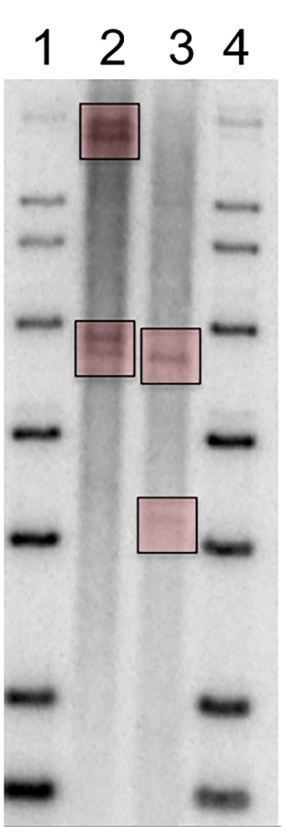
Supplementary Material

*Note: The following supplementary figures and tables are in order of their reference in the main body of the manuscript.*

Supplementary Table 1. Comparisons of amino acid compositions of shell matrix fractions and folian-cv1 by the Cornish-Bowden difference index1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | SM | 48 kDa band | 55 kDa band | folian-cv1 |
| IM | (related) | (weakly related) | (weakly related) | (unrelated) |
|  | 38.5 | 68.8 | 46.8 | 577.7 |
| SM |  | (related) | (related) | (unrelated) |
|  |  | 38.7 | 13.3 | 368.9 |
| 48 kDa |  |  | (related) | (unrelated) |
|  |  |  | 15.4 | 565.3 |
| 55 kDa |  |  |  | (unrelated) |
|  |  |  |  | 434.6 |

1Numerical values (SΔN) were determined by the equation SΔN= 1/2 ∑ (NiA-NiB)2, where NiA and NiB are the quantity (in mole percent for these calculations) of the ith type of amino acid contained in protein A or B. If SΔN< 0.42 N (N= 100 when using mole percent), then there is a 95% certainty that the proteins are related. If 0.42 N< SΔN < 0.93 N, then the proteins are weakly related. If SΔN > 0.93 N then there is a 95% certainty that the proteins are unrelated. The index is applied ideally to proteins having approximately the same number of amino acids. It is applied here as a reproducible measure of relatedness in order to reduce bias from less quantified comparisons.



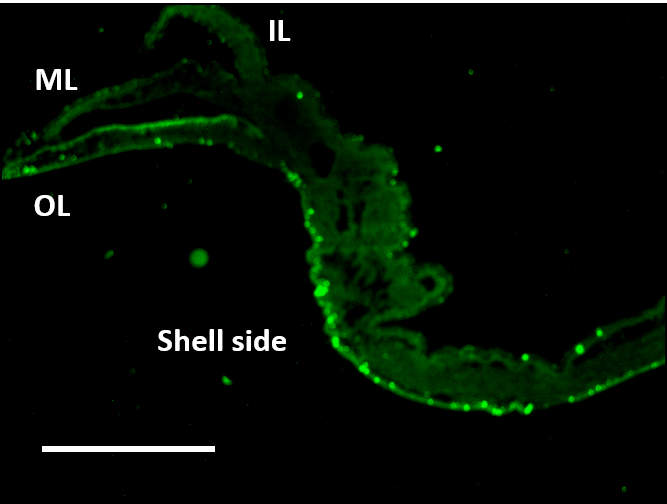
**Supplementary Figure 1.** A genomic Southern blot of *Crassostrea virginica* genomic DNA digested with two complimentary restriction enzymes and probed with the folian-cv1 gene sequence. **Lanes 1 and 4**- Lambda high range markers. **Lane 2**- BamHI digested *C. virginica* genomic DNA. **Lane 3**- HindIII digested *C. virginica* genomic DNA. The boxed regions highlight two distinct regions that are positive for the folian-cv1 probe following digestion with either restriction enzyme and indicate that there are two copies of the folian-cv1 gene.



