Molecular characterisation of biomineralising genes in the freshwater pond snail *Lymnaea stagnalis*

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm Geowissenschaften/Geographie

der Georg-August University School of Science (GAUSS)

vorgelegt von

Ines Herlitze

aus Rudolstadt

Göttingen 2017

Betreuungsausschuss:

Prof. Dr. Daniel J. Jackson, Abteilung Geobiologie, Fakultät für Geowissenschaften, Georg-August-Universität Göttingen

Dr. Frédéric Marin, BioME group, Laboratoire Biogéosciences, Universite de Bourgogne, Dijon, France.

Mitglieder der Prüfungskommission:

Referent:
Prof. Dr. Daniel J. Jackson, Abteilung Geobiologie, Fakultät für Geowissenschaften, Georg-August-Universität Göttingen

Korreferent:
Dr. Frédéric Marin, BioME group, Laboratoire Biogéosciences, Universite de Bourgogne, Dijon, France.

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Gernot Arp, Abteilung Geobiologie, Fakultät für Geowissenschaften, Georg-August-Universität Göttingen

Prof. Dr. Alexander Schmidt, Abteilung Geobiologie, Fakultät für Geowissenschaften, Georg-August-Universität Göttingen

Prof. Dr. Gregor Bucher, Abteilung Entwicklungsbiologie, Johann-Friedrich-Blumenbach- Institut für Zoologie und Anthropologie, Georg-August-Universität Göttingen

Dr. Nico Posnien, Abteilung Entwicklungsbiologie, Johann-Friedrich- Blumenbach-Institut für Zoologie und Anthropologie, Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 14.09.2017

Versicherung

Hiermit versichere ich an Eides statt, dass die Dissertation mit dem Titel „Molecular characterisation of biomineralising genes in the freshwater pond snail *Lymnaea stagnalis*" selbständig und ohne unerlaubte Hilfe angefertigt wurde.

Göttingen, den 09.08.2017

Acknowledgements

Contents

Abstract

Much of the evolutionary success of conchiferan molluscs can be traced back to the ability to construct a biocalcified shell that supports the soft molluscan body and offers protection from predation and desiccation. Only a small fraction of the shell is comprised of organic material, but these components play an important role in conferring the shells material properties and finally define shape, size and colour of the structure. Technical advances in nucleic acid sequencing and high-throughput proteomics promoted the identification of the proteinaceous components of the shell on the molecular level and enabled the collection of a large number of mineralising genes and proteins. Most studies focused on marine species with nacro-prismatic shells such as the pearl oyster *Pinctada* and the abalone *Haliotis*. To get a better understanding of the shell secretome diversity, it is important to extend the research focus to species that compose microstructures other than narco-prismatic. In my thesis, I present the shell proteome of the crossed-lamellar type. To date, the functional characterisation of the proteinaceous shell components still represents a major challenge in the field of molluscan biomineralisation. I gained insight into the function of shell-forming genes by ontogenetically and spatially characterising the crossed lamellar shell proteome of the freshwater gastropod *Lymnaea stagnalis*. The expression showed patterns of asymmetry in the shell-forming cells of larvae and hint at the potential for interactions between co-expressed genes. They furthermore revealed that many adult shell-forming genes are already present in larvae, foreshadowing the zonation of the adult mantle. I furthermore analysed the genomic architecture and tissue specific expression patterns of these genes and propose that alternative splicing significantly contributes to the molecular diversity of the *L. stagnalis* shellome.

1

General introduction

1.1 Molluscs

Molluscs have conquered almost all aquatic and terrestrial environments and are with more than 100 000 extant species the second largest phylum in the animal kingdom [1]. Their morphological appearance can vary greatly; from microscopic clams and snails to giant squids with tentacles longer than ten meter or clams weighting more than 200 kg [2, 3]. Much of the evolutionary success of molluscs can be traced back to the ability to construct a calcified shell, i.e., the protective structure present in most molluscan clades that has emerged during the late pre-Cambrian [4]. The secretion of the shell secures the support of the soft body tissue and protects the animal from predation or desiccation [5]. Modifications of this process have facilitated the evolution of a vast diversity of shapes, sizes and colourations and fascinate scientists and naturalists since centuries [6, 7].

1.2 Biomineralisation

The formation of the molluscan shell is one of many examples of biomineralisation, which is "the process by which organisms form minerals" [8]. As a result of this biologically influenced mineralisation, the minerals can appear quite different than their inorganic counterpart [9]. Biomineralisation is found in all five kingdoms and the mineralising organisms are able to form more than 60 different minerals. More than half of the biominerals are calcium based, but also phosphate, silica or iron oxides are commonly found [8]. The structures range from small-scale magnetosomes and spicules to shells, scales, bones and teeth up to mile long reef systems. The functions biominerals fulfil are not limited to structural support and protection. Tasks like grinding, cutting, filtration, light harvesting, gravity perception, magnetic field guidance and many more are accomplished by mineralised hard parts [10]. Mineralised products can also have a huge impact on the environment, such as ocean or freshwater chemistry as well as on the evolutionary history of the organisms that produce them. [11-13].

1.2.1 Biomineralisation and the advent of metazoan diversity

The first metazoan crown groups appeared during the Neoproterozoic 1000 to 541 million years ago (mya). The major diversification and the increase in morphological disparity of metazoan life occurred much later during the Cambrian 541 to 515 mya [14, 15]. Explanations for this rapid diversification event, the so-called Cambrian explosion, focus on genetic and developmental capacities of the taxa, as well as on changes in the abiotic and biotic environment [16]. Biomineralisation became an important factor in a complex interplay of these biotic and abiotic processes. In the Proterozoic world, skeletonising metazoans are documented only in a few cases. This changed radically during the Cambrian, when many metazoan groups developed the ability to form mineralised hard parts for the first time [13].

Evidence for the emergence of molluscs can be found in the small shelly fauna, a fossil assemblage of the Precambrian-Cambrian boundary that comprises representatives of molluscs and other metazoan phyla. Most likely, molluscs root back to the Precambrian [16]. Evidence for a Precambrian molluscan representative might be *Kimberella quadrata*, a prominent fossil of the Ediacaran biota, which is suspected to be a non- mineralising mollusc with a univalved shell [17].

The rise of predators and the associated need for protection is a renowned explanation for the emergence of the various forms of skeletal hard parts, as this has led to an escalatory arms race and has fuelled the metazoan diversification process [10, 15]. Another reason for the thriving metazoan biomineralisation may have been the radically increased calcium concentrations in the Cambrian oceans [15]. To cope with this environmental stress, organisms may have modified existing calcium detoxifying mechanisms, which in the end has led to the ability to form highly beneficial calcified structures [18, 19]. Either way, during the evolutionary history of metazoans, biomineralisation became a vital part of animal life.

1.2.2 Biomineralisation strategies

The influence that an organism exerts over the mineralisation process can vary greatly. Based on the level of control, the processes are classified as either 'biologically induced' or 'biologically controlled' [20, 21].

Mineralisation "as a result of the interactions between biological activity and the environment is termed 'biologically induced' mineralisation" [9]. In this case, the organism that promotes mineral precipitation does not posses a particular cellular machinery to induce mineralisation and precipitation occurs in the open environment. Depending on the surrounding, the same organism could induce the formation of different minerals [8]. Common sites of biologically induced nucleation are cell walls or polymeric materials like slimes or biofilms [22]. Not always is the emerged mineral beneficial for the organism. For example in the case of kidney stones, which is a common instance of pathological mineralisation [8, 23].

In contrast to the passive nature of the 'biologically induced' mineralisation, organisms use "cellular activities to direct the nucleation, growth, morphology and final location of the mineral that is deposited" in the case of 'biologically controlled' mineralisation [9]. In order to control the composition of the mother liquor, a fundamental feature of 'biologically controlled' mineralisation is to seal the site of mineralisation from the environment in form of lipid bilayers or water insoluble macromolecules. These delineated spaces are necessary to create an oversaturated solution, which is a cornerstone for crystal nucleation [8].

The site of 'biologically controlled' mineralisation can either be inter-, intra- or extracellular [9]. 'Biologically controlled intercellular mineralisation' is not a common phenomenon. This type of mineralisation seems similar to 'biologically induced mineralisation' but contrary to the unregulated induced process the organisms are indeed able to control the shape and the polymorph that is formed [9]. An example is the marine algae *Halimeda* that forms aragonite needles within intercellular spaces, which are completely isolated from the external medium [24]. When the biomineral is formed within the cell, the nucleation process is classified as 'biologically controlled intracellular mineralisation'. It often occurs in vesicles or vacuoles. A well-known example is the extrusion of intracellularly formed coccolith structures by Haptophyte algae [9, 25]. The mineralisation strategy employed by

molluscs to form the shell is a 'biologically controlled extracellular mineralisation' [26]. The site on which nucleation occurs is an extracellular matrix that has been secreted by the cells. This matrix consists of proteins, polysaccharides or glycoproteins which self assemble to form a three dimensional framework [9].

1.3 The molluscan shell

The formation of the molluscan shell is an example of biomineralisation, which has created a cornucopia of beauty and comes in all kind of shapes and sizes. People have appreciated this beauty throughout history and put high values to some specimen of molluscan biomineralisation. In the past, cowrie shells have served as currency in some cultures and today pearls are popular components of precious jewellery [27]. Some extraordinary pieces even reach revenues in the millions of dollars [28]. Since pearl cultivation is a profitable business, much attention has been devoted to the research of the formation of mother of pearl. The interest in the mechanisms that generate the shell has led to the collection of diverse transcriptome and proteome scale shell-forming datasets, with an emphasis on commercially valuable nacre from bivalve species. But the nacre ultrastructure is just one of more than 50 types and subtypes of molluscan shell microstructures [8]. The most common ultrastructure in bivalves and gastropods is the crossed-lamellar shell type, but little is known about the composition of the organic matrix that forms this widespread ultrastructure [29]. The model organism investigated in this study, the freshwater pulmonate *Lymnaea stagnalis*, produces a crossed-lamellar shell. *L. stagnalis* has once been a popular model to study the morphology of the shell-forming tissue in the 1960s and 1970s by Timmermans and Kniprath [30-33] and recently regained popularity as a model to study the establishment of chirality, evolutionary developmental processes or the cellular mechanism of learning and memory [34-39]. With this study, we want to continue the work of Timmermans and Kniprath and carry their knowledge and findings into the genomics era. We have combined a next generation sequencing transcriptome analysis of adult mantle tissue with a proteomic survey of the adult shell. We then compared the resulting data with similar datasets from other species, and generated the first ever *in situ* validated transcriptome-scale dataset for a species that forms a crossed-lamellar shell.

1.3.1 General aspects of molluscan shell formation

The adult molluscan shell gets secreted by the mantle tissue and comprises of an outer leathery layer and aragonite or calcite crystals. The mantle tissue is a cuticular epidermal sheet that surrounds the visceral mass and lines the inner surface of the shell [30, 40]. The molluscan shell grows in a delineated compartment- the extrapallial space. This space exhibits a supersaturated ion concentration that is necessary for crystal nucleation. The calcifying matrix is secreted into the compartment by the outer mantle epithelium. It interacts with the mineral ions and acts as a template for crystal nucleation. Components like proteins, glycoproteins, proteoglycans, polysaccharides and chitin have been identified in the extracellular matrix and their exact composition is critical for the formation of the distinct ultrastructure [5]. To delineate the extrapallial space from the surrounding environment the mantle secrets the periostracum, a leathery organic sheet that covers the $CaCO_3$ crystals and protects the shell from harmful effects of the surrounding medium [5, 32, 41].

1.3.2 The periostracum

The thin, pliable, fibrous layer that covers the molluscan shell is the periostracum [41]. It is highly proteinaceous and fulfils the important function of sealing of the extrapallial space from the environment and serves as initial support for $CaCO_3$ crystals [5, 8]. The periostracum is remarkably tough and withstands the action of peptidases like pepsin or trypsin and many hydrolytic solvents [42]. Most likely, this is due to the presence of scleroproteins, which are cross-linked structural proteins that contribute to the mechanical strength of the structure. A well accepted theory proposes that the proteins of the periostracum get scleretonised by a quinone-tanning process [42].
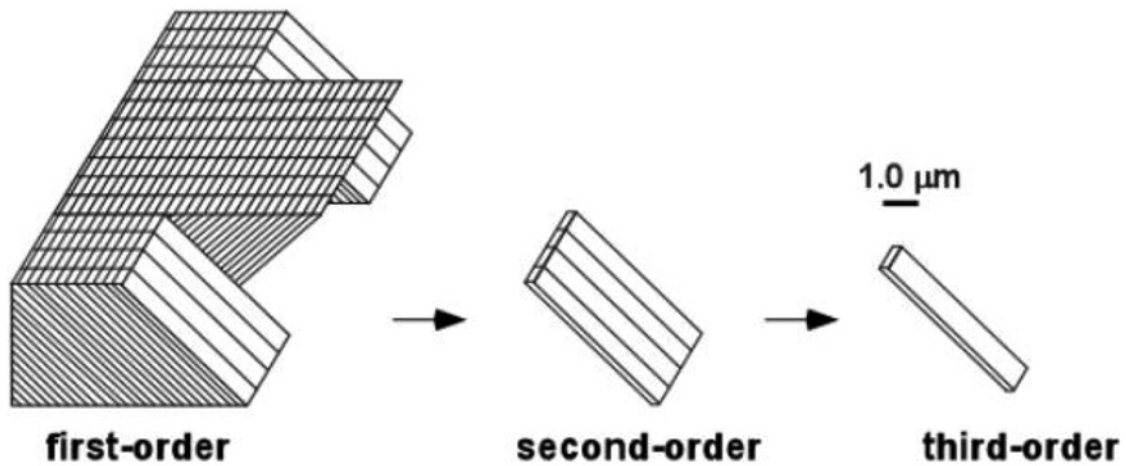
Fig 1.1. A schematic representation of the hierarchical organization of the crossed-lamellar microstructure. From Suzuki et al. 2011 [43].

1.3.3 The crossed-lamellar ultrastructure

More than 90% of molluscan species have crossed-lamellar (CL) components within their shell [44]. The freshwater pulmonate *L. stagnalis* also produces the crossed-lamellar shell type (personal observation of Daniel Jackson and Diana Bauermeister née Obermann). This microstructure shows great mechanical properties with extremely high fracture toughness that earned it its label of 'ceramic plywood' [45, 46]. That the mechanical strength is achieved despite the relatively low production costs (CL comprised about 1% organic matter compared to nacre which contains 4,5%) might be significant for the evolutionary success of the CL shell type [44, 47]. It is noteworthy that the CL microstructures comprise of a family of homeomorphic microstructures that show complex and varied crystallographies and do not represent one single microstructure [44]. The ultrastructure of the *L. stagnalis* shell has not been described yet, but the general composition of a CL shell has been studied in other species. They all consist of first, second and third order lamellae with slightly differing numbers and orientation from taxa to taxa [29]. In a nutshell, the third order lamellae consist of fibres-like structures that come together in bundles to form second order lamellae. Three-dimensional arrays of these bundles form the first order lamellae. Neighbouring first order lamellae show a proximal 90 ° inclination of the third order lamellae, as shown in Fig. 1.1 [48]. This highly hierarchical structure is able to deflected and arrested cracks at the inter-lamellar boundaries, which results in excellent fracture resistance [48]. Besides the widespread CL shell type, there is a great variety of

microstructures. Well known are the elongated prisms that either consists of aragonite or calcite or the iridescent aragonite nacre tablets that can be arranged in a 'brick wall', 'row-stack' or 'columnar' fashion [49].

1.3.4 The organic matrix

Even though the organic compounds of the shell only make up 0.1 – 5%, are these the elements that determine the $CaCO_3$ polymorph and shape the crystal ultrastructure [50]. The three-dimensional organic matrix consists of a self-assembling mixture of proteins, glycoproteins, proteoglycans, polysaccharides, free amino acids, small peptides, lipids, pigments and chitin which gets secreted by the mantle edge and forms the microenvironment that guides mineralisation [5, 49, 51]. Proteins (and the corresponding genes) are the matrix components that have received much attention, because of the diverse techniques available for high-throughput analyses. Much progress has been made in identifying components of the shell-forming proteome from a variety of gastropod and bivalve species [52-59].

A well-accepted model for the composition of a nacre-forming organic matrix has been proposed by Weiner et al. 1984 [60] and later advanced by Levi-Kalisman et al. 2001 [61] and Addadi et al. 2006 [51]. In this model, silk-like proteins form a space-filling hydrophobic hydrogel that is sandwiched between two chitin layers. Aspartic acid-rich glycoproteins are absorbed to the chitin layer to control the crystal nucleation [51, 61]. Within the gel matrix, the nacre tablets grow vertically from one chitin layer to the next and subsequently expand laterally. During lateral grow; the acidic proteins are incorporated into the nacre tablet, presumably altering the mechanical and solubility properties of the aragonite. The hydrophobic silk properties are not incorporated into the crystal. Instead they are pushed aside, finally trapped between neighbouring tablets or between tablets and the chitin layer [51]. As shown by Jackson et al. 2010 [4], different gene repertoires are employed by gastropods and bivalves to form nacre, suggesting that the architecture of the organic matrix might differ greatly between nacre-forming species. With that in mind, it seems plausible that the organic matrix of another shell ultrastructure such as crossed-lamellar might deviate considerably from the model by Levi-Kalisman et al. 2001. Uozomi et al. 1972 could observe organic

components surrounding the third order lamella within the crossed-lamellar hierarchical structure [62]. Other than that, little is know about the organic matrix that forms the crossed-lamellar ultrastructure. It has yet to be elucidated which organic components act on the crystals that form this shell type and in what architecture they might be arranged.
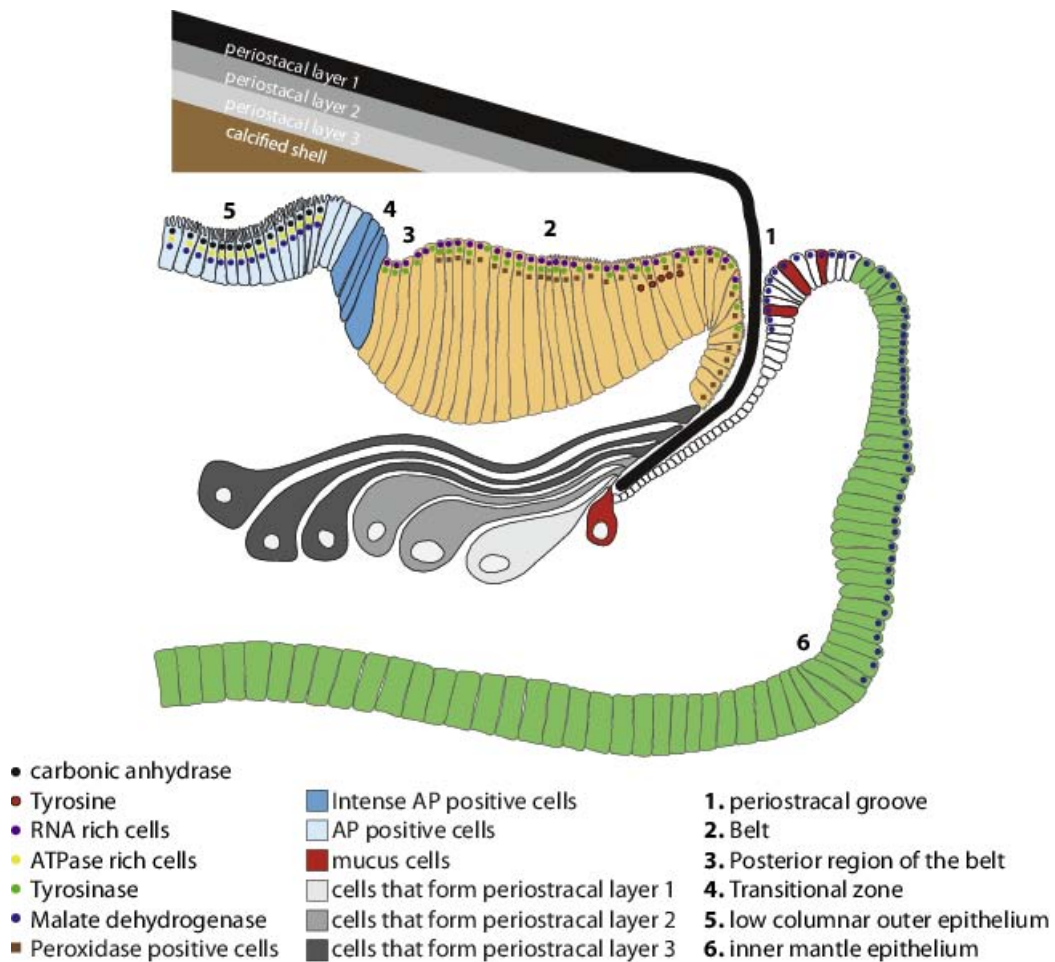
Fig. 1.2. A schematic representation of the *L. stagnalis* adult mantle adapted from Timmermans (1969) and Kniprath (1972). From Jackson et al. 2016 [63]. The activity of a variety of enzymes and the presence of various biomolecules are indicated by the colour scheme.

1.4 The shell-forming tissue of *L. stagnalis*

1.4.1 The shell-secreting adult mantle of *L. stagnalis*

The shell-secreting tissue of *L. stagnalis* has been studied extensively by Timmermans 1969 [30]. Based on a number of histochemical analyses, she was able to subdivide the mantle groove, the mantle belt, and the low columnar outer epithelium into distinct zones (Fig. 1.2). Zone one and two span the posterior wall of the groove and two-thirds of the neighbouring belt. These zones show strong activity of peroxidase and a high content of RNA [30]. The following zone three shows a similar composition as zone one and two except for the absent of peroxidase activity [30]. Zone four marks a few cells between the belt and the low columnar outer epithelium. These cells are mainly characterised by a high activity of alkaline phosphatase and glucose-6-phosphate dehydrogenase [30]. The cells of the low columnar epithelium form zone five and show the presence of glycogen, alkaline phosphatase, carbonic anhydrase, ATPase, dehydrogenase and cytochrome oxidase [30].

1.4.2 The shell-forming tissue of *L. stagnalis* larvae

The molluscan shell is a structure that is established early during development. To get an overall understanding of the *L. stagnalis* shell-forming process, it is vital to study the organ and its development throughout the life history of the animal. Much of the ontogeny of the shell-forming tissue in *L. stagnalis* has been closely observed by Kniprath 1977 [33] and Hohagen & Jackson 2013 [37]. With these studies, the authors have established a valuable foundation for further investigations concerning the shell-forming process in *L. stagnalis* during development.

The first sign of the shells precursor organ is a thickening of the dorsal ectoderm following gastrulation. Elongated ectodermal cells are clearly distinguishable from the neighbouring cells of the ectoderm and a contact with the underlying endodermal cells is established (Fig. 1.3) [33, 37]. A well-accepted theory argues that the formation of the shell field is induced by a contact between cells of the dorsal ectoderm and some endodermal cells [64]. After a number of proliferations, the epithel invaginates and forms the shell gland. The uninvaginated cells at the
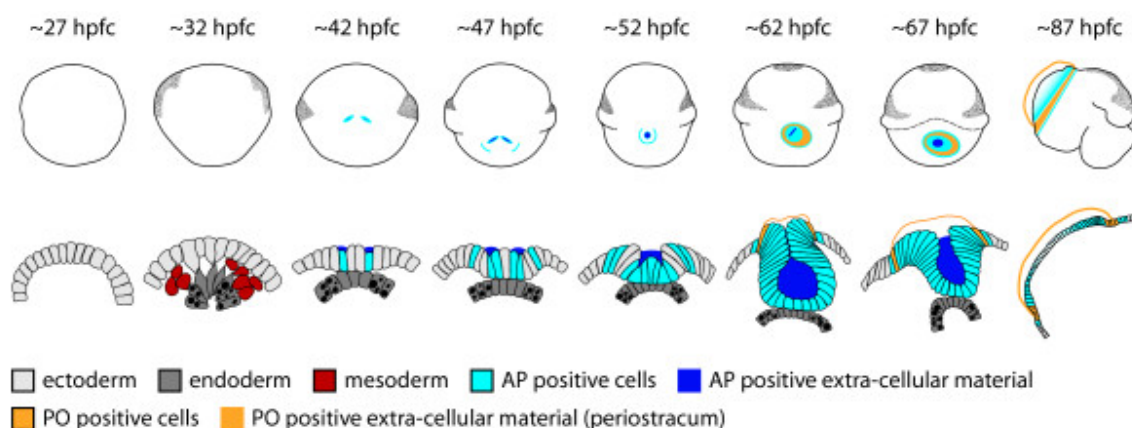
Fig. 1.3. A schematic representation of the major events during early development of the shell gland and shell field in *L. stagnalis*. AP Alkaline Phosphatase, PO Peroxidase, hpfc hours post first cleavage. From Hohagen et al. 2013 [37].

entry of the shell gland secrete a pellicle, which is the precursor of the periostracum. Subsequently, the shell gland evaginates and forms the shell field. At this stage, the first calcified material is visible. The expansion of the shell field continues until it has overgrown the visceral mass and finally forms the shell secreting mantel edge [33, 37]. The cells in the centre of the growing shell field are flat while the cells in the periphery have an elongated shape. This corresponds to the morphology of the adult mantle, where the belt region comprises of elongated cells while the cells of low columnar outer epithelium are flat [30]. In addition, also the histochemical composition of the larval shell-secreting zones correlates with the zonation of the adult mantle. The elongated cells of the larval shell field contain much RNA and are peroxidase positive, while the cells of the flattened centre contain alkaline phosphatase and glycogen. The correspondence of the larval and adult zonation of the shell-forming tissue indicates that the same processes act on the formation of the larval and the adult shell [30].

1.5 Genes and gene products involved in molluscan shell-formation

The first molluscan shell protein sequence identified in full length was from the nacre protein nacrein in 1996 [65]. In the following years, molluscan shell proteins were identified 'one-by-one' using classical molecular approaches, such as RT-PCR with degenerated primers based on known fragmentary sequences, or cDNA

library screening with oligonucleotides or antibodies. With these methods, about 40 proteins could be identified from a few shell-bearing representatives [66]. Since next-generation sequencing techniques and high-through-put proteomics are available for non-model organisms, much effort has been made to study the transcriptomes and proteomes of the mantle tissue [4, 56, 58, 67-73]. Especially studies that combine data from both the mantle transcriptome and the shell proteome obtain high-quality candidate genes. The obtained 'shellomes' represent valuable resources that allow assumptions about the evolutionary history of the molluscan shell and its organic components [66]. Shell-forming proteins often share little to no sequence similarity with proteins from model organisms, which makes any inference about their function difficult. Their functional characterisation currently represents a major challenge.

In a pioneer study, Jackson et al. 2006 investigated transcripts that putatively encode the secreted proteins that contribute to the *Haliotis asinina* organic matrix, the so-called 'secretome' [58]. 80% of the *H. asinina* 'secretome' shared no significant similarity to any sequences in GenBank and, compared to the *Lottia scutum* genome, homology to only a small fraction of transcripts could be identified. Based on these findings, it was suggested that the 'secretome' is encoded primarily by rapidly evolving genes and this rapid evolution might be the foundation for the diversity of molluscan shell shapes and patterns [58]. Similar trends could be observed by studying the proteome of *Haliotis asinina*, *Lottia gigantea* and *Cepaea nemoralis* [56, 59, 69]. To obtain more insights on the evolutionary history of the nacre-forming proteome, Jackson et al. 2010 compared genes expressed within the nacre producing cells of the bivalve *Pinctada maxima* and the gastropod *H. asinina* [4]. The authors found marked differences in gene content and expression levels and suggested that the molecular machinery used in shell construction differs fundamentally across the Mollusca. This finding hints at a significant modification of an inherited process or indicates an independent evolution of the ability to form gastropod or bivalve nacre [4]. Marie et al. 2016 counterbalances these findings by highlighting the number of similar if not homologous proteins between the nacre producing mussel *Mytilus*, the pearl oyster *Pinctada*, the freshwater mussels *Elliptio complanata* and *Villosa leinosa* as well as the non-nacreous gastropod shell of the limpet [57, 74, 75]. Furthermore,

the authors emphasise the deep molecular conservation within the bivalve nacre biomineralisation toolkit that suggests the existence of a set of Cambrian-conserved bivalve nacre proteins [74]. The evolution of molluscan shell proteins is still cryptic and many more comparison studies across the whole phylum have to be conducted to approximate its history.

Some general protein features that emerge in a number of molluscan shell proteins could be identified. These proteins often show posttranslational modifications and repetitive low complexity domains (RLCDs). These short repeated motifs consist mainly of one to three amino acids that are predominant within one module or the whole peptide sequence [49, 76]. Common are aspartic acid-rich residues that are predicted to facilitate calcium binding [77]. Many shell proteins are organized in a modular fashion and show multifunctional abilities [49]. Well-known examples are the nacre proteins lustrin A and nacrein. Lustrin A is composed of a number of alternating cysteine- and proline- rich modules that act as shock absorber, a protease inhibition domain and a short acidic domain that likely interacts with $CaCO_3$ crystals [78, 79]. Nacrein features an enzymatically functional carbonic anhydrase module and an asparagine and glycine repetitive sequence that putatively acts as a negative regulator for calcification [65, 80].

These insights focus on the secreted proteins of the organic matrix, which represent the 'downstream' molecular events that act directly on the shell-forming process. The expression of the 'downstream' shell-forming genes displays a modular arrangement within the adult mantle tissue [58, 81]. This suggests that the distinct spatial and temporal expression of the components of the organic matrix are likely coordinated by 'upstream' gene regulatory networks [63]. The modularity of the shell-forming tissue is not solely a phenomenon of the adult mantle, but can be observed already during development [30, 37]. The different shell-secreting cell types or mantle morphologies are eventually established during the embryonic development. Identifying the gene regulatory networks that coordinate the differentiation events during ontogeny is one goal of evolutionary developmental biologists. By studying the molecular nature of these events during the shell-forming process and further inter-species comparisons, they hope to provide new insights into the shell's evolutionary history [63]. Some regulatory genes that are present in the shell-forming tissue during development could

already be identified, but the molecular machinery that initiates the shell field development is still obscure. The expression of the homeodomain transcription factor *engrailed* has been shown in a wide range of molluscan taxa [82-86]. It is thought to play a conserved role in the compartment boundary formation between the embryonic shell field and the adjacent ectoderm [87]. The expression pattern of the homeobox genes *Hox1*, *Hox4* [88, 89] and the posterior *Hox* genes *Post1* and *Post2* in gastropods [89], the morphogen *dpp* in gastropods [87, 90] and bivalves [83], the inter-cellular signalling molecule *delta* in gastropods [63] and the homeodomain transcription factor *distal-less* in gastropods [63] also indicate a role associated with the patterning of the molluscan shell field. The distinct functions of these regulatory molecules are still unclear, but new molecular techniques like the gene editing CRISPR-Cas nuclease system might soon allow functional studies *in vivo* that can help elucidate their role in shell formation.

1.6 Introduction into the following chapter

The morphological properties of the shell-forming organ of molluscs have been described extensively, and a rich collection of shell-forming proteomes and transcriptomes has been collected over the past decade. A major part of these components shares little to no sequence similarity with proteins from model organisms. This makes any inference about their function very difficult. One approach to gain insights into their function is to characterise their spatial expression patterns *in vivo*. With *in situ* hybridisation experiments, we were able to trace the exact localisation of the shell-forming genes throughout the development of *L. stagnalis.*

In order to trace the expression of the shell-forming gene candidates, a robust *in situ* protocol had to be installed. Chapter 2 and 3 contain protocols for whole mount *in situ* hybridisation on *L. stagnalis* larvae*.* In chapter 4, an *in situ* protocol for *L. stagnalis* adult mantle tissue slides is presented. More than 30 candidate genes associated with the production of the *L. stagnalis* shell were identified with a proteomic survey coupled with transcriptome sequencing. We identified conserved shell-forming elements by comparing a number of biomineralising proteomes with the shell proteome of *L. stagnalis*. We furthermore analysed the

gene architecture of the biomineralisation genes to gain insight into the molecular mechanisms that might have acted on the expansion of the shell-forming molecular repertoire. This work is presented in chapter 5. To further characterise the function of the candidate genes, I attempted over-expression experiments in a heterologous expression host. This on-going project is described in chapter 6.

References

1. Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B. Phylogenomics reveals deep molluscan relationships. Nature. 2011;477:452-456.
2. Haszprunar G, Wanninger A. Molluscs. Current Biology. 2012;22:R510-R514.
3. Seed R. Structural Organization, Adaptive Radiation, and Classification of Molluscs. In: Hochachka PW, Wilbur KM, editors. Metabolic Biochemistry and Molecular Biomechanics. New York: Academic Press; 1983. p. 1-54.
4. Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C. Parallel evolution of nacre building gene sets in molluscs. Molecular Biology and Evolution. 2010;27:591-608.
5. Marin F, Luquet G. Molluscan shell proteins. Comptes Rendus Palevol. 2004;3:469-492.
6. Geubel HK: Die Gehäuse und sonstigem Gebilde der Mollusken. Johann David Sauerländer; 1845.
7. Comfort A. The pigmentation of molluscan shells. Biological Reviews. 1951;26:285-301.
8. Lowenstam HA, Weiner S: On biomineralization. Oxford University Press; 1989.
9. Weiner S, Dove PM. An overview of biomineralization processes and the problem of the vital effect. Reviews in Mineralogy and Geochemistry. 2003;54:1-29.
10. Wood R, Zhuravlev AY. Escalation and ecological selectively of mineralogy in the Cambrian Radiation of skeletons. Earth-Science Reviews. 2012;115:249-261.
11. Smetacek V. Diatoms and the ocean carbon cycle. Protist. 1999;150:25-32.
12. Finlay BJ, Hetherington NB, Da Vison W. Active biological participation in lacustrine barium chemistry. Geochimica et Cosmochimica Acta. 1983;47:1325-1329.
13. Knoll AH. Biomineralization and evolutionary history. Reviews in Mineralogy and Geochemistry. 2003;54:329-356.
14. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science. 2011;334:1091-1097.
15. Smith MP, Harper DAT. Causes of the Cambrian Explosion. Science. 2013;341:1355-1356.
16. Marshall CR. Explaining the Cambrian "Explosion" of Animals. Annual Review of Earth and Planetary Sciences. 2006;34:355-384.
17. Fedonkin MA, Waggoner BM. The Late Precambrian fossil *Kimberella* is a

mollusc-like bilaterian organism. Nature. 1997;388:868-871.

18. Brennan ST, Lowenstein TK, Horita J. Seawater chemistry and the advent of biocalcification. Geology. 2004;32:473-476.

19. Simkiss K. Biomineralization and detoxification. Calcified Tissue International. 1977;24:199-200.

20. Mann S. Mineralization in biological systems. In: editors. Springer; 1983. p. 125-174.

21. Lowenstam HA. Minerals formed by organisms. Science. 1981;211:1126-1131.

22. Frankel RB, Bazylinski DA. Biologically induced mineralization by bacteria. Reviews in Mineralogy and Geochemistry. 2003;54:95-114.

23. Wesson JA, Ward MD. Pathological biomineralization of kidney stones. Elements. 2007;3:415-421.

24. Borowitzka MA. Morphological and cytological aspects of algal calcification. International Review of Cytology. 1982;74:127-162.

25. Paasche E. Coccolith formation. Nature. 1962;193:1094-1095.

26. Falini G, Albeck S, Weiner S, Addadi L. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. Science. 1996;271:67-69.

27. Hogendorn J, Johnson M: The Shell Money of the Slave Trade. Cambridge Univ Press; 1986.

28. dpa. Eine der größten Salzwasserperlen versteigert, dpa. Die Welt. 2014;Newsticker

29. Dauphin Y, Denis A. Structure and composition of the aragonitic crossed lamellar layers in six species of Bivalvia and Gastropoda. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology. 2000;126:367-377.

30. Timmermans LPM. Studies on shell formation in molluscs. Netherlands Journal of Zoology. 1969;19:413-523.

31. Kniprath E. Das Wachstum des Mantels von *Lymnaea stagnalis* (Gastropoda). Cytobiologie. 1975;10:260-267.

32. Kniprath E. Formation and structure of the periostracum in *Lymnaea stagnalis*. Calcified Tissue Research. 1972;9:260-271.

33. Kniprath E. Zur Ontogenese des Schalenfeldes von *Lymnaea stagnalis*. Wilhelm Roux's Archives of Developmental Biology. 1977;181:11-30.

34. Kuroda R, Endo B, Abe M, Shimizu M. Chiral blastomere arrangement dictates zygotic left–right asymmetry pathway in snails. Nature. 2009;462:790-794.

35. Liu MM, Davey JW, Banerjee R, Han J, Yang F, Aboobaker A, Blaxter ML, Davison A. Fine mapping of the pond snail left-right asymmetry (chirality) locus using RAD-Seq and fibre-FISH. PLoS One. 2013;8:e71067.

36. Davison A, McDowell GS, Holden JM, Johnson HF, Koutsovoulos GD, Liu MM, Hulpiau P, Van Roy F, Wade CM, Banerjee R. Formin is associated with left-right asymmetry in the pond snail and the frog. Current Biology. 2016;26:654-660.

37. Hohagen J, Jackson DJ. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. BMC Developmental Biology. 2013;13:27.

38. Hohagen J, Herlitze I, Jackson DJ. An optimised whole mount *in situ* hybridisation protocol for the mollusc *Lymnaea stagnalis*. BMC Developmental Biology. 2015;15:19.

39. Khan AM, Spencer GE. Novel neural correlates of operant conditioning in normal and differentially reared *Lymnaea*. Journal of Experimental Biology. 2009;212:922-933.
40. Brusca RC, Brusca GJ: Invertebrates. 2nd edition. Sinauer Associates Sunderland, Massachusetts; 2002.
41. Saleuddin ASM, Petit HP. The mode of formation and the structure of the periostracum. In: Saleuddin ASM, Wilbur KM, editors. The Mollusca, vol. 4, Physiology. New York: Academic Press; 1983. p. 199-233.
42. Waite JH. Quinone-tanned scleroproteins. In: Saleuddin ASM, Wilbur KM, editors. Physiology. New York: Academic Press; 1983. p. 467-504.
43. Suzuki M, Kogure T, Weiner S, Addadi L. Formation of aragonite crystals in the crossed lamellar microstructure of limpet shells. Crystal Growth & Design. 2011;11:4850-4859.
44. Almagro I, Drzymała P, Berent K, Sainz-Díaz CI, Willinger M-G, Bonarski J, Checa AG. New crystallographic relationships in biogenic aragonite: the crossed-lamellar microstructures of mollusks. Crystal Growth & Design. 2016
45. Kuhn-Spearing LT, Kessler H, Chateau E, Ballarini R, Heuer AH, Spearing SM. Fracture mechanisms of the *Strombus gigas* conch shell: implications for the design of brittle laminates. Journal of Materials Science. 1996;31:6583-6594.
46. Kamat S, Su X, Ballarini R, Heuer AH. Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. Nature. 2000;405:1036-1040.
47. Rodriguez-Navarro AB, Checa A, Willinger M-G, Bolmaro R, Bonarski J. Crystallographic relationships in the crossed lamellar microstructure of the shell of the gastropod *Conus marmoreus*. Acta Biomaterialia. 2012;8:830-835.
48. Pokroy B, Zolotoyabko E. Microstructure of natural plywood-like ceramics: a study by high-resolution electron microscopy and energy-variable X-ray diffraction. Journal of Materials Chemistry. 2003;13:682-688.
49. Marin F, Le Roy N, Marie B. The formation and mineralization of mollusk shell. Frontiers in Bioscience. 2012;4:1099-1125.
50. Nudelman F. Nacre biomineralisation: A review on the mechanisms of crystal nucleation. Seminars in Cell & Developmental Biology. 2015;46:2-10.
51. Addadi L, Joester D, Nudelman F, Weiner S. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. Chemistry – A European Journal. 2006;12:980-987.
52. Zhao M, He M, Huang X, Wang Q, Shi Y. Functional characterization and molecular mechanism exploration of three granulin epithelin precursor splice variants in biomineralization of the pearl oyster *Pinctada fucata*. Molecular Genetics and Genomics. 2016;291:399-409.
53. Wang J, Gao J, Xie J, Zheng X, Yan Y, Li S, Xie L, Zhang R. Cloning and mineralization-related functions of the calponin gene in *Chlamys farreri*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 2016;201:53-58.
54. Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. Marine Genomics. 2016;27:9-15.
55. Gao P, Liao Z, Wang X-, Bao L-, Fan M-, Li X-, Wu C-, Xia S-. Layer-by-layer proteomic analysis of *Mytilus galloprovincialis* shell. PloS ONE. 2015;10:e0133913.

56. Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. BMC Genomics. 2014;15:249.

57. Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cleon I, Guichard N, Marin F. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. FEBS Journal. 2013;280:214-232.

58. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM. A rapidly evolving secretome builds and patterns a sea shell. BMC Biology. 2006;4:40.

59. Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C, Marin F. Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. Proteome Science. 2010;8:54.

60. Weiner S, Traub W, Parker SB. Macromolecules in mollusc shells and their functions in biomineralization [and Discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences. 1984;304:425-434.

61. Levi-Kalisman Y, Falini G, Addadi L, Weiner S. Structure of the nacreous organic matrix of a bivalve mollusk shell examined in the hydrated state using cryo-TEM. Journal of Structural Biology. 2001;135:8-17.

62. Uozumi S, Iwata K, Togo Y. The ultrastructure of the mineral in and the construction of the crossed-lamellar layer in molluscan shell. Journal of the Faculty of Science, Hokkaido University Series 4, Geology and mineralogy= 北海道大學理學部紀要. 1972;15:447-477.

63. Jackson DJ, Degnan BM. The importance of Evo-Devo to an integrated understanding of molluscan biomineralisation. Journal of Structural Biology. 2016;196:67-74.

64. Raven CP. Morphogenesis in *Limnaea stagnalis* and its disturbance by lithium. Journal of Experimental Zoology. 1952;121:1-77.

65. Miyamoto H, Miyashita T, Okushima M, Nakano S, Morita T, Matsushiro A. A carbonic anhydrase from the nacreous layer in oyster pearls. Proceedings of the National Academy of Sciences. 1996;93:9657-9660.

66. Marin F, Marie B, Hamada SB, Ramos-Silva P, Le Roy N, Guichard N, Wolf SE, Montagnani C, Joubert C, Piquemal D. 'Shellome': Proteins Involved in Mollusc Shell Biomineralization—Diversity, Functions. Recent Advances in Pearl Research. 2013149-166.

67. Clark MS, Thorne MAS, Vieira FA, Cardoso JCR, Power DM, Peck LS. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. BMC Genomics. 2010;11:362.

68. Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y, Montagnani C. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. BMC Genomics. 2010;11:613.

69. Mann K, Edsinger-Gonzales E, Mann M. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. Proteome Science. 2012;10:28.

70. Marie B, Joubert C, Tayaléa A, Zanella-Cléon I, Belliard C, Piquemal D, Cochennec-Laureau N, Marin F, Gueguen Y, Montagnani C. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of

Sciences. 2012;109:20986-20991.

71. Werner GDA, Gemmell P, Grosser S, Hamer R, Shimeld SM. Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. Marine Biotechnology. 2013;15:230-243.

72. Freer A, Bridgett S, Jiang J, Cusack M. Biomineral proteins from *Mytilus edulis* mantle tissue transcriptome. Marine Biotechnology. 2014;16:34-45.

73. Sleight VA, Thorne MAS, Peck LS, Arivalagan J, Berland S, Marie A, Clark MS. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. Marine Genomics. 2016;27:47-55.

74. Marie B, Arivalagan J, Dubost L, Berland S, Marie A, Marin F. Unveiling the Evolution of Bivalve Nacre Proteins by Shell Proteomics of Unionoidae. Key Engineering Materials. 2016;672:158-167.

75. Marie B, Le Roy N, Zanella-Cléon I, Becchi M, Marin F. Molecular evolution of mollusc shell proteins: insights from proteomic analysis of the edible mussel *Mytilus*. Journal of Molecular Evolution. 2011;72:531-546.

76. Marin F, Luquet G, Marie B, Medakovic D. Molluscan shell proteins: primary structure, origin, and evolution. Current Topics in Developmental Biology. 2008;80:209-276.

77. Addadi L, Weiner S. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. Proceedings of the National Academy of Sciences. 1985;82:4110-4114.

78. Wustman BA, Weaver JC, Morse DE, Evans JS. Structure-function studies of the lustrin A polyelectrolyte domains, RKSY and D4. Connective Tissue Research. 2003;44:10-15.

79. Smith BL, Schäffer TE, Viani M, Thompson JB, Frederick NA, Kindt J, Belcher A, Stucky GD, Morse DE, Hansma PK. Molecular mechanistic origin of the toughness of natural adhesives, fibres and composites. Nature. 1999;399:761-763.

80. Miyamoto H, Miyoshi F, Kohno J. The carbonic anhydrase domain protein nacrein is expressed in the epithelial cells of the mantle and acts as a negative regulator in calcification in the mollusc *Pinctada fucata*. Zoological Science. 2005;22:311-315.

81. Jolly C, Berland S, Milet C, Borzeix S, Lopez E, Doumenc D. Zona localization of shell matrix proteins in mantle of *Haliotis tuberculata* (Mollusca, Gastropoda). Marine Biotechnology. 2004;6:541-551.

82. Moshel SM, Levine M, Collier JR. Shell differentiation and engrailed expression in the *Ilyanassa* embryo. Development Genes and Evolution. 1998;208:135-141.

83. Kin K, Kakoi S, Wada H. A novel role for *dpp* in the shaping of bivalve shells revealed in a conserved molluscan developmental program. Developmental Biology. 2009;329:152-166.

84. Wanninger A, Haszprunar G. The expression of an engrailed protein during embryonic shell formation of the tusk-shell, *Antalis entalis* (Mollusca, Scaphopoda). Evolution & Development. 2001;3:312-321.

85. Baratte S, Andouche A, Bonnaud L. *Engrailed* in cephalopods: a key gene related to the emergence of morphological novelties. Development Genes and Evolution. 2007;217:353-362.

86. Jacobs DK, Wray CG, Wedeen CJ, Kostriken R, DeSalle R, Staton JL, Gates RD,

Lindberg DR. Molluscan *engrailed* expression, serial organization, and shell evolution. Evolution & Development. 2000;2:340-347.

87. Nederbragt AJ, van Loon AE, Dictus WJAG. Expression of *Patella vulgata* orthologs of *engrailed* and *dpp-BMP2/4* in adjacent domains during molluscan shell development suggests a conserved compartment boundary mechanism. Developmental Biology. 2002;246:341-355.

88. Hinman VF, O'Brien EK, Richards GS, Degnan BM. Expression of anterior Hox genes during larval development of the gastropod Haliotis asinina. Evolution & Development. 2003;5:508-521.

89. Samadi L, Steiner G. Involvement of Hox genes in shell morphogenesis in the encapsulated development of a top shell gastropod (*Gibbula varia* L.). Development Genes and Evolution. 2009;219:523-530.

90. Iijima M, Takeuchi T, Sarashina I, Endo K. Expression patterns of *engrailed* and *dpp* in the gastropod *Lymnaea stagnalis*. Development Genes and Evolution. 2008;218:237-251.

2

An optimised whole mount *in situ* hybridisation protocol for the mollusc *Lymnaea stagnalis*

Jennifer Hohagen, Ines Herlitze and Daniel John Jackson

2.1 Abstract

Background: The ability to visualise the expression of individual genes *in situ* is an invaluable tool for developmental and evolutionary biologists; it allows for the characterisation of gene function, gene regulation and through inter-specific comparisons, the evolutionary history of unique morphological features. For well-established model organisms (e.g., flies, worms, sea urchins) this technique has been optimised to an extent where it can be automated for high-throughput analyses. While the overall concept of *in situ* hybridisation is simple (hybridise a single-stranded, labelled nucleic acid probe complementary to a target of interest, and then detect the label immunologically using colorimetric or fluorescent methods), there are many parameters in the technique that can significantly affect the final result. Furthermore, due to variation in the biochemical and biophysical properties of different cells and tissues, an *in situ* technique optimised for one species is often not suitable for another, and often varies depending on the ontogenetic stage within a species.

Results: Using a variety of pre-hybridisation treatments we have identified a set of treatments that greatly increases both whole mount *in situ* hybridisation (WMISH) signal intensity and consistency while maintaining morphological integrity for early larval stages of *Lymnaea stagnalis*. These treatments function well for a set of genes with presumably significantly different levels of expression (*beta tubulin*, *engrailed* and *COE*) and for colorimetric as well as fluorescent WMISH. We also identify a tissue specific background stain in the larval shell field of *L. stagnalis* and a treatment, which eliminates this signal.

Conclusions: This method that we present here will be of value to investigators employing *L. stagnalis* as a model for a variety of research themes (e.g. evolutionary biology, developmental biology, neurobiology, ecotoxicology), and brings a valuable tool to a species in a much understudied clade of animals collectively known as the Spiralia.

2.2 Background

Analysing how spatial and temporal developmental gene expression profiles evolve is a powerful strategy for understanding how morphological diversity can be generated. The most commonly employed technique for the study of spatial gene expression in a given tissue or developmental stage is *in situ* hybridisation (ISH), often applied to whole embryos or larvae as whole mount *in situ* hybridisation (WMISH). WMISH provides information about the timing and localisation of a gene's expression in a developing embryo or larva, and can be used to characterise and identify cell types, tissues or organs within the whole organism and to make inferences about their function and evolutionary history [1-3]. Unfortunately, the technique can be challenging, especially when applied to an organism for which there is little knowledge regarding the multifarious conditions that optimise the balance between WMISH signal intensity and the preservation of morphological integrity, two often conflicting requirements. WMISH experiments on embryos can be further challenged by changes in the biochemical and biophysical properties of the developing tissues during ontogenesis. Thus, the procedure often needs to be adapted for distinct developmental stages within a species.

From an evo-devo perspective, the pulmonate freshwater gastropod *Lymnaea stagnalis* (Linnaeus, 1758) is a representative of a significantly understudied group of animals, the Spiralia/Lophotrochozoa. Primarily due to its availability and ease of culture, *L. stagnalis* was once a much used model for studying molluscan development [4-6] and is currently employed as a model for studies focused on various biological processes including the establishment of chirality [7], the evolution of shell formation [8] and ecologically regulated development [9]. However, *L. stagnalis* possesses certain traits that represent technical challenges to WMISH. First, *L. stagnalis* embryos develop individually within egg capsules filled

with a fluid that serves a nutritive function and is uptaken by pinocytosis during development [10-12]. This viscous intra-capsular fluid, which consists of a complex mixture of ions, polysaccharides, proteoglycans and other polymers [13], can be seen to stick to the embryo following decapsulation, and likely interferes with any WMISH procedure. Second, from 52 hours post first cleavage onwards the first insoluble material associated with shell formation is secreted [8]. This material nonspecifically binds some nucleic acid probes and generates a characteristic background signal. This phenomena is not restricted to *L. stagnalis* but can be observed in larvae of other gastropods (our unpublished data), bivalves, scaphopods and polyplacophoran molluscs (pers. comm. Tim Wollesen). Finally, *L. stagnalis* embryos and larvae undergo significant morphometric and biophysical changes in the characteristics of their tissues during the first days of development (Fig. 2.1). Previously described WMISH protocols for larvae of *L. stagnalis* produced WMISH signals with low signal to noise ratios, making some previously reported gene expression patterns difficult to interpret [14-16].

In order to achieve consistent WMISH signals in *L. stagnalis* larvae with maximum signal to noise ratios, we have systematically compared the influence of a variety of chemical and enzymatic pre-hybridisation treatments previously reported to address each of these challenges. We first evaluated the effect of the mucolytic agent N-acetyl-L-cysteine (NAC) in order to assess the possibly negative influence of the intra-capsular fluid on WMISH in *L. stagnalis*. A treatment with NAC has been shown to improve WMISH signal intensity in the platyhelminth flatworm *Schmidtea mediterranea*, presumably by degrading the mucosal layer surrounding the animal and thereby increasing accessibility of the probe to the tissue [17]. WMISH signal quality was also improved in *S. mediterranea* through the use of the reducing agent dithiothreitol (DTT) and the detergents sodium dodecyl sulfate (SDS) and NP-40, a treatment referred to by Pearson et al. as 'reduction' [17]. An alternative permeabilising treatment solely utilising SDS is commonly employed in WMISH protocols for a variety of animals such as the platyhelminth *S. mediterranea* [18] or the arthropod *Parhyale hawaiensis* [19, 20]. Here we assess the impact of different combinations of these and other standard WMISH treatments (storage, enzymatic permeabilisation by Proteinase K (Pro-K), acetylation) on the strength and consistency of the WMISH signal across early

developmental stages of *L. stagnalis*. We also systematically evaluated the effects of the Alkaline Phosphatase (AP) -conjugated anti-DIG anti- body concentration, the composition of the colour detection solution and different probe preparation approaches. We have performed these experiments with a selection of three genes, which can be reasonably assumed to have different levels of expression: *beta tubulin*, and the transcription factors *engrailed* and *COE* (*collier/olfactory-1/early B cell factor*). We also demonstrate the presence of a tissue-specific background stain, which can be abolished by treatment with triethanolamine (TEA) and acetic anhydride (AA). The optimised WMISH method we present here will allow for future molecular studies to be performed on a wide range of developmental processes within *L. stagnalis*.
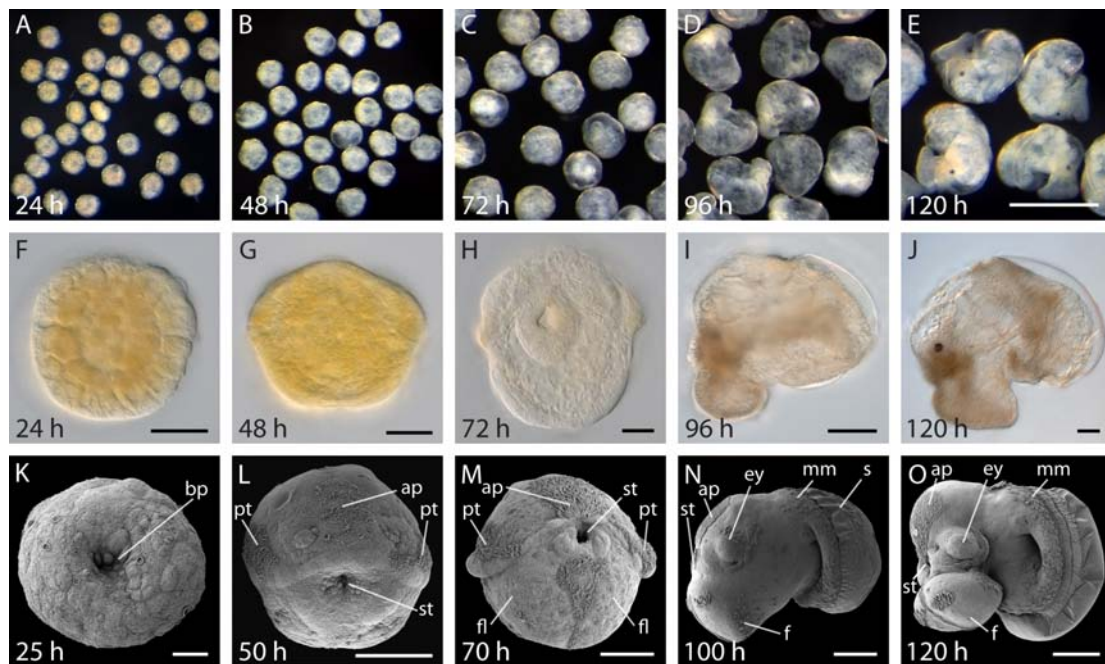
Fig. 2.1. Overview of the early larval development of *L. stagnalis*. During the first five days of development, embryos of *L. stagnalis* undergo drastic changes in size (**A-E**, images are to the same scale shown in **E**), tissue composition **(F-J)** and form all main larval structures **(K-O)**. Indicated are the positions of the apical plate (ap), the eye (ey), the foot lobe (fl) or foot (f), the developing mantle margin (mm), the prototroch (pt), the shell (s) and the blastopore (bp) or stomodaeum (st). All ages are indicated in hours post first cleavage (h). **F-H** and **K-M** are ventral views with scale bars representing 50 μm. **I, J**, **N** and **O** are lateral views with scale bars of 100 μm. The scale bar in panel **E** is 500 μm. Panels **I, J**, **N** and **O** are reflected about the vertical axis for consistency of presentation.

