2013

Suppressive subtractive hybridisation transcriptomics provides a novel insight into the functional role of the hypobranchial gland in a marine mollusc

Patrick W. Laffy
Flinders University

Kirsten Benkendorff
Southern Cross University

Catherine A. Abbott
Flinders University

Publication details

Post print of: Laffy, PW, Benkendorff, K & Abbott, CA 2013, 'Suppressive subtractive hybridisation transcriptomics provides a novel insight into the functional role of the hypobranchial gland in a marine mollusc', *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, vol. 8, no. 2, pp. 111-122.
Published version available from: http://dx.doi.org/10.1016/j.cbd.2013.01.001

Suppressive subtractive hybridization transcriptomics provides a novel insight into the functional role of the hypobranchial gland in a marine mollusc

Patrick W. Laffy[1.], Kirsten Benkendorff[2.] and Catherine A. Abbott[1.].

1. School of Biological Sciences, Flinders University, GPO Box 2100 Adelaide SA 5001, Australia.
2. Marine Ecology Research Centre, School of Environment, Science and Engineering, Southern Cross University, PO box 157, Lismore, NSW 2480, Australia

Corresponding Author:
Kirsten Benkendorff
Email: kirsten.benkendorff@scu.edu.au
Telephone: +61 2 6620 3755
Fax: +61 2 6621 2669

Additional Author email addresses:

Cathy.abbott@flinders.edu.au
Patrick.laffy@flinders.edu.au

Abstract

The hypobranchial gland present in gastropods is an organ whose function is not clearly understood. Involved in mucus production, within members of the family Muricidae it is also the source of the ancient dye Tyrian purple and its bioactive precursors. To gain further insights into hypobranchial gland biology, suppressive subtractive hybridisation was performed on hypobranchial gland and mantle tissue from the marine snail *Dicathais orbita* creating a differentially expressed cDNA library. 437 clones were randomly sequenced, analysed and annotated and 110 sequences had their functions putatively identified. Importantly this approach identified a putative gene involved in Tyrian purple biosynthesis, an arylsulfatase gene. Confirmation of the upregulation of arylsulfatase in the hypobranchial gland compared to the mantle was demonstrated using quantitative real-time PCR. Other genes identified as playing an important role in the hypobranchial gland were those involved in mucus protein synthesis, choline ester regulation, protein and energy production. This study confirms that the hypobranchial gland is involved in the production of mucus secretion and also identifies it as a site of chemical interaction and biosynthesis. This study lays the foundation for a better understanding of the enzymatic production of Tyrian purple precursors within the gland.

Keywords

Tyrian purple, mollusc, transcriptome, hypobranchial gland, EST library.

1. Introduction

The function of the hypobranchial gland, present within the mantle cavity of gastropod molluscs, has been the subject of investigation and speculation for many years. It is known that the hypobranchial gland is responsible for the amalgamation of particulate matter within the mantle cavity and it has been characterized as a predominant tissue involved in mucus production (Fretter and Graham, 1994). Further studies into the gland suggest varied roles for this pallial organ, including possible roles in aestivation and antiseptic production (Andrews and Little, 1972), olfactory response during feeding (Roller et al., 1995), the production of pheromones involved

in aggregation and spawning behaviour (Kuanpradit et al., 2010) and in the production of the embedding medium into which egg capsules are deposited onto benthic substrata (Westley, 2008). Several secondary metabolites, including pharmacologically active choline esters (Baker and Duke, 1976; Roseghini et al., 1995; Whittaker, 1960), brominated indole derivatives (Kelley et al., 2003; Naegel and Cooksey, 2002; Westley and Benkendorff, 2008) and an attractin-like protein (Kuanpradit et al., 2010) have been extracted from this gland. Furthermore, arylsulfatase and bromoperoxidase enzyme activity has been identified within the gland in the Muricidae (Erspamer, 1946; Jannun and Coe, 1987; Westley, 2008; Westley and Benkendorff, 2009).

In the Muricidae family of neogastropods, the hypobranchial gland is well known as a source of the ancient dye Tyrian purple (Baker, 1974; Cooksey, 2001). This pigment is a brominated indole that was first identified in 1909 by Friedlander (Friedlander, 1909). The dye is produced from tryptophan-derived indoxyl sulphate prochromagens, which are hydrolysed by arylsulfatase enzymes (Baker and Duke, 1976; Cooksey, 2001; Naegel and Cooksey, 2002) to yield intermediate precursors that exhibit antibacterial and anticancer activities (Benkendorff et al., 2000; Benkendorff et al., 2001b; Benkendorff et al., 2011; Edwards et al., 2012; Westley et al., 2006). *Dicathais orbita* (Figure 1A) is an ideal model species in which to investigate gene expression of the hypobranchial gland, due to its well documented production of bioactive compounds. These compounds are also found in the egg masses (Figure 1A) and their associated biological activity has led to suggestions that hypobranchial gland secretions play a role in maternal defence of the egg masses (Benkendorff et al., 2000, 2001a), as well as an immune role within adult snails (Westley et al., 2006). Bromination of secondary metabolites has also been associated with the hypobranchial gland in other families of molluscs, including at least one species in the distantly related Vetigastropoda. 6-Bromo-2-mercaptoptrytamine, isolated from the hypobranchial gland of the vetigastropod *Calliostoma canaliculatum*, has been characterized as a potent neurotoxin, targeting potassium channels (Kelley et al., 2003; Wolters et al., 2005). These secondary metabolites within the hypobranchial gland of distinct gastropod lineages suggests that one of the roles of the gland is the production of bioactive secondary metabolites, which may have key roles in the chemical defence of molluscs.

A modern approach for investigating biosynthesis and the functional role of the hypobranchial gland is to examine gene expression. Molluscs have been used historically as model organisms to study the central nervous system and several transcriptome studies investigating neuronal interactions, chronic pain and neurological disorders have been performed (Feng et al., 2009; Moroz et al., 2006; Moroz et al., 2004; Walters and Moroz, 2009). More recently, the transcriptome of several gastropods have been investigated, at the embryonic, larval and metamorphic developmental stages, in order to better understand cellular differentiation and embryogenesis (Henry et al., 2010; Heyland et al., 2011; Jackson and Degnan, 2006). Transcriptomics has also been applied to the investigation of biomineralization, shell development and deposition in molluscs (Bai et al., 2010; Berland et al., 2011; Clark et al., 2010; Fang et al., 2011; Jackson and Degnan, 2006; Joubert et al., 2010). Most recently an extensive transcriptome analysis of *Thais clavigera* using pyrosequencing has been used to identify selected genes involved in the innate immune system of this mollusc (Rhee *et al*., 2012). In all of these studies, transcriptomics has been used as a tool to better understand biological processes and has helped resolve how global gene expression influences physiological changes.

Suppressive subtractive hybridization (SSH) is a powerful technique that allows the rapid and cost effective investigation of gene expression between different tissue types or at different stages of development and produces a library of differentially expressed genes. It was first developed by Diatchenko *et. al.* (1999) and has been used for a variety of different investigations. In particular it has been used to examine the role of the cerebral ganglia in the growth, feeding behaviour and reproduction of the tropical abalone *Haliotis asinina* (York et al., 2010) and the hibernating brain of the greater horseshoe bat *Rhinolophus ferrumequinum* (Yuan et al., 2008). In these two examples, SSH has been a very useful transcriptomics tool that has allowed further understanding of the biology of these species. So far the only known genes that have been identified from *D. orbita* are highly conserved sequences used in phylogenetic studies (Colgan et al., 2003; Colgan et al., 2007; Colgan et al., 2000; Laffy et al., 2009). Thus the use of SSH in this study should allow us to identify functional genes within the hypobranchial gland, to identify genes that may be involved in Tyrian

purple synthesis and to expand on our general knowledge of this gland in gastropod molluscs.

The mantle tissue, which surrounds the mantle cavity and protects the visceral mass, including the hypobranchial gland, was used as the subtractor (driver) in this SSH study. The epithelium of the mantle tissue in gastropod molluscs is responsible for shell biomineralisation and thus the mantle has been the focus of several gene expression studies in gastropods (Jackson and Degnan, 2006; McDougall et al., 2011; Nagai et al., 2007). More importantly, histochemistry and chemical extractions demonstrate the mantle does not produce Tyrian purple (Westley and Benkendorff, 2008; Westley et al., 2010). We hypothesize that a number of different genes will be identified from a SSH cDNA library containing expressed sequence tags (EST) that are differentially or uniquely expressed within the hypobranchial gland, relative to the mantle tissue of *D. orbita*, corresponding to the roles this gland plays in organismal cell biology. Quantitative real-time PCR was applied to a subset of the identified genes to confirm differential expression.

