Doctoral Dissertation
博士論文

Proteomic analysis of shell matrix proteins in the pond snail *Lymnaea stagnalis*

（軟体動物 *Lymnaea stagnalis* の貝殻プロテオーム解析）

A Dissertation Submitted for the Degree of Doctor of Philosophy
July 2020

令和 2 年 7 月　博士（理学）申請

Department of Earth and Planetary Science, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科
地球惑星科学専攻

Akito Ishikawa
石川 彰人

Abstract

Biomineralization is an essential process for the development of hard body parts, which are integral to many organisms. The acquisition of biominerals such as bones, teeth, and shells brought about great diversification of morphology and ecology among animals. Furthermore, biominerals, which often remain as fossils, provide important records of Earth history to understand the interactions, and the histories thereof, between the global environments and life. In order to clarify those interactions and histories, it would be necessary to understand the molecular mechanisms of biomineral formation. In molluscs, their shells have been intensively studied from the viewpoints of morphology, taxonomy, phylogeny, and biomineralization because they show a considerable diversity on the present Earth and have left a rich and continuous fossil record throughout the Phanerozoic. However, detailed molecular mechanisms of molluscan shell biomineralization have not been clarified yet. Molluscan shells consist of inorganic crystals of $CaCO_3$ and organic matters such as polysaccharides and proteins, collectively known as shell organic matrices, which are secreted from the mantle tissue. Recent advances of analytical techniques brought even more drastic changes in our understanding of shell matrix proteins (SMPs). Proteomic analyses combined with genomic and transcriptomic analyses made it possible to almost comprehensively characterize protein sequences contained in biominerals, but a new question has arisen: do all those literally hundreds of SMPs contained in one kind of biomineral have a function in biominerarization?

In Chapter 2 of this study, I identified potentially important SMPs, among other SMPs, by exploiting the asymmetric shell growth of snails. Formation of an asymmetric shell would require laterally asymmetric expression of SMP genes in the mantle tissue. I examined the expression levels of the 35,951 transcripts expressed in the left and right sides of the mantle tissue in the dextral pond snail, *Lymnaea stagnalis*. This transcriptome dataset was used to identify 207 SMPs by LC-MS/MS. A total of 32 out of the 207 SMP genes show asymmetric expression patterns in the transcriptome data, and they were further verified using quantitative PCR analysis, resulting in identification of four asymmetric genes out of those 32 SMP genes. Among the asymmetrically expressed SMPs in dextral snails in the transcriptome analysis or in the combined transcriptome and qPCR analysis, those that are more highly expressed in the left side than in the right side are three times more numerous than those that are more highly expressed in the right than in the left, suggesting importance of inhibitory roles of SMPs in shell formation. This observation was unexpected because it was assumed that a dextrally coiled shell is produced by a greater shell precipitation on the right than the left side of the mantle, and that more shell precipitation-promoting SMPs would be expressed in the right than in the left.

The 32 SMPs identified have distinctive features, such as conserved domains and low complexity regions, which may be essential in biomineralization. One of the SMPs that showed higher expression in the left than in the right, in both transcriptomic and qPCR analyses (Ls-SMP-88), showed a significant sequence similarity to Pif, an SMP originally isolated from the pearl oyster *Pinctada fucata*. Ls-SMP-88 contains two chitin-binding domains (ChtBD2) and an extracellular domain (Laminin_G) as in Pif, but, it has no von Willebrand factor type A domain (VWA), which is involved in protein binding, and is always found in Pif. Phylogenetic analysis of ChtBD2 and Laminin_G domain sequences indicated that Ls-SMP-88 is closely related to Pif of bivalvia, and that it originated as Pif, but lost the VWA domain subsequently. In pearl oysters, Pif binds aragonite crystals and promotes nacre formation. Although functions of Ls-SMP-88 have yet to be clarified, one possibility is that the loss of the VWA domain led to loss of shell formation-promoting roles, and acquisition of inhibitory roles instead. These results suggest that a dextrally coiled shell is produced by inhibition of shell precipitation on the left. Inhibitory roles of SMPs have long been recognized, and they could be at work in the process of coiled shell formation.

In Chapter 3, focusing on the laterality of the coiling direction of the snail, comparative proteomic analyses between the dextral and sinistral strains of *L. stagnalis* have been performed. *L. stagnalis* shows two types of shell coiling, namely, dextral (wild type) and sinistral (mutant type), and they have a mirror-symmetrical relationship with each other. In order to minimize the differences in genetic background between the dextral and sinistral strains, those two strains were crossed, and then each of the dextral and sinistral strains was established once again before being subjected to proteomic analyses. In the proteomic analyses, about 100 individuals each of the dextral and sinistral shells were sampled and provided for the comparative analysis. As a result, a total of 443 SMPs have been identified. The 443 SMPs include all the 207 SMPs identified in the dextral strain in Chapter 2.

The comparisons of protein repertoires between the dextral and sinistral shells indicated no difference between them, but relative abundance of the proteins contained in the shell was different for some proteins: the most abundant SMP in the dextral shells, comp88734_c0_seq1, is 3.13 times more abundant in the dextral than in the sinistral shells, while the most abundant SMP in the sinistral shells, comp88616_c0_seq1, is 2.88 times more abundant in the sinistral than in the dextral snail. These results suggest that the abundance profiles of SMPs in each of the dextral and sinistral shells are not simple mirror images. This would mean that the lateral asymmetry in *L. stagnalis* is not a simple matter of asymmetry determination at the very early developmental stage, but involves some “maintenance mechanisms”, which result in different expression profiles of SMP encoding genes in the mantle between dextral and sinistral strains in late

developmental stages. The fact that the shell shapes are not exact mirror images between dextral and sinistral shells is consistent with this hypothesis. The exact mechanisms, including the gene regulatory pathways of SMP expression need to be elucidated by further analysis.

Furthermore, the expression levels of the 39,069 mantle transcripts have been compared between the right and left sides of mantle tissues in the sinistral strain of *L. stagnalis*. Contrary to expectation, the expression patterns of SMP genes in the sinistral strain did not mirror the asymmetric pattern observed in the dextral strain shown in Chapter 2. Only one SMP (comp153562_c0_seq1) gene indicated a statistically significant difference in expression level between the right and left sides of the mantle, with higher expression in the left (outer side of the sinistral shell) than in the right (axial side of the sinistral shell). The SMP comp153562_c0_seq1 has two EFh (EF-hand) domains which can interact with cations. This SMP was identified in the dextral shells, but was found to show no significant difference in the expression levels of the transcripts between the right and left sides in the dextral snails. Instead, Ls-SMP-61 and Ls-SMP-62, which are distinct from comp153562_c0_seq1, but possess components of conserved domains similar to those of comp153562_c0_seq1, have been identified in the 32 asymmetric SMPs in the dextral strain discussed in Chapter 2, and are expressed higher in the left than in the right in the dextral snails. These observations suggest that a similar (but not the same) set of SMPs control some aspects of shell formation in dextral and sinistral strains. It is notable that, although the dextral strain and sinistral strains have almost the same genome background, they indicate distinctly different expression profiles of SMP genes.

In order to understand the functions of SMPs in the future, it is necessary to analyze the *in vivo* functions of SMPs using genome editing techniques such as CRISPR/Cas9. The SMPs which have been narrowed down as potentially important in this study provide the first candidates to be analyzed in those *in vivo* functional analyses. In this context, investigations of not only the expression profiles of SMP genes but also the arrangements of SMPs in the shell and the relationships between the SMPs and the shell microstructures will be essential. In future, evolution of the molluscan shell morphology may be understood as a history of changes in the developmental programs of the shell. To this end, the lateral asymmetry of shell morphology, on which this study put a focus, provides a unique foothold for elucidating the molecular mechanisms of shell formation. It is hoped that further studies focusing on this aspect help understand the nature of hard tissue formation and evolution.

Contents

1. General Introduction

Biomineralization is an essential process for the development of integral body parts in many organisms. Acquisition of biominerals greatly facilitated life, providing physical support of the body, armor against predation, apparatus for food ingestion, attachment for muscles, gravity sensors, optical lenses, magnet for navigation, etc. It played pivotal roles in the diversification events of morphology and ecology of life, such as terrestrialization and arms races between predators and predators (Schmidt-Nielsen, 1984; Vermeiji, 1987; Coates, 1996; Kelley et al., 2003; Shubin et al., 2004; Ahlberg et al., 2005; Glenner et al., 2006; Ashley-Ross et al., 2013). Furthermore, biominerals often remain as fossils, and thus provide important records of Earth history, including the history of the interactions between global environments and life (Clack, 1997; Shu et al., 1999; Kelley et al., 2003; Knoll, 2003).

Over 62 different biominerals are known, including the hydroxyapatite of bones or teeth, calcite and aragonite of the mollusc shells, gypsum of jellyfish larvae, barite of algae, silica of plants, diatoms, or sponges, magnetite of the magnetotactic bacteria, goethite of limpet teeth, and ferrihydrite of mammalian ferritin cores (Lowenstam & Weiner, 1989). Biominerals are composed of inorganic mineral crystals and organic molecules, which are mainly polysaccharides and proteins.

The biomineralization process can be divided into the following 6 important steps (Lowenstam & Weiner, 1989): (1) space delineation, or the creation of a boundary space in which nucleation can occur; (2) matrix formation, when proteins which promote aggregation are synthesized and exported by cells to the boundary space, where they assemble and form a hydrated environment for nucleation to occur; (3) supersaturation, which is a condition where the concentration of ions in the boundary space reaches a critical threshold of insolubility, guaranteeing the nucleation of an inorganic solid; (4) control over nucleation, when agents or additives produced by cells are exported to the boundary space and control the nucleation of biominerals; (5) control over biomineral formation controlled by the cells with the assistance of agents or additives; and (6) termination of biomineralization. As the biomineralization process reaches its dimensional limit, the nucleation and mineral deposition processes are terminated. These steps are likely driven by intracellular regulation.

The organic molecules extracted from biominerals have conventionally been classified into soluble and insoluble fractions based on the solubility in water. Insoluble matrix molecules are generally much more abundant than soluble ones, and are considered to provide framework or scaffold for biomineralization, while soluble components have been considered to help control the nucleation in the above step (4), and the subsequent mineralization steps (5) and (6), as an additive to the biominerals (Lowenstam & Weiner, 1989; S. Mann, 2001).

Matrix proteins, which are considered as the main controllers of biomineralization, have been characterized by primary structure determination, first by one-by-one approaches based on Edman degradation (e.g. Sarashina and Endo 1998), followed by more comprehensive proteomic analyses (see Table 1.1). *In vitro* functional analyses of matrix proteins have also been performed to reveal promotive or inhibitory roles of shell matrix proteins (SMPs) on $CaCO_3$ crystallization, and other potential functions of SMPs. For example, the acetic acid-soluble SMP Caspartin inhibits calcium carbonate precipitation (Marin et al., 2005). The SMP Prismalin-14 also inhibits precipitation of calcium carbonates *in vitro* (Suzuki & Nagasawa, 2007). The super-acidic SMP Aspein controls the $CaCO_3$ polymorph formation *in vitro* (Takeuchi et al., 2008). The SMP Perlucin promotes the calcium carbonate precipitation and modification of crystal morphology at ambient conditions (Wang et al., 2008). The acidic SMP Pif is a key molecule in the induction of aragonite crystal formation (Suzuki et al., 2009). The SMP Pearlin modulates mineral growth/formation and recovers conformation to bind calcium or to bind alpha-chitin (Montagnani et al., 2011). The basic SMP PfN23 accelerates the deposition of calcium carbonates and induces the formation of aragonite crystals (Fang et al., 2012). Lysine-rich SMP KRMP7 inhibits $CaCO_3$ precipitation, changes the morphology of calcite, and inhibits the growth of aragonite (Liang et al. 2016). *In vivo* functional analyses of matrix proteins have been advanced mainly in the studies of bones in model organisms (e.g. Patel et al. 2006; Williams et al. 2009; G. Zhao et al. 2018). With those *in vitro* and *in vivo* functional analyses of matrix proteins, detailed molecular mechanisms of biomineralization are still unclear.

Species	Year	Refference	Numbers of reference seqs	Numbers of Proteins	Author
Pinctada margaritifera	2007	NCBI nonredundant DB		6	Bédouet et al.
Haliotis asinina	2010	EST	8,335	14	Marie et al.
Mytilus edulis	2011	EST	5,000	3	Marie et al.
Mytilus galloprovincialis	2011	EST	19,000	5	Marie et al.
Mytilus californianus	2011	EST	42,300	6	Marie et al.
Crassostrea gigas	2011	EST	220,000	8	Marie et al.
Pinctada margaritifera	2012	EST	76,790	78	Marin et al.
Pinctada fucata	2012	Transcriptome	800,982	no proteomic analysis	Takeuchi et al.
Crassostrea gigas	2012	Genome DB		259	Zhang et al.
Lottia gigantea	2012	Genome DB		569	Mann et al.
Pinctada margaritifera	2012	EST	76,790	78	Marie et al.
Pinctada maxima	2012	EST	7,272	42	Marie et al.
Mytilus edulis	2014	Transcriptome	26,785	no proteomic analysis	Freer et al.
Lottia gigantea	2014	Genome DB + UniProtKB		57	Mann & Edsinger
Cepaea nemoralis	2014	Transcriptome	676,358	59	Mann & Jackson
Pinctada fucata	2015	Genome DB		144	Liu et al.
Mytilus coruscus	2015	Transcriptome	106,452	63	Liao et al.
Mytilus galloprovincialis	2015	EST	68,209	113	Gao et al.
Mya truncata	2016	Transcriptome	20,106	71	Arivalagan et al.
Crassostrea hongkongensis	2016	SMP data of previous studies		42	Upadhyay & Thiyagarajan
Crassostrea angulate	2016	SMP data of previous studies		37	Upadhyay & Thiyagarajan
Crassostrea gigas	2017	SMP data of previous studies		53	Feng et al.
Haliotis laevigata	2018	Transcriptome	not mentioned	448	Mann et al.
Lymnaea stagnalis	2018	Transcriptome	164,254	40	Herlitze et al.
Euhadra quaesita	2019	Transcriptome	74,293	54	Shimizu et al.
Lymnaea stagnalis	2020	Transcriptome	337,195 (dextral) 222,444 (sinistral)	443	in this study

Table 1.1.

The list of recent previous studies which focused SMPs by transcriptomic or proteomic analyses.

In molluscs, their shells have been intensively studied from the viewpoints of morphology, taxonomy, phylogeny, and biomineralization because they show a considerable diversity on the present Earth and have left a rich and continuous fossil record throughout the Phanerozoic (Ponder & Lindberg, 2008; Kocot et al., 2011; Smith et al., 2011). The phylum of extant molluscs is composed of eight classes that are divided into two major lineages, the Conchifera including gastropods, bivalves, scaphopods, cephalopods and monoplacophorans, and the Aculifera including polyplacophorans, solenogastres, and caudofoveates (Kocot et al., 2011; Smith et al., 2011; Vinther et al., 2012; McDougall & Degnan, 2018). Molecular phylogenetic analyses greatly updated the relationships among those molluscan classes, however, even recent phylogenomic analyses show some conflicts, with gastropods being shown to be sister to scaphopods in one study, and to bivalves in another (Kocot et al., 2011; Smith et al., 2011; Vinther et al., 2012; McDougall & Degnan, 2018). Molluscan morphological diversity of the shell, with over 100,000 species, is among the largest, only next to arthropods, in extant animal phyla (Ponder & Lindberg, 2008). The molluscan shells also exhibit a micro-level diversity in the form of "shell microstructures", including such diverse structures as nacreous, prismatic, cross-lamellar, foliated, granular, and homogeneous structures (Lowenstam & Weiner, 1989; Chateigner et al., 2000; S. Mann, 2001). In other words, the molluscan shells are greatly diverse, from micro- to macro-levels. However, detailed molecular mechanisms of molluscan shell biomineralization have not been clarified yet.

The molluscan shells generally consist of inorganic crystals of $CaCO_3$ and organic matters such as polysaccharides and proteins, which are secreted from the mantle tissues (Lowenstam & Weiner, 1989; S. Mann, 2001). Pioneering studies on shell protein sequences have identified such matrix proteins as the carbonic anhydrase Nacrein (Miyamoto et al., 1996) as well as the likely shell framework proteins MSI60 and MSI31 from the pearl oyster *Pinctada fucata* (Sudo et al., 1997), and another framework protein Lustrin A from the abalone *Haliotis rufescens* (X. Shen et al., 1997). An unusually acidic shell protein, MSP-1, which is rich in aspartic acid, was then identified from the scallop *Patinopecten yessoensis* (Sarashina & Endo, 1998, 2001). The following decade saw a boom of isolation and sequence determination for a wealth of molluscan shell matrix proteins (Sarashina & Endo, 2006; Marin et al., 2007). In fresh water molluscs, the extracellular matrix protein known as dermatopontin was found in the shell matrix of *Biomphalaria glabrata* (Marxen et al., 2003).

Recent development of analytical techniques made it possible to investigate the transcriptome of mantle tissues and proteomes of the shell matrices from very small quantities of the sample material, potentially allowing us to detect interactions among various shell proteins. Indeed, this comprehensive approach triggered a burst of novel shell proteins identified from various molluscan species including *Pinctada margaritifera*

(Bédouet et al., 2007; Berland et al., 2011), *P. margaritifera* and *P. maxima* (Marie et al., 2012), *Crassostrea gigas* (Marie, Zanella-Cléon, et al., 2011), *Haliotis asinina* (Marie et al., 2010), *Mytilus edulis*, *Mytilus galloprovincialis*, and *Mytilus californianus* (Marie, Le Roy, et al., 2011; Freer et al., 2014), *Pinctada fucata* (Takeuchi et al., 2011; Miyamoto et al., 2013), *Lottia gigantea* (K. Mann et al., 2012; Marie et al., 2013), *Cepaea nemoralis* (K. Mann & Jackson, 2014), *Mytilus coruscus* (Liao et al., 2015), *Mya truncata* (Arivalagan et al., 2016), *Haliotis laevigata* (K. Mann et al., 2018), *Lymnaea stagnalis* (Herlitze et al., 2018), and *Euhadra quaesita* (Shimizu et al., 2019).

Functions of those proteins in shell formation, however, have remained largely unexplored. It is desirable to establish a system with which to study *in vivo* functions of shell matrix proteins systematically. Previous transcriptome studies have mainly used marine molluscs, and only a few freshwater molluscs have been subjected to transcriptomic or proteomic analysis. In this study, I characterized the shell matrix proteins of the pond snail *Lymnaea stagnalis*.

The freshwater pond snail *Lymnaea stagnalis* (Linnaeus, 1758) (Fig. 1a), also referred to as great pond snail or common pond snail, belongs to the phylum Mollusca, class Gastropoda, subclass Heterobranchia, superorder Hygrophila and family Lymnaeidae. *L. stagnalis* inhabits fresh waters, with low current, and are predominantly herbivores, feeding mostly on plants. Detailed embryonic development is summarized by (Meshcheryakov, 1990).

Piaget (1929) reported that the shell shapes of *L. stagnalis* are somewhat different between those in running water and still water environments. Arthur (1982) confirmed that this observation was statistically significant in (Piaget, 1929; Arthur, 1982). Asami et al. (2008) have established laboratory lines of enantiomorphs of the pond snail *L. stagnalis* starting from a wild population, and revealed that the early embryos of the dextral and sinistral strain of *L. stagnalis* are not mirror images of each other (Asami et al., 2008). Kuroda et al. (2009) found that when the 8-cell stage micromeres of a sinistral embryo were artificially placed so as to twist clockwise, the individual grew normally as a dextral embryo (Kuroda et al., 2009). Shimizu et al. (2011; 2013) reported the signal transduction factor Dpp showed laterally asymmetric expression patterns in the embryonic shell gland, embryonic mantle, and adult mantle, with high levels of expression in the right and left sides of the body in those shell forming tissues in the dextral and sinistral snails, respectively. When embryos of *L. stagnalis* were subjected to a chemical that inhibits the signal transduction by Dpp, they produced a shell of a cone-like shape without coiling, confirming that Dpp is responsible for shell coiling in those snails (Shimizu et al., 2011, 2013). Then, Davison et al. (2016) revealed that Formin, which is a protein involved in polymerization of actin, and is encoded by the *Lsdia1* gene, is likely the long-sought maternal factor that determine the coiling direction in *L.*

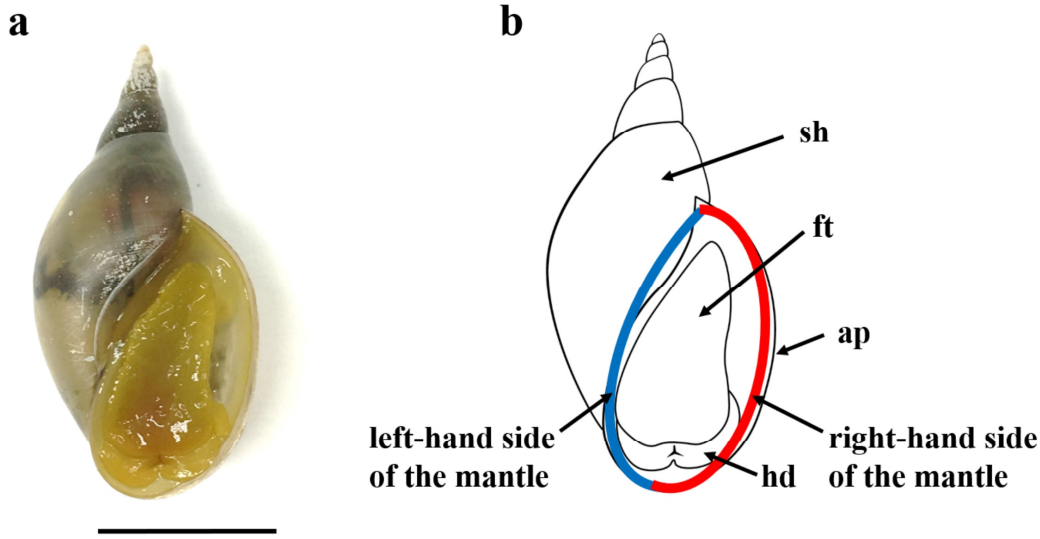

Figure 1.

An adult individual of *Lymnaea stagnalis* showing the position of mantle tissues dissected for analysis. (a) Ventral view of *L. stagnalis*. (b) A schematic diagram of *L. stagnalis*. Red line: right side, blue line: left side of the mantle, ap: shell aperture, ft: foot, hd: head, and sh: shell. Scale bar = 1 cm.

stagnalis, a discovery which was confirmed by gene knockout experiments of Abe & Kuroda (2019).

The shell microstructure of *L. stagnalis* is composed by crossed lamellar microstructure (Yonezawa et al., 2016), and shell matrix proteins have also been characterized in *L. stagnalis* (Sarashina et al., 2006; Yonezawa et al., 2016; Herlitze et al., 2018). Herlitze et al. (2018) identified 34 candidate shell-forming proteins, showing that their transcripts display a variety of spatial and temporal expression patterns at different developmental stages (Herlitze et al., 2018).

L. stagnalis has a short life cycle and can be reared easily. Besides, *L. stagnalis* has been widely used in neurophysiology, embryology, environment toxicology, and so on (de Vlieger et al., 1980; Rittschof & McClellan-Green, 2005; Bandow & Weltje, 2012; Munley et al., 2013; Atli & Grosell, 2016; Amorim et al., 2019). Therefore, *L. stagnalis* has a high potential of becoming a 'model organism' in the study of shell formation.

In this study, by investigating the shell matrix proteins of *L. stagnalis* much more comprehensively than in any previous studies, I attempted to identify potentially functionally important SMPs out of the hundreds of SMPs characterized by proteomic analysis, and to make a basis for understanding the mechanisms of biomineralization as well as the evolutionary processes of shell formation in molluscs.

2. Functional shell matrix proteins tentatively identified by asymmetric snail shell morphology

2.1. Background

Biomineralization is the process by which organisms incorporate and deposit minerals. The end products, called biominerals, are composed of both minerals and organic matrices, which are considered essential to formation of highly ordered, functional materials (Lowenstam & Weiner, 1989; S. Mann, 2001). In these organic matrices, proteins are the major components and have attracted much interest. SM50, from sea urchin larval spicules, was the first such protein sequenced among calcium carbonate biominerals (Sucov et al., 1987). Subsequently, studies of molluscan shell proteins have identified such matrix proteins as the carbonic anhydrase nacrein (Miyamoto et al., 1996) and the probable shell framework proteins, MSI60 and MSI31 (Sudo et al., 1997), from the pearl oyster, *Pinctada fucata*. The following decade saw a significant surge in sequence determination of skeletal matrix proteins (Sarashina & Endo, 2006; Marin et al., 2007; Nagasawa, 2013).

Recent advances in analytical techniques brought even more drastic changes in our understanding of matrix proteins. Proteomic analyses combined with genomic and transcriptomic analyses made it possible to almost comprehensively characterize protein sequences from biominerals. These advances triggered a burst of novel matrix proteins identified from various biominerals, including chicken eggshells (K. Mann et al., 2006), sea urchin larval spicules (K. Mann et al., 2008), shells of molluscs (G. Zhang et al., 2012; K. Mann et al., 2012, 2018; K. Mann & Jackson, 2014; Gao et al., 2015; Liao et al., 2015; Liu et al., 2015; Arivalagan et al., 2016; Upadhyay & Thiyagarajan, 2016; Di et al., 2017; Shimizu et al., 2019), and brachiopods (Isowa et al., 2015; Jackson et al., 2015; Luo et al., 2015). In these studies, literally hundreds of proteins have been identified from biominerals of individual species.

These techniques enabled a new era of proteomic biomineralization studies, however, they also raised a conundrum. Previous studies identified proteins that are not specific to biomineralization, e.g., house-keeping proteins such as EF-1α and ribosomal proteins (G. Zhang et al., 2012). Do all these matrix proteins function in biomineralization? Is there a way to identify essential SMPs among the literally hundreds of SMPs identified by omics approaches? To address these questions, we focused our attention on the pond snail, *Lymnaea stagnalis* (Fig. 1a). Because these snails produce coiled-shells, which can only be produced by asymmetric accretion of shell material to the shell aperture, we hypothesized that some genes responsible for shell formation may be differentially expressed between the left and right sides of their mantle tissues. In other words, by comparing gene expression levels between the left and right sides of the mantle, we

anticipated being able to identify functionally important proteins. We posited that enhanced and diminished SMP expression would result in more biomineralization on the right side than left side of the mantle in dextral snails (Fig. 1b).

In addition, *L. stagnalis* is an ideal organism for such a study because it has a short life cycle and can be reared easily in the laboratory. For this reason, it has been used for a wide range of studies, including neurophysiology, embryology, and environmental toxicology (Davies & Henrissat, 1995; Rittschof & McClellan-Green, 2005; Shimizu et al., 2011; Munley et al., 2013). At least three previous studies have used transcriptomic analyses to understand the snail central nervous system or responses to a pesticide (Z.-P. Feng et al., 2009; Bouétard et al., 2012; Sadamoto et al., 2012). Shell matrix proteins (SMPs) have also been characterized in *L. stagnalis* (Sarashina & Endo, 2006; Herlitze et al., 2018). Herlitze et al. (2018) identified 34 candidate shell-forming proteins, showing that their transcripts display a variety of spatial and temporal expression patterns at different developmental stages (Herlitze et al., 2018). Therefore, *L. stagnalis* is an ideal 'model organism' for biomineralization.

In this study, we identified shell matrix proteins of *L. stagnalis* using a combination of proteomic and transcriptomic analyses. Gene expression levels of shell matrix proteins have been compared between the right and left sides of the mantle. We identified 32 shell matrix protein genes that are asymmetrically expressed in the mantle transcriptome, suggesting their roles in shell formation in this species. Using quantitative PCR analysis, asymmetric expression patterns were further verified for four of these 32 SMPs. The shell proteomic and transcriptomic data presented here may support additional studies of biomineralization mechanisms, as well as evolutionary processes of shell formation in molluscs.

2.2. Results

2.2.1. Mantle transcriptomic analysis

Approximately 70 million reads were obtained for each of the 6 paired-end libraries prepared from left and right mantle tissues of three biological replicates. The read length was 200 bp and GC content was ~40% (Table 2.1). Sequence assemblages using all six pairs of samples from mantle tissues generated 337,195 contigs with a maximum contig length of 37,809 bp, an average length of 1,140 bp, and an N50 value of 2,828 bp. Local BLASTN searches of these contigs against the whole genome shotgun sequence of *L. stagnalis* (GCA_900036025.1, unpublished, Ashworth Laboratories, 2016) returned significantly similar sequences for 309,623 contigs (e-value < 10^{-10}). After an ORF search by TransDecoder, 162,121 contigs remained in the FASTA file with a maximum contig length of 32,196 bp, an average length of 1,050 bp, and an N50 value of 1,728 bp. After clustering with CD-HIT, 35,951 sequences remained with a maximum sequence length of 32,196 bp, an average length of 1,190 bp, and an N50 value of 1,974 bp. After clustering with CD-HIT, we used those sequences as references for the proteomic analysis. The sequence assembly, gene set, and transcriptome completeness of the FASTA file have been checked using BUSCO statistics (Simão et al., 2015; Waterhouse et al., 2018). The results indicated that our contig sequences are well assembled and comprise a nearly complete gene set that identified 99.1% (969) complete genes and 0.6% (6) fragmented genes among the 978 metazoan BUSCO genes (Supplementary Table S2.1).

About 7 million reads with read lengths of 35 - 309 bp and a GC content of 42% were obtained as paired-end sequences for foot tissue (Table 2.1). After removal of low quality reads, 5,965,429 reads (paired-end pairs) remained in the FASTQ file (Table 2.1). BUSCO statistics for the FASTA file obtained for the foot transcriptome (116,738 contig sequences; Fig. 2.1) identified 93.7% (917) complete genes and 5.7% (56) fragmented genes, a value slightly lower than, but comparable to the value for the mantle transcriptome shown above (Supplementary Table S2.2).

File name	Source of RNA	Direction of Paired-end reads	Number of reads in total	Sequence length	GC content (%)	DRA Accession number
Sample07_TAGCTT_ALL_R1_001.fastq	Left side of mantle from individual #1	Forward	75,993,110	101	39	SAMD00074112
Sample07_TAGCTT_ALL_R2_001.fastq	Left side of mantle from individual #1	Reverse	75,993,110	101	39	SAMD00074112
Sample08_GGCTAC_ALL_R1_001.fastq	Left side of mantle from individual #2	Forward	68,329,443	101	37	SAMD00074113
Sample08_GGCTAC_ALL_R2_001.fastq	Left side of mantle from individual #2	Reverse	68,329,443	101	37	SAMD00074113
Sample09_GTGGCC_ALL_R1_001.fastq	Left side of mantle from individual #3	Forward	66,846,658	101	39	SAMD00074114
Sample09_GTGGCC_ALL_R2_001.fastq	Left side of mantle from individual #3	Reverse	66,846,658	101	39	SAMD00074114
Sample10_GTTTCG_ALL_R1_001.fastq	Right side of mantle from individual #1	Forward	69,768,907	101	39	SAMD00074115
Sample10_GTTTCG_ALL_R2_001.fastq	Right side of mantle from individual #1	Reverse	69,768,907	101	39	SAMD00074115
Sample11_CGTACG_ALL_R1_001.fastq	Right side of mantle from individual #2	Forward	70,252,413	101	40	SAMD00074116
Sample11_CGTACG_ALL_R2_001.fastq	Right side of mantle from individual #2	Reverse	70,252,413	101	40	SAMD00074116
Sample12_GAGTGG_ALL_R1_001.fastq	Right side of mantle from individual #3	Forward	68,107,519	101	39	SAMD00074117
Sample12_GAGTGG_ALL_R2_001.fastq	Right side of mantle from individual #3	Reverse	68,107,519	101	39	SAMD00074117
Ls8_S3_L001_R1_001.fastq	Foot from individual #4	Forward	7,127,862	35 - 309	42	SAMD00106507
Ls8_S3_L001_R2_001.fastq	Foot from individual #4	Reverse	7,127,862	35 - 309	42	SAMD00106507

Table 2.1.

Details of transcriptomic data obtained in this study.

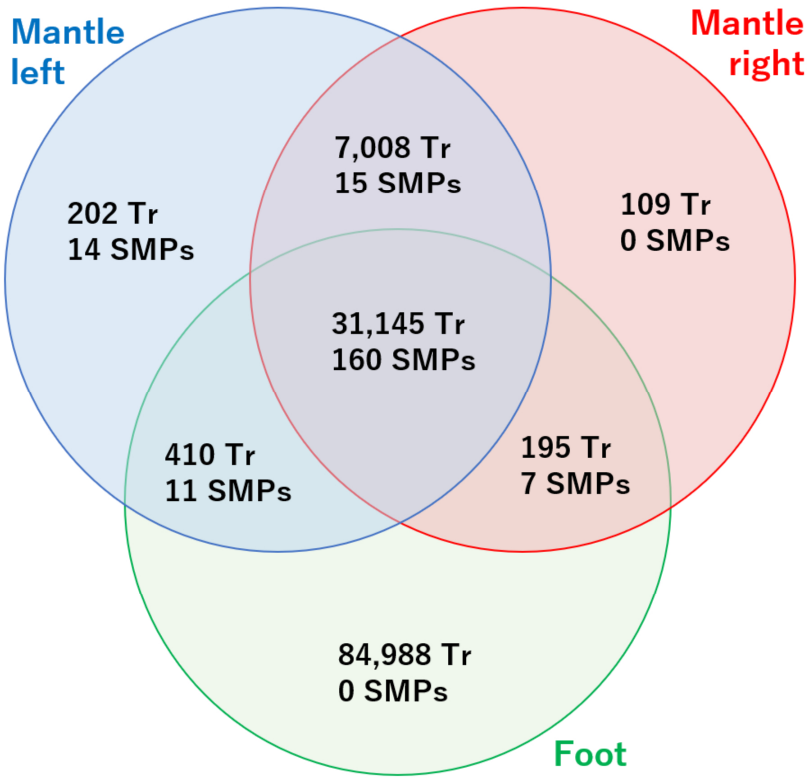

Figure 2.1.

Venn diagram showing numbers of transcripts (Tr) and SMPs identified in right mantle, left mantle, and foot using transcriptomic and proteomic analyses.

