

8-5-2017

# Novel species-specific glycoprotein on the surface of *Mytilus edulis* and *M. trossulus* eggs

Fiona M. Harper

*Rollins College, FHARPER@Rollins.edu*

Kasandra J. Riley

*kriley@rollins.edu, kriley@rollins.edu*

R. D. Rawson

*University of Maine*

Follow this and additional works at: [http://scholarship.rollins.edu/as\\_facpub](http://scholarship.rollins.edu/as_facpub)



Part of the [Biology Commons](#), and the [Cell and Developmental Biology Commons](#)

---

## Published In

F. M. Harper, K. J. Riley, and P. D. Rawson, "Novel species-specific glycoprotein on the surface of *Mytilus edulis* and *M. trossulus* eggs," *Invertebrate Reproduction & Development*, 2017, pages 1-7.

This Article is brought to you for free and open access by Rollins Scholarship Online. It has been accepted for inclusion in Faculty Publications by an authorized administrator of Rollins Scholarship Online. For more information, please contact [rwalton@rollins.edu](mailto:rwalton@rollins.edu).

# Novel species-specific glycoprotein on the surface of *Mytilus edulis* and *M. trossulus* eggs

F. M. Harper<sup>a,c</sup>, K. J. Riley<sup>b</sup> and P. D. Rawson<sup>c</sup>

<sup>a</sup>Department of Biology, Rollins College, Winter Park, Florida; <sup>b</sup>Department of Chemistry, Rollins College, Winter Park, Florida; <sup>c</sup>School of Marine Sciences, University of Maine, Orono, Maine

## ABSTRACT

Protein-protein interactions play a central role in the gamete attraction, binding, and fusion stages of gamete interactions and fertilization for broadcast spawning species, such as marine mussels in the *Mytilus edulis* species complex. Although assortative gamete interaction has been implicated in the level of reproductive isolation among the three species in this complex, the molecular basis of these interactions has not been elucidated. Using mass spectrometry peptide sequencing, cDNA sequencing, and bioinformatics approaches, we have investigated species-level variation in the proteins expressed on the surface of mussel eggs. We herein describe an extracellular protein, MESP-1, from the surface of the eggs of *M. edulis* and *M. trossulus* that has a unique domain structure when compared to protein structures that have heretofore been identified. Given variation in the size of MESP-1 predicted from cDNA sequences versus those estimated from SDS-PAGE gels, we conclude this protein is subject to significant species-specific post-translation modifications. Further, bioinformatic analysis of the novel structure of MESP-1 suggests that this protein may be an integral membrane protein involved in sperm-egg fusion, and/or released to the vitelline envelope.

Keywords (3 – 6 words): sperm-egg fusion, mussel, vitelline envelope, fertilization

## **Introduction**

Proteins direct sperm-egg interactions during fertilization and successful reproduction in many broadcast spawning invertebrates. Molecules expressed on the surface of gametes are involved in each step, from the attraction of sperm by egg compounds through to the production of the zygote (Lessios 2011). The large gastropod abalone (genus *Haliotis*) is one of few invertebrates where both sperm and egg fertilization proteins, and their interactions, have been well described. Studies on this genus have shown the plasma membrane of molluscan eggs is surrounded by a raised vitelline envelope (VE) of glycoprotein fibers which acts as a species-specific barrier to sperm entry. Prior to fertilization in abalone, the sperm releases the protein lysin, which binds to a glycoprotein on the egg VE called VERL (VE receptor for lysin; Galindo et al. 2002). Binding of lysin to VERL causes the glycoprotein fibers of the VE to separate, creating a hole for the sperm to enter (Vacquier et al. 1999; Kresge et al. 2001). While a few rapidly evolving gamete proteins identified as major loci controlling species-specific sperm-egg binding have been studied in other marine invertebrates (see reviews by Swanson and Vacquier 2002a; 2002b; Palumbi 2009), the proteins involved in regulating egg and sperm fusion have received much less attention. Most of what is known about proteins involved in gamete fusion comes from studies of mammalian gametes (e.g. Lee and Sampson 2009; Ikawa et al. 2012).