2. Methods

2.1 Tissue collection and RNA isolation

A study investigating the chemical composition of Tyrian purple precursors within *D. orbita* identified that there are distinct differences in precursor composition between the hypobranchial glands of male and female individuals (Westley and Benkendorff, 2008). In order to minimise the potential gene expression differences that may be involved in this sexual dimorphism, our study focussed only on the gene expression within female snails. Two female *D. orbita* individuals (e.g. Figure 1A) were obtained from captive populations originally collected from shallow subtidal reefs along the coast of the Fleurieu and Eyre Peninsulas, South Australia. Dissection was performed as described in Westley and Benkendorff (2008). A bench top vice was used to crack the shell and the soft tissue was removed by severing the columellar muscle from the shell. Soft tissue was transferred to a dissecting tray where the visceral mass was separated from the dorsal mantle by making an incision along the lateral margins of the columellar muscle. A clean section of dorsal mantle tissue was collected. The remaining dorsal mantle was folded back and pinned in place to reveal the pallial

gonoduct and the hypobranchial gland (Figure 1B). The egg capsule glands were then dissected away from the hypobranchial gland and the branchial and medial regions of this gland (Westley and Benkendorff, 2008) were excised and collected. All tissue samples were stored in cryovials, snap frozen in liquid nitrogen and stored at -80°C until required.

Total RNA was extracted by combining the hypobranchial tissue from two individual snails and using the Ambion RNaqueous RNA extraction kit (Applied Biosystems, Foster City, CA). Separately, the mantle tissue was combined from the same two snails for RNA extraction. Pooled tissue was immersed in 600µl of lysis binding solution on ice and homogenized using the Turrat T8 homogenizer (IKA-Werke, Germany) before RNA was extracted, in accordance with manufacturers' instructions with the following two modifications. Firstly, 700 µl of TRI reagent (ABgene, UK) was added to tissue in the lysis binding buffer, mixed by pipetting and then a 12,000g centrifugation for 10 min at 4°C was performed. This step improved the overall quality of the RNA as it helped to remove the mucus from the snail tissue before RNA extraction. The aqueous RNA containing fraction was then loaded on to a column as per manufacturers' instructions. The second alteration was that an on-column DNAse I digestion was performed between the second and third column wash steps, using 50 units of DNase I (Invitrogen, Carlsbad, CA), 18 µl DNase I buffer and 161 µl diethyl pyrocarbonate (DEPC) treated water. Columns were incubated at room temperature for 15 min before centrifugation for 30 sec at 14,000g. RNA was precipitated by treatment with 8M ammonium acetate and 100% ethanol. After overnight incubation at -20°C, the RNA was pelleted at 14,000 rpm for 20 min. The supernatant was removed and the pellet was washed with 75% ethanol to remove residual salts. Finally the RNA pellet was spun at 14,000 rpm for 10 mins, the supernatant was removed and the pellet was air-dried. RNA was rehydrated in DEPC treated water, aliquoted and stored at -80°C. Total RNA quality and integrity was assessed by agarose gel electrophoresis and spectrophotometric quantification of 260/280 ratio using the Nanodrop 1000 spectrophotometer (NanoDrop, Thermo Scientific).

2.2 Creation of subtracted cDNA library, plasmid isolation and sequencing

The SMART™ cDNA Library construction kit (Clontech, Mountain View, CA) and the Clontech PCR-Select™ cDNA subtraction kit (Clontech) were used to construct

our cDNA library containing unique and differentially expressed genes from the hypobranchial gland of *D. orbita*, in accordance with the relevant product manuals using mantle tissue as the driver. Two μg of total RNA was used in the SMART™ cDNA synthesis reaction, where mRNA sequences were used as templates to make cDNA. Final suppressive subtractive hybridization reactions were set up following the commercial protocol and resulted in driver to tester ratios of 25:1, essential for the production of true differentially expressed cDNA libraries. Our subtracted cDNA population was cloned into the pGEM®-T Easy vector system (Promega, Madison, WI) using the manufacturer's protocol. White colonies were selected and grown overnight in Luria Bertani media containing ampicillin. Colonies were stored at -80°C in 25% glycerol.

Plasmid DNA was purified at either the Australian Genome Research Facility (AGRF) (Brisbane, Australia) or using Wizard Plus SV minipreps DNA Purification system (Promega). Sequencing was performed at AGRF or Flinders and Southpath sequencing facility (Bedford Park, South Australia) using Big Dye terminator chemistry (Perkin Elmer, Waltham, MA, USA) and either M13F or M13R primer. The subsequent sequences were determined using either an ABI 3730xl 96-capillary automated DNA sequencer or an ABI 3100 sequence analyser.

2.3 EST processing, contig assembly and preliminary analysis

Vector sequence and adaptor regions were removed from our ESTs and clustered using Sequencher Version 4.1.4. (Genecodes, Ann Arbour, MI), producing a non-redundant set of EST sequences from the hypobranchial gland of *D. orbita.* Sequence assembly was performed to identify the amount of overlap in our sequences (Table 1). Perl scripts were used to manipulate the formatting of sequences before BLAST analysis was performed on each EST. ESTs were first manually analysed using the National Centre for Biotechnology (NCBI) nucleotide and protein database searches on the South Australian Partnership for Advanced Computing (SAPAC) BLAST portal, BLAST family of programs and classified in accordance with Gene Ontology (GO) defined gene functions. Manual sequence classifications were assigned, where corresponding tBLASTx sequence matches displayed 1xE values less than or equal to $1xE^{-05}$. tBLASTx analysis was performed using standard nuclear translation, before *D. orbita* sequences underwent automated sequence annotation using the program

BLAST2GO, maintaining a cutoff threshold of 1xE value less than or equal to $1xE^{-05}$ (Conesa et al., 2005). Assembled non-redundant sequence data cannot be submitted to the NCBI EST database therefore all subsequent sequence analysis was performed on processed unassembled sequences that were deposited into Genbank (accession numbers **GD253659-GD254033**, **J455778-FJ455785**, **GE905829-GE905839** and **FJ476128-FJ476170)**.

EST sequences were also compared using BLASTn analysis to the 20 most-sequenced molluscs in the EST database on NCBI as of May 2012, representing members of the taxonomic classes Gastropoda, Bivalvia and Cephalopoda. BLASTn analyses were performed on all transcripts, in addition to BLAST2GO nucleotide analysis. It should be noted that *Ilyanassa obsoleta* was the only neogastropod present in this dataset. The *Lottia gigantea* genome (Department of Energy website, http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html), the only available molluscan genome at the time of analysis, was also used in BLAST analysis in order to identify homologous sequences. While 454 pyrosequence data from *Concholepas concholepas* has been produced (Cardenas et al., 2011), this data remains largely unassembled and unannotated (average read length 160-252bp) and was not used in this study.

2.4 Gene Ontology, KEGG (Kyoto Encyclopedia of Genes and Genomes) enzymatic pathway and InterPro annotation

D. orbita sequences were annotated in May 2012, using the program BLAST2GO (Conesa et al., 2005). BLAST2GO uses GO terms (http://www.geneontology.org/) to describe biological processes, molecular function and cellular component of each gene identified. Using the NCBI BLAST server, BLASTx analysis was performed using the default parameters. For each sequence, the top 20 BLAST hits were analysed, and an E value threshold was set, discounting all BLASTx matches with 1xE values greater than $1xE^{-03}$. Furthermore, sequences less than 100 nucleotides were eliminated in this initial analysis to maximise the accuracy of annotation. Biological processes, molecular function and cellular component were assigned to sequences with significant sequence similarity, based on sequence homology matches and previously defined terms. KEGG biochemical pathways were assigned to annotated sequences using BLAST2GO. Finally, the InterProScan suite, which is

incorporated within BLAST2GO, was used to search for protein motif patterns and annotate sequences accordingly.

2.5 Real-time PCR gene expression levels analysis

Reverse transcriptase quantitative real-time PCR (qRT-PCR) was performed on hypobranchial gland and mantle tissue samples from female pre-copulatory individuals (n=3). Dissection, RNA extraction and assessment of RNA integrity were performed as described above (section 2.1). Dor- has been used to designate the gene products sequenced from *D. orbita*.