2.2.2. Proteomic analysis

LC-MS/MS identified 378 unique peptide fragments from the shell matrix of *L. stagnalis*. Of these 378 peptide fragments, 91, 233, and 55 peptides were identified from the soluble, insoluble, and both soluble and insoluble fractions, respectively (Supplementary Table S2.3). Using transcriptomic data obtained from mantle tissues of *L. stagnalis*, protein sequences identified by more than one unique peptide were employed in subsequent analyses. In all, 207 proteins were identified with 21, 116, and 70 having been identified in the soluble, insoluble, and both fractions, respectively (Supplementary Table S2.4).

2.2.3. Analysis of SMP-encoding transcripts

147 contigs encoded complete protein sequences, accounting for 71% of the 207 shell matrix proteins of *L. stagnalis*. In this study, a complete sequence refers to a gene model that has both start and stop codons. Of the remaining 60 sequences, 16, 31, and 13 had the 3' end missing, the 5' end missing, or internal sequences, respectively (Supplementary Table S2.4). The distribution of theoretical isoelectric points (pI)

estimated for all complete sequences of SMPs identified in this study indicated a bimodal pattern with acidic proteins being more numerous than basic ones (Fig. 2.2). The highest and lowest pIs were 10.90 and 3.65, respectively.

2.2.4. Similarity searches using BLAST

In order to find homologous sequences in the databases, the 207 SMPs identified in this study were searched against GenBank using BLASTP, and 165 proteins showed similarity to known proteins. Of the 165 proteins, 156 and 9 SMPs indicated high similarity to those molluscs and other invertebrates, and even to vertebrates. The remaining 42 SMPs are novel proteins, which are dissimilar from all known proteins (Supplementary Table S2.4).

2.2.5. Conserved domain search

Searches for conserved domains using SMART identified 261 domains in the 207 SMPs. Those domains were grouped into six categories: extracellular matrix (38 domains), enzyme (76), cation interaction (33), polysaccharide interaction (33), proteinase inhibitor (14), and others (67) (Fig. 2.3).

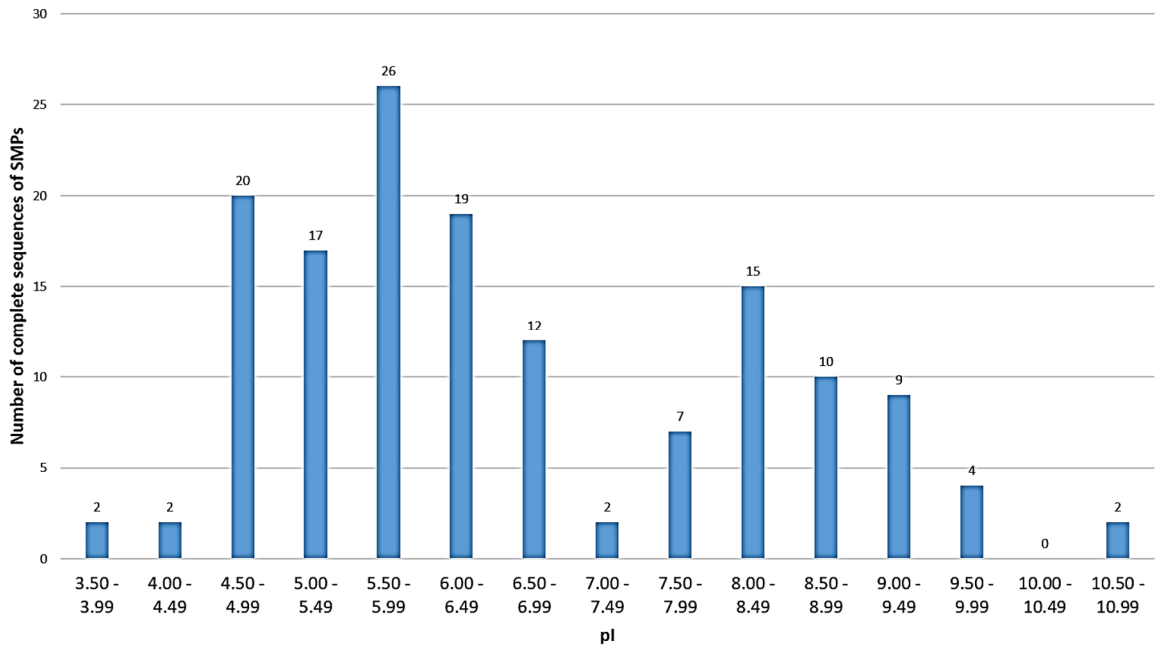

Figure 2.2.

Histogram showing the frequencies of theoretical pIs of SMPs identified in this study. Theoretical pI was estimated from the amino acid sequence translated using UniProtKB. The values of the highest and the lowest pIs were 10.90 and 3.65, respectively.

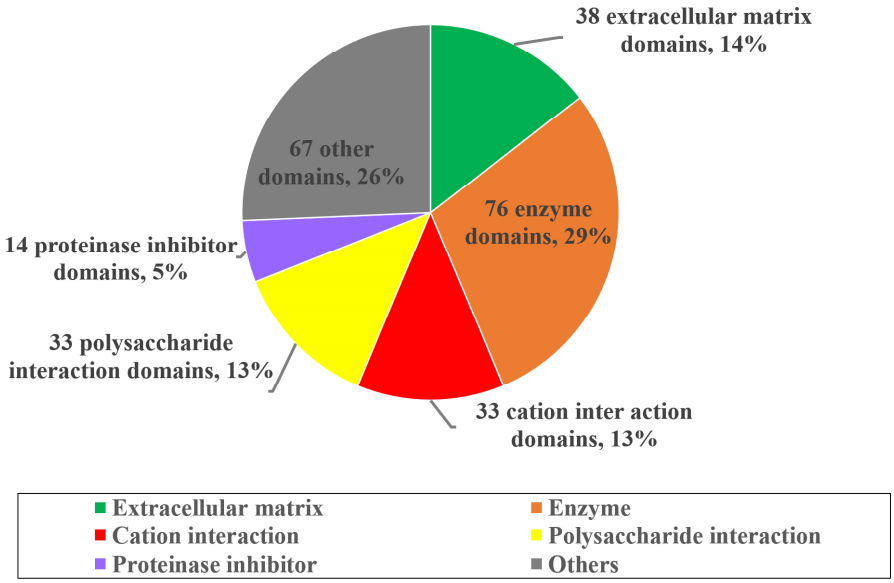

Figure 2.3.

Summary of domains identified from SMPs of *L. stagnalis*. Actual counts and frequencies of different kinds of domains observed among the 207 SMPs are shown.

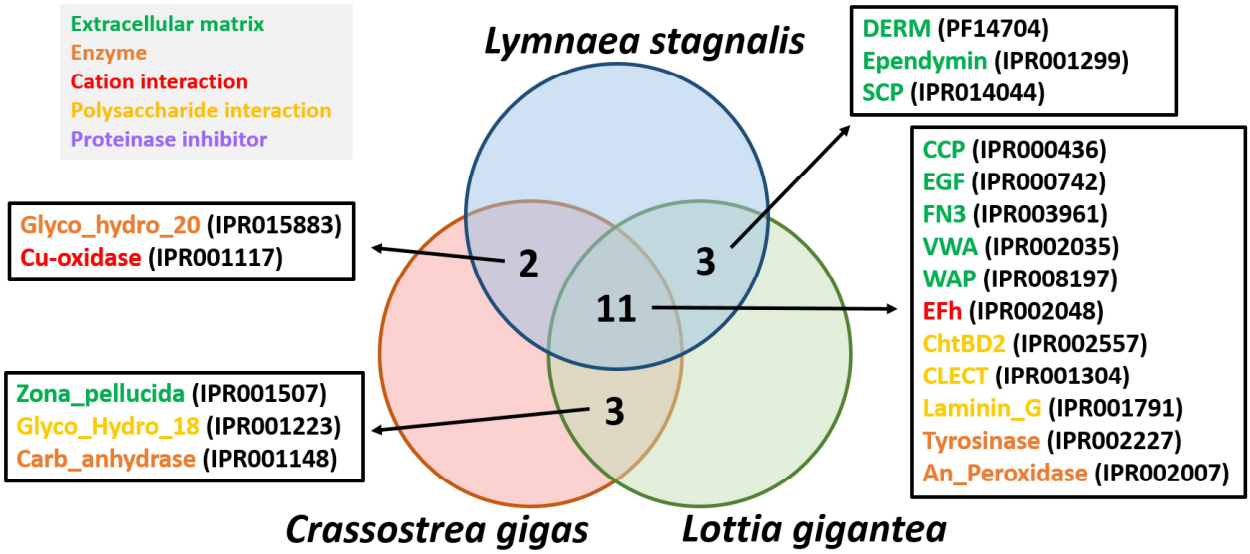

Figure 2.4.

Venn diagram showing shared conserved domains among SMPs of three mollusc species (*Lymnaea stagnalis*, *Crassostrea gigas*, and *Lottia gigantea*). Conserved domains are grouped into 5 categories. 11 domains (CCP, EGF, FN3, VWA, WAP, EFh, ChtBD2, CLECT, Laminin_G, Tyrosinase, and An_Peroxidase) are shared among all three species. 2 domains (Glyco_hydro_20, and Cu-oxidase) are shared between *L. stagnalis* and *C. gigas*, and 3 domains (Ependymin, DERM, and SCP) between *L. stagnalis* and *L. gigantea*.

115 proteins had signal peptides, including 99 complete amino acid sequences. Among the 115 proteins were 22 house-keeping, 17 room-keeping, 8 known SMPs, 37 uncharacterized proteins, and 31 novel proteins (Supplementary Table S2.4). The 99 SMPs with signal peptides and complete sequences included 17 house-keeping, 13 room-keeping, 7 SMPs, 36 uncharacterized proteins, and 26 novel proteins. Of the 92 proteins lacking signal peptides there were 48 complete and 44 partial sequences. Among them 48 house-keeping proteins, 16 room-keeping proteins, 5 SMPs, 12 uncharacterized proteins, and 11 novel proteins (Supplementary Table S2.4). The 48 complete SMP sequences without signal peptides included 29 house-keeping proteins, 8 room-keeping proteins, 3 SMPs, 5 uncharacterized proteins, and 3 novel proteins.

Eighty-one proteins contained one or more low complexity regions (LCRs). Of these, 14, 13, 6, 24, and 24 proteins were identified as house-keeping, room-keeping, SMP, uncharacterized protein, and novel proteins, respectively (Supplementary Table S2.4).

Comparisons of shared domains among the SMPs of three molluscs, *L. stagnalis*, *Lottia gigantea*, and *Crassostrea gigas*: (the latter two have decoded draft genomes (G. Zhang et al., 2012; Simakov et al., 2013)) indicated that 11 kinds of domains are shared among these species, including 5 extracellular regions, 3 polysaccharide interaction domains, 2 enzymes, and 1 cation interaction domain (Fig. 2.4).

2.2.6. Expression levels of SMP genes in mantle tissue

FPKMs for each of the 207 SMPs were calculated from transcriptomic data. Fig. 2.5 indicates the 10% of sequences with the highest FPKMs among the 207 SMPs (Fig. 2.5 and Supplementary Table S2.4). The sequence with the highest FPKM is Ls-SMP-126, annotated as adductor muscle actin by BLAST. Other sequences with top 10% expression levels are dominated by apparent house-keeping genes, including 60S acidic ribosomal P2 (Ls-SMP-173), actin (Ls-SMP-127), L-amino-acid oxidase (Ls-SMP-33), arginase-1-like isoform X2 (Ls-SMP-114), ATP-dependent RNA helicase DDX43 (Ls-SMP-175), voltage-dependent anion-selective channel protein 2-like isoform X4 (Ls-SMP-12), retrograde protein of 51 kDa (Ls-SMP-77), 14-3-3 protein epsilon (Ls-SMP-17), and YGSC-1 (Ls-SMP-176). Some proteins that had already been identified as SMPs, such as formin-like protein 2 isoform X3 (Ls-SMP-53), matrilin (Ls-SMP-90), extensin-like isoform X1 (Ls-SMP-52), and dermatopontin 1 (Ls-SMP-195) also exhibited high expression levels, and were included in the top 10% of sequences based on FPKMs. The remaining sequences in Fig. 2.5 include hemocyanin alpha D-subunit (Ls-SMP-170), a "room-keeping" gene, and an unidentified protein (Ls-SMP-172) or uncharacterized proteins including LOC106071610 (Ls-SMP-51) and LOC106053304 (Ls-SMP-85) (Fig. 2.5 and Supplementary Table S2.4).

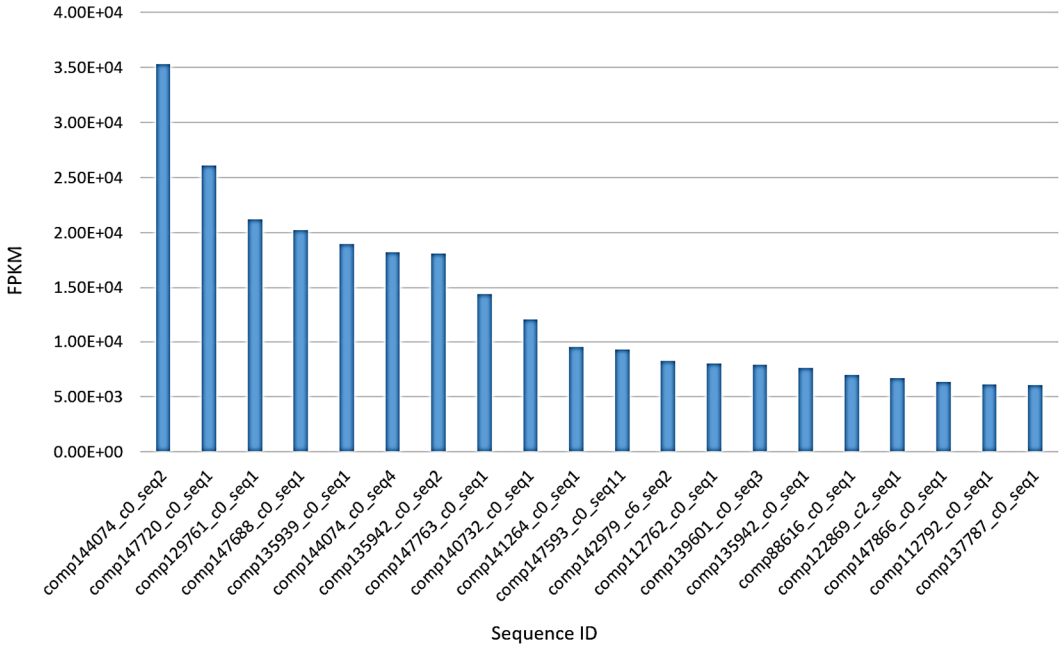

Figure 2.5

Top 10% of contigs for SMP-coding genes of *L. stagnalis* that yielded the highest FPKM values.

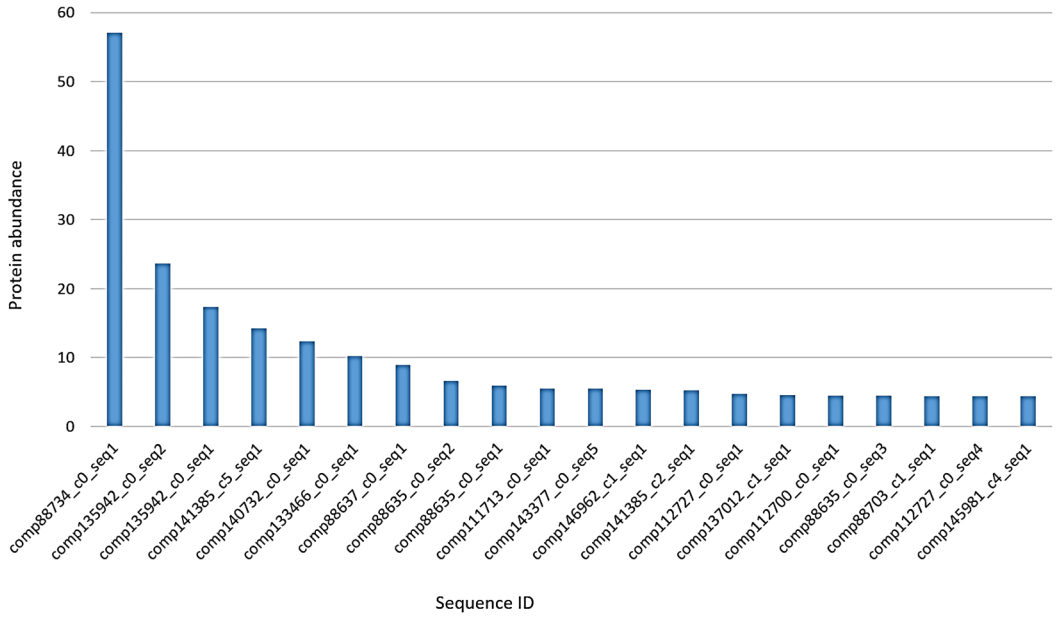

Figure 2.6

Top 10% of SMPs (shown in contig names) yielding the highest protein abundance values.

2.2.7. Abundances of SMPs in the shell

Protein abundance, or the amount of a protein in the shell matrix, was estimated from proteomic data. Fig. 2.6 shows the most abundant 10% of protein sequences among the 207 SMPs (Fig. 2.6 and Supplementary Table S2.4). The most abundant protein, Ls-SMP-203, could not be identified by BLAST or conserved domain search. Indeed, more than half of the most abundant 10% of shell proteins were either unidentified (Ls-SMP-45, Ls-SMP-4, Ls-SMP-120, Ls-SMP-161, Ls-SMP-96, Ls-SMP-7, Ls-SMP-57, Ls-SMP-5, Ls-SMP-10 and Ls-SMP-144) or uncharacterized [coiled-coil domain-containing 1-like isoform X2 (Ls-SMP-198), LOC106053304 (Ls-SMP-85), LOC106070421 (Ls-SMP-200), LOC106069873 (Ls-SMP-197), and LOC106073248 (Ls-SMP-202)] proteins. The remaining proteins had already been identified as either SMPs [formin-like protein 2 isoform X3 (Ls-SMP-53), extensin-like isoform X1 (Ls-SMP-52), and galaxin (Ls-SMP-97)], or an otolith matrix protein [starmaker-like isoform X1 (Ls-SMP-199)].

2.2.8. SDS-PAGE analysis

Matrix proteins were extracted from shells of *Lymnaea stagnalis*, which yielded 0.53 μg soluble and 1.39 μg insoluble fractions per g of shell. Extracted proteins separated by SDS-PAGE revealed 3 major (10 kDa, 20 kDa, and 22 kDa) and 2 minor (28 kDa and 220 kDa) bands for the soluble fraction and one major (20 kDa) and 3 minor (28 kDa, 32 kDa, and 220 kDa) bands for the insoluble fraction, when stained with Coomassie Brilliant Blue (Fig. 2.7). As shown later, sizes of those proteins do not match those of proteins with the highest numbers of hits in the proteomic analysis, suggesting that the major proteins seen in SDS-PAGE are products of post-translational modifications such as cleavage, glycosylation, and phosphorylation.

2.2.9. Gene ontology analysis

Gene ontologies (GO) of the mantle and SMP-encoding transcriptomes were analyzed and visualized with Blast2GO. Annotated genes were classified into three different categories, “cellular component”, “biological process”, and “molecular function”, with three different levels for each category (Figs. 2.8 - 2.13). In the categories, “cellular component” and “biological process”, no notable differences were observed between the mantle and SMP transcriptomes, except that at level 2 of “cellular component”, the term “extracellular region” is enriched in SMPs compared to the mantle (Figs. 2.8, 2.9, 2.11, and 2.12). On the other hand, some differences were seen in the category “molecular function.” Especially at level 3, the GO terms “hydrolase activity”, “protein binding”, “oxidoreductase activity”, and “carbohydrate derivative-binding” are enriched in SMPs (Figs. 2.10 and 2.13). At level 4, the GO terms “cation binding” and “anion binding” are enriched in SMPs compared to the mantle (Supplementary Figs. 2.10 and 2.13).

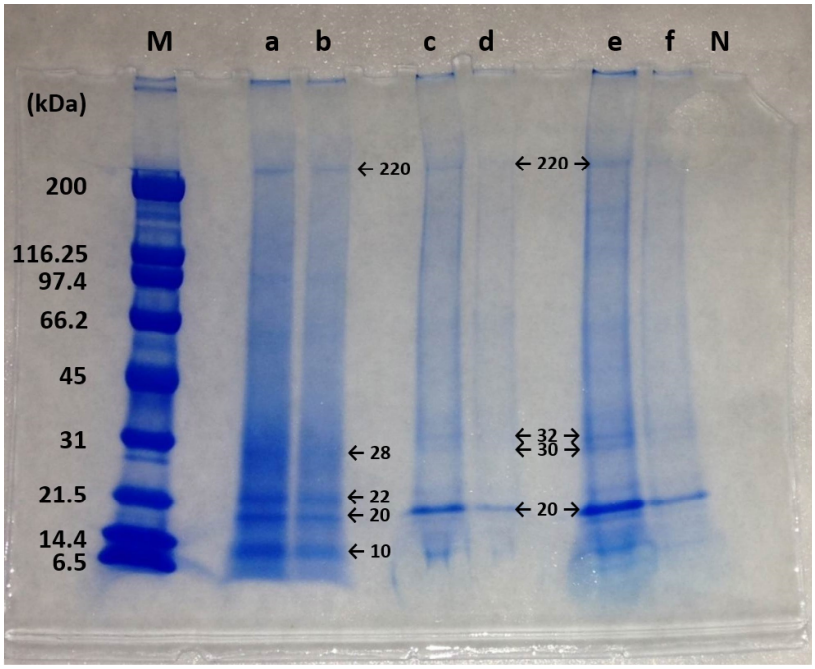

Figure 2.7.

SDS-PAGE analysis of soluble and insoluble fractions extracted from shells of *L. stagnalis*. M: Marker, a: soluble fraction (904 µg/mL), b: soluble fraction (452 µg/mL), c: insoluble fraction (40 µg/mL), d: insoluble fraction (10 µg/mL).

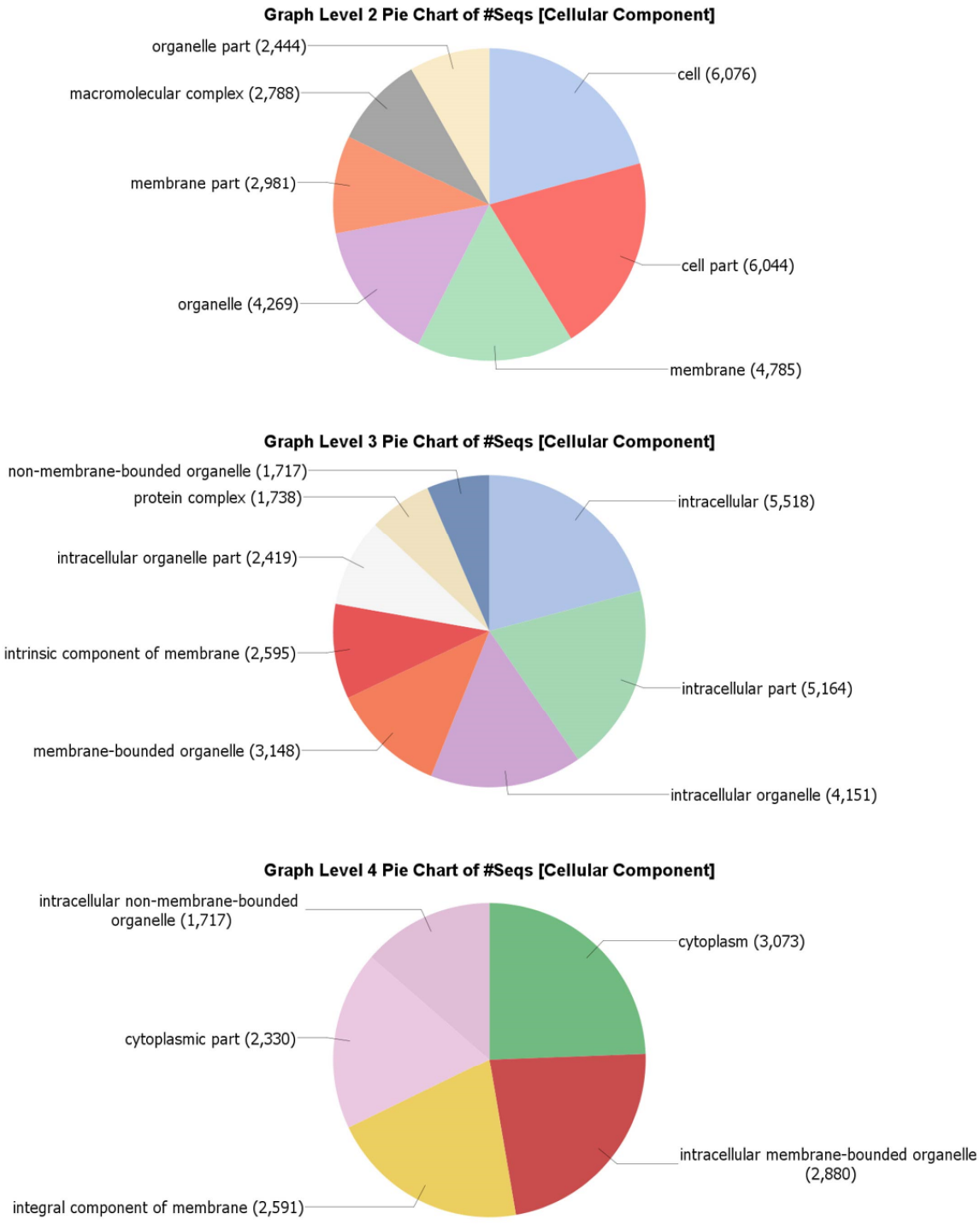

Figure 2.8.

Combined graphs for GO of mantle tissues (cellular component) produced by Blast2GO.

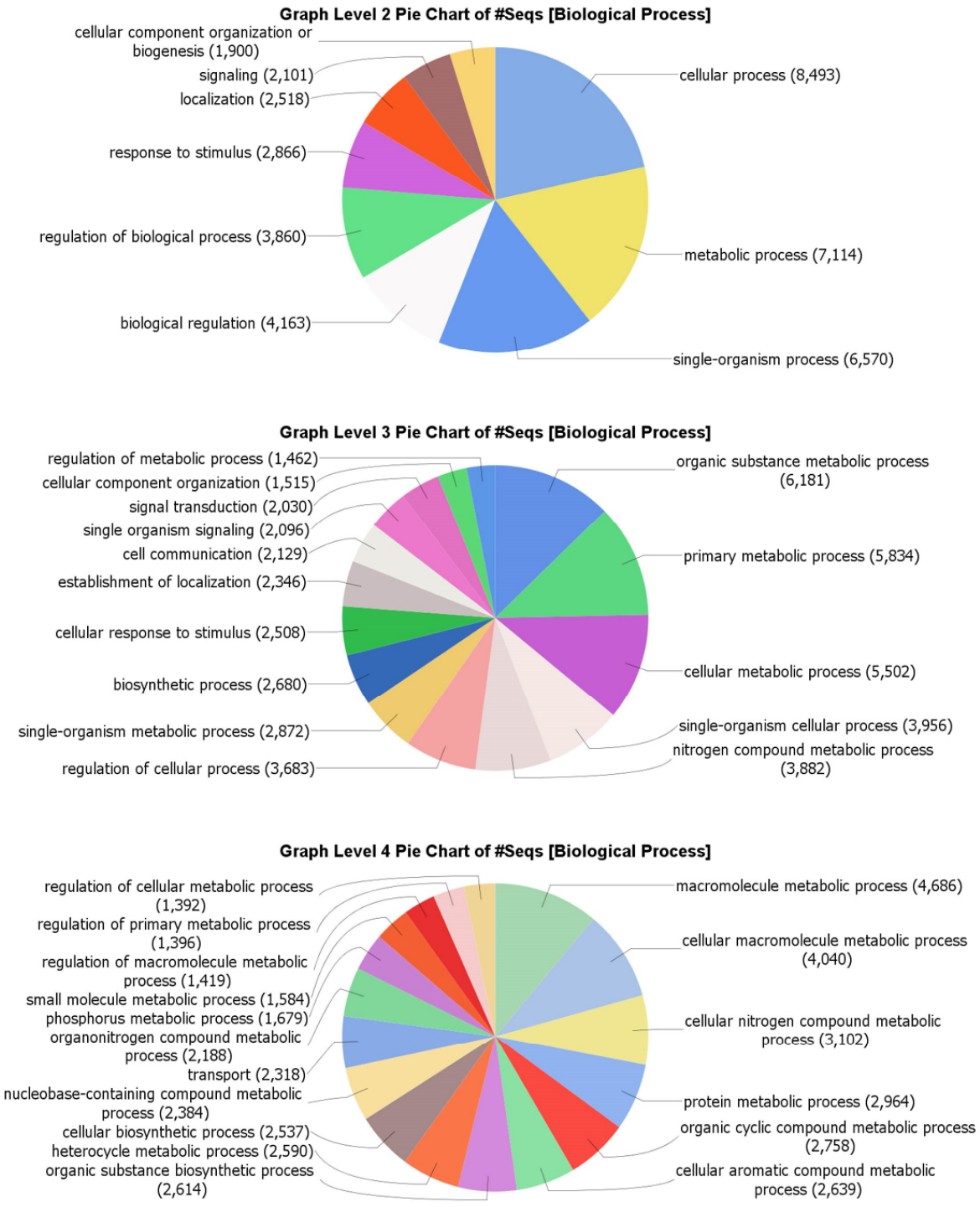

Figure 2.9.

Combined graphs for GO of mantle tissues (biological process) produced by Blast2GO.

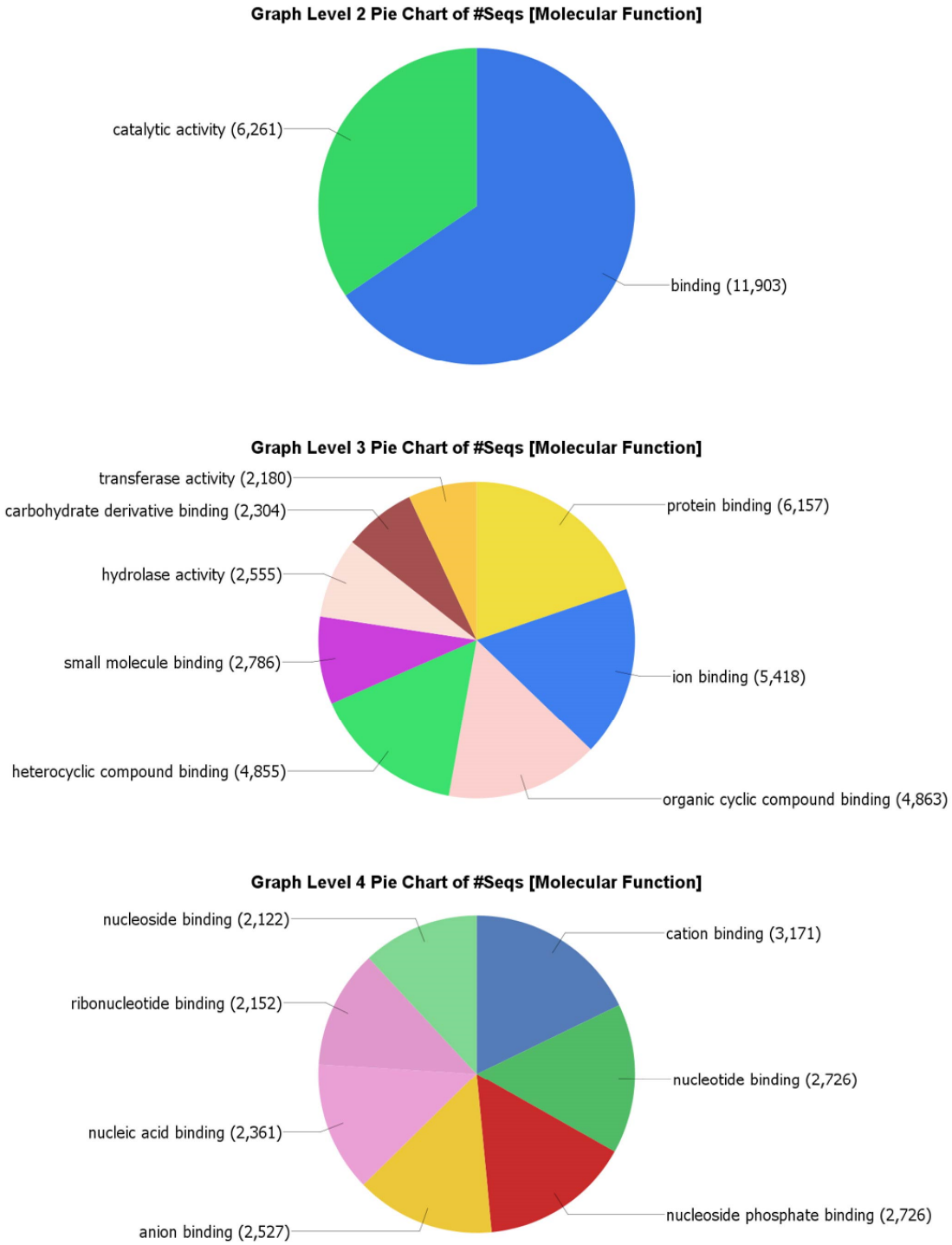

Figure 2.10.

Combined graphs for GO of mantle tissues (molecular function) produced by Blast2GO.