2.3 Methods

A detailed list of all solutions used can be found at the end of this section. If not otherwise indicated, all steps were carried out at room temperature.

2.3.1 Cultivation of adult *L. stagnalis* and preparation of embryos

Laboratory cultures derived from adult *L. stagnalis* collected from the Northeimer Seenplatte, Germany, from a pond on the North campus of the University of Göttingen, Germany, and from Nottingham, U.K. and were kept in standard tap water at 25 °C, under a 16:8 light dark regime and fed *ad libidum* with lettuce and a variety of other vegetables. Under this regime adult snails lay egg masses year round. Egg masses of diverse ages were collected and grouped into three developmental time windows: from one to two days post first cleavage (dpfc), from approximately two to three dpfc and from three to five dpfc. Individual egg capsules were freed from the surrounding jelly by rolling them over moist filter paper. Embryos were released from their egg capsules by manual dissection using forceps and mounted needles. In order to minimise experimental error, embryos for each experiment were pooled and processed up to a point when experiment specific treatments were applied.

2.3.2 NAC treatment

Freshly dissected embryos were immediately incubated in NAC solution. The duration and concentration of this treatment were age-dependent. Embryos ranging from two to three dpfc were treated for five minutes with 2.5% NAC, and samples between three and six dpfc were treated with 5% NAC twice for five minutes each. All samples were then immediately fixed for 30 minutes in 4% paraformaldehyde (PFA) in PBS.

2.3.3 Fixation

All samples were transferred into freshly prepared 4% (w/v) PFA in 1X PBS and incubated for 30 minutes at room temperature. The fixative was removed by one wash for five minutes in 1X PBTw. Samples were then subjected to a treatment with SDS.

2.3.4 SDS treatment

Following fixation, all samples were washed once in PBTw for five minutes and then incubated in 0.1% SDS, 0.5% SDS or 1% SDS in PBS for ten minutes at room temperature. Following the SDS treatment, samples were rinsed in PBTw and dehydrated through a graded ethanol (EtOH) series in PBTw; one wash in 33% (v/v) EtOH, one wash in 66% (v/v) EtOH and two washes in 100% EtOH, each wash lasting five to ten minutes. All samples were then stored at −20 °C.

2.3.5 Reduction

Following fixation and one five minutes wash in PBTw, embryos between two and three dpfc were treated with 0.1X reduction solution for ten minutes at room temperature. Embryos between three and five dpfc were incubated for ten minutes in preheated 1X reduction solution at 37 °C. All samples were carefully inverted once during this time. We found all samples to be extremely fragile in this solution and should be handled with care. After removal of the reduction solution, all samples were briefly rinsed with PBTw before being dehydrated through a graded EtOH series; one wash in 50% (v/v) EtOH, two washes in 100% EtOH, each wash lasting five to ten minutes. All samples were then stored at −20 °C. Note: this treatment replaces the SDS treatment described above.

2.3.6 RNAse treatment in order to investigate the source of non-specific WMISH staining

NAC-treated samples were fixed and dehydrated as described above and stored at −20 °C. Samples were then rehydrated through a graded EtOH series into PBTw and then incubated for 30 minutes at 37 °C in 10 μg/mL and 100 μg/mL RNAse A (Sigma, #R5503) in 2X SSC. Samples were then washed five times in PBTw for 5 minutes each before proceeding with Proteinase K digestion.

2.3.7 Protein digestion with Proteinase K (Pro-K)

Following fixation, dehydration, storage at −20 °C and any additional treatments (NAC, reduction or SDS), samples were rehydrated through a graded EtOH series into PBTw. Embryos were then treated with an agedependent concentration of Pro-K (Carl Roth, #7528) for ten minutes at room temperature. The regimes

ultimately employed are the culmination of a more exhaustive series of trials using a greater range of Pro-K concentrations (0 to 50 µg/mL Pro-K). Embryos between one and two dpfc were incubated in concentrations of Pro-K ranging from 1–15 µg/mL, embryos between two and three dpfc in concentrations of Pro-K ranging from 5–20 µg/mL and older embryos (between three and five dpfc) were treated at concentrations between 5 µg/mL and 40 µg/mL. Pro-K activity was stopped by two five minutes washes in 2 mg/mL glycine. All samples were then briefly rinsed in PBTw.

2.3.8 Triethanolamine + acetic anhydride (TEAAA) treatment

Samples were transferred into a 1% (v/v) solution of triethanolamine (TEA) in PBTw and incubated for five minutes. This step was then repeated. This solution was then replaced with a solution of 1% TEA + 0.3% (v/v) acetic anhydride (TEAAA) in PBTw. This step was repeated for some samples. All samples were then washed once with PBTw, postfixed for 15 to 20 minutes in 4% PFA in PBTw, and washed three times with PBTw before being transferred into an Intavis In situ-Pro robot for all subsequent hybridisation, antibody incubation and washing steps.

2.3.9 Riboprobe synthesis

Primers designed to amplify fragments of *beta tubulin*, *engrailed* and *COE* were designed from 454 and Illumina RNASeq data (see Additional file S2.1 for all primer sequences). These PCR products were cloned into vectors containing T7 and SP6 promotor sites and verified by Sanger sequencing. These fragments were then amplified from plasmid DNA using M13 primers, and purified using the QIAGEN QIAquick Gel Extraction Kit. Antisense riboprobes were synthesised using Promega reagents in a 10 µL-reaction containing 1X reverse transcription buffer, 10 mM Dithiothreitol, 1X Digoxigenin RNA labeling Mix (Roche, #11277073910), 0.25 - 0.5 volume PCR template and 20 Units of the appropriate RNA polymerase (SP6 or T7; Promega, #P108 or #P207). Probe synthesis reactions were carried out at 37 °C for two to four hours. For *beta tubulin*, a 702 bp long internal fragment was used for riboprobe synthesis. For *engrailed*, a 929 bp internal fragment partially covering the homeobox domain was used. The riboprobe against *COE* was generated from a 1626 bp long internal fragment covering the DNA binding

domain and the TIG/IPT domain. All riboprobes were purified by precipitation using 0.1 volume of 3 M sodium acetate pH 5.2 and 3 volumes of absolute EtOH for 15 minutes, and subsequently centrifuged for 15 minutes at 16,000 RCF. All precipitation steps were carried out at room temperature. The resulting pellets were washed once in 75% EtOH, dried and dissolved in 10 µL water at 55 °C. After quantification using a Nanodrop, 500 ng of riboprobe was denatured in 95% deionised formamide at 75 °C for 10 minutes and qualitatively assessed by agarose gel electrophoresis. The remaining riboprobe solution was adjusted to a final concentration of 300 ng/µL using deionised formamide. In order to assess the affect of probe hydrolysis on WMISH signal, some riboprobes were also hydrolysed as described by [17].

2.3.10 Probe hybridisation and antibody binding

All samples were incubated in hybridisation buffer for 15 minutes at room temperature before being heated to the hybridisation temperature of 55 °C. The hybridisation buffer was then exchanged and incubated for an additional two hours. Each riboprobe in hybridisation buffer was denatured for ten minutes at 75 °C and aliquoted into individual hybridisation reaction tubes for subsequent use in the robot. The hybridisation buffer on all samples was replaced by the riboprobe in hybridisation buffer and allowed to hybridise for 16 hours at 55 °C using the following optimised concentrations of riboprobes: *beta tubulin* 100, 150 or 200 ng/mL; *engrailed* 500 ng/mL; *COE* 100 ng/mL or 300 ng/mL. Unbound probe was washed out with three washes in 4X wash buffer for 15 minutes each, three washes in 2X wash buffer for 15 minutes each, three washes in 1X wash buffer for 15 minutes each and one wash in 1X SSC + 0.1% Tween for 15 minutes, all performed at 55 °C. Samples were then allowed to cool to room temperature and then washed twice in 1X SSC + 0.1% Tween for 15 minutes each. Two washes in maleic acid buffer (MAB) pH 7.5 were then performed for ten minutes each. All samples were then cooled to 10 °C and incubated for three hours and 30 min in pre-cooled 2% block solution (Roche, #11096176001) in MAB with one exchange. Block solution was then replaced by block solution containing a 1/10,000 dilution of anti-DIG antibody conjugated to Alkaline Phosphatase (Roche, #11093274910) and incubated for five hours followed by a renewal of this solution and a further

five hours incubation, all at 10 °C. Samples were then allowed to warm to room temperature and unbound antibody was removed by 15 washes with PBTw for ten minutes each.

2.3.11 Colour development and postprocessing

For colour development, samples were transferred into 1X Alkaline Phosphatase buffer with 0.1% (v/v) Tween 20 (APTw) and incubated with two ten minutes washes at room temperature. The 1X APTw buffer was replaced by the detection buffer and colour development was performed in the dark. Fluorescent signals were developed using the SIGMA FAST™ Fast Red TR/Napthol AS-MX Alkaline Phosphatase Substrate (Sigma, #F4648), prepared according to the manufactorer's recommendations. All reactions were stopped by replacing the colour substrate solution with three five minutes washes in PBTw each, followed by two five minutes washes in 0.1 M Glycine pH 2. Samples for direct comparisons were stopped at the same time. After three further five minutes washes in PBTw, samples were postfixed in 4% (v/v) PFA in PBTw for at least two hours at room temperature or over night at 4 °C. The fixative was removed by two five minutes washes with PBTw, followed by two washes with prewarmed deionised water for each ten minutes at 37 °C. Embryos were then dehydrated through a graded EtOH series (33%, 66%, twice with 100%) and stored at −20 °C.

2.3.12 Imaging

Prior to imaging samples were rehydrated through a graded EtOH series in PBTw (66%, 33%, twice with PBTw) and cleared at 4 °C over night in 60% (v/v) glycerol. For "bulk" imaging (where a single image of tens of embryos gives an impression of the consistency of a given treatment) embryos were mounted in a 96 well plate with U-shaped bottom and imaged under a Zeiss stereo Discovery V8 microscope running Zeiss camera software Axio Vision Rel. 4.7. For images of individual embryos, samples were mounted on glass slides and photographed using a Zeiss Axio Imager Z1 microscope running Zeiss camera software Axio Vision Rel. 4.8. Images of all samples were captured using automatic settings for exposure and white balance. Images of individual embryos were also captured at different focal planes some of which were projected using Macnification version 2.0.1. All images

were edited in Adobe Photoshop CS3 version 10.0.1 to achieve the optimal visual representation of each WMISH treatment and to facilitate qualitative comparisons. Each image was linearly adjusted for brightness, contrast and colour balance using the automatic function. These adjustments were applied to every pixel in each image and did not obscure, eliminate, or misrepresent any information.

Fluorescence detection was performed using a Zeiss LSM 510 Meta with the following microscope setup: HeNe 543 laser, HFT 488/543/633; NFT490; LP560. Both individual images and stacks were captured using the following settings: laser power of 2.9%; pinhole of 59 μm; amplifier gain of 1; amplifier offset and gain adjusted to the sample brightness; stack size of 1024 × 1024; scan speed and number of scans 7 and 4 respectively. For individual images the stack size was 2048 × 2048.

2.3.13 Solutions

1X PBS (phosphate buffered saline): 0.1 volume of 10X PBS stock (1.37 M NaCl; 27 mM KCl; 100 mM $Na_2HPO_4 \cdot 2H_2O$; 20 mM KH_2PO_4)

1X PBTw (phosphate buffered saline+Tween 20): 10% (v/v) of 10X PBS stock; 0.1% (v/v) Tween-20

2.5% NAC (N-acetyl cysteine): 50% (v/v) of 5% (w/v) NAC in 1X PBS

4% PFA (paraformaldehyde): 25% (v/v) of 16% (w/v) PFA pH 7–8; 1X PBS or 1X PBTw

0.5% SDS (sodium dodecyl sulfate): 2.5% (v/v) of 20% (w/v) SDS; 1X PBS

0.1X reduction solution: 0.1% (v/v) Tergitol (NP40); 0.05% (v/v) sodium dodecyl sulfate (SDS); 5 mM dithiothreitol (DTT)

1X reduction solution: 1% (v/v) Tergitol (NP40); 0.5% (v/v) SDS, 50 mM DTT

33% EtOH (ethanol): 33% (v/v) volume of absolute EtOH in PBTw

66% EtOH (ethanol): 66% (v/v) volume of absolute EtOH in PBTw

Pro-K (Proteinase K): Diluted from 10 mg/mL stock using PBTw

2 mg/mL glycine pH 2: Diluted from 100 mg/mL stock using PBTw

TEA (triethanolamine): 1% (v/v) TEA diluted in 1X PBTw

TEAAA (triethanolamine + acetic anhydride): 1% (v/v) TEA; 0.3% (v/v) acetic anhydride diluted in 1× PBTw

Hybridisation buffer: 25% (v/v) 20X SSC stock (3 M NaCl; 0.3 M trisodium citrate dihydrate); 5 mM ethylene diamine tetra-acetic acid (EDTA) from 500 mM stock pH 8.0; 0.5 volume deionised formamide; 100 µg/mL Heparin from 100 mg/mL stock; 0.1% (v/v) Tween-20; 1X Denhardt's from 100X stock (2% (m/v) Ficoll type 400; 2% (w/v) polyvinylpyrrolidone K30; 2% (w/v) bovine serum albumin); 100 µg/mL single-stranded salmon sperm DNA from 10 mg/mL stock

4X wash: 20% (v/v) 20X SSC stock; 50% (v/v) formamide; 0.1% (v/v) Tween-20

2X wash: 10% (v/v) 20X SSC stock; 50% (v/v) formamide; 0.1% (v/v) Tween-20

1X wash: 5% (v/v) 20X SSC stock; 50% (v/v) formamide; 0.1% (v/v) Tween-20

MAB (maleic acid buffer): 0.1 M maleic acid from 1 M stock pH 7.5; 0.15 M NaCl from 5 M stock

Block solution: 2% (v/v) block from 10% (w/v) stock in MAB

Antibody solution: AP-conjugated anti-DIG fab fragments diluted 1/10,000 in block solution

1X APTw (Alkaline Phosphatase buffer + Tween 20): 20% (v/v) 5X AP buffer stock (0.5 M Tris pH 9.5 from 1 M stock; 0.5 M NaCl from 5 M stock); 0.1% (v/v) Tween-20

Colour detection buffer: 1X APTw; 50 mM $MgCl_2$ from 1 M stock; 450 µg/mL NBT from 100 mg/mL stock in DMF; 175 µg/mL BCIP from 50 mg/mL stock in water

Colour detection buffer + PVA (polyvinyl alcohol): 1X APTw; 50 mM $MgCl_2$ from 1 M stock; 450 µg/mL NBT from 100 mg/mL stock in DMF; 175 µg/mL BCIP from 50 mg/mL stock in water; all diluted in 10% (w/v) PVA in water

Fluorescent colour detection buffer: 0.1 M Tris; 1 mg/mL Fast Red TR; 0.4 mg/mL Napthol AS-MX; 0.15 mg/mL Levamisol; final pH 7.9-8.5

Stop solution: 0.1 M glycine pH 2.2 from 1 M stock; 0.1% (v/v) Tween-20

PFA: 4% (v/v) PFA in 1x PBTw

60% glycerol: 60% glycerol (v/v) in water

2.4 Results and Discussion

Previously described WMISH protocols for molluscan embryos and larvae did not yield consistent or satisfactory WMISH signals in *L. stagnalis* [14, 21, 22]. Therefore, we focused on a few key steps of sample preparation we believed to cause background, weak WMISH signals and non-specific staining. Note that in this work we did not explore the effect of hybridisation temperature on the final result. Initial experiments with the probes used in the present study revealed that 55 °C produced consistent and acceptable results, allowing us to focus on systematically optimising other parameters. Of course, hybridisation temperature should be empirically optimised for every probe and has the potential to significantly improve or impair the final result. A summary of the treatments we found to generate the clearest WMISH signals for each developmental stage (a "protocol at a glance") is provided in Additional file S. 2.2. The results of control experiments using no antibody and no riboprobe (which generated no signals) are provided in Additional file S. 2.3.

2.4.1 The effect of NAC treatment

The fluid that bathes *L. stagnalis* larvae during their encapsulated development is characterised by a high viscosity and adheres to the embryo following de-capsulation. An incubation step with the mucolytic reagent NAC apparently leads to a superior preservation of the overall morphology (Fig. 2.2 and Additional file S. 2.3C *cf.* E and S. 2.3I *cf.* K). However, treatment with NAC resulted in a significant reduction of signal intensity for all ages and genes that we investigated (Fig. 2.2C, G, K, O and S). However, when NAC treatment was combined with a reduction step this effect was reversed for some combinations of riboprobe and developmental stage (Fig. 2.2H, P and T). The combined NAC and reduction treatment gave the best signal to noise ratio for *beta tubulin* in three to six dpfc larvae (Fig. 2.2H), and for *COE* in two to three dpfc old larvae (Fig. 2.2T). This is in contrast to the situation for *engrailed* in all investigated ages: under the appropriate reduction treatment, omitting the NAC generated a better signal to noise ratio (Fig. 2.2J and N) than including it (Fig. 2.2K, L, O and P).
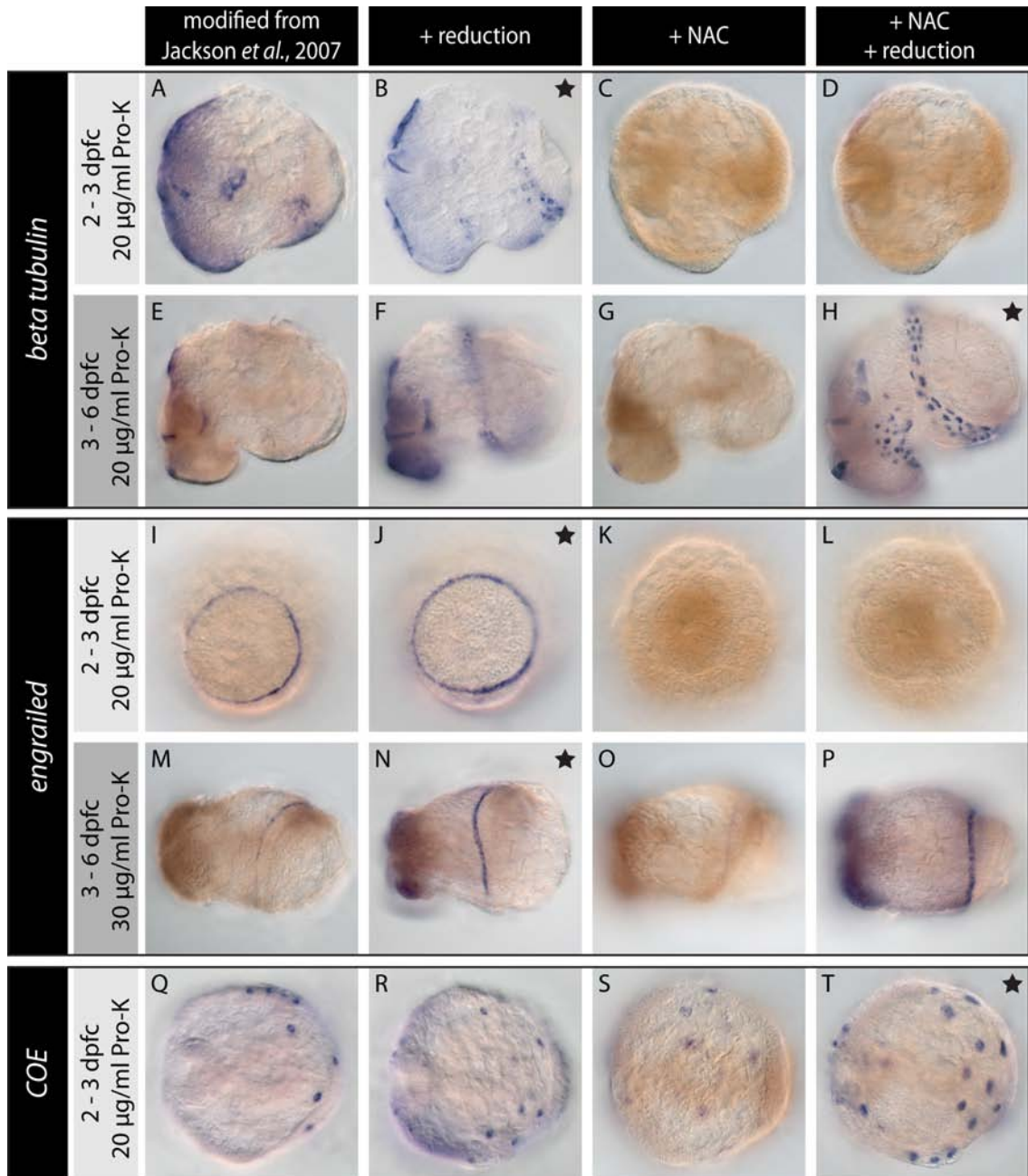
Fig. 2.2. Overview of WMISH signals generated after pre-hybridisation treatment with NAC and/or reduction. *L. stagnalis* larvae of different ages were subjected to a WMISH protocol similar to that described by Jackson *et al*. [22] **(A, E, I, M and Q)**. This protocol was then modified by the addition of a reduction treatment **(B, F, J, N and R)**, a NAC treatment **(C, G, K, O and S)** or a combination of both NAC and reduction treatment **(D, H, L, P and T)**. Using this set of pre-hybridisation treatments, the optimal sample preparation regime for WMISH varies with respect to the target gene and the developmental stage. For *engrailed* and *beta tubulin* in younger larvae, the samples that underwent a reduction treatment display the best signal to noise ratio **(B, J and N)**. Excess background that is revealed by reduced samples for *COE* and *beta tubulin* in older larvae **(F and R)** is diminished by a treatment with NAC **(H and T)**. Black stars indicate optimal results after sample preparation involving NAC treatment and reduction. Panels **A** to **H** show larvae from a lateral perspective with the shell field oriented to the right. Larvae in **I** to **P** are viewed from dorsal and **Q** to **T** are viewed from apical.

Our overall impression of treating *L. stagnalis* larvae with NAC is that this treatment may be beneficial in combination with a reduction step when working with probes, which tend to generate non-specific background. The signal diminishing effect of NAC is in contrast to the situation in the planarian *Schmidtea mediterranea*. Here, a NAC treatment is used to remove the planarian's surrounding mucous layer and generally increases the WMISH staining intensity, at least when combined with a permeabilisation step using SDS or DTT [17]. In *L. stagnalis*, NAC may be removing the intra-capsular fluid, however it appears that in our hands NAC is most likely reducing WMISH signal strength by significantly inhibiting the activity of Pro-K; larvae that were incubated in Pro-K and 1% NAC at the same time did not show any signs of compromised morphology, while larvae in control reactions with Pro-K but with less (0.1%) or without NAC were digested (Additional file S. 2.4).

2.4.2 Treatment with DTT and detergents (reduction)

A treatment using a solution containing DTT and the detergents SDS and NP-40 following fixation greatly increased WMISH signal intensity for all investigated genes and developmental stages (Fig. 2.2). The best WMISH signal for *beta tubulin* in four to six dpfc old larvae was achieved using a combination of NAC and reduction (Fig. 2. 2H). This suggests that the reduction treatment might represent a highly effective permeabilisation approach. However, this was at the cost of all material becoming highly fragile until dehydrated in ethanol. Reduced samples were also more likely to reveal unspecific background staining (Fig. 2.2F, N and R).

2.4.3 SDS treatment

Between one and five dpfc old embryos and larvae of *L. stagnalis* were treated with different amounts of the anionic detergent SDS prior to hybridisation (Fig. 2.3). A permeabilising treatment with 0.1% SDS did not produce strong WMISH signals for all studied genes and larval ages (Fig. 2.3A, D, G, J, M, P and S) whereas treatments with higher concentrations of SDS generated strong WMISH signals. For two of the genes we studied here, *beta tubulin* and *engrailed*, treating larvae between three to five dpfc with 0.5% or 1% SDS produced equally good results. In contrast, the staining was more intense after treatment with 0.5% SDS than with
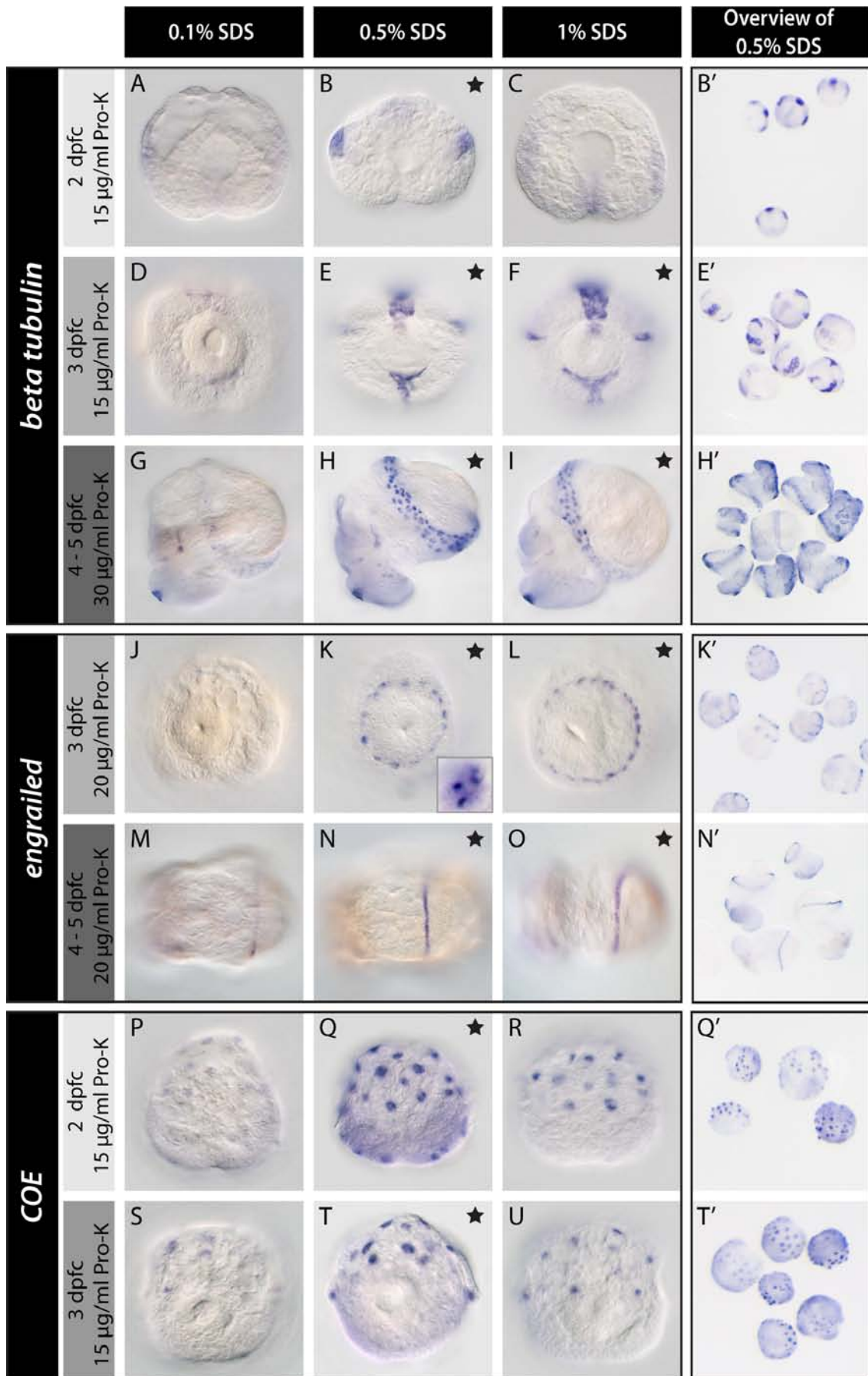
Fig. 2.3. A pre-hybridisation treatment with different SDS concentrations significantly affects the WMISH signal. *L. stagnalis* larvae of different ages were subjected to pre-hybridisation treatments with varying amounts of SDS and then hybridised with anti-sense probes to *beta tubulin* **(A-I)**, *engrailed* **(J-O)** and *COE* **(P-U)**. For all genes and larval ages, treatment with 0.1% SDS did not generate consistent or strong WMISH signals **(A, D, G, J, M, P and S)**. Treatments with both 0.5% and 1% SDS produced strong WMISH signals for *beta tubulin* and *engrailed* in larvae aged three to five days post first cleavage (dpfc), with high spatial resolution **(inlet in *K*)**. For *COE* 0.5% SDS outperformed the 1% SDS treatment **(T vs. U)**. Black stars indicate optimal treatments. Note that some treatments produced equally good results. The most consistent results (defined as constantly good signals among genes and ontogenetic stages with little variation between individuals within an experiment) were achieved with 0.5% SDS **(examples shown in B', E', H', K', N', Q' and T')**. Larvae in **A-C** and **M-R** are shown from an apical perspective, larvae in **D-F** are viewed ventrally, **G-I** laterally and **J-L** and **S-U** dorsally.

1% SDS for *COE* (Fig. 2.3Q vs. R and T vs. U) as well as for *beta tubulin* in two dpfc old larvae (Fig. 2.3B vs. C), which may suggest a loss of the target transcripts due to excess permeabilisation of these younger stages. Additionally, embryos between one and two dpfc tend to adhere toplastic surfaces in 1% SDS. While in other animal systems SDS is commonly used at a concentration of 1% [18-20], in *L. stagnalis* strong WMISH signals were most consistently achieved with 0.5% SDS across different genes and ontogenetic stages.

The WMISH signals for *engrailed*, *COE* and *beta tubulin* (the latter at least in younger larvae) in SDS-treated larvae reveal equally good or superior signal intensities compared to reduced or NAC-treated and reduced larvae (Fig. 2.2 *cf.* Fig. 2.3). The *engrailed* signal in the shell field of SDS-treated larvae also clearly exhibits a significantly improved spatial resolution compared to the signal in reduced larvae (Fig. 2.2J *cf.* Fig. 2.3K). In terms of ease of handling, non-specific background and consistency of WMISH signal among the different genes, the SDS treatment is our recommended sample preparation strategy. A reduction treatment might increase the signal intensity in WMISH experiments against genes expressed in older larvae (Fig. 2. 2H), but should then be performed in parallel to the SDS treatment to control for a possible loss of spatial resolution

2.4.4 Enzymatic permeabilisation with Pro-K

The optimal Pro-K concentration for tissue permeabilisation and mRNA target unmasking depends on the incubation temperature, incubation time and the ontogenetic stage of the material. In general, under-treatment yields weak WMISH signals while over-treatment increases background staining and leads to a compromised tissue morphology. We tested a wide range of Pro-K concentrations

in combination with different pre-hybridisation treatments and found that treatment with Pro-K at concentrations of 0–2 μg/mL drastically reduced signal intensity while treatment at concentrations of 40–50 μg/mL compromised tissue integrity (data not shown). Therefore, we finally used concentrations of 1–15 μg/mL for one to two dpfc old larvae, 5–20 μg/mL for two to three dpfc old larvae and 5–40 μg/mL for larvae between three and five dpfc in combination with different SDS concentrations (data not shown). For embryos between one and two dpfc we found the highest concentration of Pro-K of 15 μg/mL gave the strongest signals (Fig. 2.3B and Q). For two and three dpfc larvae, 15 μg/mL or 20 μg/mL Pro-K were suitable (Fig. 2.3E, F, K, L and T). In larvae between three and five dpfc the best signal to noise ratio was achieved with 30 μg/mL for *beta tubulin* (Fig. 2.3H and I) and 20 μg/mL for *engrailed* (Fig. 2.3N and O). *COE* is apparently not expressed in larvae between three and five dpfc.

2.4.5 Removal of non-specific background with triethanolamine + acetic anhydride (TEAAA)

In preliminary experiments using a wide variety of riboprobes to different genes we obtained a strong, well-defined WMISH signal located at the periphery of the shell field and in the radular sac (Fig. 2.4, black arrows). To determine whether these staining patterns represented genuine probe/target hybridisation events, or non-specific binding of the probe, we treated fixed larval material with RNAse A prior to hybridisation. As a control, all expected *beta tubulin* staining was abolished following this RNAse treatment, confirming the degradation of all target mRNA (Fig. 2.4E). In RNAse treated embryos hybridised with probes that generated the spurious shell field and radula patterns this signal was still present (Fig. 2.4F-H), indicating these WMISH patterns represent high affinity binding of the riboprobe to molecules other than RNA. In order to address this background staining, we assessed the effect of a triethanolamine + acetic anhydride (TEAAA) treatment. Treatment of biological substrates with TEAAA is a common practice for many WMISH protocols, and decreases non-specific binding of labelled probes through the acetylation of polar and charged groups [23]. For *L. stagnalis*, treatment with TEAAA following Pro-K digestion successfully abolished the non-specific WMISH signal in the larval shell and in the radular sac (Fig. 2.4J-L). Due to
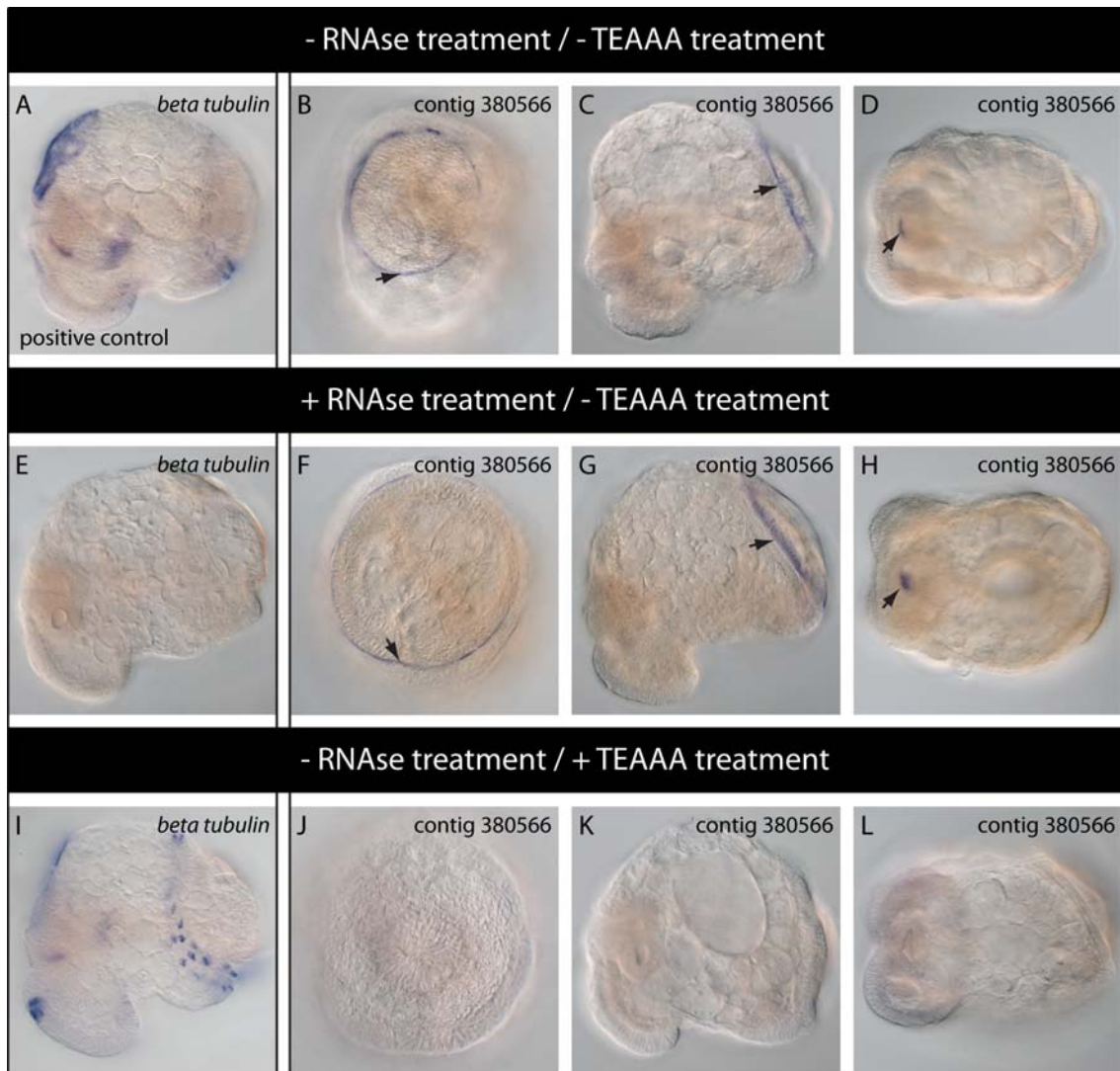
Fig. 2.4. Non-specific probe binding to the shell field and radular is eliminated by treatment with TEAAA. We observe a well-defined and consistent WMISH stain for a variety of probes (represented here with a probe against the gene "contig 380566") in the periphery of the shell field **(arrows in B and C)** and in the radular sac **(arrow in D)**. Probes against other genes (for example *beta tubulin*) do not produce these patterns **(A)**. The stain in the shell field periphery and the radular remains following a pre-hybridisation treatment with RNAse **(F-H)**, while the specific signal against *beta tubulin* is abolished **(E)**. This indicates that the signals in the shell field and the radular are the result of non-specific probe binding. Treatment with TEAAA abolishes this non-specific stain **(J-L)**, while the specific signal against *beta tubulin* remains unaltered **(I)**. **B, F** and **J** are dorsal views. **A, C, E, G, I** and **K** are lateral views of larvae with the shell gland oriented to the right. **D, H** and **L** are ventral views. Panels **C, G** and **K** are reflected about the vertical axis for consistency of presentation.

its strength, consistency and spatial definition, this tissue-specific background stain is particularly likely to be misinterpreted as genuine WMISH signal. A good example of this is *engrailed*, which is genuinely expressed in the shell field periphery (Fig. 2.3J-O) and also produces the shell field background stain (data not shown). Our riboprobe against *engrailed* (synthesised multiple times) covers the same region as the probe used in a recent study that possibly produced the same

background staining [14]. Therefore, treatment with TEAAA appears to be critical for the correct interpretation of genes with expression patterns associated with the shell gland and shell field.

General background staining and its elimination by TEAAA can also be observed for samples that underwent a NAC (Fig. 2.2) or SDS treatment (Fig. 2.3). For some genes, the TEAAA treatment can be shortened by one incubation step in TEAAA instead of two incubation steps (Additional file S. 2.5).

2.4.6 The effect of antibody concentration, PVA and riboprobe hydrolysis

Since the dilution of AP-conjugated anti-DIG antibody in our optimised protocol (1/10,000) is lower than described in many WMISH protocols, we assessed the effect of increasing the concentration of antibody to 1/3,000. While the signal intensities of *beta tubulin* and *engrailed* expression were slightly higher, more overall non-specific background staining was also evident (Fig. 2.5B, G, L and Q).

A common approach to improve WMISH sensitivity and reduce background staining is through the addition of PVA to the colour detection solution. This is thought to increase the local concentration of the colour reaction product by limiting its diffusion [24]. A direct comparison of WMISH colour development in *L. stagnalis* with and without PVA did not reveal a significant increase in the signal intensities, but rather a lower signal to noise ratio for both *beta tubulin* and *engrailed* (Fig. 2.5C, H, M and R). Furthermore, the morphological integrity of especially older larvae was compromised (Fig. 2. 5H).

An alternative strategy to increase WMISH signal sensitivity is to hydrolyse the riboprobe before use [23, 24]. Hydrolysing the riboprobe into smaller fragments is thought to facilitate better tissue penetration and to improve hybridisation kinetics (reviewed in [23, 25]). We specifically tested whether a hydrolysed riboprobe generates an improved *engrailed* signal and found that instead both the signal specificity (as revealed by the lack of the shell field lining expression pattern, Fig. 2.5S) and overall intensity were reduced compared to non-hydrolysed control reactions (Fig. 2.5P vs. S). Reduced signal sensitivity derived from hydrolysed riboprobes has been previously reported [26], highlighting the
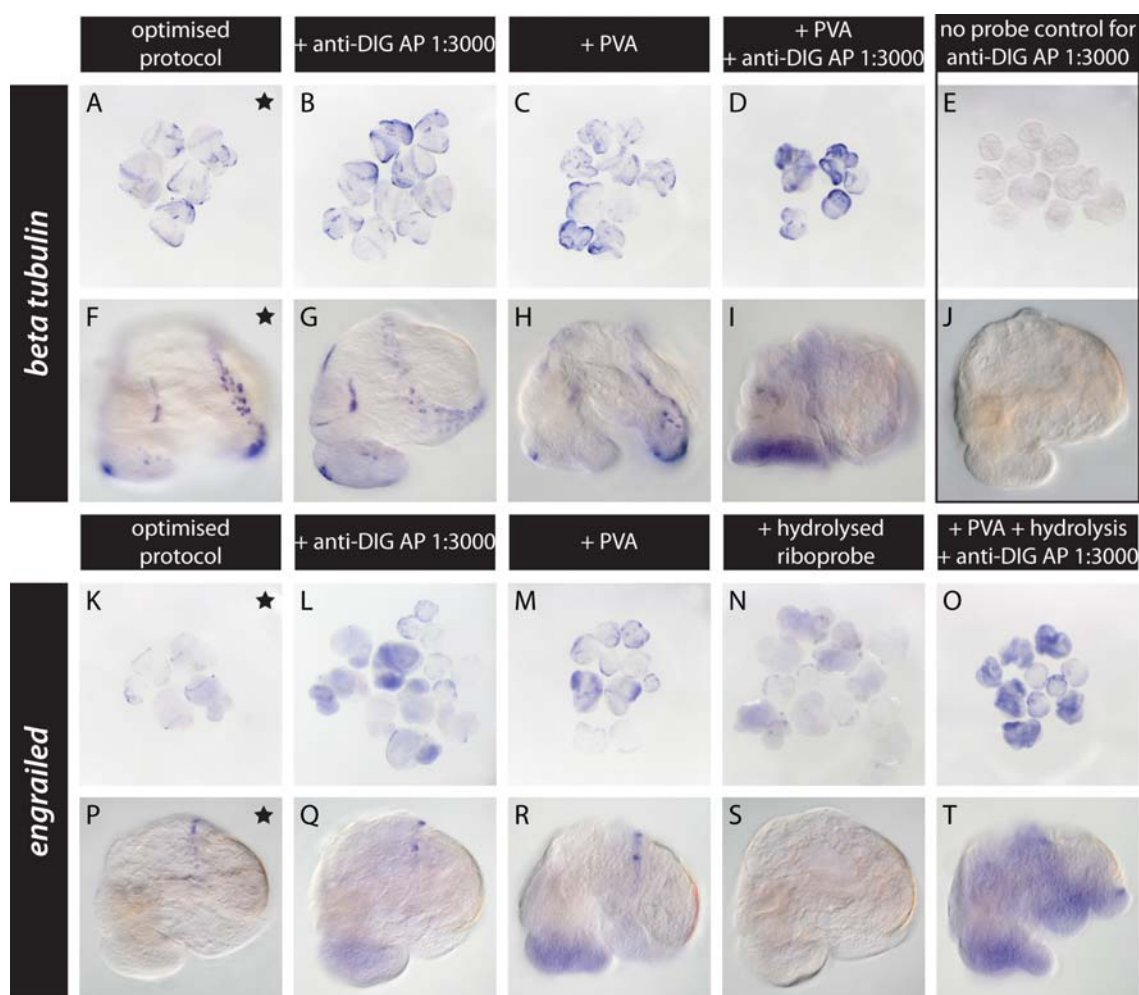
Fig. 2.5. Our optimised WMISH protocol is not improved by more antibody, PVA or hydrolysed riboprobes. Larvae three to four days post first cleavage (dpfc) were subjected to our optimised WMISH protocol **(A, F, K and P)**. Using a *beta tubulin* probe, we investigated the effect of increasing the amount of anti-DIG antibody **(B, G, L and Q)**, the addition of PVA to the colour detection solution **(C, H, M and R)** and the combination of more antibody and PVA **(D and I)**. We also assessed the effect of hydrolysing the *engrailed* riboprobe individually **(N and S)** and in combination with a higher antibody concentration and the use of PVA **(O and T)**. None of these modifications generated superior WMISH results to our baseline protocol. Samples incubated in more antibody and developed with PVA showed slightly more intense signals, but a lower signal to noise ratio **(B, C, G, H, L, M, Q and R)**. PVA also appeared to compromise the morphological integrity of older larvae **(H, I and R)**. Signals generated by the hydrolysed *engrailed* probe were much fainter and were partially masked by an increase in general background staining **(N, O, S and T)**. The optimal treatment **(A, F, K, and P)** is indicated by a black star. Control WMISH experiments lacking a riboprobe and using the increased antibody concentration do not reveal any staining **(E and J)**. All images of individual larvae are lateral views.

necessity to test the optimal probe preparation strategy for each gene individually [27].

To summarise these treatments, neither more antibody nor the use of PVA or hydrolysed riboprobes generated WMISH signals with higher sensitivity and/or signal to noise ratios.

2.4.7 Fluorescent WMISH signal detection

The optimised protocol that we have identified also allows for the visualisation of fluorescent signals using a confocal microscope. This method of detection provides a much higher spatial resolution than colorimetric methods. Using probes against all three of the genes employed in this study we were able to develop fluorescent signals using Fast Red (Fig. 2.6).
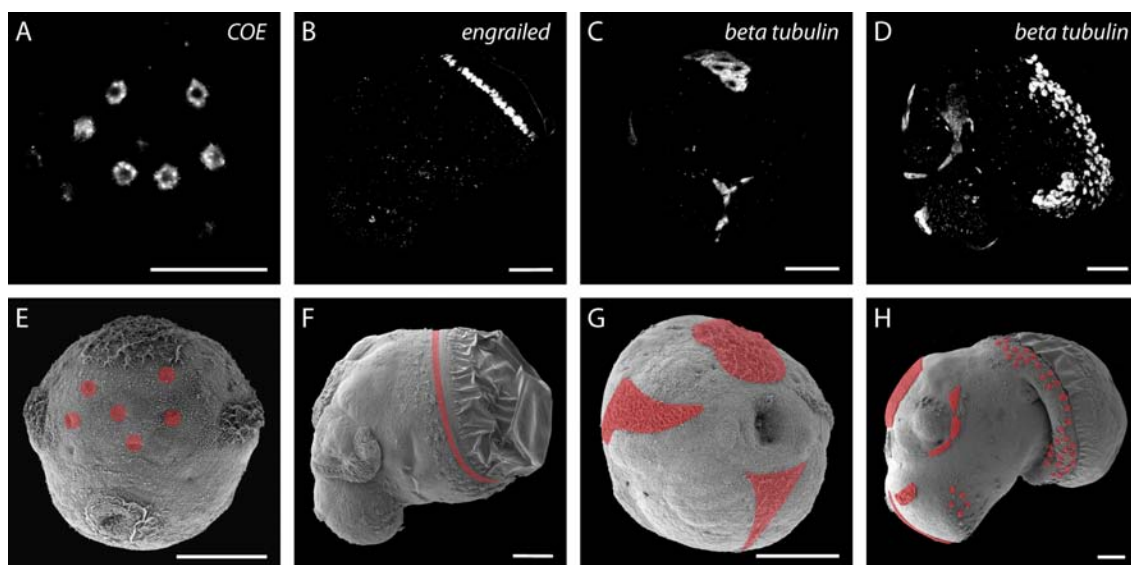
Fig. 2.6. Our WMISH protocol is suitable for fluorescent signal detection. The expression of *COE* **(A)**, *engrailed* **(B)** and *beta tubulin* **(C and D)** in larvae treated with 0.5% SDS was detected using the fluorescent substrate Fast Red **(A-D)**. Panel E is a scanning electron micrograph of a 62 hours post first cleavage (hpfc) old larva with the approximate localisation of the *COE* expression indicated in red. Panel **F** is a lateral perspective of a 90 hpfc old larva with indicated *engrailed* expression. Panels **G** and **H** show the positions of *beta tubulin* expression in a 57 hpfc old larva **(G, ventral perspective)** and a 100 hpfc old larva **(H, lateral perspective)**. Panels **B-D** are projections of confocal laser scanning micrographs. Panels **B**, **D** and **H** are reflected about the vertical axis for consistency of presentation. All scale bars are 50 μm.

2.4.8 The effect of storing fixed material in Ethanol vs. Methanol

Methanol (MeOH) is used to dehydrate and store fixed embryonic and larval material at low temperatures (−20°C) in many WMISH protocols. Due to the high toxicity of MeOH relative to EtOH we assessed the effect of storing fixed *L. stagnalis* larvae in MeOH vs. EtOH on the WMISH signal generated by *beta tubulin*, *engrailed* and *COE*. We found no consistent or significant difference with respect to any of the signals generated (data not shown).

2.5 Conclusions

Our study provides an optimised whole mount *in situ* hybridisation protocol for early larval stages of the mollusc *L. stagnalis*. Using a variety of pre-hybridisation treatments we have identified a set of conditions that allow for high WMISH signal intensity and consistency in colorimetric as well as fluorescent WMISH. These include a treatment with 0.5% SDS, treatment of one to two dpfc larvae with 15 μg/mL Pro-K, two to three dpfc larvae with 15–20 μg/mL Pro-K and three to five dpfc larvae with 20–30 μg/mL, followed by treatment with TEAAA. We also demonstrate that non-specific shell field and radula staining can easily be abolished with this TEAAA treatment. In our experience, every riboprobe/developmental stage combination benefits from an individualised protocol, which needs to be empirically determined. Nonetheless we believe that this WMISH protocol should serve as a baseline method from which consistent and clearly visible patterns of gene expression can be obtained. This method should serve to raise the profile of *L. stagnalis* as a tractable experimental molluscan model, a niche that is currently underpopulated.

Additional files

The additional files for this article can be found online at: https://bmcdevbiol.biomedcentral.com/articles/10.1186/s12861-015-0068-7 or on the attached CD in the printed version.

Additional file S. 2.1: Primer sequences used to isolate gene fragments for riboprobe syntheses.

Additional file S. 2.2: WMISH "protocol at a glance".

Additional file S. 2.3: Control experiments for the optimised sample preparations. Control WMISH experiments lacking riboprobe (A-F) or antibody (G-N) demonstrate the absence of any non-specific colour reaction for samples treated with SDS (A, B, G, H, M and N) or with reduction solution (C, D, I and J) as well as reduced + NAC treated samples (E, F, K and L) for about three dpfc old larvae (A, C, E, G, I, K and M) and about five dpfc old larvae (B, D, F, H, J, L and N). Panels A to L were colour-developed using NBT/BCIP as colour substrate and panels M and N were developed using Fast Red. All embryos are shown from a lateral perspective. Scale bars are 50 μm (A, C, E, G, I, K and M) or 100 μm (B, D, F H, J, L and N). Panels C, E, G and H are reflected about the vertical axis for consistency of presentation.

Additional file S. 2.4: Proteinase-K activity is inhibited by NAC. Larvae incubated in Pro-K without NAC (A-D) or in 0.1% NAC (E-H) are almost completely digested after 4 and 22 hours of incubation respectively. In contrast, larvae incubated in Pro-K with 1% NAC (I-L) do not show

any signs of Pro-K digestion and maintain their morphology even over extended incubation times (L). All larvae are about 4 days post first cleavage old. All images are to the same scale shown in L (1 mm).

Additional file S. 2.5: A shortened treatment with TEAAA is sufficient to minimise non-specific probe binding in SDS-treated samples. The background stain in the shell field periphery (identified in Fig. 2. 3) is also observed for SDS-treated samples (arrows in A), as represented by a probe against the gene "contig 380566". Note that after treatment with SDS, the protonephridia are stained (arrow in B). Both non-specific WMISH stains are strongly reduced after one incubation step in TEAAA (D-F) and disappear after two incubation steps (G-I). Larvae in A, D and G are shown from a dorsal perspective, and larvae in B, E and H are viewed from lateral. Panels B, E and H are reflected about the vertical axis for clarity of presentation.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JH carried out the experimental procedures. IH assisted with the experimental procedures. DJJ conceived the study. JH and DJJ drafted the manuscript. All authors participated in the design of the study. All authors read and approved the final manuscript.

Acknowledgements

References

1. Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D. Ancient animal microRNAs and the evolution of tissue identity. Nature. 2010;463:1084-1088.
2. Arendt D. The evolution of cell types in animals: emerging principles from molecular studies. Nature Reviews Genetics. 2008;9:868-882.
3. Hashimoto N, Kurita Y, Wada H. Developmental role of *dpp* in the gastropod

shell plate and co-option of the *dpp* signaling pathway in the evolution of the operculum. Developmental biology. 2012;366:367-373.

4. Raven CP. Morphogenesis in *Limnaea stagnalis* and its disturbance by lithium. Journal of Experimental Zoology. 1952;121:1-77.

5. Timmermans LPM. Studies on shell formation in molluscs. Netherlands Journal of Zoology. 1969;19:413-523.

6. Kniprath E. Zur Ontogenese des Schalenfeldes von *Lymnaea stagnalis*. Wilhelm Roux's Archives of Developmental Biology. 1977;181:11-30.

7. Liu MM, Davey JW, Banerjee R, Han J, Yang F, Aboobaker A, Blaxter ML, Davison A. Fine mapping of the pond snail left-right asymmetry (chirality) locus using RAD-Seq and fibre-FISH. PLoS One. 2013;8:e71067.

8. Hohagen J, Jackson DJ. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. BMC Developmental Biology. 2013;13:27.

9. Voronezhskaya EE, Khabarova MY, Nezlin LP. Apical sensory neurones mediate developmental retardation induced by conspecific environmental stimuli in freshwater pulmonate snails. Development. 2004;131:3671-3680.

10. Raven CP. The Development of the Egg of *Limnaea Stagnalis* L. From the First Cleavage Till the Troghophore Stage, With Special Reference To Its "Chemical Embryology". Archives Néerlandaises de Zoologie. 1946;7:353-434.

11. Elbers PF, Bluemink JG. Pinocytosis in the developing egg of *Limnaea stagnalis* L. Experimental Cell Research. 1960;21:619-622.

12. Morrill JB. Development of the pulmonate gastropod, *Lymnaea*. In: Harrison FW, Cowden RR, editors. New York: Alan R. Liss; 1982. p. 399-483.