First strand cDNA was synthesized from 1µg of total RNA from the hypobranchial gland and mantle of three female pre-copulatory individuals using Superscript II reverse transcriptase (Invitrogen Life Technologies, Grand Island, NY, USA). For each RNA sample, cDNA was synthesised at least twice. Oligo$(dT)_{12\text{-}18}$ (Invitrogen Life Technologies) was used as the priming oligonucleotide, with a final reaction volume of 20µl as per the manufacturers' instructions.

For qRT-PCR four genes were selected from our EST library as potential reference genes, *D. orbita* (Dor-)alphatubulin, Dor-COX1, Dor-18srRNA, and Dor-profilin. The Dor-arylsulfatase gene was chosen as it is putatively involved in Tyrian purple synthesis. In addition, six other genes were randomly selected to test for suitability for use in the qRT-PCR study. Primers were designed manually for all 11 sequences (see Table 1) and gradient PCR was performed at differing primer and template concentrations, in order to determine the optimal reaction conditions (data not shown). Four primer sets, (Dor-18srRNA, Dor-profilin, Dor-polyubiquitin and Dor-acetylcholinesterase) failed to amplify single PCR products from both hypobranchial and mantle cDNA, and were removed from our study. Higher levels of Dor-alphatubulin were observed in the hypobranchial gland compared to the mantle cDNA (Figure 4C). Equivalent expression of the only remaining houskeeping gene Dor-COX1 was observed in both hypobranchial and mantle cDNA (data not shown) so it was chosen as the reference gene for qRT-PCR.

qRT-PCR reactions were prepared containing 0.1 µg cDNA, 5µl 2X ABsolute SYBR green qPCR mix (Abgene, Epsom, UK), 10 ng/µl forward and 10ng/µl reverse primer, and sterile water to a final volume of 10µl and analysed in real-time on a Corbett RotorGene 3000 (Corbett Research, New South Wales, Australia). Cycling conditions were as follows: 1 cycle of 95°C for 15 minutes for the activation of Taq polymerase followed by 40 cycles of (95°C for 30 seconds; 54°C for 30 seconds; 72°C for 30 seconds) and 1 cycle for 72°C for 1 min. Samples from all three individual snails were tested as independent replicates; each qRT-PCR run was repeated twice and samples were set up in triplicate within each run. All primers amplified one band of expected size as determined by melt curve analysis and gel electrophoresis. The external PCR standards for each gene of interest consisted of known numbers of molecules of purified PCR product and a no template control was run in triplicate in each run. Standards were prepared and input copy number calculated according to published methods (Yin et al., 2001). All qRT-PCR data are represented as a ratio of the number copies of the gene of interest, to the number of Dor-COX1 products. Comparisons between mantle and hypobranchial gland tissue for qRT-PCR experiments were made using independent samples t-test in Prism statistical software for Graphpad (Graphpad software, La Jolla CA, USA). For all analyses, $P \leq 0.05$ was considered significant. All data are expressed as mean ± standard error.

3. Results

3.1 EST manual analysis

A total of 554 randomly picked EST clones were sequenced from our hypobranchial gland cDNA library, and after performing quality control on sequence chromatographs we were left with 437 usable sequences (Table 2, Table 3). When assembly of sequences was performed eliminating overlap, as well as the removal of all clones with inserts less than one hundred base pairs, we were left with 311 singletons, and 37 contigs, comprising in total 348 non-redundant sequences (Table 2).

From these 437 EST sequences, 181 showed significant identity to known sequences within the NCBI database, based on preliminary BLAST analysis using tBLASTx, BLASTn and BLASTp (Table 3). From these 181 sequences, a total of 28 ribosomal

RNA sequences were found, 24 of which showed significant nucleotide identity to gastropod sequences, and four showed significant nucleotide identity with ciliate protozoa sequences (Table 3), such as those belonging to *Paramecium tetraurelia* and *Cyclidium glaucoma*. The remaining 153 sequences returned hits to known protein coding genes. It was observed that 57 of these sequences contained matches to known ciliate genes and would only be translated into functional proteins if the ciliate alternate codon translation system was utilized. These were excluded from BLAST2GO annotation. The remaining 96 protein coding sequences, which were translated utilizing the standard codon convention, represent genes expressed within the hypobranchial gland of *D. orbita.* These protein coding genes together with the 256 EST sequences that did not show homology to any known nucleotide sequences were combined (352 sequences in total) for automated BLAST2GO annotation.

3.2 D. orbita gene BLAST2GO ontology and annotation

As outlined above, we were left with 352 non-redundant sequences which underwent BLAST2GO annotation using BLASTx. From this analysis a total of 110 *D. orbita* unique ESTs (31%) were assigned GO categories. Figure 2 shows the distributions of GO terms from our sequences. Based on the BLASTx sequence analysis results sequences were grouped and functions were assigned, reporting on the biological process(es) (level 6 GO terms, Figure 2A), molecular function(s) (level 3 GO terms, Figure 2B) and the cellular component(s) (level 5 GO terms, Figure 2C) that the genes are likely to be involved in. Sequences of particular interest, due to suggested involvement in specific gene functions including transcription, translation, protein modification, electron transport, cellular processing, neurotransmission and Tyrian purple biosynthesis, are listed in Table 4.

The most dominant biological process encountered from our EST library (177 biological process matches in total) was identified as translation (n=22), followed by cellular respiration (n=13), proton transport (n=10) and ATP synthesis coupled electron transport (n=8) (Figure 2A). Several key genes involved in protein translation and modification were also identified; genes involved in post-translational protein modification (n=4), cellular protein complex assembly (n=18), protein folding (n=5) and phosphorylation (n=12) were all expressed in the hypobranchial gland, as well as genes involved in spindle organization and elongation (n=5 and 4 respectively),

carboxylic acid, RNA, monosaccharide and hexose metabolic processes (n= 8, 4, 4 and 4 respectively).

Considering the molecular function assigned to our sequences (a total of 233 assignments), genes implicated in nucleotide and nucleoside binding dominated (n=36 and 14 respectively), with a large proportion of sequences involved in protein binding also observed (n=29), as were tetrapyrrole binding sequences (n=7) and sequences that are structural constituents of the ribosome (n=19) (Figure 2B). Numerous enzymatic activities were also observed, including hydrolase activity (n=33), oxidoreductase activity (n=18) and transferase activity (n=9). Sequences involved in transport were also observed, including substrate-specific transporter activity (n=9) and transmembrane transporter activity (n=10).

When analysing the cellular component assignments of sequences in our EST library, of all 284 matches produced, the most abundant by far were those sequences within the cytoplasm or within a defined cytoplasmic part (n=58 and 48 respectively), closely followed by sequences localized in intracellular organelles (n=71) (Figure 2C). Sequences intrinsic to membranes, organelle inner membranes and plasma membranes were also observed (n=15, 10 and 4 respectively). Sequences were also identified for genes that localized to the ribonucleotide protein complex (n=24) and ribosomal subunit (n=12).

In sequences where enzymatic pathways could be assigned, the KEGG Pathways complemented the BLAST2GO annotation results (Table 4). A total of 53 cDNA sequences were assigned to KEGG pathways, and 22 different pathways were suggested to be operating within the hypobranchial gland of *D. orbita*. Pathways of particular interest are listed in Table 5. In total, only 11 different genes were identified to have KEGG enzymatic pathways, with a total of 20 different EST sequences involved in these classifications (Table 5).

InterProScan analysis within the BLAST2GO software allowed a broader range of sequences to be putatively identified, based on protein structure and feature information. From the 352 sequences that were analysed within this study, 39 showed protein signature matches to previously defined protein domains. In addition, 122

sequences displayed conserved protein domains consistent with a function as signal peptides, and 61 sequences were identified to have transmembrane domains.

Interestingly one of our sequences (Genbank accession no. **GD253910**) matched a sulfatase gene belonging to the sea urchin *Strongylocentrotus purpuratus* with an e value of 2.11 x E^{-39} according to BLASTx analysis. InterProScan analysis failed to identify any features or motif in this sequence, but KEGG analysis identified this sequence as one involved in sphingolipid metabolism (Table 5). 1xE value cutoff values were maintained at a very low threshold ($1xE^{-03}$) in order to maximise the chances of picking up even distant sequence similarities, but we were unable to identify any other enzyme sequences related to Tyrian purple production, such as bromoperoxidases or tryptophanases. Nevertheless, an alcohol dehydrogenase involved in tryptophan and alkaloid biosythesis was identified along with Glutaryl-CoA dehydrogenase which can be involved in tryptophan metabolism (Table 5).