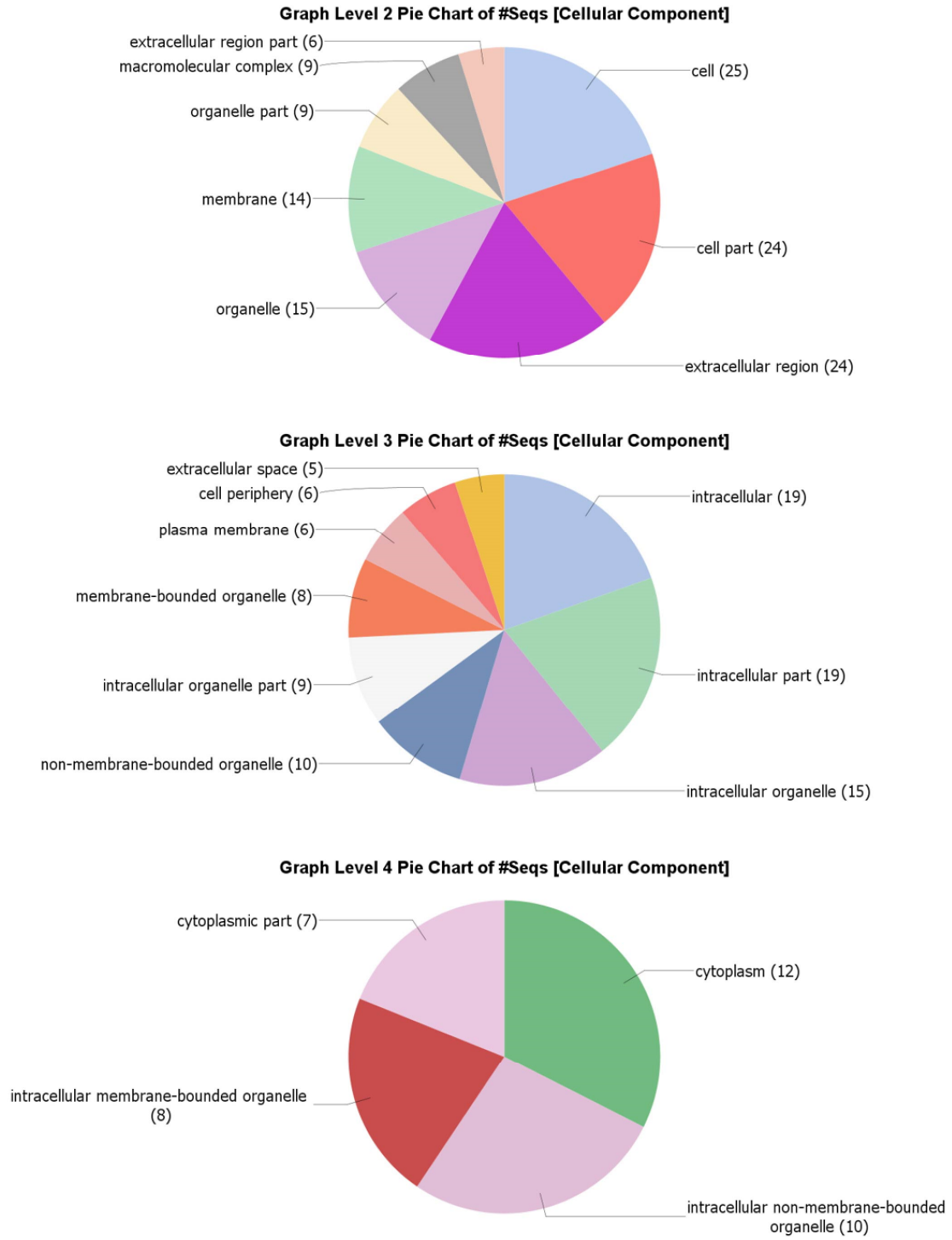

Figure 2.11.

Combined graphs for GO of SMPs (cellular component) produced by Blast2GO.

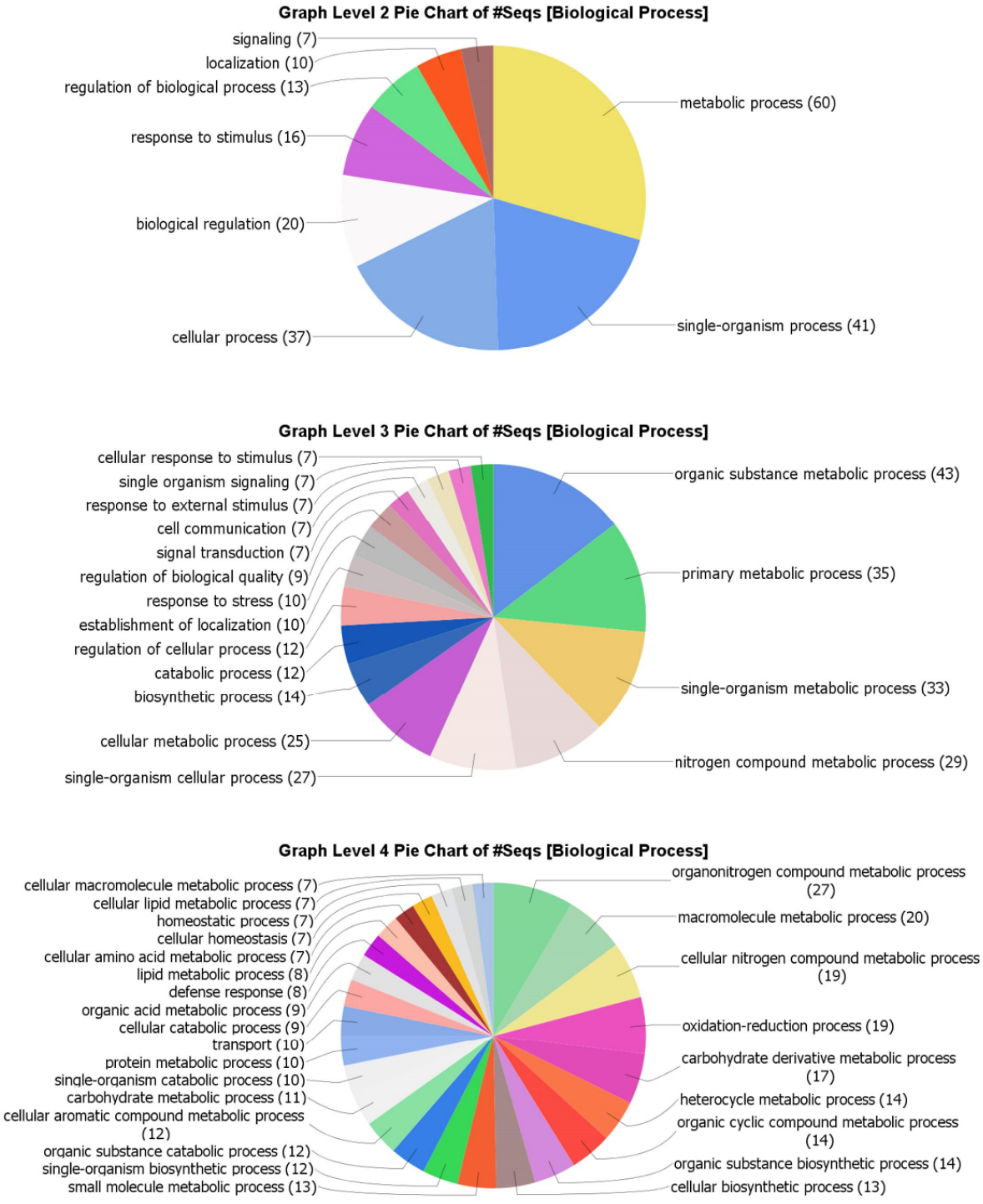

Figure 2.12.

Combined graphs for GO of SMPs (biological process) produced by Blast2GO.

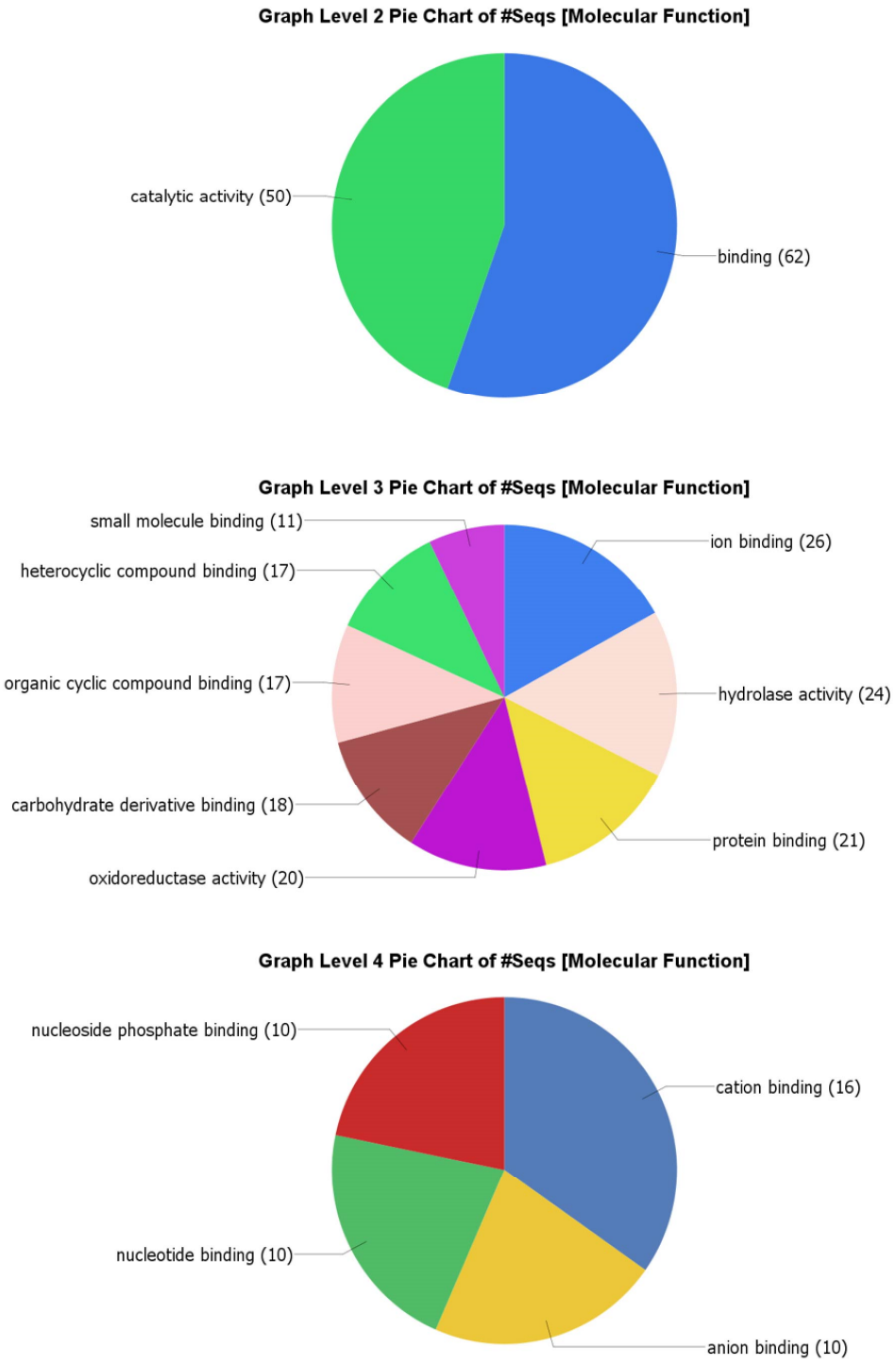

Figure 2.13.

Combined graphs for GO of SMPs (molecular function) produced by Blast2GO.

2.2.10. Phylogenetic analyses of some conserved domains

Pif proteins in general and the Pif-like protein (Ls-SMP-88) of *L. stagnalis* have two ChtBD2 domains arranged in tandem. Phylogenetic analyses of those ChtBD2 domains indicated that each forms a cluster (with bootstrap probabilities for the upstream and downstream domains being 64% and 68%, respectively), suggesting that the last common ancestor of bivalves and gastropods already had those two ChtBD2 domains in tandem (Fig. 2.14a). Phylogenetic analyses also indicated that the Laminin_G domain sequence of the Pif-like SMP of *L. stagnalis* forms a cluster with sequences of typical Pif SMPs of other molluscs, rather than with Laminin_G sequences of BMSP (blue mussel shell protein) (Fig. 2.14b).

Phylogenetic analyses of tyrosinase domains clearly discriminated three classes of TDC-SMPs: lophotrochozoan tyrosinase, vertebrate tyrosinase, and hemocyanin-tyrosinase (Fig. 2.15). Thus, TDC-SMPs of *L. stagnalis* may be classified into two groups of different origin, i.e., tyrosinase SMP and hemocyanin-tyrosinase, with the latter having greatly expanded in *L. stagnalis* (Fig. 2.15).

2.2.11. Comparison of SMP expression levels between the right and left sides of mantle tissues

The right and left sides of the mantle tissues of *L. stagnalis* should have different rates of shell growth to produce the asymmetric dextral shell (Fig. 1). Levels of shell matrix protein gene expression have been compared between the right and left sides of the mantle in three individuals, to identify functionally important proteins (Supplementary Tables S2.5 and S2.6). There are indeed differences in the patterns of gene expression between the left and right sides of the mantle. Of the 35,951 transcript sequences identified from the mantle, 916 transcripts (2.6%) indicated a statistically significant difference between the left and right sides, considering variations in gene expression levels among the three individuals studied ($q < 0.05$; Fig. 2.16, Supplementary Table S2.6). Of those 916 transcripts, 32 transcripts encode SMPs (15.5% of the 207 SMPs identified in this study). Among the 916 transcripts, 612 encoding proteins other than SMPs (64.4%), and 25 encoding SMPs (78.1% of 32) showed higher expression on the left side than on the right (Fig. 2.17). Among the 32 SMPs, 25 showed higher expression on the left than the right, with Ls-SMP-30 showing the greatest difference in expression level. In contrast, 7 SMPs showed higher expression on the right side than the left, with Ls-SMP-43 being the most distinctively different (Figs. 2.1, 2.16 and Supplementary Table S2.4).

The mantle transcriptome was compared to the foot transcriptome using local BLAST to find genes that are expressed in the mantle, but not in the foot (E value cut off $>10^{-10}$) (genes hereafter termed “mantle specific”). Of the 207 SMP genes, 29 were mantle

specific (Fig. 2.17 and Supplementary Table S2.7). 14 of the 29 mantle-specific SMP genes were significantly more highly expressed on the left side than the right. Those 14 genes encode ovipostatin (Ls-SMP-150), matrilin-like (Ls-SMP-48), cysteine-rich venom protein LO1-like protein (Ls-SMP-108), three uncharacterized proteins (Ls-SMP-30, Ls-SMP-63, and Ls-SMP-192), and 8 novel proteins (Ls-SMP-4, Ls-SMP-39, Ls-SMP-59, Ls-SMP-60 Ls-SMP-61, Ls-SMP-62, Ls-SMP-70, and Ls-SMP-206). The remaining 15 genes did not show asymmetric expression patterns, and include 5 SMP genes containing conserved domains found in other molluscs (Ls-SMP-23, Ls-SMP-24, Ls-SMP-81, Ls-SMP-82, and Ls-SMP-186) (Supplementary Table S2.4).

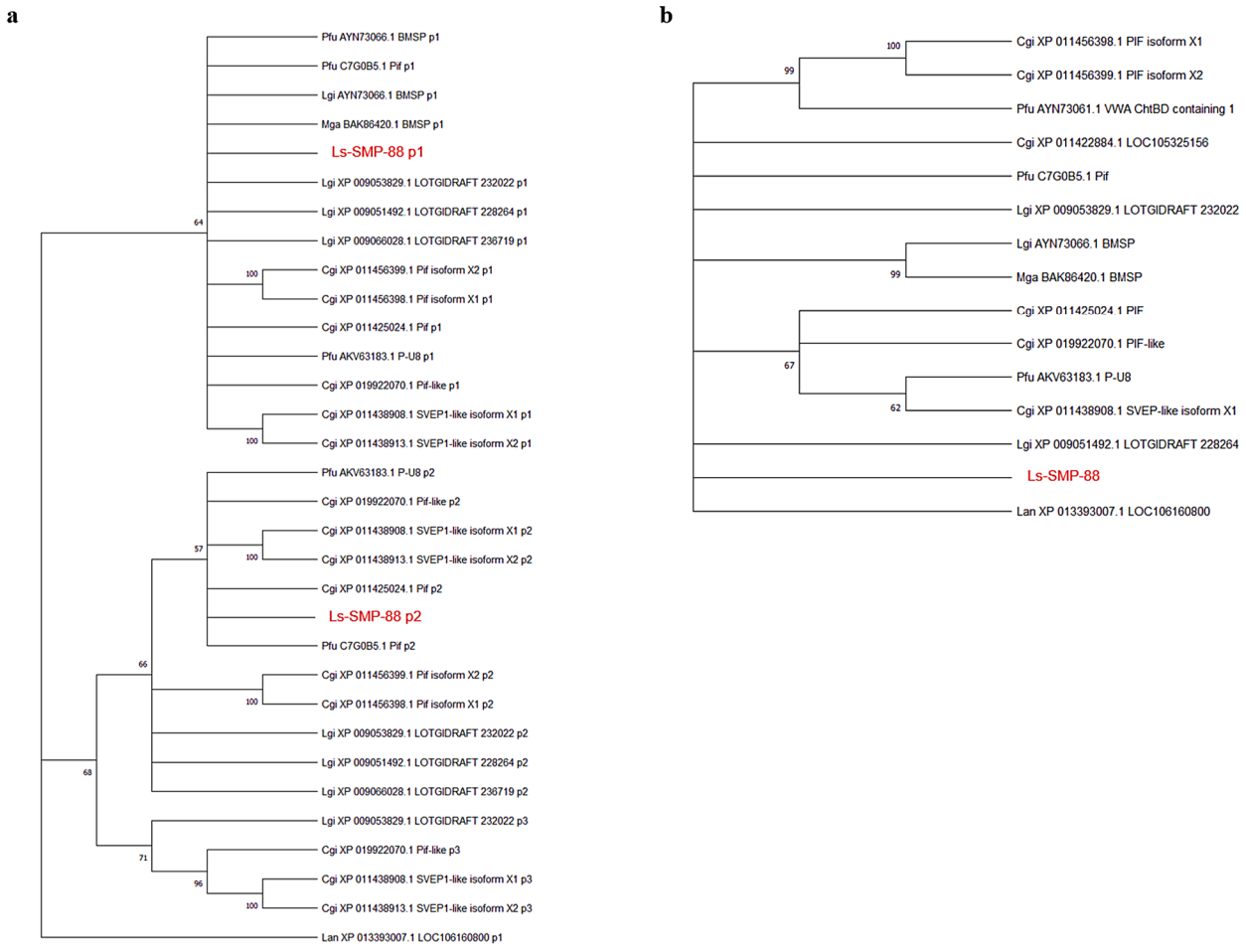

Figure 2.14.

Figure 2.14 (legend).

(a) Maximum likelihood tree of ChtBD2 domains in molluscan SMPs, including the one from *Lymnaea stagnalis*. Molluscan ChtBD2 domain sequences were retrieved from NCBI GenBank after BLAST searches using the two ChtBD2 domain sequences of Ls-SMP-88 of *L. stagnalis* as query. The ML tree was inferred from two domains of Ls-SMP-88 and 29 domains of 14 proteins, using the WAG + G + I model based on 47 amino acids. When more than one ChtBD2 domain exists in a protein, they are discriminated in order from the N-terminus, and named p1, p2, etc. (b) Maximum likelihood tree of Laminin_G domains in molluscan SMPs, including the one identified from *Lymnaea stagnalis*. Molluscan Laminin_G domain sequences were retrieved from NCBI GenBank after BLAST searches using the Laminin_G domain sequence of Ls-SMP-88 of *L. stagnalis* as a query. The ML tree was inferred from a domain of Ls-SMP-88 and 13 domains of 13 proteins, using the LG + G model, based on 128 amino acids. (a, b) Polychotomy results if the bootstrap value of the node is lower than 50%. Bootstrap values are indicated for nodes with a value greater >50%. A domain sequence of the chitin-binding domain-containing protein of the brachiopod, *Lingula anatina*, was included as outgroup. The sequence name in red indicates the sequence from *L. stagnalis*. Cgi: *Crassostrea gigas*, Lan: *Lingula anatina*, Lgi: *Lottia gigantea*, Mga: *Mytilus galloprovincialis*, Pfu: *Pinctada fucata*.

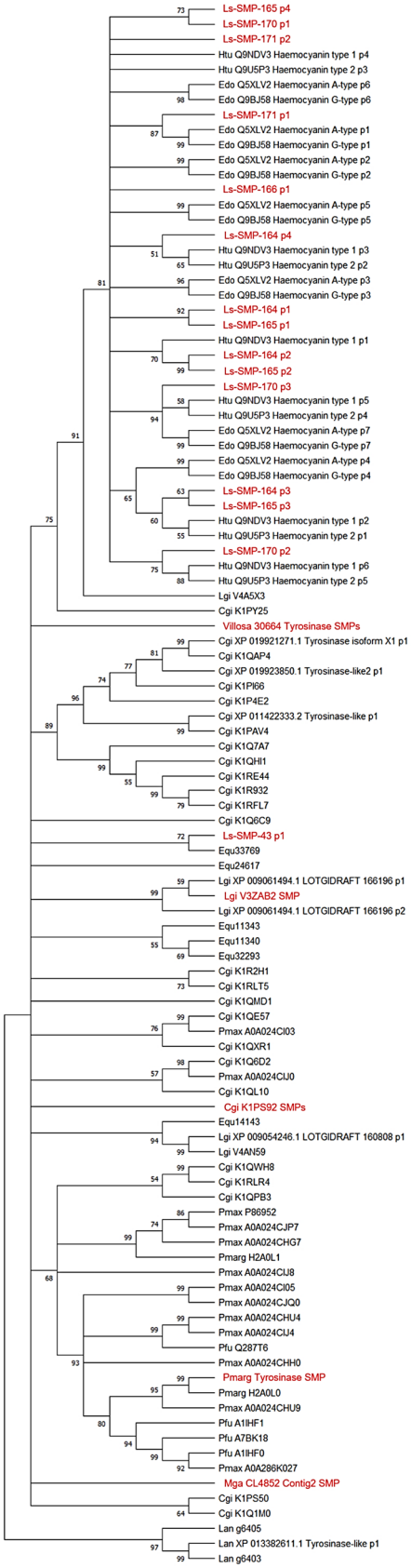
Ls-SMP-165 p4
Ls-SMP-170 p1
Ls-SMP-171 p2
Htu Q9NDV3 Haemocyanin type 1 p4
Htu Q9U5P3 Haemocyanin type 2 p3
Edo Q5XLV2 Haemocyanin A-type p6
Edo Q9BJ58 Haemocyanin G-type p6
Ls-SMP-171 p1
Edo Q5XLV2 Haemocyanin A-type p1
Edo Q9BJ58 Haemocyanin G-type p1
Edo Q5XLV2 Haemocyanin A-type p2
Edo Q9BJ58 Haemocyanin G-type p2
Ls-SMP-166 p1
Edo Q5XLV2 Haemocyanin A-type p5
Edo Q9BJ58 Haemocyanin G-type p5
Ls-SMP-164 p4
Htu Q9NDV3 Haemocyanin type 1 p3
Htu Q9U5P3 Haemocyanin type 2 p2
Edo Q5XLV2 Haemocyanin A-type p3
Edo Q9BJ58 Haemocyanin G-type p3
Ls-SMP-164 p1
Ls-SMP-165 p1
Htu Q9NDV3 Haemocyanin type 1 p1
Ls-SMP-164 p2
Ls-SMP-165 p2
Ls-SMP-170 p3
Htu Q9NDV3 Haemocyanin type 1 p5
Htu Q9U5P3 Haemocyanin type 2 p4
Edo Q5XLV2 Haemocyanin A-type p7
Edo Q9BJ58 Haemocyanin G-type p7
Edo Q5XLV2 Haemocyanin A-type p4
Edo Q9BJ58 Haemocyanin G-type p4
Ls-SMP-164 p3
Ls-SMP-165 p3
Htu Q9NDV3 Haemocyanin type 1 p2
Htu Q9U5P3 Haemocyanin type 2 p1
Ls-SMP-170 p2
Htu Q9NDV3 Haemocyanin type 1 p6
Htu Q9U5P3 Haemocyanin type 2 p5
Lgi V4A5X3
Cgi K1PY25
Villosa 30664 Tyrosinase SMPs
Cgi XP 019921271.1 Tyrosinase isoform X1 p1
Cgi K1QAP4
Cgi XP 019923850.1 Tyrosinase-like2 p1
Cgi K1PI66
Cgi K1P4E2
Cgi XP 011422333.2 Tyrosinase-like p1
Cgi K1PAV4
Cgi K1Q7A7
Cgi K1QHI1
Cgi K1RE44
Cgi K1R932
Cgi K1RFL7
Cgi K1Q6C9
Ls-SMP-43 p1
Equ33769
Equ24617
Lgi XP 009061494.1 LOTGIDRAFT 166196 p1
Lgi V3ZAB2 SMP
Lgi XP 009061494.1 LOTGIDRAFT 166196 p2
Equ11343
Equ11340
Equ32293
Cgi K1R2H1
Cgi K1RLT5
Cgi K1QMD1
Cgi K1QE57
Pmax A0A024CI03
Cgi K1QXR1
Cgi K1Q6D2
Pmax A0A024CIJ0
Cgi K1QL10
Cgi K1PS92 SMPs
Equ14143
Lgi XP 009054246.1 LOTGIDRAFT 160808 p1
Lgi V4AN59
Cgi K1QWH8
Cgi K1RLR4
Cgi K1QPB3
Pmax P86952
Pmax A0A024CJP7
Pmax A0A024CHG7
Pmarg H2A0L1
Pmax A0A024CIJ8
Pmax A0A024CI05
Pmax A0A024CJQ0
Pmax A0A024CHU4
Pmax A0A024CIJ4
Pfu Q287T6
Pmax A0A024CHH0
Pmarg Tyrosinase SMP
Pmarg H2A0L0
Pmax A0A024CHU9
Pfu A1IHF1
Pfu A7BK18
Pfu A1IHF0
Pmax A0A286K027
Mga CL4852 Contig2 SMP
Cgi K1PS50
Cgi K1Q1M0
Lan g6405
Lan XP 013382611.1 Tyrosinase-like p1
Lan g6403

Figure 2.15.

Figure 2.15 (legend).

Maximum likelihood tree of the tyrosinase domains contained in some animal tyrosinase proteins and molluscan SMPs, including the one from *Lymnaea stagnalis*. The ML tree was inferred from 15 domains in 6 SMPs of *L. stagnalis*, 25 domains in two molluscan haemocyanin proteins from NCBI GenBank and 63 domains in tyrosinase proteins identified by Shimizu et al. (2019) (Shimizu et al., 2019), using the LG + G model based on 82 amino acids. Polychotomy results if the bootstrap value of the node is lower than 50%. Bootstrap values are indicated for nodes with a value greater >50%. Sequences of tyrosinase domains of the brachiopod, *Lingula anatina*, were included as an outgroup. Sequence names in red indicate proteins that have been identified as SMPs in this study or by Shimizu et al. (2019) (Shimizu et al., 2019). Cgi: *Crassostrea gigas*, Equ: *Euhadra quaesita*, Lan: *Lingula anatina*, Lgi: *Lottia gigantea*, Mga: *Mytilus galloprovincialis*, Pmarg: *Pinctada margaritifera*, Pmax: *Pinctada. maxima*, Pfu: *Pinctada fucata*. When more than one tyrosinase domain exists in a protein, they are listed in order from the N-terminus, and named p1, p2, etc.

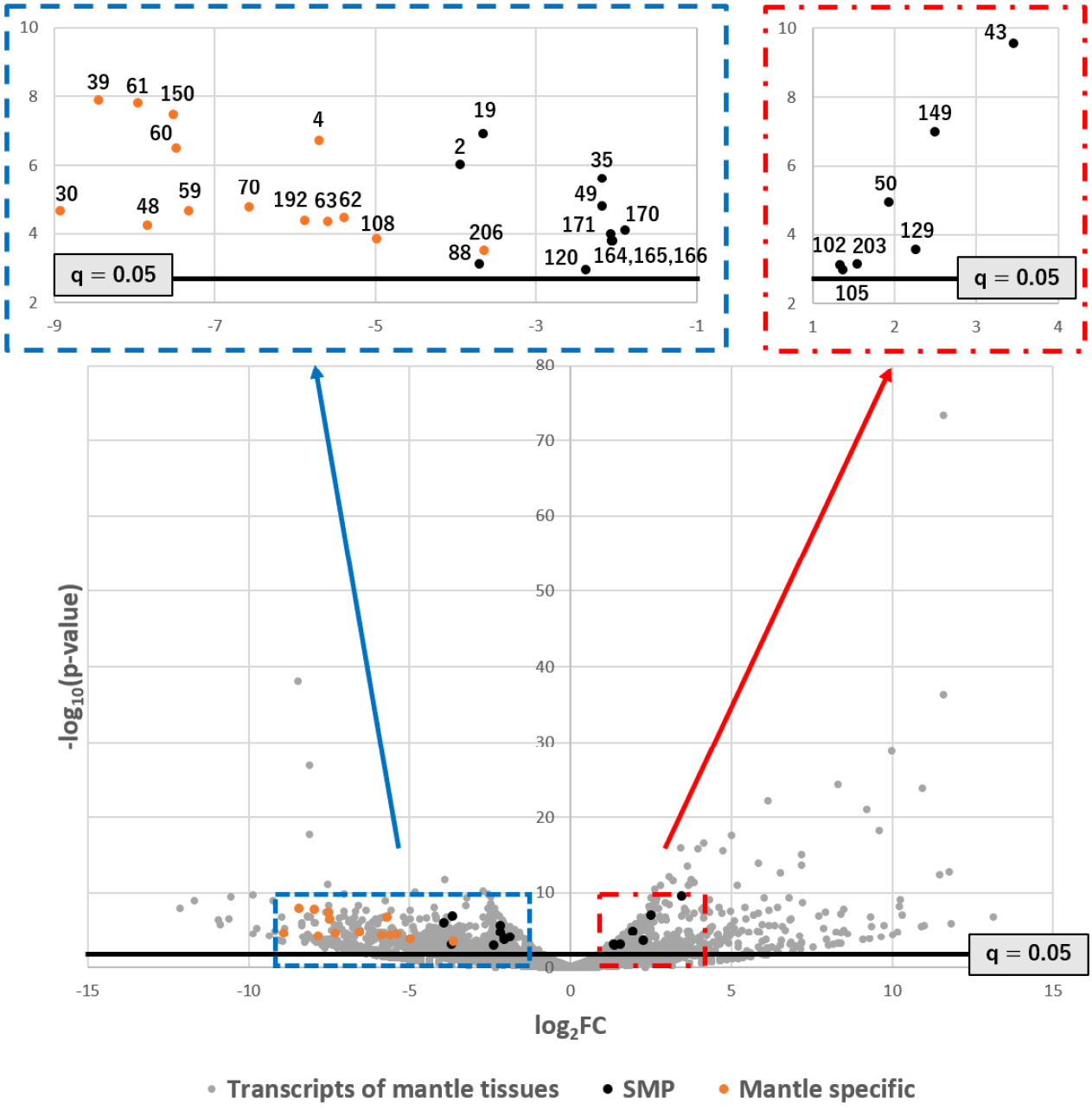

Figure 2.16.

Volcano plot showing differential expression of SMP-coding genes between right and left sides of mantle tissues of the dextral strain. The X axis represents the logarithm of the change in expression levels of the right side vs. the left side. The Y axis represents the logarithm of the significance level for each comparison of the gene. The level of significance to reject the null hypothesis (q = 0.05) is shown as a black line. Numbers denote serial numbers of contigs (genes) in Supplementary Table S2.4. Only genes that were considered significant or expressed specifically in the mantle (shown in orange dots) are numbered.

From Figure 2.16 it is evident that (1) 14 of 32 SMP transcripts are specifically expressed in the mantle, without detectable expression in the foot, (2) all mantle-specific SMP transcripts are more strongly expressed on the left side than the right, and (3) those 14 transcripts dominate the left side of Figure 2.16. The 14 SMPs include those that are homologous to the three room-keeping proteins, ovipostatin (Ls-SMP150), matrilin-like (Ls-SMP-48), and cysteine-rich venom protein LIO1-like protein (Ls-SMP-108), three uncharacterized proteins (Ls-SMP-30, Ls-SMP-192, and Ls-SMP-63), and 8 novel proteins (Ls-SMP-39, Ls-SMP-61, Ls-SMP-4, Ls-SMP-60, Ls-SMP-70, Ls-SMP-59, Ls-SMP-62, and Ls-SMP-206) (Fig. 2.17). An ELH domain is contained in the novel protein Ls-SMP-4, two EFh domains are present in each of the novel proteins Ls-SMP-61 and Ls-SMP-62, and a CLECT domain and an H_lectin domain are found in the novel protein Ls-SMP-59 and the uncharacterized protein Ls-SMP-63 (Fig. 2.17). One or more low complexity regions are seen in Ls-SMP-150, Ls-SMP-60, Ls-SMP-63, and Ls-SMP-108 (Fig. 2.17). Amino acid composition analyses of low complexity regions indicate that the low complexity regions in Ls-SMP-60, Ls-SMP-63, Ls-SMP-108, and Ls-SMP-150 are rich in acidic residues (58.3%, 33.3%, 39.3%, and 28.6% acidic residues, respectively) (Supplementary Table S2.8). Signal peptides have been identified in most of those SMPs (10/14). Potential N-glycosylation (11/14), O-glycosylation (8/14), or phosphorylation (14/14) sites have also been inferred for those proteins (Supplementary Table S2.7). Although the actual functions of those SMPs must yet be confirmed by *in vivo* functional analysis, those 14 SMPs appear to be important in shell biomineralization, especially in suppressing shell precipitation, because they are specifically expressed in the mantle and they are more highly expressed on the left than the right side.

The other 18 SMP-encoding transcripts are expressed in both mantle and foot; thus, they are not specific to the mantle. However, since they show a significant difference in expression levels between the left and right sides of the mantle, they are probably important in shell formation. Of the 18 SMP-encoding transcripts, 11 are more highly expressed on the left than the right. They include a chitin-binding domain (ChtBD2)-containing protein (Ls-SMP-88), proteins homologous to two house-keeping proteins, alkaline phosphatase (Ls-SMP-35) and pancreatic triacylglycerol lipase-like protein (Ls-SMP-49), 5 SMPs homologous to the room-keeping protein, hemocyanin (Ls-SMP-170, Ls-SMP-171, Ls-SMP-164, Ls-SMP-165, and Ls-SMP-166), and 3 novel proteins (Ls-SMP19, Ls-SMP-2, and Ls-SMP-120). These three novel proteins all have one or more low-complexity regions, including one rich in acidic residues (42.9% of Asp: Ls-SMP-120) (Supplementary Table S2.8).