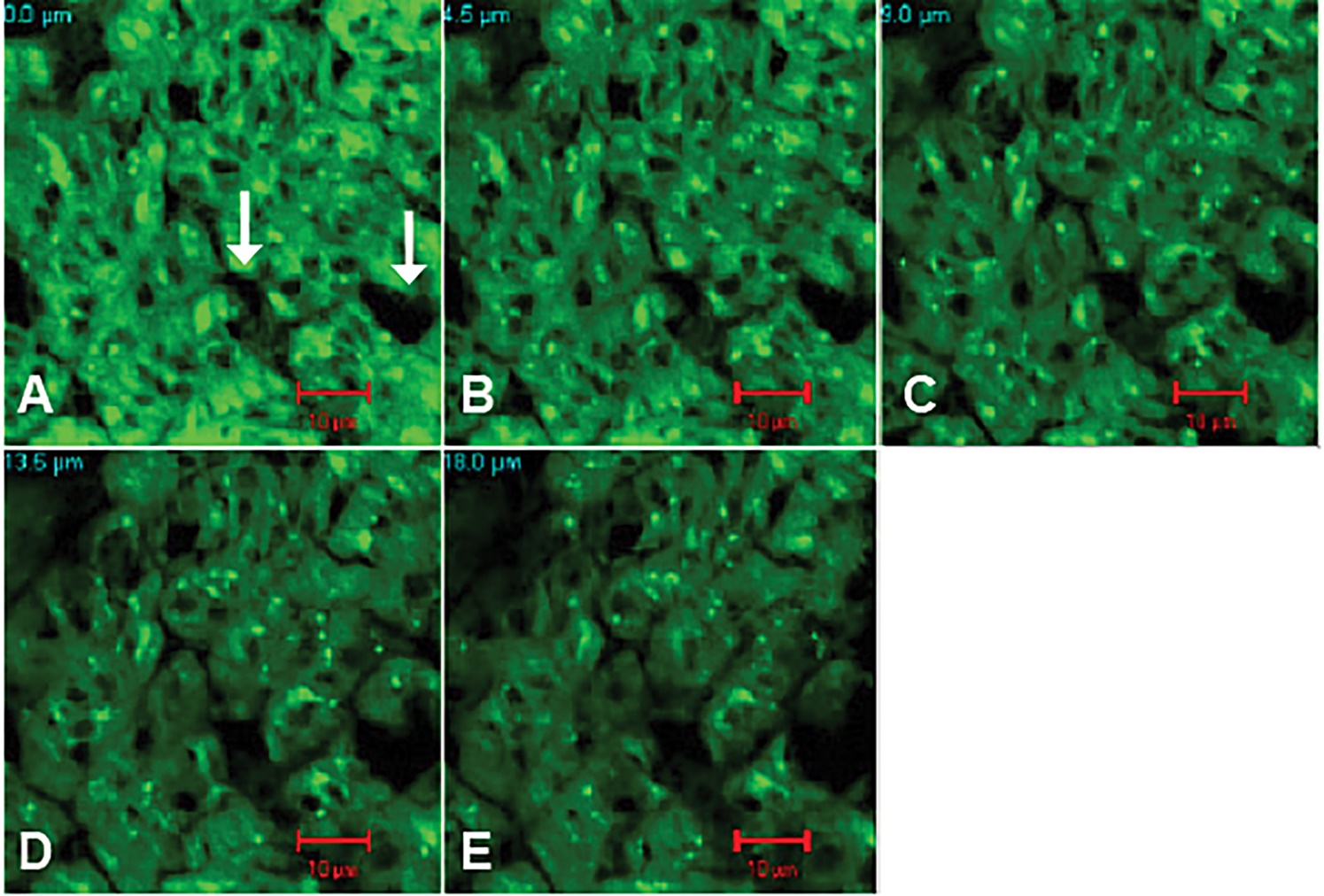
**Supplementary Figure 2.** A schematic representation of the 1,716 base genomic fragment that was sequenced and annotated for the folian-cv1 gene. Within this fragment, a promoter region (pink) and transcriptional start site (TSS: black line), start codon, two exons (purple) defined by splice boundaries, a stop codon, and a putative poly-A site (black line) were identified. The EST (Genbank accession EH645766.1) aligns to the predicted gene structure with 1 large gap representing the putative intron (white); the blue boxes show alignment. The two green boxes show the location of the forward and reverse primer designed from the EST.