13. Bayne CJ. Histochemical studies on the egg capsules of eight gastropod molluscs. Journal of Molluscan Studies. 1968;38:199-212.

14. Iijima M, Takeuchi T, Sarashina I, Endo K. Expression patterns of *engrailed* and *dpp* in the gastropod *Lymnaea stagnalis*. Development Genes and Evolution. 2008;218:237-251.

15. Shimizu K, Sarashina I, Kagi H, Endo K. Possible functions of Dpp in gastropod shell formation and shell coiling. Development Genes and Evolution. 2011;221:59-68.

16. Shimizu K, Iijima M, Setiamarga DHE, Sarashina I, Kudoh T, Asami T, Gittenberger E, Endo K. Left-right asymmetric expression of *dpp* in the mantle of gastropods correlates with asymmetric shell coiling. EvoDevo. 2013;4:15.

17. Pearson BJ, Eisenhoffer GT, Gurley KA, Rink JC, Miller DE, Sánchez Alvarado A. Formaldehyde-based whole-mount *in situ* hybridization method for planarians. Developmental Dynamics. 2009;238:443-450.

18. King RS, Newmark PA. *In situ* hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea*. BMC Developmental Biology. 2013;13:8.

19. Browne WE, Schmid BGM, Wimmer EA, Martindale MQ. Expression of *otd* orthologs in the amphipod crustacean, *Parhyale hawaiensis*. Development Genes and Evolution. 2006;216:581-595.

20. Price AL, Patel NH. Investigating divergent mechanisms of mesoderm development in arthropods: the expression of *Ph-twist* and *Ph-mef2* in *Parhyale hawaiensis*. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 2008;310:24-40.

21. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM. A

rapidly evolving secretome builds and patterns a sea shell. BMC Biology. 2006;4:40.

22. Jackson DJ, Wörheide G, Degnan BM. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. BMC Evolutionary Biology. 2007;7:160.

23. Zeller R, Rogers M, Haramis AG, Carrasceo AE. *In situ* hybridization to cellular RNA. Current Protocols in Pharmacology. 1989A-3F.

24. Jowett T. Double *in situ* hybridization techniques in zebrafish. Methods. 2001;23:345-358.

25. Hauptmann G, Gerster T. Multicolor whole-mount *in situ* hybridization. Developmental Biology Protocols. 2000139-148.

26. Shain DH, Zuber MX. Sodium dodecyl sulfate (SDS)-based whole-mount *in situ* hybridization of *Xenopus laevis* embryos. Journal of Biochemical and Biophysical Methods. 1996;31:185-188.

27. Jezzini SH, Bodnarova M, Moroz LL. Two-color *in situ* hybridization in the CNS of *Aplysia californica*. Journal of Neuroscience Methods. 2005;149:15-25.

3

A whole mount *in situ* hybridization method for the gastropod mollusc *Lymnaea stagnalis*

Daniel J. Jackson, Ines Herlitze, Jennifer Hohagen
JoVE. 2016e53968.

3.1 Abstract

Whole mount *in situ* hybridization (WMISH) is a technique that allows for the spatial resolution of nucleic acid molecules (often mRNAs) within a 'whole mount' tissue preparation, or developmental stage (such as an embryo or larva) of interest. WMISH is extremely powerful because it can significantly contribute to the functional characterization of complex metazoan genomes, a challenge that is becoming more of a bottleneck with the deluge of next generation sequence data. Despite the conceptual simplicity of the technique much time is often needed to optimize the various parameters inherent to WMISH experiments for novel model systems; subtle differences in the cellular and biochemical properties between tissue types and developmental stages mean that a single WMISH method may not be appropriate for all situations. We have developed a set of WMISH methods for the re-emerging gastropod model *Lymnaea stagnalis* that generate consistent and clear WMISH signals for a range of genes, and across all developmental stages. These methods include the assignment of larvae of unknown chronological age to an ontogenetic window, the efficient removal of embryos and larvae from their egg capsules, the application of an appropriate Proteinase-K treatment for each ontogenetic window, and hybridization, post-hybridization and immunodetection steps. These methods provide a foundation from which the resulting signal for a given RNA transcript can be further refined with probe specific adjustments (primarily probe concentration and hybridization temperature).

3.2 Video Link

The video component of this article can be found at http://www.jove.com/video/53968/

3.3 Introduction

Molluscs are a group of animals that hold the interest of a broad diversity of scientific disciplines. Despite their morphological diversity [1], species richness (second only to the Arthropods in terms of species number [2]) and relevance to a wide range of commercial [3], medical [4] and scientific issues [5-8], there are relatively few molluscan species that can claim to be both well-equipped scientific models and easy to maintain in a laboratory environment.

One mollusc that is much used by disciplines such as neurobiology [9], ecotoxicology [10] and more recently evolutionary biology [11, 12] is *Lymnaea stagnalis*, primarily because of its widespread distribution and extreme ease of maintenance. Despite its popularity as a 'model' organism and its long history of use by developmental biologists [13-16] [17-19], the range and power of molecular tools available to the *L. stagnalis* scientific community lies far behind that of more traditional animal models (*Drosophila*, mouse, sea urchin, nematodes).

Our desire to develop *Lymnaea* as a molecular model stems from an interest in the molecular mechanisms that guide shell formation. This motivated us to refine a set of techniques that would allow for the efficient, consistent and sensitive visualization of gene expression during *Lymnaea's* development. Whole mount *in situ* hybridization (WMISH) is widely employed for a variety of model organisms and has been in use for more than 40 years [20]. In its different guises, ISH can be employed to spatially localize specific loci on chromosomes, rRNA, mRNA and micro-RNAs.

One of the challenges we needed to address prior to refining a WMISH method for *L. stagnalis* was the issue of gently and efficiently extracting embryos and larvae of varying stages from the egg capsules in which they are deposited. This extraction, or 'decapsulation', needs to be achieved efficiently in order to collect adequate material for a given *in situ* experiment, while at the same time maintaining

morphological and cellular integrity. While other model organisms also undergo encapsulated development, in our hands none of the methods reported for those species could be successfully employed in *L. stagnalis.*

The overall goals of this method are therefore: to extract *L. stagnalis* embryos and larvae from their capsules in a high-throughput fashion, to apply pre-hybridization treatments that optimize the WMISH signal, to prepare embryos and larvae with satisfactory WMISH signals for imaging.

3.4 Protocol

NOTE: The following steps outline our method for conducting an *in situ* experiment on embryonic and larval stages of *L. stagnalis.* Where a step involves the use of a hazardous chemical this is indicated by the word 'CAUTION' and all appropriate safety procedures should be adopted. Links to representative MSDS sheets for hazardous chemicals are provided in Supplementary File S. 3.1. Recipes for all reagents are provided in Supplementary File S. 3.2.

3.4.1. Assembly of Decapsulation Apparatus

1. To do this, connect a 20 mL disposable syringe to silicon tubing (with an inner diameter of 1 mm and an outer diameter of 3 mm) using a P1,000 tip cut to an appropriate length as shown in Fig. 3.2A.

2. Tape a standard microscope slide to an inverted large Petri dish. Tape the silicon tubing immediately adjacent to the microscope slide as shown in Fig. 3.2C and 3.2D.

3. Rest a pulled glass needle on the microscope slide and gently insert it into the silicon tubing until the tip of the needle protrudes approximately 20% of the way across the inner diameter of the tubing (see Fig. 3.2C and 3.2D). Once the needle is in position tape it down to the Petri dish.

4. Allow the free end of the silicon tubing to rest in another Petri dish that will collect the decapsulated material.
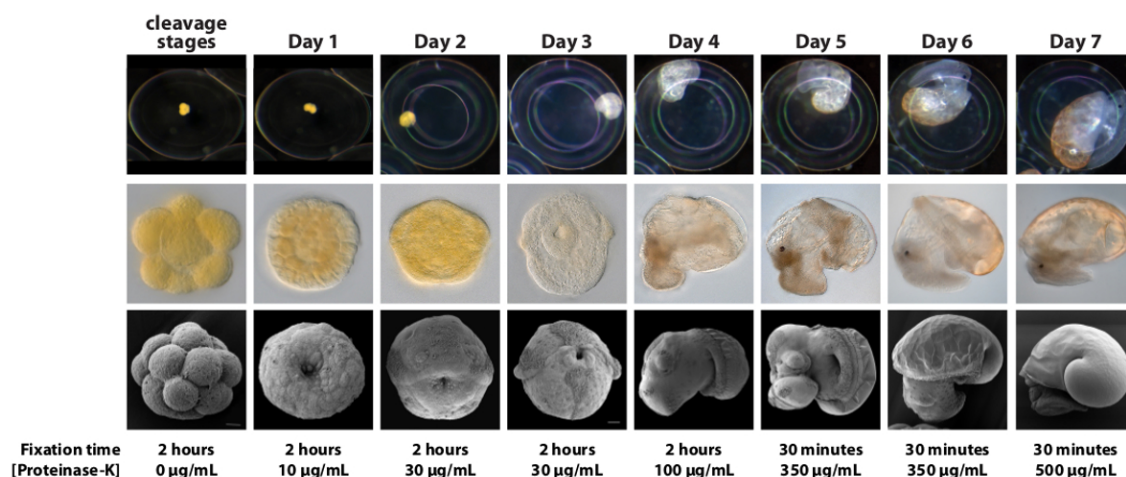
Fig. 3.1. A Summarized Ontogeny of *Lymnaea stagnalis* and Corresponding Fixation and Proteinase-K Treatments. Representative images of embryos and larvae from the first 7 days of development illustrate a significant increase in size (row 1) and morphological complexity (rows 2 and 3). These developmental changes translate into large differences in the appropriate fixation time and Proteinase-K concentrations that generate optimal WMISH signals. For WMISH all stages should be fixed in 3.7% formaldehyde in PBS at RT with gentle agitation within their egg capsules. All stages should then be treated with the appropriate Proteinase-K concentration for 10 min. Note that we observe significant inter-batch variation in the activity of Proteinase-K from our supplier. This variation must be accounted for by performing a round of 'calibrating' WMISH experiments where the activity of the new Proteinase-K is empirically determined. The Proteinase-K concentrations stated in the figure should therefore be treated as an initial guide, however the relative concentrations between developmental stages (for example 2 day old embryos require a Proteinase-K concentration 3 times higher than day 1 embryos) are set.

3.4.2. Sample Collection, Fixation and Decapsulation

NOTE: All steps are carried out at RT unless otherwise noted.

1. Carefully collect egg strings from the walls of an aquarium. To do this, use a flat piece of flexible plastic as a spatula to scrape the egg string off the substrate, and use a plastic tea strainer to fish the floating egg string out of the water. Stage and sort the material under a microscope using the guide provided in Fig. 3.1.

2. Place the egg string onto a paper towel and make a longitudinal incision along the egg mass using the featherweight forceps. Roll the egg capsules out of the egg string and remove as much of the jelly material as possible from each capsule by pushing them around the paper towel using the featherweight forceps.

3. Using the featherweight forceps, transfer the egg capsules into a Petri dish containing 5 mL of tap water. Continue to collect de-jellied egg capsules of the desired developmental stages into this dish. Collect enough capsules from all

developmental stages for the planned WMISH experiment then proceed to the next step.

4. When working with more developed larvae (5 days post first cleavage (dpfc) and older) anaesthetize them prior to fixation.

   NOTE: This will prevent muscles from contracting which makes the interpretation of *in situ* staining patterns very difficult.

   1. Relax larvae (while they are still in their capsules) in a 2% w/v solution of $MgCl_2 \bullet 6H_2O$ for 30 min prior to fixation.

   2. Assess the degree of relaxation after 30 min by submerging several larvae while still in their egg capsules in fixative solution and monitoring their response under a microscope. Incompletely relaxed larvae will retract into their shells, while fully relaxed larvae will not respond. Once these larvae have been relaxed proceed to the next step.

5. Transfer the egg capsules using a wide bore plastic pipette into a sealable tube that provides 10 times the volume of egg capsules (*e.g.*, 1 mL of settled capsules would require a 10+ mL tube).

6. Aspirate as much liquid as possible from the tube and replace with a volume of fixative solution that is 10x the volume of the settled egg capsules. Gently rotate the egg capsules in fixative at RT for the appropriate time for each developmental stage (see Fig. 3.1).

7. Discontinue rotation and allow the capsules to sink and aspirate the fixative solution into an appropriate waste container.

8. Wash the egg capsules by replacing the fixative solution with Phosphate Buffered Saline with 0.1% Tween-20 (PBTw) and rotating at RT for 5 min. Aspirate the PBTw and repeat twice.

9. Remove embryos and larvae from their capsules using the apparatus (see section 1 for details) shown in Fig. 3.2. Draw the capsules up into the 20 mL syringe, attach the tubing and then dispel the capsules through the tubing and past the glass needle out into the collection dish.
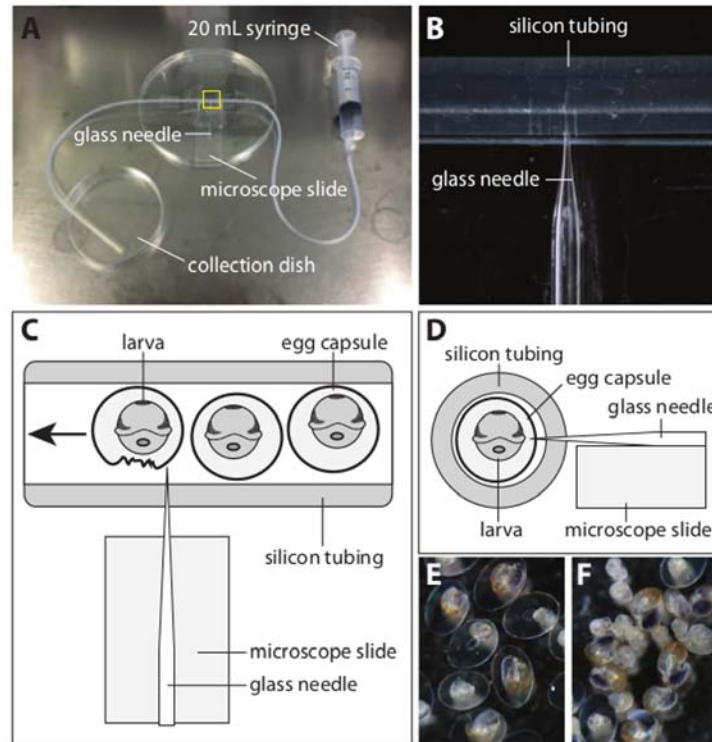
Fig. 3.2. Apparatus Used to Decapsule *L. stagnalis* Embryos. **(A)** An overview of the apparatus that can efficiently remove *L. stagnalis* embryos and larvae from their capsules. **(B)** A magnified view of the yellow boxed area in **(A)**. A sharp glass needle is placed on the microscope slide and inserted into the silicon tubing (inner diameter 1 mm, outer diameter 3 mm) such that the tip protrudes approximately 20% of the way into the cavity of the tubing. The glass needle is then also taped to the microscope slide and the Petri dish. **(C)** A schematic 'plan' view of the yellow boxed section in **A**. Egg capsules containing fixed embryos and larvae are first collected using the 20 mL syringe. The syringe is then attached to the silicon tubing and the capsules expelled through the tubing and past the needle. Egg capsule membranes are torn by the needle and the liberated embryonic and larval material can be collected from the collection dish using a micropipette. **(D)** A schematic 'cross section' view of the yellow boxed section in **A**. The microscope slide ensures that the needle enters the silicon tubing at the correct height. **(E)** A representative view of larvae prior to being processed by the decapsuling apparatus. **(F)** A representative view of larvae that have made a single pass through the apparatus. More than 90% of the material has been effectively and gently removed from their capsules.

NOTE: Normally, the majority (>90%) of all capsules should need only one pass through the device. In many cases the capsule membrane is damaged but embryos and larvae remain inside the ruptured capsule. Reprocess this material by simply drawing it up into the syringe and dispelling it again.

10. Collect the decapsulated embryos and larvae into a 1.5 mL tube using a micropipette (a P20 for 0 - 3 day old larvae, and a P200 for older larvae with the end of the tip cut off).

NOTE: A pause (for up to several months) in the protocol can be made at this stage. If this is required, the decapsuled material should be collected into a

1.5 mL tube for storage (continue to the next step). Otherwise continue immediately to Protocol 3.4.3.

11. Allow embryos and larvae to sink to the bottom of the tube and aspirate the supernatant. Replace with 33% Ethanol (EtOH) in PBTw and let sit for 5 – 10 min. Repeat with 66% EtOH in PBTw and 100% EtOH. Wash larvae twice in 100% EtOH at RT. Store the material at -20 °C in 100% EtOH.

12. When ready to continue, re-hydrate the samples by removing the 100% EtOH and replacing with 66% EtOH in PBTw, let sit for 5 - 10 min. Repeat with 33% EtOH in PBTw, let sit for 5 - 10 min. Finally wash with 3x 5 min washes of PBTw to ensure all EtOH is removed.

3.4.3. Proteinase-K, TEA and Post-fixation

NOTE: We find performing the following steps in small baskets with a mesh floor the most efficient and gentle method for quickly exchanging time critical solutions. While these can be home made, we use baskets (medium size) that are compatible with the Intavis *InSituPro*-Vsi liquid handling robot (www.intavis.de/products/automated-ish-and-ihc). Such baskets can be quickly and easily moved between the wells of a 12 well tissue culture dish (TCD) in order to exchange solutions, or the solution can be aspirated from the well using a pipette. Alternatively, all solution exchanges can be performed without these baskets by simply aspirating the supernatant from the larvae. In this case a gentle swirling motion will concentrate all embryos and larvae to the center of the dish allowing the supernatant to be removed from the edge of the well. The following assumes the user is employing baskets for solution exchanges.

1. Using a pipette, transfer embryos and larvae into a basket sitting in a 12 well TCD with 2 mL of PBTw.
   NOTE: The number of embryos and/or larvae that can be added depends on the developmental stage being investigated, however a general rule of thumb is to maintain at least 25% of the floor space free of embryos/larvae (*i.e.*, do not overcrowd the basket).

2. Prepare an adjacent well with 2 mL of the appropriate Proteinase-K solution (see Fig. 3.1), and another 2 wells with 2 mL of 0.2% Glycine each. Move each

basket into the well containing the appropriate concentration of Proteinase-K and immediately start timing.

3. After 10 min move each basket into a well containing 0.2% Glycine and incubate for 5 min. Then move each basket into the second well with 0.2% Glycine and incubate for 5 min. Wash out the Glycine with 3x 5 min exchanges of 3 mL PBTw.

4. Remove the PBTw solution and treat the samples once with freshly prepared Triethanolamine (TEA) solution for 5 min. CAUTION! Do not agitate. Replace with 3 mL of freshly prepared TEA solution for 5 min. Do not agitate. Aspirate the majority of the TEA solution from the samples. Add the freshly prepared Triethanolamine with Acetic Anhydride (TEAAA) solution and incubate for 5 min. CAUTION! Do not agitate.

5. OPTIONAL - While the above step is incubating, prepare another fresh batch of TEAAA solution and repeat the above step.
   NOTE: This second treatment with TEAAA is optional but may help to completely eliminate all background with probes prone to generating background.

6. Remove the TEAAA solution by aspirating it from the well and replace with 3 mL of PBTw. Do not agitate. Remove the PBTw and apply 3 mL of 3.7% formaldehyde in PBTw. Gently swirl occasionally during a 30 min incubation.

7. Remove the fixative by aspirating it from the well and replace with 3 mL of PBTw. Transfer the material into a 1.5 mL tube. Replace the PBTw with hybridization buffer and incubate for 5 min at RT - CAUTION.

8. Place the tubes into a hot block at RT, and set the temperature to the desired hybridization temperature.
   NOTE: The hybridization temperature is probe specific and will need to be empirically optimized, however we find 55 °C to be a good temperature to initially trial. Allow the block to come to the hybridization temperature, and allow the samples to pre-hybridize for approximately 15 min (*i.e.*, the time it takes to prepare the riboprobes, longer is not necessary). Prepare adequate volumes (normally 100 µL) of the diluted riboprobes. We typically trial probe concentrations of 100 and 500 ng/mL as an initial range. We prepare riboprobes according to [12].

9. Remove the hybridization buffer from the samples and add the probe in hybridization buffer to the samples. Overlay with 100 μL (or an adequate volume to form a phase above the hybridization buffer) of mineral oil.

    NOTE: The mineral oil prevents extensive condensation that will form during the lengthy hybridization step. Extensive condensation will significantly alter both the chemistry of the hybridization solution and the concentration of the probe.

10. Denature the probe and the target RNA by heating the samples to 75 °C for 10 min, then reduce the heat to the desired hybridization temperature. Allow hybridization to proceed for a minimum of 12 hr (O/N) or longer (24 - 48 hr).

11. During hybridization remove a single tube from the heat block, rotate it rapidly between thumb and forefinger to suspend the larvae without disturbing the oil phase, and replace it in the heat block. Repeat this once every 6 - 12 hr or so.

3.4.4. Hot Washes and Immunodetection

NOTE: While we use a liquid handling robot for the following steps, these can also easily be done manually. In this case, embryos and larvae should be kept in the 1.5 mL tubes they were hybridized in. All subsequent solution exchanges are aspirated and added with a P1,000 pipette. When performed manually each of the following steps should employ 1 mL of each solution.

1. Heat adequate volumes (3 mL each for each sample) of the 4x, 2x and 1x wash solutions to the hybridization temperature.

2. Wash all samples three times in 4x wash buffer for 15 min each at the hybridization temperature. Wash all samples three times in 2x wash buffer for 15 min each at the hybridization temperature. Wash all samples three times in 1x wash buffer for 15 min each at the hybridization temperature.

3. Wash all samples once with 1x Sodium Chloride Sodium Citrate buffer + 0.1% Tween (SSC + 0.1% Tween) at the hybridization temperature. Allow samples to cool to RT.

4. Wash all samples twice in 1x SSC + 0.1% Tween for 15 min. Replace this 1x SSC solution with Maleic Acid Buffer (MAB) and let sit for 10 min. Repeat the MAB wash. Replace MAB with block solution and incubate for 1.5 hr.

5. Exchange the block solution for antibody solution (1:10,000 dilution of antibody in block solution) and incubate for 12 hr (O/N) at RT with gentle agitation.

3.4.5. Color Development and Mounting

1. Aspirate the antibody solution and wash 15 times with PBTw for 10 min each. Replace PBTw with 1x Alkaline Phosphatase Buffer (AP) and incubate for 10 min.

2. While the above incubation step proceeds, prepare the AP staining solution (see Supplementary File S. 3.2) - CAUTION. Transfer the material into a TCD well and replace the 1x AP solution with AP staining solution. Note the time now so that the length of time the color reaction takes place for can be recorded. Monitor the development of the staining pattern until the signal to background ratio is optimal.

3. To stop color development, remove the 1x AP staining solution (dispose in the appropriate waste container), and apply 2 mL of PBTw and note the time now. Rinse twice more with PBTw for 5 min each. Use one of these rinses to transfer the material into a 1.5 mL tube.

4. Remove the PBTw and apply 1 mL of 3.7% formaldehyde in PBTw and agitate or rotate for least 30 min at RT (this can also be done O/N).

5. Wash out the fixative with 3 washes of PBT (dispose the waste fixative in an appropriate waste container). Wash the samples 3 times at 50 °C in de-ionized water.

   NOTE: These washes eliminate the precipitation of salts that may become visible during clearing and visualization steps.

6. Decide whether the samples should be mounted in Benzyl Benzoate:Benzyl Alcohol (BB:BA, also known as Murray's clear) or glycerol.

   NOTE: We prefer the more powerful clearing agent BB:BA as *L. stagnalis* embryos are somewhat opaque. Samples to be mounted in BB:BA will first need to be dehydrated through an ethanol series. Samples to be mounted in 60% glycerol can be immediately mounted.

7. Transfer the samples into the mounting solution (BB:BA or 60% glycerol) using a pipette.

NOTE: When mounting in BB:BA this must be done in a glass well (BB:BA will melt polycarbonate plastic).

8. Allow the samples to clear for 5 - 10 min, and then mount them onto a slide using an appropriate number of stacked coverslips as spacers (1 coverslip for samples <1 dpfc, 2 coverslips for samples >1 dpfc).

    NOTE: Whole mounts can now be imaged using a compound microscope with DIC optics.

3.5 Representative Results

The representative WMISH staining patterns shown in Fig. 3.3 were generated using the technique described above, and reflect a variety of spatial expression patterns for genes involved in a range of molecular processes ranging from shell formation *(Novel gene 1*, *2*, *3* and *4*), to cell- cell signaling (*Dpp*) to transcription regulation (*Brachyury*) across a range of developmental stages. While we have not quantified the expression levels of these genes we expect that they would also vary significantly, indicating that our method can be applied against a broad variety of gene products expressed in all stages of development at various levels. Only one of the genes presented here (*Dpp*) has been previously described in *L. stagnalis* [21, 22]. The results we present here are largely in keeping with these previous reports, but with significantly higher spatial resolution. The spatial expression pattern of *Brachyury* has been described in abalone [23] and limpet [24] and in both cases was also detected in mantle cells as we find for *L. stagnalis* (Fig. 3.3F). We isolated *Novel genes 1 - 4* from a proteomic screen designed to identify gene products directly involved in shell formation, and so their spatial expression patterns associated with the shell gland (Fig. 3.3A and B) or shell field (Fig. 3.3C and D) are completely congruent with shell-forming functions. These results indicate that the high throughput technique we have developed for removing embryos and larvae from the egg capsule, and the subsequent stage-specific permeabilization treatments, generate whole mount samples that will yield high quality *in situ* staining patterns for a wide variety of genes for all stages of embryonic and larval development.
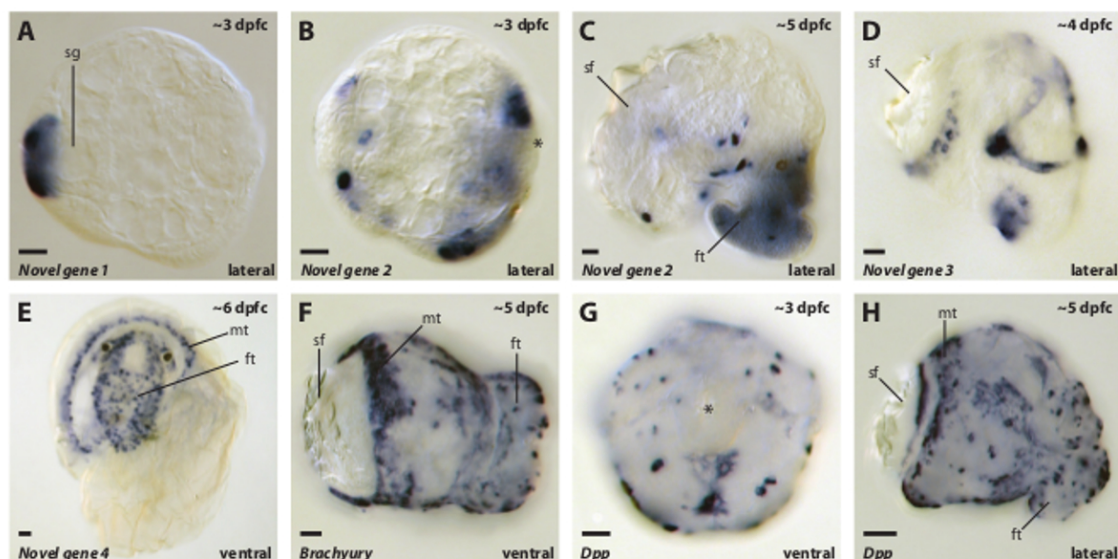
Fig. 3.3. Representative Images of WMISH Expression Patterns Against a Variety of Genes From a Range of *L. stagnalis* Developmental Stages Generated by the Method Described Here. All developmental stages were processed as described in the above method and have been mounted and imaged in BB:BA (Murray's clear). Approximate ages are indicated in the top right of each panel and the orientation is indicated in the lower right. Gene orthology (when known) is indicated in the lower left of each panel. Abbreviations: shell gland (sg); shell field (sf); mantle (mt); foot (ft); Decapentaplegic (*Dpp*); dpfc (days post first cleavage). All scale bars are 20 μm.

3.6 Discussion

The method described here allows for the efficient visualization of RNA transcripts with presumably varying expression levels within all developmental stages of *Lymnaea stagnalis*. To remove embryos and larvae from their capsules we trialed a variety of chemical, osmotic shock and physical treatments reported for other encapsulated-developing model organisms. However, in our hands the method we describe here is the only high-throughput technique that removes the tough capsular membrane without damaging the embryos and larvae. Following decapsulation, the material can either be stored, or treated with a stage specific regimen of Proteinase-K and then hybridized to a riboprobe. Additional empirical optimization efforts (typically focused on probe concentration and hybridization temperature) may be required for each probe/ target. These parameters (in addition to the fixation regimen and Proteinase-K treatments) are typically the most influential parameters of any *in situ* experiment (assuming that the quality of the fixed material and the RNA probe are of a high standard).

The importance of an appropriate Proteinase-K treatment to the final result of an *in situ* experiment is paramount for *L. stagnalis*. This is reflected in the wide range

of Proteinase-K concentrations required by distinct developmental stages (ranging from 0 μg/mL to 500 μg/mL). It is therefore important to be able to assign a given egg string to an ontogenetic window. To this end, the guideline that we provide in Fig. 3.1 allows for the staging of developmental material of unknown ages. We note that for other species of gastropods Proteinase-K treatments for WMISH can either be kept constant for a wide range of developmental stages [8, 25, 26] or can be omitted entirely [27]. This is in stark contrast to the situation in *L. stagnalis*. Furthermore, while other research groups have previously reported WMISH expression patterns for several genes in *L. stagnalis* larvae (see [21, 22, 28]) the method that we describe here yields patterns of significantly higher spatial resolution. Finally, we have observed significant inter-batch variation in the activity of the Proteinase-K from our supplier. This variation must be accounted for by performing a round of 'calibrating' WMISH experiments where the activity of the new Proteinase-K is empirically determined. All subsequent experiments with aliquots of Proteinase-K from that batch can then be freely performed.

We previously described an alternative WMISH method for *L. stagnalis* embryos and larvae elsewhere [12]. That method detailed the use of the mucolytic agent N-acetyl-L-cysteine (NAC), a reducing agent such as Dithiothreitol (DTT) and a pre-hybridization treatment with sodium dodecyl sulfate (SDS). We found those treatments enhanced the staining patterns of some genes for some developmental stages. The fixation strategy that we recently developed and describe here (fixing larvae within their capsules) simplifies and expedites the steps required to prepare material for an *in situ* experiment, and apparently negates the need for empirically determining additional optimal pre-hybridization treatments with NAC, DTT or SDS. Future refinements to the technique reported here could include the visualization of microRNAs (following modifications to standard WMISH protocols previously reported [29]), double or triple labeling of mRNA targets [30], and fluorescent visualization of WMISH signals [31]. Arguably the greatest limitation of the technique is the overall length of time it takes to go from collecting the material, to a digital image that represents a given gene expression pattern. Due to the nature of the biochemical and biophysical events that must take place during such a process this is an inherent feature of most *in situ* hybridization protocols.

Lymnaea occupies a position within the Metazoa that is extremely under-represented in terms of model organisms. As a representative Spiralian, *Lymnaea* can bring insight into the evolution of distinct morphological features such as shell formation [12] and body handedness [32-34] and is also a valuable neuroethology [35] and neurophysiology model [9, 36]. Powerful techniques such as the ability to efficiently visualize gene expression patterns *in situ* increases the functionality of *Lymnaea* as a model organism, and broadens the variety of questions that it can be used to address. At a time when the generation of large sequence datasets (complete transcriptomes and even genomes) is relatively routine, such methods will become more relevant to researchers wishing to interpret the flood of sequence data from such models. While *Lymnaea* is a relatively derived gastropod [1], and possesses what would be considered a large genome in comparison to other model organisms (1.22 Gb [37]), it has many practical and interesting features that make it an attractive model system. The methods that we describe here expand the toolbox available to *Lymnaea* and may be of use to other species that undergo encapsulated development.

Supplementary files

The supplementary files for this article can be found online at: https://www.jove.com/video/53968/a-whole-mount-situ-hybridization-method-for-gastropod-mollusc-lymnaea or via the attached CD of the printed version.

Supplementary file S. 3.1 Material, safety comments and catalogue numbers
Supplementary file S. 3.2 Recipes for all reagents

Disclosures

The authors have nothing to disclose.

Acknowledgements

Published online: 15 March 2016

References

1. Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SCS, Rouse GW, Giribet G, Dunn CW. Resolving the evolutionary relationships of molluscs with phylogenomic tools. Nature. 2011;480:364-367.
2. Brusca RC, Brusca GJ: Invertebrates. 2nd edition. Sinauer Associates Sunderland, Massachusetts; 2002.
3. Food and Agriculture Organisation of the United Nations [http://www.fao.org/fishery/statistics/global-aquaculture-production/query/en (2013)]
4. Schistosomiasis: number of people treated in 2011. Weekly Epidemiological Record. 2013;88:81-88.
5. Henry JJ, Collin R, Perry KJ. The slipper snail, *Crepidula*: an emerging lophotrochozoan model system. The Biological Bulletin. 2010;218:211-229.
6. Perry KJ, Henry JQ. CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. Genesis. 2015;53:237-244.
7. Kandel ER. The molecular biology of memory storage: a dialog between genes and synapses. Bioscience Reports. 2004;24:475-522.
8. Jackson DJ, Ellemor N, Degnan BM. Correlating gene expression with larval competence, and the effect of age and parentage on metamorphosis in the tropical abalone *Haliotis asinina*. Marine Biology. 2005;147:681-697.
9. Carter CJ, Farrar N, Carlone RL, Spencer GE. Developmental expression of a molluscan RXR and evidence for its novel, nongenomic role in growth cone guidance. Developmental Biology. 2010;343:124-137.
10. Rittschof D, McClellan-Green P. Molluscs as multidisciplinary models in environment toxicology. Marine Pollution Bulletin. 2005;50:369-373.
11. Liu MM, Davey JW, Jackson DJ, Blaxter ML, Davison A. A conserved set of maternal genes? Insights from a molluscan transcriptome. The International Journal of Developmental Biology. 2014;58:501-511.
12. Hohagen J, Herlitze I, Jackson DJ. An optimised whole mount *in situ* hybridisation protocol for the mollusc *Lymnaea stagnalis*. BMC Developmental Biology. 2015;15:19.
13. Raven CP. The development of the egg of *Limnaea stagnalis* L. from oviposition till first cleavage. Archives Néerlandaises de Zoologie. 1946;7:91-121.
14. Raven CP. The Development of the Egg of *Limnaea Stagnalis* L. From the First Cleavage Till the Troghophore Stage, With Special Reference To Its "Chemical Embryology". Archives Néerlandaises de Zoologie. 1946;7:353-434.
15. Raven CP. Morphogenesis in *Limnaea stagnalis* and its disturbance by lithium. Journal of Experimental Zoology. 1952;121:1-77.
16. Raven CP. The nature and origin of the cortical morphogenetic field in *Limnaea*. Developmental Biology. 1963;7:130-143.
17. Morrill JB, Blair CA, Larsen WJ. Regulative development in the pulmonate gastropod, *Lymnaea palustris*, as determined by blastomere deletion experiments. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology. 1973;183:47-55.
18. Van den Biggelaar JAM. Timing of the phases of the cell cycle during the period of asynchronous division up to the 49-cell stage in *Lymnaea*. Development. 1971;26:367-391.

19. Verdonk NH. Gene expression in early development of *Lymnaea stagnalis*. Developmental Biology. 1973;35:29-35.
20. Gall JG, Pardue ML. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proceedings of the National Academy of Sciences. 1969;63:378-383.
21. Iijima M, Takeuchi T, Sarashina I, Endo K. Expression patterns of *engrailed* and *dpp* in the gastropod *Lymnaea stagnalis*. Development Genes and Evolution. 2008;218:237-251.
22. Shimizu K, Sarashina I, Kagi H, Endo K. Possible functions of Dpp in gastropod shell formation and shell coiling. Development Genes and Evolution. 2011;221:59-68.
23. Koop D, Richards GS, Wanninger A, Gunter HM, Degnan BM. The role of MAPK signaling in patterning and establishing axial symmetry in the gastropod *Haliotis asinina*. Developmental Biology. 2007;311:200-212.
24. Lartillot N, Lespinet O, Vervoort M, Adoutte A. Expression pattern of *Brachyury* in the mollusc *Patella vulgata* suggests a conserved role in the establishment of the AP axis in Bilateria. Development. 2002;129:1411-1421.
25. Jackson DJ, Wörheide G, Degnan BM. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. BMC Evolutionary Biology. 2007;7:160.
26. Jackson DJ, Meyer NP, Seaver E, Pang K, McDougall C, Moy VN, Gordon K, Degnan BM, Martindale MQ, Burke RD. Developmental expression of *COE* across the Metazoa supports a conserved role in neuronal cell-type specification and mesodermal development. Development Genes and Evolution. 2010;220:221-234.
27. Perry KJ, Lyons DC, Truchado‚ÄêGarcia M, Fischer AHL, Helfrich LW, Johansson KB, Diamond JC, Grande C, Henry JQ. Deployment of regulatory genes during gastrulation and germ layer specification in a model spiralian mollusc *Crepidula*. Developmental Dynamics. 2015;244:1215-1248.
28. Shimizu K, Iijima M, Setiamarga DHE, Sarashina I, Kudoh T, Asami T, Gittenberger E, Endo K. Left-right asymmetric expression of *dpp* in the mantle of gastropods correlates with asymmetric shell coiling. EvoDevo. 2013;4:15.
29. Christodoulou F, Raible F, Tomer R, Simakov O, Trachana K, Klaus S, Snyman H, Hannon GJ, Bork P, Arendt D. Ancient animal microRNAs and the evolution of tissue identity. Nature. 2010;463:1084-1088.
30. Koga M, Kudoh T, Hamada Y, Watanabe M, Kageura H. A new triple staining method for double *in situ* hybridization in combination with cell lineage tracing in whole-mount *Xenopus* embryos. Development, Growth & Differentiation. 2007;49:635-645.
31. Lauter G, Söll I, Hauptmann G. Two-color fluorescent *in situ* hybridization in the embryonic zebrafish brain using differential detection systems. BMC Developmental Biology. 2011;11:43.
32. Davison A, Frend HT, Moray C, Wheatley H, Searle LJ, Eichhorn MP. Mating behaviour in *Lymnaea stagnalis* pond snails is a maternally inherited, lateralized trait. Biology Letters. 2009;5:20-22.
33. Kuroda R, Endo B, Abe M, Shimizu M. Chiral blastomere arrangement dictates zygotic left–right asymmetry pathway in snails. Nature. 2009;462:790-794.
34. Shibazaki Y, Shimizu M, Kuroda R. Body handedness is directed by genetically determined cytoskeletal dynamics in the early embryo. Current Biology.

2004;14:1462-1467.

35. Lu TZ, Feng Z-P. A sodium leak current regulates pacemaker activity of adult central pattern generator neurons in *Lymnaea stagnalis*. PLoS One. 2011;6:e18745.

36. Dawson TF, Boone AN, Senatore A, Piticaru J, Thiyagalingam S, Jackson D, Davison A, Spafford JD. Gene splicing of an invertebrate beta subunit (LCavβ) in the N-terminal and HOOK domains and its regulation of LCav1 and LCav2 calcium channels. PloS one. 2014;9:e92941.

37. Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD. Eukaryotic genome size databases. Nucleic Acids Research. 2007;35:D332-D338.

4

An *in situ* hybridisation protocol for *Lymnaea stagnalis* mantle tissue sections

4.1 Introduction

The wide spread gastropod *L. stagnalis* was once a popular model to investigate shell formation [1-3] as well as developmental processes in molluscs [4-8]. Recently, *L. stagnalis* regained much appreciation as a model organism. Studies concerning the establishment of chirality [9-11], the reproduction in simultaneous hermaphrodites [12], the cellular mechanisms of learning and memory [13] and studies investigating evolutionary developmental processes [14, 15] utilized the freshwater pulmonate as model.

In situ hybridisation is a common technique in molecular biology and represents an essential tool for studying molecular processes within an organism. In a nutshell, in an *in situ* hybridisation experiment a single stranded complementary nucleic acid probe is hybridised to a target of interest, and subsequently detected immunologically using a colourimetric method [15]. This powerful technique allows the spatial and temporal resolution of gene expression within a tissue of interest and has recently been optimised for *L. stagnalis* larvae by Hohagen et al. 2015 [15] (see chapter 2) and further developed by Jackson et al. 2016 [16] (see chapter 3). The goal of this study is to determine a robust *in situ* hybridisation protocol for *L. stagnalis* adult mantle tissue. As shown in Jackson et al. 2016, the success of an *in situ* hybridisation experiment carried out on *L. stagnalis* larval tissue highly depends on finding a balance between tissue fixation and subsequent permeabilisation that allows sufficient probe penetration [16]. By assuming similar premises for adult mantle paraffin sections, the influence of different formaldehyde fixation regimes and proteinase K concentrations on an *in situ* hybridisation experiment were systematically compared. The optimisation procedure was performed with the biomineralisation genes *Lstag-sfc-10* and *Lstag-sfc-23*. The gene products of *Lstag-sfc-10* and *Lstag-sfc-23* were isolated from the organic components of the adult shell and thus were expected to be active within the adult shell-forming mantle tissue. Furthermore, these genes reared strong *in*

situ signals within the shell-forming tissue of *L. stagnalis* trochophore and veliger larva (see chapter 5). The protocol developed in this study allows tracking the expression of candidate genes within the adult *L. stagnalis* mantle tissue. It represents the continuation of the optimisation experiments on *L. stagnalis* larval tissue and will be an indispensible tool when studying the molecular mechanisms of the shell-forming process.
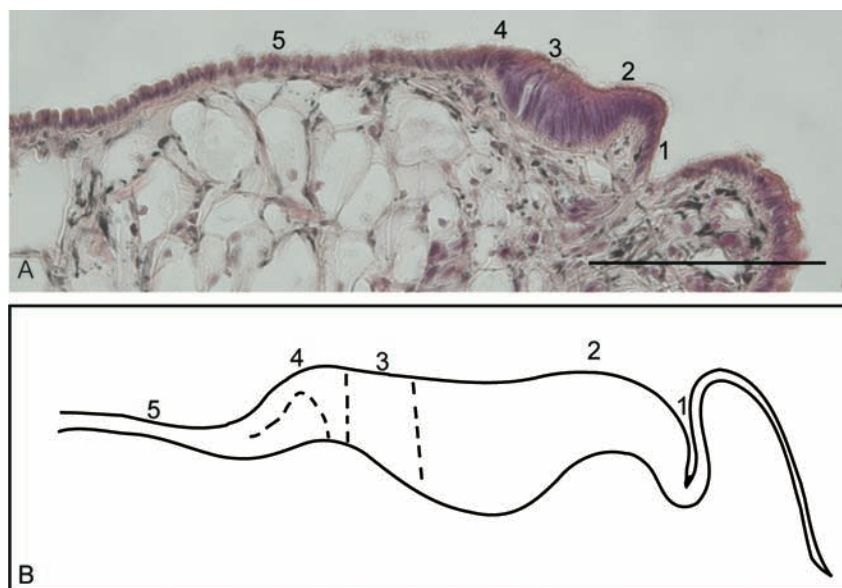
Fig. 4.1. The adult mantle tissue of *L. stagnalis.* (A) A paraffin section taken from an adult *L. stagnalis* stained with hematoxylin and eosin. **(B)** A schematic representation of the *L. stagnalis* adult mantle, redrawn from Timmermans (1969) [1]. **1** periostracal groove, **2** anterior 2/3 of the belt, **3** posterior 1/3 of the belt, **4** transitional zone, **5** low columnar outer epithelium. Scale bar represents 100 µm.

4.2 Methods

4.2.1 Cultivation of adult *L. stagnalis*

Adult specimens of *L. stagnalis* derived from animals originally collected from the Northeimer Seenplatte, Germany (51° 43' 26.5368', 9° 57' 24.75') and from a pond on the North campus of the University of Göttingen, Germany (51° 33' 23.727', 9° 57' 25.617') were kept in a "Stand-alone V30 unit" (Aqua Schwarz) in demineralized water supplemented with ReMineral+ (Dennerle, #7036) to a conductivity of 200 - 220 µS and heated to 23 °C at a pH of 7.5 to 7.9. Five to ten individuals were kept in three or five litre boxes under a constant and low flow rate. Snails were fed *ad libidum* with lettuce and a variety of other vegetables. Under this regime adult snails lay egg masses year round.

4.2.2 Fixation

Samples were anaesthetised with 2% $MgCl_2$ for 2 hours and transferred to freshly prepared 3.7% Formaldehyde (FA)(Carl Roth, 7398) in 1X phosphate buffered saline (PBS). To assess the optimal fixation time, four sample sets of snails were incubated for 10 minutes, 30 minutes, one hour and two hours. Following fixation the samples were washed three times with PBS. To allow easy embedding, the snails were cut in half across the midline. Subsequently, samples were dehydrated through a graduated ethanol (EtOH) series in PBS with one wash in 15% (v/v) EtOH, one wash in 33% (v/v) EtOH, one wash in 50% (v/v) EtOH, one wash in 66% (v/v) EtOH, one wash in 75% (v/v) EtOH and two washes in 100% EtOH.

4.2.3 Paraffin embedding and tissue sectioning

Samples were incubated in xylene for one hour twice. After incubating in paraffin at 60 °C for 24 hours, bisected samples were oriented to allow sectioning in the proper anatomical plane. Samples were serially sectioned into 10 μM thick slices and placed on polysine slides (Menzel J2800AMNZ). To facilitate adhesion, slides were stored in a heated chamber at 60 °C over night. Slides were dewaxed with three washes of xylene for five minutes each and rehydrated through a graduated ethanol series in water with two washes in 100% EtOH, one wash in 90% (v/v) EtOH, one wash in 80% (v/v) EtOH, one wash in 70% (v/v) EtOH and one wash in deionised water. Slides were assembled with counter-slides in a phosphate buffered saline + Tween 20 (PBS-T) bath and placed into the Intavis *In situ*-Pro robot.

4.2.4 Proteinase K treatment

To assess the optimal treatment with proteinase K (Carl Roth, #7528), differently fixed samples were treated with different concentration of proteinase K in PBS-T (500 μg/mL, 100 μg/mL, 50 μg/mL, 25 μg/mL, 10 μg/mL and 0 μg/mL) for ten minutes at room temperature. To stop proteinase K activity, samples were washed twice with glycine for five minutes and washed thrice with PBS-T. Then, samples were postfixed with 3.7% FA/PBS-T for 20 minutes and washed with PBS-T thrice.

4.2.5 Riboprobe synthesis

PCR products were generated using mantle 3' cDNA, a universal primer (CAGCTAC TAGGTGCATGTCGTA, CAGCTACTAGGTGCATGTCGTAGAAGGGAATCGAGTGAGGTTG AG) and gene specific primer for *Lstag-sfc-23* (GGCTTTTCGTGGAGCTCACATTTTCG CCTA) and *Lstag-sfc-10* (CCAAGATCCCATCGTCATCAGCCAAGG) and cloned into plasmids containing T6 and SP6 promoter sites. Fragments were amplified with M13 primers and purified using QIAGEN QIAquick Gel Extraction Kit. Antisense riboprobes were synthesised using Promega reagents containing 1X reverse transcription buffer, 10 mM dithiothreitol, 1X digoxigenin RNA labelling Mix (Roche, #11277073910), 0.25 - 0.5 volume PCR template and 20 units of the appropriate RNA polymerase (SP6 or T7; Promega, #P108 or #P207). Probe synthesis reactions were carried out at 37 °C for two to four hours. Riboprobes were precipitated with 0.1 volumes of 3 M sodium acetate pH 5.2 and 3 volumes of absolute EtOH for 15 minutes and centrifuged for 15 minutes at 16,000 RCF. Pellets were washed once in 75% EtOH, dried and dissolved in 10 µL deionised water at 55 °C.

4.2.6 Probe hybridisation and antibody binding

The samples were incubated with hybridisation buffer for 15 minutes at room temperature. Then, samples were heated to 55 °C and incubated with hybridisation buffer for two hours followed by 16 hours incubation of 500 ng riboprobe in hybridisation buffer. Samples were washed with 4X wash buffer for 15 minutes thrice, 2X wash buffer for 15 minutes thrice and 1X wash buffer for 15 minutes thrice and once with 1X saline-sodium citrate buffer (SSC) + 0.1% Tween for 15 minutes. All hybridisation and washing steps were carried out at 55 °C. After the hot washes, samples were allowed to cool down to room temperature. All following steps were performed at room temperature. Samples were incubated with 1X SSC + 0.1% Tween for 15 minutes and washed with maleic acid buffer for 10 minutes. Then, samples were incubated with 2% block solution (Roche #11096176001) for 1 hour 45 minutes followed by 10 hours incubation with block solution containing 1/10000 of anti-DIG antibody conjugated to alkaline phosphatase (Roche, #11093274910). Unbound antibody was removed with 15

washes of PBS-T for 10 minutes each. For colour development, samples were taken out of the Intavis *Insitu*Pro VSi.

4.2.7 Colour development

Slides and counter-slides were disassembled in a PBS-T bath and washed twice in 1X alkaline phosphatase buffer. Minimal volume of detection buffer was applied and colour development was performed in the dark. The reaction was stopped with two washes of PBS. Samples were postfixed in 3.7% FA in PBS over night at 4 °C. After fixation, samples were washed with 1X PBS and deionised water followed by dehydration through a graduated EtOH series in water and mounted with Roti- Histokitt II (Carl Roth, T160).

4.2.8 Imaging

Samples were photographed using Zeiss Axio Imager Z1 microscope running Zeiss camera software Axio Rel.4.8.

4.2.9 Solutions

1X PBS (phosphate buffered saline): 0.1 volume of 10X PBS stock (1.37 M NaCl; 27 mM KCl; 100 mM $Na_2HPO_4 \cdot 2\ H_2O$; 20 mM KH_2PO_4).

1X PBS-T (phosphate buffered saline + Tween 20): 0.1 volume of 10X PBS stock; 0.1% (v/v) Tween-20.

3.7% FA (formaldehyde): 0.1 volume of 37% FA in 1X PBS

Pro-K (proteinase-K): Diluted from 10 mg/mL stock using PBS-T.

2 mg/mL glycine pH 2: Diluted from 100 mg/mL stock using PBS-T.

Hybridisation buffer: 0.25 volume 20X SSC stock (3 M NaCl; 0.3 M trisodium citrate dihydrate); 5 mM ethylene diamine tetra-acetic acid (EDTA); 0.5 volume deionised formamide; 100 µg/mL heparin; 0.1% (v/v) Tween-20; 1X Denhardt's (2% (m/v) Ficoll type 400; 2% (w/v) polyvinylpyrrolidone K30; 2% (w/v) bovine serum albumin); 100 µg/mL single-stranded salmon sperm DNA .

4X wash: 0.2 volume 20X SSC stock; 0.5 volume formamide; 0.1% (v/v) Tween-20.

2X wash: 0.1 volume 20X SSC stock; 0.5 volume formamide; 0.1% (v/v) Tween-20.

1X wash: 0.05 volume 20X SSC stock; 0.5 volume formamide; 0.1% (v/v) Tween-20.

MAB (maleic acid buffer): 0.1 M maleic acid; 0.15 M NaCl.

Block solution: 2% (v/v) block in MAB.

Antibody solution: Anti-DIG fab fragments diluted 1:10,000 in block solution.

1X AP (alkaline Phosphatase buffer): 0.2 volume 5X AP buffer stock (0.5 M Tris pH 9.5; 0.5 M NaCl).

Colour detection buffer: 1X AP; 50 mM $MgCl_2$; 450 µg/mL NBT; 175 µg/mL BCIP.
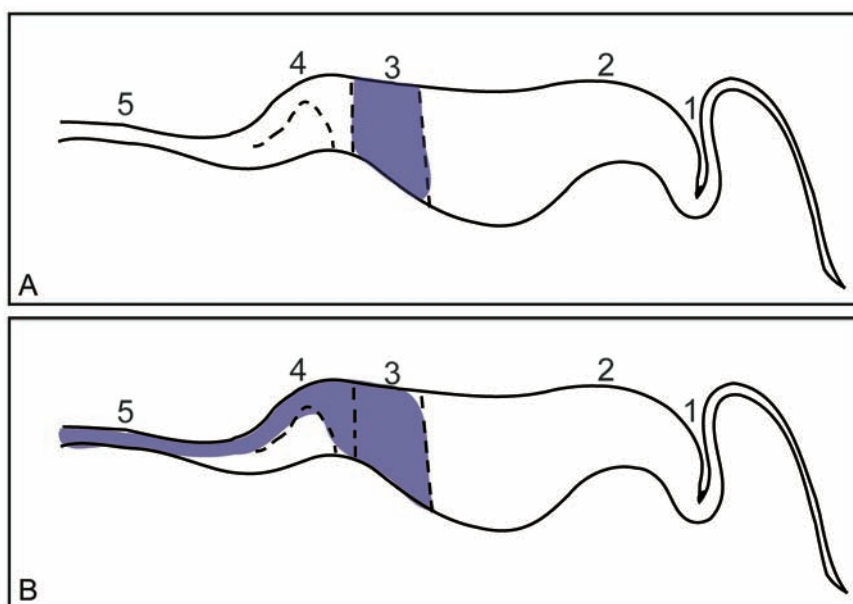
Fig. 4.2. The approximate localisation of the expression *Lstag-sfc-10* (A) and *Lstag-sfc-23* (B) within the *L. stagnalis* mantle indicated in blue.

4.3 Results and Discussion

One major aspect of an *in situ* experiment is the treatment with proteinase K. Determining the appropriate proteinase K parameter is crucial, since under treatment of proteinase K can hinder probe penetration and result in an absent or weak signal, while overtreatment can lead to background staining and severely altered tissue morphology [15, 16]. The adjustable proteinase K parameters include concentration as well as incubation time and temperature. To determine the optimal proteinase treatment for sections of *L. stagnalis* adult mantle tissue the focus was on altering the proteinase K concentration, while incubation time and temperature remained constant.

The *L. stagnalis* mantle edge comprises of three histological distinct zones that are responsible for shell secretion [1]. The mantle groove is located on the outermost

edge, followed by a narrow zone of high columnar cells (belt) that gradually merge into a low columnar epithelium (Fig. 4.1). These regions have previously been subdivided based on their biochemical properties. The periostracal groove forms zone one. The mantle belt is subdivided into zone two which spans the first two thirds and zone three which occupies the rest of the belt. The transitional zone between the belt and the low columnar outer epithelium is coined zone four and the low columnar outer epithelium forms zone five (Fig. 4.1) [1]. The approximate localisation of *Lstag-sfc-10* expression is indicated in Fig. 4.2A and located in zone three. The approximate localisation of *Lstag-sfc-23* is indicated in Fig. 4.2B and located within zone three, four and five.