Lastly, genes involved in carbohydrate metabolism and biopolymer synthesis were identified within our cDNA library. While specific mucin genes were not identified within our cDNA library, a gene involved in the regulation of mucins (Jonckheere et al., 2004), Smad4, was shown to be expressed within the hypobranchial gland.

3.3 Molluscan sequence comparisons

Our 352 hypobranchial gland ESTs were compared to other molluscan EST library sequences and 141 sequences were found to display significant matches to known mollusc ESTs (Figure 3). 87 of these sequences had already been putatively identified based on BLAST2GO Genbank sequence similarity analysis (Figure 3), so EST analysis against the available molluscan sequence libraries contributed an additional 54 sequence identity matches (Figure 3). The majority of these sequences showed sequence similarity to known gastropod ESTs, nearly half of which were sequences from *Ilyanassa obsoleta* (Figure 3). The remaining 211 sequences did not have any significant sequence similarity to publically available molluscan ESTs. Furthermore 188/211 did not display sequence similarity to any published sequence in the Genbank databases, confirming the uniqueness of our dataset (Figure 3).

The novelty of our dataset was further supported by the comparison of our 352 ESTs to the *Lottia gigantea* genome. Only 41 ESTs shared significant sequence identity with sequences in the *L. gigantea* genome. All of these EST sequences also had sequence identity matches in at least one of the previous analyses and 39 of the ESTs that matched with the *L. gigantea* genome showed significant sequence similarity with sequences in both the molluscan ESTs and the BLAST2GO annotation.

3.4 Gene expression comparisons

qRT-PCR was performed on six genes to confirm differential expression. Our results confirmed that Dor-arylsulfatase was upregulated in the hypobranchial gland compared to the mantle in *D. orbita* (Figure 4, P=0.03). In addition Dor-ACE was not expressed in the mantle at all (Figure 4.). This confirms the ability of our technique to detect gene expression differences between the mantle and the hypobranchial gland. There was substantial inter-individual variation in the gene expression of the four other genes (Dor-alphatubulin, Dor-arginine kinase, Dor-serine protease and Dor-ribosomalP0) chosen for qRT-PCR. However, on average there is a trend towards greater expression of all these genes in the hypobranchial gland.

4. Discussion

This study resulted in a total of 437 high quality EST sequences that are up-regulated or uniquely expressed in the hypobranchial gland, relative to mantle tissue, in *D. orbita*. When sequences underwent preliminary manual sequence analysis, the putative functions of 96 *D. orbita* protein coding sequences were identified, as well as 28 ribosomal RNA sequences and 256 novel sequences that were not assigned a function. A further 57 sequences were identified as having significant similarities to known ciliate protozoan genes (Laffy, 2011; Laffy et al., 2009). The broad range of functional genes identified within our cDNA library greatly increases the knowledge of Gastropoda gene expression, particularly regarding gene expression within the hypobranchial gland.

A notable proportion of the genes sequenced did not share similarity with any sequence in the current databases, with a total of 53% of the transcripts sequenced from *D. orbita* not showing any similarity to known sequences in Genbank. Similarly, less than ten percent of our sequences were found to have homologs in the *L. gigantea*

genome, further emphasising the number of novel sequences present in our EST library. The hypobrancial gland is absent or highly reduced in Pattellogastropods, such as *Lottia*, whereas a single left hypobranchial gland is characteristic of the Orthogastropoda (Ponder and Lindberg, 1997). This structure has increased in complexity through-out gastropod evolution and is a highly developed structure in *D. orbita*, with 9 distinct cell types (Westley et al., 2010). The high level of novel and non-homologous sequences in our *D. orbita* EST is likely to be influenced by the fact that no sequence data has ever been attained from any hypobranchial gland sample to date, and there is very little known about the gland itself.

Analysis of GO classifications of matched sequences from our hypobranchial gland library reveals a higher level of protein translation and modification is being performed within this gland, in comparison to the mantle. This increase in genes involved in protein translation and modification may be influenced by the presence of proteins specifically localised in the hypobranchial glands of gastropods, including attractin-like pheromones (Kuanpradit et al., 2010) and biosynthetic enzymes (Erspamer, 1946; Jannun and Coe, 1987; Westley and Benkendorff, 2009). At least 50 biological process categorizations were made for genes involved in regulation of gene expression, cellular protein complex assembly, translation or post translational protein modification. GTPase activity and GTP binding are both intrinsic to the process of ribosomal protein translation (Voigt and Nagel, 1993). The apparent upregulation of genes involved in electron transport and aerobic respiration within the hypobranchial gland relative to the mantle, including those with cytochrome-c oxidase, ATPase and ATP binding activity, is likely to be related to the high metabolic requirements of the diverse secretory cells which dominate this gland (Westley et al., 2010). ATP synthesis has been directly correlated with ATP demand (Hochachka, 1994), and one of the most prominent consumers of ATP in cells is protein synthesis (Wieser and Krumschnabel, 2001). As such, the expression of these ATP synthesis genes in the hypobranchial gland further supports an increase in protein production in this gland relative to the mantle tissue.

It is likely that numerous proteins are involved in mucin production, both as mucus proteins, as well as in the synthesis and processing of these proteins. The presence of genes involved in monosaccharide metabolic processing and hexose metabolic

processing are likely to be related to the production of mucus within the hypobranchial gland. It has been well documented that the hypobranchial gland of muricids can secrete large volumes of mucus (Naegel, 2005). Hunt and Jevon's investigations into the compositions and characteristics of the hypobranchial mucin of *Buccinum undatum* (Hunt, 1967; Hunt and Jevons, 1965), has given the most thorough analysis of muricid mucus currently available. It was determined that 41% of the mucus is protein, with 30% carbohydrate, 2% sugar (hexosamine), 13.2% sulphate and 0.5% calcium, with an ash composition of 9% (Hunt and Jevons, 1965). An ash composition of 9% suggests a considerable proportion of hypobranchial gland mucus contains inorganic compounds. A high proportion of our identified sequences have shown ion and cofactor binding capacity. Mucus is also secreted from the pedal gland of molluscs (Wright et al., 1997) and the associated genes might also be expressed in the mantle tissue. Thus it is possible that some genes involved in mucus biosynthesis and secretion from the hypobranchial gland may not be present in our cDNA library, due to subtraction from gene transcripts also expressed within the mantle tissue. A more thorough analysis of the hypobranchial gland transcriptome would be required to identify all sequences involved in mucus secretion in the future.

Genes associated with Tyrian purple biosynthesis were predicted to be expressed in the hypobranchial gland of *D. orbita*. Arylsulfatase has been shown to facilitate the conversion of tyrindoxyl sulphate into the biologically active precursors of Tyrian purple within the hypobranchial gland of muricid molluscs (Cooksey, 2001; Erspamer, 1946; Westley et al., 2006), including *D. orbita* (Baker and Sutherland, 1968). The presence of an arylsulfatase sequence within our cDNA library is the first identification of a transcript responsible for this enzyme within the Muricidae. In addition, the identification of an arylsulfatase gene sequence is the first confirmation that the arylsulfatase activity previously identified in these molluscs is not obtained from dietary sources or symbiotic micoroorganisms, but is in fact a molluscan gene transcribed within the hypobranchial gland. We have subsequently obtained the full length arylsulfatase gene from *Dicathais orbita* (Genbank accession no. **HM246144**) (Laffy, 2011).

Some other genes involved in tryptophan metabolism and alkaloid biosynthesis were identified by BLAST2GO analysis. The discovery of transcripts for these genes in our

subtracted library further supports *de novo* synthesis of the tryptophan derived indole alkaloids known to be produced by *D. orbita* and other Muricidae in the hypobrancial gland. However, current analysis of our cDNA library did not identify the expression of any halogenase, bromoperoxidases or any other enzyme likely to be involved in the bromination of these Tyrian purple precursors. Due to the high level of novel, unidentified sequences within our cDNA library, it is possible that we have obtained one or more of these sequences, but they are not homologous to the sequences present in the current database. It is possible that these enzymes have evolved to perform similar function without any sequence similarity to the bromoperoxidases that have been previously identified from bacteria and marine algae (Isupov et al., 2000; Pelletier et al., 1994; Weyand et al., 1999). It is also possible that homologous bromoperoxidase genes are expressed in the hypobranchial gland, but were not sequenced in our cDNA library, although preliminary attempts to amplify bromoperoxidase sequences from hypobranchial gland cDNA also failed to identify relevant sequences (data not shown). Further research is required to investigate any evidence for molluscan-specific brominating enzymes.