Our research investigates gamete interactions among blue mussels in the *Mytilus edulis* species complex (*M. edulis*, *M. trossulus*, *M. galloprovincialis*). In contrast to abalone, extensive hybrid zones occur whenever two of the three species co-occur. In the northwest Atlantic, *M. edulis* and *M. trossulus* hybrids comprise  $\approx$  15% of sympatric populations, indicating strong but incomplete

gamete compatibility (Rawson et al. 2003; Slaughter et al. 2008). In contrast, gene flow is very high in Baltic Sea populations between these same two species, and no pure *M. trossulus* remain (Riginos and Cunningham 2005). Assortative gamete interaction has been identified as playing a role in reproductive isolation in these species (Rawson et al. 2003), but its molecular basis is not well understood. In the current study, we isolated and sequenced a novel protein from eggs of *M. edulis*, here referred to as Mytilus Egg Surface Protein 1 (med-MESP-1). The same protein was also isolated from eggs of *M. trossulus* (mtr-MESP-1). Although sequence length is conserved among the two homologs, they have distinct electrophoretic mobilities, suggesting differential post-translational modification. The identification of MESP-1 represents one of relatively few egg surface proteins described in mussels to date and its varied gel mobility can be used to differentiate *M. edulis*, *M. trossulus*, and interspecies hybrids, which we demonstrate can express both homologs of MESP-1.

## **Methods**

### ***Isolation and sequence analysis of egg vitelline envelope proteins***

Sexually mature blue mussels were collected from a population in the low intertidal zone near Lubec, Maine, USA (latitude 44°81'78"N; longitude 66°98'09"W) in 2004 and Lubec and Whiting Bay, ME (latitude 44°82'57"N; longitude 67°15'26"W) in 2005 where appreciable frequencies of *Mytilus edulis*, *M. trossulus*, and their hybrids have been observed (e.g. Rawson et al. 2001). DNA was isolated from mantle tissue (Qiagen DNA Mini Kit), following the manufacturers' protocol and used as the template in seven nuclear DNA, PCR-based markers diagnostic for *M. edulis* and *M. trossulus* (Table S1). Individual females were identified as either

*M. edulis* or *M. trossulus* if they were homozygous for *M. edulis*- or *M. trossulus*-specific alleles at all seven nuclear DNA markers.

The spawning and collection of eggs followed the methods outlined in Rawson et al. (2003). To prepare vitelline envelope proteins, a 500 µl suspension of unfertilized eggs (~1000 eggs) from each female was suspended in 100 µl of cold, 1 µm filtered sea water, and treated with 20 ml of egg dissolving medium (EDM; 0.5 M NaCl, 10 µM EDTA, 20mM Tris pH 8.0, 1% Triton-X100; Vacquier, pers. comm.) to remove cytoplasmic material. The egg-EDM mixture was centrifuged at 1000 rpm for 3 min at 4°C. The supernatant was removed, leaving a clear pellet of cell membranes and vitelline envelopes. An additional 1.25 ml of EDM and 1.25 ml of cold, 1 µm filtered sea water was added, the pellet gently resuspended, and the VE suspensions centrifuged a second time. The resulting pellets were resuspended and homogenized in cold protein homogenization buffer (50mM Tris-Cl pH6.8, 4% SDS, 1mM EDTA with Roche Complete Protease Inhibitors at 100 µl buffer per 1 mg of isolated vitelline envelopes). The homogenized VE suspensions were heated to 100°C for 5 min, centrifuged at 14,000 rpm for 15 min at 20°C, the supernatant containing VE proteins collected, and the concentration of protein estimated using a Coomassie Plus Protein Analysis Kit (Pierce).

VE proteins were analysed by polyacrylamide gel electrophoresis by combining 2-5 µl of homogenized VE protein (~50 mg total protein) with 1.25 µl of 10X sample reducing agent and 3 µl of 4X LDS loading dye in a total volume of 12.5 µl. Each mixture was heated to 95°C for 5 min and loaded onto a 4-12% Bis-Tris NuPage gel (Invitrogen). Electrophoresis was performed in 1X MOPS-SDS running buffer; each gel included BSA and Sypro Candy Cane (Molecular Probes) protein standards for comparison. After electrophoresis, gels were stained to detect glycoproteins (Pro-Q Emerald Green 300 Glycoprotein Gel Stain Kit; Molecular Probes) and

total proteins (Brilliant Blue Coomassie R-250; ThermoScientific).

