The Molecular Basis of Sex: Linking Yeast to Human

Willie J. Swanson,1 Jan E. Aagaard,1 Victor D. Vacquier,2 Magnus Monné,†3 Hamed Sadat Al Hosseini,3 and Luca Jovine*3

1Department of Genome Sciences, University of Washington
2Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego
3Department of Biosciences and Nutrition and Center for Biosciences, Karolinska Institutet, Huddinge, Stockholm, Sweden
†Present address: Department of Chemistry, University of Basilicata, Potenza, Italy

*Corresponding author: E-mail: luca.jovine@ki.se.
Associate editor: Michael Nachman

Abstract

Species-specific recognition between egg and sperm, a crucial event that marks the beginning of fertilization in multicellular organisms, mirrors the binding between haploid cells of opposite mating type in unicellular eukaryotes such as yeast. However, as implied by the lack of sequence similarity between sperm-binding regions of invertebrate and vertebrate egg coat proteins, these interactions are thought to rely on completely different molecular entities. Here, we argue that these recognition systems are, in fact, related: despite being separated by 0.6–1 billion years of evolution, functionally essential domains of a mollusc sperm receptor and a yeast mating protein adopt the same 3D fold as egg zona pellucida proteins mediating the binding between gametes in humans.

Key words: fertilization, egg–sperm interaction, egg coat, zona pellucida domain, yeast mating, protein structure.

Introduction

Like their counterparts in the vitelline (egg) envelope (VE) of other vertebrates as well as invertebrates such as the mollusc abalone (Aagaard et al. 2006), mammalian zona pellucida (ZP) subunits ZP1–4 assemble into the nascent egg coat using a common C-terminal “ZP domain” (Bork and Sander 1992; Jovine et al. 2002). This conserved polymerization module consists of two domains, ZP-N and ZP-C (Jovine et al. 2004; Wassarman and Litscher 2008) (fig. 1). Recent crystallographic studies of sperm receptor ZP3 revealed that the ZP-N domain defines a new subtype of the immunoglobulin (Ig) superfamily of proteins, characterized by two disulfide bonds with invariant 1–4, 2–3 connectivity, a unique E’ strand implicated in polymerization, and a conserved Tyr residue in strand F (Monné et al. 2008). Moreover, they showed that—despite having a very different sequence—ZP-C also adopts a β-sandwich fold with the same basic topology as ZP-N, suggesting that the two moieties of the ZP module might have originated by duplication of a single ancestral Ig-like domain (Han et al. 2010). Additional copies of ZP-N are found within the N-terminal region of some vertebrate ZP/VE components (Callebaut et al. 2007; Monné et al. 2008) (fig. 1), where—as in the case of mammalian ZP2—they can bind sperm (Tsubamoto et al. 1999) and regulate gamete recognition in the case of mammalian ZP2—they can bind sperm (Bleil et al. 1981; Gahlay et al. 2010). Notably, repeated sequences located within the N-terminal region of abalone VE subunits VERL and VEZP14 (fig. 1) are also thought to bind sperm (Swanson and Vacquier 1997; Aagaard et al. 2010), but because of very low-sequence similarity, no connection could be made between molluscan and mammalian repeats.