Sequence ID	Contig Name	Result of comparison of expression levels in transcriptomic analysis	Result of qPCR	FPKM	Protein abundance	Mantle specific	Presence of signal peptide	Presence of low complexity region	Gene type	Conserved domain	Sequence similarity
Ls-SMP-2	comp111558_c0_seq1	Left mantle	Left mantle	6.5E+02	2.00	No	Yes	Yes			
Ls-SMP-4	comp111713_c0_seq1	Left mantle	No difference	2.2E+02	5.43	Yes	Yes	No		ELH (IPR003424)	
Ls-SMP-19	comp123495_c0_seq1	Left mantle	Left mantle	7.6E+02	0.75	No	5' end missing	Yes			
Ls-SMP-30	comp128349_c0_seq1	Left mantle	No amplification	2.9E+01	0.41	Yes	Yes	No			PREDICTED: uncharacterized protein LOC106054306 [*Biomphalaria glabrata*]
Ls-SMP-35	comp130259_c0_seq1	Left mantle	No difference	2.9E+02	0.49	No	Yes	No	house-keeping	alkPPc (IPR001952)	PREDICTED: alkaline phosphatase, tissue-nonspecific isozyme-like [*Aplysia californica*]
Ls-SMP-39	comp132293_c0_seq2	Left mantle	No amplification	1.1E+02	0.94	Yes	No	No			
Ls-SMP-43	comp132880_c1_seq1	Right mantle	No difference	2.8E+02	0.15	No	Yes	Yes		Tyrosinase (PF00264)	PREDICTED: uncharacterized protein LOC106066173 [*Biomphalaria glabrata*]
Ls-SMP-48	comp134060_c0_seq1	Left mantle	No difference	5.4E+00	0.76	Yes	Yes	No	room-keeping	VWA (IPR002035)	matrilin [*Biomphalaria glabrata*]
Ls-SMP-49	comp134361_c0_seq1	Left mantle	No difference	1.8E+02	0.29	No	5' end missing	No	house-keeping	Lipase (PF00151)	PREDICTED: pancreatic triacylglycerol lipase-like [*Aplysia californica*]
Ls-SMP-50	comp135841_c0_seq2	Right mantle	No difference	3.5E+03	1.00	No	No	Yes			PREDICTED: uncharacterized protein LOC106075959 [*Biomphalaria glabrata*]
Ls-SMP-59	comp137266_c0_seq1	Left mantle	No difference	2.6E+01	1.31	Yes	Yes	No		CLECT (IPR001304)	
Ls-SMP-60	comp137504_c0_seq2	Left mantle	No amplification	1.2E+02	0.78	Yes	Yes	Yes			
Ls-SMP-61	comp137504_c0_seq3	Left mantle	No difference	3.0E+02	1.73	Yes	Yes	No		EFh (IPR011992) * 2	
Ls-SMP-62	comp137606_c0_seq3	Left mantle	No difference	4.0E+01	0.36	Yes	Yes	No		EFh (IPR011992) * 2	
Ls-SMP-63	comp137606_c0_seq4	Left mantle	No difference	2.2E+01	1.08	Yes	5' end missing	Yes		H_lectin (IPR019019)	PREDICTED: uncharacterized protein LOC101863352 [*Aplysia californica*]
Ls-SMP-70	comp139037_c0_seq1	Left mantle	No amplification	2.4E+01	1.56	Yes	5' end missing	No			
Ls-SMP-88	comp141104_c0_seq1	Left mantle	Left mantle	1.5E+01	1.41	No	Yes	Yes	known SMP	ChtBD2 (IPR002557) * 2	PREDICTED: protein PIF-like [*Aplysia californica*]
Ls-SMP-102	comp141907_c1_seq10	Right mantle	No difference	1.7E+02	0.20	No	5' end missing	No	house-keeping	Cu-oxidase_2 (PF07731), Cu-oxidase_3 (PF07732)	PREDICTED: hephaestin-like protein 1, partial [*Biomphalaria glabrata*]
Ls-SMP-105	comp141907_c1_seq7	Right mantle	No difference	3.3E+01	0.28	No	5' end missing	No	house-keeping	Cu-oxidase_2 (PF07731), Cu-oxidase_3 (PF07732)	PREDICTED: hephaestin-like protein 1, partial [*Biomphalaria glabrata*]
Ls-SMP-108	comp142576_c3_seq8	Left mantle	No difference	1.9E+01	1.42	Yes	5' end missing	Yes	room-keeping	ShKT (IPR003582) * 3	PREDICTED: cysteine-rich venom protein LIO1-like [*Biomphalaria glabrata*]
Ls-SMP-120	comp143377_c0_seq5	Left mantle	No difference	4.7E+01	5.43	No	Yes	Yes			
Ls-SMP-129	comp144225_c2_seq13	Right mantle	No difference	1.9E+02	2.73	No	5' end missing	Yes	house-keeping		PREDICTED: bromodomain-containing protein DDB_G0280777-like [*Biomphalaria glabrata*]
Ls-SMP-149	comp146497_c0_seq3	Right mantle	Right mantle	8.8E+02	0.93	No	Yes	No	room-keeping	EGF (IPR000742), WAP (IPR008197) * 2	PREDICTED: neurogenic locus notch homolog protein 1-like isoform X1 [*Biomphalaria glabrata*]
Ls-SMP-150	comp146552_c3_seq2	Left mantle	No amplification	2.2E+02	3.06	Yes	Yes	Yes	room-keeping		ovipostatin [*Lymnaea stagnalis*]
Ls-SMP-164	comp147522_c0_seq1	Left mantle	No difference	3.9E+03	0.39	No	5' end missing	No	room-keeping	Tyrosinase (PF00264) * 4, Heamocyan_bet_s (PF14830) * 3	PREDICTED: hemocyanin G-type, units Oda to Odg-like [*Aplysia californica*]
Ls-SMP-165	comp147522_c0_seq3	Left mantle	No difference	3.5E+03	0.32	No	5' end missing	No	room-keeping	Tyrosinase (PF00264) * 4, Heamocyan_bet_s (PF14830) * 3	PREDICTED: hemocyanin G-type, units Oda to Odg-like [*Aplysia californica*]
Ls-SMP-166	comp147522_c0_seq4	Left mantle	No difference	5.7E+03	0.67	No	5' end missing	No	room-keeping	Tyrosinase (PF00264)	PREDICTED: hemocyanin G-type, units Oda to Odg-like [*Aplysia californica*]
Ls-SMP-170	comp147593_c0_seq11	Left mantle	No difference	9.3E+03	0.36	No	No	Yes	room-keeping	Tyrosinase (PF00264) * 3, Heamocyan_bet_s (PF14830) * 3	hemocyanin alphaD-subunit [*Helix lucorum*]
Ls-SMP-171	comp147593_c0_seq8	Left mantle	No difference	4.1E+03	0.81	No	Yes	Yes	room-keeping	Tyrosinase (PF00264) * 2, Heamocyan_bet_s (PF14830)	PREDICTED: hemocyanin G-type, units Oda to Odg-like [*Aplysia californica*]
Ls-SMP-192	comp64206_c0_seq1	Left mantle	No difference	3.9E+01	0.67	Yes	Yes	No			PREDICTED: uncharacterized protein LOC106054306 [*Biomphalaria glabrata*]
Ls-SMP-203	comp88734_c0_seq1	Right mantle	No difference	3.3E+03	57.08	No	Yes	Yes			
Ls-SMP-206	comp93327_c0_seq1	Left mantle	No difference	5.6E+01	1.14	Yes	Yes	No			

Figure 2.17.

Features of the 32 SMPs displaying a significant difference in gene expression between left and right sides of the mantle. "Room-keeping" genes are those specific to a particular functional unit of the body, such as the nervous system, blood, or the immune system. See Supplementary Table S2.7 for further details of those 32 SMPs.

The remaining 7 SMPs, which are expressed more strongly on the right than the left, comprise three house-keeping proteins, the bromodomain-containing protein, DDB_G0280777-like protein (Ls-SMP-129) and two hephaestin-like proteins, which include Cu-oxidase domains (Ls-SMP-102 and Ls-SMP-105), one room-keeping neurogenic locus notch homolog protein (Ls-SMP-149), two uncharacterized proteins (Ls-SMP-43 and Ls-SMP-50), and one novel protein (Ls-SMP-203). These three novel and uncharacterized proteins (Ls-SMP-43, Ls-SMP-50, and Ls-SMP-203) have one or more low-complexity regions, and Ls-SMP-43 has a tyrosinase domain.

As discussed above, the 32 SMPs identified by left-right comparisons include a number of proteins that contain apparently important domains. In addition, the 207 SMPs identified in this study include 65 proteins (31.4%) that are homologous to known house-keeping proteins, but among the 32 left-right asymmetric SMPs, the number of proteins homologous to known house-keeping proteins is only 5 (15.6% of the 32 SMPs). Thus, the exercise of left-right comparisons appears to have successfully narrowed the list of potentially important SMPs.

In order to validate the above observations of asymmetric left-right expression of SMP genes in the mantle, levels of gene expression were also compared using quantitative PCR (qPCR). Overall, qPCR results did not show any trends significantly contradictory to those of transcriptome comparisons (Fig. 2.18 and Supplementary Table S2.9). However, some individual biological replicates showed contradictory patterns, resulting in detection of only four SMPs (Ls-SMP-2, Ls-SMP-19, Ls-SMP-88, and Ls-SMP-149) that indicated statistically significant differences between left and right, with the former three (Ls-SMP-2, Ls-SMP-19, and Ls-SMP-88) being more highly expressed on the left than the right, and Ls-SMP-149 more highly expressed on the right than the left. Four other SMPs (Ls-SMP-48, Ls-SMP-59, Ls-SMP-129, and Ls-SMP-203) also indicated patterns of gene expression perfectly consonant with those of the transcriptome analysis; however, since they did not show statistically significant differences between left and right, we focused on the former four SMPs as the most important SMPs identified in our study.

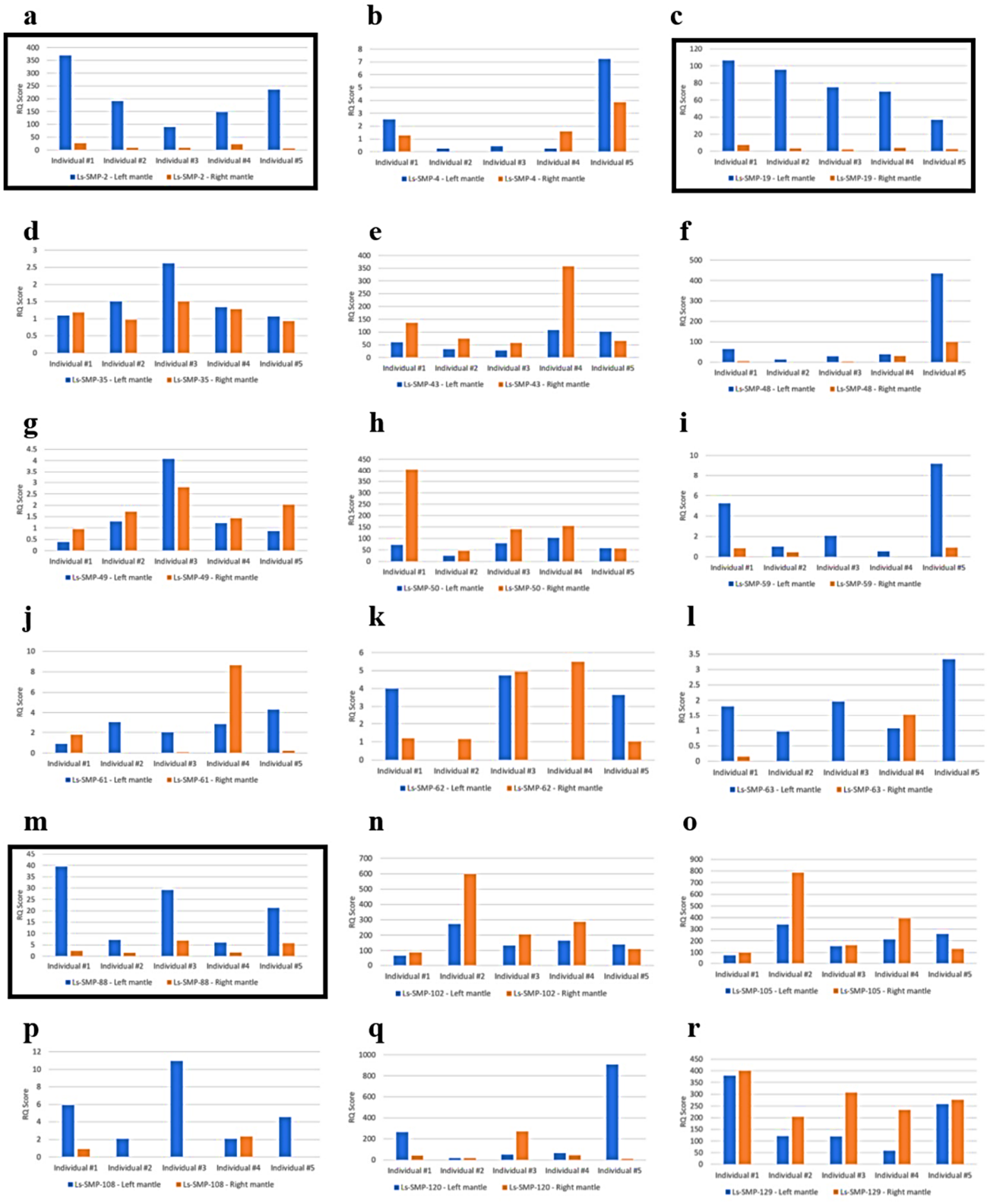

Figure 2.18 (continued).

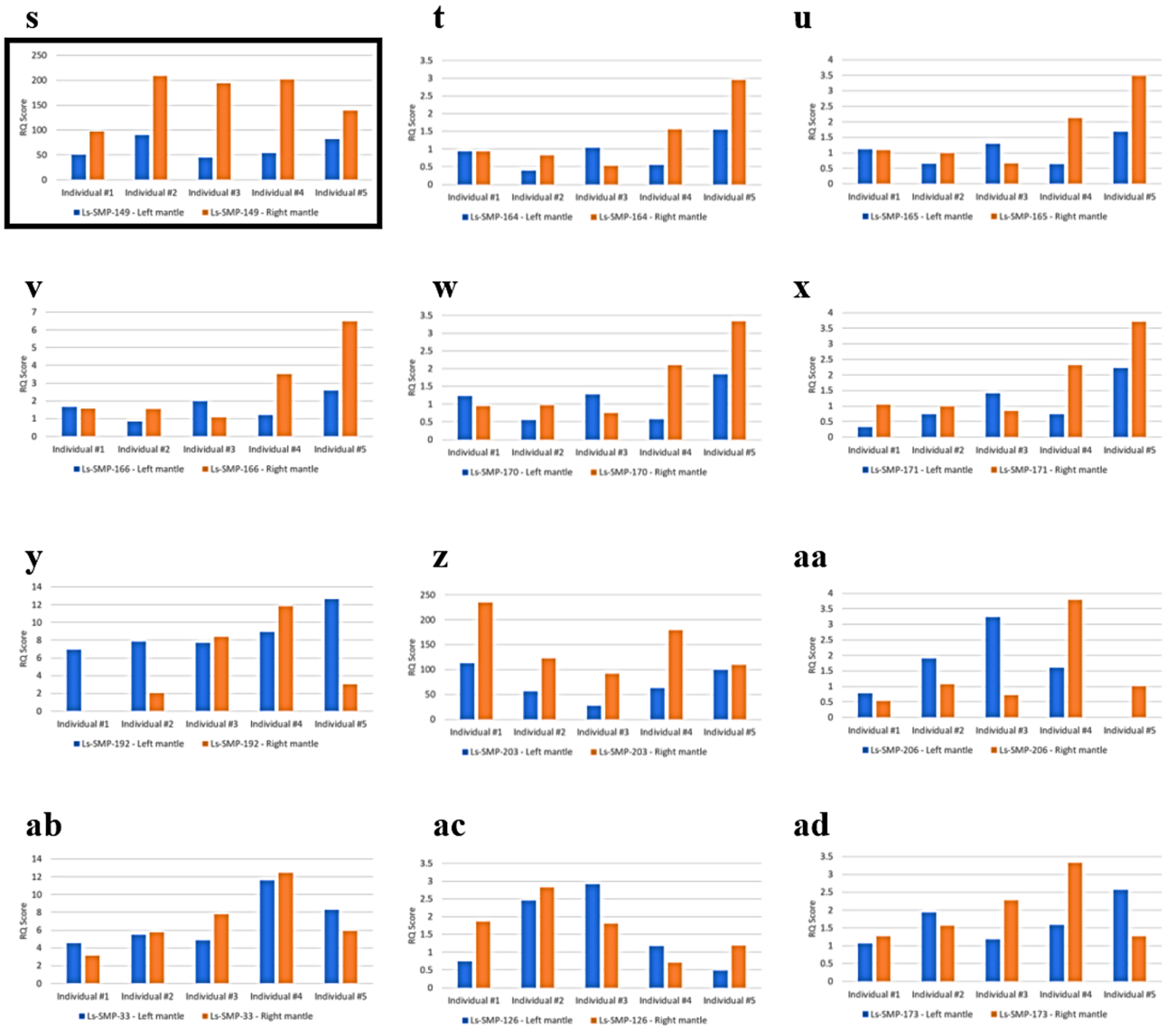

Figure 2.18 (continued).

Results of qPCR analysis. RQ (Relative Quantitation) scores are shown for 27 successfully amplified asymmetrically expressed SMP genes in the transcriptomic analysis (a to aa) as well as the three symmetrically expressed SMP genes (as control; ab to ad). Quantitative PCR was performed on each left and right mantle sample for 5 biological replicates. Relative expression levels (RQ Scores) of two technical replicates for each individual are shown. The four SMP genes (a: Ls-SMP-2, c: Ls-SMP-19, m: Ls-SMP-88, and s: Ls-SMP-149) that indicated exactly the same asymmetric trends as in transcriptomic analysis are shown by a black rectangle. Five SMP genes (Ls-SMP-30, Ls-SMP-39, Ls-SMP-60, Ls-SMP-70, and Ls-SMP-150) out of the 32 asymmetrically expressed SMP genes in the transcriptomic analysis did not show detectable signals (Fig. 2.17).

2.3. Discussion

2.3.1. Comprehensive identification and sequence annotation of SMPs

Traditionally, prediction of functionally important SMPs identified by proteomic analysis has relied on sequence similarities, including the presence of domains conserved among SMPs and other proteins (e.g. Isowa et al. 2015; Jackson et al. 2015; Takeuchi et al. 2016)) and the abundance of SMPs contained in the shells (Jackson et al., 2015). These two approaches have been effective in identifying dozens of potentially important SMPs (see Supplementary Material for detailed discussion). However, since the *in vivo* functions of those conserved domains or known SMPs in biomineralization have yet to be clarified, identification of sequences homologous to those domains or proteins is not promising. In addition, the mere abundance of SMPs may be an inadequate measure of importance.

Since shell matrix protein are secreted by mantle epithelial cells, genes with high expression levels in the mantle would result in high abundances of gene products contained in the shell, reflecting their involvement with shell formation. Contrary to this assumption, mean gene expression level of SMPs in the whole mantle is not correlated with the number of peptides in the shell (Fig. 2.19a: $r_s = 0.054$, $p = 0.78$; Spearman's rank correlation coefficient). The 4 proteins with the highest gene expression levels (FPKM: > 2.0×10^4) are rarely detected in the shell (protein abundance: < 5.00). Highly expressed SMP genes with FPKM values greater than 2.0×10^4 include those encoding a house-keeping protein homolog that lacks a signal peptide, such as Ls-SMP-173 (60S ribosomal protein P2) and Ls-SMP-175 (ATP-dependent RNA helicase DDX43) (Supplementary Table S2.4). Notably, these genes are not differentially expressed between left and right sides of the mantle. These results imply that these SMPs are accidentally embedded in the shell, merely because they are abundant in epithelial cells.

2.3.2. Inference of important SMPs based on traditional sequence-based approaches

In order to identify functionally important SMPs based on sequence similarities, the 207 SMPs identified in this study were first searched against GenBank using BLAST, and a total of 165 proteins showed similarity to known proteins. Of the 165 proteins, 70 were categorized as house-keeping proteins, 40 as room-keeping, 13 as SMPs, and 49 as uncharacterized proteins (Supplementary Table S2.4). Seven proteins were categorized as both room-keeping proteins and SMPs. This indicates that many SMPs identified in this study are house-keeping proteins that are probably irrelevant to shell formation. Some of the SMPs categorized as room-keeping proteins are also known as SMPs in other taxa or as an otolith matrix protein (Ls-SMP-199; starmaker-like), suggesting that those proteins could well be important in shell formation.

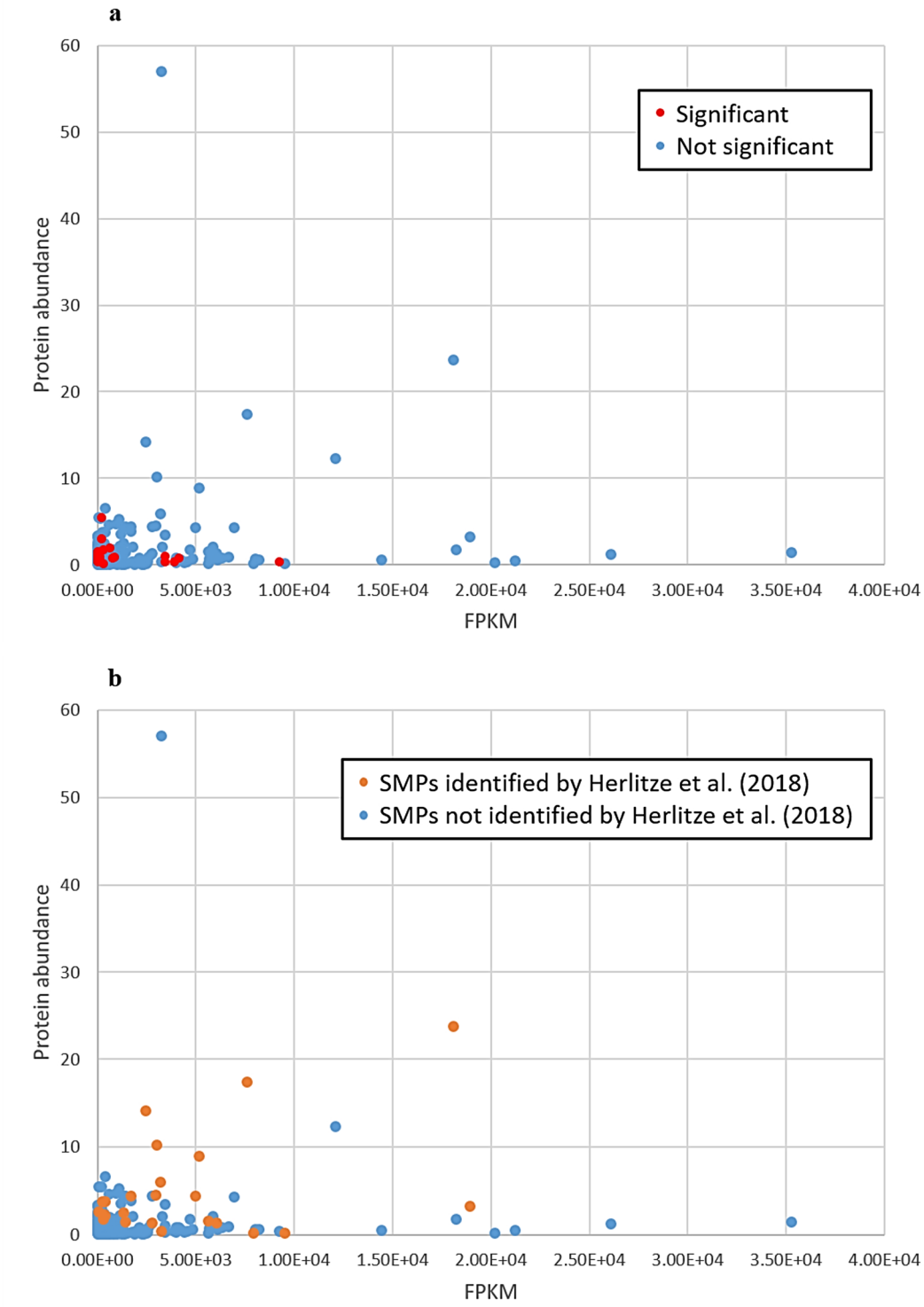

Figure 2.19.

Graphs showing comparisons of FPKM and protein abundance for identified SMPs. (a) SMPs displaying significant differences between left and right sides of the mantle are shown in red. (b) SMPs that were also identified by Herlitze et al. (2018) are shown in orange (Herlitze et al., 2018).

In order to further identify possibly important SMPs based on sequence similarities, searches for domains conserved among molluscan SMPs were carried out. Out of the 261 domains identified in the 207 SMPs of *L. stagnalis* using SMART and InterProScan 5, fifteen domains are shared with at least one of the two other molluscs compared, the limpet, *Lottia gigantea* (K. Mann et al., 2012) and the oyster, *Crassostrea gigas* (G. Zhang et al., 2012; D. Feng et al., 2017). They include 8 extracellular matrix, two polysaccharide interaction, two cation interaction, and three enzyme domains (Fig. 2.4). Of these 16 domains, 11 domains [5 extracellular matrix (CCP, EGF, FN3, VWA, and WAP), three polysaccharide interaction (CLECT, Cht_BD2, and Laminin_G), two enzyme (Tyrosinase and An_Peroxidase), and one cation interaction (EFh) domains] are shared among all three species. Two domains [one enzyme (Glyco_hydro_20) and one cation interaction (Cu-oxidase) domains] are shared only by *L. stagnalis* and *C. gigas*, and three domains [three extracellular matrix (Ependymin, DERM, and SCP) domains] by only *L. stagnalis* and *L. gigantea* (Fig. 2.4). SMPs containing one of those domains may be functionally important (Jackson et al., 2015; Takeuchi et al., 2016).

The 16 common domains identified in this study can be classified into different categories according to their known functions. Eight domains (EGF, FN3, VWA, WAP, SCP, DERM, Ependymin and CCP) are categorized as extracellular matrix domains. EGF (Epidermal Growth Factor) (IPR000742) domain is a short peptide with a distinctive six-cysteine motif. The main structure involves a two-stranded beta-sheet followed by a loop and a C-terminal short, two-stranded sheet. Both calcium-binding EGF repeats and non-calcium-binding EGF repeats are known (Davis, 1990). FN3 (Fibronectin type 3) (IPR003961) domain is one of the three types of internal repeat in the plasma protein, fibronectin. Fibronectin is involved in cell adhesion, cell morphology, thrombosis, cell migration, and embryonic differentiation (Sottrup-Jensen et al., 1984). VWA (von Willebrand factor type A) (IPR002035) domains in extracellular eukaryotic proteins mediate adhesion via metal ion-dependent adhesion sites, and von Willebrand factor is a large multimeric glycoprotein found in blood plasma (Ruggeri, 1993). WAP (Whey Acidic Protein) (IPR008197) domain constitutes the four-disulfide core of WAP. A number of proteins that have WAP domains have been shown to exhibit antiproteinase activity (Bingle et al., 2002). SCP (sperm-coating glycoprotein) (IPR014044) is also known as CAP (cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 protein) domain, and is found in a wide range of organisms, including prokaryotes and eukaryotes, where it may function as an endopeptidase (Gibbs et al., 2008). Thus, SCP domain can also be categorized as an enzyme domain. DERM (dermatopontin) (PF14704) is known as a low-molecular-weight protein in the extracellular matrix that has been reported to mediate cell adhesion by cell surface integrin binding (Okamoto & Fujiwara, 2006; Okamoto et al., 2010). Ependymin (IPR001299) is known as a secretory and calcium-binding

meningeal glycoprotein found in the cerebrospinal fluid of teleost fish and has a bound form of glycoprotein associated with the extracellular matrix, including collagen fibrils (Shashoua, 1991). CCP (IPR000436) domain is also known as Sushi domain. It has a beta-sandwich structure, and is found in selectin, which can bind to C-type lectins (CLECT) (Sharon & Lis, 2001).

Two domains (CLECT, and ChtBD2) are categorized as polysaccharide interaction domains. CLECT (C-type lectin) (IPR001304) is a Ca^{2+} dependent carbohydrate-binding domain, and has functions in cell adhesion, immune response to pathogens, and apoptosis (K. Mann et al., 2000; Sharon & Lis, 2001). ChtBD2 (chitin-binding 2) (IPR002557), containing six conserved cysteine residues, is a chitin-binding domain, and is also known as an extracellular domain (Z. Shen & Jacobs-Lorenat, 1998).

Two domains (EFh and Cu-oxidase) are categorized as a cation-interaction domain. EFh (EF-hand) (IPR002048) has a calcium-binding motif that is shared by many calcium-binding proteins (Ban et al., 1994). Cu-oxidase (multicopper oxidase) (IPR001117) oxidizes substrate molecules by accepting electrons at a mononuclear copper center and transferring them to a trinuclear center (Bento et al., 2005). Cu-oxidase is also categorized as an enzyme domain.

Three domains (Tyrosinase, An_peroxidase and Glyco_hydro_20) are categorized as enzyme domains. Tyrosinase (IPR002227) is an oxidase that controls production of melanin and other pigments. A tyrosinase domain has three histidine residues for interaction with copper atoms that are shared by some hemocyanins, which are copper-containing oxygen carriers from the hemolymph of many molluscs and arthropods (Kato et al., 2018). An_peroxidase (IPR002007) is known as animal heme-dependent peroxidase, the heme-containing enzyme that uses hydrogen peroxide as an electron acceptor to catalyze various oxidative reactions (Kimura & Ikeda - Saito, 1988; H. Li & Poulos, 1994; Nelson et al., 1994). Glyco_hydro_20 (glycoside hydrolase family 20) (IPR015883) is a widespread group of enzymes that hydrolyze the glycosidic bond between two or more carbohydrate moieties (Vinet & Zhedanov, 2011). In *L. stagnalis*, SMPs containing DERM, An_peroxidase, CLECT, CCP, or VWA domains have been identified in previous studies (Sarashina & Endo, 2006; Herlitze et al., 2018).

Tyrosinase domain-containing SMPs dominate the SMPs with an asymmetric gene expression pattern in the mantle, representing 6 of the 32 asymmetric SMPs (Ls-SMP-43, Ls-SMP-164, Ls-SMP-165, Ls-SMP-166, Ls-SMP-170, and Ls-SMP-171). Phylogenetic analyses demonstrated that those proteins comprise two distinct classes, tyrosinase (Ls-SMP-43) and hemocyanin (Ls-SMP-164, Ls-SMP-165, Ls-SMP-166, Ls-SMP-170, and Ls-SMP-171) (Fig. 2.15).

Of SMPs containing at least one of those 15 domains, 9 SMPs are specific to the mantle transcriptome, showing no detectable expression in the foot (Supplementary

Table S2.4). They include Ls-SMP-23 (DERM), Ls-SMP-24 (DERM), Ls-SMP-48 (VWA), Ls-SMP-59 (CLECT), Ls-SMP-61 (EFh), Ls-SMP-62 (EFh), Ls-SMP-81 (CLECT), Ls-SMP-82 (CLECT), and Ls-SMP-186 (EFh) (shared domain names are indicated in parentheses). Ls-SMP-23 and Ls-SMP-24 are dermatopontin homologs that are inferred to be splicing variants (Fig. 2.20). Sarashina et al. (2006) reported three dermatopontin genes from *L. stagnalis*, but considered only one as an SMP (Sarashina & Endo, 2006). In this study, all of these dermatopontin homologs were identified as SMPs, but their transcripts were also expressed in the foot. Ls-SMP-23 and Ls-SMP-24 are newly identified in this study as SMPs, of which transcripts are specific to the mantle. Ls-SMP-81 was identified as a perlucin-like protein, and Ls-SMP-82 is a likely splicing variant of Ls-SMP-81. Perlucin is an SMP that may promote nucleation and/or growth of calcium carbonate crystals with an ability to bind D-galactose and D-mannose/D-glucose. It has been identified in many other molluscs (K. Mann et al., 2000; Weiss et al., 2000; Wang et al., 2008); thus, it appears likely to be actively involved in shell formation. Ls-SMP-61, Ls-SMP-62, and Ls-SMP-186 have EFh domains, which could mediate calcium ion binding during biomineralization. Ls-SMP-48 has a VWA domain, and is been identified as matrilin-like by BLAST. Matrilin is known as an extracellular protein.

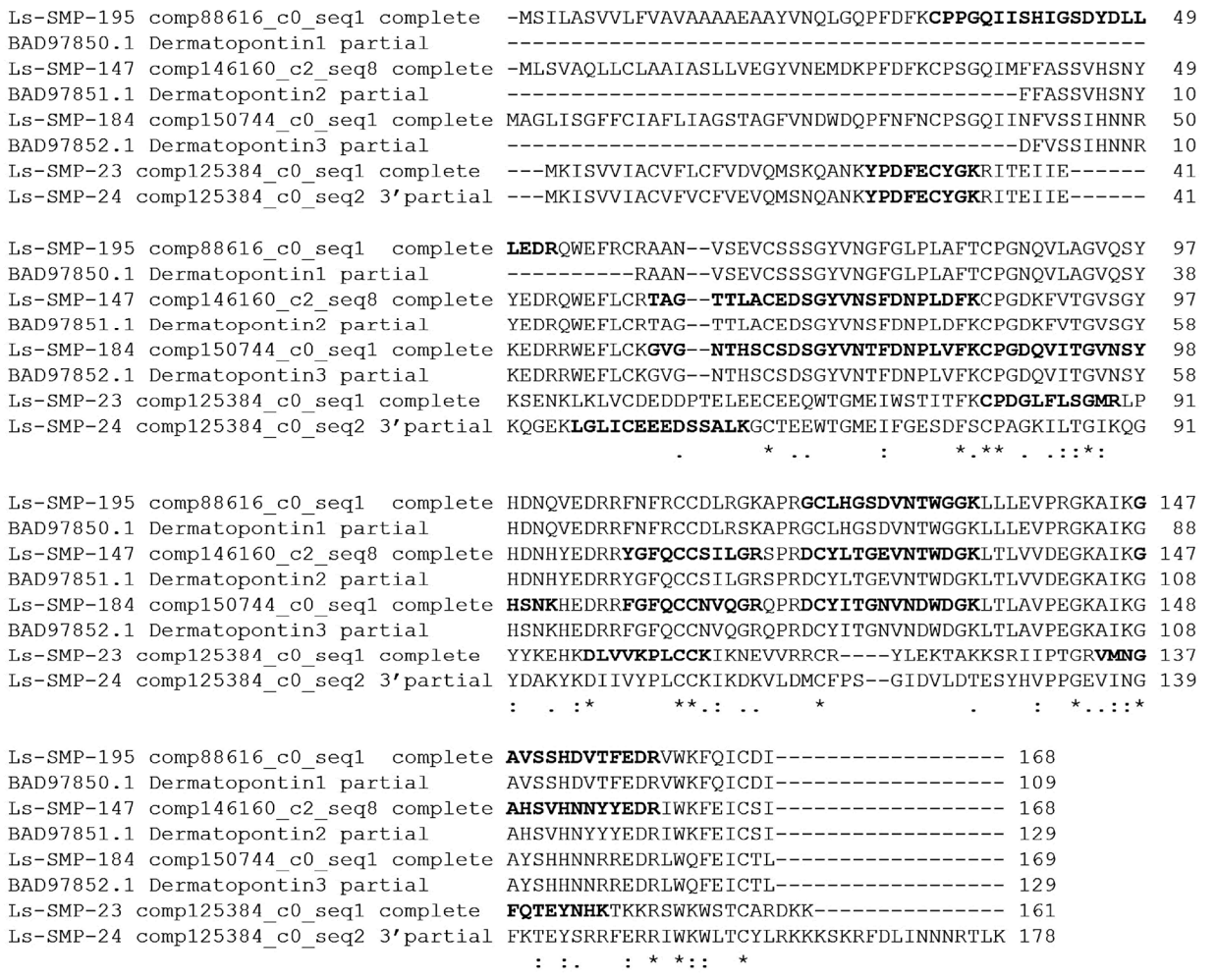

Figure 2.20.