Two species-specific protein bands, resolved on both Coomassie R-250 and Emerald Green 300 stained gels, were selected for study for two reasons: the significant difference in electrophoretic mobility of these high molecular weight proteins and because the hybrid individual expressed both. These two bands were separately excised from a Coomassie R-250 stained gel, washed with water, dehydrated in acetonitrile, washed in 0.1 M ammonium bicarbonate, dehydrated a second time in acetonitrile and then dried. The protein in each band was digested overnight with 10-15  $\mu$ l 20 ng/ $\mu$ l trypsin in 50 mM ammonium bicarbonate. The resulting peptides were extracted and analyzed using standard protocols on a Finnigan LCQ-Deca ion mass spectrometer system at the Lerner Research Institute's Mass Spectrometry Core Facility (<http://www.lerner.ccf.org>).

#### ***Amplification and analysis of med-MESP-1 and mtr-MESP-1 cDNAs***

Degenerate oligonucleotides (URP-1 and URP-2; Table S2) for use in 3' RACE-PCR experiments were designed from the peptides obtained for med-MESP-1. Total RNA was isolated from the mature gonad of a single adult *M. edulis* female mussel using an RNeasy Kit (Qiagen). First strand cDNA was synthesized with a BD SMART cDNA kit (Clontech) using only the 3' BD SMART CDS primer IIA. A two-step PCR amplification was performed in 25  $\mu$ l reactions containing 1  $\mu$ l of first strand cDNA, 1X PCR Buffer (Invitrogen), 1.5 mM MgCl<sub>2</sub> (Invitrogen), 0.2 mM of each dNTP (Invitrogen), 0.6 mM URP-1, and 1 unit Taq polymerase (Invitrogen). Thermocycling conditions consisted of a 94°C soak for 2 min, followed by 1 cycle of 94°C for 30 s, annealing at 55°C for 20 s, and extension at 72°C for 4 min and nine additional cycles in which annealing temperature was reduced by 1°C every cycle. The PCR product was

diluted 10-fold and 1  $\mu$ l of the dilution used as template in a nested amplification with a total volume of 46  $\mu$ l containing 1X PCR Buffer, 1.6 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 0.65  $\mu$ M URP-2, 0.65  $\mu$ M adaptor primer II A (Clontech), and one unit Taq polymerase. Thermocycling conditions included a 94°C soak for 2 min, and 25 cycles of 94°C for 30 s, 55°C for 20 s, and 72°C for 4 min. A single band ~1,200 bp in size was obtained and cloned using the TOPO TA cloning kit (Invitrogen). Cloned inserts were PCR amplified using vector primers and sequenced in both directions using the BigDye Terminator v3.1 Cycle Sequencing Kit on an ABI 377 DNA sequencer (Applied Biosystems).

For isolation of the 5' end of the gene for med-MESP-1 by 5' RACE (Rapid Amplification of cDNA ends), we constructed full-length double stranded cDNA from total RNA using the BD SMART cDNA kit (Clontech). Two new gene-specific oligonucleotides (URP-1R and URP-2R; Table S2) were used in nested amplifications with the 3' BD SMART CDS IIA and BD SMART IIA oligonucleotides. The two-step protocol for 5' RACE followed the same procedure as described above, except the first reaction cocktail used 2  $\mu$ l of 5'RACE-cDNA, 0.24  $\mu$ M URP-1R, and 0.24  $\mu$ M adaptor primer II A (Clontech). In the second amplification step, the reaction was diluted 100-fold with H<sub>2</sub>O and 1  $\mu$ l of the diluted product was used as template in a nested PCR amplification using 0.26  $\mu$ M URP-1R. The nested 5' RACE PCR produced a single band ~1,800 bp in size which was cloned and sequenced as above, with additional sequences generated using gene-specific primers URP-2R, URP-3R, and URP-4R (Table S2). The resulting chromatograms were proofread and contiguous sequences assembled using the program Vector NTI (Invitrogen).