No preliminary information was available for an *L. stagnalis in situ* procedure on adult mantle sections. Therefore, the rough range for the optimal proteinase K treatment had to be obtained in the first optimisation *in situ* experiment. An extremely broad range of proteinase K concentrations from 10 µg/mL to 50 µg/mL, 100 µg/mL and 500 µg/mL proteinase K was applied to samples fixed for either 10 minutes, 30 minutes, 1 hour or 2 hours in 3.7% FA. Samples treated with very high concentrations of proteinase K (100 µg/mL and 500 µg/mL) showed strong signal intensity but suffered from compromised tissue morphology (Figs. 4.3I - P and 4.4I –P). This was also true for samples that were fixed lightly for 10 minutes and treated with any of the above-mentioned amounts of proteinase K (Figs. 4.3A, E, I, M and 4.4A, E, I, M). In the case of *Lstag-sfc-10*, similar signal intensities while remaining tissue integrity were obtained for the samples that were fixed for 30 minutes, 1 hour or 2 hours and treated with 10 µg/mL or 50 µg/mL proteinase K (Fig. 4.3B, C, D, F, G, H). In the case of *Lstag-sfc-23*, the treatments with the best signal intensity while remaining tissue integrity were obtained when fixed for 30 minutes or 1 hour and treated with 50 µg/mL proteinase K (Fig. 4.4F, G). All samples that derived from one *L. stagnalis* individual fixed for one hour showed a unique staining for both *Lstag-sfc-10* and *Lstag-sfc-23* along the cell surface of the cells of zone one and two (Figs. 4.3C, G, K, O and 4.4G, K, O). This unique staining was never obtained again with any other *L. stagnalis* individual for both genes (personal observation) and could hint at gene expression polymorphism or indicate background staining potentially caused by damaged tissue.

Fig. 4.3. Overview of the *Lstag-sfc-10 in situ* hybridisation signals produced by different proteinase K treatments and fixation incubation times. Proteinase K concentrations from 10 µg/mL to 50 µg/mL, 100 µg/mL and 500 µg/mL were applied to samples fixed for either 10 minutes, 30 minutes, 1 hour or 2 hours in 3.7% FA. Samples treated with very high concentrations of proteinase K (100 µg/mL and 500 µg/mL) as well as samples fixed for 10 minutes suffered from compromised tissue morphology **(A, E, I, J, K, L, M, N, O P)**. Sufficient signal intensities while remaining tissue integrity were obtained when the samples were fixed for 30 minutes, 1 hour or 2 hours and treated with 10 µg/mL or 50 µg/mL proteinase K **(B, C, D, F, G, H)**. The asterisks indicate the best results.

Fig. 4.4. Overview of the *Lstag-sfc-23 in situ* hybridisation signals produced by different proteinase K treatments and fixation incubation times. Proteinase K concentrations from 10 µg/mL to 50 µg/mL, 100 µg/mL and 500 µg/mL were applied to samples fixed for either 10 minutes, 30 minutes, 1 hour or 2 hours in 3.7% FA. Samples treated with very high concentrations of proteinase K (100 µg/mL and 500 µg/mL) as well as samples fixed for 10 minutes suffered from compromised tissue morphology **(A, E, I, J, K, L, M, N, O, P)**. Sufficient signal intensities while remaining tissue integrity were obtained when the samples were fixed for 30 minutes or 1 hour and treated with 50 µg/mL Proteinase K **(F, G)**. The asterisks indicate the best results.

Fig. 4.5. Overview of the *Lstag-sfc-10 in situ* hybridisation signals produced by different proteinase K treatments and fixation incubation times. Proteinase K concentrations between 0μg/mL, 10 μg/mL, 25 μg/mL and 50 μg/mL were applied to material that was fixed for one and two hours. The best results were obtained when fixed for one hour and treated with 25 μg/mL or 50 μg/mL proteinase K **(E, G)** and are indicated by an asterisk.
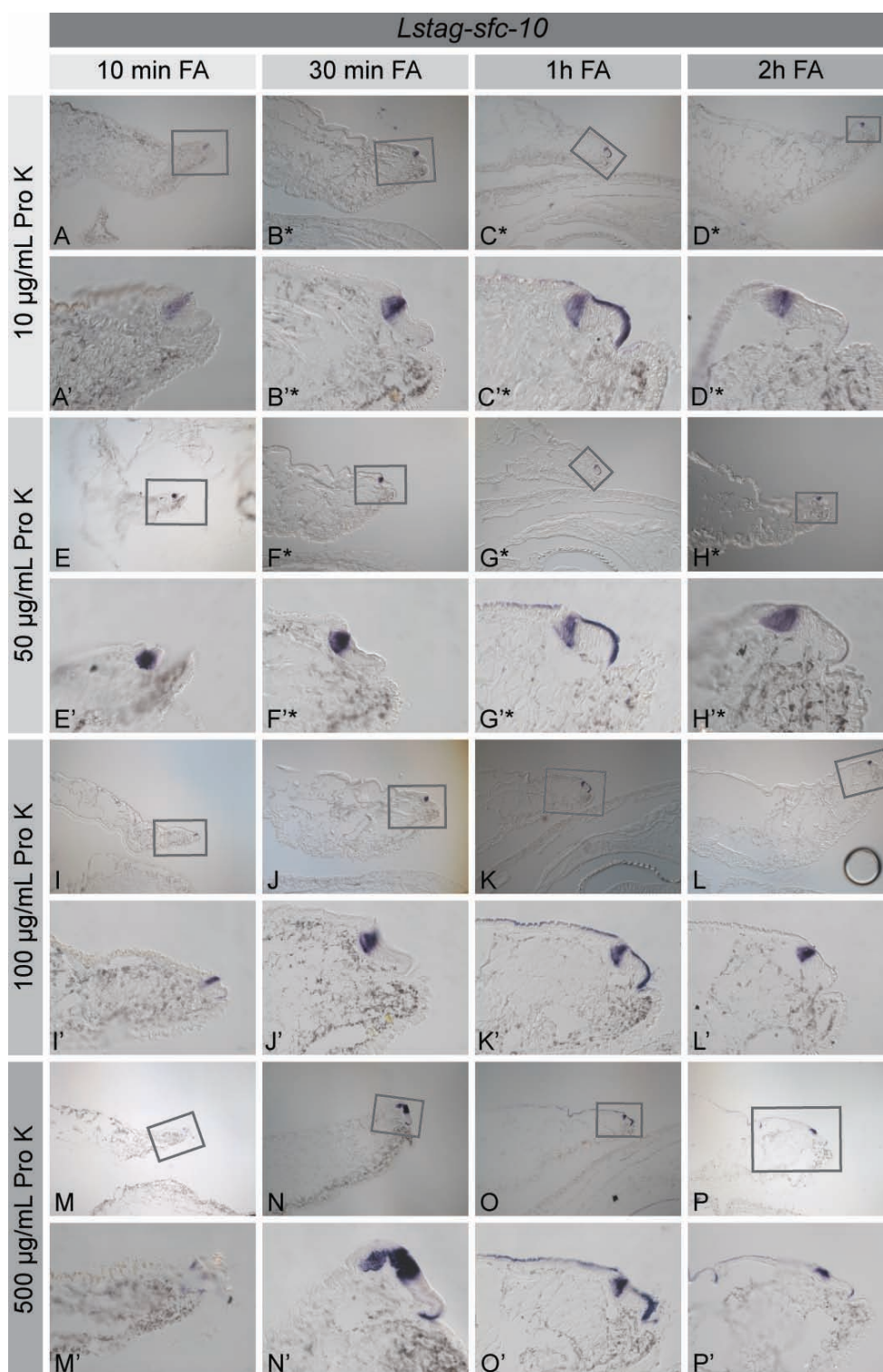
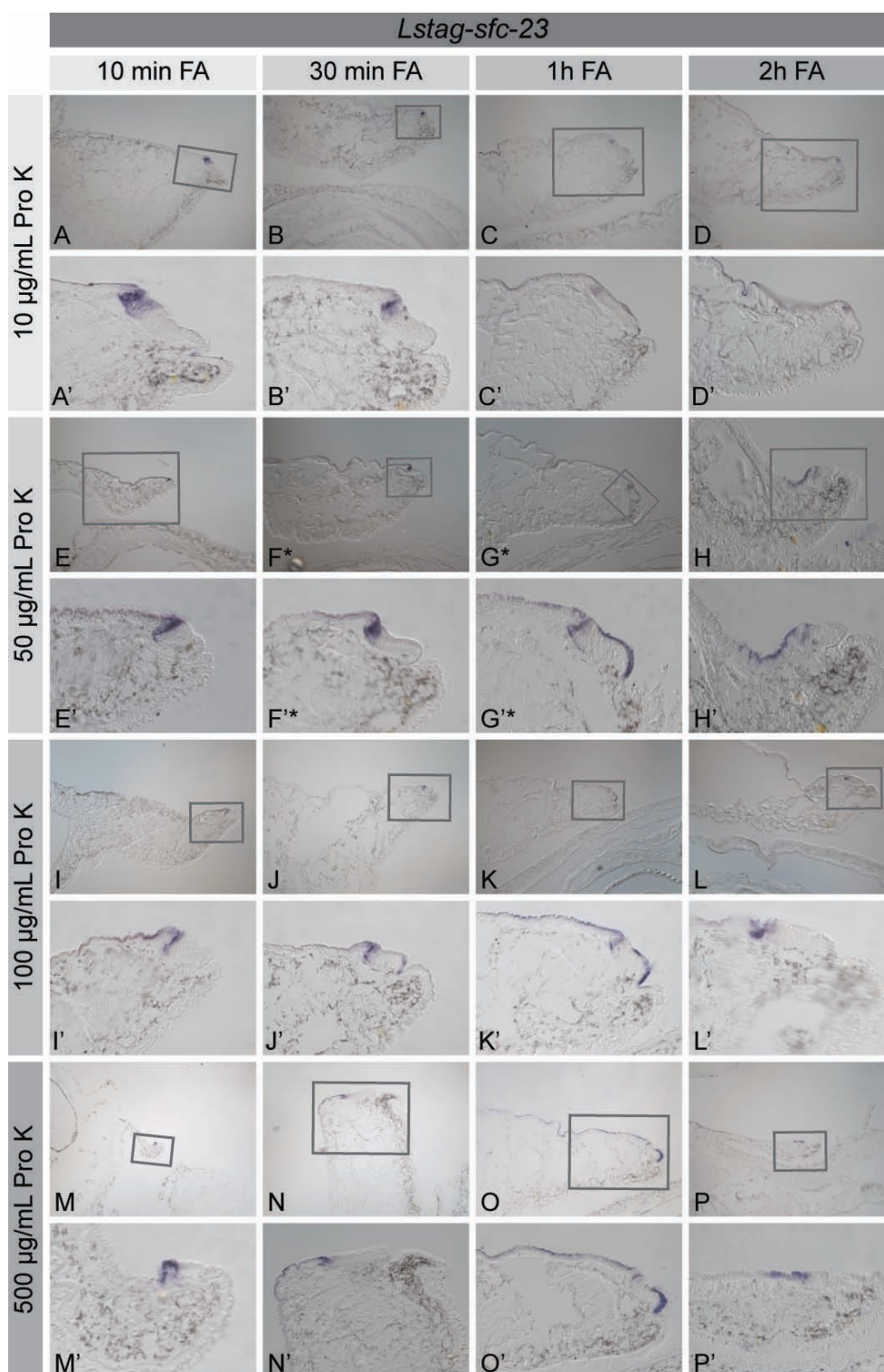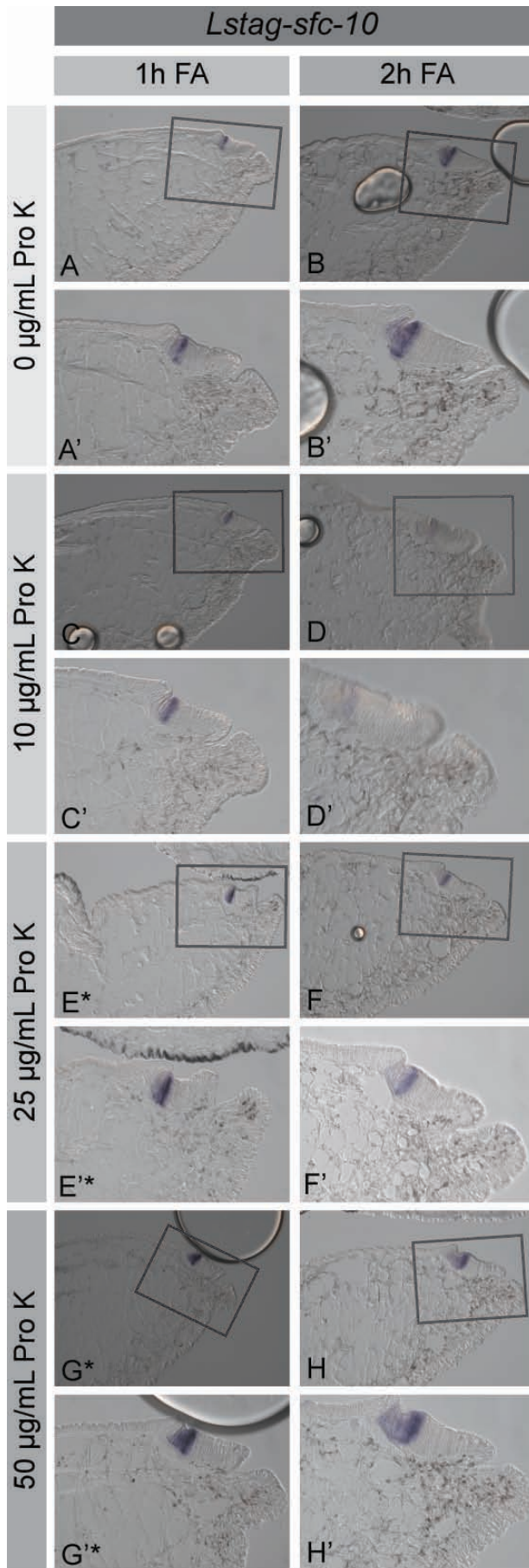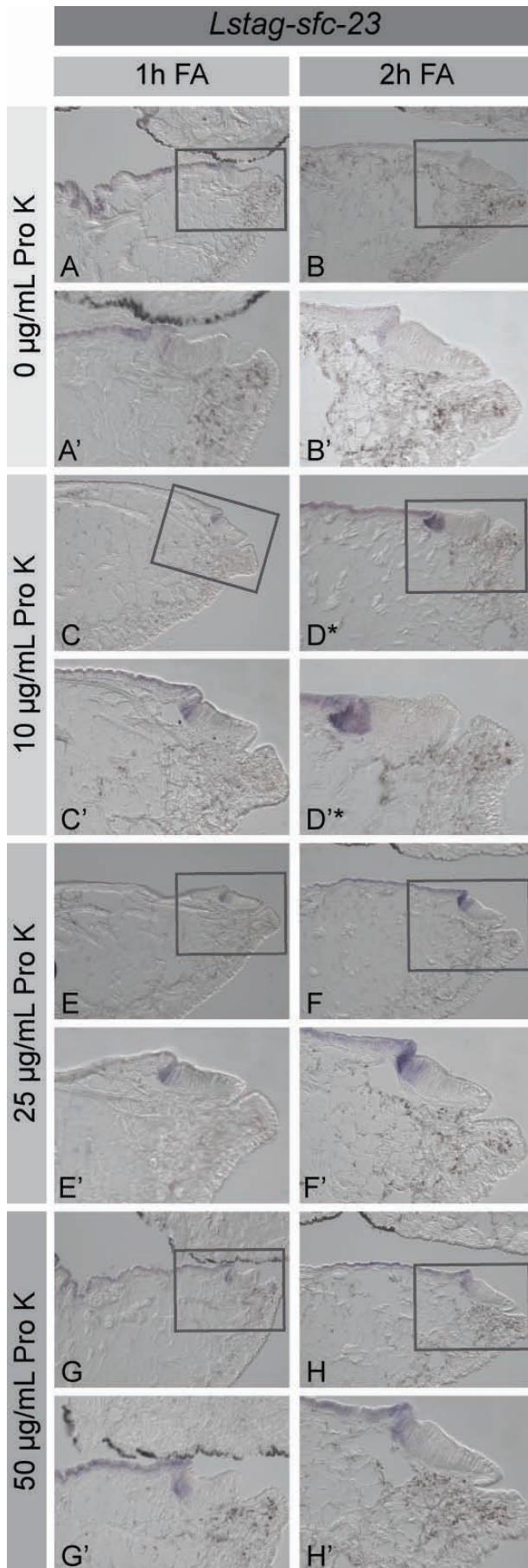**Fig. 4.6. Overview of the _Lstag-sfc-23 in situ_ hybridisation signals produced by different proteinase K treatments and fixation incubation times.** Proteinase K concentrations between 0 µg/mL, 10 µg/mL, 25 µg/mL and 50 µg/mL were applied to material that was fixed for one and two hours. The best results were obtained when fixed for two hours and treated with 10 µg/mL **(D)** and is indicated by an asterisk.
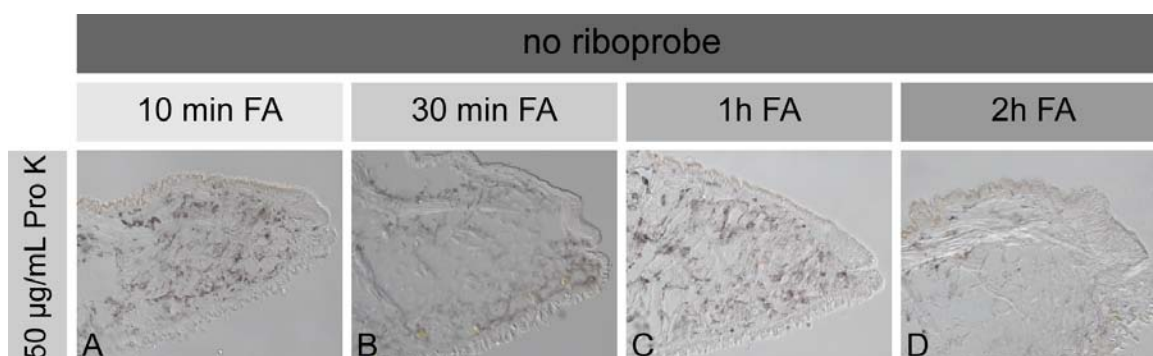
Fig. 4.7. No Probe control experiment. Control experiments lacking riboprobe demonstrate the absence of any non-specific colour reaction when fixed for 10 minutes, 30 minutes, 1 hour or 2 hours and treated with 50 µg/mL proteinase K.

The first experiment showed that the optimal proteinase K concentration should not exceed 50 µg/mL and the samples should be fixed for 30 minutes, 1 hour or 2 hours. Based on these results a follow up *in situ* was performed to further narrow down the optimal fixation and permeabilisation regime. Proteinase K concentrations of 0 µg/mL, 10 µg/mL, 25 µg/mL and 50 µg/mL on material that was fixed for one and two hours was tested. In the case of *Lstag-sfc-10*, all treatments reared at least a faint signal. The best results could be obtained when fixed for one hour and treated with 25 µg/mL or 50 µg/mL proteinase K (Fig. 4.5E, G). In the case of *Lstag-sfc-23*, all samples reared at least a faint signal, whereas the signal intensity in samples treated with no proteinase K and fixed for one or two hours was not satisfying (Fig. 4.6A, B). The best *Lstag-sfc-23* signal was reared when fixed for two hours and treated with 10 µg/mL proteinase K (Fig. 4.6D). The conditions that reared the best results for *Lstag-sfc-23* produce the weakest signal for *Lstag-sfc-10* (Figs. 4.5D and 4.6D). This emphasizes that the best conditions for one gene might not be the optimal settings for another and that adjustments concerning the proteinase K treatment might be necessary if an experiment yields insufficient results. In general, the experiment suggests that if 10 µg/mL to 50 µg/mL proteinase K are applied to samples fixed for one or two hours an *in situ* signal can be yielded without compromising the tissue integrity (Figs. 4.5C - H and 4.6C - H). Samples treated with 50 µg/mL proteinase K and fixed for one hour reared good results for both genes and is the recommended treatment for an *in situ* experiment on adult mantle sections (Figs. 4.5G and 4.6G). Control experiments lacking riboprobe demonstrate the absence of any non-specific colour reaction within the *L. stagnalis* mantle tissue (Fig. 4.7).

4.4 Conclusion

The *in situ* hybridisation protocol for adult mantle sections represents an indispensable resource when studying the molecular properties of the mantle tissue and expands the molecular toolbox available for the aspiring gastropod model system *L. stagnalis*. Especially investigations focusing on the molecular mechanisms of the shell-forming process in adult *L. stagnalis* snails will profit from this technique. With this method it is possible to link the expression of a shell-forming gene to a distinct position within the shell-forming organ. Based on this localisation first assumptions about their putative function can be drawn

References

1. Timmermans LPM. Studies on shell formation in molluscs. Netherlands Journal of Zoology. 1969;19:413-523.
2. Kniprath E. Formation and structure of the periostracum in *Lymnaea stagnalis*. Calcified Tissue Research. 1972;9:260-271.
3. Kniprath E. Das Wachstum des Mantels von *Lymnaea stagnalis* (Gastropoda). Cytobiologie. 1975;10:260-267.
4. Raven CP. The Development of the Egg of *Limnaea Stagnalis* L. From the First Cleavage Till the Troghophore Stage, With Special Reference To Its "Chemical Embryology". Archives Néerlandaises de Zoologie. 1946;7:353-434.
5. Raven CP. Morphogenesis in *Limnaea stagnalis* and its disturbance by lithium. Journal of Experimental Zoology. 1952;121:1-77.
6. Kniprath E. Zur Ontogenese des Schalenfeldes von *Lymnaea stagnalis*. Wilhelm Roux's Archives of Developmental Biology. 1977;181:11-30.
7. Verdonk NH. Gene expression in early development of *Lymnaea stagnalis*. Developmental Biology. 1973;35:29-35.
8. Morrill JB. Development of the pulmonate gastropod, *Lymnaea*. In: Harrison FW, Cowden RR, editors. New York: Alan R. Liss; 1982. p. 399-483.
9. Kuroda R, Endo B, Abe M, Shimizu M. Chiral blastomere arrangement dictates zygotic left–right asymmetry pathway in snails. Nature. 2009;462:790-794.
10. Liu MM, Davey JW, Banerjee R, Han J, Yang F, Aboobaker A, Blaxter ML, Davison A. Fine mapping of the pond snail left-right asymmetry (chirality) locus using RAD-Seq and fibre-FISH. PLoS One. 2013;8:e71067.
11. Davison A, McDowell GS, Holden JM, Johnson HF, Koutsovoulos GD, Liu MM, Hulpiau P, Van Roy F, Wade CM, Banerjee R. Formin is associated with left-right asymmetry in the pond snail and the frog. Current Biology. 2016;26:654-660.
12. Nakadera Y, Swart EM, Hoffer JNA, den Boon O, Ellers J, Koene JM. Receipt of seminal fluid proteins causes reduction of male investment in a simultaneous hermaphrodite. Current Biology. 2014;24:859-862.
13. Khan AM, Spencer GE. Novel neural correlates of operant conditioning in normal and differentially reared *Lymnaea*. Journal of Experimental Biology. 2009;212:922-933.

14. Hohagen J, Jackson DJ. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. BMC Developmental Biology. 2013;13:27.

15. Hohagen J, Herlitze I, Jackson DJ. An optimised whole mount *in situ* hybridisation protocol for the mollusc *Lymnaea stagnalis*. BMC Developmental Biology. 2015;15:19.

16. Jackson DJ, Herlitze I, Hohagen J. A whole mount *in situ* hybridization method for the gastropod mollusc *Lymnaea stagnalis*. JoVE. 2016e53968.

5

Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas

Ines Herlitze, Benjamin Marie, Frédéric Marin, Daniel J. Jackson
GigaScience. 2018;7:giy056.

5.1 Abstract

Background: Conchiferan molluscs construct a biocalcified shell that likely supported much of their evolutionary success. However beyond broad proteomic and transcriptomic surveys of molluscan shells and the shell-forming mantle tissue, little is known of the spatial and ontogenetic regulation of shell fabrication. In addition, most efforts have been focused on species that deposit nacre, which is at odds with the majority of conchiferan species that fabricate shells using a crossed lamellar microstructure, *sensu lato*.

Results: By combining proteomic and transcriptomic sequencing with *in situ* hybridisation we have identified a suite of gene products associated with the production of the crossed-lamellar shell in *Lymnaea stagnalis*. With this spatial expression data we are able to generate novel hypotheses of how the adult mantle tissue coordinates the deposition of the calcified shell. These hypotheses include functional roles for unusual and otherwise difficult to study proteins such as those containing repetitive low complexity domains (RLCDs). The spatial expression readouts of shell- forming genes also reveal cryptic patterns of asymmetry and modularity in the shell-forming cells of larvae and adult mantle tissue

Conclusions: This molecular modularity of the shell-forming mantle tissue hints at an intimate association between structure, function and evolvability and may provide an elegant explanation for the evolutionary success of the second largest phylum among the Metazoa.

5.2 Introduction

Due to its evolutionary significance, impressive materials properties and aesthetic beauty the molluscan shell has long received attention from a wide variety of scientific disciplines [1–6]. Although molluscan shells are constructed from a complex mixture of $CaCO_3$, carbohydrates [7] [8] and lipids [9], proteins have received the most attention arguably for two main reasons: they can provide deep insight into the evolutionary history of this composite structure; and the techniques for the high throughput study of these molecules are well established and are technically straight forward. Much progress has been made in identifying the components of the shell forming proteome from a variety of gastropod and (primarily) bivalve species (for example [10–17]) largely due to advances in nucleic acid sequencing technologies which, when coupled with high-throughput proteomic surveys of the biomineralised proteome, allow for the rapid generation of extensive lists of shell-associated proteins. However without further validation, genes identified in this way should only be considered as candidate biomineralising molecules. This problem is often compounded by the fact that these proteins often share little to no sequence similarity with proteins from conventional model organisms, making any inference about their function very difficult. This bottleneck represents one of the current major challenges for scientists interested in understanding the mechanisms and evolution of molluscan biomineral formation. While knock-down of individual shell-forming genes via RNAi has been reported in some species of bivalves [10, 18, 19], these assays are rarely validated by protein immuno-detection, and levels of penetrance or statistical quantitation of knock-down phenotypes are rarely reported.

Another approach to gain insight into the function of shell-forming genes is to characterise their spatial expression patterns *in situ*. We previously adopted this approach in the tropical abalone *Haliotis asinina* with a Sanger EST dataset and characterised the spatial expression patterns of over 20 putative shell-forming genes in juvenile snails [16]. While a spatial expression pattern in the mantle is not direct evidence of a functional role in calcification, we were able to assign putative functions to genes involved in shell pigmentation [16] and ecological and mineralogical transitions [20]. Here we have combined an NGS transcriptome analysis of adult mantle tissue with a proteomic survey of the adult shell of the

freshwater pulmonate gastropod *Lymnaea stagnalis* in order to both compare the resulting data with other similar datasets, and to generate the first ever *in situ* validated ontogenetic transcriptome-scale dataset for a species that forms the most common molluscan shell microstructure, crossed lamellar [21–23]. The high order structure of crossed- lamella which allows it to efficiently deflect and arrest cracks [24–27], coupled with its extremely low organic content (typically <0.5%), has been suggested to be one reason it has enjoyed so much evolutionary success (reviewed in [28]). Recent proteomic studies have been reported for molluscs that build crossed lamellae shells (*Helix aspersa maxima* [29] and *Cepaea nemoralis* [14]), however those studies did not conduct any spatial expression analyses for the shell-forming proteins they identified. In addition to characterising the spatial expression patterns of more than 30 shell- forming candidates in the adult mantle tissue of *L. stagnalis* we have also investigated their spatial expression patterns during development.

Our analyses hint at the potential pleiotropic nature of some of these shell-forming genes, and highlight the dynamic and asymmetric natures of their spatial regulation. A striking result of our analyses in the adult mantle tissue is the degree of spatial modularity displayed by distinct sets of genes. This general observation may contribute to an explanation of why the molluscan shell is apparently so evolvable. With the availability of a draft *L. stagnalis* genome and transcriptome data from a variety of adult tissues we have also investigated the genetic architectures of our biomineralisation candidates and explored to what extent alternative splicing plays a role in shell formation in *L. stagnalis*. These genes can also be compared with similar datasets from distantly related molluscs that build shells with alternative polymorphs of calcium carbonate (calcite vs. aragonite) and textures (prismatic vs. nacreous vs. crossed lamellae). Such comparisons can generate testable hypotheses regarding which components of the shell-forming toolkit contribute to these differences and which components are required for more fundamental aspects of shell formation.

5.3 Methods

5.3.1 Cultivation of adult *L. stagnalis*

Lymnaea stagnalis (Mollusca; Gastropoda; Heterobranchia; Euthyneura; Panpulmonata; Hygrophila; Lymnaeoidea; Lymnaeidae; Lymnaea) does not fall under the German animal protection act §8 and is listed as 'least concern' under the IUCN's list of threatened species. This work was therefore exempt from the University of Göttingen Ethics Committee. Adult specimens of *L. stagnalis* derived from animals originally collected from the Northeimer Seenplatte, Germany (51° 43' 26.5368', 9° 57' 24.75') and from a pond on the north campus of the University of Göttingen, Germany (51° 33' 23.727', 9° 57' 25.617') were kept in a "Stand-alone V30 unit" (Aqua Schwarz) in demineralized water supplemented with ReMineral+ (Dennerle, #7036) to a conductivity of 200 - 220 µS, 23 °C and a pH of 7.5 to 7.9 and a 16:8 light regimen. Five to ten individuals were kept in three or five litre boxes under a constant and low flow rate. Snails were fed *ad libitum* with lettuce and a variety of other vegetables. Under this regime adult snails lay egg masses year round.

5.3.2 Organic matrix extraction from calcified shells

Twelve shells of adult *L. stagnalis* (larger than 3-4 cm in length) were selected for extraction. Prior to further treatment, the columella was delicately cut and removed from each shell. Superficial organic contaminants were removed by incubating pooled shell fragments in 10%, v/v sodium hypochlorite (NaOCl) for 24 h. Fragments were then thoroughly rinsed with water and subsequently ground into a fine powder that was sieved (> 200 µM). This biomineral powder was incubated in 5% v/v NaOCl for 5 h and rinsed twice with MilliQ water. Powdered samples were decalcified overnight at 4 °C in cold 5%, v/v acetic acid which was slowly added by an automated burette (Titronic Universal, Mainz, Germany) at a flow rate of 100 µL every 5 s. The solution (final pH ~4.2) was centrifuged at 3,900 g for 30 min. The resulting acid insoluble matrix (AIM) pellet was rinsed six times with MilliQ water, freeze-dried and weighed. The supernatant containing acetic acid- soluble matrix (ASM) was filtered (Millipore, 5 µM) and concentrated in an Amicon ultra-filtration stirred cell (model 8400, 400 mL) on a Millipore membrane (10 kDa cut-off). The final solution (> 5 mL) was extensively dialysed

against 1 L of MilliQ water (six water changes) before being freeze- dried and weighed.

5.3.3 Sample preparation for proteomic analysis

In-solution digestion of unfractionated ASM (0.1 mg) and AIM (1 mg) material was performed as follows. Samples were reduced with 50 µL of 10 mM dithiothreitol in 50 mM NH_4HCO_3 for 30 min at 50 °C. Alkylation was performed with 50 µL of 100 mM iodoacetamide in 50 mM NH_4HCO_3 for 30 min at room temperature in the dark. The solution was then treated with 1 µg of trypsin (proteomic grade; Promega) in 10 µL of 50 mM NH_4HCO_3 overnight at 37 °C. Samples were then dried in a vacuum concentrator and re-suspended in 30 µL of 0.1% trifluoroacetic acid and 2% CH_3CN.

5.3.4 Peptide fractionation and data acquisition

Mass spectrometry (MS) was performed using a 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, The Netherlands). One microliter of each sample was loaded onto a trap column (PepMap100 C18; 5 µm; 100 Å; 300 µM x 5 mm; Dionex), washed for 3 min at 25 µL/min with 0.05% trifluoroacetic acid/2% acetonitrile, then eluted onto a C18 reverse phase column (PepMap100 C18; 3 µm; 100 Å; 75 µM x 150 mm; Dionex). Peptides were separated at a flow rate of 0.300 µL/min with a linear gradient of 5– 80% acetonitrile in 0.1% formic acid over 120 min. MS data were acquired automatically using ANALYST QS 1.1 software (Applied Biosystems). Following a MS survey scan over *m/z* 400–1600 range, MS/MS spectra were sequentially and dynamically acquired for the three most intense ions over *m/z* 65–2000 range. The collision energy was set by the software according to the charge and mass of the precursor ion. MS and MS/MS data were recalibrated using internal reference ions from a trypsin autolysis peptide at *m/z* 842.51 [M + H]+ and *m/z* 421.76 [M + 2H]2+.

5.3.5 Mass spectrometry data analysis

Protein identification was performed using the MASCOT search engine (version 2.1; Matrix Science, London, UK) against translations in all 6 frames of our mantle transcriptomes which possessed BUSCO completeness scores of >98% (see below). LC-MS/MS data were searched using carbamido-methylation as a fixed modification, and methionine oxidation as a variable modification. The peptide mass and fragment ion tolerances were set to 0.5 Da. The peptide hits (protein score >50; FDR <0.05; 1 missed cleavage allowed) were manually confirmed by the observation of the raw LC-MS/MS spectra with ANALYST QS software (Version 1.1). Quality criteria were the peptide MS value, the assignment of major peaks to uninterrupted y- and b-ion series of at least three to four consecutive amino acids and the match with the *de novo* interpretations proposed by the software. All mass spectrometry data has been deposited with the ProteomeXchange Consortium via PRIDE [30] with the dataset identifiers PXD008547 and 10.6019/PXD008547. Shell-forming candidates *Lstag-sfc*-7, *Lstag-sfc*-8 and *Lstag-sfc*-9 were bioinformatically selected for analysis based on the presence of a signal peptide and their glycine-rich sequences (i.e. they were not detected using the proteomic methods described above).

5.3.6 Bioinformatic analysis of protein sequences

Using the peptides identified from the proteomic survey described above, partial, or in most cases full length, coding sequences were isolated by standard or RACE PCR as described in [31]. In some cases Illumina transcriptome data (see below) was used to clarify the putative complete mRNA. Open reading frames were translated with the ExPASy translate tool [32]. Protein sequences were searched for signal sequences with SignalP 4.1 [33]. The theoretical pI, amino acid composition and number of amino acids were determined using ExPasy ProtParam tool [34]. Tandem repeats were identified with the T-REKS tool [35]. Sequence similarities searches were performed the BLAST algorithm [36] with tBLASTx against nr and dbEST, and BLASTx against SwissProt. Domain searches were performed with CD search [37]. Molecular function was predicted with InterPro Scan [38]. GalNAc O-glycosylation sites were predicted using the NetOGlyc 4.0 Server [39]. Scaled schematics of protein sequences were generated using Gene

Structure Draw [40]. Intron-exon boundaries were mapped to a draft genome of *L. stagnalis* (Dryad accession http://doi.org/10.5061/dryad.r4342) originally reported in [41] using Splign [42]. Similar transcripts were retrieved from the assembled transcriptomes of mantle zones 1-5 combined, mantle zone 5 alone, cephalic tentacle, cephalic lobe, CNS, foot, buccal mass and larval stages 42 hours post first cleavage (hpfc), 52 hpfc and 67 hpfc using BLASTn searches (see below for NGS details). All transcripts with complete open reading frames (ORF) were considered. Only candidates yielding an mRNA coverage of >98% and an overall identity of >98% are documented. Scaled schematics of the gene architecture were generated using Gene Structure Draw [40]. Protein patterns were searched for using a modified local installation of PatMatch [43].

2.3.7 NGS sequencing

Total RNA was extracted from the mantle edge and the proximal mantle tissue of a single adult *L. stagnalis* using TriReagent following the manufacturers instructions. The resulting RNA was processed by the sequencing center at the IKMB at the University of Kiel (Germany). Paired end, stranded TrueSeq RNA libraries were constructed and sequenced for 101 bases from both ends using the Illumina HiSeq2000 platform. More than 99 million and 100 million reads were generated from each of these libraries respectively. These Illumina reads were processed using our pipeline as previously described [44]. Briefly, raw reads were adapter trimmed and quality filtered using Trimmomatic V0.32. Filtered reads were then assembled with Trinity V2.0.3, CLC Genomics Workbench *de novo* assembler (V8.5), and IDBA-tran. The resulting assemblies were then merged and filtered for redundancy using our pipeline [44]. Mantle transcriptome assemblies and cDNA and protein translations of the 34 shell-forming genes are available in the corresponding GigaScience Database publication [45]. In addition, transcriptomes from 5 adult tissues (cephalic tentacle, cephalic lobe, CNS, foot and buccal mass) and 3 larval stages (42 hours post first cleavage (hpfc), 52 hpfc and 67 hpfc) were sequenced and assembled as described above. These transcriptomes were used to assess the tissue specific alternative splicing characteristics of all shell-forming genes. All raw NGS data has been deposited in the SRA with BioSample accession numbers SAMN08117214, SAMN08117215, SAMN08709370, SAMN08709371,

SAMN08709372, SAMN08709373, SAMN08709374, SAMN08709375, SAMN08709376 and SAMN08709377.

5.3.8 *In situ* hybridisation on whole mounts and sections

Larvae were prepared for whole mount *in situ* hybridisation as described in [46]. Sections (10 µM) were taken from *L. stagnalis* (shell length 10-50 mm), which had been fixed in formaldehyde for 1 hour and embedded in paraffin. Riboprobes were prepared as described in [16] and were used at concentrations of 100 ng – 500 ng/mL. Whole mounts and tissue sections were processed for hybridisation, the colour reaction developed and photo-documented as described in [46].

5.3.9 Comparisons of molluscan shell forming proteomes

BLASTp based comparisons of the *L. stagnalis* shell proteome were performed against a variety of calcifying proteomes reported in a wide phylogenetic range of metazoans as described in [14]. These included: 42 proteins from the oyster *Pinctada maxima* reported in [47]; 78 proteins from the oyster *Pinctada margaritifera* reported in [47]; 94 proteins from the abalone *Haliotis asinina* reported in [17] and [16]; 80 proteins from the abalone *H. laevigata* reported in [48]. 63 protein from the limpet *Lottia gigantea* reported in [49]; 53 proteins from the oyster *Crassostrea gigas* reported in [50]; 71 proteins from the mussel *Mya truncata* reported in [51]; 59 proteins from the grove snail *Cepaea nemoralis* reported in [14]; 44 proteins from the oyster *Pinctada fucata* reported in [52]; 53 proteins from the mussel *Mytilus coruscus* reported in [53]; 66 proteins from the brachiopod *Magellania venosa* reported in [54]; 139 proteins from the sea urchin *Strongylocentrotus purpuratus* reported in [55]; 37 proteins from the coral *Acropora millepora* reported in [56].

5.3.10 Analysis of the saccharide moieties of the shell matrix

The monosaccharide content of AIM and ASM was obtained by suspension and homogenization (vortex and ultrasound) of lyophilates in 2 M trifluoroacetic acid (TFA) and subsequent hydrolysis at 105 °C for 4 h under a nitrogen atmosphere. This treatment allows for the release of most monosaccharides from complex mixtures, except sialic acids which are destroyed, and the acetylated forms of

glucosamine and galactosamine, which are converted to their respective non-acetylated forms. Samples were then centrifuged for 5 min, at 15,000g and evaporated to dryness (using a SpeedVac) before being dissolved in 100 μL of 20 mM NaOH and homogenized. After a short centrifugation (2 min), 80 μL of supernatant was injected into the chromatograph system: the neutral, amino and acidic sugar contents of hydrolysates were determined by HPAE-PAD (High Pressure Anion Exchange - Pulsed Amperometric Detection) on a CarboPac PA 100 column (Dionex Corp., Sunnyvale, CA, USA). As blank controls, non-hydrolysed AIMs were analysed, in order to detect potential free monosaccharides that may lead to an over-representation of some sugar residues.

5. 4 Results

5.4.1 Proteomic analysis of the biomineralised matrix of *L. stagnalis* shells

More than 1,230 peptides were analyzed by HPLC-mass spectrometry and subsequently used for protein identification using Mascot against our translated mantle transcriptomes. Of these 1,230 peptides 329 of them returned significant matches. From these 329 matches a total of 40 shell-forming candidate transcripts were identified (see Additional file 5.1). Of these 40 gene products, 31 (78%) could be cloned and exhibit *in situ* hybridisation signals compatible with a role in shell formation (either in larval stages and/or in the adult mantle tissue). Seven of these 40 candidates (18%) could be cloned from *L. stagnalis* cDNA, but did not produce a positive or consistent *in situ* signal in any tissue. Three of the 40 candidate genes (8%) could not be amplified by gene specific PCR or RACE PCR. In addition to the 31 proteomically identified candidates that generated positive *in situ* signals, three candidates that were identified via *in silico* methods (based purely on the presence of a signal sequence and their glycine-rich protein sequences) also generated *in situ* signals compatible with a role in shell formation and are reported here.

5.4.2 A brief morphological description of *L. stagnalis* shell ontogeny and the adult mantle

We previously described the ontogeny of the shell gland and shell field in *L. stagnalis* [57]. In order to aid the interpretation of our *in situ* patterns the

following is a summary of the main developmental stages that we focused on. The first visible sign of differentiation of the shell-forming tissue in *L. stagnalis* is a thickening of the dorsal ectoderm that begins at approximately 29 hours post first cleavage (hpfc) [57, 58]. These cells subsequently invaginate and by 2 days post first cleavage (dpfc) a clearly visible 'shell gland' is present [57, 58]. By 3 dpfc, the shell gland has formed a sealed lumen, and displays the first signs of outward signs of asymmetry [57]. The marginal cells that border the shell gland remain uninvaginated and form a ring-like structure, the rosette [2]. During this time the first extracellular organic material is secreted and is clearly visible by SEM (Fig. 5.1; [57]). By 3 dpfc, the shell gland has evaginated to form the shell field. The former rosette cells remain highly elongated while the central cells take on a low columnar appearance. Over the next several days the shell field continues to expand until it has overgrown the visceral mass and will eventually become the adult mantle tissue [2, 57, 58].

The adult mantle covers the inner surface of the shell and is responsible for shell growth and repair. The free edge of the mantle is responsible for the growth of the outer lip of the shell. Timmermans conducted an extensive histochemical characterisation of the mantle tissue of *L. stagnalis* and she was able to categorise the free edge of the adult mantle into six distinct zones based on their morphology, enzymatic activities and biochemical signatures [59]. We largely follow this categorisation of the adult mantle tissue. Parallel to the mantle edge runs the mantle groove (also know as the pallial groove) defined as zone 1 (Fig. 5.2). Several high-resolution microscopy and histological studies on a variety of molluscs have demonstrated that it is from within the pallial groove that the periostracum is formed and secreted [59–64]. We detected a sub-regionalisation of the pallial groove (zone 1) into proximal and distal zones. Immediately adjacent to the pallial groove is a broad region of high columnar cells referred to by Timmermans [59] as the 'belt' that can be subdivided into three distinct zones (zones 2-4). Zone 2 is immediately adjacent to the posterior wall of the pallial groove and comprises the anterior (or distal) portion of the belt (Fig. 5.2). Zone 3 consists of the posterior portion of the belt, while zone 4 represents the transitional zone between the high columnar cells of the belt proper and the more posterior low columnar cells of the outer epithelium which comprise zone 5 (Fig.
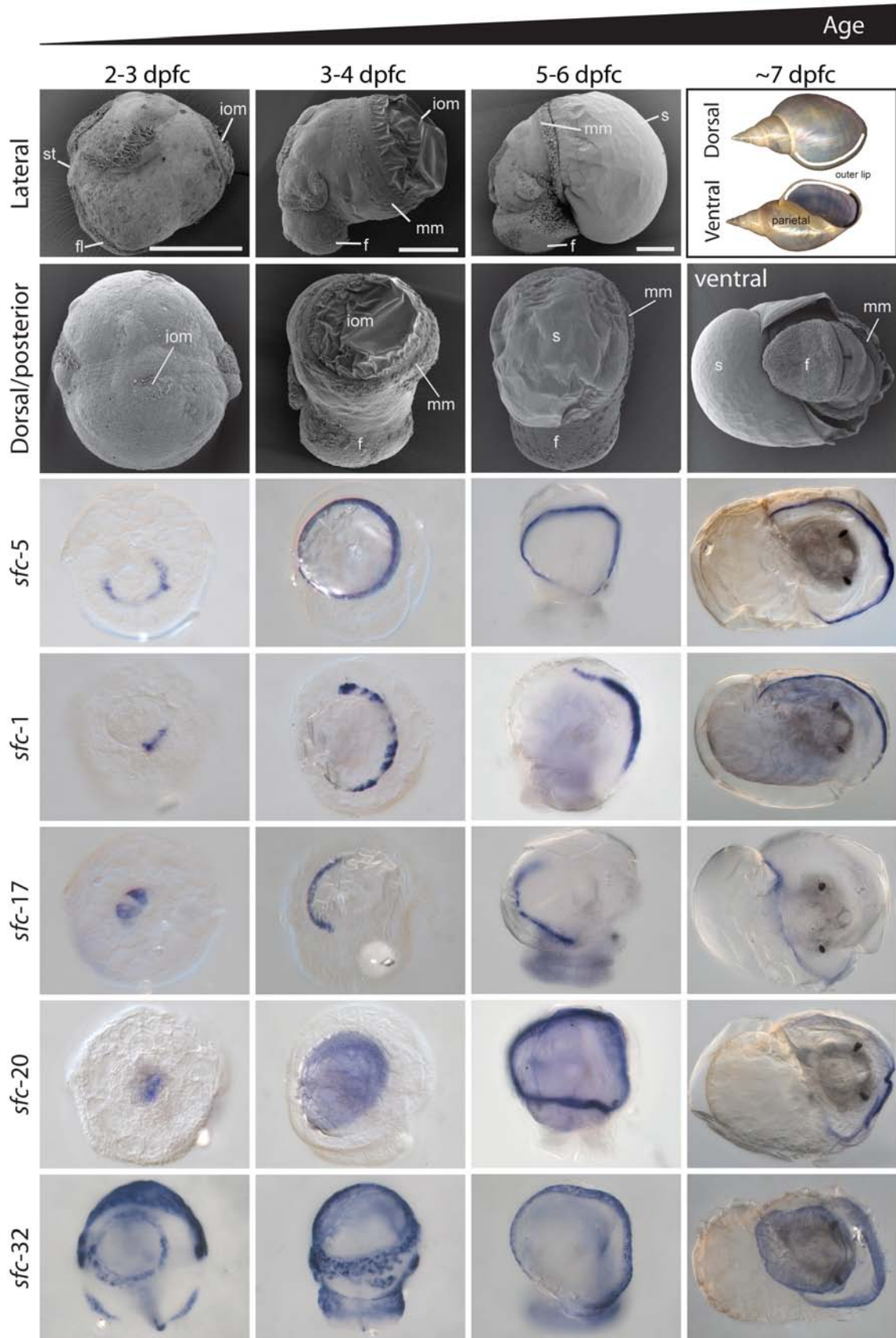
Fig. 5.1. Overview of four developmental stages and representative shell-forming gene expression patterns in *L. stagnalis.* The first two rows provide a set of reference SEM images and adult shells (top-right-most panel) against which the images of the *in situ* results can be oriented. All *in situ* panels are from a dorsal view except the right-most column which is from a ventral view. Here we present the expression patterns of a selection of 5 different shell-forming genes. These include genes with expression patterns in shell-forming cells that display evidence of symmetry (*sfc-5*), right asymmetry (*sfc-1*), left-asymmetry (*sfc-17*), expression entirely throughout the shell field and dorsal mantle epithelium (*sfc-20*) and expression in additional non-shell-forming cells. This last expression pattern provides evidence of genes involved in shell-formation that have pleiotropic functions. The scale bars in the first row are 100 μm. Indicated in the SEM images are the positions of the foot lobe (fl), foot (f), mantle margin (mm), calcified shell (s), stomodeum (st) and insoluble organic material (iom) of the shell.

5.2) [59].

5.4.3 Spatial expression patterns and molecular features of shell-forming candidate genes

We performed *in situ* hybridisation for 34 distinct shell-forming genes on 4 distinct developmental stages and on adult mantle tissue. The detailed results of these analyses are presented in Additional files 5.2 – 5.35 with an extensive summary presented in Additional file 5.36 (the raw image files that constitute these figures are available in the associated GigaDB repository [45]. In Figure 5.1 (for larvae) and Figure 5.2 (for adult mantle tissue) we present a selection of these results that highlight some prominent features of these expression patterns. In trochophore and veliger larval stages (2-6 dpfc) all genes could be categorised as either being expressed in: cells that symmetrically or asymmetrically border the shell gland or shell field (15/34); cells that lay within the shell gland or shell field (9/34); a pattern that did not fit into our classification scheme (1/34); or were not expressed in any detectable way (9/34). In later stages (~7 dpfc) all genes were either expressed: uniformly along the outer edge of the mantle (10/34); asymmetrically in the outer edge of the mantle (18/34); throughout the entire mantle tissue (2/34); a pattern that did not fit into our classification scheme (1/34); or were not expressed in any detectable way (3/34). Finally in adult mantle tissue genes were either expressed: in one or more of the 5 zones described by Timmermans [59] (32/34); a pattern that did not fit into our classification scheme (1/34); or were not expressed in any detectable way (1/34). We have schematically summarised all of these results in Fig. 5.3.
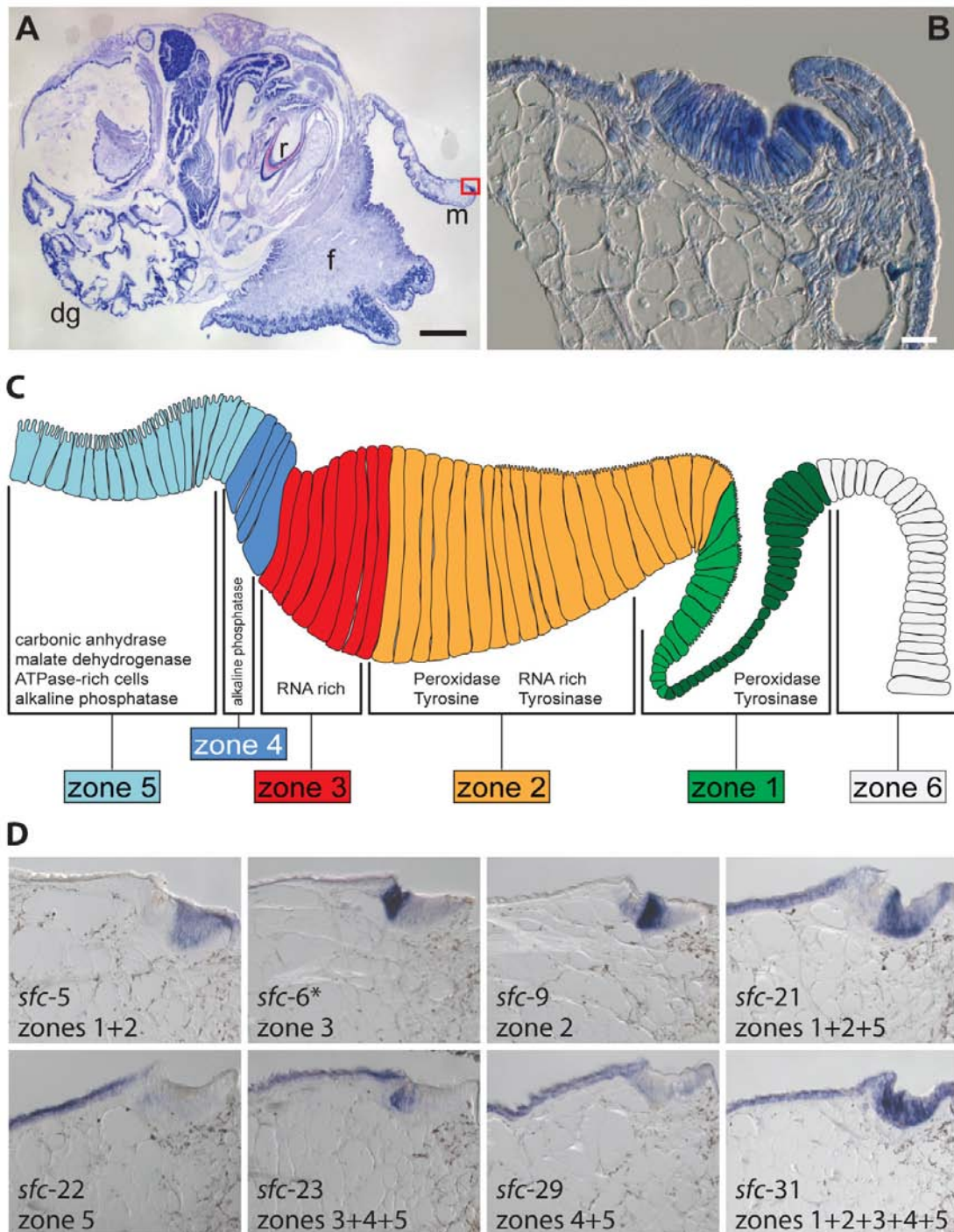
Fig. 5.2. Overview of the adult *L. stagnalis* shell-forming mantle tissue and representative shell-forming gene expression patterns that reveal its molecular modularity. A. A semi-thin sagittal section of an adult *L. stagnalis* stained with Giemsa. The foot (f), mantle (m), digestive gland (dg) and radula (r) are indicated. The mantle tissue is a thin sheet of epithelium that covers the dorsal surface of the adult animal and is responsible for fabricating the shell. **B.** A magnified view of the red-boxed region in A reveals some of the cellular morphology of the adult mantle tissue. **C.** A schematic representation of the mantle tissue divided into 6 zones as described by Timmermans [59]. The spatial distribution of enzymatic activities and biochemicals indicated in this schematic are adapted from [59]. We detect a sub-regionalisation of the pallial groove (zone 1) into proximal (light green) and distal (dark green) zones. **D.** The spatial expression patterns of 8 representative shell-forming genes in the adult mantle tissue. The asterisk indicates that *sfc*-6 was identified using *in silico* methods rather than proteomic methods.
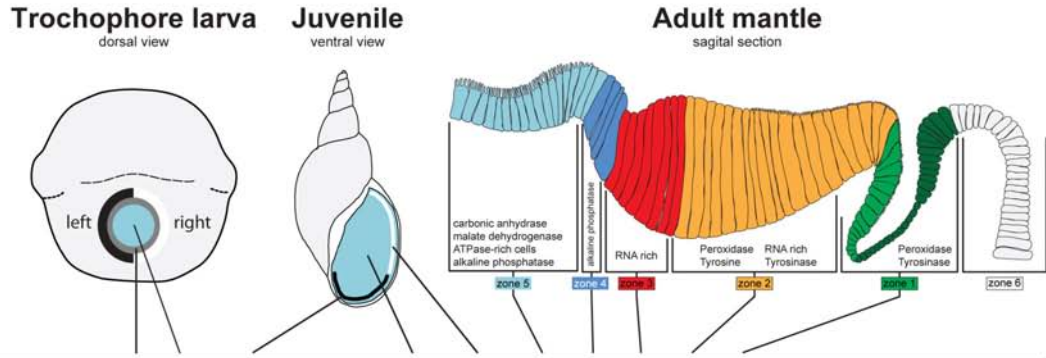
Shell-forming candidate	Asymmetric border	Symmetric border	Interior	Left side of outer lip	Left + right side of outer lip	Entire mantle	Right side of outer lip	Zone 5	Zone 4	Zone 3	Zone 2	Zone 1	Sequence similarity or conserved domain
Lstag-sfc-1	right			✓							✓	✓	
Lstag-sfc-2a	right			✓							✓	✓	Early nodulin-12A (SP)
Lstag-sfc-2b	right			✓							✓	✓	
Lstag-sfc-3	right			✓							✓	✓	None
Lstag-sfc-4	right				✓							✓	None
Lstag-sfc-5		✓			✓						✓	✓	Animal haem peroxidase (CD)
Lstag-sfc-6a*		✓					✓			✓			None
Lstag-sfc-6b*		✓					✓			✓			None
Lstag-sfc-7*		✓					✓			✓			None
Lstag-sfc-8*	✗	✗	✗				✓			✓			None
Lstag-sfc-9a		✓			✓						✓		None
Lstag-sfc-9b		✓			✓						✓		None
Lstag-sfc-10a		✓			✓					✓			None
Lstag-sfc-10b		✓			✓					✓			None
Lstag-sfc-11a		✓					✓			✓			None
Lstag-sfc-11b		✓					✓			✓			None
Lstag-sfc-11c		✓					✓			✓			None
Lstag-sfc-12	left						✓			✓			None
Lstag-sfc-13			✓			✓				✓			None
Lstag-sfc-14			✓				✓			✓			None
Lstag-sfc-15	✗	✗	✗				✓		✓	✓	✓	✓	None
Lstag-sfc-16	✗	✗	✗	✗	✗	✗	✗			✓	✓	✓	None
Lstag-sfc-17	left						✓		✓				None
Lstag-sfc-18	?	?	?	?	?	?	?		✓	✓	✓		Immunoglobulin domain (CD)
Lstag-sfc-19	✗	✗	✗	✗	✗	✗	✗		✓				Perlucin-like protein (SP), C-type lectin (CD)
Lstag-sfc-20a			✓				✓		✓				None
Lstag-sfc-20b			✓				✓		✓				None
Lstag-sfc-21	✗	✗	✗		✓			✓			✓	✓	Putative chitin deacetylase-like domain (CD)
Lstag-sfc-22	✗	✗	✗	✗	✗	✗	✗	✓					Pif97 Aragonite-binding protein (SP)
Lstag-sfc-23a			✓				✓	✓	✓	✓			Otoancorin (SP)
Lstag-sfc-23b			✓				✓	✓	✓	✓			Otoancorin (SP)
Lstag-sfc-24a			✓				✓	✓					None
Lstag-sfc-24b			✓				✓	✓					None
Lstag-sfc-25	✗	✗	✗		✓			✓	✓				None
Lstag-sfc-26			✓				✓	✓					PREDICTED: uncharacterized protein (NR)
Lstag-sfc-27a			✓				✓	✓					PREDICTED: extensin-like isoform X1 (NR)
Lstag-sfc-27b			✓				✓	✓					PREDICTED: formin-like protein 2 (NR)
Lstag-sfc-28a			✓		✓			✓					None
Lstag-sfc-28b			✓		✓			✓					None
Lstag-sfc-28c			✓		✓			✓					None
Lstag-sfc-29	✗	✗	✗				✓	✓	✓				Galaxin (SP)
Lstag-sfc-30			✓			✓		✓					None
Lstag-sfc-31	✗	✗	✗				✓	✓	✓	✓	✓	✓	SUSHI repeat + short complement-like repeat (CD)
Lstag-sfc-32		✓			✓			✓	✓	✓	✓	✓	Intermediate filament protein (CD)
Lstag-sfc-33		✓			✓			?	?	?	?	?	vWA type A domain + collagen alphaI-XII-like (CD)
Lstag-sfc-34		✓			✓			✗	✗	✗	✗	✗	None

Figure 5.3. Summary of the spatial gene expression profiles and conserved features of 34 ***L. stagnalis*** **shell-forming candidates.** Schematically represented in a trochophore larva are genes with an asymmetric expression profile (dark grey), as well as genes expressed broadly across the shell field (light blue). Cells in this region of the trochophore are likely to give rise to cells in zone 5 of the adult mantle, and we have maintained that colour scheme to suggest this. Although we schematically present a trochophore larva here (2-3 dpfc) the summarised expression patterns also include veliger stages (3-6 dpfc). Cells bordering the larval shell gland and shell field (black ring in the trochophore) are likely to give rise to one or more zones 1-4 in the adult mantle. In juveniles (~7 dpfc) many genes were either expressed in the left, right, or continuously throughout the free edge of the mantle that produces the outer lip of the shell. Question marks indicate expression patterns that could not be categorised according to our scheme. An "x" indicates no expression was detected. A "?" indicates that the expression pattern could not be categorised according to our scheme. The names of enzymes and other molecular features indicated in zones 1-5 on the schematic of the adult mantle are summarised from [59] and [3]. Sequence similarity and conserved domains in the final column of the table are summarised from a number of BLAST searches against SwissProt (SP), the non-redundant (NR) NCBI database and the Conserved Domain (CD) database. See Additional files 40-42 for the results of all BLAST and domain searches. The lineages of the top BLAST hits are listed in Additional file 43. A version of this figure that includes a more complete summary of the molecular features of each gene is provided in Additional file 36. The asterisks indicate that *sfc-6, -7* and *-8* were identified using *in silico* methods rather than proteomic methods (as was the case for all other gene products presented here).