The ultimate precursor to Tyrian purple is stored as an indoxyl sulphate choline ester salt in the hypobranchial gland of muricids (Baker and Duke, 1976). Murexine and urocanylcholine are pharmacologically active choline esters present within the hypobranchial glands of muricids (Whittaker, 1960). Acetylcholinesterases are vital for the control of choline esters in biological systems and play a role in cholinergic neurotransmission in molluscs and other invertebrates (Hussein et al., 2002). We detected an acetylcholinesterase gene sequence in the hypobranchial gland of *D. orbita* and this was assigned a gene function in neurotransmission in our BLAST GO analysis. However, recent histochemical analysis of the hypobranchial gland reported that no nerve cells are present within the hypobranchial gland of *D. orbita* (Westley et al., 2010). Consequently, acetylcholinesterases present in the gland must be expressed by another cell type other than neurons, and these proteins may be involved in the maintenance and regulation of the bioactive murexine and urocanylcholine ester compounds present within the gland.

Ciliated supportive cells have been identified as the most abundant of all cell types present within the hypobranchial gland (Westley and Benkendorff, 2008; Westley et

al., 2010). The protein products of particular genes, namely dynein, alpha tubulin and beta tubulin, are known to be key components within cilia, and our SSH library suggests that these gene sequences are upregulated in the hypobranchial gland. The ciliated supportive cells of the hypobranchial gland have been identified as the localized site of Tyrian purple precursor production, which does not occur until arylsulfatase enzymes are released from within the supportive cells (Westley, 2008). The identification of these structural genes is important as it suggests that SSH can be used to differentiate between physiologically different tissue types. This can be seen in the hypobranchial gland of *D. orbita*, where highly specialized cellular functions, such as the release of Tyrian purple precursors and biosynthetic enzymes, are observed. Studies on the hypobranchial glands from other Muricidae have also identified highly specialized secretory cells (Hunt, 1973; Naegel and Aguilar-Cruz, 2006; Roller et al., 1995; Westley et al., 2010). The individual intracellular granules identified within the different cells displayed varied secretions, including eosinophilic and basophilic spherules and granules of differing dimensions. This large and varied number of secretory cells may account for the number of membrane intrinsic, cell wall and external side of the plasma membrane sequences that were identified within the gland. This remarkable diversity in cell types and secretory phenotypes is likely to contribute to the large number of varied and novel sequences identified within the hypobranchial gland of *D. orbita*.

The location of the hypobranchial gland in relation to other organs may also explain some of the gene expression identified in this study. Peptidyl-dipeptidase A expression (Table 5) suggests a link between the hypobranchial gland and vasoconstriction/vasodilation (Salzet et al., 2001). The hypobranchial gland has been described as highly vascularised tissue (Fretter and Graham, 1994), located in close proximity to the vascular sinus (Westley et al., 2010). Consequently these genes could be involved in regulating blood vessels and blood supply to this metabolically active biosynthetic organ.

From our SSH results, we are able to make several inferences into the biological functions and processes that occur within the hypobranchial gland of the marine snail *D. orbita*. It is important to note that these hypobranchial gland gene expression patterns are relative to expression within the mantle tissue, and subsequent

comparisons to other tissues would result in a more complete analysis of differentially expressed genes within the hypobranchial gland of *D. orbita*. Genes that may play a role in mucus production and secretion were expressed within the gland, as were sequences involved in protein production and modification. This supports current knowledge regarding hypobranchial gland physiology and function, including its role in the production and secretion of mucus into the mantle cavity. Clearly, high levels of protein production and processing are taking place within the gland in comparison to the mantle. The role of the hypobranchial gland in Tyrian purple production is also confirmed with the identification of an arylsulfatase gene, in addition to acetylcholinesterase, which is likely to help regulate the indoxyl sulphate choline ester precursors. Further research is required to identify the remaining enzymes involved in Tyrian purple biosynthesis. The high proportion of novel sequences present within our cDNA library suggests there is an abundance of unique *D. orbita* genes, and their expression within the hypobranchial gland may suggest an involvement in key biochemical processes that are specific to this gland in molluscs. Further characterization of these genes, by obtaining full length sequences, would be required to obtain a more thorough understanding of the transcriptome of the hypobranchial gland. In-situ hybridization studies would provide additional information regarding the localisation of these transcripts within specific cell types and would further enrich our understanding of cellular processing within the hypobranchial gland. This work demonstrates that the SSH approach holds great potential for the discovery and classification of novel genes, including useful biosynthetic enzymes, and lays a strong foundation for future studies investigating Tyrian purple biogenesis and hypobranchial gland functionality.

5. Acknowledgements

Patrick Laffy was supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. This research was funded by a grant from a philanthropic foundation to KB and CAA. We would also like to thank the South Australian Partnership for Advanced Computing, for the use of their BLAST portal and Dr Chantel Westley for her assistance in specimen dissection.

References

Andrews, E.B., Little, C., 1972. Structure and function in the excretory system of some terrestrial prosobranch snails (Cyclophoridae). J. Zool. 168, 395-422.

Bai, Z., Yin, Y., Hu, S., Wang, G., Zhang, X., Li, J., 2010. Identification of genes potentially involved in pearl formation by expressed sequence tag analysis of mantle from freshwater pearl mussel (*Hyriopsis Cumingii Lea*). J. Shellfish Res. 29, 527-534.

Baker, J., Duke, C., 1976. Isolation of choline and choline ester salts of tyrindoxyl sulphate from the marine molluscs *Dicathais orbita* and *Mancinella keineri*. Tetrahedron Lett., 1233-1234.

Baker, J., Sutherland, M., 1968. Pigments of marine animals VIII. Precursors of 6,6'-dibromoindigotin (Tyrian Purple) from the mollusc *Dicathais orbita* Gmelin. Tetrahedron Lett., 43-46.

Baker, J.T., 1974. Tyrian purple: an ancient dye, a modern problem. Endeavour (Oxford) 33, 11-17.

Benkendorff, K., Bremner, J.B., Davis, A.R., 2000. Tyrian purple precursors in the egg masses of the Australian muricid, *Dicathais orbita*: A possible defensive role. J. Chem. Ecol. 26, 1037-1050.

Benkendorff, K., Bremner, J.B., Davis, A.R., 2001a. Indole derivatives from the egg masses of Muricid molluscs. Molecules 6, 70-78.

Benkendorff, K., Davis, A.R., Bremner, J., 2001b. Chemical defense in the egg masses of benthic invertebrates: An assessment of antibacterial activity in 39 mollusks and 4 polychaetes. J. Invertebr. Pathol. 78, 109-118.

Benkendorff, K., McIver, C.M., Abbott, C.A., 2011. Bioactivity of the Murex homeopathic remedy and of extracts from an Australian muricid mollusc against human cancer cells. Evid. Based Complementary Altern. Med., Article ID 879585, 879512 pages, 872011. doi:879510.871093/ecam/nep879042.

Berland, S., Marie, A., Duplat, D., Milet, C., Sire, J.Y., Bedouet, L., 2011. Coupling proteomics and transcriptomics for the identification of novel and variant forms of mollusk shell proteins: A study with *P. margaritifera*. Chembiochem 12, 950-961.

Cardenas, L., Sanchez, R., Gomez, D., Fuenzalida, G., Gallardo-Escarate, C., Tanguy, A., 2011. Transcriptome analysis in *Concholepas concholepas* (Gastropoda, Muricidae): Mining and characterization of new genomic and molecular markers. Marine Genomics 4, 197-205.

Clark, M.S., Thorne, M.A.S., Vieira, F.A., Cardoso, J.C.R., Power, D.M., Peck, L.S., 2010. Insights into shell deposition in the Antarctic bivalve *Laternula elliptica*: gene discovery in the mantle transcriptome using 454 pyrosequencing. BMC Genomics 11, 362

628 Colgan, D., Ponder, W., Beacham, E., Macaranas, J., 2003. Gastropod phylogeny
629 based on six segments from four genes representing coding or non-coding and
630 mitochondrial or nuclear DNA. Molluscan Research, 23, 123-148.

631 Colgan, D.J., Ponder, W.F., Beacham, E., Macaranas, J., 2007. Molecular
632 phylogenetics of Caenogastropoda (Gastropoda: Mollusca). Mol. Phylogenet. Evol.
633 42, 717-737.

634 Colgan, D.J., Ponder, W.F., Eggler, P.E., 2000. Gastropod evolutionary rates and
635 phylogenetic relationships assessed using partial 28S rDNA and histone H3
636 sequences. Zool. Scr. 29, 29-63.