The primers RPF<sub>or</sub> and RPRE<sub>v</sub> (Table S2), were designed to amplify full length med-MESP-1

and mtr-MESP-1 genes. Total RNA was isolated from *M. trossulus* eggs and double stranded cDNA synthesized using the protocols described above. Full length mtr-MESP-1 was amplified from *M. trossulus* ~100 pg ds-cDNA in 45 µl PCR reactions containing 1X PCR Buffer, 1.7 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 0.67 µM RPF<sub>or</sub> and RPRE<sub>v</sub>, and 1 unit Taq polymerase.

Thermocycling conditions, cloning and sequencing methods were as described above.

The nucleotide and deduced amino acid sequences for MESP-1 from *M. edulis* and *M. trossulus* were aligned using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). The sequences were compared to nucleotide and protein sequences in the NCBI database (GenBank) using the Blast and Blastp algorithms. Domain predictions for med-MESP-1 were made using the NCBI Conserved Domain search (Marchler-Bauer et al. 2009; 2011). Predictions of transmembrane helices in the proteins were performed using the TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/>), and SignalP 4.1 was used to predict the signal peptide (Petersen et al. 2011).

### ***MESP-1 variation in allopatric populations***

Variation in med-MESP-1 and mtr-MESP-1 among allopatric *M. edulis* and *M. trossulus* populations was assessed by sequencing and restriction fragment length polymorphism (RFLP) analyses targeting a cluster of non-synonymous substitutions toward the 5' end of MESP-1. A 138 bp fragment spanning bases 119 to 257 of the med-MESP-1 mRNA was amplified from genomic DNA isolated from individual mussels. PCR amplifications (25 µl reactions) containing 1 µl of template DNA and MESP-1 specific primers RPS<sub>hort</sub>For and RPS<sub>hort</sub>Rev (Table S2) were conducted at an annealing temperature of 55°C to generate PCR products for 12 *M. edulis* from Lewes, DE and 11 *M. trossulus* from Newport, OR. Gel purified PCR products were cloned into pGEM-T vector (Promega) and 2-6 clones from each cloning reaction were sequenced using

primers M13For and M13Rev. We aligned the sequences using Clustal Omega (Li et al. 2015) and used MEGA (v6; Tamura et al. 2013) to estimate sequence divergence and construct an unrooted neighbor-joining phylogeny based on the Kimura 2-parameter model. We analysed an additional 60 mussels using a restriction fragment length polymorphism approach utilizing a *MboII* restriction site common to *M. trossulus* sequences, but missing in all *M. edulis* sequences. PCR products (5 µl) for the 138 bp fragment of MESP-1 from 8 *M. edulis* from Lewes, DE and 20 *M. edulis* from Narragansett, RI and 9 and 23 *M. trossulus* from Newport, OR and Little Tutka Bay, AK, respectively, were digested with 10 units of MboII enzyme (NEB) and fragment size variation assessed using agarose gel electrophoresis (3% agarose gel with Tris-Acetate-EDTA buffer).

## Results

### *Analysis of vitelline envelope proteins*

We identified 13 *M. edulis*, 5 *M. trossulus* and two hybrid female mussels among our sample of mussels from a hybrid zone population in eastern Maine. We consistently resolved 10-13 distinct protein bands from vitelline envelope preparations from the eggs of these females using Coomassie blue R250 and Pro-Q Emerald 300 staining (Fig. 1). The former stain typically detected proteins ranging from 60->100 kDa in size (Fig. 1A). The latter stain detects proteins that are glycosylated, is 333 times more sensitive than Coomassie blue staining, and detected bands of lower-abundance glycosylated proteins ranging from 40->200 kDa in size in both species (Fig. 1B). For this study, we targeted two species-specific protein bands of similar weight for mass-spectral analysis and gene sequencing. One protein band (65 kDa; med-MESP-

1) was observed in egg-VE preparations from all 13 *M. edulis* females. The other protein band (70 kDa; mtr-MESP-1) was observed in VE preparations from all 5 *M. trossulus* females. Eggs of hybrid females expressed both med-MESP-1 and mtr-MESP-1 protein bands (Fig. 1A).