5.4.4 Alternative splicing increases the diversity of shell-forming proteins

Via alternative splicing of mRNAs, transcripts with a variety of functions can be generated from a single genomic locus [65]. With a draft genome for *L. stagnalis* available we were able to perform some preliminary investigations into alternative splicing of our shell-forming candidates. While some candidate genes displayed the same or very similar exon-splicing patterns in all surveyed tissues (for example Additional files 5.5, 5.9, 5.14, 5.15 and 5.16), most candidates are apparently alternatively spliced depending on the tissue they are expressed in (Additional files 5.4, 5.11, 5.12, 5.18, 5.19, 5.22, 5.23, 5.25 and 5.33). Striking examples include *Lstag-sfc-21* and *Lstag-sfc-24* which are expressed in many tissues, but display significant alternative splicing in the adult mantle (Additional files 5.22 and 5.25). All splice variants of candidate *Lstag-sfc-24* encode proteins with the same aspartic acid-rich motif (Additional files 5.25, 5.37 and 5.38). Aspartic acid-rich proteins have been suggested to act as an organic template for epitaxial crystal growth [66, 67]. It is tempting to speculate that the three additional domains only present in adult mantle *Lstag-sfc-24* contigs confer a specific shell-forming function to this protein. The putative chitin-interacting candidate *Lstag-sfc-21* presented in Additional file 5.22 carries a signal sequence and is predicted to possess a catalytic activity. Intriguingly a number of splice variants of this gene within the adult

mantle are predicted to lack a signal sequence, the chitin-binding or catalytic ability (Additional file 5.37).

A number of shell-forming gene candidates produce alternatively spliced transcripts that encode proteins with differences regarding the presence/absence of a signal sequence (Additional files 5.11, 5.12, 5.22, 5.25 and 5.37). Some shell-forming genes also produce alternatively spliced transcripts that encode proteins with similar coding features, but radically different 5' or 3' UTRs (Additional files 5.12, 5.19, 5.33 and 5.37). While UTRs do not contain protein-coding information, they can be critical for localisation of the mRNA [68, 69] and post- transcriptional gene regulation by molecules such as miRNAs [70]. Indeed several miRNAs have now been associated with the targeting and regulation of biomineralising proteins [71, 72].

5.4.5 Comparisons of molluscan shell forming proteomes

We conducted a broad comparison of our *L. stagnalis* shell forming genes against a wide phylogenetic range of twelve other biomineralising proteomes comprising in total 879 proteins (sequences used in this analysis are provided in Additional file 5.39). Of all *L. stagnalis* shell proteins, 27 shared significant sequence similarity with one or more proteins derived from one of these twelve proteomes (Fig. 5.4). The highest degree of overall similarity was found with the shell forming proteome of the common groove snail *C. nemoralis* (Fig. 5.4), the closest phylogenetic relative to *L. stagnalis* of all species in this comparison. The next highest level of similarity shared with the *L. stagnalis* shell forming proteome (15.9% of the *L. gigantea* shell forming proteome) was markedly lower. *L. stagnalis* shell-forming proteins that shared significant similarity with biomineralizing proteins from other species and that also returned a significant match against a SwissProt entry included *Lstag*-sfc-32 (with similarity to *C. nemoralis* contig 572), which appears to be an intermediate filament protein. *Lstag*-sfc-22 (a gene expressed exclusively in zone 5, Additional file 5.23) shared relatively weak similarity with *C. nemoralis* contig 821, and shares significant sequence similarity with PIF, an aragonite binding protein reported to be involved in nacre formation in the oyster *P. fucata* [19]. Interestingly, of the 12 candidates expressed in the matrix-secreting zone 5, nine showed similarity with other shell proteins (Figs. 5.4, and Additional files 5.23-

5.33). Eight of them were shared with *C. nemoralis* (Fig. 5.4). In contrast, none of the asymmetrically expressed or glycine-rich candidates were found in any of the other biomineralizing proteomes (low complexity filtering was inactivated in these comparison; Fig. 5.4).

5.4.6 Glycosylation patterns of the shell matrix

The monosaccharides profiles of ASMs and AIMs extracted from adult *L. stagnalis* shells were peculiar, with less than half of the dozen standard monosaccharides represented. For ASM, these include galactosamine, glucosamine, galactose, glucuronic acid and glucose, while AIM lacked glucuronic acid (Table 5.1). We also found marked differences in the glycosylation rates of acid-soluble and acid-insoluble matrices. In general the absolute degree of glycosylation of the ASM was 1-3 orders of magnitude greater that of the AIM. The largest difference was in the amount of galactosamine (10.6 ng/µL vs. 0.03 ng/µL). While glucosamine was more abundant absolutely in the ASM (9.41 ng/µL vs. 0.13 ng/µL), the proportional difference was not so extreme (34.8% vs. 54.2%).

5.5 Discussion

5.5.1 Molecular modularity of the adult molluscan mantle

Two of the most striking features of the phylum Mollusca are its size (in terms of number of species) and its diversity. Widely accepted to be second only to the Arthropoda in terms of number of living species [73, 74], molluscs arguably display the greatest diversity of body forms of all metazoan phyla, and have successfully colonized all kinds of environments. While there currently exists no consensus as to why molluscs have enjoyed such deep evolutionary success (one interesting suggestion includes a plastic nervous system, [75]) we believe the mantle tissue (an apomorphy of the phylum) and its ability to prolifically evolve new shell phenotypes must contribute to an explanation of this success. A logical extension of this question would therefore be, "what is it about the molluscan mantle tissue that makes it so evolutionarily plastic?" For arthropods, segmentation and body plan modularisation (and the underlying gene regulatory networks that control appendage identity within each segment) are widely
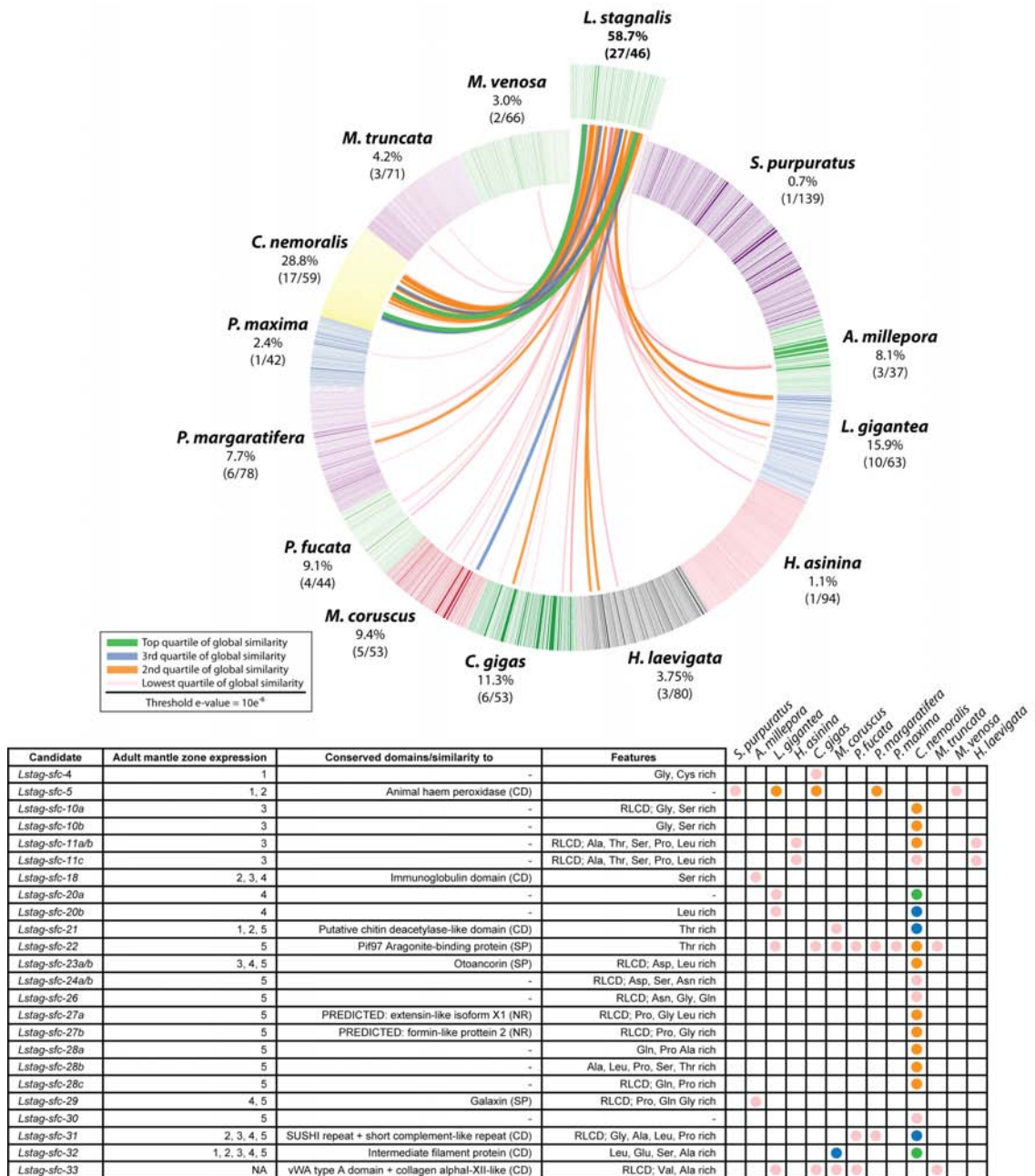
L. stagnalis
58.7%
(27/46)

M. venosa
3.0%
(2/66)

M. truncata
4.2%
(3/71)

S. purpuratus
0.7%
(1/139)

C. nemoralis
28.8%
(17/59)

A. millepora
8.1%
(3/37)

P. maxima
2.4%
(1/42)

L. gigantea
15.9%
(10/63)

P. margaratifera
7.7%
(6/78)

H. asinina
1.1%
(1/94)

P. fucata
9.1%
(4/44)

M. coruscus
9.4%
(5/53)

C. gigas
11.3%
(6/53)

H. laevigata
3.75%
(3/80)

Top quartile of global similarity
3rd quartile of global similarity
2nd quartile of global similarity
Lowest quartile of global similarity
Threshold e-value = 10e^{-6}

Candidate	Adult mantle zone expression	Conserved domains/similarity to	Features	S. purpuratus	A. millepora	L. gigantea	H. asinina	C. gigas	M. coruscus	P. fucata	P. margaratifera	P. maxima	C. nemoralis	M. truncata	M. venosa	H. laevigata
Lstag-sfc-4	1	-	Gly, Cys rich													
Lstag-sfc-5	1, 2	Animal haem peroxidase (CD)	-	●		●	●			●			●			
Lstag-sfc-10a	3	-	RLCD; Gly, Ser rich								●					
Lstag-sfc-10b	3	-	Gly, Ser rich								●					
Lstag-sfc-11a/b	3	-	RLCD; Ala, Thr, Ser, Pro, Leu rich				●				●			●		
Lstag-sfc-11c	3	-	RLCD; Ala, Thr, Ser, Pro, Leu rich				●				●					
Lstag-sfc-18	2, 3, 4	Immunoglobulin domain (CD)	Ser rich	●												
Lstag-sfc-20a	4	-	-			●					●					
Lstag-sfc-20b	4	-	Leu rich	●		●					●					
Lstag-sfc-21	1, 2, 5	Putative chitin deacetylase-like domain (CD)	Thr rich					●			●					
Lstag-sfc-22	5	Pif97 Aragonite-binding protein (SP)	Thr rich		●		●	●	●	●	●	●				
Lstag-sfc-23a/b	3, 4, 5	Otoancorin (SP)	RLCD; Asp, Leu rich								●					
Lstag-sfc-24a/b	5	-	RLCD; Asp, Ser, Asn rich								●					
Lstag-sfc-26	5	-	RLCD; Asn, Gly, Gln								●					
Lstag-sfc-27a	5	PREDICTED: extensin-like isoform X1 (NR)	RLCD; Pro, Gly Leu rich								●					
Lstag-sfc-27b	5	PREDICTED: formin-like protein 2 (NR)	RLCD; Pro, Gly rich								●					
Lstag-sfc-28a	5	-	Gln, Pro Ala rich								●					
Lstag-sfc-28b	5	-	Ala, Leu, Pro, Ser, Thr rich								●					
Lstag-sfc-28c	5	-	RLCD; Gln, Pro rich								●					
Lstag-sfc-29	4, 5	Galaxin (SP)	RLCD; Pro, Gln Gly rich	●												
Lstag-sfc-30	5	-	-								●					
Lstag-sfc-31	2, 3, 4, 5	SUSHI repeat + short complement-like repeat (CD)	RLCD; Gly, Ala, Leu, Pro rich							●	●		●			
Lstag-sfc-32	1, 2, 3, 4, 5	Intermediate filament protein (CD)	Leu, Glu, Ser, Ala rich					●					●			
Lstag-sfc-33	NA	vWA type A domain + collagen alpha1-XII-like (CD)	RLCD; Val, Ala rich			●	●	●	●		●			●		●

Figure. 5.4. BLASTp comparisons of the *L. stagnalis* shell proteome against 879 biocalcifying proteins derived from 6 bivalves, 4 gastropods, 1 brachiopod, 1 sea urchin and 1 coral. Individual lines spanning the ideogram connect proteins that share significant similarity (e values < 10e^{-6}). Transparent red lines connect proteins with the lowest quartile of similarity (with a threshold of 10e^{-6}) and green lines with the highest quartile of similarity. The percentage of each shell proteome that shared similarity with the *L. stagnalis* proteome is indicated. The table provides further information for those candidates that share sequence similarity with another species. Abbreviations: conserved domain database (CD); Genbank non-redundant protein database (NR); SwissProt database (SP).

 thought to have played leading roles in supporting the diversification of insects, spiders and crustaceans [76]. The importance of establishing segmentation at a very early developmental age in prominent phyla such as annelids, chordates and arthropods has caused much effort to be spent on identifying the causal molecular

mechanisms that may have common evolutionary histories [77–79]. As recently reviewed by Esteve-Altava [80], the presence of morphological modules can help us to understand the evolvability of body form, but the identification of such modules has so far been biased towards mammals, arthropods and plants. Following Esteve-Altava's and Eble's [81] definition of a morphological module (a group of body parts that are more integrated among themselves than they are to other parts outside the group) we propose that the molluscan mantle is a prime example of such a morphological module. This modular nature of the molluscan mantle is not unique to *Lymnaea* [82–84]. Although the precise functions of these zones (and of the individual gene products that define them) await the development of targeted genome editing methods, it is clear that they must act in a coordinated way to deposit the shell. We predict that there are related modules of gene regulatory networks (GRNs) that act to specify each zone of the molluscan mantle, and that it is the modular nature of these GRNs and the resulting morphological modularity of the mantle tissue that supported the diversification of the phylum Mollusca. Characterisation of the spatial expression patterns of shell-forming proteomes from a selection of other molluscan lineages will contribute to a more refined understanding of molluscan shell evolution.

Table 5.1: Glycosylation analysis of acid-soluble and acid-insoluble matrices extracted from adult *L. stagnalis* shells.

Monosaccharide	ASM		AIM	
	ng/μg	%	ng/μg	%
Fucose	ND	-	TR	-
Rhamnose	ND	-	TR	-
Galactosamine	10.60	39.2	0.03	12.5
Arabinose	ND	-	ND	-
Glucosamine	9.41	34.8	0.13	54.2
Galactose	4.34	16.0	0.04	16.7
Glucose	0.32	1.2	0.04	16.7
Mannose	ND	-	TR	-
Xylose	ND	-	ND	-
Galacturonic acid	ND	-	ND	-
Glucuronic acid	2.39	8.8	ND	-
Total	**27.06**	**100.0**	**0.24**	**100.0**

ND = not detected, TR = trace.

5.5.2 Ontogenetic expression of shell-forming candidates

A prominent outcome of our survey of the adult shell proteome is that many of the genes that encode these proteins are not only regulated spatially, but also temporally. Many shell-forming candidates are expressed in the invaginated larval shell gland of the trochophore (Additional files 5.14, 5.15, 5.21, 5.24, 5.25, 5.27 and 5.28) or in cells that border it (Additional files 5.2, 5.3, 5.6, 5.7, 5.8, 5.10, 5.11 and 5.18). Only two candidates were solely expressed in the adult mantle tissue and not in any larval stage (Additional files 5.17 and 5.23). Timmermans [59] concluded that the spatial patterning of larval shell-forming cells persists throughout development and foreshadows the zonation observable in the adult mantle. We also observed this phenomenon at the molecular level. All candidate genes that were expressed in the margin of the shell gland or the shell field were expressed in the belt (zones 2, 3 and 4) of the adult mantle (Summarised in Fig. 5.3, Additional files 5.2-5.8, 5.10-5.12 and 5.18). Most candidates expressed in the invaginated cells of the shell gland or throughout the developing shell field were subsequently expressed in the low columnar outer epithelium of zone 5 in adult mantle tissue (Summarised in Fig. 5.3, Additional files 5.24, 5.25, 5.27, 5.28, 5.29, 5.31). However three genes conspicuously deviate from this pattern. *Lstag-sfc-13, -14* and *-20* display a broad expression pattern in the invaginated cells of the shell gland and throughout the entire shell field in larvae, but were not detected in the low columnar outer epithelium (zone 5) of the adult mantle tissue (Additional files 5.14, 5.15 and 5.21). However we should point out that for all candidate shell-forming genes we did not consider the potential effect of a diurnal rhythm on gene expression. All samples for *in situ* hybridisation were taken during daylight hours, and so genes with activity during the night would be missed.

5.5.3 Asymmetric expression of shell-forming genes

The expression of *Lstag-sfc-1, Lstag-sfc-2* and *Lstag-sfc-3* in zones 1 and 2 of the adult mantle suggests they may be involved in the formation of the periostracum (Fig 5.1 and Additional files 5.2, .5.3 and 5.4), however it is their larval expression patterns that are more striking. *Lstag-sfc-1, -2* and *-3* display a right-sided asymmetric expression pattern in cells bordering the shell gland and shell field. In contrast *Lstag-sfc-17* is expressed on the left side (Fig 5.1 and Additional file 5.18).

Following the expression of these genes ontogenetically into older larvae that begin to display the coiled phenotype of the adult, it is apparent that right-sided cells in the trochophore are likely to be those that give rise to the right + anterior region of the adult mantle that will produce the outer lip of the shell, while left-sided cells will give rise to posterior mantle tissue responsible for forming the left + parietal region of the shell (Fig. 5.1). We therefore suggest that *Lstag-sfc-1, -2* and *-3* are in some way associated with producing thinner, more rapidly produced shell at the outer shell lip than in the thicker parietal region while *Lstag-sfc-17* may inhibit the rapid deposition of shell. Exactly how this is achieved awaits more specific gene function assays.

In addition to the trochophore left/right asymmetry corresponding to the left + parietal/right + outer lip regions of the shell, there is a second axis of symmetry that becomes apparent in 7-day old juvenile snails. Many shell-forming candidates are initially symmetrically expressed in or surrounding the shell gland of 2-3 dpfc trochophores, but then become asymmetrically expressed in the mantle of older animals. For example *Lstag-sfc-6, -7, -8, -12, -14, -15, -17, -18, -20, -23, -24, -26, -27, -29 and -31* are expressed in the left side of the free mantle edge in 7 dpfc juveniles (summarised in Fig. 5.3). In contrast, relatively few shell-forming gene candidates (*Lstag-sfc-*5, *-9*, *-10*, *-21* and *-25)* are expressed evenly along the free edge of the mantle in 7 dpfc juveniles (summarised in Fig. 5.3).

5.5.4 The spatial expression of a peroxidase in the adult mantle allows a model of shell formation to be developed

In agreement with Timmermans histochemical study of peroxidase activity [59], the expression of *Lstag-sfc-5*, a shell-forming candidate with an "Animal heme-dependent peroxidase" domain (Pfam PF03098; Additional file 5.41B) is localised to zones 1 and 2. Peroxidases may be involved in periostracum formation by cross-linking fibrous proteins rich in reactive quinones to form water insoluble, protease-resistant polymers [85–87]. This process, also referred to as tanning or sclerotisation, can also be catalysed by tyrosinase (also known as catechol oxidase, catecholase, polyphenoloxidase, phenoloxidase and phenolase), and within the molluscan biomineralisation literature sclerotisation by tyrosinase appears to be the more commonly assumed mechanism, rather than by peroxidase. Nonetheless,

Timmermans demonstrated that heat inactivation clears the periostracal groove and belt of both peroxidase activity and the ability to form melanin (a typical assay used to test for tyrosinase activity), while specific tyrosinase inhibitors (NaHSO$_3$ and KCN) did not affect its ability to produce melanin [59]. The spatial expression pattern of *Lstag-sfc-5*, coupled with the observations that newly secreted periostracum itself also displays peroxidase activity [57], and Timmermans experiments [59], strongly suggests that the peroxidase we have identified here plays a key role in cross-linking the periostracum in *L. stagnalis* rather than a tyrosinase, as also supposed for other gastropods such as *Lottia* [15].

5.5.6 Glycine-rich shell-forming candidates are likely to be substrates for the peroxidase

An important aspect of scleroprotein formation is its spatial coordination. The cross-linking reaction often generates cytotoxic intermediates, and the end products cannot be easily degraded or resorbed [88]. Furthermore the uncontrolled formation of extensive scleroprotein polymers prior to secretion would clearly be detrimental to the cell. One common strategy to avoid these events is to compartmentalize the scleroprotein precursor (that is unable to spontaneously polymerize) away from the cross-linking enzyme. Following secretion, the precursors are activated and enzymatically cross-linked [88]. Such a scenario would suggest that the substrate upon which the peroxidase acts are not located within the same cells.

Three candidates expressed in zone three (*Lstag-sfc-6, -7* and *-8*) encode secreted, basic proteins that are dominated by repetitive low complexity domains (RLCDs) and anomalous amino acid contents (high glycine, tyrosine, asparagine and leucine contents; Additional files, 5.37 and 5.38). All of these glycine-rich proteins carry tyrosine residues flanked by glycine. This arrangement has been shown to be favourable for the formation of cross-links between tyrosine residues by peroxidase [89]. Waite, in his review of natural quinone-tanned glues [85], highlighted the typical DOPA-containing consensus precursor peptide sequences from a number of marine invertebrates. Allowing for a single mismatch, these substrate peptides (VGG**Y**G**Y**GK, GGGFGG**Y**GK and GGG**Y**GG**Y**GK, crosslinking tyrosine residues in bold) can be found within *Lstag*-sfc-6, *Lstag*-sfc-7 and *Lstag-*

sfc-8. Interestingly, these glycine-rich candidates are expressed exclusively in zone three (Additional files 5.7-5.9) immediately adjacent to zone 2, the region in which the peroxidase *Lstag*-sfc-5 is expressed (Additional file 5.6). Theoretically, once these proteins are secreted, the secreted peroxidase would be in very close proximity to the glycine-rich proteins and could act on the favoured tyrosine residues to form di-tyrosine cross-links extracellularly.

5.5.7 A role for immunity and signalling in shell formation

Lstag-sfc-18 contains two Ig superfamily domains (Additional file 5.19; [90]) and displays sequence similarity with the IMP-L2 like proteins (Additional files 5.40 and 5.41), an insulin-like growth factor binding protein (IGF-BP) that carries two immunoglobulin-like domains and is able to bind insulin-like growth factors (IGF) [91]. Several studies by Dogterom and colleagues demonstrated the influence of a growth hormone secreted by the cerebral ganglia specifically on shell formation in *L. stagnalis* [92–95]. The authors conclude that this growth hormone acts on cells in the belt region to control shell extension and periostracum formation, but not on shell thickening. Interestingly Perlustrin, a protein associated with nacre in abalone shells, contains an IGF-BP domain and was also shown to bind IGFs and insulin [96]. An intriguing idea for the presence of IGF-BP in the abalone shell is that it would allow the shell to signal to the underlying mantle epithelium. According to this hypothesis, IGFs present in the extrapallial fluid are bound by IGF-BP during calcification and incorporated into the shell. Should the shell dissolve or be locally damaged, these IGFs would be released and subsequently stimulate the underlying mantle epithelium to re-calcify. One line of evidence that strongly supports this hypothesis is provided by the osteogenic activity of mollusc shells [97]. This hypothesis implies that although the shell is acellular, it is able to actively communicate and provide real-time feedback to the mantle epithelium [98]. This interesting hypothesis awaits the development of gene specific knock-down or knock-out assays.

5.5.8 RLCDs are an abundant feature of *L. stagnalis* shell proteins

Proteins containing repetitive low-complexity domains (RLCDs) are a prominent feature of molluscan shell-forming proteomes [15, 99, 100], and *L. stagnalis* is no

exception. More than half of the *L. stagnalis* shell-forming candidates we identified possess RLCDs. Proteins containing these domains were present in the belt and the low columnar outer epithelium of the adult mantle and in a wide variety of patterns of the larval stages we investigated. The motif complexity, motif length and number of motif repeats can vary greatly, from stretches consisting of a single amino acid (for example Additional file 5.24), to motifs that exceed ten amino acids (for example Additional files 5.3, 5.4, 5.29 and 5.34). In some cases, almost the whole protein is composed of RLCDs (Additional files 5.7-5.10). Repeated motifs are a common feature of structural proteins such as collagens, keratins, silk and cell wall proteins, as well as structural modules in functional proteins such as receptors, histones, ion channels and transcription factors [101, 102]. RLCDs are often part of intrinsically unstructured regions that lack a fixed or ordered three-dimensional structure [101]. In some cases, these regions define the functionality of the protein. As a general rule, unstructured proteins interact readily with other proteins [103], and the highly repetitive, modular and biased amino acid compositions can confer strength and elasticity [104]. It will be extremely informative to selectively remove RLCDs from shell-forming proteins and to study the resulting shell phenotypes once genome modification tools become broadly available to molluscs.

5.5.9 Broad sequence similarity comparisons of metazoan biomineralizing proteomes

The crossed-lamellar microstructure is fabricated by phylogenetically diverse molluscan taxa and is by far the most commonly employed shell design of the Conchifera [21, 22, 28]. While much attention has been dedicated to the characterisation of nacre-forming bivalve shell proteomes, technical advances in nucleic acid sequencing and proteome-scale surveys has seen a rapid growth in the number and diversity of molluscan shell-forming proteomes, and allows broad comparisons of these datasets to be performed. These comparisons can provide insight into the degree of evolutionary conservation that exists across shell-forming proteomes [50]. In general, molluscan shell-forming proteomes are markedly different, with some deeply conserved elements such as alkaline phosphatases, peroxidases and carbonic anhydrases [28, 57, 105, 106]. The

significant diversity of molluscan shell ultrastructures, crystal textures, colours, and materials properties therefore cannot be explained by the use of the same genes in different ways. Rather, each lineage has uniquely evolved a large fraction of its shell-forming proteome [14–16, 100, 107]. To expand on this comparative theme we collected a total of 879 biomineralising proteins validated by proteomics from a total of ten molluscs, one brachiopod, one sea urchin and one coral, and performed sequence similarity comparisons against our *L. stagnalis* dataset. Two of the ten molluscs, *Cepaea nemoralis* and *Mya truncata* [51, 108], construct shells that contain crossed lamellar texture. Interestingly our comparative analyses show that *L. stagnalis* and *M. truncata* have only three proteins that share relatively low degrees of sequence similarity, while *L. stagnalis* and *C. nemoralis* share seventeen proteins (some of these with very high degrees of sequence similarity), the highest extent of similarity between all species surveyed (Fig. 5.4). Both *L. stagnalis* and *C. nemoralis* inhabit non-marine environments, and the similarities in their shell proteomes may either be a manifestation of this, and/or a reflection of their crossed lamellar shells. The similarity of their shell protein content may also reflect the relatively recent divergence time (Meso-Cenozoic) of these two clades (Stylommatophoran, *i.e. C. nemoralis* vs. hygrophilid, *i.e., L. stagnalis*) within the monophyletic order of pulmonate gastropods, in comparison to the other species. One of the most striking observations we made in these comparisons was that almost all *L. stagnalis* shell-forming candidates expressed in zone 5 share sequence similarity with *C. nemoralis*. Conversely, *L. stagnalis* shell-forming candidates expressed asymmetrically on the right side in larvae were not present in any other biomineralising proteomes.

Some *L. stagnalis* shell-forming proteins contain domains found in a number of the biomineralizing proteins present in the dataset we assembled, or are known to play a role in processes other than biomineralization such as the Sushi domain, the von Willebrand factor A domain, the immunoglobulin domain and the filament protein domain [14, 109]. The Pif-like protein is prevalent in both bivalve and gastropod nacreous shell proteomes, and is known to bind aragonite crystals and to regulate nacre formation [110]. However limpets, which construct crossed lamellar structures, also contain Pif in their shells [15, 110, 111]. Our results

further demonstrate that Pif is not limited to nacreous matrices, and that it is likely to be a deeply conserved element of the molluscan biomineralizing proteome.

5.5.10 Glycosylation patterns

Our preliminary analysis of the sugar moieties associated with shell-forming proteins revealed an interesting dichotomy between the ASM and AIM; the population of ASM proteins appear to be far more glycosylated than AIM proteins (Table 5.1). Whether this difference is generated by a heavily glycosylated subset of the ASM, or if it reflects a general trend of most ASM proteins being glycosylated remains unknown. We also cannot determine whether there are any spatial biases within the adult mantle tissue with regards to the location of glycosylated proteins. The high percentage of glucosamine identified in AIM and ASM suggests that chitin, or its deacetylated derivative chitosan, is present in both extracts but this hypothesis requires further testing. Despite their likely importance to the functional mechanisms of shell formation, post-translational modifications of molluscan shell-forming proteins remain relatively understudied, and we predict that research efforts in these directions would yield interesting functional insights into the mechanisms of shell fabrication.

5.6 Conclusion

By characterising the spatial expression patterns of 34 genes associated with shell-formation we have revealed patterns of asymmetry that presumably contribute to the coiled phenotype of *Lymnaea's* shell. Our broad survey of these genes in the adult mantle tissue also highlight the morphological modularity of this phylum specific organ, and allude to an explanation as to why the Mollusca have evolved so many successful shell morphologies. While gene co-option, domain shuffling and gene family expansion are mechanisms that have clearly contributed to the great diversity of molluscan shell-forming proteins, our analyses also suggest that alternative splicing acts to significantly expand the shell-forming molecular repertoire. Comparing the results of spatial gene expression surveys focused on shell-formation from a broad range of molluscan taxa will shed further light on the evolutionary story of this fascinating structure.

Availability of supporting data

All raw NGS data has been deposited with the SRA with BioSample accession numbers SAMN08117214, SAMN08117215, SAMN08709370, SAMN08709371, SAMN08709372, SAMN08709373, SAMN08709374, SAMN08709375, SAMN08709376, and SAMN08709377. Individual image files for the *in situ* hybridization gene expression patterns and the sense strand cDNA sequences used to generate the *in situ* hybridization riboprobes can be accessed from the associated GigaDB repository [45]. The mantle transcriptome assemblies are also available via GigaDB [45] (file names: C2844 CLC idba Trinity for annotation.fasta and C2845 CLC idba Trinity for annotation.fasta). All MS data have been deposited with the ProteomeXchange Consortium with the dataset identifiers PXD008547 and 10.6019/PXD008547. Other supporting data are available from additional files, also including an extended description of the in situ hybridization results (see additional file 5.47).

Additional files

Additional file 5.1. Summarized results of MASCOT searches.

Additional files 5.2–5.35. Whole mount in situ hybridisation results and molecular features of 34 shell-forming gene candidates.

Additional file 5.36. A more comprehensive summary of the results presented in Fig. 5.3.

Additional file 5.37. Detailed table of the molecular features of all shell-forming protein candidates.

Additional file 5.38. Results of repetitive motif searches using Xstream.

Additional file 5.39. A FASTA formatted file of the 879 protein se- quences used to construct Fig. 5.4.

Additional file 5.40. Detailed results of tBLASTx similarity searches for all shell-forming candidate genes against nr database.

Additional file 5.41. Detailed results of protein family and protein domain similarity searches for all shell-forming candidate genes against CDD database.

Additional file 5.42. Detailed results of BLAST similarity searches for all shell-forming candidate genes against SwissProt database.

Additional file 5.43. Lineages for all shell-forming candidates that returned positive BLAST results.

Additional file 5.44. Nucleotide sequences of 34 families of shell forming candidate genes.

Additional file 5.45. Translated sequences of 34 families of shell forming candidate genes.

Additional file 5.46. mRNA regions targeted by riboprobes.

Additional file 5.47. Extended Results and Discussion.

The additional files can be found via the attached CD of the printed version or online at https://academic.oup.com/gigascience/article/7/6/giy056/4997018. Additional files 2-35 can be fount in the appendix 1.

Abbreviation

AIM: acid-insoluble matrix; ASM: acid-soluble matrix; BLAST: Basic Local Alignment Search Tool; CNS: central nervous sys- tem; dpfc: days post first cleavage; GRN: gene regulatory net- work; hpfc: hours post first cleavage; IGF-BP: insulin-like growth factor–binding protein; LC: Liquid chromatography; MS: mass spectrometry; NGS: next-generation sequencing; PCR: polymerase chain reaction; RLCD: repetitive low-complexity do- main; SEM: Scanning Electron Microscopy; SRA: Sequence Read Archive; UTR: untranslated region.

Competing interests

The authors declare that they have no competing interests.

Author contributions

I.H. carried out the molecular work, bioinformatic analyses, and co-wrote and drafted the manuscript. F.M. and B.M. performed the proteomic analyses and drafted the manuscript. D.J.J. conceived and supervised the study, contributed to the molecular work, contributed to the bioinformatic analyses, performed the histological sections, and co-wrote and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

References

1. Mao L-B, Gao H-L, Yao H-B, Liu L, Cölfen H, Liu G, Chen S-M, Li S-K, Yan Y-X, Liu Y-Y. Synthetic nacre by predesigned matrix-directed mineralization. Science. 2016;354:107-110.
2. Kniprath E. Ontogeny of the molluscan shell field: a review. Zool Script. 1981;10:61-79.
3. Jackson DJ, Degnan BM. The importance of Evo-Devo to an integrated understanding of molluscan biomineralisation. J Struct Biol. 2016;196:67-74.
4. Okabe T, Yoshimura J. Optimal designs of mollusk shells from bivalves to snails. Sci Rep. 2017;7:42445.
5. Aguilera F, McDougall C, Degnan BM. Co-option and de novo gene evolution underlie molluscan shell diversity. Mol Biol Evol. 2017;34:779-792.
6. Liang J, Xie J, Gao J, Xu C-Q, Yan Y, Jia G-C, Xiang L, Xie L-P, Zhang R-Q. Identification and Characterization of the Lysine-Rich Matrix Protein Family in *Pinctada fucata*: Indicative of Roles in Shell Formation. Mar Biotech. 2016;18:645-658.
7. Marxen JC, Hammer M, Gehrke T, Becker W. Carbohydrates of the organic shell matrix and the shell-forming tissue of the snail *Biomphalaria glabrata* (Say). Biol Bull. 1998;194:231-240.
8. Arias JL, Fernández MS. Polysaccharides and proteoglycans in calcium carbonate-based biomineralization. Chem Rev. 2008;108:4475-4482.
9. Farre B, Dauphin Y. Lipids from the nacreous and prismatic layers of two Pteriomorpha Mollusc shells. Comp Biochem Physiol B. 2009;152:103-109.
10. Zhao M, He M, Huang X, Wang Q, Shi Y. Functional characterization and molecular mechanism exploration of three granulin epithelin precursor splice variants in biomineralization of the pearl oyster *Pinctada fucata*. Mol Genet Genom. 2016;291:399-409.
11. Wang J, Gao J, Xie J, Zheng X, Yan Y, Li S, Xie L, Zhang R. Cloning and mineralization-related functions of the calponin gene in *Chlamys farreri*. Comp Biochem Physiol B. 2016;201:53-58.
12. Yarra T, Gharbi K, Blaxter M, Peck LS, Clark MS. Characterization of the mantle transcriptome in bivalves: *Pecten maximus*, *Mytilus edulis* and *Crassostrea gigas*. Marine Genomics. 2016;27:9-15.
13. Gao P, Liao Z, Wang X-x, Bao L-f, Fan M-h, Li X-m, Wu C-w, Xia S-w. Layer-by-

layer proteomic analysis of *Mytilus galloprovincialis* shell. PloS ONE. 2015;10:e0133913.

14. Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. BMC Genom. 2014;15:249.

15. Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cleon I, Guichard N, Marin F. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. FEBS J. 2013;280:214-232.

16. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM. A rapidly evolving secretome builds and patterns a sea shell. BMC Biol. 2006;4:40.

17. Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C, Marin F. Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. Prot Sci. 2010;8:54.

18. Zhao M, He M, Huang X, Wang Q. A homeodomain transcription factor gene, *PfMSX*, activates expression of *Pif* gene in the pearl oyster *Pinctada fucata*. PloS ONE. 2014;9:e103830.

19. Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science. 2009;325:1388-1390.

20. Jackson DJ, Wörheide G, Degnan BM. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. BMC Evol Biol. 2007;7:160.

21. Dauphin Y, Denis A. Structure and composition of the aragonitic crossed lamellar layers in six species of Bivalvia and Gastropoda. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology. 2000;126:367-377.

22. de Paula SM, Silveira M. Studies on molluscan shells: contributions from microscopic and analytical methods. Micron. 2009;40:669-690.

23. Almagro I, Drzymała P, Berent K, Sainz-Díaz CI, Willinger M-G, Bonarski J, Checa AG. New crystallographic relationships in biogenic aragonite: the crossed-lamellar microstructures of mollusks. Crystal Growth & Design. 2016

24. Kuhn-Spearing LT, Kessler H, Chateau E, Ballarini R, Heuer AH, Spearing SM. Fracture mechanisms of the *Strombus gigas* conch shell: implications for the design of brittle laminates. Journal of Materials Science. 1996;31:6583-6594.

25. Kamat S, Su X, Ballarini R, Heuer AH. Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. Nature. 2000;405:1036-1040.

26. Pokroy B, Zolotoyabko E. Microstructure of natural plywood-like ceramics: a study by high-resolution electron microscopy and energy-variable X-ray diffraction. Journal of Materials Chemistry. 2003;13:682-688.

27. Rodriguez-Navarro AB, Checa A, Willinger M-G, Bolmaro R, Bonarski J. Crystallographic relationships in the crossed lamellar microstructure of the shell of the gastropod *Conus marmoreus*. Acta Biomaterialia. 2012;8:830-835.

28. Marin F, Luquet G, Marie B, Medakovic D. Molluscan shell proteins: primary structure, origin, and evolution. Curr Top Dev Biol. 2008;80:209-276.

29. Pavat C, Zanella-Cléon I, Becchi M, Medakovic D, Luquet G, Guichard N, Alcaraz G, Dommergues J-L, Serpentini A, Lebel J-M, Marin F. The shell matrix of the pulmonate land snail *Helix aspersa maxima*. Comp Biochem Physiol B.

2012;161:303-314.

30. Vizcaíno JA, Csordas A, Del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T. 2016 update of the PRIDE database and its related tools. Nucleic Acids REesearch. 2016;44:D447-D456.

31. Jackson DJ, Ellemor N, Degnan BM. Correlating gene expression with larval competence, and the effect of age and parentage on metamorphosis in the tropical abalone *Haliotis asinina*. Mar Biol. 2005;147:681-697.

32. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nuc Acids Res. 2003;31:3784-3788.

33. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth. 2011;8:785-786.

34. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A: Protein identification and analysis tools on the ExPASy server. Springer; 2005.

35. Jorda J, Kajava AV. T-REKS: identification of Tandem REpeats in sequences with a K-meanS based algorithm. Bioinformatics. 2009;25:2632-2638.

36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403-410.

37. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI. CDD: NCBI's conserved domain database. Nuc Acids Res. 2014gku1221.

38. Mitchell A, Chang H-Y, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S. The InterPro protein families database: the classification resource after 15 years. Nuc Acids Res. 2014gku1243.

39. Steentoft C, Vakhrushev SY, Joshi HJ, Kong Y, Vester-Christensen MB, Katrine T, Schjoldager BG, Lavrsen K, Dabelsteen S, Pedersen NB. Precision mapping of the human *O*-GalNAc glycoproteome through SimpleCell technology. The EMBO journal. 2013;32:1478-1488.

40. Gene Structure Draw. [http://www.compgen.uni-muenster.de/tools/strdraw/index.hbi?] [http://www.compgen.uni-muenster.de/tools/strdraw/index.hbi?]

41. Davison A, McDowell GS, Holden JM, Johnson HF, Koutsovoulos GD, Liu MM, Hulpiau P, Van Roy F, Wade CM, Banerjee R. Formin is associated with left-right asymmetry in the pond snail and the frog. Curr Biol. 2016;26:654-660.

42. Kapustin Y, Tőzsér J, Souvorov A, Tatusova T, Lipman D. Splign: algorithms for computing spliced alignments with identification of paralogs. Biology Direct. 2008;3:20.

43. Yan T, Yoo D, Berardini TZ, Mueller LA, Weems DC, Weng S, Cherry JM, Rhee SY. PatMatch: a program for finding patterns in peptide and nucleotide sequences. Nuc Acids Res. 2005;33:W262-W266.

44. Cerveau N, Jackson DJ. Combining independent de novo assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms. BMC Bioinformatics. 2016;17:525.

45. Herlitze I, Marie B, Marin F, Jackson DJ. Supporting data for "Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas". GigaScience Database. 2018

46. Jackson DJ, Herlitze I, Hohagen J. A whole mount *in situ* hybridization method for the gastropod mollusc *Lymnaea stagnalis*. JoVE. 2016e53968.

47. Marie B, Joubert C, Tayaléa A, Zanella-Cléon I, Belliard C, Piquemal D,

Cochennec-Laureau N, Marin F, Gueguen Y, Montagnani C. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences. 2012;109:20986-20991.

48. Karlheinz Mann, Nicolas Cerveau, Meike Gummich, Fritz M, Jackson DJ. In-depth proteomic analyses of *Haliotis laevigata* (greenlip abalone) nacre and prismatic organic shell matrix. Submitted.

49. Mann K, Edsinger E. The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. Prot Sci. 2014;12:28.

50. Feng D, Li Q, Yu H, Kong L, Du S. Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: characterization of genetic bases regulating shell formation. Sci Rep. 2017;7:45754.

51. Arivalagan J, Marie B, Sleight VA, Clark MS, Berland S, Marie A. Shell matrix proteins of the clam, *Mya truncata*: Roles beyond shell formation through proteomic study. Marine Genomics. 2016;27:69-74.

52. Liu C, Li S, Kong J, Liu Y, Wang T, Xie L, Zhang R. In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. Sci Rep. 2015;5:17269.

53. Liao Z, Bao L-f, Fan M-h, Gao P, Wang X-x, Qin C-l, Li X-m. In-depth proteomic analysis of nacre, prism, and myostracum of Mytilus shell. Journal of Proteomics. 2015;122:26-40.

54. Jackson DJ, Mann K, Häussermann V, Schilhabel MB, Lüter C, Griesshaber E, Schmahl W, Wörheide G. The *Magellania venosa* biomineralizing proteome: a window into brachiopod shell evolution. Gen Biol Evol. 2015;7:1349-1362.

55. Mann K, Poustka AJ, Mann M. In-depth, high-accuracy proteomics of sea urchin tooth organic matrix. Prot Sci. 2008;6:33.

56. Ramos-Silva P, Kaandorp J, Herbst F, Plasseraud L, Alcaraz G, Stern C, Corneillat M, Guichard N, Durlet C, Luquet G, Marin F. The Skeleton of the Staghorn Coral *Acropora millepora*: Molecular and Structural Characterization. PLoS ONE. 2014;9:e97454.

57. Hohagen J, Jackson DJ. An ancient process in a modern mollusc: early development of the shell in *Lymnaea stagnalis*. BMC Dev Biol. 2013;13:27.

58. Kniprath E. Zur Ontogenese des Schalenfeldes von *Lymnaea stagnalis*. Wilhelm Roux's Archives of Developmental Biology. 1977;181:11-30.

59. Timmermans LPM. Studies on shell formation in molluscs. Neth J Zool. 1969;19:413-523.

60. Saleuddin ASM. An electron microscopic study on the formation of the periostracum in *Helisoma* (Mollusca). Calcified Tissue International. 1975;18:297-310.

61. Kniprath E. Formation and structure of the periostracum in *Lymnaea stagnalis*. Cal Tiss Res. 1972;9:260-271.

62. Bubel A. An electron-microscope study of periostracum formation in some marine bivalves. I. The origin of the periostracum. Mar Biol. 1973;20:213-221.

63. Bevelander G, Nakahara H. An electron microscope study of the formation of the periostracum of *Macrocallista maculata*. Cal Tiss Res. 1967;1:55-67.

64. Bubel A. An electron-microscope study of periostracum formation in some marine bivalves. II. The cells lining the periostracal groove. Mar Biol. 1973;20:222-234.

65. Ast G. How did alternative splicing evolve. Nat Rev Gen. 2004;5:773-782.
66. Weiner S, Traub W, Parker SB. Macromolecules in mollusc shells and their functions in biomineralization [and Discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences. 1984;304:425-434.
67. Addadi L, Weiner S. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. Proceedings of the National Academy of Sciences. 1985;82:4110-4114.
68. Kingsley EP, Chan XY, Duan Y, Lambert JD. Widespread RNA segregation in a spiralian embryo. Evol Dev. 2007;9:527-539.
69. Lambert JD, Nagy LM. Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages. Nature. 2002;420:682-686.
70. Bartel DP, Chen C-Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Gen. 2004;5:396-400.
71. Zheng Z, Jiao Y, Du X, Tian Q, Wang Q, Huang R, Deng Y. Computational prediction of candidate miRNAs and their potential functions in biomineralization in pearl oyster *Pinctada martensii*. Saudi Journal of Biological Sciences. 2016;23:372-378.
72. Jiao Y, Zheng Z, Tian R, Du X, Wang Q, Huang R. MicroRNA, pm-miR-2305, participates in nacre formation by targeting pearlin in pearl oyster *Pinctada martensii*. Int J Molec Sci. 2015;16:21442-21453.
73. Brusca RC, Brusca GJ: Invertebrates. 2nd edition. Sinauer Associates Sunderland, Massachusetts; 2002.
74. Rosenberg G. A new critical estimate of named species-level diversity of the recent Mollusca. Am Malacol Bull. 2014;32:308-322.
75. Hochner B, Glanzman DL. Evolution of highly diverse forms of behavior in molluscs. Curr Biol. 2016;26:R965-R971.
76. Williams TA, Nagy LM. Developmental modularity and the evolutionary diversification of arthropod limbs. Journal of Experimental Zoology Part A: Ecological Genetics and Physiology. 2001;291:241-257.
77. Peel AD, Chipman AD, Akam M. Arthropod segmentation: beyond the *Drosophila* paradigm. Nat Rev Gen. 2005;6:905-916.
78. Tautz D. Segmentation. Developmental Cell. 2004;7:301-312.
79. Raff RA: The shape of life: genes, development, and the evolution of animal form. University of Chicago Press; 1996.
80. Esteve-Altava B. In search of morphological modules: a systematic review. Biol Rev. 2017;92:1332-1347.
81. Eble GJ. Morphological modularity and macroevolution. In: Callebaut W, Rasskin-Gutman D, editors. Modularity: understanding the development and evolution of natural complex systems. MIT Press, Cambridge; 2005. p. 221-238.
82. McDougall C, Green K, Jackson DJ, Degnan BM. Ultrastructure of the mantle of the gastropod *Haliotis asinina* and mechanisms of shell regionalization. Cells Tissues Organs. 2011;194:103-107.
83. Jolly C, Berland S, Milet C, Borzeix S, Lopez E, Doumenc D. Zonal localization of shell matrix proteins in mantle of *Haliotis tuberculata* (Mollusca, Gastropoda). Mar Biotech. 2004;6:541-551.
84. Sleight VA, Marie B, Jackson DJ, Dyrynda EA, Marie A, Clark MS. An Antarctic molluscan biomineralisation tool-kit. Sci Rep. 2016;6:36978.
85. Waite JH. The phylogeny and chemical diversity of quinone-tanned glues and

varnishes. Comp Biochem Physiol B. 1990;97:19-29.

86. Waite JH. Evidence for the mode of sclerotization in a molluscan periostracum. Comp Biochem Physiol B. 1977;58:157-162.

87. Waite JH, Wilbur KM. Phenoloxidase in the periostracum of the marine bivalve *Modiolus demissus* Dillwyn. J Exp Zool. 1976;195:359-367.

88. Waite JH. Quinone-tanned scleroproteins. In: Saleuddin ASM, Wilbur KM, editors. Physiology. New York: Academic Press; 1983. p. 467-504.

89. Michon T, Chenu M, Kellershon N, Desmadril M, Guéguen J. Horseradish peroxidase oxidation of tyrosine-containing peptides and their subsequent polymerization: a kinetic study. Biochem. 1997;36:8504-8513.

90. Halaby DM, Mornon JPE. The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. Journal of Molecular Evolution. 1998;46:389-400.

91. Honegger B, Galic M, Köhler K, Wittwer F, Brogiolo W, Hafen E, Stocker H. Imp-L2, a putative homolog of vertebrate IGF-binding protein 7, counteracts insulin signaling in *Drosophila* and is essential for starvation resistance. Journal of biology. 2008;7:1.

92. Dogterom AA, Doderer A. A hormone dependent calcium-binding protein in the mantle edge of the freshwater snail *Lymnaea stagnalis*. Calcified Tissue International. 1981;33:505-508.

93. Dogterom AA, Jentjens T. The effect of the growth hormone of the pond snail *Lymnaea stagnalis* on periostracum formation. Comparative Biochemistry and Physiology Part A: Physiology. 1980;66:687-690.

94. Dogterom AA, van Loenhout H, van der Schors RC. The effect of the growth hormone of *Lymnaea stagnalis* on shell calcification. Gen Comp Endocrinol. 1979;39:63-68.