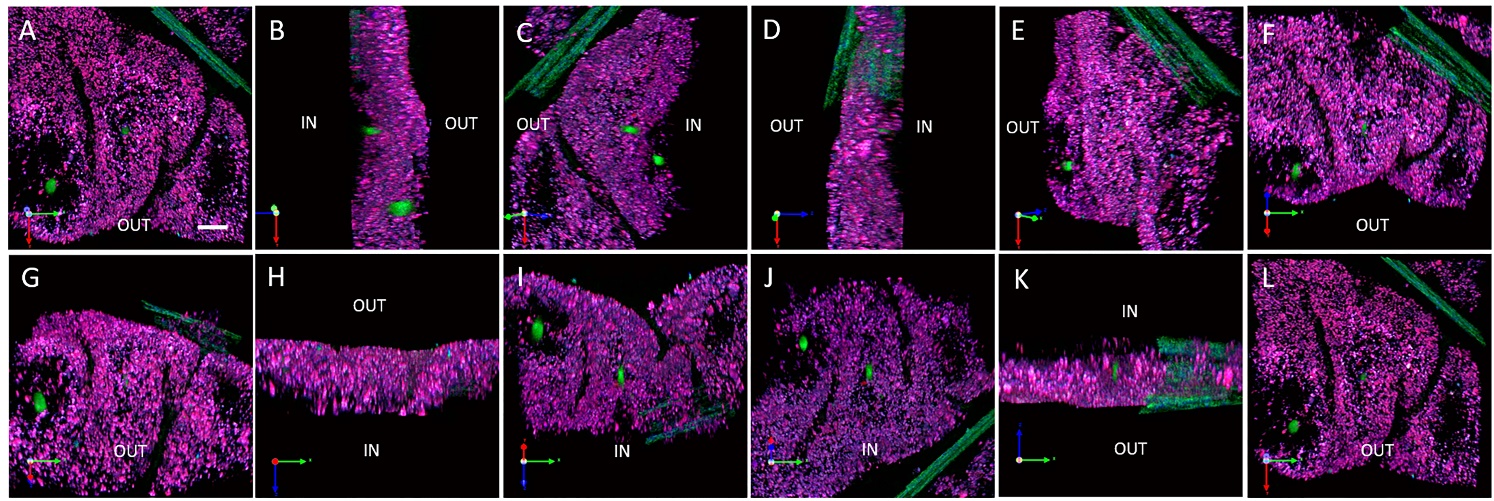
**Supplementary Figure 3.** Alignment of folian-cv1 deduced from the BAC sequence from South Carolina oysters with two deduced variants identified from the *C. virginica* genome from Delaware oysters using Clustal Omega. The two Delaware variants were identified using the Blastx search engine and the folian-cv1 protein sequence as a probe. The variants were identified on chromosome 1 at positions 111101504 and 111119246. Blue shaded areas indicate exact match regions; “----” indicates gap/insert regions. Substitutions are indicated with lower case letters in line with gray-shaded matched sequence.