Alignment of amino acid sequences of dermatopontin of *L. stagnalis*. Bold letters denote sequences detected by LC-MS/MS. Asterisks (*) indicate amino acid residues conserved among the 8 sequences. Colons (:) indicate synonymous substitutions. Periods (.) indicate nonsynonymous substitutions.

Our signal peptide search indicated that of the 147 SMPs with complete sequences, only 99 are predicted to have signal peptides. In total, 115 SMPs were predicted to have signal peptides. That those SMPs have signal peptides is consistent with the fact that SMPs are secretory proteins. On the other hand, for the remaining 92 of the 207 SMPs, or 48 of the 147 SMPs with a complete sequence, signal peptides were not predicted. SMPs without signal peptides may have originated from either extracellular regions of proteins with transmembrane regions or from contaminating cells (Marin et al., 2016). However, since no transmembrane regions have been identified in SMPs with complete sequences, contaminating cells, such as hemocytes, are the more likely source. The fact that SMPs without signal peptides are dominated by house-keeping proteins, such as 60S ribosomal proteins and histone (Supplementary Table S2.4), tends to support this

interpretation. Thus, those SMPs may have accidently leaked from cells such as dying or randomly trapped cell remnants and may not be important in shell formation.

Our domain searches revealed that SMPs categorized as novel or uncharacterized proteins by BLAST, often contain one or more low-complexity regions (LCRs), with 14, 13, 6, 24, and 24 LCR-containing SMPs having been identified among 70 house-keeping, 33 room-keeping, 13 known SMPs, 49 uncharacterized, and 42 novel proteins (Supplementary Table S2.4). LCRs are often present in SMPs and are considered important in shell precipitation (Jackson et al., 2010; Marin et al., 2016). Therefore, novel or uncharacterized SMPs that contain LCRs may also be important in shell formation.

In order to further characterize SMP-encoding genes, based on sequence comparisons, we compared gene ontology (GO) between SMP-encoding genes and mantle transcripts. The results indicated that in the category "molecular function", notable differences were observed in hydrolase activity, protein binding, oxidoreductase activity, and carbohydrate derivative-binding at level 3, and in cation binding and anion binding at level 4. Those terms are enriched in SMP-encoding genes (Figs. 2.10 and 2.13). For the other two categories "cellular component" and "biological process", notable differences were not observed between the mantle and SMP transcriptomes, except for the term "extracellular region" at level 2 of "cellular component", a term enriched in SMP transcripts (Supplementary Figs. 2.8, 2.9, 2.11, and 2.12). Proteins that indicated enrichment in GO terms typically seen in SMP-encoding genes could be important in shell formation. Enrichment of the term "extracellular region" is concordant with the fact that SMPs are secretory proteins.

Marin et al. (2016) classified the sources of organic matrices in animal calcium carbonate skeletons into three categories: (1) secretome, (2) cleaved extracellular domains of transmembrane proteins, and (3) cellular contaminants. They distinguished SMPs originating from (1) and (2) (skeletome) from those in (3) (entrapped contaminants) (Marin et al., 2016). A secretome, by definition, is comprised of proteins with a signal peptide. Of the 32 left-right asymmetric SMPs, 18 SMPs have a signal peptide, and are thought to have originated from the secretome of this species. On the other hand, 14 SMPs did not show potential signal peptides, although 11 of those 14 SMPs are represented by 5' partial or internal sequences, and could have signal peptides. Thus, three SMPs (Ls-SMP-39, Ls-SMP-50, and Ls-SMP-170) remain as complete sequences without signal peptides. They cannot be identified as belonging to the above category (2) because they do not have transmembrane regions in domain searches using InterProScan5, SMART, and Blast2GO. Ls-SMP-39 is a novel protein, showing a high value of logFC and the lowest p value in the test of the difference in gene expression levels between the left and right (Fig. 2.17). Ls-SMP-50 is homologous to an uncharacterized protein identified from *Biomphalaria glabrata*, and has two low

complexity regions, but has no known functional domains. It is more strongly expressed in the right mantle than in the left (Fig. 2.17). Ls-SMP-170 has been identified as a hemocyanin homolog, and has a low-complexity region, although its function in shell formation is unknown (Fig. 2.17). Incidentally, Marin et al. (2016) considered only coral skeletal proteins as included in this category having a transmembrane protein extracellular domain. Isowa et al. (2015) and Jackson et al. (2015) reported SMPs containing one or more transmembrane domains from the brachiopods *Laqueus rubellus* and *Magellania venosa*, respectively, and they may represent other SMPs belonging to this category (Isowa et al., 2015; Jackson et al., 2015). According to the classification of Marin et al. (2016), therefore, the three proteins without signal peptides may be regarded as cellular contaminants. Although we cannot completely exclude this possibility, since they showed asymmetric gene expression between left and right, they may not be just cellular contaminants. One possibility is that they function to maintain the asymmetric nature of the mantle in regard to growth and metabolism, indirectly contributing to asymmetric shell formation.

2.3.3. Inferring important SMPs based on traditional abundance-based approaches

Other traditional approaches for prediction of important SMPs involve quantification of abundances of SMPs in the shells. When we look at amounts of SMPs, quantified for each protein by the number of times constituent peptides appeared in the spectrograms standardized by the length of the protein, the highest abundance was observed for Ls-SMP-203, which accounted for 14.7% of all 207 SMPs, and was annotated as a novel protein, followed by Ls-SMP-53 and Ls-SMP-52, which accounted for 6.0% and 4.5% of the total, respectively, and which were annotated as a formin-like protein and an extensin-like protein. All were represented by complete sequences in both the soluble and insoluble fractions. A common feature of these three proteins is that they contain at least one low-complexity region. Ls-SMP-203 [theoretical molecular mass (Mm) = 20094.3; theoretical isoelectric point (pI) = 8.04] has a signal peptide, a potential O-glycosylation site, and 16 potential phosphorylation sites. Ls-SMP-53 (Mm = 19811.7; pI = 7.01) has no signal peptide with 11 potential phosphorylation sites. Ls-SMP-52 (Mm = 24230.1; pI = 8.89) has no signal peptide, a potential O-glycosylation site, and 13 potential phosphorylation sites (Supplementary Table S2.4).

The results of SDS-PAGE analysis revealed three major bands (10 kDa, 20 kDa, and 22 kDa) for the soluble fraction and one major band (20 kDa) for the insoluble fraction (Fig. 2.7). Since all three of the highest abundance SMPs (Ls-SMP-203, Ls-SMP-53, and Ls-SMP-52) have molecular masses of about 20 kDa and are represented in both the soluble and insoluble fractions, it appears likely that they correspond to the 20 kDa bands seen in SDS-PAGE gels in the soluble and insoluble fractions. Those proteins may be

important in shell formation by virtue of their high abundance, but details should be experimentally studied.

2.3.4. Asymmetrical expression of SMP encoding gene in the mantle tissues

Instead of averaged gene expression levels, asymmetric gene expression patterns in the mantle regions may offer a better measure of the functional importance of SMPs. Because snail shells are laterally asymmetric, accretionary shell growth implies asymmetric expression of functional SMP genes. For example, SMPs that suppress shell precipitation should be more highly expressed in the mantle region at the inner side of the shell than in that corresponding to the outer edge of the shell. In dextral shells, the outer and inner sides correspond to the right and left sides of the mantle, respectively. In this study, the dextral pond snail, *L. stagnalis*, has been studied, and expression levels have been compared between the right and left sides of the mantle tissues in three individuals, in order to identify functionally important SMPs.

Both transcriptomic and qPCR analyses of gene expression levels of SMPs between the right and left sides of mantle tissues showed that among asymmetrically expressed SMPs, those that are more highly expressed on the left than the right are three time more abundant than those that are more highly expressed on the right than the left (Fig. 2.17). This observation was unexpected because we assumed that a dextrally coiled shell is produced by a greater shell precipitation on the right than the left side of the mantle, and that more shell precipitation-promoting SMPs would be expressed on the right than the left. Our results, however, suggest that a dextrally coiled shell is produced by inhibition of shell precipitation on the left. Inhibitory roles of SMPs have long been recognized (Wheeler et al., 1981; Westbroek, 1991) and could be at work in production of coiled shells.

2.3.5. Candidates for potentially functional SMPs

One of the three SMPs that showed higher expression on the left than the right, in both transcriptomic and qPCR analyses (Ls-SMP-88), indicated a significant sequence similarity to Pif (Fig. 2.17), an SMP originally isolated from the pearl oyster *Pinctada fucata* (Suzuki et al., 2009). It contains two chitin-binding domains (ChtBD2) and an extracellular domain (Laminin_G), as in Pif. However, it has no von Willebrand factor type A domain (VWA), which is involved in protein binding, and is always found in Pif (Suzuki et al., 2009, 2013; R. Zhao et al., 2018). There exists a Pif-like SMP, known as BMSP (Blue Mussel Shell Protein) (Suzuki et al., 2011), originally isolated from the bivalve, *Mytilus galloprovincialis*. It has four VWA domains, a ChtBD2 domain, and a Laminin_G domain. Phylogenetic analysis of ChtBD2 and Laminin_G domain sequences indicated that Ls-SMP-88 is closer to Pif than to BMSP, suggesting that it originated as

Pif, but lost the VWA domain subsequently (Figs. 2.14 and 2.21). In pearl oysters, Pif binds aragonite crystals and promotes nacre formation (Suzuki et al., 2009). Although functions of Ls-SMP-88 have yet to be clarified, one possibility is that the loss of the VWA domain led to loss of shell formation-promoting roles and acquisition of inhibitory roles instead.

Two other SMPs that indicated higher expression on the left than the right in both transcriptomic and qPCR analyses (Ls-SMP-2 and Ls-SMP-19) do not show any similarity to known proteins or domains. Except for the fact that Ls-SMP-2 and Ls-SMP19 have two and one low complexity region(s) (LCRs), respectively (Supplementary Table S2.8), they do not to show any apparent features characteristic of SMPs (Fig. 2.22). Further studies are required to confirm their importance in biomineralization.

Ls-SMP-149 was the only SMP more highly expressed on the right than the left in both transcriptomic and qPCR analyses. It showed significant sequence similarity to the neurogenic locus Notch (NlN) homolog protein-like isoform X1 of the pond snail, *Biomphalaria grabrata* (Fig. 2.17). Although it is annotated as a homolog of NlN in *B. grabrata*, its overall domain composition and domain arrangements are entirely different from those of other typical NlN proteins in *B. grabrata*, *Crassostrea gigas*, *Drosophila melanogaster*, and *Homo sapiens* (Fig. 2.23). Ls-SMP-149 has a single EGF domain (which exists in multiple copies in typical NlNs) and two Whey Acidic Protein (WAP) domains (that do not exist in NlN). Ls-SMP-149, therefore, is unlikely to function as a NlN. Since it has two WAP domains that exhibit an antiproteinase function (Bingle et al., 2002), it may act as a proteinase inhibitor, although its function in biomineralization needs to be verified. A WAP domain-containing SMP (isotig_7807) has been isolated from the land snail, *Cepaea nemoralis* (K. Mann & Jackson, 2014); however, isotig_7807 has only one WAP domain and no EGF domain. Thus, Ls-SMP-149 and isotig_7807 are not very similar. Phylogenetic analysis of the WAP domain sequences tends to support different origins (Fig. 2.24). Iwamoto et al. (2020) reported that an EGF-like domain containing protein, which has two EGF domains in tandem, promotes aggregations of polycrystalline calcite *in vitro* (Iwamoto et al., 2020). Ls-SMP-149 has only one EGF domain and is different in origin from this EGF-like domain containing protein (Fig. 2.25). Online tBLASTn searches using NCBI GenBank, EMBL EnsemblMetazoa, and OIST genome browser with the Ls-SMP-149 (or Ls-SMP-58) sequence as the query found a similar sequence in the *B. grabrata* genome (XP_013065122.1), but returned no similar sequences in *C. gigas*, *P. fucata*, and *L. gigantea* genomes. It appears likely that the homologs of Ls-SMP-149 evolved in pulmonates independent from the EGF domain containing SMPs of other molluscs.

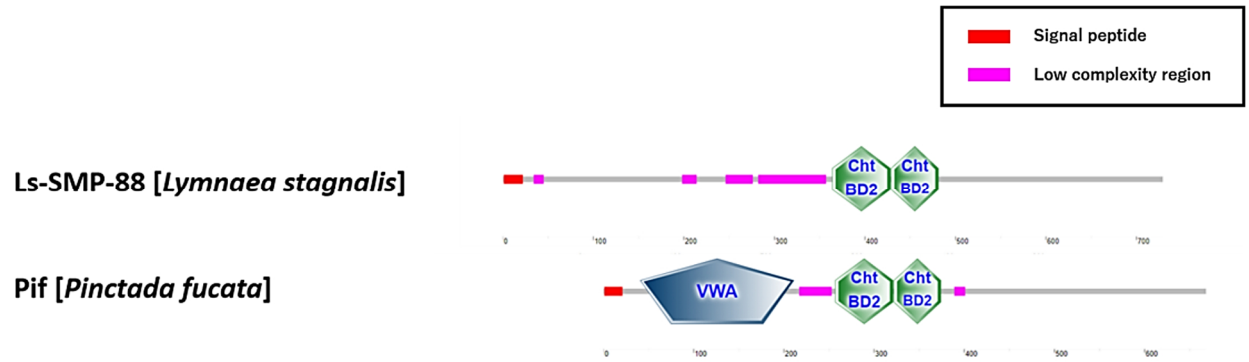

Figure 2.21.

Schematic representations of the conserved domain configuration of Ls-SMP-88 of *Lymnaea stagnalis* and Pif of *Pinctada fucata* (results from SMART domain searches). A Laminin_G domain was not identified by SMART domain searches in those proteins, but was identified by Pfam domain searches in the region just downstream of the ChtBD2 domains in both Ls-SMP-88 and Pif.

>Ls-SMP-2 comp111558_c0_seq1 type:3prime_partial len:296

```
1      MMTSLLLLSCCCLSLAMSQKMSHILMNMVAQDPYARNMMRLPLFQELFDIEIPPTSEPSA  60
61     KQLGVASLFGMAGNGIFGGPSVGGNAPRAPPTGSSTPSQSQTSESMLGASSWDRSGFGSG  120
121    QPVGRVTGPSQFDGGYQMTGSVPDTSQNGPPQLGGSLTWPQQNNGRSFGSSPVGNGLNWP  180
181    PTVPSNSQMGGGNQMTWPQQPSGNQQMGDGSSWPPQMGGVIAWPPFGGNMPGNQQMGNNI  240
241    PGNAQPQWPAQMGSNQQLGYPANQGNSQPTWPSQSNAQSGNPPNQGNAQPQWPAQ       296
```

>Ls-SMP-19 comp123495_c0_seq1 type:5prime_partial len:215

```
1      NPPNQGNAQPQWPAQMGGNQQIGNGVPENNNPIWSGMARPPNTGDSMQGNNQMVGSMPGN  60
61     PQMGGPMPGNNQMGGAMPGNNQMGGPMPGNNQMAGPMPVNNHMGGPMPGNNQMSSSPPYG  120
120    ISSNTRPKIPLDFIVQLYNTASNPTQGEPTSGNSAMEGQVSQTDPSTMASFSNTPRQPRA  180
180    VIQAGQVILPKGPPSRNFIDRLMALGGAEGIEIV                            215
```

Figure 2.22.

Amino acid sequences of Ls-SMP-2 and Ls-SMP-19. The signal peptide of LS-SMP-2 is shown in red. Low complexity regions are highlighted by yellow shades. The MS/MS peptides are indicated by gray shades.

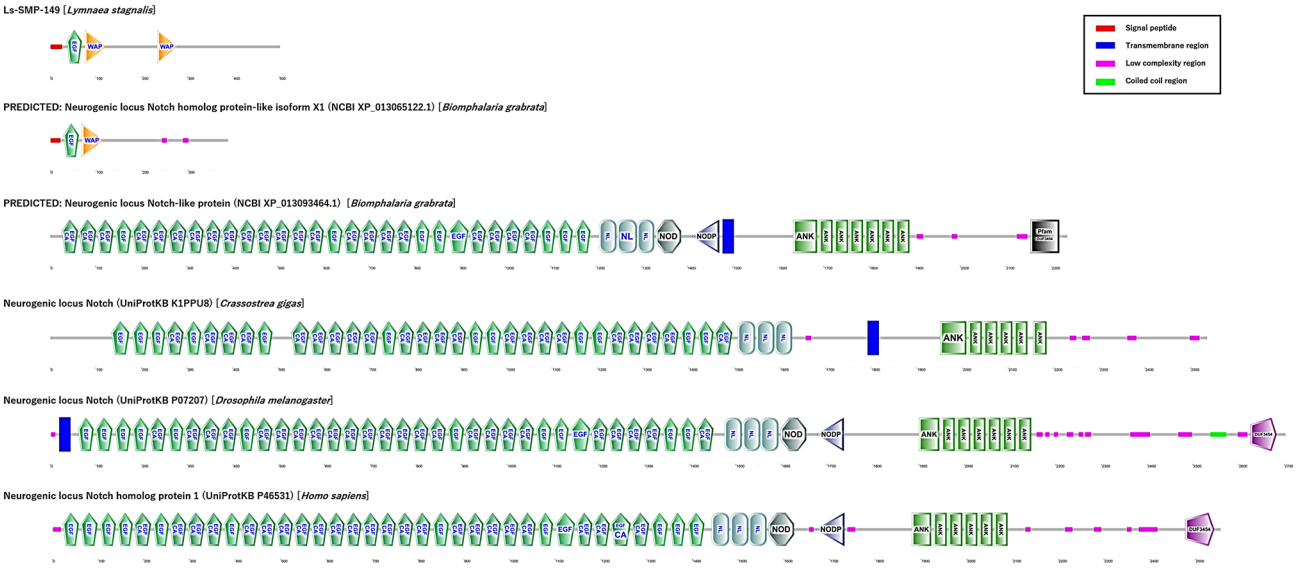

Figure 2.23.

Schematic representations of the conserved domain configuration of Ls-SMP-149 and neurogenic locus Notch proteins and a related protein of other species (*Biomphalaria grabrata*, *Crassostrea gigas*, *Drosophila melanogaster*, and *Homo sapiens*) (results from SMART domain searches).

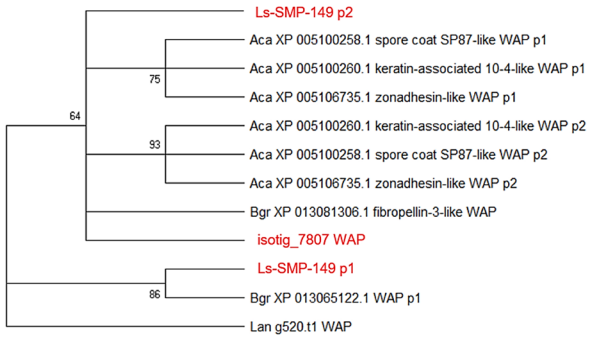

Figure 2.24.

Maximum likelihood tree of the WAP domains in molluscan proteins, including SMPs identified from *Lymnaea stagnalis* in this study and *Cepaea nemoralis* by Mann and Jackson (2014) (K. Mann & Jackson, 2014). The ML tree was inferred from 12 domain sequences of 7 proteins, including one domain sequence from one protein of the brachiopod, *Lingula anatina* as an outgroup, using the JTT + G + I model based on 35 amino acids. Polychotomy results if the bootstrap value of the node is lower than 50%. Bootstrap values are indicated for nodes with a value greater >50%. Molluscan WAP domain sequences were retrieved from NCBI GenBank after BLAST searches using the WAP domain sequence of Ls-SMP-149 of *L. stagnalis* as a query. Sequence names in red indicate SMP sequences from *L. stagnalis* (Ls-SMP-149), and the land snail, *C. nemoralis* (isotig_7807). Ana: *Aplysia californica*, Bgr: *Biomphalaria grabrata*, Cne: *Cepaea nemoralis*, Lan: *Lingula anatina*. When more than one WAP domain exists in a protein, they are given in order from the N-terminus, and named p1, p2, etc.

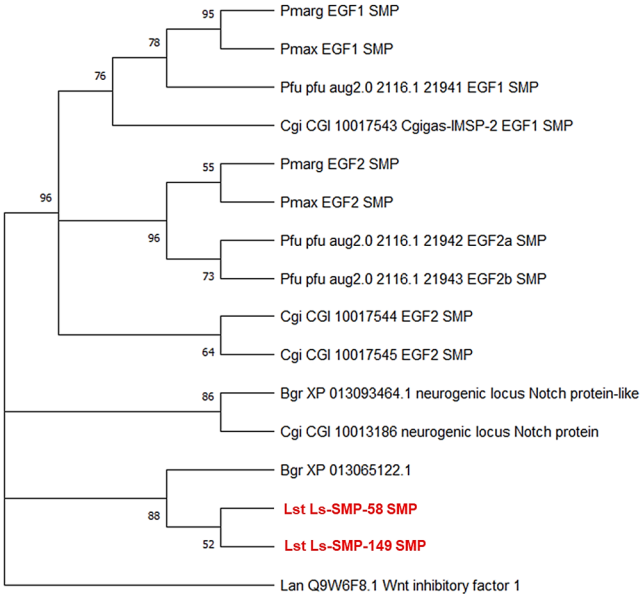

Figure 2.25.

Maximum likelihood tree of the EGF domains in molluscan proteins, including SMPs identified from *Lymnaea stagnalis* in this study, *Crassostrea gigas* by Marie et al. (2011), Zhang et al. (2012). and Iwamoto et al. (2020), *Pinctada margaritifera* and *Pinctada maxima* by Marie et al. (2012), and *Pinctada fucata* by Zhao et al. (2018). The ML tree was inferred from 16 proteins, including one protein of the brachiopod, *Lingula anatina* as an outgroup, using the WAG+ G model based on 138 amino acids. Polychotomy results if the bootstrap value of the node is lower than 50%. Bootstrap values are indicated for nodes with a value greater than 50%. Molluscan EGF domain sequences were retrieved from NCBI GenBank, EMBL EnsemblMetazoa, or OIST genome browser after BLAST searches using both Ls-SMP-58 and Ls-SMP-149 of *L. stagnalis* as a query. Sequence names in red indicate SMP sequences from *L. stagnalis*. Bgr: *Biomphalaria grabrata*, Cgi: *Crassostrea gigas*, Pfu: *Pinctada fucata*, Pmarg: *Pinctada margaritifera*, Pmax: *Pinctada maxima*, Lst: *Lymnaea stagnalis*, Lan: *Lingula anatina*.

2.3.6. Comparison of previous shell proteome study of *L. stagnalis*

In a previous shell proteome study of *L. stagnalis*, Herlitze et al. (2018) identified 46 shell-forming protein candidates and analysed their expression patterns in trochophore larvae and in the outer mantle lip of juveniles (Herlitze et al., 2018). Of those 46 sequences, 30 have homologs (corresponding to 24 SMPs) in our SMP data, but we could not find an SMP in our data homologous to the remaining 16 sequences (Supplementary Table S2.10). Those 16 sequences comprise 9 that were found in our transcriptome data, but were not found in our proteome data (Case I), and 7 sequences that were not found in our data at all (Case II). Case I may have arisen due to the more stringent conditions, under which our proteome analyses were conducted. We analyzed only protein sequences that were identified by more than one unique peptide, while Herlitze et al. (2018) accepted those identified by only one peptide fragment. Case II can be explained by (1) the incompleteness of the transcriptomic data, (2) differences in methods of sequence assembly, and (3) genetic differences of the strains used. Between the homologous sequences of their 30 candidates and our 24 SMPs, only one SMP [Ls-SMP-61: *Lstag-sfc*-16 of Herlitze et al. (2018)] displayed an asymmetric gene expression pattern in the mantle in our transcriptome analysis. This SMP has two EFh domains, and it is more highly expressed on the left side of the mantle. However, this SMP gene did not show detectable signals in the expression analysis of Herlitze et al. (2018). Of their remaining 29 candidate sequences found among our sequences, 17 sequences are more highly expressed on the right side of the outer lip of juvenile mantle. However, corresponding SMP genes in our data did not show significant differences between left and right (Supplementary Table S2.10). These differences may be explained by the fact that the "left-right" direction of Herlitze et al. (2018) corresponds to the anterior-posterior direction in our study, as evidenced by the illustration for the ventral view of a juvenile in Figure 2.3 of Herlitze et al. (2018). On the other hand, four sequences [*Lstag-sfc-1*, *Lstag-sfc-2a*, *Lstag-sfc-2b*, and *Lstag-sfc-3* of Herlitze et al. (2018)], which were not identified as SMPs in our study, showed a similar expression pattern in that are expressed more strongly on the right side of the mantle in our study, and are expressed on the right side of the asymmetric border of the shell field in trochophore larvae in Herlitze et al. (2018) (Supplementary Table S2.10). This commonality may reflect the fact that the left-right direction in trochophore larvae in Herlitze et al. (2018) corresponds to the left-right direction in our study.

2.4. Conclusions

In this study, we performed a comprehensive study of SMPs by combining transcriptomic analyses using NGS and proteomic analyses using mass spectrometry to identify 207 SMPs from the pond snail, *Lymnaea stagnalis*. We focused on the fact that rates of shell formation differ between the helical axis and the outer edges of snail shells, and we compared levels of gene expression, including those of 207 SMPs between the left and right sides of the mantle tissues. As a result, 32 SMPs that indicated significant differences between left and right have been identified. Quantitative PCR performed on those 32 SMP genes revealed that four SMPs were consistently asymmetric. Although expression levels are low, those four and the 28 other SMPs identified in this study are clearly first candidates for further functional characterizations for the following reasons. In addition to their left-right gene expression asymmetry, they showed (1) sequence similarities to known and likely important SMPs, such as Pif, (2) possession of potentially important functional domains, such as EFh, CLECT, and to chitin binding, and (3) mantle-specific expression of the genes when they showed a large change of gene expression between left and right (see Supplementary Material for detailed discussion).

Snails are well suited to the study of not only biomineralization, but also development of asymmetric morphologies due to the intrinsic nature exploited in this study. *L. stagnalis* will be a good model organism with a relevance to a wide range of research fields, such as neurophysiology, embryology, environmental toxicology, and biomineralization (Davies & Henrissat, 1995; Rittschof & McClellan-Green, 2005; Shimizu et al., 2011; Munley et al., 2013; Herlitze et al., 2018). In the future, it will be necessary to perform more detailed spatial and temporal gene expression analyses, and *in vivo* functional analysis of SMPs using a genome editing method such as CRISPR/Cas9, for the candidates identified in this study (Perry & Henry, 2015).

3. Proteomic analysis of chiral shells: comparisons between the dextral and sinistral shells

3.1. Background

Three major types of coiling, namely, dextral, sinistral, and planispiral types, exist among the spiral shell morphologies of gastropods, with respect to lateral asymmetry. Out of about 6,000 extant gastropod species, the proportion of the species with a dextral shell is reported to be about 80% (Robertson, 1993). Are there any selective reasons why dextral species have dominated in snail evolution? Does the production of dextral or sinistral shells have something to do with biomineralization? Those are only a few questions, among many others, remained to be addressed concerning lateral asymmetry in gastropods.

It has long been known that the direction of shell coiling is determined by a single gene with delayed maternal inheritance (Sturtevant, 1923; Hierck et al., 2005; Asami et al., 2008). In early embryos of the pond snail *Lymnaea stagnalis*, Asami et al. (2008) reported that the blastomeres are twisted 45 degrees to the right in the dextral strain of *L. stagnalis* at the time of the third cleavage, but the split occurs in the sinistral strain without twisting. Subsequently, Kuroda et al. (2009) found that when the 8-cell stage micromeres of a sinistral embryo were artificially placed so as to twist clockwise, the individual grew normally as a dextral embryo. Then, Davison et al. (2016) revealed that Formin, which is a protein involved in polymerization of actin, and is encoded by the *Lsdia1* gene, is likely the long-sought maternal factor that determine the coiling direction in *L. stagnalis*, a discovery which was confirmed by gene knockout experiments of Abe & Kuroda (2019).

Concerning shell coiling mechanisms, it has been revealed that the signal transduction factor Dpp is likely one of the controlling factors (Shimizu et al., 2011, 2013). It shows laterally asymmetric expression patterns in the embryonic shell gland, embryonic mantle, and adult mantle, with high levels of expression in the right and left sides of the body in those shell forming tissues in the dextral and sinistral snails, respectively, while it shows symmetric expression patterns in both the embryonic shell gland and adult mantle in the limpet *Nipponacmea fuscoviridis*, which has a non-coiled shell. When embryos of *L. stagnalis* were subjected to a chemical that inhibits the signal transduction by Dpp, they produced a shell of a cone-like shape without coiling, confirming that Dpp is responsible for shell coiling in those snails.

The dextral and sinistral shells produced by the same species, such as those of *L. stagnalis*, are usually believed to be mirror images to each other. However, it has been reported that they do not show exact mirror images, even when the shells of dextral and sinistral embryos with minimal genotypic differences are compared (Utsuno et al., 2011).

Since the genotypic differences are minimal, and the maternal effect, which represents among the few differences between the dextral and sinistral individuals, does not have effect after the 16-cell stage, it appears possible that the expression patterns of the genes responsible for shell formation are altered between dextral and sinistral snails, after determination of lateral asymmetry at the 8-cell stage.

In order to address the possibility of lateral asymmetry, a comparative analysis of the expression profiles of the factors involved in shell formation (both the SMPs and the transcripts theirof) in the dextral and sinistral snails was performed in this chapter. The results of the proteome analysis of dextral snails in Chapter 2 of this thesis revealed a left-right asymmetric expression profile of SMP genes in the mantle. If a mirror image of this profile is observed also in the sinistral snails, it would give support to the underlying premise that was employed to identify functionally important SMPs in Chapter 2. On the other hand, if the observed profile in the sinistral snails does not mirror the profile in the dextral snails, then it would provide some clues to understand the reasons why the sinistral shell shape does not mirror dextral one, and also to have deeper insight into the mechanisms of shell morphogenesis and biomineralization. The tentative results indicate that the latter is the case. But, since some aspects of the experiments carried out in this study are problematic, discussion shall be given as to how the experimental designs can be improved in the future study, in addition to the discussion concerning the implications of the results obtained in this study.

3.2. Results & Discussion

3.2.1. Analysis of SMP-encoding transcripts of sinistral snail *L. stagnalis*

The mantle tissue from each of three individuals of the sinistral strain was divided into left and right halves, and transcriptome analysis was performed using Illumina MiSeq (Illumina, California, USA). As a result of the assemblage of the obtained data using Trinity under the same conditions as in the dextral strain (cf. Chapter 2), a total of 222,444 contigs were obtained, with a total contig size of 183,290,765 bp, a maximum contig size of 29,507 bp, a minimum contig size of 201 bp, and an N50 length of 1,750 bp. In the dextral transcriptome data, which was obtained using the Illumina HiSeq, the total number of contigs was 37,809, with a total contig size of 384,388,773 bp, and an N50 length of 2,828 bp. Although the total contig size obtained for the sinistral individuals was about the half of that obtained for the dextral individuals, the N50 value, which is commonly used as a measure of data quality, is comparable between the two, indicating that the assembled contigs are reasonably long for both dextral and sinistral data. These transcriptomic data from the dextral and sinistral strains were integrated to form a single dataset, which was used as a reference to identify the protein sequences

contained in the matrix of the sinistral shells. After searching for the longest ORFs by TransDecoder and clustering together the sequences of more than 90% similarities in amino acid sequences by CD-HIT, the integrated reference FASTA file included 37,394 sequences; with a total length of 14,721,755 amino acids (AAs), the maximum length of 10,731 AAs, the minimum length of 86 AAs, an average length of 394 AAs. In order to make a more accurate reference file, it is desirable to integrate the original transcript reads from both the dextral and sinistral strains, and to perform *de novo* assembling, however, even if the 256GB memory is used in the supercomputer system of the National Institute of Genetics, Mishima, it will exceed the memory limit. We need to wait for future developments in both hardware and software to process such large-scale transcriptome data.