95. Dogterom AA, van der Schors RC. The effect of the growth hormone of *Lymnaea stagnalis* on (bi) carbonate movements, especially with regard to shell formation. Gen Comp Endocrinol. 1980;41:334-339.

96. Weiss IM, Göhring W, Fritz M, Mann K. Perlustrin, a *Haliotis laevigata (*abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates. Biochem Biophys Res Commun. 2001;285:244-249.

97. Zhang G, Willemin AS, Brion A, Piet M-H, Moby V, Bianchi A, Mainard D, Galois L, Gillet P, Rousseau M. A new method for the separation and purification of the osteogenic compounds of nacre Ethanol Soluble Matrix. J Struct Biol. 2016;196:127-137.

98. Marin F, Luquet G. Molluscan shell proteins. Comptes Rendus Palevol. 2004;3:469-492.

99. Shen X, Belcher AM, Hansma PK, Stucky GD, Morse DE. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. J Biol Chem. 1997;272:32472-32481.

100. Kocot KM, Aguilera F, McDougall C, Jackson DJ, Degnan BM. Sea shell diversity and rapidly evolving secretomes: insights into the evolution of biomineralization. Front Zool. 2016;13:23.

101. Luo H, Nijveen H. Understanding and identifying amino acid repeats. Briefings in Bioinformatics. 2014;15:582-591.

102. Albà M, Tompa P, Veitia R. Amino acid repeats and the structure and evolution of proteins. In: editors. Gene and Protein Evolution. Karger

Publishers; 2007. p. 119-130.

103. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: Molecular Biology of the Cell. 5th revised edition. New York: Garland Science; 2007.

104. McDougall C, Woodcroft BJ, Degnan BM. The widespread prevalence and functional significance of silk-like structural proteins in metazoan biological materials. PLoS ONE. 2016;11:e0159128.

105. Le Roy N, Jackson D, Marie B, Ramos-Silva P, Marin F. Carbonic anhydrase and metazoan biocalcification: a focus on molluscs. In: Marin F, Brümmer F, Checa A, Furtos G, Lesci IG, Šiller L, editors. Biomineralization: from fundamentals to biomaterials & environmental issues. Trans Tech Publications Ltd; 2015. p. 151-157.

106. Sun X, Yang A, Wu B, Zhou L, Liu Z. Characterization of the mantle transcriptome of Yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. PloS one. 2015;10:e0122967.

107. Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C. Parallel evolution of nacre building gene sets in molluscs. Mol Biol Evol. 2010;27:591-608.

108. Sleight VA, Thorne MAS, Peck LS, Arivalagan J, Berland S, Marie A, Clark MS. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. Marine Genomics. 2016;27:47-55.

109. Arivalagan J, Yarra T, Marie B, Sleight VA, Duvernois-Berthet E, Clark MS, Marie A, Berland S. Insights from the shell proteome: biomineralization to adaptation. Mol Biol Evol. 2016msw219.

110. Suzuki M, Kogure T, Weiner S, Addadi L. Formation of aragonite crystals in the crossed lamellar microstructure of limpet shells. Crystal Growth & Design. 2011;11:4850-4859.

111. Mann K, Edsinger-Gonzales E, Mann M. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. Prot Sci. 2012;10:28.

6

Expression and purification of the *Lymnaea stagnalis* shell-forming candidate 10 in *E. coli*

6.1 Introduction

A great number of genes and gene products involved in shell formation has been identified due to next-generation sequencing techniques and high-through-put proteomics [1-10]. The resulting protein catalogues are valuable resources for broad comparisons analyses between different shell secretomes and are useful when questions relating the evolutionary conservation or innovation of shell proteins are addressed [1, 9, 11]. Besides that, it is still challenging to assign a function to each of these proteins. To shed light on the role of these components, *in vitro* approaches such as calcification assays with native [12-16] or recombinant proteins [17-21] can be applied. With these assays, the effects of a shell protein fraction, individual proteins or specific protein domains on calcium carbonate crystallisation can be studied. The production of recombinant proteins in a heterologous expression host such as *Escherichia coli* has the advantage that large quantities can be produced at relatively low costs. This is beneficial if the protein derives from a species that is rare or hard to access. Furthermore, it is possible to produce single domains of a protein that can be studied independently in respect to their effect on mineralization. For the production of a protein in a heterologous expression host, the general approach involves the acquisition of cDNA, the engineering of an expression vector, the cloning into competent cells, the expression of the protein and subsequent purification of the target protein. The aim is to produce a soluble protein that reflects the native condition in the animal. However, the expression and purification of a heterologous protein in *E. coli* can be tricky. The behaviour of each protein is different and there is no protocol that works for all. Optimisation experiments regarding vector design, choice of expression host and culture condition have to be conducted in order to obtain the purified protein. Only 10% of eukaryotic proteins can be expressed in *E. coli* in the soluble form [22]. Proteins that are not expressed in the soluble form are likely not modified or folded properly and can precipitate in the cell through the formation

of inclusion bodies [22]. Unfortunately, this is common if the protein is produced in large quantities [23]. This chapter summarises the on-going project of the expression and purification of the *Lymnaea stagnalis* shell-forming candidate 10 (Lstag-sfc-10). In order to express and purify this protein in the soluble form, a number of optimisation experiments were necessary.
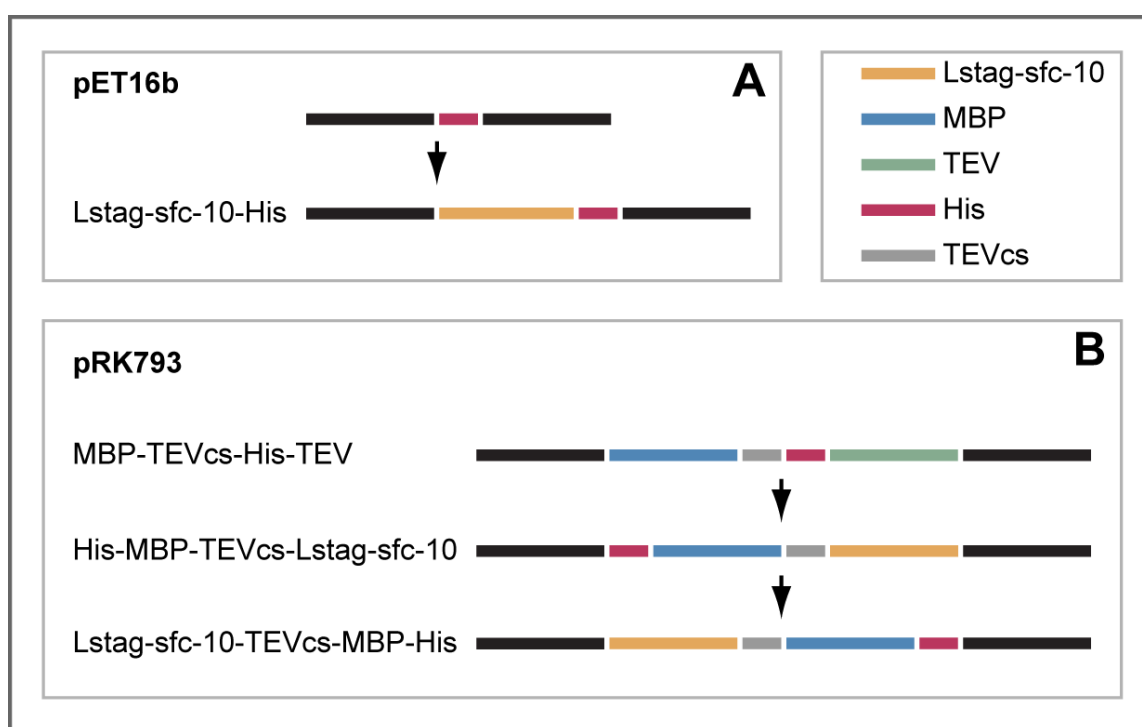

Fig. 6.1. Schematic representation of the expression constructs. (A) Schematic representation of the expression construct composed of the Lstag-sfc-10 fragment ligated into the pET16b expression vector. **(B)** Schematic representation of the expression construct that was engineered from a modified pRK793 and the Lstag-sfc-10 fragment. Both constructs encode MBP-Lstag-sfc-10 fusion proteins, but vary in the arrangement of the MBP, Lstag-sfc-10, TEV cleavage site and His-tag. His: oligohistidine tag, MBP: maltose binding protein, TEV: TEV protease, TEVcs: cleavage site targeted by TEV protease.

6.2 Methods

6.2.1 Cultivation of adult *L. stagnalis*

Adult *L. stagnalis* were cultured as described in chapter 5.

6.2.2 Gene isolation

RNA was extracted from the adult mantle with TRI reagent according to the manufactures instruction including an additional precipitation step with high salt solution (0.8 M sodium citrate, 1.2 M NaCl). A cDNA library was constructed using 1 µg total RNA, 2 µM oligo dT primer (20mer) and water in a volume of 10 µL and heated to 70 °C for 5 minutes (min). After cooling to 42 °C, 15 µL reaction mix comprising 1X reverse transcription buffer, 0.5 mM dNTP and 200 U MMLV-reverse transcriptase H- (Promega, #M3682) were added and incubated at 42 °C for 90 min. The reaction was stopped by heating to 70 °C for 15 min.

6.2.3 Expression construct preparation

pET16b+Lstag-sfc-10-His

Using mantle cDNA as template and gene specific primers Lst_c1222_OE_F1 (see supplementary Table S. 6.1 for all primer sequences) and Lst_c1222_OE_R1 that introduced NcoI restriction sites, a fragment of *Lstag-sfc-10* that spans the open reading frame (ORF) excluding the signal sequence, was isolated. 44 µL of the PCR product were incubated with CutSmart Buffer (NEB, #B7204S) and 1 µL NcoI-HF (NEB, #R3193S) for 1 hour (h) at 37 °C. 1.5 µL oft the NcoI digested PCR product was incubated with 0.5 µL of NcoI digested and dephosphorylated pET16b (Addgene, #69662-3), 2.5 µL of 2x ligation buffer and 0.5 µL T4 DNA ligase (Promega, #M1801). For plasmid proliferation, pET16b+Lstag-sfc-10-His was cloned into DH 10B. Plasmids were isolated using the High Pure Plasmid Isolation Kit (Roche, #11754777001). The correct fragment insertion was verified with Sanger sequencing. A schematic of pET16b+Lstag-sfc-10-His is presented in Fig. 6.1A.

Modified pRK793+ Lstag-sfc-10

A modified pRK793 producing a MBP-TEVcs-His-Lstag-sfc-10 fusion protein and lacking the TEV protease domain was prepared via a PCR using the pRK793 plasmid (Addgene, #8827) as template and gene specific primers pRK793_Gib_F1 and pRK793_Gib_R1. To obtain the Lstag-sfc-10 insert with overhang sequence that allows Gibson assembly into the modified vector, a PCR was prepared using pET16b+Lstag-sfc-10-His as template and gene specific primers pRK793_Gib_c1222_F1 and pRK793_Gib_c1222_R1. Gibson assembly with the modified vector and the Lstag-sfc-10 insert was performed according to manufactures instructions (NEB, #E2611S)[25]. This expression construct produces a MBP-TEVcs-His-Lstag-sfc-10 fusion protein that is suitable for TEV cleavage but cannot be purified from the likewise his tagged TEV protease. To engineer a vector producing a His-MBP-TEVcs-Lstag-sfc-10 fusion protein suitable for TEV cleavage and subsequent protein purification, further vector modifications were performed.

The modified pRK793 vector backbone containing the Lstag-sfc-10 fragment and lacking the TEV protease, the MBP, the His-tag and the TEVcs was prepared via a PCR using the modified pRK793 MBP-TEVcs-His-Lstag-sfc-10 as template and gene specific primers 793+c1222_Gib_F1 and 793+c1222_Gib_R1. To obtain the MBP insert with flanking His-tag and TEVcs, a PCR was prepared using the modified pRK793 MBP-TEVcs-His-Lstag-sfc-10 as template and gene specific primers His-MBP-TEVcl_F and His-MBP-TEVcl_R. To introduce overhang sequence that allows Gibson assembly, a PCR was prepared using the previous PCR product as template and gene specific primers Gib_His-MBP_F1 and Gib_His-MBP_R1. Gibson assembly with the modified vector backbone and MBP insert was performed according to manufactures instructions (NEB, #E2611S)[25]. For plasmid proliferation, the modified pRK793His-MBP-TEVcs-Lstag-sfc-10 was cloned into NEB 5-alpha (NEB, #C2987I). Plasmids were isolated using the High Pure Plasmid Isolation Kit (Roche, #11754777001). The correct fragment insertion was verified with Sanger sequencing. A schematic of the modified pRK793 His-MBP-TEVcs-Lstag-sfc-10 is presented in Fig. 6.1B.

As positive control served the expression vector containing a chitin-binding domain-GFP (CBD-GFP) [26] that was a kind gift of Prof. Dr. Ingrid Weiss (University of Stuttgart, Germany).

6.2.4 Prediction of protein molecular weight

The molecular weight of the recombinant proteins were predicted with the ExPasyProtParam tool [27, 28]

6.2.5 Protein production

For protein production, the expression vectors were cloned into a variety of *E. coli* cell strains. Expression pET16b+Lstag-sfc-10-His was cloned into JM 09 (DE3), Origami B (DE3) and BL 21 (DE3). The modified pRK793+His-MBP-TEVcs-Lstag-sfc-10 was cloned into JM 09 (DE3), Rosetta 2 (DE3), BL 21 (DE3), BL 21 Codon Plus (DE3) RIPL and BL 21 Gold (DE3) pLysS. ZYM5052 or MTBau medium [29] including the appropriate antibiotics were inoculated with transformed cells and incubated overnight at 37 °C at 750 rpm. Optimisation experiments were performed with strains that expressed the protein most efficiently at 37 °C, 30 °C and 25 °C at 750 rpm.

6.2.6 Cell culture preparation for SDS-PAGE and Coomassie staining

To prepare the soluble and insoluble cell fractions of the expression hosts, cultures were centrifuged at 1000 relative centrifugal force (rcf) at 4 °C for 5 min. The supernatant was removed and cell pellets were stored at -80 °C. Pellets were resuspended in RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 1% NP-40, 1% sodium deoxycholate) and sonicated in 3 x 20 seconds pulses at 130 W and 20% amplitude. Then, samples were centrifuged at maximum speed at 4 °C for >20 min. The supernatant contained the soluble proteins and was separated from the pellet. The pellet contained the insoluble proteins and was resuspended in RIPA buffer. 4x loading dye (40% glycerol, 8% SDS, 1.75 mM Tris pH 6.8, 0.1 M dithiothreitol (DTT), 0.01% bromphenol blue) was added to the soluble and insoluble fractions and heat denatured at 85 °C for 10 min. For SDS polyacrylamide gel electrophoresis (PAGE), a 12% or 10% acrylamide gel (Rotiphorese Gel 30 (Roth #3029.3), 0.37 M Tris-HCl pH 8.8, 0.1%

SDS (Roth #3060.2), 0.08 % TEMED (Roth #2367.3), 0.05% ammoniumperoxodisulfat (Roth #9592.3), H_2O) was prepared, depending on the size oft the produced protein. After electrophoresis, the gel was stained with Coomassie stain (40% methanol, 10% acetic acid, 20% Coomassie staining solution (Roth #A152.2), 30% H_2O) and de-stained over night with 25% methanol in H_2O. The gel was photo documented with a standard image scanner.

6.2.7 Protein purification

His-tagged proteins were purified with the Ni-NTA Spin Kit (Qiagen, #31314). Proteins that were present in the insoluble fraction of the cell lysate were purified under denaturing conditions. Thawed cells were resuspendet in 700 µL Buffer B (7 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl pH 8.0) and sonicated in 3 x 20 seconds pulses at 130 W and 20% amplitude. The lysate was centrifuged at 12000 x g for 15-30 min. The supernatant was applied onto a pre-equilibrated Ni-NTA spin column and centrifuged at 270 x g until all liquid passed the column. The column was washed twice with 600 µL Buffer C (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-Cl pH 6.3) at 890 x g for 2 min each. The protein was eluted twice in 150 µL buffer E (8 M urea, 0.1 M NaH_2PO_4, 0.01 M Tris-HCl pH 4.5) with a 2 min spin at 890 x g. Fractions of each step were taken for SDS PAGE. Electrophoresis was carried out as described in the previous section.

Proteins that were present in the soluble fraction of the cell lysate were purified under native conditions. Bacterial pellets from a 5 mL cell culture were resuspendet in 630 µL lysis buffer NPI-10 (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated in 3 x 20 seconds pulses at 130 W and 20 % amplitude. The lysate was centrifuged at 12000 x g for 35 min. The cleared lysate was applied onto a pre- equilibrated NiNTA spin column and centrifuged for 5 min at 270 x g. The column was washed twice with 600 µL wash buffer NPI-20 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) by centrifuging for 2 min at 890 x g. The protein was eluted twice with 150 µL elution buffer NPI-500 (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0) by centrifuging for 2 min at 890 x g. Fractions of each step were taken for SDS PAGE. Electrophoresis was carried out as described in the previous section.

6.3 Results and discussion

By combining proteomic analyses, transcriptome sequencing and *in situ* hybridisation, more than 30 candidate genes associated with the production of the crossed-lamellar *L. stagnalis* shell were identified (see chapter 5). *Lstag-sfc-10* is one of these candidate genes and encodes a secreted protein that has a glycine-rich repeated motif and an anomalous glycine content of 20%. The shell-forming candidate *Lstag-sfc-10* is expressed in a symmetrical ring in the cells bordering the shell gland in trochophore larvae and along the entire shell field margin in veliger larvae. In adults, the gene displays a distinct and clear signal in zone three of the mantle belt (see chapter 5). These expression patterns suggest that Lstag-sfc-10 plays a vital role in shell-formation throughout *L. stagnalis* development. However, the concrete function of the candidate remains elusive.

6.3.1 Lstag-sfc-10 is expressed in the insoluble cell fraction

In the first attempt, Lstag-sfc-10 (containing the open reading frame except the signal sequence) was ligated into the expression vector pET16b, which supplies a His-tag (Fig. 6.1A). Creating a His-tagged recombinant protein is necessary, when the construct is purified with immobilized metal affinity chromatography (IMAC) [31]. His-tags are relatively small and do not effect the folding of the attached protein or alter its solubility [22, 23]. The predicted 36.64 kDa Lstag-sfc-10-His (S. 6.2) was successfully expressed in the insoluble protein fraction of the *E. coli* strain BL21 (DE3) at a growing temperature of 37 °C (Fig. 6.2A). JM 09 (DE3) and Origami B (DE3) did not express Lstag-sfc-10-His (data not shown).

A common strategy to force an insoluble target protein into the soluble cell fraction is by decreasing the growing temperature. This change of culturing condition increases the stability and correct folding patterns of the recombinant protein [23]. At a growing temperature of 30 °C, less target protein was present in the insoluble cell fraction and no target protein was present in the soluble cell fraction (Fig. 6.2A). At a growing temperature of 25 °C, no target protein was present in any cell fraction (Fig. 6.2A). The insoluble target protein could be purified with an IMAC approach under denaturing conditions (Fig. 6.2B).
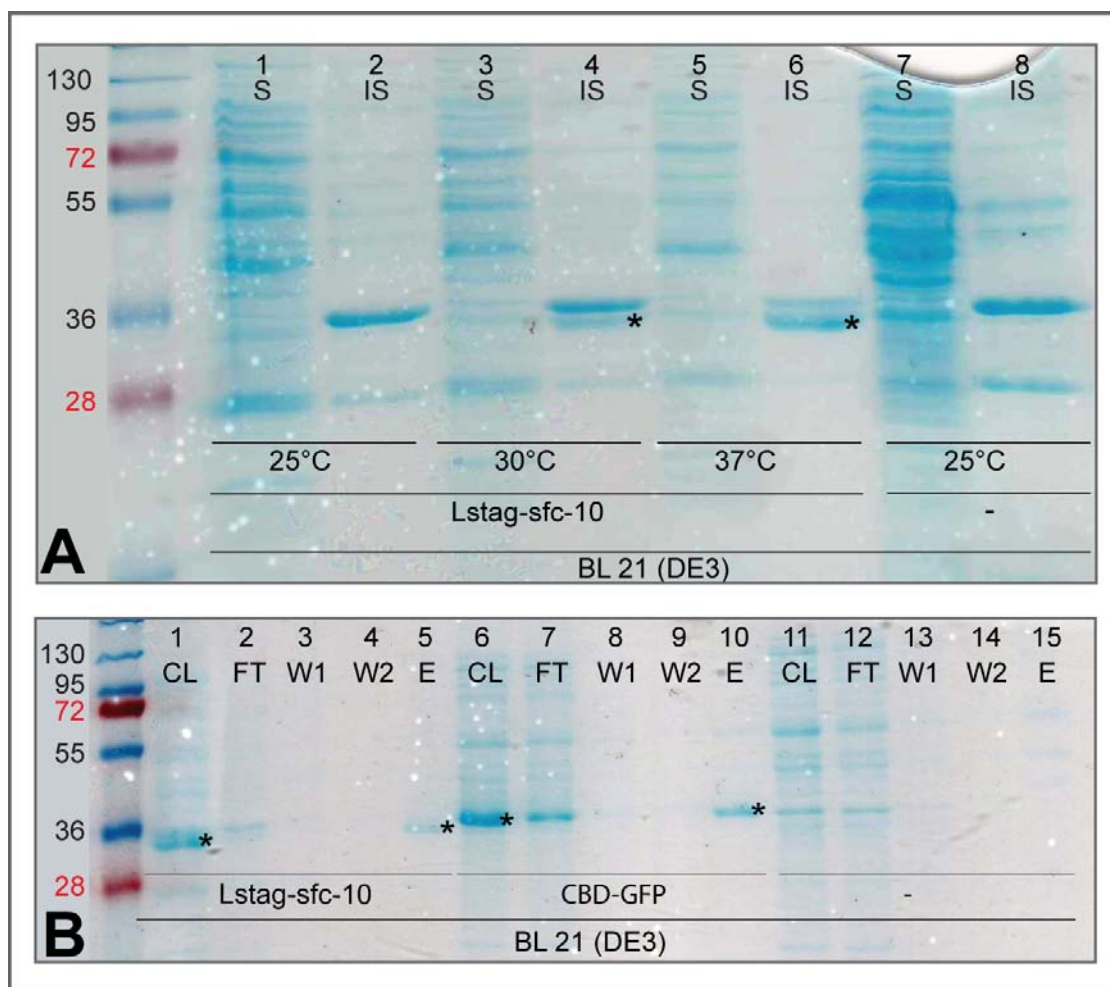
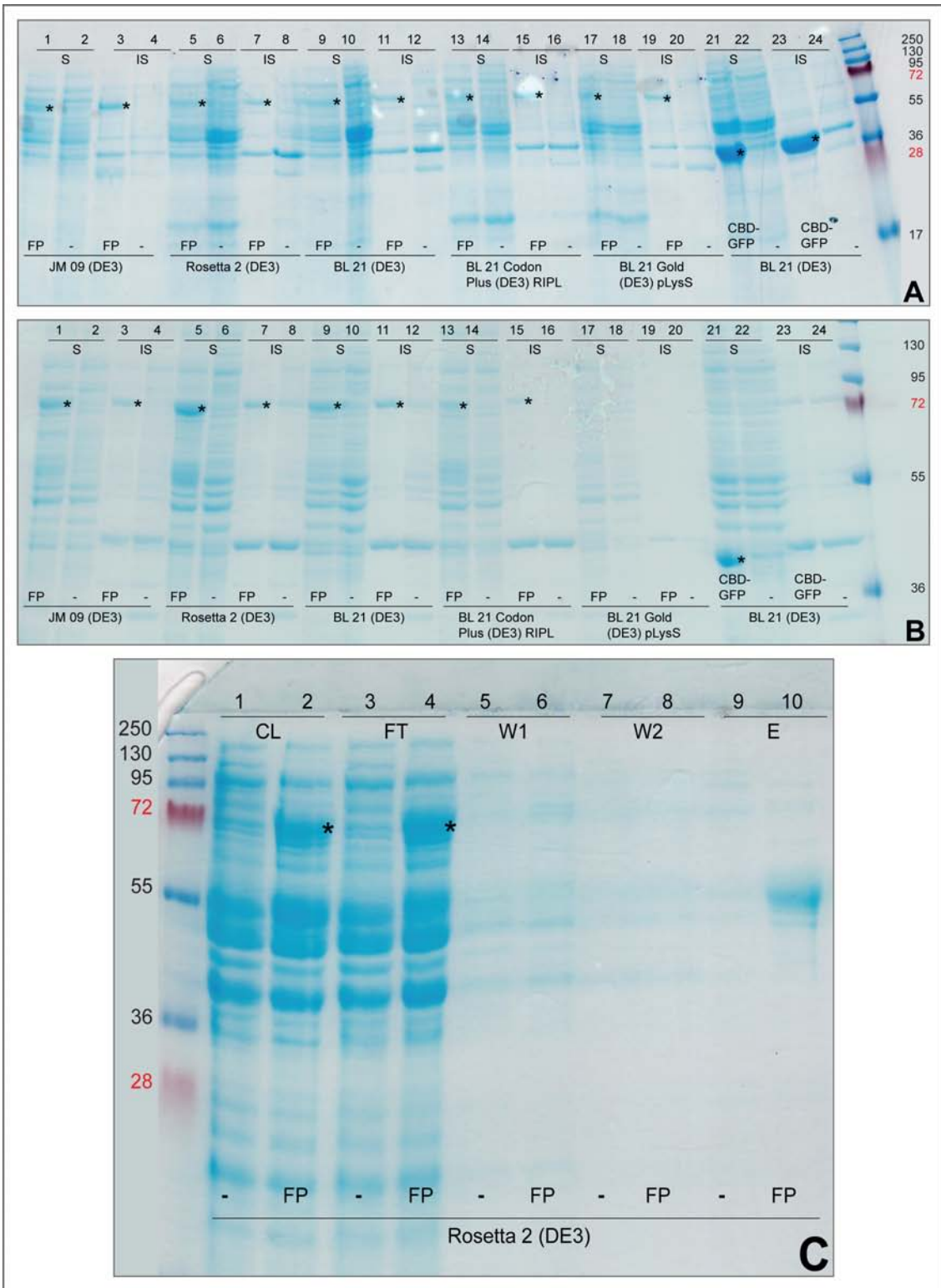**Fig. 6.2. Over Expression of Lstag-sfc-10-His. (A)** Coomassie stained SDS PAGE of the soluble and insoluble fractions of BL21 (DE3) cell lysates. Cells were transformed with pET16b Lstag-sfc-10-His and cultured at the growing temperature of 25 °C (lane 1 & 2), 30 °C (lane 3 & 4) and 37 °C (lane 5 & 6). Negative control: BL21 (DE3) lacking the expression vector (lane 7 & 8). The asterisk indicates the target protein Lstag-sfc-10-His with a molecular weight of about 36 kDa. The target protein was present in the insoluble fractions of the cell lysates. With decreasing growing temperature, less target protein was produced in the expression host. **(B)** IMAC based protein purification under denaturing conditions of the target protein Lstag-sfc-10-His (lane 1-5), CBD-GFP as the positive control (lane 6-10) and BL21 (DE3) lacking the expression vector as negative control (lane 11-15). The asterisk indicates the produced protein. S: soluble cell fraction, IS: insoluble cell fraction, CBD-GFP: chitin binding domain-green fluorescent protein, - : negative control , CL: cell lysate, FT: flow-through, W1: first wash, W2: second wash, E: eluate. Protein ladder indicates the molecular weight in kDa.

6.3.2 Solubilisation of Lstag-sfc-10 through fusion with the maltose binding protein

Proteins with unstructured polypeptide chain segments such as Lstag-sfc-10 can often not be expressed in the soluble form in *E. coli* [22]. Forcing the target protein into the soluble cell fraction could not be achieved by reducing the growing temperature (Fig. 6.2A). Another approach to solubilise a target protein is by fusing it with a highly soluble protein [23]. The maltose binding protein (MBP) is a fusion partner that is commonly used to enhance protein solubility [32-35]. The MBP-TEV encoding expression vector pRK793 was altered to create a vector that encodes a His-MBP-TEVcs-Lstag-sfc-10 fusion protein (Fig. 6.1B) that is predicted to have a molecular weight of 75.97 kDa (S. 6.2). The fusion protein contains a short amino acid sequence (TEVcs) that is recognised by the *Tobacco Etch Virus* (TEV) protease and allows separation of the Lstag-sfc-10 from its fusion partner MBP via protease cleavage. The location of the His-tag is defined by the clean-up strategy of the protein. In case of the fusion protein His-MBP-TEVcs-Lstag-sfc-10, it involves purification from other bacterial components via IMAC and subsequent cleavage with a His-tagged TEV protease. The target protein is then separated from the MBP and TEV protease by IMAC. In order to allow the proper separation of the MBP and TEV protease from the target protein, the MBP has to carry the His-tag (Fig. 6.1B).

The fusion protein His-MBP-TEVcs-Lstag-sfc-10 could be expressed in the insoluble and soluble protein fraction of the *E. coli* strains JM 09 (DE3), Rosetta 2 (DE3), BL 21 (DE3), BL 21 Codon Plus (DE3) RIPL and BL 21 Gold (DE3) pLysS at a growing temperature of 37 °C (Fig. 6.3A). To force more fusion protein into solubility, the culturing temperature was decreased to 30 °C. A markedly improvement of solubility was reached in the *E. coli* strain Rosetta 2 (DE3) (Fig. 6.3B).

6.3.3 Purification of the fusion protein

Subsequently, the soluble fusion protein His-MBP-TEVcs-Lstag-sfc-10 produced in Rosetta 2 (DE3) was purified with IMAC under native condition (Fig. 6.3C). The *in silico* predicted molecular weight of the His-MBP-TEVcs-Lstag-sfc-10 is 75.97 kDa (S. 6.2) and SDS PAGE indicated a molecular weight of about 72 kDa (Fig. 6.3). The

Fig. 6.3. Over Expression of the His-MBP-TEVcs-Lstag-sfc-10 fusion protein. (A) Coomassie stained SDS PAGE of the soluble and insoluble fractions of cell lysate of different cell hosts transformed with a modified pRK793 expressing the fusion protein His-MBP-TEVcs-Lstag-sfc-10 and cultured at the growing temperature of 37 °C. The asterisk indicates the fusion protein. The His-MBP-TEVcs-Lstag-sfc-10 fusion protein has a molecular weight of about 75 kDa. Positive control: CBD-GFP, negative controls: respective cell host lacking expression vector. **(B)** Coomassie stained SDS PAGE of the soluble and insoluble fraction of cell lysate of different cell hosts transformed with a modified pRK793 expressing the fusion protein His-MBP-TEVcs-Lstag-sfc-10 and cultured at the growing temperature of 30 °C. The asterisk indicates the fusion protein. The His-MBP-TEVcs-Lstag-sfc-10 fusion protein has a molecular weight of about 75 kDa. Positive control: CBD-GFP, negative controls: respective cell host lacking expression vector. Reduction of the growing temperature to 30 °C improved the solubility of the fusion protein. Best results were obtained when the fusion protein was produced in Rosetta 2 (DE3) (lane 5). **(C)** IMAC based clean-up under native conditions of the fusion protein His-MBP-TEVcs-Lstag-sfc-10 expressed in Rosetta 2 (DE3) at 30 °C. Negative control: cell host lacking expression vector. The asterisk indicates the produced protein. The fusion protein does not bind the nickel resin and is abundant in the flow through (lane 4). The 55 kDa protein in the eluate might be a truncated version of the fusion protein (lane 10). FP: fusion protein His-MBP-TEVcs-Lstag-sfc-10, CBD-GFP: chitin binding domain-green fluorescent protein, - : negative control, S: soluble cell fraction, IS: insoluble cell fraction, CL: cell lysate, FT: flow-through, W1: first wash, W2: second wash, E: eluate. Protein ladder indicates the molecular weight in kDa.

72 kDa fusion protein did not bind the nickel resin and was abundant in the flow-through, whereas the eluate contained a protein with a molecular weight of about 55 kDa (Fig. 6.3C). The negative control did not contain the 55 kDa protein in the eluate (Fig. 6.3C). This indicates that the 55 kDa protein is not an endogenous histidine-rich protein. Likely, it is a truncated version of the fusion protein that bound the nickel resin with greater affinity than the complete His-MBP-TEVcs-Lstag-sfc-10 fusion protein. The His-tag of the complete fusion protein might have been less accessible due to its tertiary structure.

Truncated target proteins are common contaminants and can result from wrong initiation of translation, premature translation termination or the degradation during protein expression or purification [36]. It is unlikely that the truncated version resulted from an internal translation initiation, because the resulting protein would lack the 5' His-tag and improbable bind the nickel resin. Premature translation termination can be excluded because Sanger sequencing of the expression vector did not reveal an internal stop codon and abundant amounts of the complete fusion protein were present in the cell lysates. This hints at protein degradation during expression or purification.

6.4 Future experiments

In order to obtain the native fusion protein, the purification protocol has to be optimised. Future purifications will include protease inhibitors that might prevent the formation of truncated proteins. To increase the binding affinity of the fusion protein to the nickel resin, imidazole concentration and pH of the binding buffer have to be optimised. A rearrangement of the expression vector that produces the protein Lstag-sfc-10-TEVcs-MBP-His might lead to the production of a protein that can be purified with IMAC (Fig. 6.1B). In this scenario, Lstag-sfc-10 is located at the N-terminus of the protein, followed by a TEV cleavage site and the solubilising MBP, which is His-tagged at the C-terminus. Potentially, the His-tag will not be hidden by the tertiary structure and is fully accessible for resin binding. Another approach to produce soluble protein involves expression in a eukaryotic expression hosts such as insect cells [37] or the yeasts *Pichia pastoris* [38] and *Saccharomyces cerevisiae* [39]. If it is not possible to produce the target protein in the soluble form, calcification assays can be performed with the insoluble protein as produced from the pET16b+Lstag-sfc-10-His (Fig. 6.2B). Previously, a biomineralising protein that was purified under denaturing conditions and refolded *in vitro* was successfully deployed in an calcification assays [20].

An optimised protocol that allows the production of Lstag-sfc-10 in the heterologous expression host *E. coli* will be useful to provide sufficient protein for various protein assays, that will help elucidate the role of the protein in the shell forming process.

Supplementary files

The supplementary files can be found in the appendix 2.

Supplementary S. 6.1: List of all primers used in this study

Supplementary S. 6.2: Insert and protein sequences

References

1. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM. A rapidly evolving secretome builds and patterns a sea shell. BMC Biology. 2006;4:40.
2. Clark MS, Thorne MAS, Vieira FA, Cardoso JCR, Power DM, Peck LS. Insights

into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. BMC Genomics. 2010;11:362.

3. Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C. Parallel evolution of nacre building gene sets in molluscs. Molecular Biology and Evolution. 2010;27:591-608.

4. Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y, Montagnani C. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. BMC Genomics. 2010;11:613.

5. Mann K, Edsinger-Gonzales E, Mann M. In-depth proteomic analysis of a mollusc shell: acid-soluble and acid-insoluble matrix of the limpet *Lottia gigantea*. Proteome Science. 2012;10:28.

6. Marie B, Joubert C, Tayaléa A, Zanella-Cléon I, Belliard C, Piquemal D, Cochennec-Laureau N, Marin F, Gueguen Y, Montagnani C. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences. 2012;109:20986-20991.

7. Werner GDA, Gemmell P, Grosser S, Hamer R, Shimeld SM. Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. Marine Biotechnology. 2013;15:230-243.

8. Freer A, Bridgett S, Jiang J, Cusack M. Biomineral proteins from *Mytilus edulis* mantle tissue transcriptome. Marine Biotechnology. 2014;16:34-45.

9. Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. BMC Genomics. 2014;15:249.

10. Sleight VA, Thorne MAS, Peck LS, Arivalagan J, Berland S, Marie A, Clark MS. Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, *Mya truncata*. Marine Genomics. 2016;27:47-55.

11. Marie B, Arivalagan J, Dubost L, Berland S, Marie A, Marin F. Unveiling the Evolution of Bivalve Nacre Proteins by Shell Proteomics of Unionoidae. Key Engineering Materials. 2016;672:158-167.

12. Falini G, Albeck S, Weiner S, Addadi L. Control of aragonite or calcite polymorphism by mollusk shell macromolecules. Science. 1996;271:67-69.

13. Levi Y, Albeck S, Weiner S, Brack A, Addadi L. Control over aragonite crystal nucleation and growth: an *in vitro* study of biomineralization. Chemistry-A European Journal. 1998;4:389-396.

14. Gotliv B-A, Addadi L, Weiner S. Mollusk shell acidic proteins: in search of individual functions. ChemBioChem. 2003;4:522-529.

15. Suzuki M, Kogure T, Weiner S, Addadi L. Formation of aragonite crystals in the crossed lamellar microstructure of limpet shells. Crystal Growth & Design. 2011;11:4850-4859.

16. Montagnani C, Marie B, Marin F, Belliard C, Riquet F, Tayalé A, Zanella-Cléon I, Fleury E, Gueguen Y, Piquemal D. *Pmarg*-Pearlin is a matrix protein involved in nacre framework formation in the pearl oyster *Pinctada margaritifera*. ChemBioChem. 2011;12:2033-2043.

17. Politi Y, Mahamid J, Goldberg H, Weiner S, Addadi L. Asprich mollusk shell protein: *in vitro* experiments aimed at elucidating function in CaCO 3

crystallization. CrystEngComm. 2007;9:1171-1177.

18. Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. Science. 2007;316:1893-1895.

19. Wang N, Lee Y-H, Lee J. 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 Part B: Biochemistry and Molecular Biology. 2008;149:354-361.

20. Natalio F, Corrales TP, Panthöfer M, Schollmeyer D, Lieberwirth I, Müller WEG, Kappl M, Butt H-J, Tremel W. Flexible minerals: self-assembled calcite spicules with extreme bending strength. Science. 2013;339:1298-1302.

21. Dodenhof T, Dietz F, Franken S, Grunwald I, Kelm S. Splice variants of Perlucin from *Haliotis laevigata* modulate the crystallisation of CaCO 3. PloS one. 2014;9:e97126.

22. Gräslund S, Nordlund P, Weigelt J, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J. Protein production and purification. Nature methods. 2008;5:135-146.

23. Khow O, Suntrarachun S. Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pacific journal of tropical biomedicine. 2012;2:159-162.

24. Kapust RB, T√∂zs√©r J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein engineering. 2001;14:993-1000.

25. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods. 2009;6:343-345.

26. Weiss IM, Schönitzer V. The distribution of chitin in larval shells of the bivalve mollusk *Mytilus galloprovincialis*. Journal of Structural Biology. 2006;153:264-277.

27. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. 2003;31:3784-3788.

28. Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A: Protein identification and analysis tools on the ExPASy server. Springer; 2005.

29. Studier FW. Protein production by auto-induction in high-density shaking cultures. Protein Expression and Purification. 2005;41:207-234.

30. Cohen SN, Chang ACY, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids *in vitro*. Proceedings of the National Academy of Sciences. 1973;70:3240-3244.

31. Porath J. Immobilized metal ion affinity chromatography. Protein Expression and Purification. 1992;3:263-281.

32. Fox JD, Waugh DS. Maltose-binding protein as a solubility enhancer. In: editors. *E. coli* Gene Expression Protocols. Humana Press; 2003. p. 99-117.

33. Waugh DS. Making the most of affinity tags. TRENDS in Biotechnology. 2005;23:316-320.

34. Nallamsetty S, Waugh DS. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. Protein expression and

purification. 2006;45:175-182.

35. Nallamsetty S, Waugh DS. A generic protocol for the expression and purification of recombinant proteins in Escherichia coli using a combinatorial His6-maltose binding protein fusion tag. Nature protocols. 2007;2:383-391.

36. QIAGEN: The QIAexpressionist TM A handbook for high-level expression and purification of 6xHis-tagged proteins. 18. A handbook for high-level expression and purification of 6xHis-tagged proteins 2003.

37. Smith GE, Summers MD, Fraser MJ. Production of human beta interferon in insect cells infected with a baculovirus expression vector. Molecular and Cellular Biology. 1983;3:2156-2165.

38. Boettner M, Prinz B, Holz C, Stahl U, Lang C. High-throughput screening for expression of heterologous proteins in the yeast *Pichia pastoris*. Journal of Biotechnology. 2002;99:51-62.

39. Holz C, Hesse O, Bolotina N, Stahl U, Lang C. A micro-scale process for high-throughput expression of cDNAs in the yeast *Saccharomyces cerevisiae*. Protein Expression and Purification. 2002;25:372-378.

7

General discussion

Much of the evolutionary success of molluscs can be traced back to the ability to form a calcified shell that supports the soft molluscan body and offers protection from predation and desiccation [1]. The molluscan shell is comprised of $CaCO_3$ and a matrix of organic macromolecules such as proteins, glycoproteins, chitin and acidic polysaccharides [1]. This matrix gets secreted by the mantle tissue and constitutes a scaffold on which the $CaCO_3$ minerals form [2, 3]. Although the organic components make up less than 5% of the shell weight, they play a fundamental role in conferring the shells material properties and finally define shape, size and colour of the mature structure [4]. Technical advances in nucleic acid sequencing and high throughput proteomics promoted the identification of the proteinaceous components of the extracellular matrix on the molecular level and enabled the collection of a large number of mineralising genes and their encoded proteins. The bulk of these proteins were identified from marine species with a nacro-prismatic shell and great economic value [5-19]. However, besides the prism and nacre microstructures, the shell can be composed of many other textures such as spherulitic, laminar, foliated, helical or crossed [20]. To get a better understanding of the shell secretome diversity and further elucidate the evolutionary history of molluscan shell formation, it is important to extend the research focus to species that compose microstructures other than narco-prismatic. In my thesis, I study the shell proteome of the freshwater pulmonate *Lymnaea stagnalis* that secrets a shell of the crossed-lamellar type, with special regards to its ontogenetic characterisation.

The functional characterisation of the proteinaceous matrix components still represents a major challenge in the field of molluscan biomineralisation. I aim to gain insight into the function of shell-forming genes by analysing their expression with *in situ* hybridisation. This technique allows the spatial and temporal resolution of gene expression within a tissue and can significantly contribute to the functional characterisation of next generation sequence data. We developed *in situ*

hybridisation protocols for *L. stagnalis* larval (chapter 2, 3) and adult mantle (chapter 4) tissues to trace the expression of the biomineralisation gene candidates during all ontogenetic stages. *L. stagnalis* larvae are deposited in an egg capsuled filed with a viscous fluid that likely interferes with the WMISH procedure. The extraction of the embryos and larvae from the capsule is a delicate and time-consuming procedure. In chapter 2 we present a protocol that challenges inhibiting factors by applying the mucolytic agent N-acetyl-L-cysteine (NAC), a reducing agent dithiothreitol (DTT) and a pre-hybridization treatment with sodium dodecyl sulfate (SDS). Further optimisation efforts led to the development of a method that includes the high-throughput removal of embryos and larvae from the egg capsules and a preservation regime that fixes larvae prior to decapsulation. This advanced method presented in chapter 3 simplifies the WMISH procedure and apparently negates the need for treatments with NAC, DTT or SDS.

In chapter 5 we present more than 30 candidate genes associated with the production of the crossed-lamellar *L. stagnalis* shell that were identified with a proteomic survey coupled with transcriptome sequencing. The spatial characterisation of the shell-forming proteome highlighted its dynamic and modular regulation, and the potential for interactions between co-expressed genes. We could show that many adult shell-forming genes are expressed in larval shell-forming cells, foreshadowing the zonation of the adult mantle. The *in situ* hybridisation experiments revealed patterns of asymmetry that presumably contribute to the coiled phenotype of the shell. We were able to identify a peroxidase that likely plays a role in cross-linking periostracum proteins and present three glycine-rich proteins that might serve as substrate for the cross-linking reaction [21, 22]. The presence of chitin interacting proteins in the *L. stagnalis* shell proteome indicates that chitin plays a role in scaffolding the crossed-lamellar matrix [23]. We furthermore identified aspartic acid-rich proteins that likely serve as organic template for epitaxial crystal growth [24, 25] and found a number of proteins that are predicted to be highly glycosylated, a protein feature that is assumed to enhance biomineral nucleation [26]. Each molluscan lineage has uniquely evolved a large fraction of its shell-forming proteome with only some deeply conserved elements. This protein diversity led to the astonishing diversity of molluscan shell ultrastructures, crystal textures and colours [9, 17, 27-29]. By

comparing a number of biomineralising proteomes with the shell proteome of *L. stagnalis*, we were able to identify conserved elements (i.a. Pif-like, peroxidase, von Willenbrand factor A domain, Sushi domain) and discovered that almost all *L. stagnalis* shell-forming candidates expressed in the mantle low columnar outer epithelium share sequence similarity with *Cepaea nemoralis* shell proteins. Gene co-option, domain shuffling and gene family expansion are molecular mechanisms that have contributed to the diversity of molluscan shell-forming proteins [4, 29]. Our analysis of the gene architecture of the biomineralisation genes suggests that alternative splicing acts to significantly expand the shell-forming molecular repertoire.

7.1 Outlook

The generation of a *L. stagnalis* shell protein catalogue is important to allow the broad comparisons between shell-forming secretomes and can aid to solve questions related to evolutionary conservation and innovation. The expression analyses are useful to indicate co-expression and can emphasise unique or prominent expression patterns. Nonetheless, the exact role of the encoded protein cannot be determined by solely studying gene expression. To shed light on the role of the proteinaceous shell components, *in vitro* approaches such as calcification assays with recombinant proteins are useful [30-34]. These assays can approximate the effects of individual shell proteins or specific domains on calcium carbonate crystallisation. Furthermore, protein assays can be applied to investigate functions besides crystallisation. For example, an enzyme assay could be used to test the supposed cross-linking ability of the peroxidase on the *L. stagnalis* glycine-rich shell-forming candidates *Lstag*-sfc-6, *Lstag*-sfc-7 and *Lstag*-sfc-8 [35-37]. Furthermore, a modifying effect of the putative chitin deacetylase-like domain containing *Lstag*-sfc-21 on chitin could be determined with an enzyme assay containing the shell-forming candidate and chitin [38]. The insulin-like growth factor (IGF) binding ability of the *Haliotis laevigata* Perlustrin was previously shown *in vitro* by Weiss et al. 2001 [39]. Following the experimental set-up of that study, the putative IGF binding ability of the IMP-L2-like *Lstag*-sfc-18 could be tested.

In order to provide sufficient protein for the various protein assays, it will be useful to have protocols at hand that allow the production of the protein in question in a heterologous expression host. In Chapter 6, I present the on-going work that describes the protein production of shell- forming candidate *Lstag*-sfc-10 in *Escherichia coli*.

Vast opportunities of studying the molecular mechanisms of shell formation in *L. stagnalis* will arise once the CRISPR-Cas nuclease system is established [40, 41]. This gene editing approach is markedly easier to apply than other technologies and has proven to be highly specific and efficient for a variety of organisms [42]. It allows the knock-out of an entire gene or the removal of specific domains in a multidomain protein and has recently been demonstrated to function in molluscs [43].

The selective removal of a putative mineral interacting domain could unravel the effect of this domain on the mineralised structure. This approach represents an advantage to RNAi or morpholinos that knock-down the abundance of an entire transcript posttranslationally [6, 44, 45]. Furthermore, RNAi does not allow the generation of gene knock-outs and can show substantial off-target effects [46].

Aspartic acid-rich proteins are common biomineral components and are suspected to be involved in epitaxial crystal growth [24, 25]. We identified two aspartic acid-rich proteins in *L. stagnalis* (chapter 5). By removing the aspartic acid-rich repeats, the effect of these regions on mineralisation could be monitored. The results might give valuable insights to further elucidate the effect of aspartic acid-rich proteins on mineralisation.

Another exciting possibility of the CRISPR-Cas nuclease system is to create gene knock-ins. With this approach, *L. stagnalis* biomineralising proteins could be tagged by fusing them with fluorescent proteins. This allows the *in vivo* monitoring of the shell proteins throughout the entire life of the animal and within the mature biomineral. In chapter 5, we identified a number of candidates that show unilateral expression along the larval mantle margin, which hints at unilateral distribution of the gene products within the larval shell. If the encoded proteins show unilateral distribution within the coiled adult shell could be unravelled by creating fluorescent knock-ins. Undoubtedly the exploitation of the CRISPR-Cas nuclease

system would promote the functional characterisation of the *L. stagnalis* shell proteins immensely.

References

1. Marin F, Luquet G. Molluscan shell proteins. Comptes Rendus Palevol. 2004;3:469-492.
2. Levi-Kalisman Y, Falini G, Addadi L, Weiner S. Structure of the nacreous organic matrix of a bivalve mollusk shell examined in the hydrated state using cryo-TEM. Journal of Structural Biology. 2001;135:8-17.
3. Addadi L, Joester D, Nudelman F, Weiner S. Mollusk shell formation: a source of new concepts for understanding biomineralization processes. Chemistry – A European Journal. 2006;12:980-987.
4. Marin F, Luquet G, Marie B, Medakovic D. Molluscan shell proteins: primary structure, origin, and evolution. Current Topics in Developmental Biology. 2008;80:209-276.
5. Yano M, Nagai K, Morimoto K, Miyamoto H. Shematrin: a family of glycine-rich structural proteins in the shell of the pearl oyster *Pinctada fucata*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 2006;144:254-262.
6. Suzuki M, Saruwatari K, Kogure T, Yamamoto Y, Nishimura T, Kato T, Nagasawa H. An acidic matrix protein, Pif, is a key macromolecule for nacre formation. Science. 2009;325:1388-1390.
7. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012;490:49-54.
8. Joubert C, Piquemal D, Marie B, Manchon L, Pierrat F, Zanella-Cléon I, Cochennec-Laureau N, Gueguen Y, Montagnani C. Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. BMC Genomics. 2010;11:613.
9. Jackson DJ, McDougall C, Woodcroft B, Moase P, Rose RA, Kube M, Reinhardt R, Rokhsar DS, Montagnani C, Joubert C. Parallel evolution of nacre building gene sets in molluscs. Molecular Biology and Evolution. 2010;27:591-608.
10. Marie B, Joubert C, Tayaléa A, Zanella-Cléon I, Belliard C, Piquemal D, Cochennec-Laureau N, Marin F, Gueguen Y, Montagnani C. Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. Proceedings of the National Academy of Sciences. 2012;109:20986-20991.
11. Kinoshita S, Wang N, Inoue H, Maeyama K, Okamoto K, Nagai K, Kondo H, Hirono I, Asakawa S, Watabe S. Deep sequencing of ESTs from nacreous and prismatic layer producing tissues and a screen for novel shell formation-related genes in the pearl oyster. PLoS One. 2011;6:e21238.
12. Miyamoto H, Endo H, Hashimoto N, Isowa Y, Kinoshita S, Kotaki T, Masaoka T, Miki T, Nakayama S, Nogawa C. The diversity of shell matrix proteins: genome-wide investigation of the pearl oyster, *Pinctada fucata*. Zoological Science. 2013;30:801-816.
13. Shen X, Belcher AM, Hansma PK, Stucky GD, Morse DE. Molecular cloning and characterization of lustrin A, a matrix protein from shell and pearl nacre of

Haliotis rufescens. Journal of Biological Chemistry. 1997;272:32472-32481.

14. Weiss IM, Kaufmann S, Mann K, Fritz M. Purification and characterization of perlucin and perlustrin, two new proteins from the shell of the mollusc *Haliotis laevigata*. Biochemical and Biophysical Research Communications. 2000;267:17-21.

15. Bédouet L, Schuller MJ, Marin F, Milet C, Lopez E, Giraud M. Soluble proteins of the nacre of the giant oyster Pinctada maxima and of the abalone *Haliotis tuberculata*: extraction and partial analysis of nacre proteins. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology. 2001;128:389-400.

16. Jolly C, Berland S, Milet C, Borzeix S, Lopez E, Doumenc D. Zona localization of shell matrix proteins in mantle of *Haliotis tuberculata* (Mollusca, Gastropoda). Marine Biotechnology. 2004;6:541-551.

17. Jackson DJ, McDougall C, Green K, Simpson F, Wörheide G, Degnan BM. A rapidly evolving secretome builds and patterns a sea shell. BMC Biology. 2006;4:40.

18. Jackson DJ, Wörheide G, Degnan BM. Dynamic expression of ancient and novel molluscan shell genes during ecological transitions. BMC Evolutionary Biology. 2007;7:160.

19. Marie B, Marie A, Jackson DJ, Dubost L, Degnan BM, Milet C, Marin F. Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. Proteome Science. 2010;8:54.

20. Marin F, Le Roy N, Marie B. The formation and mineralization of mollusk shell. Frontiers in Bioscience. 2012;4:1099-1125.

21. Waite JH. The phylogeny and chemical diversity of quinone-tanned glues and varnishes. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry. 1990;97:19-29.

22. Michon T, Chenu M, Kellershon N, Desmadril M, Guéguen J. Horseradish peroxidase oxidation of tyrosine-containing peptides and their subsequent polymerization: a kinetic study. Biochemistry. 1997;36:8504-8513.

23. Yonezawa M, Sakuda S, Yoshimura E, Suzuki M. Molecular cloning and functional analysis of chitinases in the fresh water snail, *Lymnaea stagnalis*. Journal of Structural Biology. 2016;196:107-118.

24. Addadi L, Weiner S. Interactions between acidic proteins and crystals: stereochemical requirements in biomineralization. Proceedings of the National Academy of Sciences. 1985;82:4110-4114.

25. Weiner S, Traub W, Parker SB. Macromolecules in mollusc shells and their functions in biomineralization [and Discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences. 1984;304:425-434.

26. Nudelman F, Gotliv BA, Addadi L, Weiner S. Mollusk shell formation: mapping the distribution of organic matrix components underlying a single aragonitic tablet in nacre. Journal of Structural Biology. 2006;153:176-187.

27. Mann K, Jackson DJ. Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. BMC Genomics. 2014;15:249.

28. Marie B, Jackson DJ, Ramos-Silva P, Zanella-Cleon I, Guichard N, Marin F. The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. FEBS Journal. 2013;280:214-232.

29. Kocot KM, Aguilera F, McDougall C, Jackson DJ, Degnan BM. Sea shell diversity

and rapidly evolving secretomes: insights into the evolution of biomineralization. Frontiers in Zoology. 2016;13:23.

30. Politi Y, Mahamid J, Goldberg H, Weiner S, Addadi L. Asprich mollusk shell protein: *in vitro* experiments aimed at elucidating function in CaCO 3 crystallization. CrystEngComm. 2007;9:1171-1177.

31. Jackson DJ, Macis L, Reitner J, Degnan BM, Wörheide G. Sponge paleogenomics reveals an ancient role for carbonic anhydrase in skeletogenesis. Science. 2007;316:1893-1895.

32. Wang N, Lee Y-H, Lee J. 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 Part B: Biochemistry and Molecular Biology. 2008;149:354-361.

33. Natalio F, Corrales TP, Panthöfer M, Schollmeyer D, Lieberwirth I, Müller WEG, Kappl M, Butt H-J, Tremel W. Flexible minerals: self-assembled calcite spicules with extreme bending strength. Science. 2013;339:1298-1302.