Molluscan Egg Coat Protein Repeats Adopt the ZP-N Fold of Mammalian ZP Proteins

Because rapid sequence divergence could mask potential relationships between reproductive proteins from evolutionary distant species (Swanson and Vacquier 2002), we performed a fold recognition analysis using sequence–structure comparison in FUGUE (Shi et al. 2001). Molluscan repeat sequences were threaded against a local copy of the HOMSTRAD database of structural profiles (de Bakker et al. 2001) that included an entry for the canonical ZP-N domain of VERL (Galindo et al. 2002), generated on the basis of the crystal structure of ZP3 ZP-N (Monné et al. 2008; Han et al. 2010). A high-confidence match was found between the sequence of VERL repeat 10 and the Ig-like fold variant specific to ZP-N (supplementary fig. S1, Supplementary Material online). An homology model of repeat 10 created on the basis of this sequence–structure alignment is structurally sound and exposes Asn side chains expected to be glycosylated (Swanson and Vacquier 1997) (fig. 2). Moreover, it can be readily extended to all other VERL repeats, as well as the VERL-like repeat of VEZP14 (Aagaard et al. 2010), because of significant sequence similarity (supplementary figs. S1 and S2a–b, Supplementary Material online). This suggests that all Cys within the repeat array of VERL are engaged in ZP-N-specific Cys1–4, Cys2–3 disulfide bonds, with the exception of C201 and C294 (supplementary fig. S3a, Supplementary Material online). These additional Cys, located in repeat 2, may therefore be responsible for forming the intermolecular disulfides that have been shown to mediate homodimerization of VERL (Swanson and Vacquier 1997). This prediction was experimentally confirmed by loss of VERL dimerization upon
introduction of C201D, C294S substitutions within a repeat 1–4 fragment secreted by insect cells (supplementary fig. S3b, Supplementary Material online). Considering that all other abalone VE subunits also contain a ZP module (Aagaard et al. 2010), this data collectively suggest that, as in mammals, the ZP-N domain accounts for the majority of the structure of the molluscan egg coat.

A Protein Domain Essential for Mating in Yeast Also Shares ZP-N-Specific Features

Domain analysis with SMART (Letunic et al. 2009) indicates that the N-terminal region of VEZP14, which contains the protein’s VERL-like ZP-N repeat (Aagaard et al. 2010) (fig. 1), is in turn related to yeast agglutinin-like proteins (supplementary fig. S1, Supplementary Material online). These highly glycosylated adhesion molecules mediate extracellular interactions, such as mating in Saccharomyces cerevisiae and host invasion in Candida albicans, mainly using the last of three N-terminal Ig domains (Ig III; fig. 1) (Dranginis et al. 2007). Although Ig III was initially modeled on the basis of Ig Kol—the best template available at the time—(de Nobel et al. 1996), FUGUE threading of Ig III sequences against the current protein fold database identifies the ZP-N Ig subtype as the top hit for this domain (supplementary fig. S1, Supplementary Material online), a prediction supported by I-TASSER (Roy et al. 2010). Most importantly, our ZP-N-based model of S. cerevisiae mating protein α-agglutinin/Sag1p Ig III is not only physically realistic (supplementary fig. S4, Supplementary Material online) but also completely consistent with a large amount of available biochemical data (fig. 3). Specifically, the model agrees with circular dichroism spectroscopy studies of the N-terminal half of α-agglutinin (Chen et al. 1995); accounts for the experimentally determined disulfide bond between C202 and C300 (Chen et al. 1995), which corresponds to the canonical Cys1–4 disulfide of the ZP-N fold; predicts burial of C227 and C256 (Cys 2,3)(Chen et al. 1995); and is consistent with exposure of residues that were shown experimentally to be accessible to proteases (Chen et al. 1995), glycosylated (Chen et al. 1995), or involved in binding to α-agglutinin (Cappellaro et al. 1991; de Nobel et al. 1996). Moreover, Y270 of α-agglutinin is positioned in correspondence of the conserved F-strand Tyr that lies next to invariant Cysα within the E’-F-G extension of the ZP-N fold (Monné et al. 2008; Han et al. 2010). Taken together, these considerations suggest that this particular type of Ig-like domain may not be restricted to multicellular eukaryotes as previously thought but also

![Fig. 1. Domain architecture of human ZP subunits, mollusc VERL and VEZP14, and yeast α-agglutinin/Sag1p. Pink: ZP-N domain; cyan: ZP-C domain; yellow: trefoil domain; violet: S/T-rich sequence repeat; dark blue: Sag1p Ig-like domains I, II; and dashed red box: SMART Pfam:Candida_ALS match in VEZP14.](https://academic.oup.com/mbe/article-abstract/28/7/1963/1048809)

![Fig. 2. Homology model of abalone VERL repeat 10 ZP-N domain, shown in side view using a cartoon representation with relevant residues depicted as sticks. The model is consistent with burial of hydrophobic residues (a; brown), exposure of positively charged, negatively charged, and polar side chains (b; blue, red, and cyan, respectively) and exposure of consensus sites for N-glycosylation (c; green).](https://academic.oup.com/mbe/article-abstract/28/7/1963/1048809)
exist in specialized extracellular proteins of yeast that play key roles in mating (\textit{S. cerevisiae} Sag1p) or adhesion to human tissues and biofilm formation (\textit{C. albicans} Als1p and Als3p).