3.2.2. Proteomic analysis and comparison of protein profiles and abundances between the dextral and sinistral shells of *L. stagnalis*

In the proteomic analysis, about 100 individuals each of dextral and sinistral shells were collected and subjected to the comparative proteomic analysis. As a result, a total of 443 SMPs were identified. The 443 SMPs include all the 207 SMPs identified in the dextral strain in Chapter 2. Comparisons of the repertoires of the SMPs contained in the dextral shells and the sinistral shells indicated that shells of both strains have the same set of SMPs, but the relative abundance of each SMP was considerably different between the two. The most abundant SMP in the dextral shells (comp88734_c0_seq1) is 3.13 times more abundant than that in the sinistral shells, while the most abundant SMP in the sinistral shells (comp88616_c0_seq1) is 2.88 times more abundant than that in the dextral shells. Furthermore, the SMP that indicated the largest difference of protein abundance between the dextral and sinistral shells was TRINITY_DN114176_c0_g1_il, which is 138 times more abundant in the dextral shells than in the sinistral shells. On the other hand, comp123694_c0_seq1 is 415 times more abundant in the sinistral shells than in the dextral shells (Supplementary Table 3.1). Among the SMPs that were annotated by BLAST, the following 9 SMPs indicated 10 times or more difference in the abundance between the dextral and sinistral shells : CD109 antigen-like protein (TRINITY_DN114176_c0_g1_il: 138 times; TRINITY_DN_552224_c0_g1_i1: 22.9 times), Slit-like protein (TRINITY_DN11292_c0_g1_i1: 27.2 times), myosin heavy chain (comp134017_c0_seq3: 26.3 times), and glutamate receptor 3-like (comp147489_c9_seq3: 17.7 times) (those 5 are more abundant in the dextral shells), and cytochrome b5 (comp123694_c0_seq1: 415 times), ferritin (comp131604_c0_seq1: 34.1 times), hemocyanin 2 (TRINITY_DN58078_c0_g1: 26.6 times), perlucin-like (comp1046506_c0_seq1: 22.2 times), and calmodulin-A-like (TRINITY_DN113838_c0_g1_i1: 17.6 times) (those 4 are more abundant in the sinistral

shells). The CD109 antigen has serine-type endopeptidase inhibitor activity and negatively regulates signaling by transforming growth factor beta in keratinocytes and in human (Finnson et al., 2006). The Slit protein is an extracellular matrix protein which has signaling role in the neural development in bilaterians, and has four leucine-rich repeat domains and a beta-sandwich domain similar to Laminin_G domain (Brose et al., 1999). The myosin heavy chain is a motor protein of muscle thick filaments and conserved in vertebrates and invertebrates. The calcium ion binding activity exists not in the myosin heavy chain but in the myosin light chain (Craig & Woodhead, 2006). The glutamate receptor 3 is a receptor for guanine nucleotide-binding proteins and has a calcium channel regulator activity in cell membrane (Makoff et al., 1996). The ferritin stores iron in a soluble, non-toxic, readily available form in extracellular region. In *L. stagnalis*, two groups of ferritin were found in soma and oocyte yolk (Bottke & Sinha, 1979; Andrews et al., 1992). A protein of with a sequence similar to soma ferritin was found as an SMP from the pearl oyster *Pinctada fucata*. The pearl oyster ferritin mRNA are expressed at the highest level at the mantle fold, which is an essential region for metal accumulation and contribution to metal incorporation (Y. Zhang et al., 2003). The hemocyanin is a copper-containing oxygen carrier occurring freely dissolved in the hemolymph of many molluscs and arthropods. The hemocyanin 2 is an isoform of hemocyanin, which is found from the keyhole limpet *Megathura crenulate* (Swerdlow et al., 1996). In this study, hemocyanin was found as an SMP of *L. stagnalis* (Chapter 2). The perlucin has C-type lectin domain, binding to D-galactose, D-mannose, and D-glucose, and was found from molluscan shells as an SMP. The perlucin was reported to have a potential to promote nucleation and/or growth of shell calcium carbonate crystals (K. Mann et al., 2000; Weiss et al., 2000; Wang et al., 2008). The calmodulin mediates the control of a large number of enzymes, ion channels and other proteins by calcium ions, and is known as an intracellular receptor. There are four EFh domains which are widely conserved in eukaryotes (Kaetzel & Dedman, 2003). The CD109 antigen, the Slit protein, and the myosin heavy chain are “room-keeping” proteins, the glutamate receptor 3 and the calmodulin are house-keeping proteins, and the ferritin, the hemocyanin 2, and the perlucin are known SMPs. It has been shown that the SMPs which are more abundant in the dextral than in the sinistral shells include three “room-keeping” proteins and one house-keeping protein. Concerning those “room-keeping” proteins, which are used for the maintenance of specific tissues or cells responsible for other functions than shell formation, their importance in shell formation needs to be verified by future studies. On the other hand, the SMPs which are more abundant in the sinistral than in the dextral shells include three known and one house-keeping proteins. Although a similar amount of the dextral and the sinistral shells (100 individuals each) was subjected to proteomic analysis, abundances of those known SMPs were higher in the sinistral shells than the

dextral shells. It appears possible that the protein concentration in the shell is higher in the sinistral shells than in the dextral shells in *L. stagnalis*. It is necessary to compare the shell thickness between the sinistral and dextral shells and to observe the arrangements of SMPs in the shell microstructure in both sinistral and dextral shells. Among those SMPs, perlucin-like, hemocyanin 2, and ferritin appear to be particularly important, because they indicated similar expression profiles in the dextral strain in Chapter 2. These SMPs were not identified as SMPs in Chapter 2, probably due to the differences in the sensitivity of the mass spectrometer used for the proteome analysis. Since the conditions for the transcriptome analysis differed between the coiling types (sinistral or dextral), direct comparisons of the expression levels are not possible. It is desired to perform the transcriptome analysis of sinistral and dextral individuals once again under the same conditions.

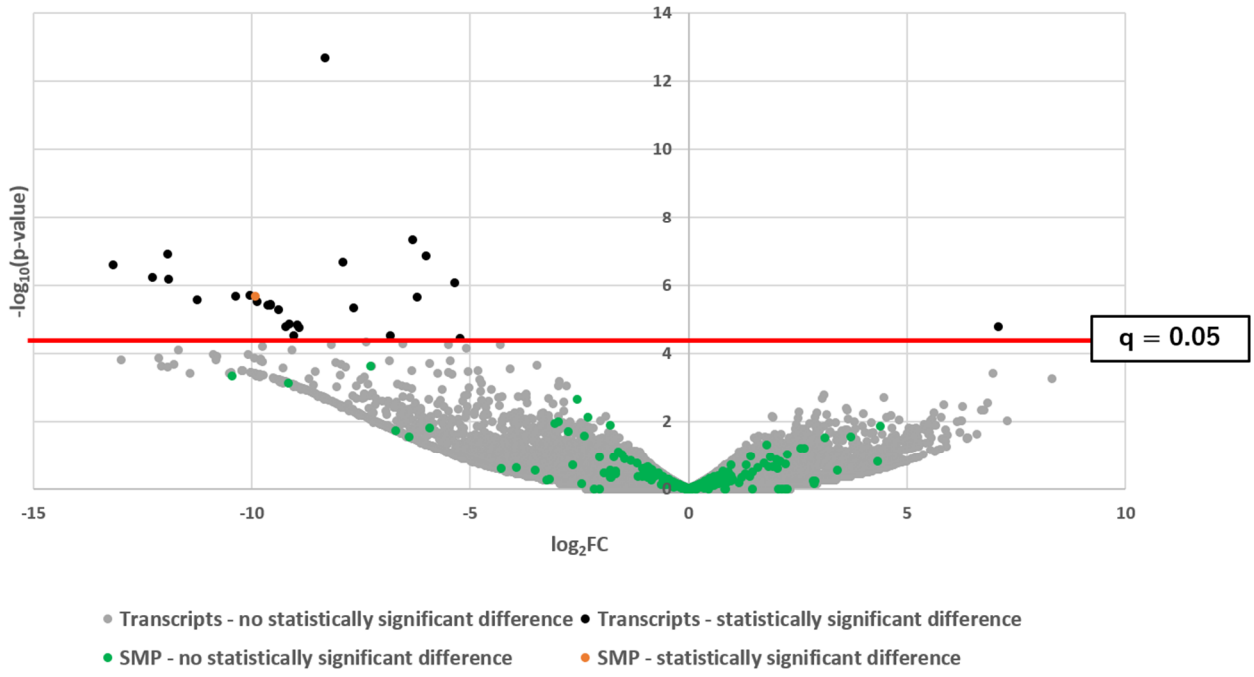

Figure 3.1.

Volcano plot showing differential expression of SMP-coding genes between right and left sides of mantle tissues of the sinistral strain. The X axis represents the logarithm of the change in expression levels of the right side vs. the left side. The Y axis represents the logarithm of the significance level for each comparison of the gene. The level of significance to reject the null hypothesis (q = 0.05) is shown as a red line.

3.2.3. Comparison of the expression levels of SMP genes between the right and left sides of mantle tissues in the sinistral snail

Next, data from the dextral and sinistral strains were used to compare the expression profiles of the SMP genes in the left and right sides of the mantle between the dextral and sinistral strains. Transcriptome data were compared using edgeR as in Chapter 2, and q value of 0.05 or less was judged as having a statistically significant difference. As a result, among the 37,395 transcripts, a total of 14 transcripts indicated a statistically significant difference in the levels of expression between the left and right sides of the mantle (Fig. 3.1). All those transcripts indicated a higher expression in the left side (outer side of the helix), and only one of which 1 represented an SMP. No transcripts, including those for SMPs, indicated statistically higher expression in the right side of the mantle (helical axis side) (Fig. 3.1).

Out of the 442 SMPs that indicated no statistically significant difference in gene expression levels between left and right sides of the mantle, a total of 177 SMP genes indicated higher expression (albeit without statistical significance) in the right than in the left, and 239 SMP genes indicated higher expression in the left than in the right. It seems that the expression levels of SMP genes are higher in the outer side of the helix than in the inner side in the sinistral strain. But the high levels of variations in gene expression between individuals appear to have hindered detection of statistically significant differences for many SMPs. The problem is how we can reduce the variations between individuals, and I will discuss in detail on this in Chapter 4. The remaining 27 SMP genes indicated zero logFC, having no expression in the sinistral strain or very low expression in the dextral strain. Those 27 SMPs are not found in the 207 SMPs identified in Chapter 2. It is suggested that even the SMPs with a very low abundance can be detected by using a reference database of enhanced quality. By combining whole genome analysis and transcriptome analysis, more accurate identification of SMP and analysis of its expression profile should be possible.

Thus, the expression patterns of the SMP genes in the left and right sides of the mantle does not mirror the patterns observed in the dextral strain in Chapter 2, but indicated a marked difference between the sinistral and dextral strains. The only SMP with a statistically significant difference is comp153562_c0_seq1, which was found to be a novel protein by BLAST. Although it is not included in the SMP identified in Chapter 2 in the dextral strain, it has two EFh (EF-hand) domains, thus has the same domain configuration as Ls-SMP-61 (comp137504_c0_seq2) and Ls-SMP-62 (comp137504_c0_seq3), which were identified as SMPs showing significantly higher gene expression in the right side than in the left side of the mantle in the dextral strain. EFh is a domain that has a function in cation-binding. Calmodulin, which has conserved EFh domains, is known as an intracellular protein, with a role to store calcium ions, and is

widely conserved among vertebrate and invertebrate animals (Friedberg & Rhoads, 2001). In molluscs, calmodulin was reported to be expressed in the mantle and in the gill of *P. fucata* (S. Li et al., 2004). Although a typical calmodulin has four EFh domains and no signal peptide, the three SMPs (Ls-SMP-61, Ls-SMP-62, and comp153562_c0_seq1) identified in this study have two EFh domains and signal peptides. Huang et al. (2007) reported that an EF-hand calcium-binding protein (EFCBP), which is an SMP with two EFh domains and a signal peptide, possibly has a role in promotion of shell formation because EFCBP is up-regulated during shell regeneration process (Huang et al., 2007). The other EFh domain containing protein named Sarcoplasmic Ca-binding Protein (SCP) from *P. fucata* is reported to control precipitation of calcium carbonate *in vitro* (Matsuura et al., 2018). Phylogenetic analysis among EFh containing SMPs, EFCBP, SCP, and calmodulin suggested that these three SMPs have a similar sequence with that of EFCBP distinct from SCP and calmodulin (Fig. 3.2, Fig. 3.3). Thus, comp153562_c0_seq1 in the sinistral strain and Ls-SMP-61 and Ls-SMP-62 in the dextral strain may have an important function, by virtue of possession of two EFh domains, possibly in promoting shell precipitation in the outer side of the helix (left side in the sinistral shells and right side in the dextral shells).

In the dextral strain, three times more SMPs genes indicated statistically higher expression levels in the left (spiral axis) side than in the right (outer side of the helix) side of the mantle, whereas in the sinistral strain, the only SMP gene that indicated a significant difference in the expression levels between left and right is expressed more strongly in the left (outer side of the helix) side of the mantle. This observation suggests, as a general role of SMPs in shell formation, that in the dextral shells, suppression of shell precipitation is more important than promotion of shell precipitation, whereas in the sinistral shells, promotion of shell precipitation is much more important than suppression of shell precipitation. The reasons for this discrepancy are unknown, awaiting further studies.

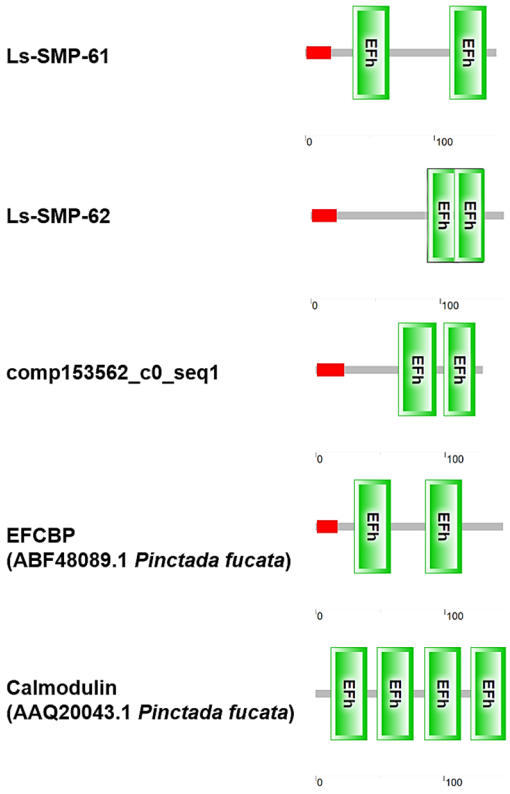

Figure 3.2.

Schematic representations of the conserved domain configuration of Ls-SMP-61, Ls-SMP-62, comp153562_c0_seq1, EFCBP of *Pinctada fucata*, and Calmodulin of *Pinctada fucata* (results from SMART domain searches).

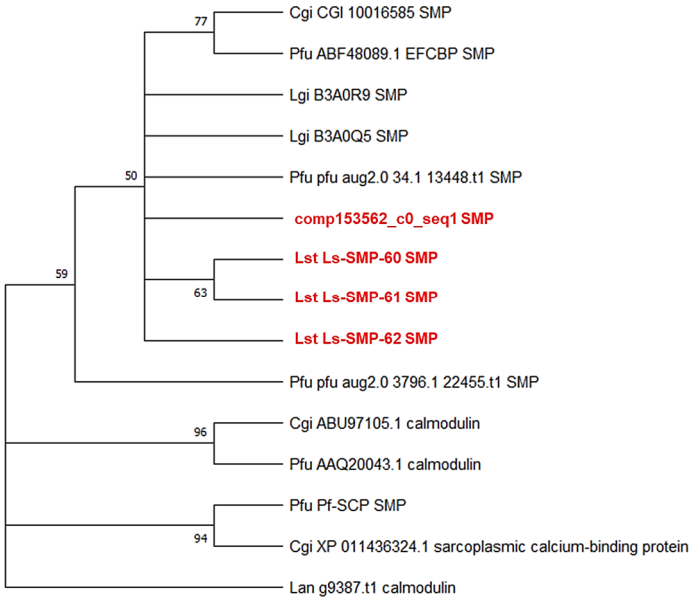

Figure 3.3.

Figure 3.3 (legend).

Maximum likelihood tree of the EFh containing proteins in molluscan proteins, including SMPs identified from *Lymnaea stagnalis* in this study, *Crassostrea gigas* by Marie et al. (2011) and Zhang et al. (2012), *Lottia gigantea* by Mann et al. (2012; 2014), and *Pinctada fucata* by Zhao et al. (2018). The ML tree was inferred from 15 proteins, including one protein of the brachiopod, *Lingula anatina* as an outgroup, using the WAG model based on 22 amino acids. Polychotomy results if the bootstrap value of the node is lower than 50%. Bootstrap values are indicated for nodes with a value greater than 50%. Molluscan EFh containing proteins were retrieved from NCBI GenBank, EMBL EnsemblMetazoa, or OIST genome browser after BLAST searches using Ls-SMP-60, Ls-SMP-61, Ls-SMP-62, and comp153562_c0_seq1 of *L. stagnalis* as a query. Sequence names in red indicate SMP sequences from *L. stagnalis*. Cgi: *Crassostrea gigas*, Pfu: *Pinctada fucata*, Lgi: *Lottia gigantea*, Lst: *Lymnaea stagnalis*, Lan: *Lingula anatina*.

	Dextral 1 Left mantle	Dextral 2 Left mantle	Dextral 3 Left mantle	Dextral 1 Right mantle	Dextral 2 Right mantle	Dextral 3 Right mantle
Ls-SMP-2 (comp111558_c0_seq1)	8700.633925	1333.488433	2880.940872	59.082928	54.413462	761.828672
Ls-SMP-19 (comp123495_c0_seq1)	6447.995485	3187.974714	2146.052803	100.347538	62.638881	833.386637
Ls-SMP-88 (comp141104_c0_seq1)	116.960184	25.566795	371.985197	7.254581	0.795632	28.774373
Ls-SMP-149 (comp146497_c0_seq3)	1624.288456	1230.074312	910.705266	8778.124852	9725.979662	2501.945605

	Sinistral 1 Left mantle	Sinistral 2 Left mantle	Sinistral 3 Left mantle	Sinistral 1 Right mantle	Sinistral 2 Right mantle	Sinistral 3 Right mantle
Ls-SMP-2 (comp111558_c0_seq1)	0	0	54.746397	284.957812	97.417102	428.078099
Ls-SMP-19 (comp123495_c0_seq1)	0	0.995312	34.97994	210.569239	83.876188	314.265813
Ls-SMP-88 (comp141104_c0_seq1)	1.813822	0.618836	0.568627	19.029445	27.100556	65.74113
Ls-SMP-149 (comp146497_c0_seq3)	0	0	111.965327	107.900158	139.606562	153.732572

Table 3.1.

The list of gene expression levels (FPKMs) of four SMPs (Ls-SMP-2, Ls-SMP-19, Ls-SMP-88, and Ls-SMP-149) in each the dextral and sinistral individuals.

Four SMPs (Ls-SMP-2, Ls-SMP-19, Ls-SMP-88, Ls-SMP-149) indicated a statistically significant difference in gene expression levels between right and left sides of the mantle in both transcriptome and qPCR analyses in the dextral snails (Chapter 2), and the expression patterns of them have been examined in the transcriptome data of the sinistral snails. The expression patterns of three SMPs (Ls-SMP-2, Ls-SMP-19, Ls-SMP-88), which showed higher expression in the left (spiral axis) side in the dextral strain, indicated the same patterns of higher expression in the spiral axis (right) side in the sinisral strain, although the differences are not statistically significant. However, the remaining SMP (Ls-SMP-149), which indicated higher expression in the right (outer side of helix) in the dextral strain, indicated higher expression in the spiral axis (right) side in the sinistral strain. As regard to Ls-SMP-2, Ls-SMP-19, and Ls-SMP-88, they may share the same function of suppressing shell formation in both dextral and sinistral strains. However, the expression patterns of Ls-SMP-149 relative to coiling direction are reversed in sinistral and dextral strains. This observation suggest that the Ls-SMP-149 gene could be a gene that is differentially expressed in the right side of the body axis regardless of the coiling direction. It is also possible that it has functions other than shell formation, such as maintaining the lateral asymmetry of the mantle (Table 3.1).

The results obtained in this study suggest that the expression profiles of SMPs in dextral and sinistral strains are not simple mirror images. Previous studies on the early development of *L. stagnalis* have shown that the coiling direction is practically determined at the 2-cell stage and that the physical position of blastomeres is important for the determination of the coiling direction (Asami et al., 2008; Kuroda et al., 2009; Davison et al., 2016). It has also been shown that Dpp, which is involved in the control of subsequent coiling direction of the shell, is expressed in a mirror symmetric way between the dextral and sinistral snails (Shimizu et al., 2013). From these facts, it was assumed that, in later developmental stages, the genes downstream of Dpp are expressed keeping a certain pattern of left-right asymmetry in the mantle, leading to mirror image expression profiles of SMP genes between dextral and sinistral snails. But, the observed profiles are clearly different between the dextral and sinistral strains. As Utsuno et al. (2011) showed, the shells of *L. stagnalis* are not mirror images between the dextral and the sinistral shells: the dextral shells are rather elongated and the sinistral shells have a wider and rounder aperture (Utsuno et al., 2011). It is considered that this is due to the differences in gene expression after the 16-cell stage, when the maternal effect becomes no longer effective.

Although detailed processes and mechanisms that brought about the different right-left asymmetric expression profiles of SMP genes between the dextral and sinistral strains, it appears possible that SMPs are responsible for the slight morphological differences between the dextral and sinistral shells. Since many more SMPs have been

shown to have statistically different expression levels between left and right sides of the mantle in the dextral strain than in the sinistral strain, dextral shells could require more control (inhibition, activation) of the growth of calcium carbonate crystals that constitute the shells than the sinistral shells to construct the final spiral structure of the shell. In addition, compared with the dextral shells, the sinistral shells show shorter shell length relative to the shell width, and can be considered as somewhat closer to a flat-coiled shells. In this regard, it would be interesting to see the SMP gene expression profiles in the planorbid *Biomphalaria glabrata*, a closely related species of *L. stagnalis*, because it has a flat-coiled shell, and is confirmed to be a sinistral species (Strong et al., 2008; Bouchet et al., 2017). If the SMPs of *L. stagnalis* are involved in the "fine-tuning" of the shell coiling direction, SMP genes in *B. glabrata* would show a similar expression profile to that of the sinistral strain of *L. stagnalis*, or even more symmetric profile in the right side and left side of the mantle. Transcriptome analysis of *B. glabrata* was performed by Lockyer et al. (2008) (Lockyer et al., 2008), but SMP genes are yet to be identified.

The results presented in this study provide a first step toward more accurate analyses, and some problems need to be fixed in future. First of all, there is a problem of intraspecific variations. More than 100 generations have passed since each of the dextral and sinistral strains was established as a pure strain, and thus it is possible that the genetic background has differentiated between the two strains by genetic drift. In this study, the dextral and the sinistral individuals were crossed, and each pure line re-established to minimize the differences in genetic background between the dextral and sinistral individuals. However, the experimental results showed that the transcriptome data tended to be affected by the differences in the condition of the individuals. To make a more accurate comparison of the dextral and sinistral strains, the heterozygous F1 generation born from each parent, which is obtained by crossing the dextral pure line and the sinistral pure line, should be used by utilizing the property of the delayed inheritance of the *L. stagnalis* coiling direction. By doing so, it is possible to obtain samples in which only the cytoplasmic transcripts and proteins responsible for the determination of the coiling direction are different. The dextral and sinistral mRNAs used in this study were extracted on different dates. They should be prepared on the same day and time, by growing the above F1 generation under the same conditions and perform RNA extractions at the same growth stage.

A second problem with the current study is that the transcriptome analyses of the dextral and sinistral strains were performed by using the Illumina Hiseq and Miseq, respectively. It has been considered that the data qualities of those analyses are comparable, but, the same analytical methods should be used in order to make more precise comparisons.

3.3. Conclusion

The comparative transcriptomic and proteomic analysis of the sinistral strain of *L. stagnalis* indicated somewhat unexpected results, showing contradictory profiles to those expected from the assumptions in Chapter 2. The estimation of the functionally important SMP appears more difficult. On the other hand, it opened up a drastically novel possibility that SMPs could be responsible for the morphological difference between the dextral and sinistral shells in *L. stagnalis*. Further analyses are obviously needed, but studies in this direction would be fruitful not only in deciphering the mechanisms of biomineralization, but also in gaining insights into the nature of left-right asymmetry in snails and in animals at large.

4. General Discussion

4.1. Quality of the transcriptomic and proteomic data

In order to understand the evolutionary diversity of molluscs, it would be essential to understand the molecular and genetic control mechanisms of shell biomineralization. In other words, it is necessary to have a perspective from both genes and proteins. By combining the transcriptome analysis of the mantle tissue and the proteome analysis of shells used in this study, more accurate primary structure analysis of SMP can be performed than by doing one of those analyses alone. Transcriptome analysis can comprehensively identify mRNAs in specific tissues, but it is difficult to narrow down those involved in shell formation, while shell proteomic analysis comprehensively identifies proteins contained in shells directly, but only fragmentary sequences can be obtained by mass spectrometry. However, by combining these two kinds of analyses, complete, or nearly complete nucleotide sequences encoding SMPs incorporated in the shell can be obtained.

There has been a proteome analysis using expressed sequence tag (EST) data, instead of transcriptome data, as a reference, but according to Lowe et al. (2017), EST can allow us to obtain only short sequences due to technical restrictions (Lowe et al., 2017). It has been reported, for example, that the average length of about 270,000 EST sequences of *Pinctada margaritifera* used by Joubert et al. (2010) and that of about 220,000 EST sequences of *Crassostrea gigas* used by Marie et al. (2011) were 214 bp and 413 bp, respectively (Joubert et al., 2010; Marie, Zanella-Cléon, et al., 2011) (Table 1.1).

On the other hand, the number of transcript sequences used as a reference for the proteomic analysis of the dextral shell in this study was 337,195, with an average length of 1,140 bp, and an N50 length of 2,828 bp. The number of sequences in the transcriptome data of the sinistral snails was 222,444, with an average length of 824 bp, and an N50 length of 1,750 bp. Thus, the sequences of the transcriptome data used in this study are much longer than those of EST data. The N50 value represents a weighted median length of contigs such that 50% of the entire assembly is contained in contigs larger than or equal to this value, and is often used as a statistic to evaluate the quality of assembly of -omics data, such as transcriptome data. Recent studies combining transcriptome and proteome analyses reported a total of 676,358 transcripts and 59 SMPs for *Cepaea nemoralis* (K. Mann & Jackson, 2014), 106,452 transcripts and 63 SMPs for *Mytilus coruscus* (Liao et al., 2015), 20,106 transcripts and 71 SMPs for *Mya truncate* (Arivalagan et al., 2016), and 74,293 transcripts and 54 SMPs for *Euhadra quaesita* (Shimizu et al., 2019). These studies did not report the N50 scores. It seems that the number of identified SMPs has no correlation with the transcriptome reference size. However, previous studies using a genome database as the reference reported a total of

259 SMPs for *Crassostrea gigas* (G. Zhang et al., 2012), 569 SMPs for *Lottia gigantea* (K. Mann et al., 2012), and 144 SMPs for *Pinctada fucata* (Liu et al., 2015) (Table 1.1). In this study, a total of 443 SMPs has been identified from a total of 337,195 + 222,444 transcripts. Although it is clear that use of genomic data for reference gives the most comprehensive results, a deep transcriptome analysis combined with proteome analysis performed in this study yielded comprehensive SMP data comparable to those of previous studies using genome databases. Moreover, the transcriptome analysis has a great advantage in that it allows us to obtain data of some *in vivo* conditions such as the gene expression levels. Therefore, combination of the whole genome and transcriptome analyses is a more effective approach.

Although Herlitze et al. (2018) analyzed the shell proteome of the same species as in the present study (*L. stagnalis*) using methods similar to the present study (Herlitze et al., 2018), the number of SMPs identified in their study (40) is much smaller than that in the present study (443). This difference may be explained by a number of reasons, including the differences in the analytical equipments used for mass spectrometry and DNA sequencing, as well as protocols for mRNA/protein extractions, transcriptome analysis, and proteome analysis. Comparing the results of MASCOT searches, the number of peptides detected in the proteomic analysis in their study (1,230) is smaller than that in the present study (8,061). Since the instruments used are different with each other [Q-Star XL nanospray quadrupole/time-of-flight tandem mass spectrometer, nanospray-Qq-TOF-MS/MS (Applied Biosystems, Villebon-sur-Yvette, France) coupled to an online nanoLC system (Ultimate Famos Switchos from Dionex, Amsterdam, Netherlands) in Herlitze et al. (2018); Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a UltiMate 3000 RSLCnano (Thermo Fisher Scientific, Waltham, MA, USA) in this study], detailed analytical conditions of peptide mass spectrometry cannot be compared, but it can be concluded that the analysis performed in this study has been much more sensitive, resulting in identification of many more SMPs from the same species.

In addition, because the LC-MS/MS analysis of proteins entails fragmentation of the proteins, typically by trypsin treatment, prior to the analysis, proteins that have no trypsin cleavage site (the basic amino acid residues of lysine and arginine) cannot be analyzed by this method. Thus, it should be noted that super-acidic proteins such as those found in bivalve shells and assumed to have important functions as SMP (Tsukamoto et al., 2004; Takeuchi et al., 2008; Isowa et al., 2012) can be overlooked by this method. Indeed, no SMPs like Aspein, which contains 60% of Asp, has been identified in this study. Comp132732_c0_seq1 is the most acidic SMP identified in this study. It is 129 AAs in length, with a molecular weight of 14,233, and a theoretical pI of 3.65. This protein is composed of 18.0% Asp and 10.9% Glu.

Another potential problem with SMP characterization concerns *de novo* assembling, which is used in *de novo* sequencing. In *de novo* assembling, sequence reads are computationally linked together, rather than aligned onto known sequences, and thus it is vulnerable to repeated sequences, which are often found in SMP genes. In order to solve this problem, acquisition of long reads using such platforms as those developed by Pacific Biosciences and Oxford Nanopore Technology, as well as elucidation of the whole genome sequences of *L. stagnalis* would be a useful approach.

Despite the above problems inherent to the proteomic study of SMPs, the qualities of the sequence data for both transcripts and proteins obtained in this study are arguably better than those of hitherto reported datasets of molluscan SMPs, providing a reliable basis for the current study and for any future studies.

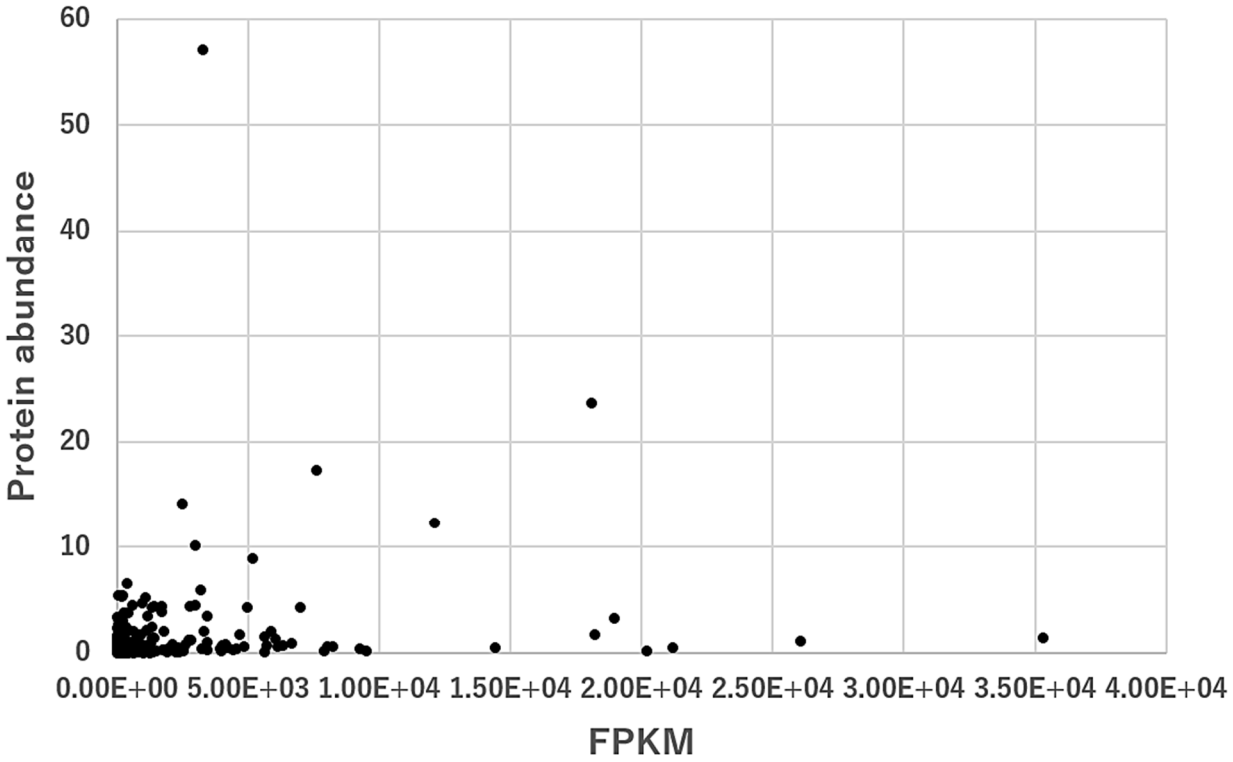

Figure 4.1.

Graph showing comparisons of gene expression levels (FPKM) and protein abundance for identified 207 SMPs from the dextral strains. The correlation coefficient between gene expression levels and protein abundance is 0.143.

4.2. General features of the SMPs identified in this study

The results of Chapter 2 in this study showed that, compared with the composition of the protein functions in the whole transcriptome of the mantle, the gene ontology of the identified 207 SMPs show more proportions of proteins with such molecular functions as ion binding, protein binding, carbohydrate derivative binding, oxidoreductase activity, and hydrolase activity. Ion binding is assumed to be responsible for the recruitment of calcium ions and interactions with calcium carbonate crystals, while protein binding, carbohydrate derivative binding is likely involved in the formation of supramolecular structures of shell organic matrices made of proteins and carbohydrates. On the other hand, oxidoreductase activity and hydrolase activity could be involved in the regulation of functionalities of SMPs, suggesting importance of the interactions between SMPs, although the details are unknown.

A closer look at the 207 SMPs revealed that not only known SMPs and novel proteins but also house-keeping proteins and room-keeping proteins which are found in other tissues than the mantle are contained. These house-keeping and room-keeping proteins tend to be very abundant at the transcript expression level but scarce at the shell protein level, that is, they are abundantly expressed in the mantle tissues but are not incorporated into the shells (correlation coefficient: 0.143, Fig. 4.1, Supplementary Table S2.4). It cannot be denied that all the SMPs identified in this study are involved, actively or passively, in shell formation, but the fact that the transcript sequences for SMPs annotated as house-keeping proteins, such as ribosomal proteins, do not have a signal peptide suggests that they do not contribute to shell formation as they are not secreted proteins. One of the initial motivations to carry out the current study (this thesis) was to develop a way to discriminate such apparently unfunctional SMPs from those potentially functional in shell formation. This goal has been achieved in Chapter 2 of this thesis, but further studies, including direct functional analyses of SMPs, are required to verify the results obtained in this study.

4.3. Asymmetry of snail shells as a mean to narrow down potentially functional SMPs

In this study, it was predicted that the morphological laterality of snail shells results in asymmetric gene expression patterns in the mantle responsible for shell formation, and the asymmetric gene expression patterns were indeed detected in 32 SMPs in the comparative transcriptome data for the left and right sides of the mantles. These 32 SMPs do not contain house-keeping proteins, but contain proteins with apparently important domains found in other SMPs identified in previous studies, such as ChtBD2, CLECT, and EFh (Suzuki et al., 2009; G. Zhang et al., 2012; K. Mann et al., 2012; R. Zhao et al., 2018). These observations strengthen the validity of the approach taken in this study, indicating that focusing on the asymmetry of the snail shells is appropriate

for narrowing down functionally important SMPs out of the numerous SMPs identified by a proteomic approach.