34. Dodenhof T, Dietz F, Franken S, Grunwald I, Kelm S. Splice variants of Perlucin from *Haliotis laevigata* modulate the crystallisation of CaCO 3. PloS one. 2014;9:e97126.

35. Stahmann MA, Spencer AK, Honold GR. Cross linking of proteins *in vitro* by peroxidase. Biopolymers. 1977;16:1307-1318.

36. Andersen SO. The cross-links in resilin identified as dityrosine and trityrosine. Biochimica et Biophysica Acta (BBA)-General Subjects. 1964;93:213-215.

37. Elvin CM, Carr AG, Huson MG, Maxwell JM, Pearson RD, Vuocolo T, Liyou NE, Wong DCC, Merritt DJ, Dixon NE. Synthesis and properties of crosslinked recombinant pro-resilin. Nature. 2005;437:999-1002.

38. Caufrier F, Martinou A, Dupont C, Bouriotis V. Carbohydrate esterase family 4 enzymes: substrate specificity. Carbohydrate research. 2003;338:687-692.

39. Weiss IM, Göhring W, Fritz M, Mann K. Perlustrin, a *Haliotis laevigata (*abalone) nacre protein, is homologous to the insulin-like growth factor binding protein N-terminal module of vertebrates. Biochemical and Biophysical Research Communications. 2001;285:244-249.

40. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337:816-821.

41. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819-823.

42. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8:2281.

43. Perry KJ, Henry JQ. CRISPR/Cas9-mediated genome modification in the mollusc, *Crepidula fornicata*. Genesis. 2015;53:237-244.

44. Zhao M, He M, Huang X, Wang Q, Shi Y. Functional characterization and molecular mechanism exploration of three granulin epithelin precursor splice variants in biomineralization of the pearl oyster *Pinctada fucata*. Molecular Genetics and Genomics. 2016;291:399-409.

45. Jiao Y, Wang H, Du X, Zhao X, Wang Q, Huang R, Deng Y. Dermatopontin, a shell matrix protein gene from pearl oyster Pinctada martensii, participates

in nacre formation. Biochemical and biophysical research communications. 2012;425:679-683.

46. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014;346:1258096.

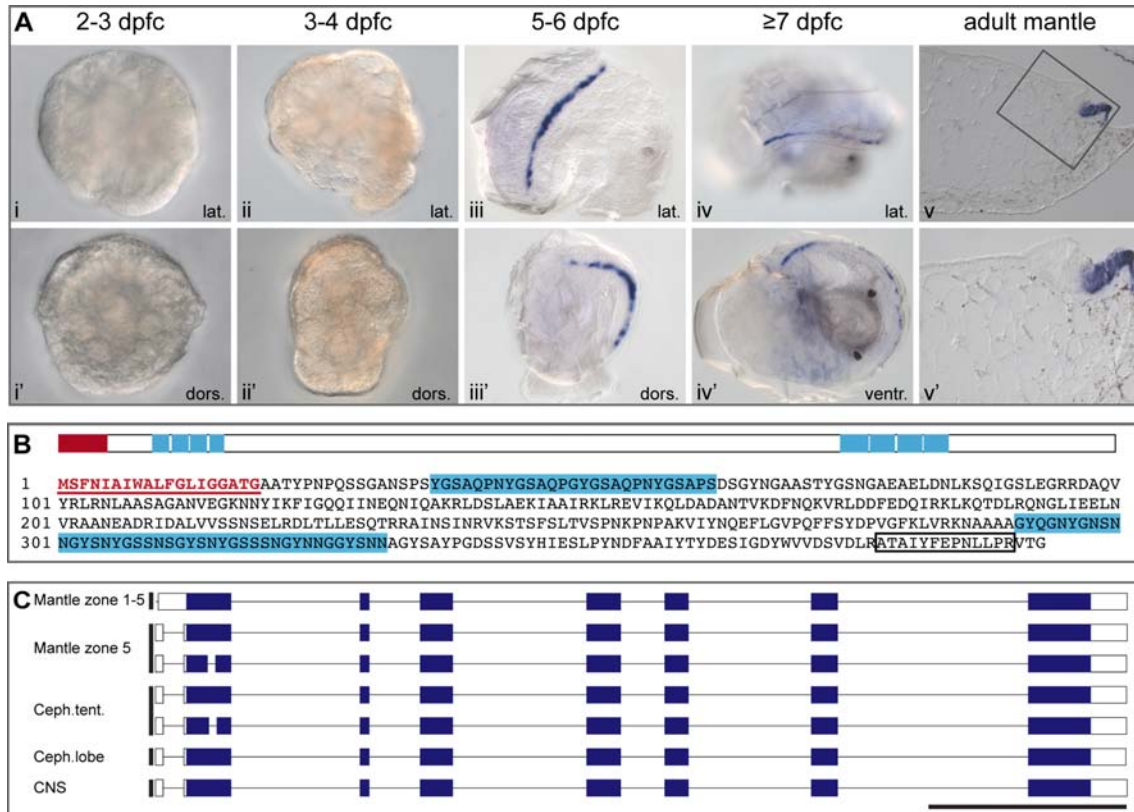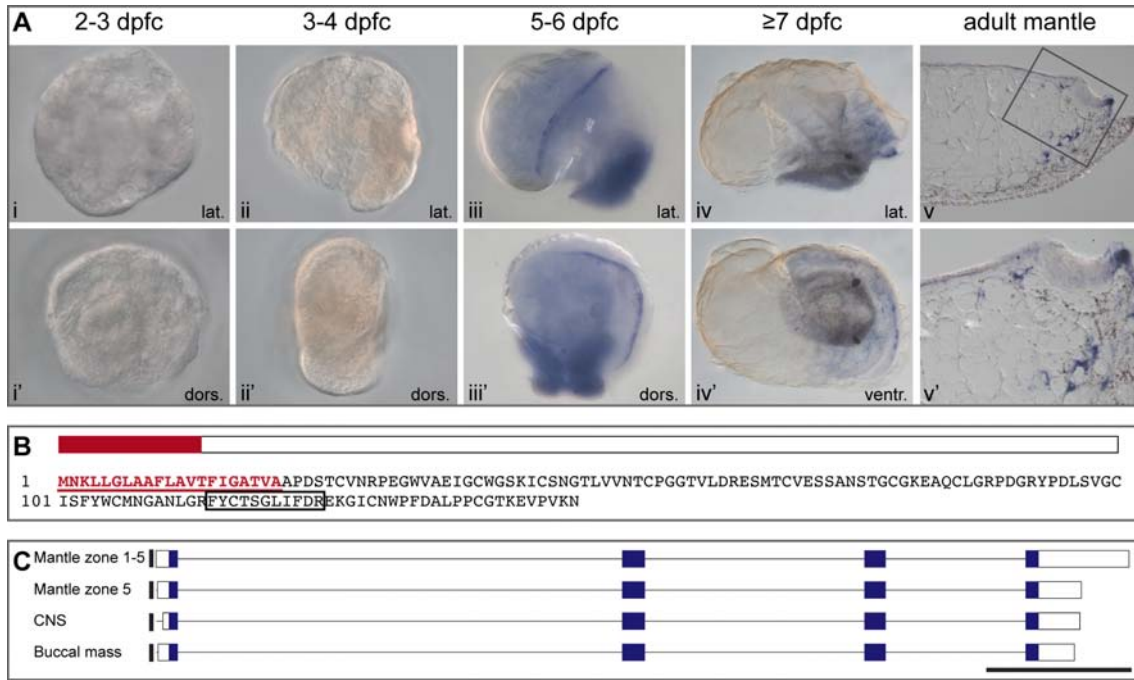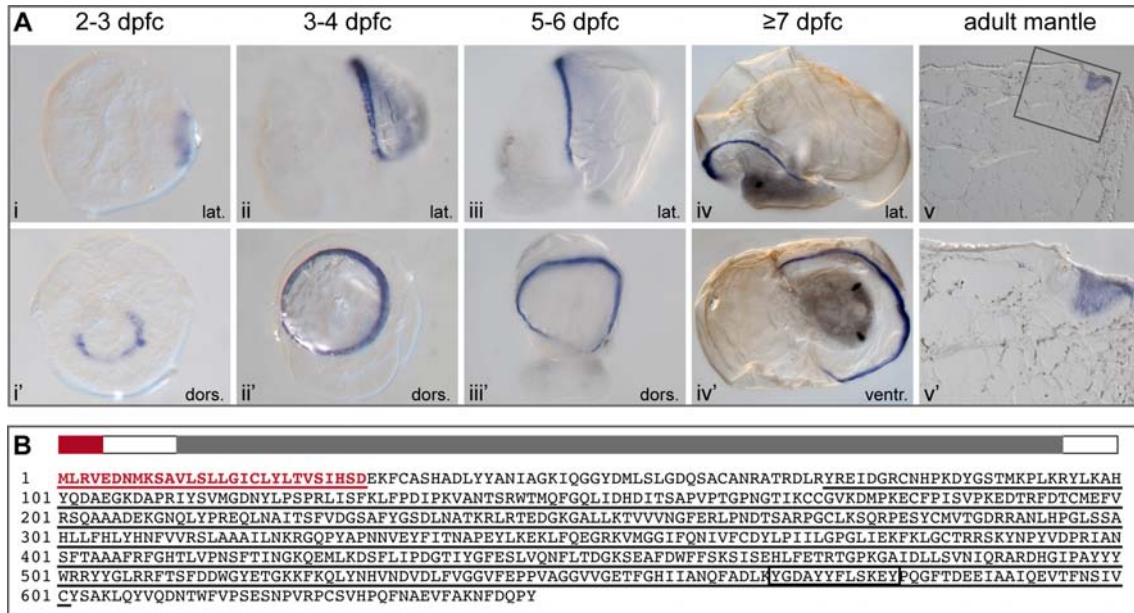# Whole mount *in situ* hybridisation results and molecular features of 34 shell-forming gene candidates
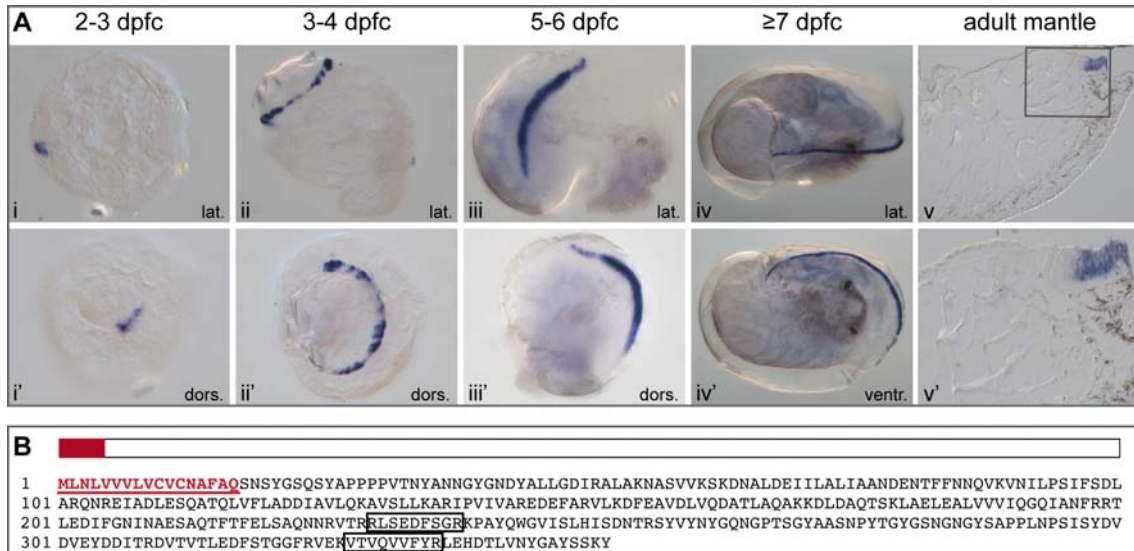
Additional file 5.2. Spatial expression and molecular features of *Lstag-sfc-1*. (A) *In situ* hybridisation against *Lstag-sfc-1* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and in zone one and two of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat), dorsal (dors), ventral (ventr). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_252293. The signal sequence (red) and the peptides identified with MS/MS (boxed sections) are indicated.
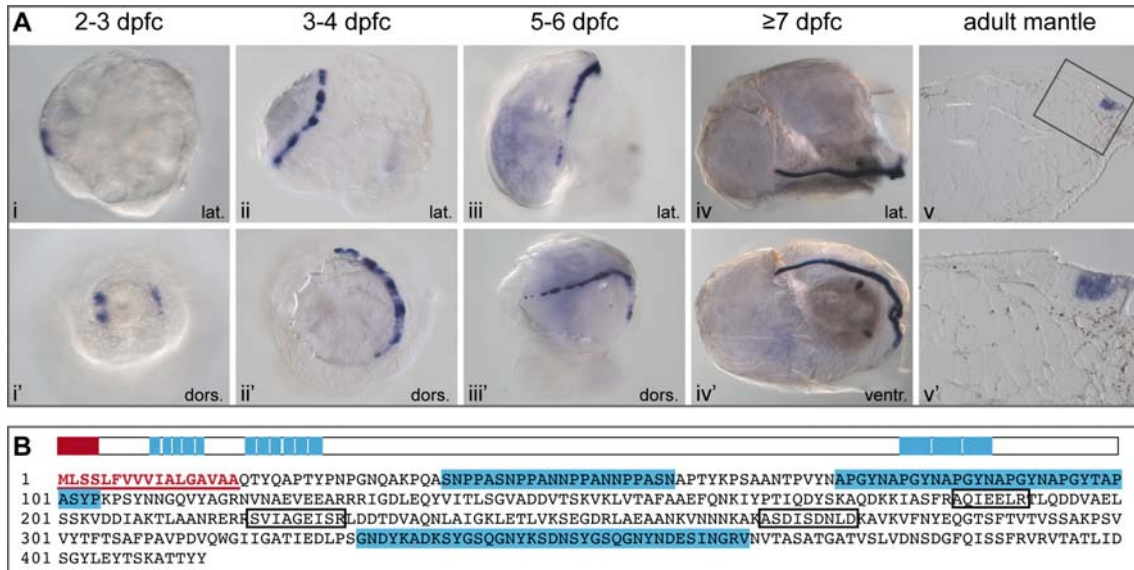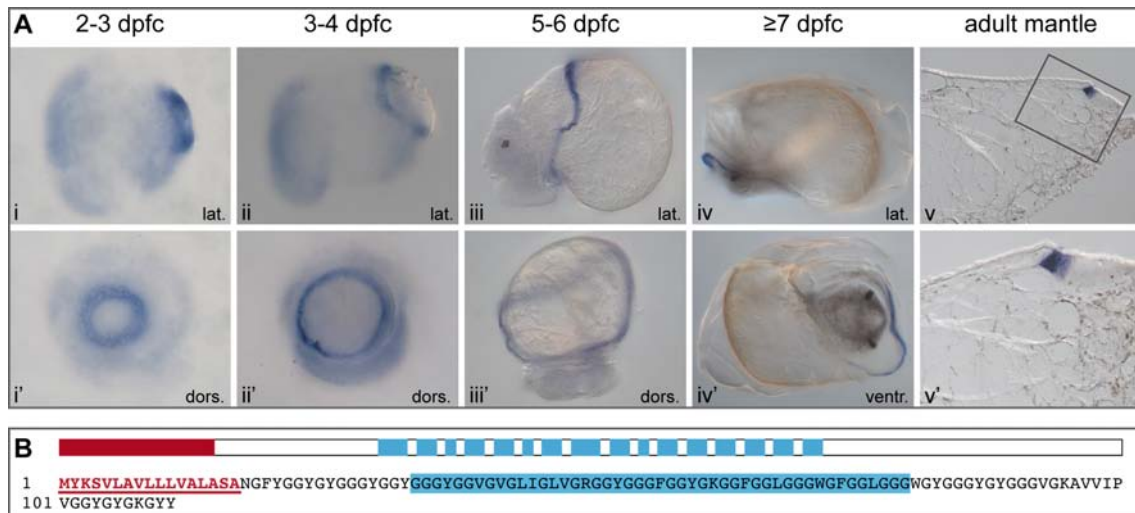
Additional file 5.3. Spatial expression and molecular features of *Lstag-sfc-2*. (A) *In situ* hybridisation against *Lstag-sfc-2* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (ii-iv, ii'-iv') and zone one and two of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_77206. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed sections) are indicated.
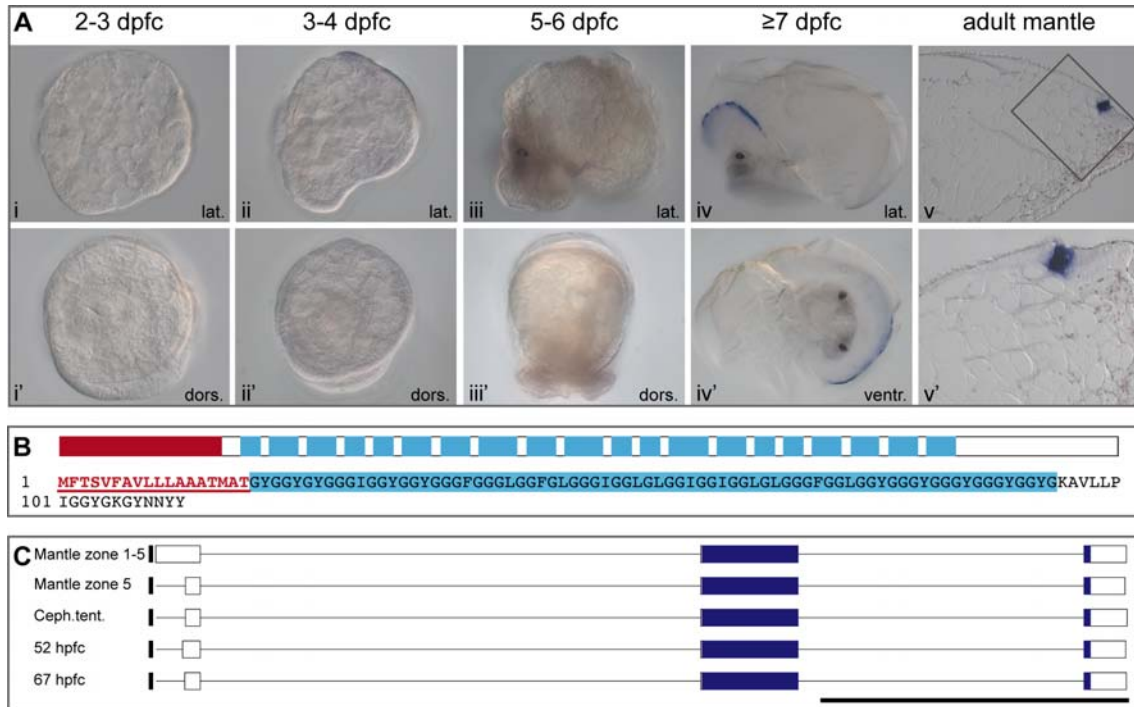
Additional file 5.4. Spatial expression and molecular features of *Lstag-sfc-3*. (A) *In situ* hybridisation against *Lstag-sfc-3* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iii, iii', iv, iv') and zone one and two of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Edge_v2_CLC_107. The signal sequence (red), the repeated motifs (blue), and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-3*. Transcriptomic contigs similar to Mantle_Edge_v2_CLC_107 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01097. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
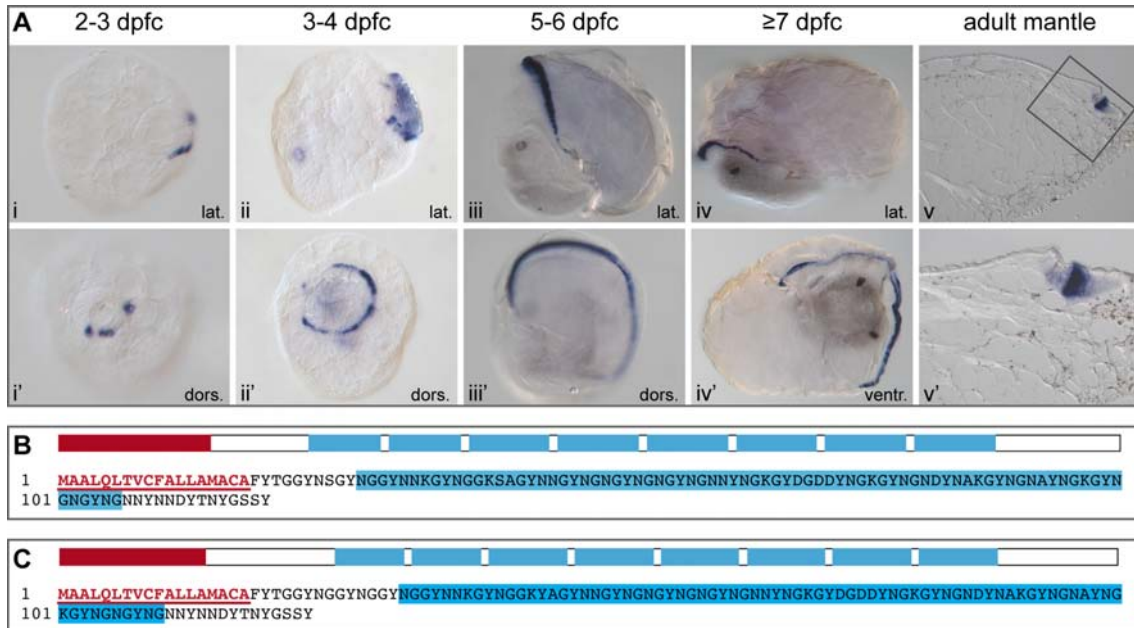
Additional file 5.5. Spatial expression and molecular features of *Lstag-sfc-4*. (A) *In situ* hybridisation against *Lstag-sfc-4* during development and in the fully differentiated adult mantle.. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iii, iv, iii', iv') and zone one at the posterior wall of the mantle groove (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_255684 and Mantle_Edge_v2_idb_73304. The signal sequence (red) and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of the *Lstag-sfc-4*. Transcriptomic contigs similar to Mantle_Edge_v2_Tri_255684 and Mantle_Edge_v2_idb_73304. were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic gLs.1.0.scaf01511. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
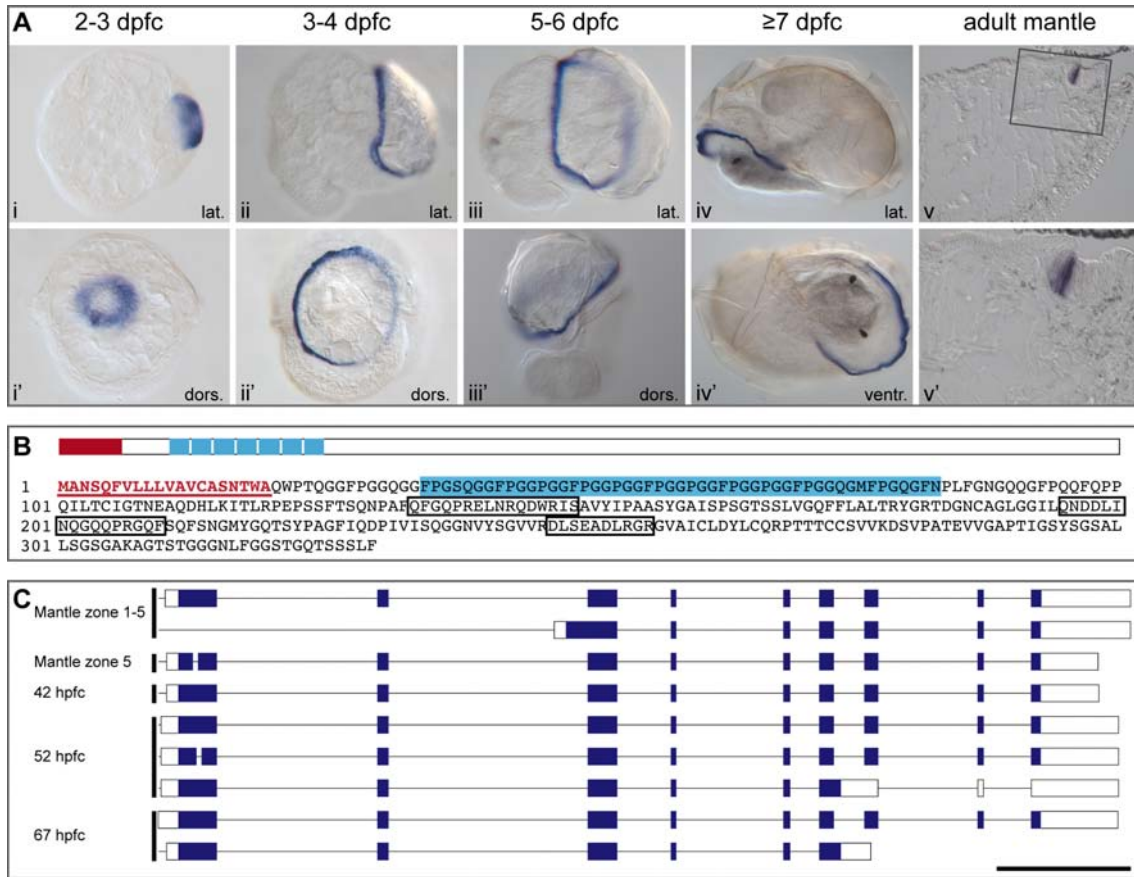
Additional file 5.6. Spatial expression and molecular features of *Lstag-sfc-5*. (A) *In situ* hybridisation against *Lstag-sfc-5* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone one and two of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_CLC_241. The signal sequence (red), the conserved animal haem peroxidase domain (underlined black and highlighted grey) and the peptide identified with MS/MS (boxed section) are indicated.
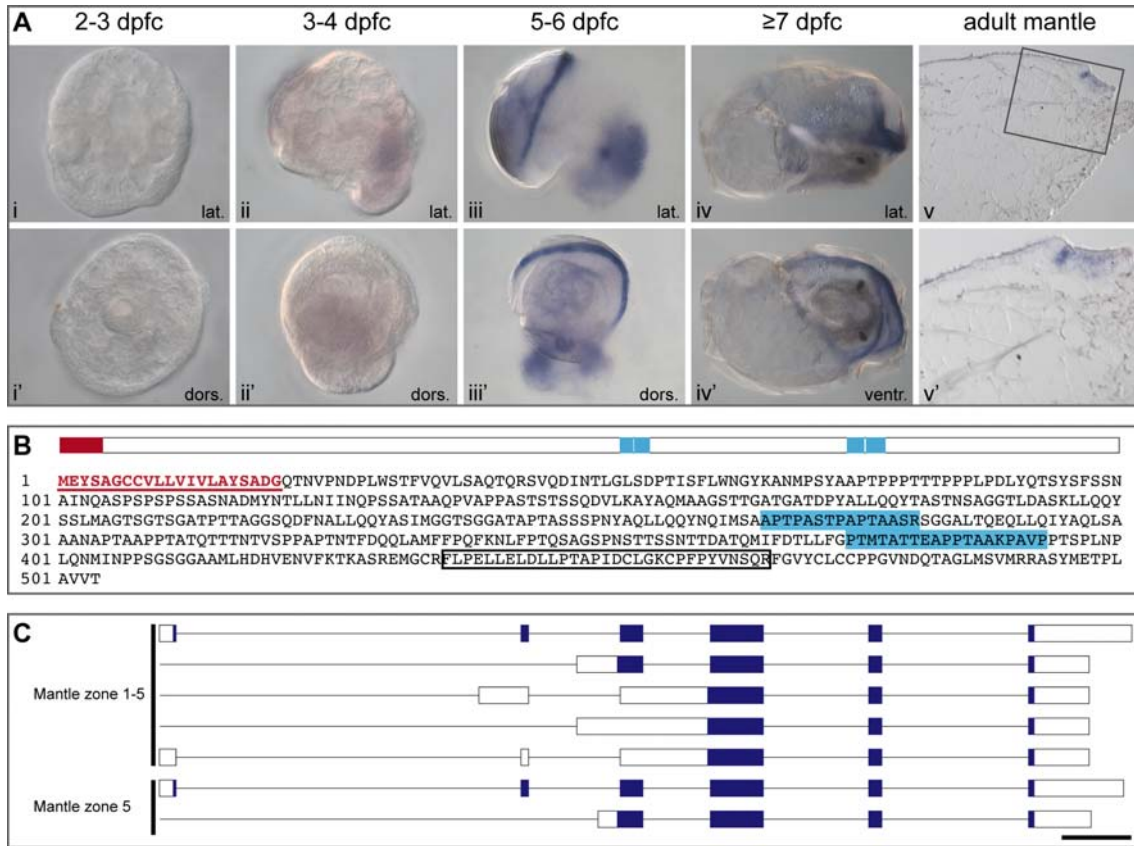
Additional file 5.7. Spatial expression and molecular features of *Lstag-sfc-6*. (A) *In situ* hybridisation against *Lstag-sfc-6* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_232353. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated. (C) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_111343. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated.
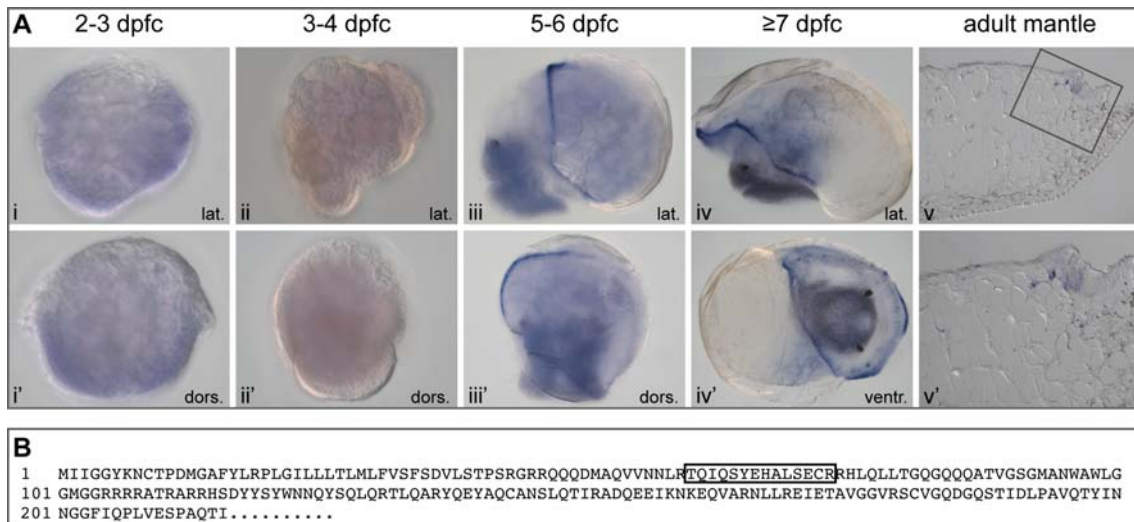
Additional file 5.8. Spatial expression and molecular features of *Lstag-sfc*-7. (A) *In situ* hybridisation against *Lstag-sfc-7* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_209370 and Mantle_Edge_v2_idb_133123. The signal sequence (red), the repeated motif (blue) are indicated.

Additional file 5.9. Spatial expression and molecular features of *Lstag-sfc-8*. (A) *In situ* hybridisation against *Lstag-sfc-8* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iv, iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_126421. The signal sequence (red) and the repeated motif (blue) are indicated. (D) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-8*. Transcriptomic contigs similar to Mantle_Edge_v2_Tri_126421 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf03406. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
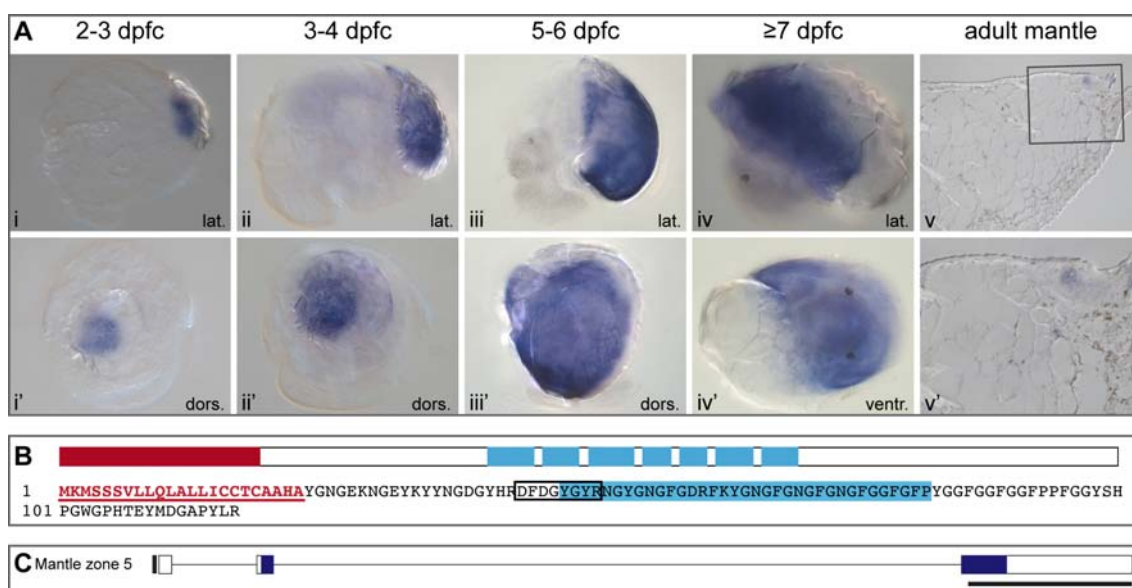
Additional file 5.10. Spatial expression and molecular features of *Lstag-sfc-9*. (A) *In situ* hybridisation against *Lstag-sfc-9* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_86937. The signal sequence (red) and the repeated motif (blue) are indicated. (C) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_86933. The signal sequence (red) and the repeated motif (blue) are indicated.

Additional file 5.11. Spatial expression and molecular features of *Lstag-sfc-10*. (A) *In situ* hybridisation against *Lstag-sfc-10* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_67296. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of the *Lstag-sfc-10*. Transcriptomic contigs similar to Mantle_Edge_v2_idb_67296 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01041. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
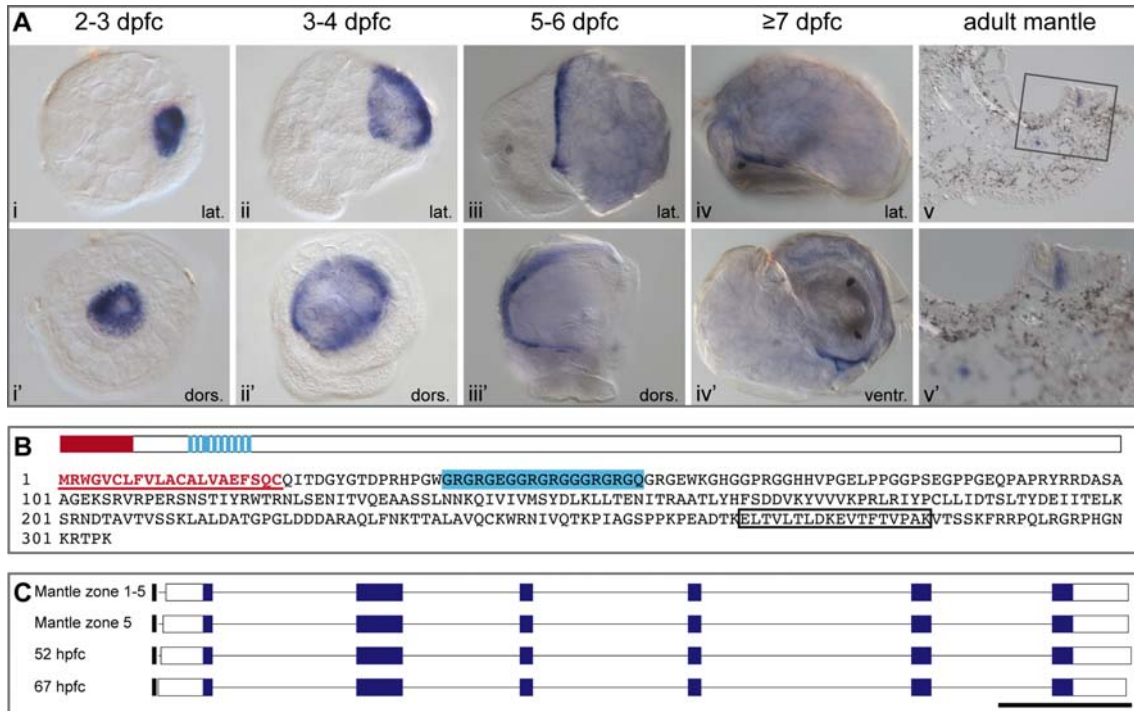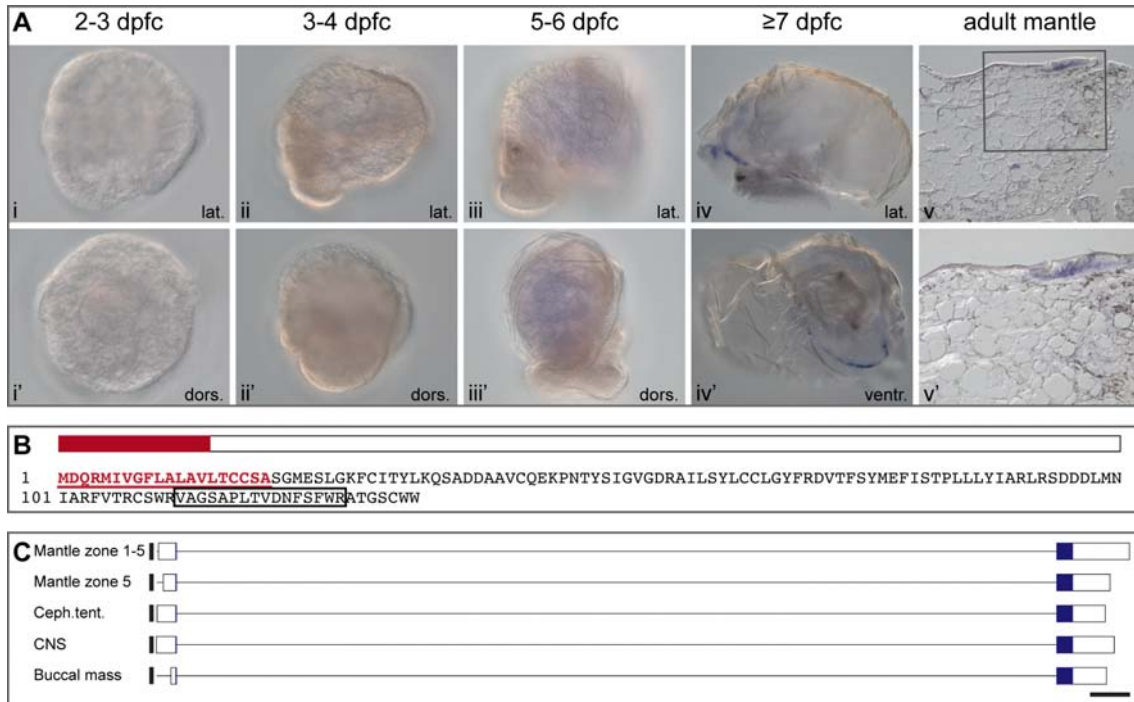
Additional file 5.12. Spatial expression and molecular features of *Lstag-sfc-11*. (A) *In situ* hybridisation against *Lstag-sfc-11* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iii, iv, iii', iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_CLC_58. The signal sequence (red), the repeated motifs (blue), C-terminal cysteine residues (highlighted grey) and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-11*. Transcriptomic contigs similar to Mantle_Edge_v2_CLC_58 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01302. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
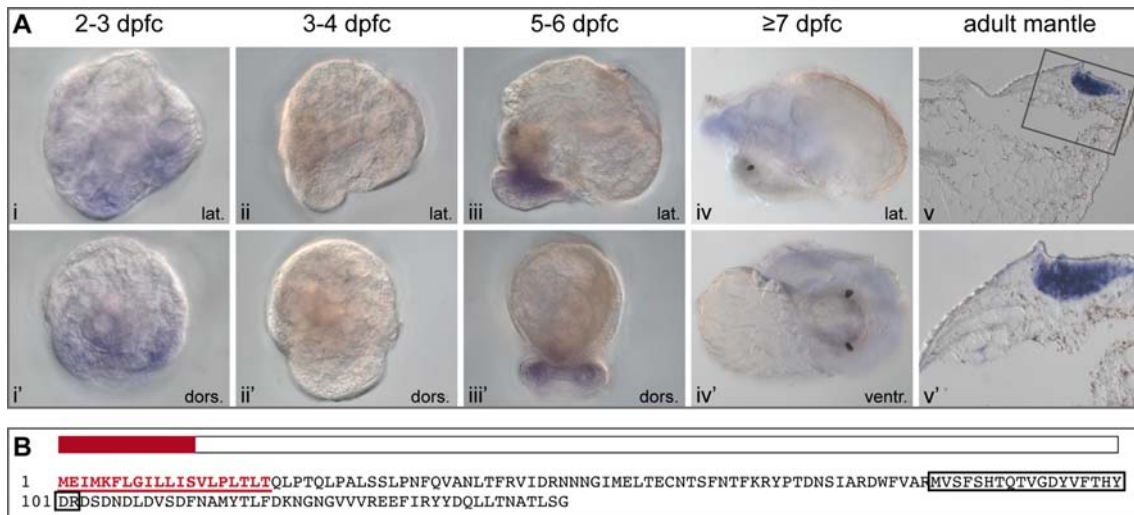
Additional file 5.13. Spatial expression and molecular features of *Lstag-sfc-12*. (A) *In situ* hybridisation against *Lstag-sfc-12* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iii, iv, iii', iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Incomplete protein sequence of the translated Mantle_Edge_v2_idb_103149. The peptide identified with MS/MS (boxed section) is indicated.
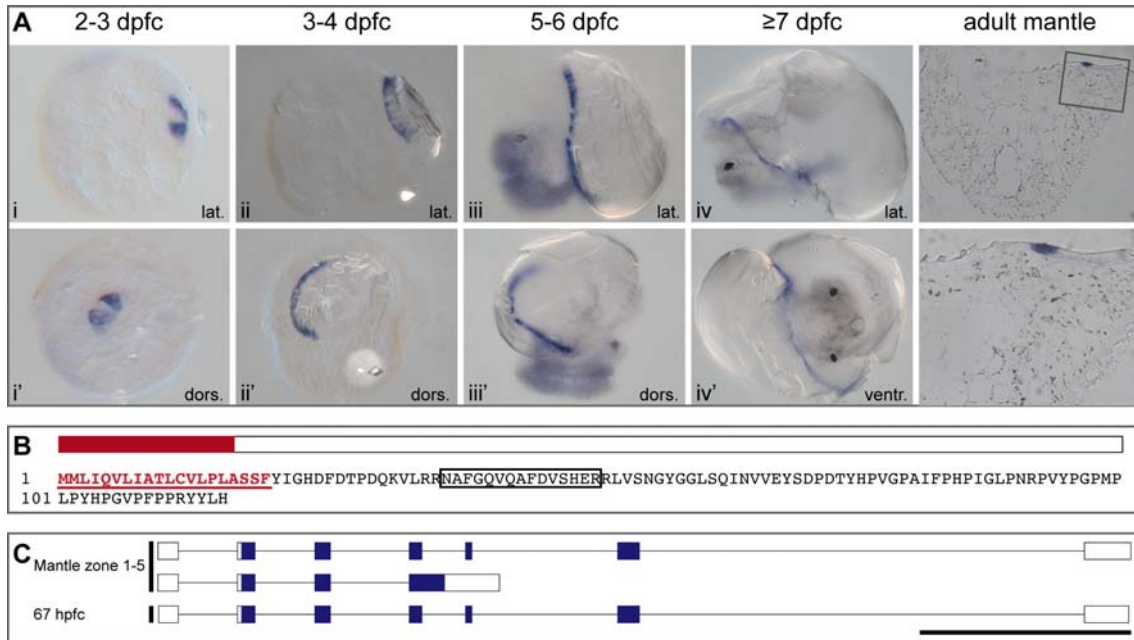
Additional file 5.14. Spatial expression and molecular features of *Lstag-sfc-13*. (A) *In situ* hybridisation against *Lstag-sfc-13* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Proximal_v2_idb_67016. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-13*. Transcriptomic contigs similar to Mantle_Proximal_v2_idb_67016 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf00172. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.

Additional file 5.15. Spatial expression and molecular features of *Lstag-sfc-14*. (A) *In situ* hybridisation against *Lstag-sfc-14* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone three of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_28903. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-14*. Transcriptomic contigs similar to Mantle_Edge_v2_Tri_28903 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf02426. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
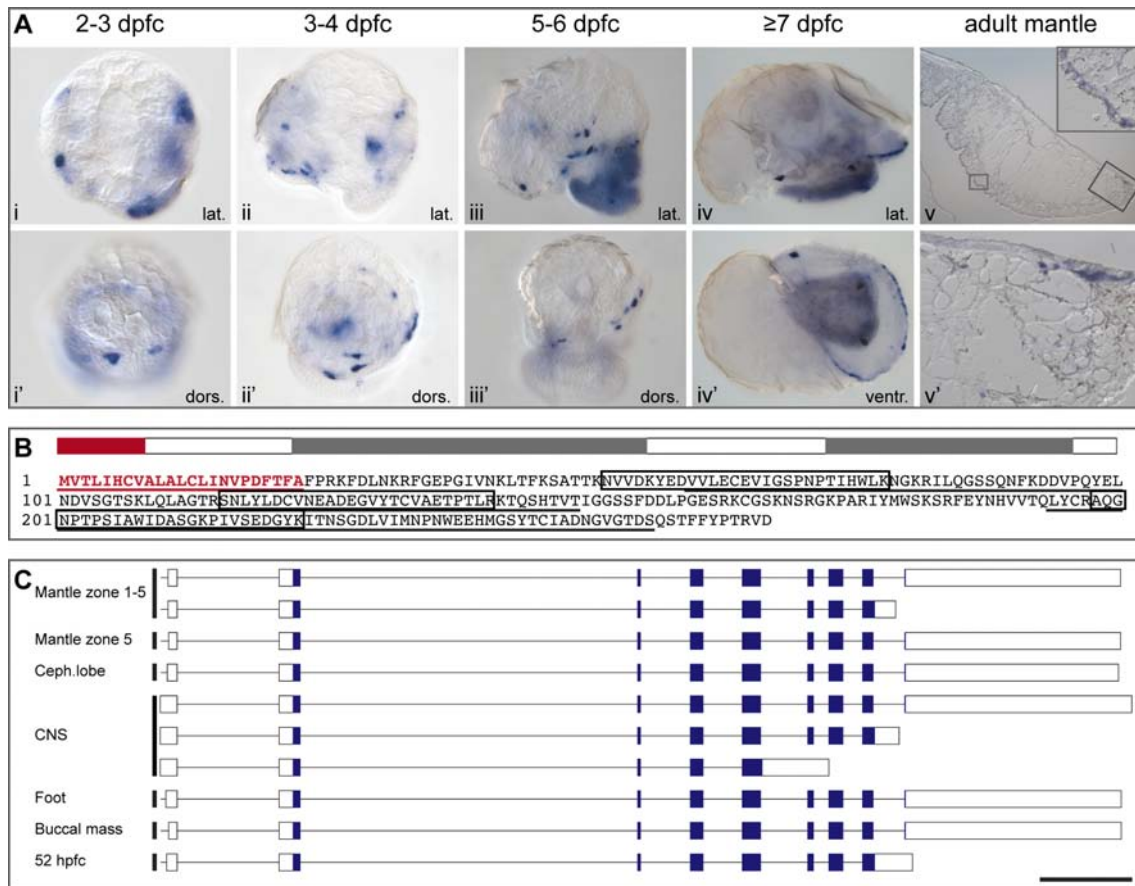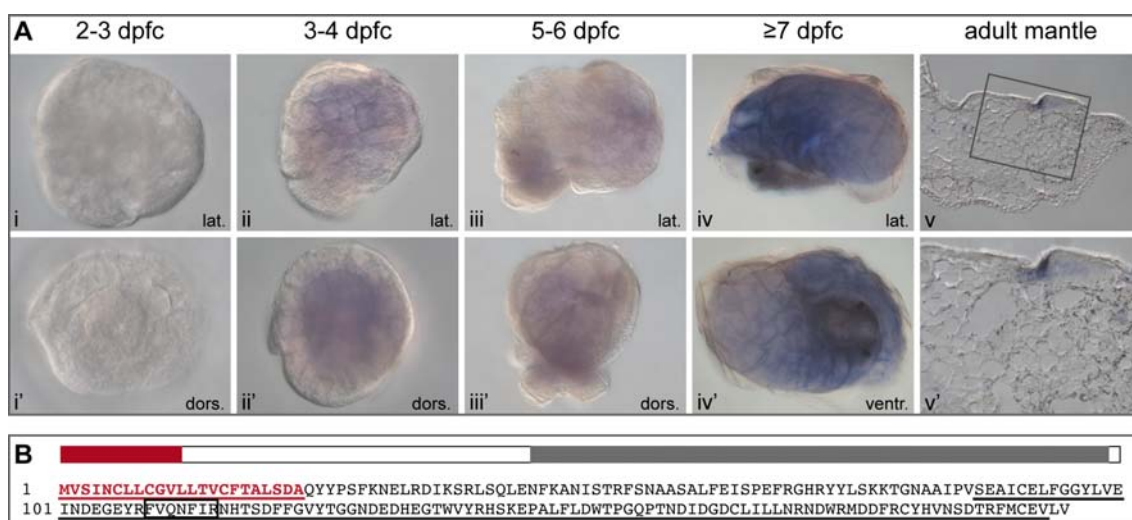
Additional file 5.16. Spatial expression and molecular features of *Lstag-sfc-15*. (A) *In situ* hybridisation against *Lstag-sfc-15* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (iv, iv') and the adult mantle belt (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_CLC_8979. The signal sequence (red) and the peptide identified with MS/MS (boxed section) are highlighted. (ç) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-15*. Transcriptomic contigs similar to Mantle_Edge_v2_CLC_8979 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01427. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.