Mass spectral analysis of med-MESP-1 resolved 14 distinct peptide fragments and 17 peptide fragments for mtr-MESP-1 (Table S3). Although none of the peptide fragments from either med-MESP-1 or mtr-MESP-1 matched protein sequences in GenBank, six of the peptide fragments in med-MESP-1 had an unambiguous match in mtr-MESP-1. The degenerate oligonucleotide primers URP-1 and URP-2 were designed from the largest peptide fragment and facilitated the amplification and sequencing of full-length coding sequence for the MESP-1 gene in both species. The cDNA sequences for MESP-1 from both *M. edulis* and *M. trossulus* indicate that the gene contains an open reading frame of 2,166 nt with a predicted amino acid sequence of 722 residues (GenBank Acc. Nos. KY274455 and KY274456). We also obtained 67 and 370 bases of sequence from the 5' and 3' untranslated regions, respectively, of med-MESP-1. We unambiguously aligned each of the peptide fragments from the mass-spectral analysis of both med-MESP-1 and mtr-MESP-1 to their respective gene sequence.

We found amino acid substitutions at 54 of the 722 (7.5%) residues in a comparison of the deduced amino acid sequences for *M. edulis* and *M. trossulus* MESP-1 between the two species (Fig. 2). The substitutions were mostly conservative (67%) or semi-conservative (19%) and scattered throughout the protein. The predicted molecular weight of unmodified MESP-1 is approximately 79 kDa for both species, larger than those estimated by SDS-PAGE analysis (Fig. 1), suggesting additional species-specific post-translational processing of MESP-1 that differentially affects size variation in the final mature proteins for each species. Given that residues differing between the two species include several that could be subject to glycosylation

(e.g. asparagine, threonine, etc.), the electrophoretic mobility difference could be due to unique glycosylation.

We conducted an NCBI Conserved Domain Search for the MESP-1 primary protein sequence; the resulting domain map (Fig. 2) contains three c-2 type IgC2 domains, an immunoglobulin-like domain (Ig-like), and a C-terminal domain with homology to the extracellular ligand-binding domain (LBD) for an acetylcholine receptor. An NCBI protein-protein BLAST (Blastp) search using the 722 amino acid sequence of med-MESP-1 indicated the highest identity (32%) with 509 amino acids associated with the Ig and Ig-like domains within the predicted hemicentin-1-like protein from the Pacific oyster *Crassostrea gigas*. Med-MESP-1 also has high identity (31%) with the titin protein from *C. gigas*, for the same Ig and Ig-like domains. BLAST searches of the last 212 amino acids of med-MESP-1 had the highest identity similarity (32%) to the predicted neuronal acetylcholine receptor subunit alpha-2-like protein from the oyster *C. gigas*.

The N-terminus of MESP-1 bears two key features. Bioinformatic prediction of transmembrane helices predicted a single-pass transmembrane domain, with a helix spanning AA 12-31, and a large extracellular domain (AA 32-722; Fig. 2). However, there is a predicted cleavage site between residues 26-27 that would remove an N-terminal signalling peptide, and no peptides corresponding to the first 26 amino acids were detected by mass spectrometry (Table S3). Our protein isolation procedure may include both VE and egg membrane proteins. Thus, MESP-1 may be anchored in the egg membrane and/or be released into the VE.

Analysis of sequence variation for a 138 bp portion of the coding region of MESP-1 identified 6 unique alleles and 5 unique alleles among mussels from allopatric *M. edulis* and *M.*

*trossulus* populations, respectively (GenBank Acc. Nos. MF467726 to MF467736). The average pairwise differences among sequences obtained from *M. edulis* ( $0.019 \pm 0.008$ ) and *M. trossulus* ( $0.017 \pm 0.010$ ) individuals were considerably less than the average sequence divergence between species ( $0.114 \pm 0.037$ ) as reflected in the neighbour-joining phylogeny in which sequences from Lewes, DE (*M. edulis*) and Newport, OR (*M. trossulus*) form two well-supported clades (Fig. 3). Digestion of the 138 bp product with *MboII* generated fragments of 97 and 41 bp for all *M. trossulus* sampled from Oregon (n=20) and Alaska (n=23) while all *M. edulis* individuals from Delaware (n=20) and Rhode Island (n=20) tested negative for the presence of the restriction site.