The 32 SMPs also include novel sequences that could not be annotated by BLAST or conserved domain searches, such as Ls-SMP-2 and Ls-SMP-19, and their left-right asymmetric expression was confirmed by qPCR. Proteins without specific motifs are usually difficult to pick up as an important SMP in conventional sequence-based studies, but I was able to pick them up by the approach developed in this study. Functions of these proteins cannot be estimated by their sequences alone, therefore future functional analysis is essential.

Comprehensive proteomic studies of SMPs have been made in gastropod molluscs (K. Mann et al., 2012, 2018; K. Mann & Jackson, 2014; Di et al., 2017; Herlitze et al., 2018; Shimizu et al., 2019), but no study focused on the morphological asymmetry of the shell. Herlitze et al. (2018) reported uneven gene expression patterns in the outer lip of the mantle of *L. stagnalis*, dividing the outer edge of the mantle tissue into left and right (Herlitze et al., 2018), but this superficial left-right axis actually corresponds to the anterior-posterior axis of the body.

Westbroek (1991) argued that polysaccharides control the formation of the complex structures made of calcite in the coccolith of the calcareous algae *Emiliania huxleyi* by inhibiting crystal growth. In molluscs, Wheeler et al. (1981) reported that the soluble matrix protein from the oyster *Crassostrea virginica* has the potential to regulate $CaCO_3$ deposition, where acidic proteins and peptides generally inhibit crystal nucleation and/or growth. Recent *in vitro* SMP functional analyses revealed that some SMPs have promotive effects, while others have inhibitory effects on $CaCO_3$ crystallization. For instance, the acetic acid-soluble matrix Caspartin inhibits calcium carbonate precipitation (Marin et al., 2005). Prismalin-14 also inhibits calcium carbonate precipitation (Suzuki & Nagasawa, 2007). Perlucin promotes the calcium carbonate precipitation and modification of crystal morphology at ambient conditions (Wang et al., 2008). The acidic SMP Pif is a key molecule in the induction of aragonite crystal formation (Suzuki et al., 2009). The basic SMP PfN23 accelerates the deposition of calcium carbonates and induces the formation of aragonite crystals (Fang et al., 2012). Lysine-rich SMP KRMP7 inhibits $CaCO_3$ precipitaion, changes the morphology of calcite, and inhibits the growth of aragonite (Liang et al., 2016).

These previous studies of *in vitro* functional analysis indicate that SMPs have roles in promotion/inhibition of crystallization and control at the molecular levels of biomineralization. Yonezawa et al. (2016) reported that chitinases might regulate the formation of the organic shell layer in *L. stagnalis* by *in vivo* functional suppression experiments (Yonezawa et al., 2016). In some of these previous studies, the relationships between SMPs and the microstructures such as the nacreous layer and the prismatic

layer have been discussed, but the relationships between SMPs and shell morphogenesis have not been discussed. The results shown in Chapter 2 confirmed the presence of asymmetrically expressed SMPs in both transcriptomic and qPCR analyses of gene expression between the right side (outer side of the shell helix) and the left side (axis side of the shell helix) of mantle tissues, showing that three times more SMPs are more highly expressed in the left than in the right in dextral shells (Fig. 2.17). On the other hand, in the sinistral strain studied in Chapter 3, only one SMP is expressed asymmetrically, with a higher expression in the left side of mantle (outer side of the shell helix). Whether promotion or inhibition of shell precipitation is generally more important, and exactly how shell precipitation is controlled in snail shell morphogenesis will be a next question to be answered. To this end, *in vivo* functional analysis would be useful.

4.4. Molecular evolution of the conserved domains in some SMPs

From the results of molecular phylogenetic analyses of the conserved domain sequences, Ls-SMP-88, which is more highly expressed in the axial than the outer side of the spiral of the shell, is considered to have lost the VWA domain. The VWA domain is a conserved domain integral to the SMP Pif. Suzuki et al. (2009) reported that Pif plays a role in promoting shell nacre formation (Suzuki et al., 2009), while the fact that the Pif-like Ls-SMP-88 is more highly expressed in the axial side suggests that it is involved in the inhibition of shell formation. One possibility is that Pif promotes shell formation, but if the VWA domain of Pif is lost, this Pif-like protein act as an inhibitor of shell formation. The VWA domain is known to be involved in multiprotein complexes. Probably, when Pif exists as multiprotein complexes and is anchored to the shell matrix framework, its acidic region can act as nucleation sites of calcium carbonate crystals, but when it is released from the framework and exists freely in the fluid, then it could act as an inhibitor of shell formation (cf. Wheeler et al. 1981).

Molecular phylogenetic analyses were also attempted for the Tyrosinase and WAP domain sequences contained in Ls-SMP-43 and Ls-SMP-149, respectively, and in other molluscan SMPs. However, the results were not very robust, with the best bootstrap branch support rate of around 60%. This may be because the domain sequences subjected to analysis were too short (82 aa for Tyrosinase and 35 aa for WAP domains).

In some SMPs, “conserved” domains have been shown to be recruited in SMPs independently in different lineages. For example, the carbonic anhydrase family of SMPs found in the bivalves *Crassostrea gigas* and *Pinctada fucata* and the gastropod *Lottia gigantea* has been reported to have experienced gene duplication in each species independently (R. Zhao et al., 2018). Therefore, caution must be exercised in linking the “conservation” of a domain across species with a single origin of the evolutionary acquisition of SMPs.

4.5. Differences in SMP gene expression profiles between the dextral and sinistral snails

It is clear from previous studies that the laterality of shell coiling direction in *L. stagnalis* is determined at the early developmental stage (Asami et al., 2008; Kuroda et al., 2009; Shimizu et al., 2011, 2013; Davison et al., 2016). But, since Utsuno et al. (2011) suggested that the dextral and sinistral shells do not have perfect mirror symmetry (Utsuno et al., 2011), it appears possible that biomineralization in late developmental stages controls shell coiling direction and morphogenesis. The results of Chapter 3 in this study revealed that the expression profiles of SMPs are significantly different between right and left strains having almost the same genetic background. This observation suggests that, the expression profile of SMPs in the mantle in later developmental stages is not directly induced by the genes that were expressed asymmetrically in the early developmental stages. The factors controlling the gene expression in the pathway of SMP induction need for further analysis.

Furthermore, by comparing the shell proteomes between the dextral and sinistral strains, it was suggested that some SMPs could be responsible for the differences in the shell morphologies between dextral and sinistral shells. In other words, the factor supporting the "F molecule" hypothesis (Brown & Wolpert, 1990) could be hidden in the SMPs. F molecules produce left-right asymmetry at the molecular level, which leads to the left-right asymmetry at the tissue/morphological level.

In this study, the only sinistral strain SMP gene that indicated a statistically significant difference (with a higher expression in the left) in the expression levels between left and right sides of the mantle encode a protein (comp153562_c0_seq1) with two EFh domains that can interact with cations. The same SMP was identified in the dextral strain, but it indicated no statistically significant difference in gene expression between left and right sides of the mantle. But other SMPs, namely, Ls-SMP-61 and Ls-SMP-62, show a conserved domain structure similar to ID（comp153562_c0_seq1）with double EFh domains, and indicate a statistically higher expression in the left side of the mantle in the dextral strain. These results indicate that those proteins with two EFh domains are highly expressed in the left side of the body axis regardless of the coiling types, suggesting that the factor acting as the F molecule is involved in the induction of these SMPs.

Answering to the question of how those SMPs are involved in shell formation and shell morphogenesis requires *in vivo* functional analysis and expression analysis using in situ hybridization. Gene knockout experiments in molluscs have been reported by Perry & Henry (2014) (on *Crepidula fornicata*) and by Abe & Kuroda (2019) (on *L. stagnalis*), and are considered applicable to SMP functional analysis (Perry & Henry, 2015; Abe & Kuroda, 2019).

Recent comprehensive identification revealed that many SMPs may complementarily

act in the formation. Therefore, in order to analyze the functions of SMPs and elucidate the molecular mechanisms of shell formation, the next step would be to investigate not only the expression levels but also the arrangements of SMPs in the dextral and sinistral shells and their relationships with the shell fine structures.

In the context of evolutionary developmental biology, the asymmetry of shell morphology provides not only a foothold for elucidating the molecular mechanisms of shell formation and evolution but also clues to understand evolution and development of lateral asymmetry in nature.

5. Material and Methods

5.1. Animals and protocol for extracting RNA

Pond snails, *Lymnaea stagnalis* (strain GSS7-1), originally collected in Neustadt, Donau, Germany, were reared in deionized and calcium carbonate-saturated tap water at about 23±1 °C (Fig. 1a). Mantle tissues were excised from three individuals, each having been separated into left and right portions using scissors (Fig. 1b). Foot tissues were excised from another individual using scissors. Total RNA was extracted from each sample using ISOGEN (Nippon Gene, Tokyo, Japan), following the protocol of Isowa et al. (2015) (Isowa et al., 2015).

In Chapter 3, the sinistral strain of *Lymnaea stagnalis* established and provided by Asami Laboratory in Shinshu University have been reared in the laboratory at 23±1 ºC. In order to "homogenize" the genetic background between the two, an individual from each strain was crossed with each other, and then the sinistral strain was re-established before subjected to transcriptome and proteome analyses.

5.2. Transcriptome analyses

Transcriptomic analyses of mantle and foot RNA were carried out using HiSeq 2500 (Illumina, California, USA) and MiSeq (Illumina, California, USA) sequencers, respectively. Twelve mantle libraries were prepared from left and right mantle tissues for each of three biological replicates from each the dextral and sinistral strains. The library for foot tissues was prepared from one individual each for the dextral and sinistral strains. Low-quality regions were removed from the FASTQ files before assembling the nucleotide sequences in FASTA format using Trinity (v2.5.1) (Grabherr et al., 2011; Haas et al., 2013) and the DDBJ Read Annotation Pipeline (http://p.ddbj.nig.ac.jp/pipeline/; last accessed April 5, 2018) (Kaminuma et al., 2009; Nagasaki et al., 2013) with the default setting of k-mer (= 25). Gene models were predicted using TransDecoder (v4.1.0) (Haas et al., 2013), and redundant sequences were removed with CD-HIT (v4.6.1) (W. Li & Godzik, 2006; Fu et al., 2012). Sequence assembly, gene set, and transcriptome completeness of the FASTA file were verified with BUSCO (v3.3) (Simão et al., 2015; Waterhouse et al., 2018) using the metazoan data set.

5.3. Protocol for extracting shell proteins

Shell proteins were extracted from 20 g of shells from about 60 individuals in Chapter 2, and from 30 g each of shells from about 100 individuals each of the dextral and sinistral strains in Chapter 3. First, shells were treated with sodium hypochlorite to remove any surface contaminants and then washed with ultrapure water to remove sodium hypochlorite. Second, washed shells were dissolved in an aqueous solution of 460 mL of

0.5 M ethylenediaminetetraacetic acid (EDTA), the pH adjusted to 8.0 using NaOH. After 48 h of incubation at 4°C, the supernatant was decanted, and EDTA was removed with Amicon ultra-15 centrifugal filter units (Merck Millipore, Burlington, USA). This EDTA-soluble protein mixture is denominated the soluble fraction. The precipitate was dissolved in an aqueous solution containing 7M urea, 2M thiourea, 3% 3-[(3-cholamidopropyl)dimetilaminio] propanesulfonate (CHAPS), and 1% Triton X-100), and after 24 h of incubation at 4°C, the preparation was centrifuged at 20,000g, at 4°C for 1 h. The supernatant was decanted, and urea and other salts were removed with Amicon ultra-15 centrifugal filter units (Merck Millipore, Burlington, USA). Th protein mixture solubilized by this procedure is referred to as the insoluble fraction. The amounts of soluble and insoluble fractions, in aqueous and CHAPS solutions, respectively, were quantified using a Qubit Fluorometer (Thermo Fisher Scientific, Massachusetts, USA).

5.4. Proteomic analysis

In Chapter 2, each of the protein samples extracted using the above protocol was dissolved in 200 µL of 0.1 M Tris-HCl buffer (pH 8.5). After adding 600 µL of methanol and 150 µL of chloroform, the sample was centrifuged (13,000 g, 4°C, 10 min.). After removing the supernatant, 500 µL of methanol was added to the sample and centrifuged again (20,400 g, 4°C, 10 min.). The supernatant was then removed, and the sample was dried by Speed Vac (EYELA, Tokyo, Japan). The dried pellet was dissolved in an aqueous solution containing 8M urea, 0.1 M Tris-HCl (pH 8.5) and 0.1 M DTT, and was incubated for 1 h at 37°C. A volume of 0.5 µL 208 mM iodoacetamide was added to the sample, which was incubated for 1 h at room temperature in the dark. After adding 0.1 M Tris-HCl and ultrapure water, it was treated with trypsin (Promega, Wisconsin, USA) in a 20-fold excess over sample protein and incubated overnight at 37°C. Tryptic peptides were analyzed with an LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled with a DiNa nanoLC system (KYA Technologies, Tokyo, Japan). Precursor ions were detected over a range of 400–1,500 m/z, and the top four high-intensity ions were selected for MS/MS analyses in data-dependent mode. Acquired MS/MS spectra were subjected to a database search against the protein sequence database translated from transcriptome data from the mantle tissues of *L. stagnalis* with the SEQUEST program using Proteome Discoverer software version 1.2 (Thermo Fisher Scientific, Waltham, MA, USA). Parameters were set as follows: the charge state of the precursor ions: automatically recognized; the mass range of tryptic peptides: 800 - 4,500; mass tolerances for precursor ions: 10 ppm; mass tolerances for fragment ions: 1 Da. Up to two missed cleavages and modifications of carbamidomethylation (+57.021) of cysteine and oxidation (+15.995) of methionine were considered for calculation of theoretical masses. False discovery rates (FDRs) were calculated based on a decoy database using

Proteome Discoverer software. A list of identified peptides that include a false discovery rate <1% was obtained after filtering low-confidence identification. Protein sequences identified by more than one unique peptide were retained.

In Chapter 3, proteomic analysis was performed by Medical ProteoScope Co., Ltd. using the following protocol. Each of the protein samples extracted using the above protocol was dissolved in 200 µL of 0.1 M Tris-HCl buffer (pH 8.5). After adding 600 µL of methanol and 150 µL of chloroform, the sample was centrifuged (13,000 g, 4°C, 10 min.). After removing the supernatant, 500 µL of methanol was added to the sample and centrifuged again (20,400 g, 4°C, 10 min.). The supernatant was then removed, and the sample was dried by a vacuum dryer. The dried pellet was dissolved in an aqueous solution containing 8M urea, 0.1 M Tris-HCl (pH 8.5) and 0.1 M DTT, and was incubated for 1 h at 37°C. An iodoacetamide was added to the sample, which was incubated for 1 h at room temperature in the dark. After adding 0.1 M Tris-HCl (pH 8.0) and ultrapure water, it was treated with trypsin in a 20-fold excess over sample protein and incubated 16 h at 37°C. Tryptic peptides were analyzed with an Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled with a UltiMate 3000 RSLCnano (Thermo Fisher Scientific, Waltham, MA, USA). Precursor ions were detected over a range of 300–1,500 m/z, and the top 10 high-intensity ions were selected for MS/MS analyses in data-dependent mode (Top 10 method). Mass resolutions were set 70,000 in MS, and 17,500 in MS/MS. Progenesis QI for proteomics (v2.0) (Waters Corporation, Milford, MA, USA) was used for spectra peak alignment and integration of each LC-MS/MS data and normalization of peak intensity. Acquired MS/MS spectra were subjected to a database search against the protein sequence database translated from transcriptome data from the mantle tissues of *L. stagnalis* with the MASCOT (v2.5) (Matrix Science Inc, Boston, MA, USA). Parameters were set as follows: peptide tolerance: ±5 ppm; MS/MS tolerance: ±0.02 Da. Up to two missed cleavages and modifications of carbamidomethylation (+57.021) of cysteine and oxidation (+15.099) of methionine were considered for calculation of theoretical masses. A list of identified peptides that include a false discovery rate (FDR) <1% was obtained after filtering low-confidence identification. Protein sequences identified by more than one unique peptide were retained.

5.5. Sequence annotation

I performed BLAST searches against GenBank at NCBI (National Center for Biotechnology Information; http://blast.ncbi.nlm.nih.gov; last accessed April 10, 2019) (Altschul et al., 1990; Boratyn et al., 2012) using the non-redundant database (all organisms) with an e-value cutoff of 10^{-10}. Conserved domains were searched using Simple Modular Architecture Research Tool (SMART; v8.0; http://smart.embl-

heidelberg.de; last accessed April 10, 2019) (Schultz et al., 1998; Letunic & Bork, 2018) provided by EMBL (European Molecular Biology Laboratory), including optional searches for outlier homologs and homologs of known structure, Pfam domains, and signal peptides. Gene Ontology (GO) terms were assigned using Blast2GO software (version 4.1; https://www.blast2go.com) (Conesa et al., 2005) for the categories cellular component, biological process, and molecular function. For each category, GO terms are assigned at 4 hierarchical levels, and term frequencies at three levels (levels 2 - 4) are visualized and considered in this study. Theoretical isoelectric points (pI) and theoretical molecular masses of identified proteins were estimated using UniProt (http://www.uniprot.org/; last accessed August 20, 2018) (Bateman, 2019). Potential O-glycosylation, N-glycosylation, and phosphorylation sites were predicted using the servers of DTU Bioinformatics (http://www.cbs.dtu.dk/services/; last accessed September 10, 2018).

5.6. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Soluble and insoluble fractions were separated by SDS-PAGE, in a 10% polyacrylamide slab gel. After electrophoresis, the gel was stained with Coomassie Brilliant Blue (CBB) to visualize proteins.

5.7. Phylogenetic analyses of some conserved domains

Molecular phylogenetic analyses have been conducted on conserved domains in Pif-like and tyrosinase-domain-containing SMPs (TDC-SMPs) identified in this study. The Pif-like SMP was analyzed because it contains ChtBD2 and Laminin_G domains, as in typical Pifs (Suzuki et al., 2009), but it lacks the VWA domain, unlike typical Pifs. Thus, the ChtBD2 and Laminin_G domain sequences of *L. stagnalis* have been compared with those of typical Pifs in other species, so as to infer how the Pif-like SMP of *L. stagnalis* originated. The TDC-SMPs were analyzed because two types of TDC-SMPs, namely those also containing Hemocyanin_bet_s domains and those without, were identified, and their relationships remain uncertain.

Conserved domains were identified by searches with the Simple Modular Architecture Research Tool (SMART; v8.0; http://smart.embl-heidelberg.de; last accessed April 10, 2019) (Schultz et al., 1998; Letunic & Bork, 2018) provided by EMBL (European Molecular Biology Laboratory), including optional searches for outlier homologs and homologs of known structure, Pfam domains, and signal peptides. Conserved domain sequences were aligned using ClustalW (Thompson et al., 1994) embedded in the phylogenetic analysis tool MEGA X (v10.1.7) (Kumar et al., 2018) with default settings. Resulting alignments were submitted to the trimAl (v1.2) (Capella-Gutiérrez et al., 2009) to remove poorly aligned regions and divergent regions of protein alignment while

allowing smaller final blocks, gap positions within the final blocks, and less strict flanking positions. The best-fit amino acid substitution model was inferred using the "find best protein models" function fitted in MEGA X, and phylogenetic analysis was performed using the maximum-likelihood (ML) method with bootstrap iterations of 1,000 replicates on MEGA X. Polychotomies were generated by collapsing nodes with a bootstrap value <50%.

5.8. Gene expression analysis

Analyses of gene expression levels were performed with Bowtie2 (v2.3.4.3; http://bowtie-bio.sourceforge.net/bowtie2/index.shtml; last accessed November 21, 2018) (Langmead & Salzberg, 2012) and eXpress (v1.5.0; https://pachterlab.github.io/eXpress/; last accessed November 21, 2018) (Roberts & Pachter, 2013) using the DDBJ Read Annotation Pipeline. Differential gene expression was evaluated using edgeR (v3.18.1; https://bioconductor.org/packages/release/bioc/html/edgeR.html; last accessed November 25, 2018) in the Bioconductor package of the R project (https://www.r-project.org/; last accessed November 25, 2018) (Robinson et al., 2009; Ramos et al., 2017). Statistical analysis of differential expression between right and left mantle tissues was performed based on the nonparametric Fisher's exact test, using edgeR. Statistical significance levels were calculated and corrected using the FDR (false discovery rate) to avoid false positives arising from multiple tests. The q value was set at 0.05.

5.9. Quantitative PCR (qPCR)

Total RNA was extracted from left and right mantle tissues of 5 biological replicates of *L. stagnalis* using ISOGEN (Nippon Gene, Tokyo, Japan) following the protocol of Isowa et al. (2015) (Isowa et al., 2015). Total RNA was also extracted from the foot tissue of one specimen of *L. stagnalis* using the method above. Total RNA samples were treated with DNase I using RQ1 RNase-Free DNase (Promega, Madison, USA). To confirm the effectiveness of the DNase I treatment, PCR was performed with primer sets for EF1α of *L. stagnalis* with total RNA samples as a template following the protocol of Young et al. (2019) (Young et al., 2019).

Complementary DNA was prepared from 1 μg of each total RNA sample using iScript RT Supermix for RT-qPCR (Bio-Rad, Hercules, USA), which contained a mixture of oligo(dT) and random primers. Reverse Transcription PCR was performed using a GeneAmp PCR system 9700 (Applied Biosystems, Foster City, USA) following the protocol of Young et al. (2019) (Young et al., 2019).

Quantitative PCR was performed using a StepOne Real-time PCR system (Applied Biosystems, Foster City, USA). Amplification of SMP genes was detected with SyBR Green dye. Reactions contained 2 μL of cDNA sample (25 ng/μL) with 5 μL of SsoAdvanced Universal SyBR Green

Mix (Bio-Rad, Hercules, USA), 250 nM forward and reverse primer concentrations, and were brought to 10 µL with DNase- and RNase-free water. Two technical replicates were performed for each of the 5 biological replicates. The custom qPCR method consisted of 95°C for 30s; 40 cycles of 95°C for 15s, 60°C for 30s. Gene expression was analysed with the comparative C_T method. Foot tissues were selected as references and *Lst-EF1a* (EF1α gene of *L. stagnalis*) designed by Young et al. (2019) (Young et al., 2019) was selected as an endogenous control. Primers for the 32 "asymmetric" SMP genes were designed using Primer3Plus (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi; last accessed November 11, 2019) to have a length of 19-23 nucleotides, a melting temperature between 57-62°C, a GC content from 40-60%, and a product size range of 70-150 bp (Supplementary Table S2.11). As a comparison, we also analysed three "symmetric" SMP genes, which indicated the highest FPKM values in the transcriptome data, by qPCR. Primers for those three SMP genes were also designed using the same method.

5.10. Data Availability

Transcriptome data from this study are available in the DDBJ Sequence Read Archive (DRA) under accession numbers DRA005517 and DRA006373.

6. Acknowledgements

I would like to show my greatest appreciation to my dissertation supervisor Professor Kazuyoshi Endo for constructive suggestions and encouragement for my research. I would like to express my sincere gratitude to my advisors, Associate Professor Takenori Sasaki (The University Museum, the University of Tokyo), Dr. Takanobu Tsuihiji (National Museum of Nature and Science), and Associate Professor Michio Suzuki (Graduate School of Agriculture, The University of Tokyo). My deepest appreciation goes to Drs. Keisuke Shimizu (the University of Tokyo), Yukinobu Isowa (the University of Tsukuba), Takeshi Takeuchi (Okinawa Institute of Science and Technology), Ran Zhao, (Shenzhen MSU-BIT University) and Professor Makiko Ishikawa (Yamazaki University of Animal Health Technology) for technical supports and constructive pieces of advice. I wish to express my gratitude to Akane Shingu, Tatsushi Kobayashi, Nanami Suzuki, and Akari Harada for many supports and encouragements. I would also like to express my deepest appreciation to Dr. Takashi Toyofuku (JAMSTEC) and Tatsuya Ogawa (National Museum of Nature and Science) for many supports to for inspiring me to have creative ideas.

This study has been supported through advice and assistance from all members of Paleobiology Lab., the University of Tokyo. Lastly, I wish to express my sincere thanks to my family for their patient support and encouragement during the course of my research.

7. References

Abe, M., & Kuroda, R. (2019). The development of CRISPR for a mollusc establishes the formin Lsdia1 as the long-sought gene for snail dextral/sinistral coiling. *Development*, **146**(9), dev175976. https://doi.org/10.1242/dev.175976

Ahlberg, P. E., Clack, J. A., & Blom, H. (2005). The axial skeleton of the Devonian tetrapod Ichthyostega. *Nature*, **437**(7055), 137–140. https://doi.org/10.1038/nature03893

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amorim, J., Abreu, I., Rodrigues, P., Peixoto, D., Pinheiro, C., Saraiva, A., Carvalho, A. P., Guimarães, L., & Oliva-Teles, L. (2019, June 15). *Lymnaea stagnalis* as a freshwater model invertebrate for ecotoxicological studies. *Science of the Total Environment*, Vol. **669**, pp. 11–28. https://doi.org/10.1016/j.scitotenv.2019.03.035

Andrews, S. C., Harrison, P. M., Yewdall, S. J., Arosio, P., Levi, S., Bottke, W., von Darl, M., Briat, J. F., Laulhère, J. P., & Lobreaux, S. (1992). Structure, function, and evolution of ferritins. *Journal of Inorganic Biochemistry*, **47**(1), 161–174. https://doi.org/10.1016/0162-0134(92)84062-R

Arivalagan, J., Marie, B., Sleight, V. A., Clark, M. S., Berland, S., & Marie, A. (2016). Shell matrix proteins of the clam, *Mya truncata*: Roles beyond shell formation through proteomic study. *Marine Genomics*, **27**, 69–74. https://doi.org/10.1016/j.margen.2016.03.005

Arthur, W. (1982). Control of shell shape in *Lymnaea Stagnalis*. *Heredity*, **49**(2), 153–161. https://doi.org/10.1038/hdy.1982.81

Asami, T., Gittenberger, E., & Falkner, G. (2008). Whole-Body Enantiomorphy and Maternal Inheritance of Chiral Reversal in the Pond Snail *Lymnaea stagnalis*. *Journal of Heredity*, **99**(5), 552–557. https://doi.org/10.1093/jhered/esn032

Ashley-Ross, M. A., Hsieh, S. T., Gibb, A. C., & Blob, R. W. (2013). Vertebrate Land Invasions-Past, Present, and Future: An Introduction to the Symposium. *Integrative and Comparative Biology*, **53**(2), 192–196. https://doi.org/10.1093/icb/ict048

Atli, G., & Grosell, M. (2016). Characterization and response of antioxidant systems in the tissues of the freshwater pond snail (*Lymnaea stagnalis*) during acute copper exposure. *Aquatic Toxicology*, **176**, 38–44. https://doi.org/10.1016/j.aquatox.2016.04.007

Ban, C., Ramakrishnan, B., Ling, K.-Y., Kung, C., & Sundaralingam, M. (1994). Structure of the recombinant *Paramecium tetraurelia* calmodulin at 1.68 Å

resolution. *Acta Crystallographica Section D Biological Crystallography*, **50**(1), 50–63. https://doi.org/10.1107/S0907444993007991

Bandow, C., & Weltje, L. (2012). Development of an embryo toxicity test with the pond snail *Lymnaea stagnalis* using the model substance tributyltin and common solvents. *Science of the Total Environment*, **435–436**, 90–95. https://doi.org/10.1016/j.scitotenv.2012.07.005

Bateman, A. (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Research*, **47**(D1), D506–D515. https://doi.org/10.1093/nar/gky1049

Bédouet, L., Marie, A., Dubost, L., Péduzzi, J., Duplat, D., Berland, S., Puisségur, M., Boulzaguet, H., Rousseau, M., Milet, C., & Lopez, E. (2007). Proteomics analysis of the nacre soluble and insoluble proteins from the oyster *Pinctada margaritifera*. *Marine Biotechnology*, **9**(5), 638–649. https://doi.org/10.1007/s10126-007-9017-1

Bento, I., Martins, L. O., Gato Lopes, G., Arménia Carrondo, M., & Lindley, P. F. (2005). Dioxygen reduction by multi-copper oxidases; a structural perspective. *Dalton Transactions*, **4**(21), 3507. https://doi.org/10.1039/b504806k

Berland, S., Marie, A., Duplat, D., Milet, C., Sire, J. Y., & Bédouet, L. (2011). Coupling Proteomics and Transcriptomics for the Identification of Novel and Variant Forms of Mollusk Shell Proteins: A Study with P. margaritifera. *ChemBioChem*, **12**(6), 950–961. https://doi.org/10.1002/cbic.201000667

Bingle, L., Singleton, V., & Bingle, C. D. (2002). The putative ovarian tumour marker gene HE4 (WFDC2), is expressed in normal tissues and undergoes complex alternative splicing to yield multiple protein isoforms. *Oncogene*, **21**(17), 2768–2773. https://doi.org/10.1038/sj.onc.1205363

Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, **7**(1), 12. https://doi.org/10.1186/1745-6150-7-12

Bottke, W., & Sinha, I. (1979). Ferritin as an exogenous yolk protein in snails. *Wilhelm Roux's Archives of Developmental Biology*, **186**(1), 71–75. https://doi.org/10.1007/BF00848109

Bouchet, P., Rocroi, J.-P., Hausdorf, B., Kaim, A., Kano, Y., Nützel, A., Parkhaev, P., Schrödl, M., & Strong, E. E. (2017). Revised Classification, Nomenclator and Typification of Gastropod and Monoplacophoran Families. *Malacologia*, **61**(1–2), 1–526. https://doi.org/10.4002/040.061.0201

Bouétard, A., Noirot, C., Besnard, A.-L., Bouchez, O., Choisne, D., Robe, E., Klopp, C., Lagadic, L., & Coutellec, M.-A. (2012). Pyrosequencing-based transcriptomic resources in the pond snail *Lymnaea stagnalis*, with a focus on genes involved in molecular response to diquat-induced stress. *Ecotoxicology*, **21**(8), 2222–2234. https://doi.org/10.1007/s10646-012-0977-1

Brose, K., Bland, K. S., Kuan, H. W., Arnott, D., Henzel, W., Goodman, C. S., Tessier-Lavigne, M., & Kidd, T. (1999). Slit proteins bind robo receptors and have an evolutionarily conserved role in repulsive axon guidance. *Cell*, **96**(6), 795–806. https://doi.org/10.1016/S0092-8674(00)80590-5

Brown, N. A., & Wolpert, L. (1990). The development of handedness in left/right asymmetry. *Development*, **109**(1).

