:5431–36
157. Wassarman PM, Qi H, Litscher ES. 1997. *Proc. R. Soc. London Ser. B* 264:323–28
158. Rankin TL, O'Brien M, Lee E, Wigglesworth K, Eppig J, Dean J. 2001. *Development* 128:1119–26
159. Rankin T, Talbot P, Lee E, Dean J. 1999. *Development* 126:3847–55
160. Wassarman PM, Mortillo S. 1991. *Int. Rev. Cytol.* 130:85–110
161. Sugiyama H, Iuchi I. 2000. *Rec. Res. Dev. Comp. Biochem. Physiol.* 1:139–61
162. Yurewicz EC, Sacco AG, Gupta SK, Xu N, Gage DA. 1998. *J. Biol. Chem.* 273:7488–94
163. Guraya SS. 1986. *Monogr. Dev. Biol.* 18: 1–223
164. Bonsignorio D, Perego L, Del Giacco L, Cotelli F. 1996. *Zygote* 4:101–8
165. Bjugn R, Flood PR. 1988. *Scand. J. Urol. Nephrol.* 22:313–15
166. Knipper M, Richardson G, Mack A, Muller M, Goodyear R, et al. 2001. *J. Biol. Chem.* 276:39046–52
167. Zhao M, Gold L, Ginsberg AM, Liang LF, Dean J. 2002. *Mol. Cell. Biol.* 22:3111–20
168. Tsubamoto H, Hasegawa A, Nakata Y, Naito S, Yamasaki N, Koyama K. 1999. *Biol. Reprod.* 61:1649–54
169. Wierzbicka-Patynowski I, Schwarzbauer JE. 2003. *J. Cell Sci.* 116:3269–76

170. Moradian-Oldak J. 2001. *Matrix Biol.* 20:293–305
171. Hulmes DJ. 2002. *J. Struct. Biol.* 137:2–10
172. Kielty CM, Baldock C, Lee D, Rock MJ, Ashworth JL, Shuttleworth CA. 2002. *Philos. Trans. R. Soc. London Ser. B* 357: 207–17
173. Hohenester E, Engel J. 2002. *Matrix Biol.* 21:115–28
174. Sasanami T, Pan J, Doi Y, Hisada M, Kohsaka T, Toriyama M. 2002. *Eur. J. Biochem.* 269:2223–31
175. Kiefer SM, Saling P. 2002. *Biol. Reprod.* 66:407–14
176. Harris JD, Seid CA, Fontenot GK, Liu HF. 1999. *Protein Expr. Purif.* 16:298–307
177. Zhao M, Gold L, Dorward H, Liang LF, Hoodbhoy T, et al. 2003. *Mol. Cell. Biol.* 23:8982–91
178. Handford PA, Downing AK, Reinhardt DP, Sakai LY. 2000. *Matrix Biol.* 19:457–70
179. Gamblin TC, Chen F, Zambrano A, Abraha A, Lagalwar S, et al. 2003. *Proc. Natl. Acad. Sci. USA* 100:10032–37
180. Mosesson MW, Siebenlist KR, Meh DA. 2001. *Ann. NY Acad. Sci.* 936:11–30
181. Taylor KM, Trimby AR, Campbell AK. 1997. *Immunology* 91:20–27
182. Bourne Y, Watson MH, Arvai AS, Bernstein SL, Reed SI, Tainer JA. 2000. *Struct. Fold. Des.* 8:841–50
183. Goodstadt L, Ponting CP. 2001. *Bioinformatics* 17:845–46

Annual Review of Biochemistry
Volume 74, 2005

CONTENTS