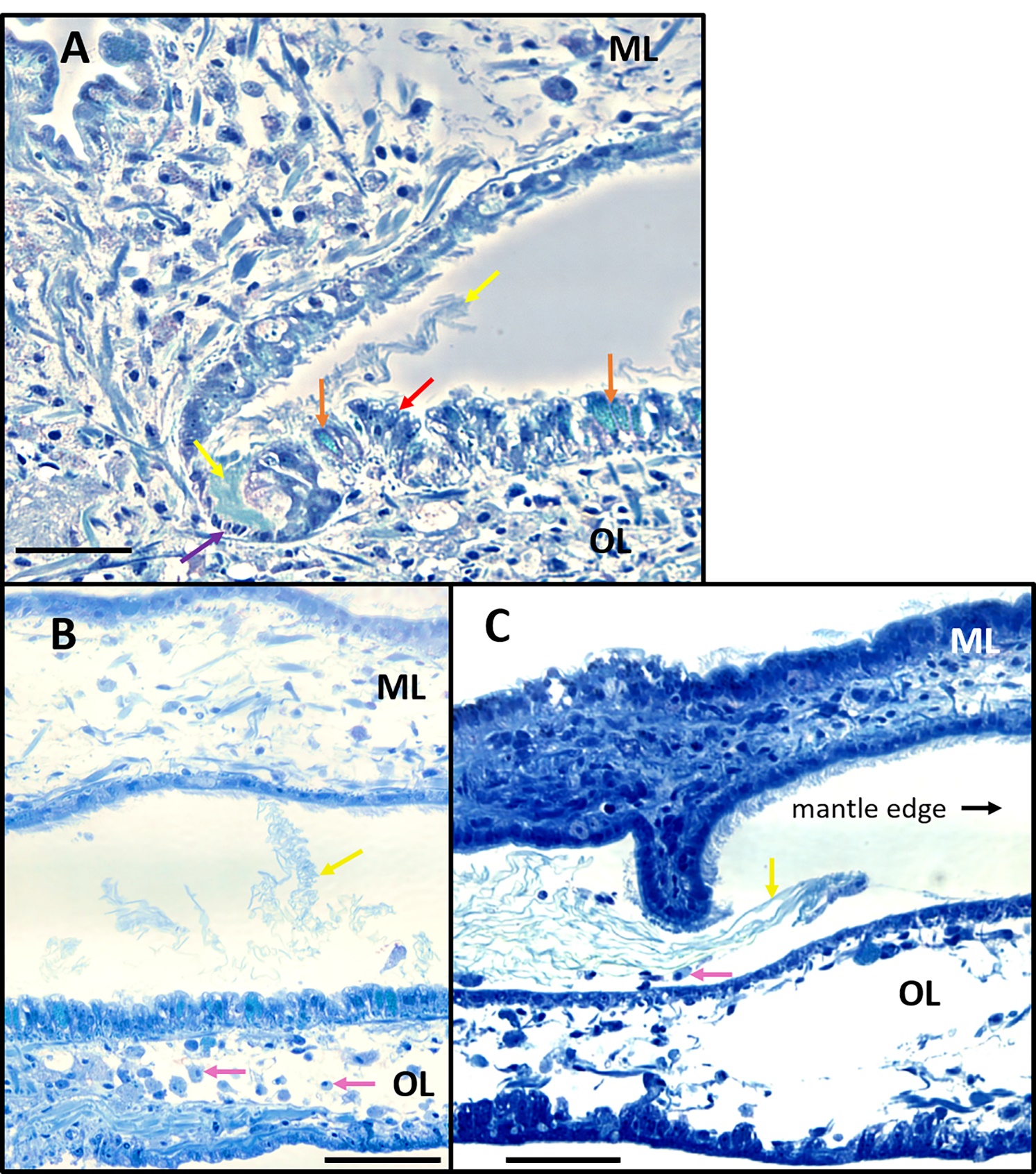
**Supplementary Figure 4.** Folian-stained mantle section showing the sheet structure extending along outer mantle surface toward the umbo. Scale bar 750 µm.

****

**Supplementary Figure 5**. Confocal laser scanning microscopy of optical cross sections of living outer (shell-facing) epithelium of the outer mantle lobe. The image was obtained by a Zeiss 510 LSM in the FITC channel from an 18 µm living autofluorescent oyster mantle tissue section. **(A)** The epithelial surface. The two white arrows indicate openings in the surface. **(B-E)** The openings continue as channels from 4.5 to 18 µm depth (towards the basal lamina). Scale bars **A-E** 10 µm.



**Supplementary Figure 6.** Images from a confocal stack of a living mantle epithelium following re-injection of calcein AM-stained hemocytes. The tissue was excited at 488 and 594 nm. Panels A – F show the rotation of the mantle along the y axis. Panels G-L show the rotation of the mantle along the x axis. Two re-injected hemocytes, (green), are observed moving from the hemocoel (labeled IN) through the epithelial layer towards the shell facing extrapallial space (labeled OUT). Mantle epithelium is autofluorescent (purple). An autofluorescent fibrous sheet (green) can be seen between two mantle folds. The white scale bar in panel A is 40 µm and applies to all panels. The confocal stack can be dynamically viewed in the **Supplementary Video 1**.

****

**Supplementary Figure 7.** Histochemical staining using Azure II of the periostracal groove region between the outer (OL) and middle (ML) mantle lobes. **(A)** The middle lobe epithelium facing the groove is lined with ciliated simple cuboidal cells while columnar pseudostratified secretory cells line the outer lobe. Periostracal material collects at the base of the groove and is extruded into the periostracal groove (yellow arrows). Secretory cells that stain light blue (orange arrows) are distinct from other secretory cells that stain dark blue (red arrow). The lighter staining cells contain multiple vesicles and are distributed similarly to the folian-positive cells in Figure 7. They have a morphology consistent with the phosphoprotein (folian) secreting cells described in Myers et al. (2007). Some of the dark blue cells appear to correlate to the non-immuno-reactive cells, which appear black in Figure 7. These include the simple cuboidal cells (purple arrow) and secretory cells (red arrow). Together, these cells may constitute the “periostracal gland” previously reported by Galstoff (1964) and Neff (1972b). Scale bar 60 µm. **(B)** Periostracum collects in the groove (yellow arrow) and hemocytes are visible in the lobes (pink arrows) where hemolymph circulates *in vivo*. Scale bar 80 µm. **(C)** Hemocytes (pink arrow) are evident in association with nascent periostracal material (yellow arrow) and are likely transported together with the material toward the mantle edge (black arrow), presumably through the action of the ciliated cells and the inflation and contraction of the mantle lobes. Scale bar 80 µm.