## **Discussion**

We identified and sequenced a novel cell surface protein, MESP-1, from the eggs of two closely related marine mussels. This protein is expressed in eggs from *Mytilus edulis*, *M. trossulus* and interspecies hybrids. Comparison of cDNA sequences from the two species indicates a large number of non-synonymous substitutions; although restricted to a small portion of the 5' end of MESP-1 gene, analysis of genomic sequences suggests that these differences are fixed in the two species. Despite the extensive amino acid divergence in MESP-1 from the two species and extensive differences in electrophoretic mobility on SDS-PAGE gels, the amino acid sequences predicted from cDNA sequencing contained the same total number of amino acids (722). Our analyses of MESP-1 predict that it has a short (11 amino acid) intracellular domain at the N-terminus, with a single-pass transmembrane helix, followed by a longer extracellular domain containing Ig and Ig-like domains. At the C-terminus, MESP-1 is homologous to the extracellular ligand-binding domain of neurotransmitter-gated ion channels. There was

divergence in amino acid sequences of the two species of 7.5%. Though their unmodified, unprocessed predicted molecular weights and amino acid lengths were virtually the same, differential mobility of the species-specific homologs in PAGE gel electrophoresis suggests differences in post-translational modification, likely glycosylation.

Blue mussels are a non-model species for which genomic resources are currently limited.

Although our understanding of the *Mytilus* genome has improved recently with publication of the first whole-genome sequencing study (Murgarella et al. 2016), it was a low-coverage study and the draft genome contains a large number of unknown repetitive sequences. In addition, we have not conducted a functional analysis of the MESP-1 protein. The predicted protein structure of MESP-1, however, provides some insight into its potential function. Domain predictions identified five Ig domains: three C-2 type immunoglobulin domains (IgC2), an immunoglobulin-like domain (Ig-like), and an undefined region in the Ig superfamily.

MESP-1 may be an integral membrane protein within the egg plasma membrane and involved in sperm-egg fusion. Immunoglobulin superfamily proteins (IgSF) are abundant on the surface of both eggs and sperm, and play key roles in fusion of sperm and egg plasma membranes. In mammalian systems, the IgSF protein *Izumo*, expressed on the surface of acrosome-reacted sperm is essential for fusion: male mice in which the *Izumo* gene has been knocked out are infertile due to an inability to fuse of their sperm to fuse with the egg membrane (Inoue et al. 2005). Similarly, CD9, a member of a family of tetraspan-membrane proteins expressed on the mammalian egg surface appears essential for ability of the egg to fuse with the sperm (see review by Jankovicova et al. 2015).

Ion gated channels are integral membrane proteins, typically composed of a large extracellular

glycosylated N-terminal ligand-binding domain, followed by three hydrophobic transmembrane regions which form the ionic channel, followed by an intracellular region of variable length and a fourth hydrophobic region at the C-terminus (Purves et al. 2001). MESP-1 has a truncated ion-channel ligand binding domain and does not contain the consensus sequence for the ion channel pore. Blast searches using the last 212 amino acids of MESP-1 isolated from *M. edulis* eggs indicate this region has the highest similarity (32%) to the predicted neuronal acetylcholine receptor subunit alpha-2-like protein from the oyster *C. gigas* (Zhang et al. 2012). Acetylcholine receptors on the egg surface of sea urchins, upon excitation by specific ligands, can regulate Na<sup>+</sup> entry into the egg, and affect membrane polarization, a crucial step in preparing the egg membrane for fusion with the sperm (Angelini et al. 2004).

MESP-1 isolated from *Mytilus edulis* eggs shares the highest identity with the predicted hemicentin-1-like protein from another bivalve mollusc, the Pacific oyster *Crassostrea gigas*, particularly among the 509 amino acids associated with the immunoglobulin (Ig) and immunoglobulin-like domains of MESP-1. Hemicentin is a large molecule, containing approximately 50 Ig domains, found on the cell surface of many cell types. Studies in vertebrate and invertebrate models suggest hemicentin functions include cell-cell adhesion (see Xu et al. 2013). The IgC2 domain is found in a diverse range of protein families including glycoproteins, growth factor receptors, and cell adhesion molecules (Aricescu and Jones 2007). The presence of Ig domains suggests a role for MESP-1 in cell adhesion.