**Conclusions and Functional Implications**

Although the rapid evolution of reproductive proteins makes the direct comparison of their sequences generally uninformative, relationships between these molecules could in principle be recognized by identifying suitably intermediate sequences that connect them (Park et al. 1997) or relying on conservation of common higher-order structural features. In this report, we have combined these approaches to detect unexpected structural similarities between reproductive proteins from both vertebrates and invertebrates, as well as yeast mating proteins. These findings suggest that some of the molecular features that regulate sexual interaction may be much more conserved during evolution than previously appreciated (fig. 1 and supplementary fig. S5, Supplementary Material Online). In this regard, it is particularly remarkable that \(\alpha\)-agglutinin amino acids essential for interaction with \(\alpha\)-agglutinin (de Nobel et al. 1996) (fig. 4c) are positioned so that they are exposed on the same face of the ZP-N fold as ZP2 and VERL residues implicated in sperm binding (fig. 4a–b). Moreover, specific adherence of \textit{Candida} to human endothelial and epithelial cells requires Als1p Ig III amino acids centered around V285 (Fu et al. 1998; Loza et al. 2004; Sheppard et al. 2004; Dranginis et al. 2007), a residue that is also predicted to be exposed on the same region of the ZP-N domain (fig. 4c). Because threading per se simply estimates the likeness that a known 3D fold is adopted by a given

**Fig. 3.** Stereograph of the model of \textit{Saccharomyces cerevisiae} \(\alpha\)-agglutinin/Sag1p Ig III. Conserved Cys residues: magenta; protease-accessible residues: red; N-glycosylated residues: light blue; O-glycosylated residues: violet; residues that are essential, important, or play a minor role in binding to \(\alpha\)-agglutinin: cyan, orange, and yellow, respectively; and Y270: dark gray.

**Fig. 4.** Mapping of functionally important residues on the homology models of human ZP2 repeat 1 ZP-N (a), mollusc VERL repeat 1 ZP-N (b), and yeast \(\alpha\)-agglutinin/Sag1p Ig III ZP-N (c). \textit{Saccharomyces cerevisiae} Sag1p Ig III residue V287, corresponding to functionally crucial residue V285 of \textit{Candida albicans} Als1p, is also indicated in (c). A top view of the proteins is shown, with N termini and Ig fold \(\beta\)-strands marked by uppercase letters.
sequence profile, it does not directly address the point of whether the corresponding proteins share common ancestry or just adopt a similar tertiary structure. Although this is currently unfeasible due to lack of abalone genome sequences and absence of significant conserved synheny between yeast and human, future identification of related sequences from additional lineages may make it possible to assess whether the similarity that we have uncovered reflects direct homology or is instead the result of convergent evolution. Nevertheless, considering the widespread distribution of the Ig fold, it is striking that reproductive protein sequences from mollusc and yeast specifically match its ZP-N variant, repeats of which had previously only been detected in vertebrate egg coat proteins.

**Supplementary Material**

Supplementary methods and figures S1–S5 are available at Molecular Biology and Evolution online (http://mbi.oxfordjournals.org/).

**Acknowledgments**

This work was supported by National Institute of Health (NIH) Grants HD057974, HD042563, and HD 054631 (W.J.S.); NIH Grant HD12986 (V.D.V.); the Center for Biosciences, Swedish Research Council grant 2009-5193, an EMBO Young Investigator award, and the European Research Council under the European Union’s Seventh Framework Program (FP7/2007–2013)/ERC grant agreement 260759 (L.J.). We thank Tsukasa Matsuda, Stevan Springer, and members of our laboratories for comments and discussions.

**References**