Additional file 5.17. Spatial expression and molecular features of *Lstag-sfc-16*. (A) *In situ* hybridisation against *Lstag-sfc-16* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the adult mantle belt (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated CNS_v2_Tri_113901. The signal sequence (red) and the peptide identified with MS/MS (boxed section) are indicated.
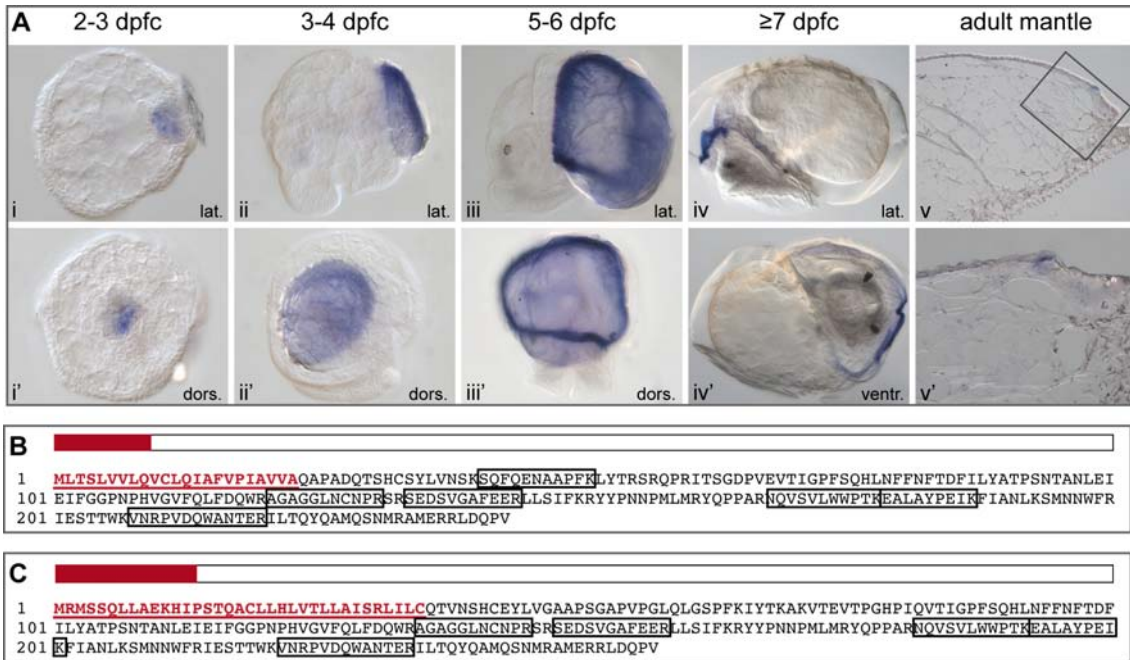
Additional file 5.18. Spatial expression and molecular features of *Lstag-sfc-17*. (A) *In situ* hybridisation against *Lstag-sfc-17* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the periphery of the shell forming tissue in larvae (i-iv, i'-iv') and zone four of the adult mantle (v, v'). (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_90504. The signal sequence (red) and the peptide identified with MS/MS (boxed section) are highlighted. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-17*. Transcriptomic contigs similar to _Mantle_Edge_v2_idb_90504 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01608. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
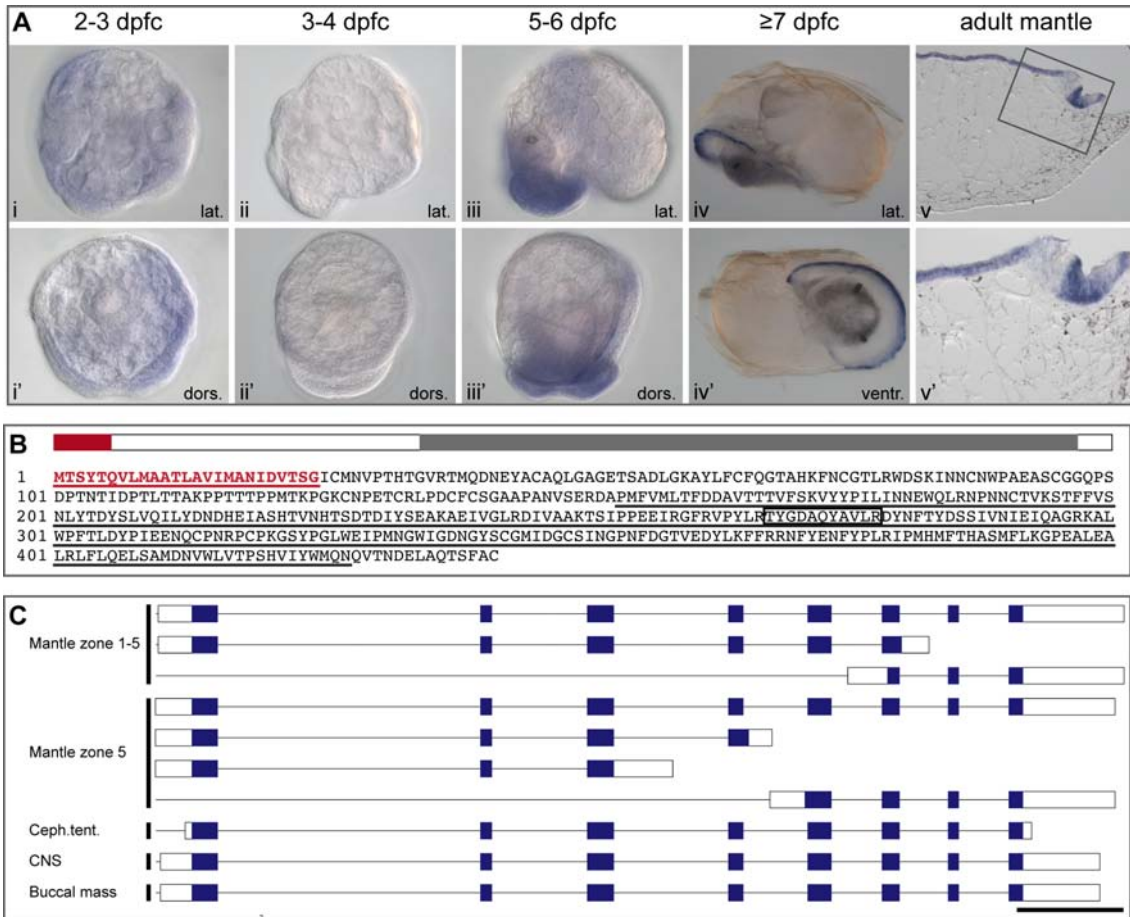
Additional file 5.19. Spatial expression and molecular features of *Lstag-sfc-18*. (A) *In situ* hybridisation against *Lstag-sfc-18* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in larvae (i-iv, i'-iv') and zones two, three and four of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_32853. The signal sequence (red), the two conserved immunoglobulin domains (underlined black and highlighted grey) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Schematic representation of the gene architecture and splice variation of the *Lstag-sfc-18*. Transcriptomic contigs similar to Mantle_Edge_v2_idb_32853 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01043. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
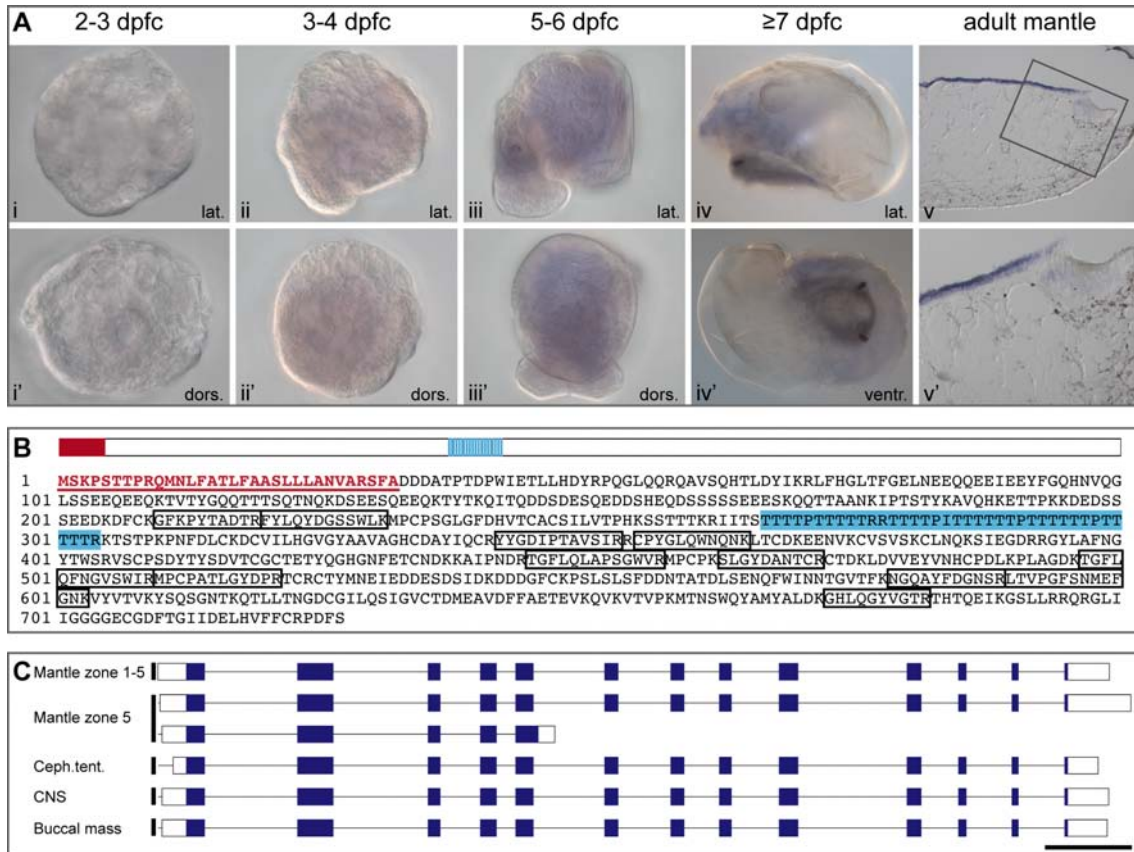
Appendix 1

Additional file 5.20. Spatial expression and molecular features of *Lstag-sfc-19*. (A) *In situ* hybridisation against *Lstag-sfc-19* during development and in the fully differentiated adult mantle. We interpret the staining in juvenile animal to be non-specific background. The blue staining indicates the expression in zone four of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_42787. The signal sequence (red), the conserved C-type lectin/C-type lectin-like domain (underlined black and highlighted grey) and the peptide identified with MS/MS (boxed section) are indicated.
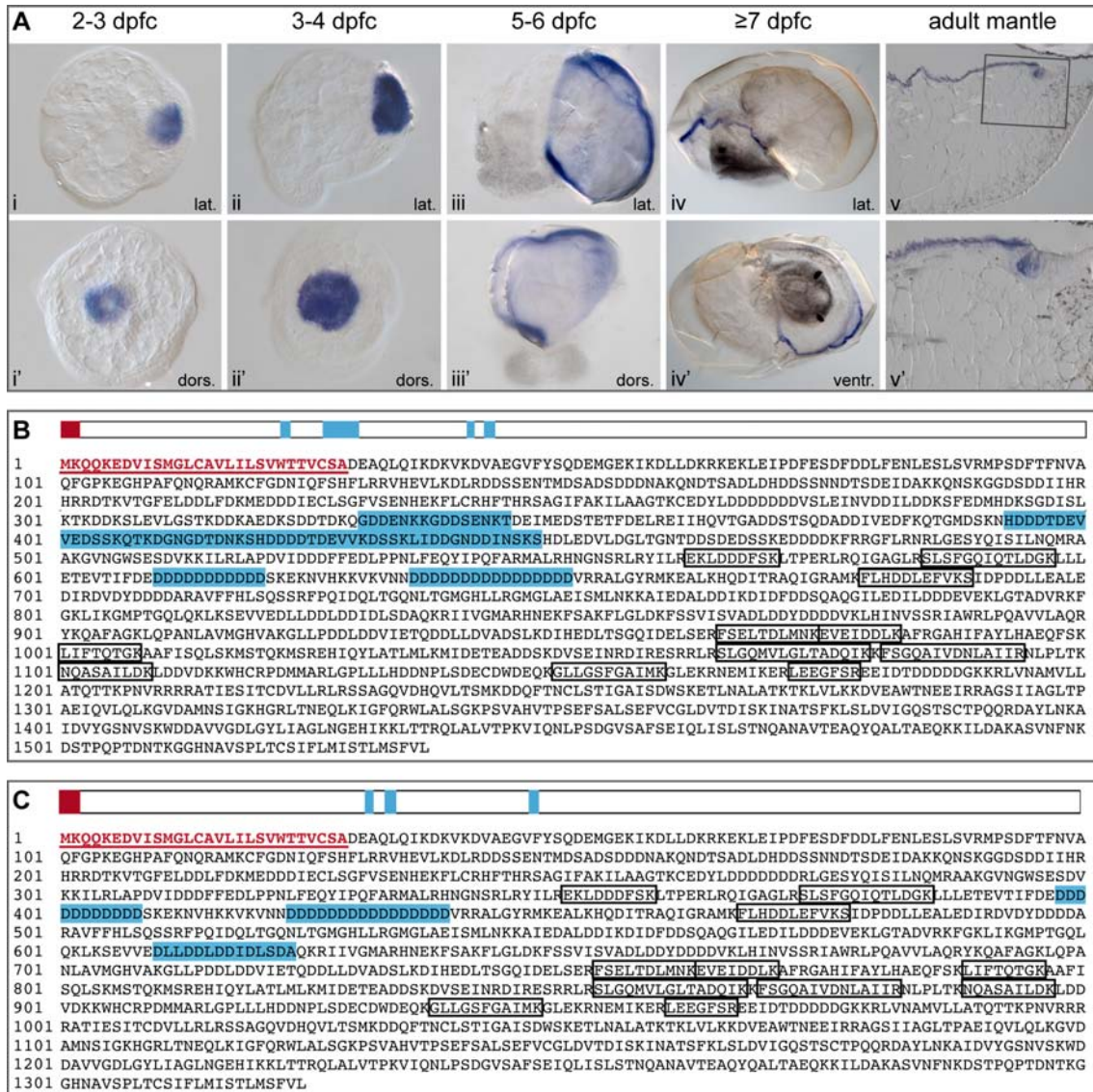
Additional file 5.21. Spatial expression and molecular features of *Lstag-sfc-20*. (A) *In situ* hybridisation against *Lstag-sfc-20* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone four of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_71081. The signal sequence (red) and the peptides identified with MS/MS (boxed sections) are highlighted. (C) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_71083. The signal sequence (red) and the peptides identified with MS/MS (boxed sections) are indicated.
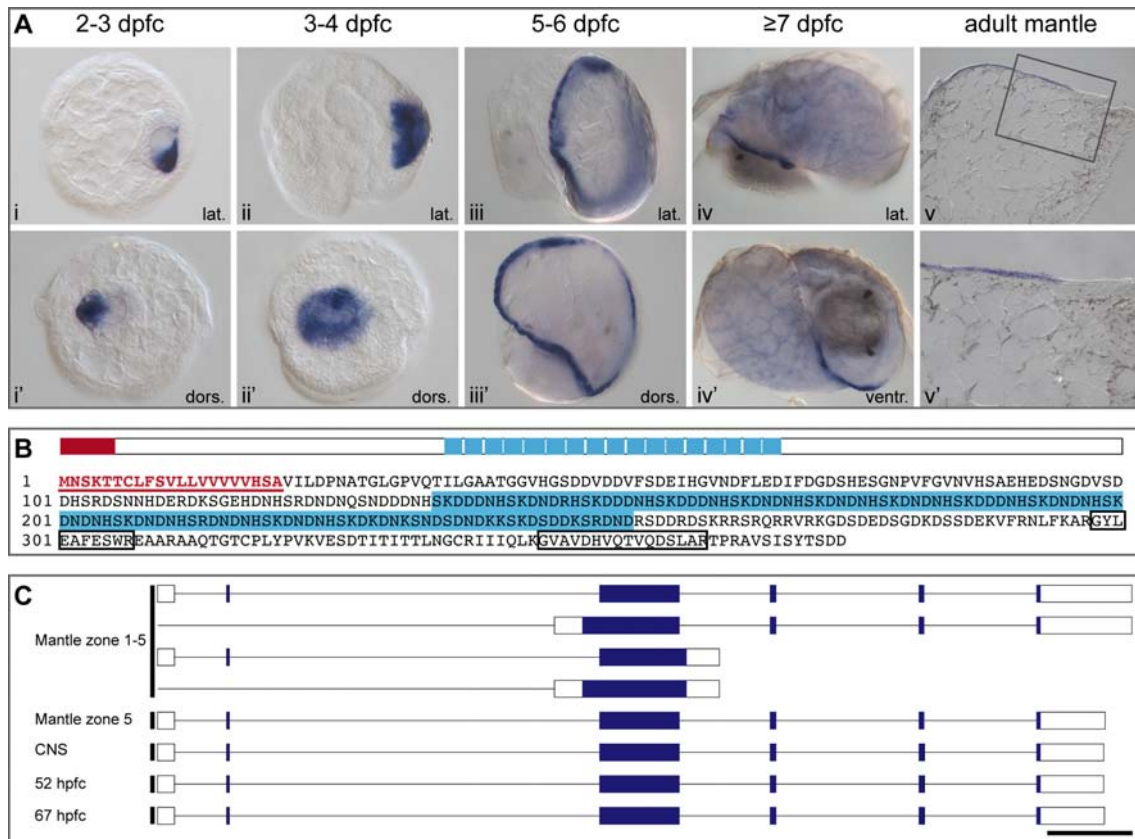
Additional file 5.22. Spatial expression and molecular features of *Lstag-sfc-21*. (A) *In situ* hybridisation against *Lstag-sfc-21* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the larval mantle edge (iv, iv') and zone one, two and five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_97925. The signal sequence (red), the putative catalytic domain of chitin deacetylase-like proteins (underlined black and highlighted grey) and the peptide identified with MS/MS (boxed section) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-21*. Transcriptomic contigs similar to Mantle_Edge_v2_idb_97925 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf00440. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
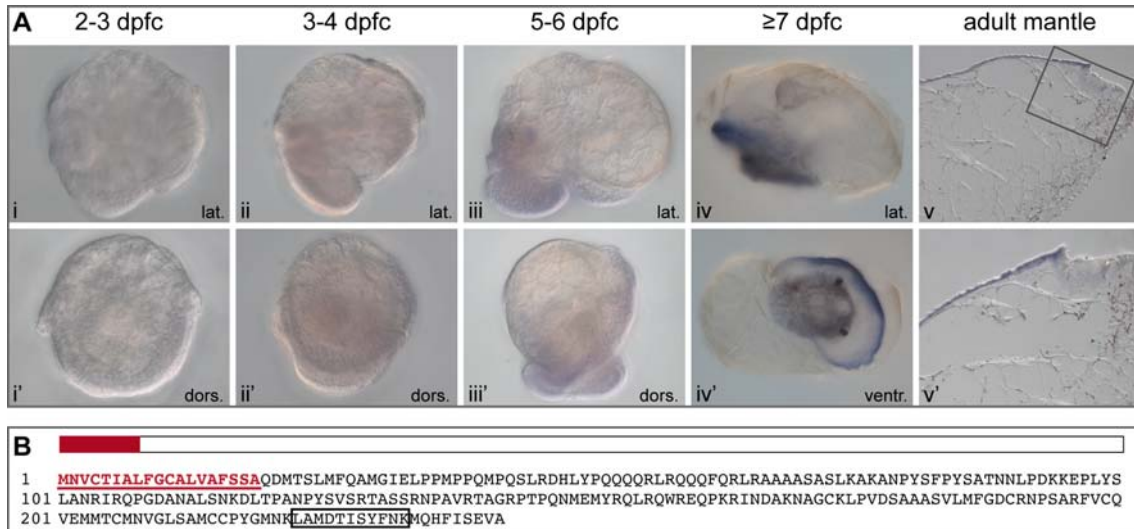
Additional file 5.23. Spatial expression and molecular features of *Lstag-sfc-22*. (A) *In situ* hybridisation against *Lstag-sfc-22* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_283967. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-22*. Transcriptomic contigs similar to Mantle_Edge_v2_Tri_283967 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf01583. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.
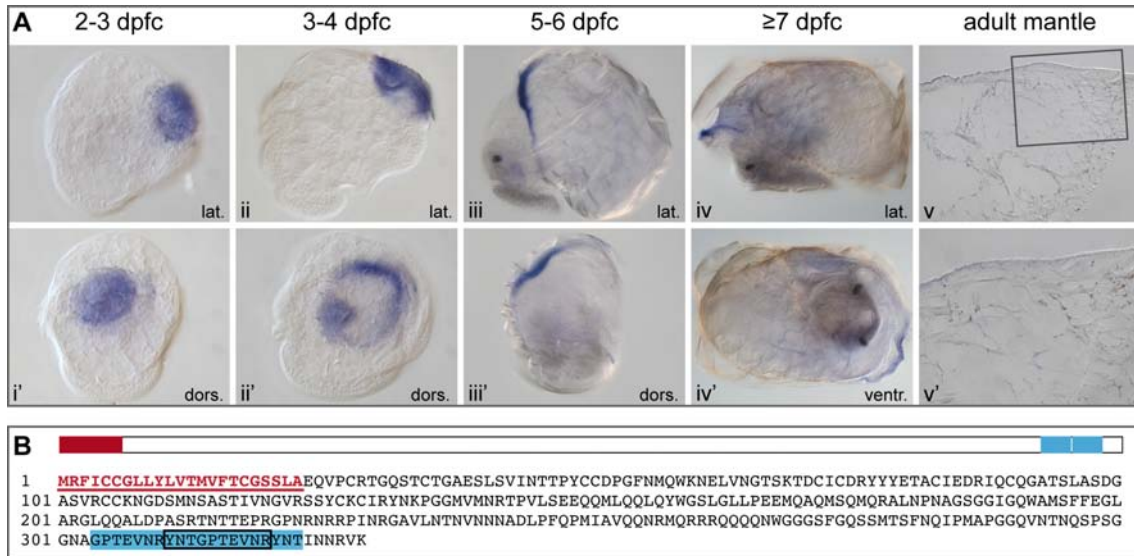
Additional file 5.24. Spatial expression and molecular features of *Lstag-sfc-23*. (A) *In situ* hybridisation against *Lstag-sfc-23* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zones three, four and five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_102017. The signal sequence (red), the repeated motifs (blue) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_102018. The signal sequence (red), the repeated motifs (blue) and the peptides identified with MS/MS (boxed sections) are indicated.
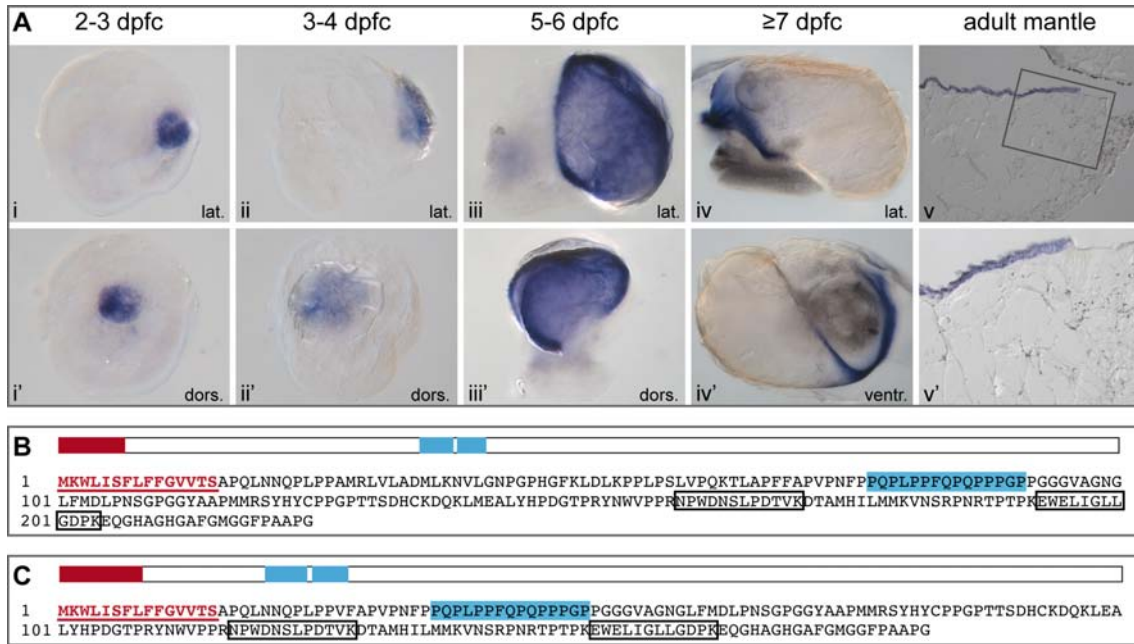
Additional file 5.25. Spatial expression and molecular features of *Lstag-sfc-24*. (A) *In situ* hybridisation against *Lstag-sfc-24* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_51055. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-24*. Transcriptomic contigs similar to Mantle_Edge_v2_idb_51055 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf00135. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.

Additional file 5.26. Spatial expression and molecular features of *Lstag-sfc*-25. (A) *In situ* hybridisation against *Lstag-sfc*-25 during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the larval mantle edge (iv, iv') and zone four and five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_44802. The signal sequence (red) and the peptide identified with MS/MS (boxed section) is highlighted.
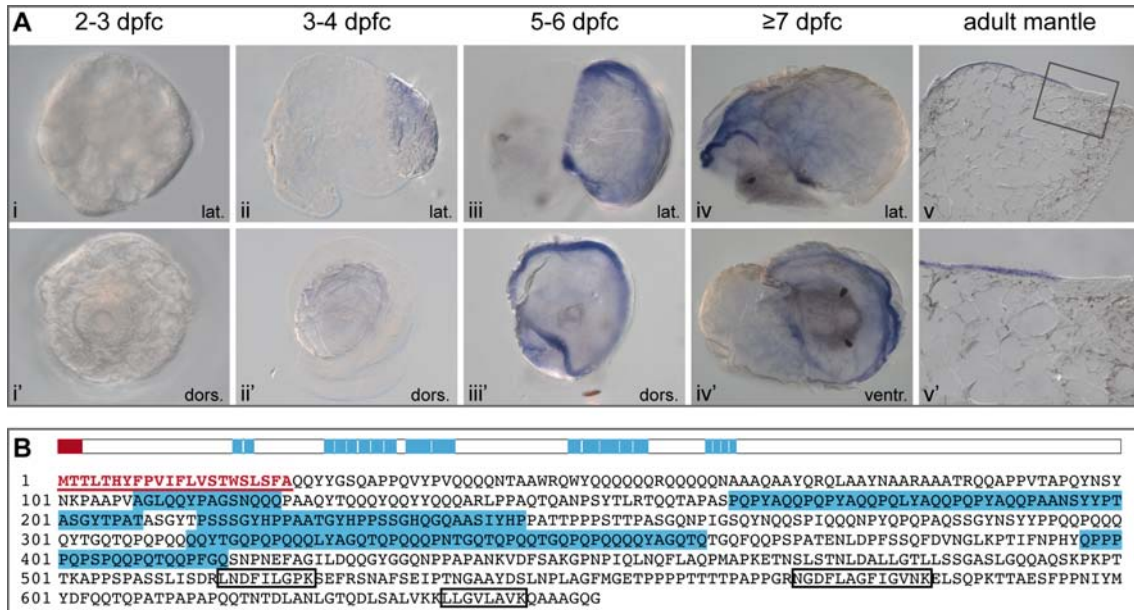
Additional file 5.27. Spatial expression and molecular features of *Lstag-sfc-26*. (A) *In situ* hybridisation against *Lstag-sfc-26* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_114584. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated.
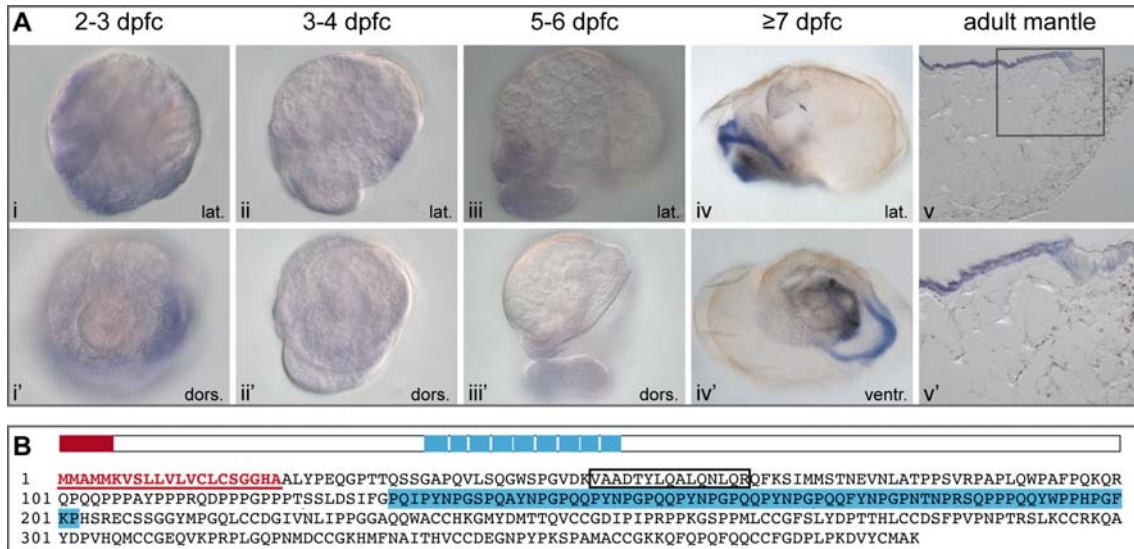
Additional file 5.28. Spatial expression and molecular features of *Lstag-sfc*-27. (A) *In situ* hybridisation against *Lstag-sfc*-27 during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (i-iv, i'-iv') and zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_25973. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_25972. The signal sequence (red), the repeated motif (blue) and the peptides identified with MS/MS (boxed sections) are indicated.
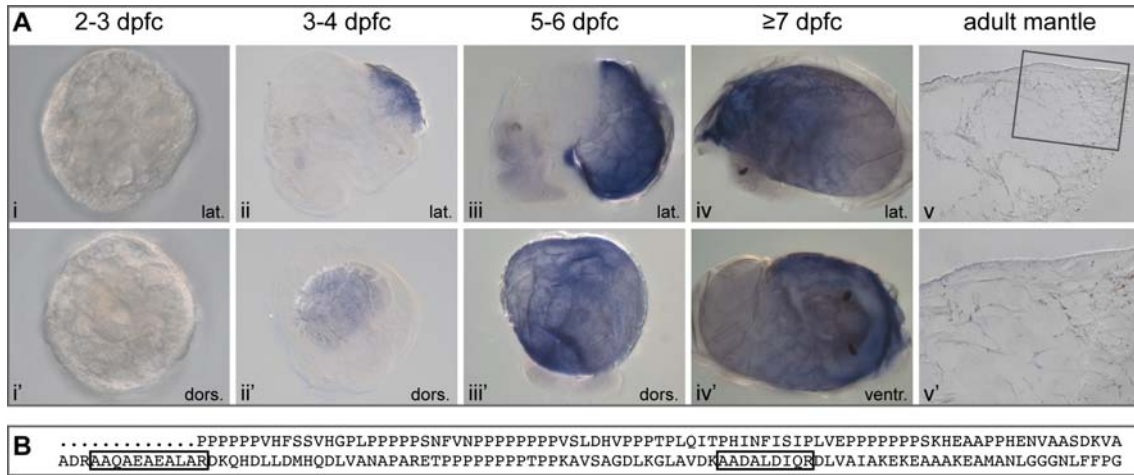
Additional file 29. Spatial expression and molecular features of *Lstag-sfc-28*. (A) *In situ* hybridisation against *Lstag-sfc-28* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (ii-iv, ii'-iv') and zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_25543. The signal sequence (red), the repeated motifs (blue) and the peptides identified with MS/MS (boxed sections) are indicated.
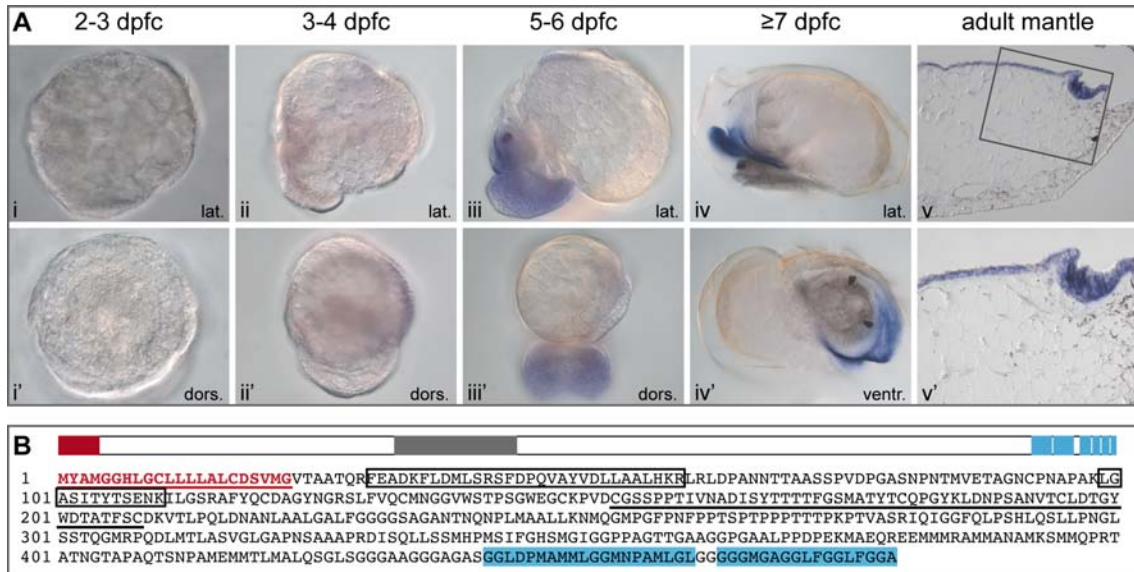
Additional file 5.30. Spatial expression and molecular features of _Lstag-sfc-29_. (A) _In situ_ hybridisation against _Lstag-sfc-29_ during development and in the fully differentiated adult mantle. blue staining indicates the expression in the larval mantle edge (iv, iv') and zone four and five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_75685. The signal sequence (red), the repeated motif (blue) and the peptide identified with MS/MS (boxed section) are indicated.
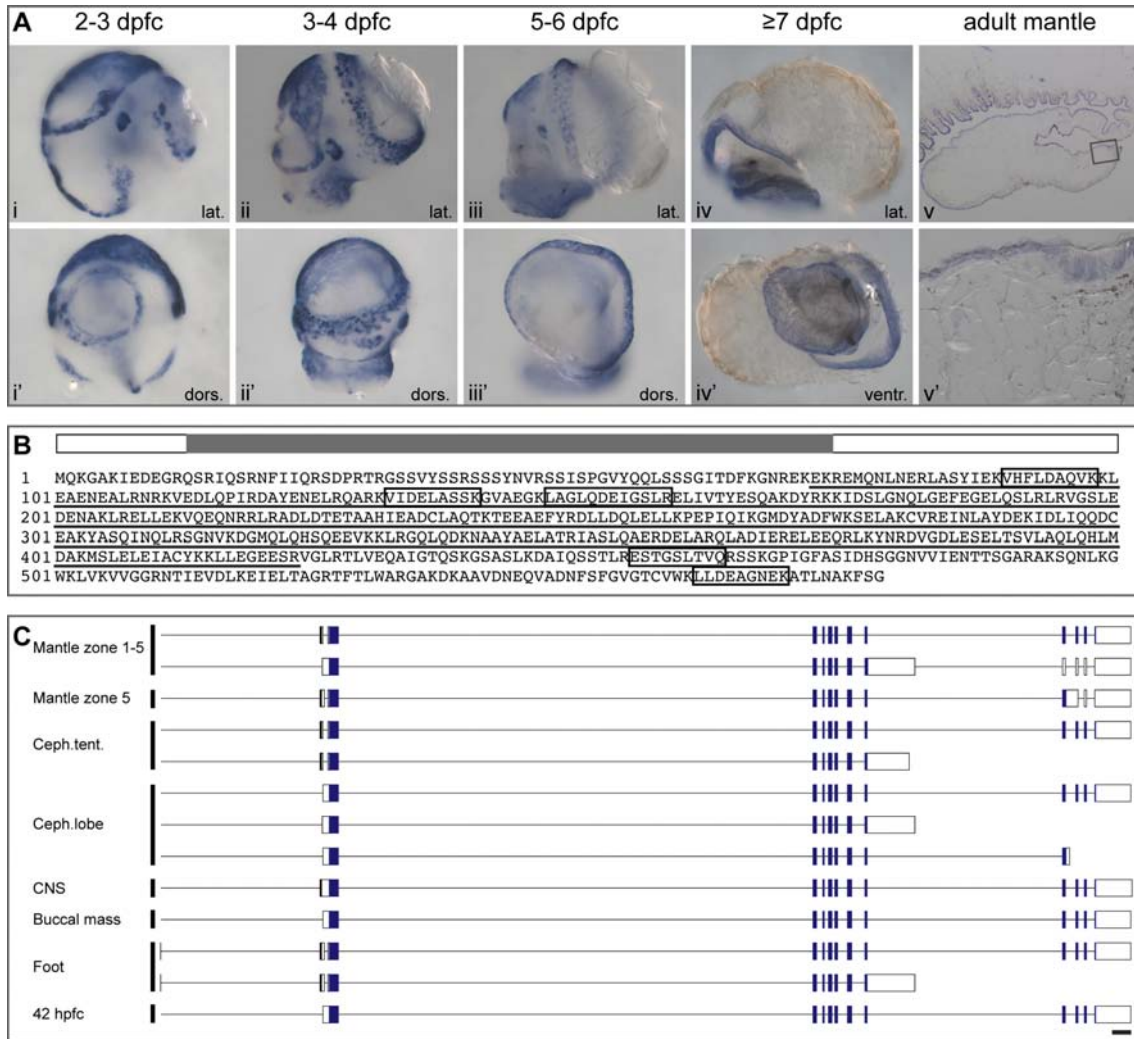
Additional file 5.31. Spatial expression and molecular features of *Lstag-sfc-30*. (A) *In situ* hybridisation against *Lstag-sfc-30* during development and in the fully differentiated adult mantle. The blue staining indicates the expression across the shell forming tissue in larvae (ii-iv, ii'-iv') and zone five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Incomplete protein sequence of the translated Mantle_Proximal_v2_Tri_90027. The peptides identified with MS/MS (boxed sections) are indicated.
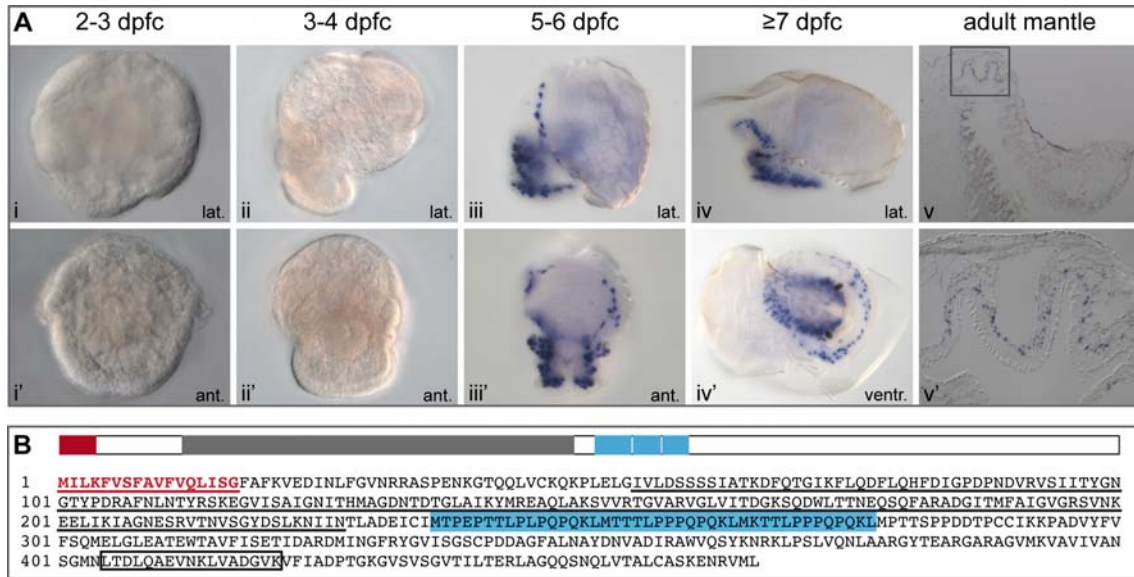
Additional file 5.32. Spatial expression and molecular features of *Lstag-sfc-31*. (A) *In situ* hybridisation against *Lstag-sfc-31* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in the larval mantle edge (iv, iv') and zone one to five of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_9904. The signal sequence (red), the repeated motifs (blue), the conserved sushi domain (underlined black and highlighted grey) and the peptides identified with MS/MS (boxed sections) are indicated.
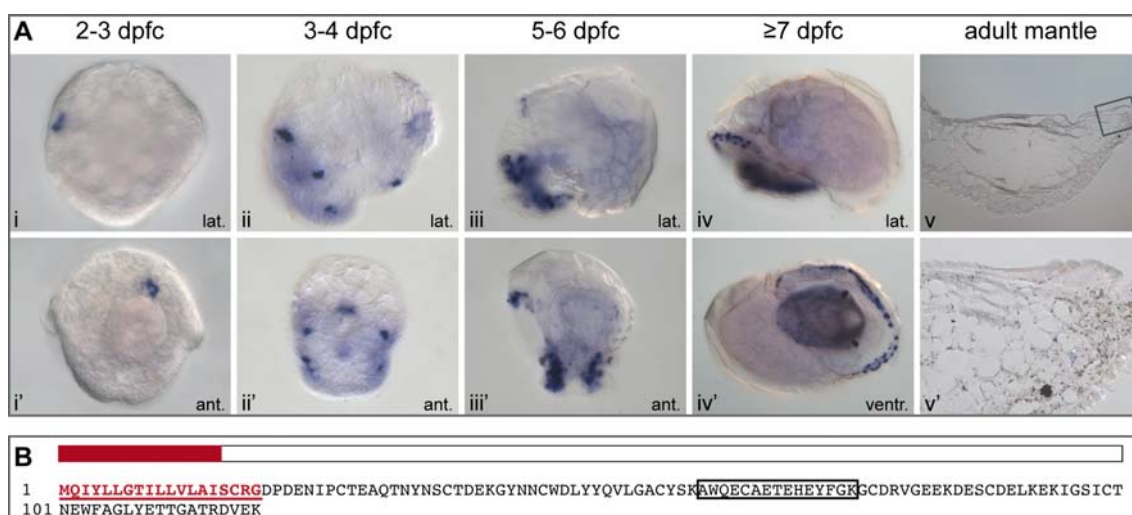
Additional file 5.33. Spatial expression and molecular features of *Lstag-sfc-32*. (A) *In situ* hybridisation against *Lstag-sfc-32* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in larvae (i-iv, i'-iv') and along the inner and outer epithelium of the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), dorsal (dors.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_Tri_206671. The conserved intermediate filament protein domain (underlined black and highlighted grey) and the peptides identified with MS/MS (boxed sections) are indicated. (C) Schematic representation of the gene architecture and splice variation of *Lstag-sfc-32*. Transcriptomic contigs similar to Mantle_Edge_v2_Tri_206671 were obtained from various *L. stagnalis* transcriptomes and aligned to the corresponding genomic contig gLs.1.0.scaf00652. The genomic contig is indicated by the horizontal black line, the aligned transcripts are indicated by the boxed sections, white boxes represent UTR region, blue boxes represent coding region. Scale bar represents 1000 bp.

Additional file 5.34. Spatial expression and molecular features of *Lstag-sfc-33*. (A) *In situ* hybridisation against *Lstag-sfc-33* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in larvae (iii, iv, iii', iv') and the adult mantle (v, v'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), anterior (ant.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_53476. The signal sequence (red), the repeated motif (blue) and the conserved von Willebrand factor type A domain (underlined black and highlighted grey) and the peptides identified with MS/MS (boxed sections) are indicated.

Additional file 5.35. Spatial expression and molecular features of *Lstag-sfc-34*. (A) *In situ* hybridisation against *Lstag-sfc-34* during development and in the fully differentiated adult mantle. The blue staining indicates the expression in larvae (i-iv, i'-iv'). Each column represents one developmental stage and shows two perspectives. The developmental stage is indicated in the header. The perspective is indicated in the lower right corner. v' shows a magnification of the boxed section in v. Days post first cleavage (dpfc), lateral (lat.), anterior (ant.), ventral (ventr.). (B) Protein sequence and schematic representation of the translated Mantle_Edge_v2_idb_111997. The signal sequence (red) and the peptide identified with MS/MS (boxed section) are highlighted.

Supplementary files Chapter 6

Supplementary S. 6.1: List of all primers used in chapter 6

Lst_c1222_OE_F1	AAAAAACCATGGGACAGTGGCCGACTCAAGGTGG
Lst_c1222_OE_R1	AAAAAACCATGGCGAAGAGGCTACTGGATGTTT
pRK793_Gib_F1	TAATAAGGATCCTCTAGAGTCGACC
pRK793_Gib_R1	ATGATGATGATGATGATGACCTTG
pRK793_Gib_c1222_F1	GGTCATCATCATCATCATCATCAGTGGCCGACTCAAGGTGGC
pRK793_Gib_c1222_R1	ACTCTAGAGGATCCTTATTAGAAGAGGCTACTGGATGTTTGGC
793+c1222_Gib_F1	CAGTGGCCGACTCAAGGTGGCTTCC
793+c1222_Gib_R1	AATCTATGGTCCTTGTTGGTGAAGT
His-MBP-TEVcl_F	ATGcatcatcatcatcatcatcatATGAAAATCGAAGAAGGTAA
His-MBP-TEVcl_R	ACCTTGAAAATAAAGATTTTCTCCCCTT
Gib_His-MBP_F1	ACCAACAAGGACCATAGATTATGCATCATCATCATCATCA
Gib_His-MBP_R1	CCACCTTGAGTCGGCCACTGACCTTGAAAATAAAGATTTT

Supplementary S. 6.2: Insert and protein sequences of chapter 6

>*Lstag-sfc-10*
AAAAAACCATGGACAGTGGCCGACTCAAGGTGGCTTCCCTGGAGGGCAGGGTGGCTTC
CCTGGAAGTCAGGGTGGCTTCCCCGGAGGTCCCGGTGGCTTTCCAGGAGGTCCCGGTGG
CTTCCCCGGAGGTCCCGGTGGCTTTCCCGGAGGCCCTGGCGGCTTTCCCGGGGGTCAAG
GAATGTTCCCTGGACAAGGATTTAATCCACTTTTCGGTAATGGACAACAAGGATTTCCA
CAACAATTCCAACCTCCTCAAATTTTGACCTGTATTGGAACTAACGAGCCCAAGATCA
TCTAAAAATCACCCTCAGACCCGAGCCAAGCAGTTTCACCTCCCAGAACCCAGCCTTCC
AGTTCGGACAGCCAAGGGAGCTCAACCGTCAAGACTGGAGAATCAGCGCCGTTTACATC
CCCGCCGCCAGCTACGGGGCCATCAGCCCATCAGGCACCAGTTCTCTCGTTGGTCAGTTC
TTCCTCGCCCTGACCCGCTACGGCCGCACTGACGGCAACTGTGCTGGTCTTGGTGGCAT
TCTCCAGAACGATGACCTCATCAACCAAGGGCAACAACCTAGAGGTCAGTTCTCCCAAT
TCAGTAACGGAATGTACGGCCAAACATCATATCCAGCTGGGTTCATCCAAGATCCCATC
GTCATCAGCCAAGGTGGCAACGTCTACTCCGGTGTTGTCAGAGATCTCTCCGAGGCTGA
CCTGAGGGGTCGTGGTGTCGCTATCTGCTTGGACTACTTGTGCCAACGTCCCACTACAA
CCTGCTGCTCTGTGGTCAAAGATTCAGTCCCCGCCACTGAAGTTGTTGGTGCTCCAACC
ATCGGCTCTTACTCAGGAAGTGCACTCCTCTCCGGATCCGGAGCTAAAGCCGGTACTAG
CACCGGAGGAGGTAACCTATTTGGTGGCTCCACCGGCCAAACATCCAGTAGCCTCTTCG
CCATGGTTTTTT

>Lstag-sfc-10
QWPTQGGFPGGQGGFPGSQGGFPGGPGGFPGGPGGFPGGPGGFPGGQGMFP
GQGFNPLFGNGQQGFPQQFQPPQILTCIGTNEAQDHLKITLRPEPSSFTSQNPAFQFGQP
RELNRQDWRISAVYIPAASYGAISPSGTSSLVGQFFLALTRYGRTDGNCAGLGGILQNDDL
INQGQQPRGQFSQFSNGMYGQTSYPAGFIQDPIVISQGGNVYSGVVRDLSEADLRGRGVAI
CLDYLCQRPTTTCCSVVKDSVPATEVVGAPTIGSYSGSALLSGSGAKAGTSTGGGNLFGGS
TGQTSSSLF

Number of amino acids: 310
Molecular weight: 31552.87 Da
Theoretical pI: 5.78

>*Lstag-sfc-10-His*
ATGGGACAGTGGCCGACTCAAGGTGGCTTCCCTGGAGGGCAGGGTGGCTTCCCTGGAAG
TCAGGGTGGCTTCCCCGGAGGTCCCGGTGGCTTTCCAGGAGGTCCCGGTGGCTTCCCCG
GAGGTCCCGGTGGCTTTCCCGGAGGCCCTGGCGGCTTTCCCGGGGGTCAAGGAATGTTC
CCTGGACAAGGATTTAATCCACTTTTCGGTAATGGACAACAAGGATTTCCACAACAATT
CCAACCTCCTCAAATTTTGACCTGTATTGGAACTAACGAGCCCAAGATCATCTAAAAA
TCACCCTCAGACCCGAGCCAAGCAGTTTCACCTCCCAGAACCCAGCCTTCCAGTTCGGAC
AGCCAAGGGAGCTCAACCGTCAAGACTGGAGAATCAGCGCCGTTTACATCCCCGCCGCC
AGCTACGGGGCCATCAGCCCATCAGGCACCAGTTCTCTCGTTGGTCAGTTCTTCCTCGC
CCTGACCCGCTACGGCCGCACTGACGGCAACTGTGCTGGTCTTGGTGGCATTCTCCAGA
ACGATGACCTCATCAACCAAGGGCAACAACCTAGAGGTCAGTTCTCCCAATTCAGTAAC
GGAATGTACGGCCAAACATCATATCCAGCTGGGTTCATCCAAGATCCCATCGTCATCAG
CCAAGGTGGCAACGTCTACTCCGGTGTTGTCAGAGATCTCTCCGAGGCTGACCTGAGGG
GTCGTGGTGTCGCTATCTGCTTGGACTACTTGTGCCAACGTCCCACTACAACCTGCTGC
TCTGTGGTCAAAGATTCAGTCCCCGCCACTGAAGTTGTTGGTGCTCCAACCATCGGCTC
TTACTCAGGAAGTGCACTCCTCTCCGGATCCGGAGCTAAAGCCGGTACTAGCACCGGAG

GAGGTAACCTATTTGGTGGCTCCACCGGCCAAACATCCAGTAGCCTCTTCGCCATGGGC
CATCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATAT
GCTCGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCG
CTGAGCAA

>Lstag-sfc-10-His
MGQWPTQGGFPGGQGGFPGSQGGFPGGPGGFPGGPGGFPGGPGGFPGGQGM
FPGQGFNPLFGNGQQGFPQQFQPPQILTCIGTNEAQDHLKITLRPEPSSFTSQNPAFQFG
QPRELNRQDWRISAVYIPAASYGAISPSGTSSLVGQFFLALTRYGRTDGNCAGLGGILQND
DLINQGQQPRGQFSQFSNGMYGQTSYPAGFIQDPIVISQGGNVYSGVVRDLSEADLRGRG
VAICLDYLCQRPTTTCCSVVKDSVPATEVVGAPTIGSYSGSALLSGSGAKAGTSTGGGNLF
GGSTGQTSSSLFAMGHHHHHHHHHHSSGHIEGRHMLEDPAANKARKEAELAAATAEQ

Number of amino acids: 357
Molecular weight: 36743.53 Da
Theoretical pI: 6.48

>*MBP-TEVcs-His-Lstag-sfc-10*
ATGAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACG
GTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGA
GCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTGCGGCAACTGGCGATGGCCCTG
ACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGGCCTGTTGGCT
GAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGT
ACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTT
ATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGAT
AAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACT
TCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAG
TACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGGGTCTGACCTTCCT
GGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGAA
GCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAA
CATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAAC
CATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAA
GAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGG
TTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAGTTGGC
GAAAGATCCACGTATTGCCGCCACCATGGAAAACGCCCAGAAAGGTGAAATCATGCCGA
ACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCC
AGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAA
CAACAACAACAATAACAATAACAACAACCTCGGGATCGAGGGAAGGGGAGAAAATCTT
TATTTTCAAGGTCATCATCATCATCATCATCATCAGTGGCCGACTCAAGGTGGCTTCCC
TGGAGGGCAGGGTGGCTTCCCTGGAAGTCAGGGTGGCTTCCCCGGAGGTCCCGGTGGCT
TTCCAGGAGGTCCCGGTGGCTTCCCCGGAGGTCCCGGTGGCTTTCCCGGAGGCCCTGGC
GGCTTTCCCGGGGGTCAAGGAATGTTCCCTGGACAAGGATTTAATCCACTTTTCGGTAA
TGGACAACAAGGATTTCCACAACAATTCCAACCTCCTCAAATTTTGACCTGTATTGGAA
CTAACGAGGCCCAAGATCATCTAAAAATCACCCTCAGACCCGAGCCAAGCAGTTTCACC
TCCCAGAACCCAGCCTTCCAGTTCGGACAGCCAAGGGAGCTCAACCGTCAAGACTGGAG
AATCAGCGCCGTTTACATCCCCGCCGCCAGCTACGGGGCCATCAGCCCATCAGGCACCA
GTTCTCTCGTTGGTCAGTTCTTCCTCGCCCTGACCCGCTACGGCCGCACTGACGGCAACT
GTGCTGGTCTTGGTGGCATTCTCCAGAACGATGACCTCATCAACCAAGGGCAACAACCT
AGAGGTCAGTTCTCCCAATTCAGTAACGGAATGTACGGCCAAACATCATATCCAGCTGG

...

GTTCATCCAAGATCCCATCGTCATCAGCCAAGGTGGCAACGTCTACTCCGGTGTTGTCA
GAGATCTCTCCGAGGCTGACCTGAGGGGTCGTGGTGTCGCTATCTGCTTGGACTACTTG
TGCCAACGTCCCACTACAACCTGCTGCTCTGTGGTCAAAGATTCAGTCCCCGCCACTGA
AGTTGTTGGTGCTCCAACCATCGGCTCTTACTCAGGAAGTGCACTCCTCTCCGGATCCG
GAGCTAAAGCCGGTACTAGCACCGGAGGAGGTAACCTATTTGGTGGCTCCACCGGCCAA
ACATCCAGTAGCCTCTTC

>MBP-TEVcs-His-Lstag-sfc-10
MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPD
IIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYN
KDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKY
DIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTINGPWAWS
NIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAV
NKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAAS
GRQTVDEALKDAQTNSSSNNNNNNNNNNLGIEGRGENLYFQGHHHHHHHQWPTQGG
FPGGQGGFPGSQGGFPGGPGGFPGGPGGFPGGPGGFPGGQGMFPGQGFNPLF
GNGQQGFPQQFQPPQILTCIGTNEAQDHLKITLRPEPSSFTSQNPAFQFGQPRELNRQD
WRISAVYIPAASYGAISPSGTSSLVGQFFLALTRYGRTDGNCAGLGGILQNDDLINQGQQP
RGQFSQFSNGMYGQTSYPAGFIQDPIVISQGGNVYSGVVRDLSEADLRGRGVAICLDYLCQ
RPTTTCCSVVKDSVPATEVVGAPTIGSYSGSALLSGSGAKAGTSTGGGNLFGGSTGQTSSS
LF

Number of amino acids: 712
Molecular weight: 75885.79 Da
Theoretical pI: 5.51

>*His-MBP-TEVcs-Lstag-sfc-10*
ATGCATCATCATCATCATCATCATATGAAAATCGAAGAAGGTAAACTGGTAATCTGGA
TTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAAGA
TACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAG
GTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGG
CTACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGC
TGTATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATC
GCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAAC
CTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTG
ATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGGTTA
TGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCT
GGCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACATACACATGAATGC
AGACACCGATTACTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCA
TCAACGGCCCGTGGGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACG
GTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGG
TATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGC
TGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCT
GAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACCATGGAAAAC
GCCCAGAAAGGTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGT
GCGTACTGCGGTGATCAACGCCGCCAGCGGTCGTCAGGCTGTCGATGAAGCCCTGAAAG
ACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAATAACAACAACCTCGGG
ATCGAGGGAAGGGGAGAAAATCTTTATTTTCAAGGTCAGTGGCCGACTCAAGGTGGCT
TCCCTGGAGGGCAGGGTGGCTTCCCTGGAAGTCAGGGTGGCTTCCCCGGAGGTCCCGGT

GGCTTTCCAGGAGGTCCCGGTGGCTTCCCCGGAGGTCCCGGTGGCTTTCCCGGAGGCCC
TGGCGGCTTTCCCGGGGGTCAAGGAATGTTCCCTGGACAAGGATTTAATCCACTTTTCG
GTAATGGACAACAAGGATTTCCACAACAATTCCAACCTCCTCAAATTTTGACCTGTATT
GGAACTAACGAGGCCCAAGATCATCTAAAAATCACCCTCAGACCCGAGCCAAGCAGTTT
CACCTCCCAGAACCCAGCCTTCCAGTTCGGACAGCCAAGGGAGCTCAACCGTCAAGACT
GGAGAATCAGCGCCGTTTACATCCCCGCCGCCAGCTACGGGGCCATCAGCCCATCAGGC
ACCAGTTCTCTCGTTGGTCAGTTCTTCCTCGCCCTGACCCGCTACGGCCGCACTGACGGC
AACTGTGCTGGTCTTGGTGGCATTCTCCAGAACGATGACCTCATCAACCAAGGGCAACA
ACCTAGAGGTCAGTTCTCCCAATTCAGTAACGGAATGTACGGCCAAACATCATATCCAG
CTGGGTTCATCCAAGATCCCATCGTCATCAGCCAAGGTGGCAACGTCTACTCCGGTGTT
GTCAGAGATCTCTCCGAGGCTGACCTGAGGGGTCGTGGTGTCGCTATCTGCTTGGACTA
CTTGTGCCAACGTCCCACTACAACCTGCTGCTCTGTGGTCAAAGATTCAGTCCCCGCCA
CTGAAGTTGTTGGTGCTCCAACCATCGGCTCTTACTCAGGAAGTGCACTCCTCTCCGGA
TCCGGAGCTAAAGCCGGTACTAGCACCGGAGGAGGTAACCTATTTGGTGGCTCCACCGG
CCAAACATCCAGTAGCCTCTTC

>His-MBP-TEVcs-Lstag-sfc-10
MHHHHHHHMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQ
VAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPI
AVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGGY
AFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNIHMNADTDYSIAEAAFNKGETAMTI
NGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLL
TDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAV
RTAVINAASGRQAVDEALKDAQTNSSSNNNNNNNNNNLGIEGRGENLYFQGQWPTQG
GFPGGQGGFPGSQGGFPGGPGGFPGGPGGFPGGPGGFPGGQGMFPGQGFNPL
FGNGQQGFPQQFQPPQILTCIGTNEAQDHLKITLRPEPSSFTSQNPAFQFGQPRELNRQD
WRISAVYIPAASYGAISPSGTSSLVGQFFLALTRYGRTDGNCAGLGGILQNDDLINQGQQP
RGQFSQFSNGMYGQTSYPAGFIQDPIVISQGGNVYSGVVRDLSEADLRGRGVAICLDYLCQ
RPTTTCCSVVKDSVPATEVVGAPTIGSYSGSALLSGSGAKAGTSTGGGNLFGGSTGQTSSS
LF

Number of amino acids: 713
Molecular weight: 75971.94 Da
Theoretical pI: 5.44