MESP-1 may not be functional in the egg plasma membrane but instead involved in sperm-egg binding as a VE protein, or incorporated into the VE following secretion from the egg membrane. In the abalone, more than 30 VE proteins have a polymerization module known as the zona pellucida (ZP) domain, commonly found in invertebrate and vertebrate egg coats

(Jovine et al. 2005). Like the VE of molluscs, the extracellular matrix surrounding the egg in mammals is referred as the ZP and is similarly composed of glycoproteins. Biochemical assays by Aagaard et al. (2013) found that the ZP-N domain was sufficient for binding lysin. MESP-1 has none of the ZP domains associated with VERL and does not appear to be a gamete recognition protein. Interestingly however, one scenario postulated by Claw and Swanson (2012) to explain differences in binding of rat sperm to mouse and human ZP (Hoodbhoy et al. 2005), was that rat sperm may recognize residues other than ZP domains. Specifically, they proposed that binding may be “mediated by species-specific posttranslational modification (e.g. glycosylation) patterns, regardless of the amino acid sequences”. Species-specific differences in the posttranslational modifications noted in MESP-1 may play an important role in sperm-egg interactions in mussels. Based on these patterns, we conclude MESP-1 is a novel species-specific glycoprotein which may be anchored in the egg plasma membrane and/or be released into the vitelline envelope, whose function requires further study.

### **Acknowledgments**

We would like to thank Beth Ann Pomerlau whose assistance was invaluable in the completion of this research and Dr. Victor Vacquier for sharing his expertise and protocols for isolating egg coat proteins. This research was partially funded by support to PDR from the National Science Foundation (DEB0315891) and Hatch Funds provided by the Maine Agricultural and Forest Experiment Station (Project ME08510).

### **Disclosure statement**

The authors declare they have no known or potential conflicts of interest.

## References

- Aagaard JE, Springer SA, Soelberg SD, Swanson WJ. 2013. Duplicate abalone egg coat proteins bind sperm lysin similarly, but evolve oppositely, consistent with molecular mimicry at fertilization. *PLoS Genetics*. e1003287.
- Angelini C, Baccetti B, Piomboni P, Trombino S, Aluigi MG, Stringara S, Gallus L, Falugi C. 2004. Acetylcholine synthesis and possible functions during sea urchin development. *European Journal of Histochemistry*. 49:235–244.
- Aricescu AR, Jones EY. 2007. Immunoglobulin superfamily cell adhesion molecules: zippers and signals. *Current Opinion in Cell Biology*. 19:543–550.
- Claw KG, Swanson WJ 2012. Evolution of the egg: new findings and challenges. *Annual Review of Genomics and Human Genetics*. 13:109–25.
- Galindo BE, Moy GW, Swanson WJ, Vacquier VD. 2002. Full-length sequence of VERL, the egg vitelline envelope receptor for abalone sperm lysin. *Gene*. 288:111–117.
- Heath DD, Rawson PD, Hilbish TJ. 1995. PCR-based nuclear markers identify alien blue mussel *Mytilus* spp. genotypes on the west coast of Canada. *Canadian Journal of Aquatic Science*. 52:2621-2625.
- Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J. 2005. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. *Journal of Biological Chemistry*. 280:12721–12731.
- Ikawa M, Inoue N, Benham AM, Okabe M. 2012. Fertilization: a sperm's journey to and interaction with the oocyte. *Journal of Clinical Investigation*. 120:1–11.
- Inoue N, Ikawa M, Isotani A, Okabe M. 2005. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature*. 434:234–238.
- Jankovicova J, Simon M, Antalikova J, Cupperovai P, Michalkova K. 2015. Role of tetraspanin CD9 molecule in fertilization of mammals. *Physiological Research*. 64:279–293.
- Jovine L, Darie CC, Litscher ES, Wassarman PM. 2005. Zona pellucida domain proteins. *Annual Review of Biochemistry*. 74:83–114.
- Kresge N, Vacquier VD, Stout CD. 2001. Abalone lysin: The dissolving and evolving sperm protein. *BioEssays*. 23:95–103.
- Lee Y, Sampson NS. 2009. Polymeric ADAM protein mimics interrogate mammalian sperm-egg binding. *ChemBioChem*. 10:929–937.