637 Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005.
638 Blast2GO: a universal tool for annotation, visualization and analysis in functional
639 genomics research. Bioinformatics 21, 3674-3676.

640 Cooksey, C.J., 2001. Tyrian purple: 6,6 '-dibromoindigo and related compounds.
641 Molecules 6, 736-769.

642 Diatchenko, L., Lukyanov, S., Lau, Y.F.C., Siebert, P.D., 1999. Suppression
643 subtractive hybridization: A versatile method for identifying differentially expressed
644 genes, Methods in Enzymology: Volume 303. cDNA Preparation And
645 Characterization. Academic Press Inc, San Diego, 349-380.

646 Edwards, V., Benkendorff, K., Young, F., 2012. Marine compounds selectively
647 induce apoptosis in female reproductive cancer cells but not in primary-derived
648 human reproductive granulosa cells. Marine Drugs 10, 64-83.

649 Erspamer, V., 1946. Ricerche chimiche e farmacologiche sugli estratti di ghiandola
650 ipobranchiale di *Murex (Truncularia) trunculus (L.)*, *Murex (Biolinus) brandaris (L.)*
651 *e Tritonalia erinacei (L.)* Pub. Stax, Zool, Napoli 20, 91-101.

652 Fang, D., Xu, G., Hu, Y., Pan, C., Xie, L., Zhang, R., 2011. Identification of genes
653 directly involved in shell formation and their functions in pearl oyster, *Pinctada*
654 *fucata*. Plos one 6, e21860, 1-13.

655 Feng, Z.P., Zhang, Z., van Kesteren, R.E., Straub, V.A., van Nierop, P., Jin, K.,
656 Nejatbakhsh, N., Goldberg, J.I., Spencer, G.E., Yeoman, M.S., Wildering, W.,
657 Coorssen, J.R., Croll, R.P., Buck, L.T., Syed, N.I., Smit, A.B., 2009. Transcriptome
658 analysis of the central nervous system of the mollusc *Lymnaea stagnalis*. BMC
659 Genomics 10, 15.

660 Fretter, V., Graham, A., 1994. British Prosobranch Molluscs. The Ray Society,
661 Dorset.

662 Friedlander, P., 1909. Uber den Farbstoff des antiken Purpurs aus *Murex brandaris*.
663 Berichte Der Deutschen Chemischen Gesellschaft, 42, 765-770.

664 Gestal, C., Costa, M.M., Figueras, A., Novoa, B., 2007. Analysis of differentially
665 expressed genes in response to bacterial stimulation in hemocytes of the carpet-shell
666 clam *Ruditapes decussatus*: Identification of new antimicrobial peptides. Gene 406,
667 134-143.

668 Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J.,
669 Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional
670 annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420-
671 3435.

672 Henry, J.J., Perry, K.J., Fukui, L., Alvi, N., 2010. Differential localization of mRNAs
673 during early development in the mollusc, *Crepidula fornicata*. Integ. Comp. Biol. 50,
674 720-733.

675 Heyland, A., Vue, Z., Voolstra, C.R., Medina, M., Moroz, L.L., 2011. Developmental
676 transcriptome of *Aplysia californica*. J. Exp. Zool. B. Mol. Dev. Evol. 316B, 113-134.

677 Hochachka, P., 1994. Solving the common problem: matching ATP synthesis to ATP
678 demand during exercise. Adv. Vet. Sci. Comp. Med. 38, 41-56.

679 Hunt, S., 1967. Secretion of mucopolysaccharide in the whelk *Buccinum undatum L*.
680 Nature 214, 395-396.

681 Hunt, S., 1973. Fine structure of the secretory epithelium in the hypobranchial gland
682 of the prosobranch gastropod mollusc *Buccinum undatum L*. J. Mar. Biol. Assoc. U.K.
683 53, 59-71.

684 Hunt, S., Jevons, F.R., 1965. The hypobranchial mucin of the whelk *Buccinum*
685 *undatum L*. Properties of the mucin and of the glycoprotein component. Biochem. J.
686 97, 701-709.

687 Hussein, A.S., Harel, M., Selkirk, M.E., 2002. A distinct family of
688 acetylcholinesterases is secreted by *Nippostrongylus brasiliensis*. Mol. Biochem.
689 Parasitol. 123, 125-134.

690 Isupov, M.N., Dalby, A.R., Brindley, A.A., Izumi, Y., Tanabe, T., Murshudov, G.N.,
691 Littlechild, J.A., 2000. Crystal structure of dodecameric vanadium-dependent
692 bromoperoxidase from the red algae *Corallina officinalis*. J. Mol. Biol. 299, 1035-
693 1049.

694 Jackson, D.J., Degnan, B.M., 2006. Expressed sequence tag analysis of genes
695 expressed during development of the tropical abalone *Haliotis asinina*. J. Shellfish
696 Res. 25, 225-231.

697 Jacob, J., Mitreva, M., Vanholme, B., Gheysen, G., 2008. Exploring the transcriptome
698 of the burrowing nematode *Radopholus similis*. Mol. Genet. Genomics 280, 1-17.

699 Jannun, R., Coe, E.L., 1987. Bromoperoxidase from the marine snail *Murex*
700 *trunculus*. Comp. Biochem. Physiol. 88B, 917-922.

701 Jonckheere, N., Perrais, M., Mariette, C., Batra, S.K., Aubert, J.P., Pigny, P., Van
702 Seuningen, I., 2004. A role for human MUC4 mucin gene, the ErbB2 ligand, as a
703 target of TGF-beta in pancreatic carcinogenesis. Oncogene 23, 5729-5738.

704 Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F., Zanella-Cleon, I.,
705 Cochennec-Laureau, N., Gueguen, Y., Montagnani, C., 2010. Transcriptome and

proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: focus on biomineralization. BMC Genomics 11, 613.

Kelley, W.P., Wolters, A.M., Sack, J.T., Jockusch, R.A., Jurchen, J.C., Williams, E.R., Sweedler, J.V., Gilly, W.F., 2003. Characterization of a novel gastropod toxin (6-bromo-2-mercaptotryptamine) that inhibits shaker K channel activity. J. Biol. Chem. 278, 34934-34942.

Kuanpradit, C., Cummins, S.F., Degnan, B.M., Sretarugsa, P., Hanna, P.J., Sobhon, P., Chavadej, J., 2010. Identification of an attractin-like pheromone in the mucus-secreting hypobranchial gland of the abalone *Haliotis asinina linnaeus*. J. Shellfish Res. 29, 699-704.

Laffy, P.W., 2011. Evolution, gene expression and enzymatic production of Tyrian purple: A molecular study of the Australian muricid *Dicathais orbita* (Neogastropoda; Muricidae). Ph.D. Thesis: School of Biological Sciences. Flinders University, Adelaide.

Laffy, P.W., Benkendorff, K., Abbott, C.A., 2009. Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae). Nautilus 123, 154-158.

McDougall, C., Green, K., Jackson, D.J., Degnan, B.M., 2011. Ultrastructure of the Mantle of the gastropod *Haliotis asinina* and mechanisms of shell regionalization. Cells Tissues Organs 194, 103-107.

Moroz, L.L., Edwards, J.R., Puthanveettil, S.V., Kohn, A.B., Hla, T., Heyland, A., Knudsen, L., Sahni, A., Yu, F.H., Liu, L., Jezzini, S., Lovell, P., Iannucculli, W., Chen, M.C., Nguyen, T., Sheng, H.T., Shaw, R., Kalachikov, S., Panchin, Y.V., Farmerie, W., Russo, J.J., Ju, J.Y., Kandel, E.R., 2006. Neuronal transcriptome of *Aplysia*: neuronal compartments and circuitry. Cell 127, 1453-1467.

Moroz, L.L., Ju, J., Russo, J.J., Puthanveetti, S., Kohn, A., Medina, M., Walsh, P.J., Birren, B., Lander, E.S., Kandel, E.R., 2004. Sequencing the *Aplysia* Genome: a model for single cell, real-time and comparative genomics, Submission of intent to complete sequencing project from the *Aplysia* genome project http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/AplysiaSeq.pdf.

Naegel, L.C.A., 2005. The effect of periodically "milking" to obtain Tyrian purple from *Plicopurpura pansa* (Gould, 1853) on the frequency of expulsion and mortality. J. Shellfish Res. 24, 85-90.

Naegel, L.C.A., Aguilar-Cruz, C.A., 2006. Hypobranchial gland from the purple snail *Plicopurpura pansa* (Gould, 1853) (Prosobranchia: Muricidae). J. Shellfish Res. 25, 391-394.