**Supplementary Table 2.** Percent similarity scoring matrix following pairwise alignments of putative homologs from oysters and from a scallop.1

|  |  |  |
| --- | --- | --- |
| Pairwise Alignments | % Similarity2 | |
| *Oysters* | *Gap penalty 10* | *Gap penalty 5* |
| folian-cv1 and DSPcg | 51.9 | 53.3 |
| folian-cv1 and MPP1 | 32.0 | 38.9 |
| DSPcg and MPP1 | 51.5 | 53.3 |
| *Oysters with scallop* |  |  |
| folian-cv1 and MSP-1 | 24.5 | 26.3 |
| DSPcg and MSP-1 | 34.9 | 40.5 |
| MPP1 and MSP-1 | 38.0 | 41.1 |

1Folian-cv1 from *C. virginica* (BAC copy 1; 307 amino acids), DSPcg from *C. gigas* (461 amino acids),MPP1 from *C. nippona* (497 amino acids)and MSP-1from thescallop, *Patinopecten yessoensis* (820 amino acids).

2Pairwise alignment was determined with EMBOSS Needle alignment tool available from EMBL- EBI (<https://www.ebi.ac.uk/>).

**Supplementary Table 3.** Comparison of deduced amino acid compositions of folian-cv1 with those of putative homologs from two other oyster species and a species of scallop1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Amino Acid | *C. virginica*  Folian-cv1 | *C. gigas*  DSPcg | *C. nippona*  MPP1 | *Scallop*2  MSP-1 |
| Asp | 43.3 | 36.2 | 20.9 | 20.1 |
| Asn | 0.7 | 1.1 | 5.8 | 7.4 |
| Thr | 1.0 | 0.9 | 0.4 | 1.0 |
| Ser | 32.6 | 45.3 | 33.8 | 32.3 |
| Glu | 9.1 | 7.4 | 3.4 | 2.7 |
| Gln | 0.0 | 0.0 | 0.0 | 0.5 |
| Pro | 1.0 | 1.1 | 0.6 | 1.5 |
| Gly | 4.2 | 3.5 | 29.0 | 25.0 |
| Ala | 2.3 | 0.9 | 0.4 | 4.1 |
| Cys | 0.0 | 0.0 | 2.6 | 0.0 |
| Val | 1.3 | 0.7 | 0.6 | 0.1 |
| Met | 0.0 | 0.0 | 0.0 | 0.0 |
| Ile | 0.0 | 0.0 | 0.0 | 0.0 |
| Leu | 0.0 | 0.0 | 0.0 | 1.2 |
| Tyr | 4.2 | 2.8 | 1.8 | 0.5 |
| Phe | 0.0 | 0.0 | 0.0 | 0.1 |
| His | 0.0 | 0.0 | 0.0 | 0.1 |
| Lys | 0.0 | 0.0 | 0.4 | 2.1 |
| Arg | 0.3 | 0.2 | 0.2 | 1.2 |
| Molecular weight in kDa (number of amino acids) | 32.4  (307) | 47.1  (461) | 44.5  (497) | 74.4  (820) |
| Theoretical pI | 2.19 | 2.09 | 2.61 | 3.34 |
| Max. % residues phosphorylated3 | 37.5 | 48.8 | 35.4 | 32.7 |
| Net charge per amino acid4 | -1.08 | -1.16 | -0.77 | -0.69 |
| Hydrophobicity2 | -33.3 | -30.5 | -16.6 | -20.9 |

1 Compositions, molecular weights, pIs and hydrophobicity were determined from deduced genes and do not include potential post-translational modifications such as phosphorylation and glycosylation. Amino acids are expressed as mole percent.

2*Patinopecten yessoensis*

3 Based on predictions using NetPhos 3.1 (www.cbs.dtu.dk/services/NetPhos) for phosphoserine, phosphothreonine and phosphotyrosine.