- Lessios HA. 2011. Speciation genes in free-spawning marine invertebrates. *Integrative and Comparative Biology*. 51:456–465.
- Li W, Cowley A, Uludag M, et al. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*. 43:W580–W584.
- Marchler-Bauer A, Anderson JB, Chitsaz F, et al. 2009. CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Research*. 37:205–210.
- Marchler-Bauer A, Lu S, Anderson JB, et al. 2011. CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research*. 39:225–229.
- Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C. 2016. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. *PLoS One*. 11:e0151561.
- Palumbi SR. 2009. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity*. 102:66–76.
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. 8:785–786.
- Purves D, Augustine GJ, Fitzpatrick D. 2001. *Neuroscience*, 2nd edn. Sinauer Associates, Sunderland MA.
- Rawson PD, Hayhurst S, Vanscoyoc B. 2001. Species composition of blue mussel populations in the northeastern Gulf of Maine. *Journal of Shellfish Research*. 20:31–38.
- Rawson PD, Secor C, Hilbish TJ. 1996. Effects of natural hybridization on the regulation of doubly uniparental mtDNA inheritance in blue mussels *Mytilus* spp. *Genetics*. 144:241–248.
- Rawson PD, Slaughter C, Yund PO. 2003. Patterns of gamete incompatibility between the blue mussels *Mytilus edulis* and *M. trossulus*. *Marine Biology*. 143:317–325.
- Riginos C, Cunningham CW. 2005. Local adaptation and species segregation in two mussel *Mytilus edulis* x *Mytilus trossulus* hybrid zones. *Marine Ecology*. 14:381–400.
- Slaughter C, McCartney MA, Yund PO. 2008. Comparison of gamete compatibility between two blue mussel species in sympatry and in allopatry. *Biological Bulletin*. 214:57–66.
- Swanson WJ, Vacquier VD. 2002a. Reproductive protein evolution. *Annual Review of Ecological Systematics*. 33:161–179.

- Swanson WJ, Vacquier VD. 2002b. The rapid evolution of reproductive proteins. *Genetics*. 3:137–144.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*. 30:2725–2729.
- Vacquier VD, Swanson WJ, Metz EC, Stout CD. 1999. Acrosomal proteins of abalone spermatozoa. *Advances in Developmental Biochemistry*. 5:49–81.
- Xu X, Xu M, Zhou X, Jones OB, Moharomd E, Pan Y, Yan G, Anthony DD, Isaacs WB. 2013. Specific structure and unique function define the hemicentin. *Cell & Bioscience*. 3:27.
- Zhang G, Fang X, Guo X, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*. 490:49–54.

## Figure Legends

**Figure 1.** Coomassie blue staining for total (A) and emerald green staining for glycosylated (B) proteins isolated from the egg vitelline envelopes of *M. edulis* (E) and *M. trossulus* (T) and hybrid (H) mussels. Approximate molecular weights (kDa) are given on the left side of each image. Bands marked with an asterisk were cut from the SDS-PAGE Coomassie gel for mass spectrometry analysis.

**Figure 2.** Comparison of the predicted amino acid sequence (total 722 residues) for MESP-1 from *M. edulis* (E) and *M. trossulus* (T) is shown in panel A. Residues at which amino acid substitutions were observed are highlighted by an asterisk. Domain predictions for med-MESP (panel B) identified five immunoglobulin-like domains (three C-2 type immunoglobulin domains (IgC2), an immunoglobulin-like domain (Ig-like), an undefined region in the Ig superfamily (IgSF)), and a C-terminal domain with homology to the extracellular ligand-binding domain (LBD) of neurotransmitter-gated ion channels in MESP-1. The location of a predicted transmembrane domain based on analysis through the TMHMM Server v. 2.0 is given in panel C.

**Figure 3.** Phylogenetic relationships among unique 138 base pair MESP alleles amplified from allopatric *M. edulis* and *M. trossulus* mussels. The sequences from an allopatric population of *M. edulis* (Lewes, DE, ME-A to ME-F) form a separate clade from those obtained from an allopatric population of *M. trossulus* (Newport, OR; MT-A to MT-E). Bootstrap support for the two clades, based on 500 resamplings, is indicated at the node for the *M. trossulus* clade.