Naegel, L.C.A., Cooksey, C., 2002. Tyrian purple from marine muricids, especially *Plicopurpura pansa* (Gould, 1853). J. Shellfish Res. 21, 193-200.

Nagai, K., Yano, M., Morimoto, K., Miyamoto, H., 2007. Tyrosinase localization in mollusc shells. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 146, 207-214.

747 Pelletier, I., Pfeifer, O., Altenbuchner, J., Van Pee, K.H., 1994. Cloning of a second
748 non-haem bromoperoxidase gene from *Streptomyces aureofaciens* ATCC 10762:
749 sequence analysis, expression in *Streptomyces lividans* and enzyme purification.
750 Microbiology 140, 509-516.

751 Ponder, W.F., Lindberg, D.R., 1997. Towards a phylogeny of gastropod molluscs: an
752 analysis using morphological characters. Zoo. J. Linn. Soc. 119, 83-265.

753 Quilang, J., Wang, S.L., Li, P., Abernathy, J., Peatman, E., Wang, Y.P., Wang, L.L.,
754 Shi, Y.H., Wallace, R., Guo, X.M., Liu, Z.J., 2007. Generation and analysis of ESTs
755 from the eastern oyster, *Crassostrea virginica* Gmelin and identification of
756 microsatellite and SNP markers. BMC Genomics 8, 11.

757 Rhee, J.S., Kim, B.M., Jeong, C.B., Horiguchi, T., Lee, Y.M., Kim, I.C., Lee, J.S.,
758 2012. Immune gene mining by pyrosequencing in the rockshell, *Thais clavigera*. Fish
759 Shellfish Immunol. 32, 700-710.

760 Roller, R.A., Rickett, J.D., Stickle, W.B., 1995. The hypobranchial gland of the
761 estuarine snail *Stramonita haemastoma canaliculata* (Gray) (Prosobranchia:
762 Muricidae): a light and electron microscopical study. Am. Malacol. Bull. 11, 177-190.

763 Roseghini, M., Severini, C., Erspamer, G., Erspamer, V., 1995. Choline esters and
764 biogenic amines in the hypobranchial gland of 55 molluscan species of the
765 neogastrapod muricoidea superfamily. Toxicon 34, 33-55.

766 Salzet, M., Deloffre, L., Breton, C., Vieau, D., Schoofs, L., 2001. The angiotensin
767 system elements in invertebrates. Brain Res. Rev. 36, 35-45.

768 Voigt, J., Nagel, K., 1993. Regulation of elongation factor G GTPase activity by the
769 ribosomal state. The effects of initiation factors and differentially bound tRNA,
770 aminoacyl-tRNA, and peptidyl-tRNA. J. Biol. Chem. 268, 100-106.

771 Walters, E.T., Moroz, L.L., 2009. Molluscan Memory of Injury: Evolutionary Insights
772 into Chronic Pain and Neurological Disorders. Brain Behav. Evol. 74, 206-218.

773 Westley, C., 2008. The distribution, biosynthetic origin and functional significance of
774 Tyrian purple precursors in the Australian muricid *Dicathais orbita* (Neogastropoda:
775 Muricidae). Ph.D. Thesis. School of Biological Sciences. Flinders University,
776 Adelaide.

777 Westley, C., Benkendorff, K., 2008. Sex-specific Tyrian purple genesis: Precursor
778 and pigment distribution in the reproductive system of the marine mollusc, *Dicathais*
779 *orbita*. J. Chem. Ecol. 34, 44-56.

780 Westley, C., Benkendorff, K., 2009. The distribution of precursors and biosynthetic
781 enzymes required for Tyrian purple genesis in the hypobranchial gland, gonoduct, an
782 egg masses of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae). Nautilus
783 123, 148-153.

784 Westley, C., Benkendorff, K., Vine, K.L., 2006. A proposed functional role for indole
785 derivatives in reproduction and the defense system of Muricidae

(Neogastropoda:Mollusca) in: L. Meijer, N. Guyard, L. Skaldsounis, G. Eisenbrand (Eds.). Life in Progress editions, Roscoff, France, 31-44.

Westley, C.B., Lewis, M.C., Benkendorff, K., 2010. Histomorphology of the hypobranchial gland in *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae). J. Moll. Stud. 76, 186-195.

Weyand, M., Hecht, H.J., Kiess, M., Liaud, M.F., Vilter, H., Schomburg, D., 1999. X-ray structure determination of a vanadium-dependent haloperoxidase from *Ascophyllum nodosum* at 2.0 angstrom resolution. J. Mol. Biol. 293, 595-611.

Whittaker, V.P., 1960. Pharmacologically active choline esters in marine gastropods. Ann. NY Acad. Sci. 90, 695-705.

Wieser, W., Krumschnabel, G., 2001. Hierarchies of ATP-consuming processes: direct compared with indirect measurements, and comparative aspects. Biochem. J. 355, 389-395.

Wolters, A.M., Jayawickrama, D.A., Sweedler, J.V., 2005. Comparative analysis of a neurotoxin from *Calliostoma canaliculatum* by on-line capillary isotachophoresis/H-1 NMR and diffusion H-1 NMR. J. Nat. Prod. 68, 162-167.

Wright, D.J., Matthews, C.L., Donaldson, I., Martin, A.P., Henderson, I.F., 1997. Immunocytochemical localisation of glycosaminoglycan-like activity in the mucus and associated tissues of terrestrial slug species (Gastropoda : Pulmonata). Comp. Biochem. Physiol. B Biochem. Mol. Biol. 118, 825-828.

Yin, J.L., Shackel, N.A., Zekry, A., McGuinness, P.H., Richards, C., Van Der Putten, K., McCaughan, G.W., Eris, J.M., Bishop, G.A., 2001. Real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for measurement of cytokine and growth factor mRNA expression with fluorogenic probes or SYBR Green I. Immunol. Cell Biol. 79, 213-221.

York, P.S., Cummins, S.F., Degnan, S.M., Woodcroft, B.J., Degnan, B.M., 2010. Identification of genes differentially expressed in the ganglia of growing *Haliotis asinina*. J. Shellfish Res. 29, 741-752.

Yuan, L.H., Zuo, X.G., Chen, J.P., Lin, B.F., Zhang, J.P., Sun, M., Zhang, S.Y., 2008. Differential expression analysis of Liprin-alpha 2 in hibernating bat (*Rhinolophus ferrumequinum*). Prog. Nat. Sci. 18, 27-32.

819 **Table 1. Quantitative real-time PCR (qRT-PCR) primers.** Primers were designed from EST
820 sequences produced. In all analyses, cytochrome oxidase subunit I (Dor-COXI) was used as a
821 housekeeper gene to normalize expression levels. Dor-acetylcholinesterase, Dor-18srRNA, Dor-
822 polyubiquitin and Dor-profilin primer sets failed to produce a single PCR band and so were not used in
823 qRT-PCR analysis.
824

Gene Name Accession no	Primer orientation	PCR Product Size	Primer sequence (5'-3')
Angiotensin converting enzyme (Dor-ACE) GD254022	Forward Reverse	172	GCTGCACGCCTTCGTTCGTC CCAGATCCTCGATGCCCTCCC
Ribosomal protein P0 (Dor-ribosomalP0) GD253988	Forward Reverse	144	GCCCTGCAGTCCTCGACATC GTCTCTTGAAGCCATTGACG
Serine protease (Dor-serineprotease) GD253884	Forward Reverse	87	GCCGATGGTTTAGCAAAG GGGTGTAGGGGCCGCTG
Alphatubulin (Dor-alphatubulin) GD253901	Forward Reverse	137	GGGCTCGAAGCAGGCGTTGG CCACTTCCCTCTGGCCACC
Arginine kinase (Dor-argininekinase) GD253925	Forward Reverse	81	CCCCAACAGCCTCCTCAAAC GCACCATTGAAGCGGGTCC
Arylsulfatase (Dor-arylsulfatase) GD253910	Forward Reverse	119	CCGTCGGGATGTGACTCCAC GGTCAAAGTCTTCATGGCTTGC
Cytochrome oxidase subunit I (Dor-COXI) FJ476170	Forward Reverse	97	GCTCCGGACATGGCCTTTC CGCTTTCAACAGCAGCTGAGG
Acetylcholinesterase (Dor-acetylcholinesterase) GD254024	Forward Reverse	107	CCTGATGGGCATCAACG CCTTGAACTGTCTGTCAGAG
18s ribosomal RNA (Dor-18srRNA) FJ476168	Forward Reverse	124	CCCATGTACGGTGAAACC GGAGGGGTCGGAGCTGG
Polyubiquitin (Dor-polyubiquitin) GD254024	Forward Reverse	117	CCCCCAGATCAGCAGCGTC CCCACGCAGACGCAGCACC
Profilin (Dor-profilin) GD253839	Forward Reverse	117	GGAGGAAGTTATCAAGTTG GCCCACTTCTTCCCGCCC