4 Sum of charges at neutral pH including maximum phosphorylated amino acids divided by total amino acids. Each phosphorylated amino acid was assigned a value of -1.5.

2 Hydrophobicity calculated as described in Supplementary Table 1.

**Supplementary Discussion- Comparison of Proteins from Foliated Microstructures**

A gene sequence encoding a protein annotated as “dentin sialophosphoprotein” (herein abbreviated DSPcg) was identified in the Pacific oyster, *Crassostrea gigas,* genome (NCBI accession number XP\_019924846; gene ID 105333218) using folian-cv1 as a probe. Based on its 52-53% similarity following pairwise alignment with folian-cv1 (**Supplementary Table 2**) and supported by their similar deduced amino acid composition (**Supplementary Table 3**), these two proteins could be homologs and, by our proposed convention, DSPcg could be named folian-cg1. The prominent acidic phosphoprotein (MPP1) from the foliated shell of, *C. nippona*, (Samata et al., 2008) also has some sequence similarity to folian-cv1 (32-38.9%), and even more similarity with DSPcg (51.5-53.3%; **Supplementary Table 2**). While the composition of MPP1 is similar in some ways to folian-cv1 and DSPcg (**Supplementary Table 3**) it is clearly distinguished by its substantially greater glycine content, its inclusion of the potential cross-linker cysteine and its lower aspartic acid content. To include a relative taxonomic outlier from the same order, the composition and sequence of the protein MSP-1 from the foliated layer of the scallop *Patinopecten yessoensis* (Sarashino and Endo, 2001)was also compared to the oyster proteins(**Supplementary Tables 2 and 3**). This protein has a respectable similarity to all the oyster foliated shell proteins but shares the higher glycine content and lower aspartic acid content only with MPP1. All the foliated shell proteins share a high potential for phosphorylation, a low pI, a high net negative charge per amino acid and a low hydrophobicity (**Supplemental Table 3**).However, folian-cv1 and DSPcg rank more extreme in these categories than MPP1 and MSP-1.

Evaluating the patterns in the sequences of these foliated shell proteins in more detail reveals some common themes and some distinct differences among them. All of them contain a signal sequence, confirming they are secreted proteins and likely would not be included incidentally in shell. The signal sequence of folian-cv1 and DSPcg differ by only one amino acid, while that of the others differ significantly from each other and these two proteins. Following the signal sequence, an N-terminal sequence of 23-38 amino acids is hydrophobic (50% or greater hydrophobic amino acids) compared to the composition of the rest of the proteins. Following near the hydrophobic regions, the proteins have distinctive sequences and compositions. For example, folian-cv1 has a glu-rich region with a polyglu sequence, MPP1 has a region with a 7 tandem repeat of NGD and MSP-1 has a lysine-rich region. As to be expected from their compositions, the remainder of the proteins are significantly dominated by asp, ser and gly in various combinations. All of them except MSP-1 have some runs of aspn>2 with the longest run asp10 in folian-cv1. All of them have several runs of sern>2 with the longest being ser15 in DSPcg. In this general asp- ser-, gly-rich region, MSP-1 and MMP-1 both have sequences of multiple tandem repeats of SG with MMP-1 having sequences of ser2-4-gly and tandem repeats of NGD as well. In this same region folian-cg1 and DSPcg have sequences of multiple tandem repeats of DS with some DDS repeats for folian-cg1 and DSS repeats for DSPcg. The proteins also contain longer runs having similar or identical sequences, such as those found in the DE rich region of MMP-1 and the three tandem repeats of DSDSGSDSDS found in folian-cv1. The most remarkable in this regard is MSP-1, for which most of the protein is made up of four substantially identical units each containing separate SG-, D- and K- rich regions. The C-termini of the proteins are all different from each other and distinctive in some way from their general makeup.

Overall, despite some similarities in regions of these various foliated proteins, there is clearly a significant divergence of sequence, even among species in the genus *Crassostrea.* This may not be surprising, as proteins with low complexity regions have been known to evolve rapidly (Marcotte et al., 1999) by a variety of genetic mechanisms mentioned above. Such rapid evolution has been suggested for shell matrix proteins having these regions (Kocot et al., 2016; Aguilera et al., 2017), resulting in shell-related secretomes that differ even among closely related species. Aguilera et al. (2017) postulate that these differences may help create the vast subtleties in shell diversity. Some of these subtleties may not be readily evident. On the other hand, some of the sequence differences may be structurally and functionally neutral.