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chateigner, D., Hedegaard, C., & Wenk, H. R. (2000). Mollusc shell microstructures and crystallographic textures. *Journal of Structural Geology*, **22**(11–12), 1723–1735. https://doi.org/10.1016/S0191-8141(00)00088-2

Clack, J. A. (1997, May 1). Devonian tetrapod trackways and trackmakers; a review of the fossils and footprints. *Palaeogeography, Palaeoclimatology, Palaeoecology*, Vol. **130**, pp. 227–250. https://doi.org/10.1016/S0031-0182(96)00142-3

Coates, M. I. (1996). The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod interrelationships and patterns of skeletal evolution. *Transactions of the Royal Society of Edinburgh: Earth Sciences*, **87**(3), 363–421. https://doi.org/10.1017/S0263593300006787

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**(18), 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Craig, R., & Woodhead, J. L. (2006, April 1). Structure and function of myosin filaments. *Current Opinion in Structural Biology*, Vol. **16**, pp. 204–212. https://doi.org/10.1016/j.sbi.2006.03.006

Davies, G., & Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**(9), 853–859. https://doi.org/10.1016/S0969-2126(01)00220-9

Davis, C. G. (1990). The many faces of epidermal growth factor repeats. *The New Biologist*, **2**(5), 410–419. Retrieved from http://europepmc.org/abstract/MED/2288911

Davison, A., McDowell, G. S., Holden, J. M., Johnson, H. F., Koutsovoulos, G. D., Liu, M. M., Hulpiau, P., Van Roy, F., Wade, C. M., Banerjee, R., Yang, F., Chiba, S., Davey, J. W., Jackson, D. J., Levin, M., & Blaxter, M. L. (2016). Formin Is Associated with Left-Right Asymmetry in the Pond Snail and the Frog. *Current Biology*, **26**(5), 654–660. https://doi.org/10.1016/j.cub.2015.12.071

de Vlieger, T. A., Kits, K. S., Lodder, J. C., ter Maat, A., & Lodder, J. C. (1980). Morphology and electrophysiology of the ovulation hormone producing neuro-

endocrine cells of the freshwater snail *Lymnaea stagnalis* (L.). *The Journal of Experimental Biology*, **84**(1), 259 LP – 271. Retrieved from http://jeb.biologists.org/content/84/1/259.abstract

Di, G., Kong, X., Miao, X., Zhang, Y., Huang, M., Gu, Y., You, W., Zhang, J., & Ke, C. (2017). Proteomic analysis of trochophore and veliger larvae development in the small abalone *Haliotis diversicolor*. *BMC Genomics*, **18**(1), 809. https://doi.org/10.1186/s12864-017-4203-7

Fang, D., Pan, C., Lin, H., Lin, Y., Zhang, G., Wang, H., He, M., Xie, L., & Zhang, R. (2012). Novel basic protein, PfN23, functions as key macromolecule during nacre formation. *Journal of Biological Chemistry*, **287**(19), 15776–15785. https://doi.org/10.1074/jbc.M112.341594

Feng, D., Li, Q., Yu, H., Kong, L., & Du, S. (2017). Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. *Scientific Reports*, **7**(1), 45754. https://doi.org/10.1038/srep45754

Feng, Z.-P., Zhang, Z., van Kesteren, R., Straub, V., van Nierop, P., Jin, K., Nejatbakhsh, N., Goldberg, J., Spencer, G., Yeoman, M., Wildering, W., Coorssen, J., Croll, R., Buck, L., Syed, N., & Smit, A. (2009). Transcriptome analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. *BMC Genomics*, **10**(1), 451. https://doi.org/10.1186/1471-2164-10-451

Finnson, K. W., Tam, B. Y. Y., Liu, K., Marcoux, A., Lepage, P., Roy, S., Bizet, A. A., Philip, A., Finnson, K. W., Tam, B. Y. Y., Liu, K., Marcoux, A., Lepage, P., Roy, S., Bizet, A. A., & Philip, A. (2006). Identification of CD109 as part of the TGF - ß receptor system in human keratinocytes. *The FASEB Journal*, **20**(9), 1525–1527. https://doi.org/10.1096/fj.05-5229fje

Freer, A., Bridgett, S., Jiang, J., & Cusack, M. (2014). Biomineral Proteins from *Mytilus edulis* Mantle Tissue Transcriptome. *Marine Biotechnology*, **16**(1), 34–45. https://doi.org/10.1007/s10126-013-9516-1

Friedberg, F., & Rhoads, A. R. (2001). Evolutionary Aspects of Calmodulin. *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)*, **51**(4), 215–221. https://doi.org/10.1080/152165401753311753

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**(23), 3150–3152. https://doi.org/10.1093/bioinformatics/bts565

Gao, P., Liao, Z., Wang, X., Bao, L., Fan, M., Li, X., Wu, C., & Xia, S. (2015). Layer-by-Layer Proteomic Analysis of *Mytilus galloprovincialis* Shell. *PLoS ONE*, **10**(7), e0133913. https://doi.org/10.1371/journal.pone.0133913

Gibbs, G. M., Roelants, K., & O'Bryan, M. K. (2008). The CAP Superfamily: Cysteine-

Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins—Roles in Reproduction, Cancer, and Immune Defense. *Endocrine Reviews*, **29**(7), 865–897. https://doi.org/10.1210/er.2008-0032

Glenner, H., Thomsen, P. F., Hebsgaard, M. B., Sørensen, M. V., & Willerslev, E. (2006, December 22). The origin of insects. *Science*, Vol. **314**, pp. 1883–1884. https://doi.org/10.1126/science.1129844

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**(7), 644–652. https://doi.org/10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., … Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, **8**(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Herlitze, I., Marie, B., Marin, F., Jackson, D. J., & Jackson, D. J. (2018). Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. *GigaScience*, **7**(May 2018), 1–15. https://doi.org/10.1093/gigascience/giy056/4997018

Hierck, B. P., Witte, B., Poelmann, R. E., Gittenberger-de Groot, A. C., & Gittenberger, E. (2005). Chirality in snails is determined by highly conserved asymmetry genes. *Journal of Molluscan Studies*, **71**(2), 192–195. https://doi.org/10.1093/mollus/eyi022

Huang, J., Zhang, C., Ma, Z., Xie, L., & Zhang, R. (2007). A novel extracellular EF-hand protein involved in the shell formation of pearl oyster. *Biochimica et Biophysica Acta - General Subjects*. https://doi.org/10.1016/j.bbagen.2007.03.006

Isowa, Y., Sarashina, I., Oshima, K., Kito, K., Hattori, M., & Endo, K. (2015). Proteome analysis of shell matrix proteins in the brachiopod *Laqueus rubellus*. *Proteome Science*, **13**(1), 1–10. https://doi.org/10.1186/s12953-015-0077-2

Isowa, Y., Sarashina, I., Setiamarga, D. H. E., & Endo, K. (2012). A comparative study of the shell matrix protein Aspein in pterioid bivalves. *Journal of Molecular Evolution*, **75**(1–2), 11–18. https://doi.org/10.1007/s00239-012-9514-3

Iwamoto, S., Shimizu, K., Negishi, L., Suzuki, N., Nagata, K., & Suzuki, M. (2020). Characterization of the chalky layer-derived EGF-like domain-containing protein (CgELC) in the pacific oyster, *Crassostrea gigas*. *Journal of Structural Biology*, **212**(1), 107594. https://doi.org/10.1016/j.jsb.2020.107594

Jackson, D. J., Mann, K., Häussermann, V., Schilhabel, M. B., Lüter, C., Griesshaber,

E., Schmahl, W., & Wörheide, G. (2015). The *Magellania venosa* biomineralizing proteome: A window into brachiopod shell evolution. *Genome Biology and Evolution*, **7**(5), 1349–1362. https://doi.org/10.1093/gbe/evv074

Jackson, D. J., McDougall, C., Woodcroft, B., Moase, P., Rose, R. A., Kube, M., Reinhardt, R., Rokhsar, D. S., Montagnani, C., Joubert, C., Piquemal, D., & Degnan, B. M. (2010). Parallel Evolution of Nacre Building Gene Sets in Molluscs. *Molecular Biology and Evolution*, **27**(3), 591–608. https://doi.org/10.1093/molbev/msp278

Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F., Zanella-Cléon, I., Cochennec-Laureau, N., Gueguen, Y., & Montagnani, C. (2010). Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: Focus on biomineralization. *BMC Genomics*, **11**(1), 613. https://doi.org/10.1186/1471-2164-11-613

Kaetzel, M. A., & Dedman, J. R. (2003). Calmodulin. In *Encyclopedia of Hormones* (pp. 241–245). https://doi.org/10.1016/B0-12-341103-3/00041-3

Kaminuma, E., Mashima, J., Kodama, Y., Gojobori, T., Ogasawara, O., Okubo, K., Takagi, T., & Nakamura, Y. (2009). DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Research*, **38**(SUPPL.1), D33–D38. https://doi.org/10.1093/nar/gkp847

Kato, S., Matsui, T., Gatsogiannis, C., & Tanaka, Y. (2018). Molluscan hemocyanin: structure, evolution, and physiology. *Biophysical Reviews*, **10**(2), 191–202. https://doi.org/10.1007/s12551-017-0349-4

Kelley, P. H., Kowalewski, M., & Hansen, T. A. (2003). *Predator—Prey Interactions in the Fossil Record* (P. H. Kelley, M. Kowalewski, & T. A. Hansen, Eds.). https://doi.org/10.1007/978-1-4615-0161-9

Kimura, S., & Ikeda - Saito, M. (1988). Human myeloperoxidase and thyroid peroxidase, two enzymes with separate and distinct physiological functions, are evolutionarily related members of the same gene family. *Proteins: Structure, Function, and Bioinformatics*, **3**(2), 113–120. https://doi.org/10.1002/prot.340030206

Knoll, A. H. (2003). Biomineralization and Evolutionary History. *Reviews in Mineralogy and Geochemistry*, **54**(1), 329–356. https://doi.org/10.2113/0540329

Kocot, K. M., Cannon, J. T., Todt, C., Citarella, M. R., Kohn, A. B., Meyer, A., Santos, S. R., Schander, C., Moroz, L. L., Lieb, B., & Halanych, K. M. (2011). Phylogenomics reveals deep molluscan relationships. *Nature*, **477**(7365), 452–456. https://doi.org/10.1038/nature10382

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and*

Evolution, **35**(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Kuroda, R., Endo, B., Abe, M., & Shimizu, M. (2009). Chiral blastomere arrangement dictates zygotic left-right asymmetry pathway in snails. *Nature*, **462**(7274), 790–794. https://doi.org/10.1038/nature08597

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4), 357–359. https://doi.org/10.1038/nmeth.1923

Letunic, I., & Bork, P. (2018). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, **46**(D1), D493–D496. https://doi.org/10.1093/nar/gkx922

Li, H., & Poulos, T. L. (1994). Structural variation in heme enzymes: a comparative analysis of peroxidase and P450 crystal structures. *Structure*, **2**(6), 461–464. https://doi.org/10.1016/S0969-2126(00)00046-0

Li, S., Xie, L., Zhang, C., Zhang, Y., Gu, M., & Zhang, R. (2004). Cloning and expression of a pivotal calcium metabolism regulator: Calmodulin involved in shell formation from pearl oyster (*Pinctada fucata*). *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, **138**(3), 235–243. https://doi.org/10.1016/j.cbpc.2004.03.012

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Liang, J., Xie, J., Gao, J., Xu, C. Q., Yan, Y., Jia, G. C., Xiang, L., Xie, L. P., & Zhang, R. Q. (2016). Identification and Characterization of the Lysine-Rich Matrix Protein Family in *Pinctada fucata*: Indicative of Roles in Shell Formation. *Marine Biotechnology*, **18**(6), 645–658. https://doi.org/10.1007/s10126-016-9724-6

Liao, Z., Bao, L. fei, Fan, M. hua, Gao, P., Wang, X. xing, Qin, C. li, & Li, X. min. (2015). In-depth proteomic analysis of nacre, prism, and myostracum of Mytilus shell. *Journal of Proteomics*, **122**, 26–40. https://doi.org/10.1016/j.jprot.2015.03.027

Liu, C., Li, S., Kong, J., Liu, Y., Wang, T., Xie, L., & Zhang, R. (2015). In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Scientific Reports*, **5**, 1–14. https://doi.org/10.1038/srep17269

Lockyer, A. E., Spinks, J., Kane, R. A., Hoffmann, K. F., Fitzpatrick, J. M., Rollinson, D., Noble, L. R., & Jones, C. S. (2008). *Biomphalaria glabrata* transcriptome: cDNA microarray profiling identifies resistant- and susceptible-specific gene expression in haemocytes from snail strains exposed to *Schistosoma mansoni*. *BMC Genomics*, **9**(1), 634. https://doi.org/10.1186/1471-2164-9-634

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., & Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology*, **13**(5). https://doi.org/10.1371/journal.pcbi.1005457

Lowenstam, H. A., & Weiner, S. (1989). *On Biomineralization*. Oxford University Press.

Luo, Y. J., Takeuchi, T., Koyanagi, R., Yamada, L., Kanda, M., Khalturina, M., Fujie, M., Yamasaki, S. I., Endo, K., & Satoh, N. (2015). The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nature Communications*, **6**, 1–10. https://doi.org/10.1038/ncomms9301

Makoff, A., Volpe, F., Lelchuk, R., Harrington, K., & Emson, P. (1996). Molecular characterization and localization of human metabotropic glutamate receptor type 3. *Molecular Brain Research*, **40**(1), 55–63. https://doi.org/10.1016/0169-328X(96)00037-X

Mann, K., Cerveau, N., Gummich, M., Fritz, M., Mann, M., & Jackson, D. J. (2018). In-depth proteomic analyses of *Haliotis laevigata* (greenlip abalone) nacre and prismatic organic shell matrix. *Proteome Science*, **16**(1), 11. https://doi.org/10.1186/s12953-018-0139-3

Mann, K., Edsinger-Gonzales, E., & Mann, M. (2012). In-depth proteomic analysis of a mollusc shell: Acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. *Proteome Science*, **10**(1), 1–18. https://doi.org/10.1186/1477-5956-10-28

Mann, K., & Jackson, D. J. (2014). Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics*, **15**(1), 249. https://doi.org/10.1186/1471-2164-15-249

Mann, K., Maček, B., & Olsen, J. V. (2006). Proteomic analysis of the acid-soluble organic matrix of the chicken calcified eggshell layer. *Proteomics*, *6*(13), 3801–3810. https://doi.org/10.1002/pmic.200600120

Mann, K., Poustka, A. J., & Mann, M. (2008). The sea urchin (*Strongylocentrotus purpuratus*) test and spine proteomes. *Proteome Science*, **6**, 1–10. https://doi.org/10.1186/1477-5956-6-22

Mann, K., Weiss, I. M., André, S., Gabius, H. J., & Fritz, M. (2000). The amino-acid sequence of the abalone (*Haliotis laevigata*) nacre protein perlucin: Detection of a functional C-type lectin domain with galactose/mannose specificity. *European Journal of Biochemistry*, **267**(16), 5257–5264. https://doi.org/10.1046/j.1432-1327.2000.01602.x

Mann, S. (2001). *Biomineralization*. Oxford University Press.

Marie, B., Jackson, D. J., Ramos-Silva, P., Zanella-Cléon, I., Guichard, N., & Marin, F. (2013). The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS Journal*, **280**(1), 214–232. https://doi.org/10.1111/febs.12062

Marie, B., Joubert, C., Tayalé, A., Zanella-Cleón, I., Belliard, C., Piquemal, D., Cochennec-Laureau, N., Marin, F., Gueguen, Y., & Montagnani, C. (2012). Different secretory repertoires control the biomineralization processes of prism and

nacre deposition of the pearl oyster shell. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(51), 20986–20991. https://doi.org/10.1073/pnas.1210552109

Marie, B., Le Roy, N., Zanella-Cléon, I., Becchi, M., & Marin, F. (2011). Molecular evolution of mollusc shell proteins: Insights from proteomic analysis of the edible mussel mytilus. *Journal of Molecular Evolution*, **72**(5–6), 531–546. https://doi.org/10.1007/s00239-011-9451-6

Marie, B., Marie, A., Jackson, D. J., Dubost, L., Degnan, B. M., Milet, C., & Marin, F. (2010). Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Proteome Science*, **8**(1), 54. https://doi.org/10.1186/1477-5956-8-54

Marie, B., Zanella-Cléon, I., Guichard, N., Becchi, M., & Marin, F. (2011). Novel Proteins from the Calcifying Shell Matrix of the Pacific Oyster *Crassostrea gigas*. *Marine Biotechnology*, **13**(6), 1159–1168. https://doi.org/10.1007/s10126-011-9379-2

Marin, F., Amons, R., Guichard, N., Stigter, M., Hecker, A., Luquet, G., Layrolle, P., Alcaraz, G., Riondet, C., & Westbroek, P. (2005). Caspartin and calprismin, two proteins of the shell calcitic prisms of the Mediterranean fan mussel *Pinna nobilis*. *Journal of Biological Chemistry*, **280**(40), 33895–33908. https://doi.org/10.1074/jbc.M506526200

Marin, F., Bundeleva, I., Takeuchi, T., Immel, F., & Medakovic, D. (2016). Organic matrices in metazoan calcium carbonate skeletons: Composition, functions, evolution. *Journal of Structural Biology*, **196**(2), 98–106. https://doi.org/10.1016/j.jsb.2016.04.006

Marin, F., Luquet, G., Marie, B., & Medakovic, D. (2007). Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. In *Current Topics in Developmental Biology* (Vol. **80**, pp. 209–276). https://doi.org/10.1016/S0070-2153(07)80006-8

Marxen, J. C., Nimtz, M., Becker, W., & Mann, K. (2003). The major soluble 19.6 kDa protein of the organic shell matrix of the freshwater snail *Biomphalaria glabrata* is an N-glycosylated dermatopontin. *Biochimica et Biophysica Acta - Proteins and Proteomics*, **1650**(1–2), 92–98. https://doi.org/10.1016/S1570-9639(03)00203-6

Matsuura, A., Yoshimura, K., Kintsu, H., Atsumi, T., Tsuchihashi, Y., Takeuchi, T., Satoh, N., Negishi, L., Sakuda, S., Asakura, T., Imura, Y., Yoshimura, E., & Suzuki, M. (2018). Structural and functional analyses of calcium ion response factors in the mantle of *Pinctada fucata*. *Journal of Structural Biology*, **204**(2), 240–249. https://doi.org/10.1016/j.jsb.2018.08.014

McDougall, C., & Degnan, B. M. (2018). The evolution of mollusc shells. *Wiley Interdisciplinary Reviews: Developmental Biology*, **7**(3), e313. https://doi.org/10.1002/wdev.313

Meshcheryakov, V. N. (1990). The Common Pond Snail *Lymnaea stagnalis*. In *Animal*

Species for Developmental Studies (pp. 69–132). https://doi.org/10.1007/978-1-4613-0503-3_5

Miyamoto, H., Endo, H., Hashimoto, N., Limura, K., Isowa, Y., Kinoshita, S., Kotaki, T., Masaoka, T., Miki, T., Nakayama, S., Nogawa, C., Notazawa, A., Ohmori, F., Sarashina, I., Suzuki, M., Takagi, R., Takahashi, J., … Watabe, S. (2013). The Diversity of Shell Matrix Proteins: Genome-Wide Investigation of the Pearl Oyster, *Pinctada fucata*. *Zoological Science*, **30**(10), 801. https://doi.org/10.2108/zsj.30.801

Miyamoto, H., Miyshita, T., Okushima, M., Nakano, A., Morita, T., & Matsushiro, A. (1996). A carbonic anhydrase from the nacreous layer in oyster pearls. *Proceedings of the National Academy of Sciences*, **93**(September), 9657–9660. https://doi.org/93(18), 9657-9660

Montagnani, C., Marie, B., Marin, F., Belliard, C., Riquet, F., Tayalé, A., Zanella-Cléon, I., Fleury, E., Gueguen, Y., Piquemal, D., & Cochennec-Laureau, N. (2011). Pmarg-Pearlin is a Matrix Protein Involved in Nacre Framework Formation in the Pearl Oyster *Pinctada margaritifera*. *ChemBioChem*, **12**(13), 2033–2043. https://doi.org/10.1002/cbic.201100216

Munley, K. M., Brix, K. V., Panlilio, J., Deforest, D. K., & Grosell, M. (2013). Growth inhibition in early life-stage tests predicts full life-cycle toxicity effects of lead in the freshwater pulmonate snail, *Lymnaea stagnalis*. *Aquatic Toxicology*, **128–129**, 60–66. https://doi.org/10.1016/j.aquatox.2012.11.020

Nagasaki, H., Mochizuki, T., Kodama, Y., Saruhashi, S., Morizaki, S., Sugawara, H., Ohyanagi, H., Kurata, N., Okubo, K., Takagi, T., Kaminuma, E., & Nakamura, Y. (2013). DDBJ read annotation pipeline: A cloud computing-based pipeline for high-throughput analysis of next-generation sequencing data. *DNA Research*, **20**(4), 383–390. https://doi.org/10.1093/dnares/dst017

Nagasawa, H. (2013). The molecular mechanism of calcification in aquatic organisms. *Bioscience, Biotechnology and Biochemistry*, **77**(10), 1991–1996. https://doi.org/10.1271/bbb.130464

Nelson, R. E., Fessler, L. I., Takagi, Y., Blumberg, B., Keene, D. R., Olson, P. F., Parker, C. G., & Fessler, J. H. (1994). Peroxidasin: a novel enzyme-matrix protein of Drosophila development. *The EMBO Journal*, **13**(15), 3438–3447. https://doi.org/10.1002/j.1460-2075.1994.tb06649.x

Okamoto, O., & Fujiwara, S. (2006). Dermatopontin, a novel player in the biology of the extracellular matrix. *Connective Tissue Research*, **47**(4), 177–189. https://doi.org/10.1080/03008200600846564

Okamoto, O., Hozumi, K., Katagiri, F., Takahashi, N., Sumiyoshi, H., Matsuo, N., Yoshioka, H., Nomizu, M., & Fujiwara, S. (2010). Dermatopontin promotes epidermal keratinocyte adhesion via α3β1 integrin and a proteoglycan receptor.

Biochemistry, **49**(1), 147–155. https://doi.org/10.1021/bi901066f

Patel, V. V., Zhao, L., Wong, P., Pradhan, B. B., Bae, H. W., Kanim, L., & Delamarter, R. B. (2006). An in vitro and in vivo analysis of fibrin glue use to control bone morphogenetic protein diffusion and bone morphogenetic protein-stimulated bone growth. *Spine Journal*, **6**(4), 397–403. https://doi.org/10.1016/j.spinee.2005.11.006

Perry, K. J., & Henry, J. Q. (2015). CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. *Genesis*, **53**(2), 237–244. https://doi.org/10.1002/dvg.22843

Piaget, J. (1929). *Les races lacustres de la" Limnaea stagnalis" L.: recherches sur les rapports de l'adaptation héréditaire avec le milieu*.

Ponder, W., & Lindberg, D. R. R. (2008). *Phylogeny and Evolution of the Mollusca* (First; W. Ponder & D. R. R. Lindberg, Eds.). Retrieved from https://www.ucpress.edu/book/9780520250925/phylogeny-and-evolution-of-the-mollusca

Ramos, M., Schiffer, L., Re, A., Azhar, R., Basunia, A., Rodriguez, C., Chan, T., Chapman, P., Davis, S. R., Gomez-Cabrero, D., Culhane, A. C., Haibe-Kains, B., Hansen, K. D., Kodali, H., Louis, M. S., Mer, A. S., Riester, M., … Waldron, L. (2017). Software for the Integration of Multiomics Experiments in Bioconductor. *Cancer Research*, **77**(21), e39–e42. https://doi.org/10.1158/0008-5472.CAN-17-0344

Rittschof, D., & McClellan-Green, P. (2005). Molluscs as multidisciplinary models in environment toxicology. *Marine Pollution Bulletin*, **50**(4), 369–373. https://doi.org/10.1016/j.marpolbul.2005.02.008

Roberts, A., & Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. *Nature Methods*, **10**(1), 71–73. https://doi.org/10.1038/nmeth.2251

Robertson, R. (1993). Snail handedness. *Research Explore*, **9**(1), 104–119.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Ruggeri, Z. M. (1993). Von Willebrand factor and fibrinogen. *Current Opinion in Cell Biology*, **5**(5), 898–906. https://doi.org/10.1016/0955-0674(93)90041-N

Sadamoto, H., Takahashi, H., Okada, T., Kenmoku, H., Toyota, M., & Asakawa, Y. (2012). *De novo* sequencing and transcriptome analysis of the central nervous system of Mollusc *Lymnaea stagnalis* by deep RNA sequencing. *PLoS ONE*, **7**(8), e42546. https://doi.org/10.1371/journal.pone.0042546

Sarashina, I., & Endo, K. (1998). Primary structure of a soluble matrix protein of scallop shell: Implications for calcium carbonate biomineralization. *American Mineralogist*, **83**(11), 1510–1515. https://doi.org/10.2138/AM-1998-1112

Sarashina, I., & Endo, K. (2001). The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. *Marine Biotechnology*, **3**(4), 362–369. https://doi.org/10.1007/s10126-001-0013-6

Sarashina, I., & Endo, K. (2006). Skeletal matrix proteins of invertebrate animals: Comparative analysis of their amino acid sequences. *Paleontological Research*, **10**(4), 311–336. https://doi.org/10.2517/prpsj.10.311

Sarashina, I., Yamaguchi, H., Haga, T., Iijima, M., Chiba, S., & Endo, K. (2006). Molecular evolution and functionally important structures of molluscan dermatopontin: Implications for the origins of molluscan shell matrix proteins. *Journal of Molecular Evolution*, **62**(3), 307–318. https://doi.org/10.1007/s00239-005-0095-2

Schmidt-Nielsen, K. (1984). *Scaling*. https://doi.org/10.1017/CBO9781139167826

Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(11), 5857–5864. https://doi.org/10.1073/pnas.95.11.5857

Sharon, N., & Lis, H. (2001). The structural basis for carbohydrate recognition by lectins. *Advances in Experimental Medicine and Biology*, **491**, 1–16. https://doi.org/10.1007/978-1-4615-1267-7_1

Shashoua, V. E. (1991). Ependymin, a Brain Extracellular Glycoprotein, and CNS Plasticity. *Annals of the New York Academy of Sciences*, **627**(1 Activity-Driv), 94–114. https://doi.org/10.1111/j.1749-6632.1991.tb25916.x

Shen, X., Belcher, A. M., Hansma, P. K., Stucky, G. D., & Morse, D. E. (1997). Molecular cloning and characterization of Lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *Journal of Biological Chemistry*, **272**(51), 32472–32481. https://doi.org/10.1074/jbc.272.51.32472

Shen, Z., & Jacobs-Lorenat, M. (1998). A type I peritrophic matrix protein from the malaria vector *Anopheles gambiae* binds to chitin. Cloning, expression, and characterization. *Journal of Biological Chemistry*, **273**(28), 17665–17670. https://doi.org/10.1074/jbc.273.28.17665

Shimizu, K., Iijima, M., Setiamarga, D. H. E., Sarashina, I., Kudoh, T., Asami, T., Gittenberger, E., & Endo, K. (2013). Left-right asymmetric expression of dpp in the mantle of gastropods correlates with asymmetric shell coiling. *EvoDevo*, **4**(1), 15. https://doi.org/10.1186/2041-9139-4-15

Shimizu, K., Kimura, K., Isowa, Y., Oshima, K., Ishikawa, M., Kagi, H., Kito, K., Hattori, M., Chiba, S., & Endo, K. (2019). Insights into the Evolution of Shells and Love Darts of Land Snails Revealed from Their Matrix Proteins. *Genome Biology*

and Evolution, **11**(2), 380–397. https://doi.org/10.1093/gbe/evy242

Shimizu, K., Sarashina, I., Kagi, H., & Endo, K. (2011). Possible functions of Dpp in gastropod shell formation and shell coiling. *Development Genes and Evolution*, **221**(2), 59–68. https://doi.org/10.1007/s00427-011-0358-4

Shu, D. G., Luo, H. L., Conway Morris, S., Zhang, X. L., Hu, S. X., Chen, L., Han, J., Zhu, M., Li, Y., & Chen, L. Z. (1999). Lower Cambrian vertebrates from south China. *Nature*, **402**(6757), 42–46. https://doi.org/10.1038/46965

Shubin, N. H., Daeschler, E. B., & Coates, M. I. (2004). The Early Evolution of the Tetrapod Humerus. *Science*, **304**(5667), 90–93. https://doi.org/10.1126/science.1094295

Simakov, O., Marletaz, F., Cho, S. J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D. H., Larsson, T., Lv, J., Arendt, D., Savage, R., Osoegawa, K., De Jong, P., Grimwood, J., Chapman, J. A., Shapiro, H., Aerts, A., … Rokhsar, D. S. (2013). Insights into bilaterian evolution from three spiralian genomes. *Nature*, **493**(7433), 526–531. https://doi.org/10.1038/nature11696

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smith, S. A., Wilson, N. G., Goetz, F. E., Feehery, C., Andrade, S. C. S., Rouse, G. W., Giribet, G., & Dunn, C. W. (2011). Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, **480**(7377), 364–367. https://doi.org/10.1038/nature10526

Sottrup-Jensen, L., Folkersen, J., Kristensen, T., & Tack, B. F. (1984). Partial primary structure of human pregnancy zone protein: Extensive sequence homology with human α2-macroglobulin. *Proceedings of the National Academy of Sciences of the United States of America*, **81**(23 I), 7353–7357. https://doi.org/10.1073/pnas.81.23.7353

Strong, E. E., Gargominy, O., Ponder, W. F., & Bouchet, P. (2008, January 18). Global diversity of gastropods (Gastropoda; Mollusca) in freshwater. *Hydrobiologia*, Vol. **595**, pp. 149–166. https://doi.org/10.1007/s10750-007-9012-6

Sturtevant, A. H. (1923). Inheritance of direction of coiling in Limnaea. *Science*, **58**(1501), 269–270.

Sucov, H. M., Benson, S., Robinson, J. J., Britten, R. J., Wilt, F., & Davidson, E. H. (1987). A lineage-specific gene encoding a major matrix protein of the sea urchin embryo spicule. II. Structure of the gene and derived sequence of the protein. *Developmental Biology*, **120**(2), 507–519. https://doi.org/10.1016/0012-1606(87)90254-5

Sudo, S., Fujikawa, T., Nagakura, T., Ohkubo, T., Sakaguchi, K., Tanaka, M., Nakashima, K., & Takahashi, T. (1997). Structures of mollusc shell framework proteins [6]. *Nature*, **387**(6633), 563–564. https://doi.org/10.1038/42391

Suzuki, M., Iwashima, A., Kimura, M., Kogure, T., & Nagasawa, H. (2013). The Molecular Evolution of the Pif Family Proteins in Various Species of Mollusks. *Marine Biotechnology*, **15**(2), 145–158. https://doi.org/10.1007/s10126-012-9471-2

Suzuki, M., Iwashima, A., Tsutsui, N., Ohira, T., Kogure, T., & Nagasawa, H. (2011). Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. *ChemBioChem*, **12**(16), 2478–2487. https://doi.org/10.1002/cbic.201100317

Suzuki, M., & Nagasawa, H. (2007). The structure-function relationship analysis of Prismalin-14 from the prismatic layer of the Japanese pearl oyster, *Pinctada fucata*. *FEBS Journal*, **274**(19), 5158–5166. https://doi.org/10.1111/j.1742-4658.2007.06036.x

Suzuki, M., Saruwatari, K., Kogure, T., Yamamoto, Y., Nishimura, T., Kato, T., & Nagasawa, H. (2009). An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science*, **325**(5946), 1388–1390. https://doi.org/10.1126/science.1173793

Swerdlow, R. D., Ebert, R. F., Lee, P., Bonaventura, C., & Miller, K. I. (1996). Keyhole limpet hemocyanin: Structural and functional characterization of two different subunits and multimers. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, **113**(3), 537–548. https://doi.org/10.1016/0305-0491(95)02091-8

Takeuchi, T., Kawashima, T., Koyanagi, R., Gyoja, F., Tanaka, M., Ikuta, T., Shoguchi, E., Fujiwara, M., Shinzato, C., Hisata, K., Fujie, M., Usami, T., Nagai, K., Maeyama, K., Okamoto, K., Aoki, H., Ishikawa, T., … Satoh, N. (2011). Draft Genome of the Pearl Oyster *Pinctada fucata*: A Platform for Understanding Bivalve Biology. *DNA Research*, **18**(1), 12–14. https://doi.org/10.1093/dnares

Takeuchi, T., Koyanagi, R., Gyoja, F., Kanda, M., Hisata, K., Fujie, M., Goto, H., Yamasaki, S., Nagai, K., Morino, Y., Miyamoto, H., Endo, K., Endo, H., Nagasawa, H., Kinoshita, S., Asakawa, S., Watabe, S., … Kawashima, T. (2016). Bivalve-specific gene expansion in the pearl oyster genome: implications of adaptation to a sessile lifestyle. *Zoological Letters*, **2**(1), 3. https://doi.org/10.1186/s40851-016-0039-2

Takeuchi, T., Sarashina, I., Iijima, M., & Endo, K. (2008). *In vitro* regulation of $CaCO_3$ crystal polymorphism by the highly acidic molluscan shell protein Aspein. *FEBS Letters*, **582**(5), 591–596. https://doi.org/10.1016/j.febslet.2008.01.026

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the

sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Tsukamoto, D., Sarashina, I., & Endo, K. (2004). Structure and expression of an unusually acidic matrix protein of pearl oyster shells. *Biochemical and Biophysical Research Communications*, **320**(4), 1175–1180. https://doi.org/10.1016/j.bbrc.2004.06.072

Upadhyay, A., & Thiyagarajan, V. (2016). Proteomic characterization of oyster shell organic matrix proteins (OMP). *Bioinformation*, **12**(05), 266–278. https://doi.org/10.6026/97320630012266

Utsuno, H., Asami, T., Van Dooren, T. J. M., & Gittenberger, E. (2011). Internal Selection Against the Evolution of Left-Right Reversal. *Evolution*, **65**(8), 2399–2411. https://doi.org/10.1111/j.1558-5646.2011.01293.x

Vermeiji, G. J. (1987). *Evolution and Escalation*. Retrieved from https://press.princeton.edu/books/paperback/9780691000800/evolution-and-escalation

Vinet, L., & Zhedanov, A. (2011). A "missing" family of classical orthogonal polynomials. *Journal of Physics A: Mathematical and Theoretical*, **44**(8), 085201. https://doi.org/10.1088/1751-8113/44/8/085201

Vinther, J., Sperling, E. A., Briggs, D. E. G., & Peterson, K. J. (2012). A molecular palaeobiological hypothesis for the origin of aplacophoran molluscs and their derivation from chiton-like ancestors. *Proceedings of the Royal Society B: Biological Sciences*, **279**(1732), 1259–1268. https://doi.org/10.1098/rspb.2011.1773

Wang, N., Lee, Y. H., & Lee, J. (2008). Recombinant perlucin nucleates the growth of calcium carbonate crystals: Molecular cloning and characterization of perlucin from disk abalone, *Haliotis discus discus*. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, **149**(2), 354–361. https://doi.org/10.1016/j.cbpb.2007.10.007

Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, **35**(3), 543–548. https://doi.org/10.1093/molbev/msx319

Weiss, I. M., Kaufmann, S., Mann, K., & Fritz, M. (2000). Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*. *Biochemical and Biophysical Research Communications*, **267**(1), 17–21. https://doi.org/10.1006/bbrc.1999.1907

Westbroek, P. (1991). *Life as a Geological Force : Dynamics of the Earth*. WW Norton & Co.

Wheeler, A. P., George, J. W., & Evans, C. A. (1981). Control of Calcium Carbonate Nucleation and Crystal Growth by Soluble Matrx of Oyster Shell. *Science*, **212**(4501), 1397–1398. https://doi.org/10.1126/science.212.4501.1397

Williams, G. A., Wang, Y., Callon, K. E., Watson, M., Lin, J., Lam, J. B. B., Costa, J. L., Orpe, A., Broom, N., Naot, D., Reid, I. R., & Cornish, J. (2009). *In Vitro* and *in Vivo* Effects of Adiponectin on Bone. *Endocrinology*, **150**(8), 3603–3610. https://doi.org/10.1210/en.2008-1639

Yonezawa, M., Sakuda, S., Yoshimura, E., & Suzuki, M. (2016). Molecular cloning and functional analysis of chitinases in the fresh water snail, *Lymnaea stagnalis*. *Journal of Structural Biology*, **196**(2), 107–118. https://doi.org/10.1016/j.jsb.2016.02.021

Young, A. P., Landry, C. F., Jackson, D. J., & Wyeth, R. C. (2019). Tissue-specific evaluation of suitable reference genes for RT-qPCR in the pond snail, *Lymnaea stagnalis*. *PeerJ*, **7**(10), e7888. https://doi.org/10.7717/peerj.7888

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., Xiong, Z., Que, H., Xie, Y., Holland, P. W. H., Paps, J., Zhu, Y., Wu, F., … Wang, J. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. *Nature*, **490**(7418), 49–54. https://doi.org/10.1038/nature11413

Zhang, Y., Meng, Q., Jiang, T., Wang, H., Xie, L., & Zhang, R. (2003). A novel ferritin subunit involved in shell formation from the pearl oyster (*Pinctada fucata*). *Comparative Biochemistry and Physiology Part B*, **135**(1), 43–54. https://doi.org/10.1016/S1096-4959(03)00050-2

Zhao, G., Huang, B.-L., Rigueur, D., Wang, W., Bhoot, C., Charles, K. R., Baek, J., Mohan, S., Jiang, J., & Lyons, K. M. (2018). CYR61/CCN1 Regulates Sclerostin Levels and Bone Maintenance. *Journal of Bone and Mineral Research*, **33**(6), 1076–1089. https://doi.org/10.1002/jbmr.3394

Zhao, R., Takeuchi, T., Luo, Y.-J. J., Ishikawa, A., Kobayashi, T., Koyanagi, R., Villar-Briones, A., Yamada, L., Sawada, H., Iwanaga, S., Nagai, K., Satoh, N., & Endo, K. (2018). Dual gene repertoires for larval and adult shells reveal molecules essential for molluscan shell formation. *Molecular Biology and Evolution*, **35**(11), 2751–2761. https://doi.org/10.1093/molbev/msy172