825
826 **Table 2. Summary statistics of ESTs generated from the hypobranchial gland of *D. orbita***
827

Feature	Value
Total number of clones sequenced	554
Average length of high quality ESTs (bp)	502
Number of high quality sequences	437
Number of contigs	37 (containing 126 sequences)
Number of singletons	311

828
829
830
831
832
833

834 **Table 3. Statistics of BLASTx searches, performing manual and automated analysis using**
835 **BLAST2GO.**

Feature	**Manual analysis***	**BLAST2GO analysis**
Number of sequences analysed	437	352
Number of sequences with significant BLAST hits	181	110
Percentage of sequences with unique hits	41 %	31 %
Number of unidentified sequences	256	242
Classification of Annotated sequences from Manual sequence analysis	**Number of sequences (Overall Percentage)**	
D. orbita ribosomal RNA sequences	24 (13%)	
D. orbita expressed genes	96 (53%)	
Ciliate protozoan ribosomal RNA sequences	4 (2%)	
Ciliate protozoan expressed genes	57 (31%)	

836
837 *Sequences in manual analysis included sequences with significant similarity to ciliate sequences and
838 ribosomal RNA sequences, which were not included in BLAST2GO analysis
839

840 Table 4. Genes expressed in the hypobranchial gland of *D. orbita* identified by
841 BLAST2GO.
842 Sequences matching ribosomal RNA and ribosomal proteins were excluded from this table

Gene Function	**Genbank Accession Number**	**Gene annotation**	**EC Number**
transcription	GD253679	mad homolog 3	-
translation	GD253962	H2A histone member V	EC:3.6.5.3
	GD253823, GD253818	translation initiation factor subunit 1 alpha	EC:3.6.5.1-4
	FJ455780, FJ476164, FJ476158, FJ476156, FJ476138	cytochrome B	-
protein modification	GD253702, GD253925	arginine kinase	EC:2.7.3.3, EC:5.3.4.1, EC:2.7.3.3., EC:2.3.2.13
	GD253956, GD253968, GD253975, GD254001, GE905837, GD253853, GD253849, GD253848, GD253837, GD253807, GD253748, GD253737	calmodulin	-
	GD253866	dual-specific tyrosine-phosphorylation regulated kinase	EC:2.7.10-11
	GD253864, GD253881	glyceraldehyde-3-phosphate dehydrogenase	EC:1.2.1.12
	GD253884	serine protease	-
	GD253872	heat shock protein 90	-
	GD253861, GD253664, GD254022	angiotensin converting enzyme	EC:3.4.15.1
	GD254009	sec11 protein	-
	GD254006	hydrocephalus inducing protein	-
	GD253865	phosphatase catalytic gamma isoform	-
electron transport	FJ476170, FJ476167, FJ476148, FJ476141, FJ476128	cytochrome c oxidase subunit I	EC:1.9.3.1
	GD253781	glutaryl-coenzyme a dehydrogenase	EC:1.3.99.7
	FJ455785	NADH dehydrogenase subunit I	EC:1.6.5.3
cellular processing	GD254015, GD253983, GD253924, GD253901, GD253894, GD253867, GD253857, GD253696	alpha-tubulin	EC:3.6.5.1-4
	GD253691, GD253871	dynein light chain family protein	-
	GD253991, GD253936, GD253913, GD253906, GD253890, GD253885, GD253803, GD253838, GD253801, GD253779, GD253693	beta tubulin	EC:3.6.5.1-4
	GD253783	tbc1 domain member 13	-
	GD253713, GD253852, GD253945	sodium potassium-dependant atpase beta-2 subunit	EC:3.6.3.9
	GD253868	pol-like protein	-
	GD253671, GD253955	histidine triad nucleotide binding protein 1	-
	GD253670, GD253772, GD253845, GD253879	Actin	-
neurotransmission	GD253726	agrin isoform 1	-
	GD254024	acetylcholinesterase	-

Tyrian purple biosynthesis	GD253910	arylsulfatase J	EC:3.1.6.1

843

844 **Table 5. Key KEGG enzymatic pathways involved or active within the hypobranchial gland of *D.***
845 ***orbita* as identified through BLAST2GO analysis.**
846 * Glutaryl-CoA dehydrogenase and alcohol dehydrogenase are involved in more than three KEGG
847 enzymatic pathways therefore are grouped together at the bottom of this Table. The genes and
848 pathways in bold are directly relevant to Tyrian purple biosynthesis.
849

KEGG enzymatic pathway	Gene(s) identified	Sequence(s) identified
Oxidative phosphorylation	ATPase	GD253852
	cytochrome C oxidase	FJ476138, FJ476156, FJ476164, FJ476170, FJ476128
	NADH dehydrogenase	FJ455785
Vibrio cholera infection	disulfide-isomerase	GD253702
	ATPase	GD253852
Flagellar assembly	ATPase	GD253852
Type III secretion system	ATPase	GD253852
Sphingolipid metabolism	**Arylsulfatase**	GD253910
Androgen and estrogen metabolism	**Arylsulfatase**	GD253910
Arginine and proline metabolism	arginine kinase	FJ455783, GD253702
Glycolysis/gluconeogenesis	glyceraldehyde-3-phosphate dehydrogenase	GD253864, GD253881
Ubiquinone biosynthesis	NADH dehydrogenase	FJ455785
Renin-angiotensin system	peptidyl-dipeptidase A	GD254022, GD253664
Complement and coagulation cascades	protein-glutamine gama-glutamyltransferase	GD253702
Gene identified *	**KEGG enzymatic pathways**	
glutaryl-CoA dehydrogenase (GD253781)	Benzoate degradation, Lysine degradation and **Tryptophan metabolism**, Fatty acid metabolism	
alcohol dehydrogenase (GD253933)	1- and 2-methyl naphthalene degradation, drug metabolism, D-alanine metabolism, Cytochrome p450 metabolism, alanine and aspartate metabolism, 3-chloroacrylic acid degradation, Aminoacyl-tRNA biosynthesis, phenylalanine, **tyrosine and tryptophan biosynthesis**, aminophosphonate metabolism, Nicotinate and nicotinamide metabolism, Bile acid biosynthesis, Tyrosine metabolism and **Alkaloid biosynthesis**, Fatty acid metabolism	

850
851
852
853

FIGURE LEGENDS

Figure 1 *D. orbita*, its egg capsules and a partially dissected hypobranchial gland.
A) shows two captive specimens of *D. orbita* surrounded by several egg capsule clutches. B) shows the three regions of the hypobranchial gland; branchial, medial and rectal, in reference to the ctenidium, osphradium and columellar muscle within the mantle cavity of *D. orbita*. Photos courtesy of Chantel Westley.

Figure 2 Gene ontology (GO) assignment (3rd level GO terms) of differentially expressed hypobranchial gland genes expressed in *D. orbita* using BLAST2GO automated sequence annotation.
A total of 110 sequences from our EST library, displaying BLASTx 1x values < $1xE^{-05}$ underwent GO assignment. A) Biological processes; categories with more than three positive assignments are displayed. B) Molecular function, C) Cellular compartment. It is important to note that multiple assignments are possible for each individual sequence and as such sequence function assignments vary between each GO category.

Figure 3 Distribution of both nucleotide and sequence matches separated by abundance and taxonomic Class.
Manual tBLASTx analysis was performed in May 2012, comparing sequences to ESTs from the 20 most sequenced molluscs in the NCBI non-redundant (nr) EST database. An E-value cut-off threshold of $1xE^{-05}$ was applied to this analysis. Although the 20 most-sequenced mollusc EST libraries were used in the analysis, only 13 species were represented when the best blast hits were observed.

Figure 4 qRT-PCR results for selected genes, relative to the expression levels of housekeeper gene Dor-COXI (cytochrome oxidase subunit I). All assays were performed in triplicate on RNA isolated from three pre-copulatory female *D. orbita*, and were repeated at least twice. Final values were expressed as mean ± SEM. *$P < 0.05$ for arylsulfatase mRNA transcripts in hypobranchial gland tissue compared to mantle tissue.