Flinders
UNIVERSITY
ADELAIDE • AUSTRALIA

Evolution, gene expression and enzymatic production of Tyrian purple: A molecular study of the Australian muricid *Dicathais orbita* (Neogastropoda: Muricidae)

Patrick Laffy

Faculty of Science and Engineering

School of Biological Sciences

Flinders University

November 2011

Declaration

I certify that this thesis does not incorporate, without acknowledgment, any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge it does not contain any material previously published or written by another person except where due reference is made in the text

…….. ……..

Patrick William Laffy Date

Acknowledgements

Where to begin? First and foremost I would like to thank my supervisors Associate Professor Cathy Abbott and Dr Kirsten Benkendorff for their guidance and support throughout this project and the last 5 years of work. Your feedback, musing of ideas and scientific knowledge has made me into a much better writer and scientist than I was 5 years ago. Not only did you conceptualize this project before I started my candidature, but you provided the financial funding to make it all possible.

I also need to thank several people who were lent their scientific knowledge and support to me. To Dr Melissa Pitman, Dr Melanie Sulda, Dr Ana Glavinic, Dr Chantel Westley and Associate Professor Michael Schwarz, I have appreciated your feedback, advice and your tireless support to the work that I have produced. I wouldn't have been able to complete this thesis without you. Special thanks to Melissa, Melanie, and Ana for the friendship you offered at the same time. And another thank you to Melissa (my, she is popular!) and Dr Simon Schmidt, who over the last year or so have turned Thursday night dinner into much needed impromptu lab meetings. We have a very supportive group of students (and now post-docs) in the School of Biological Sciences at Flinders, and I have learnt a lot from all of you, in particular, Dr Chevaun Smith, Dr Melissa Gregory, Lexi Young, Dr Simon Williams, Emma Decourcey–Ireland, Nick Warnock, Pradeep Sornaraj, Bianca Kyriacou, Dr Peter Bain, Drew Sutton and Crystal Sweetman. I am glad to call you my friends and you have all made coming to work everyday a wonderful thing, and I am really going to miss you guys when I'm gone (fingers crossed it will be soon). Special thanks to Lexi for all her help editing and formatting my thesis.

To two great labs at Flinders who have welcomed me in, despite the fact that my project didn't exactly conform to the typical lab experiments. To A/Prof Cathy Abbott, Dr Tong Chen, Dr Roger Yazbek, Dr Melissa Pitman, Dr Melanie Sulda and Dono Indarto, thank you for exposing me to the world of dipeptidyl peptidases and lending me your molecular knowledge. And to all the members of the Mollusc lab; Dr Kirsten Benkendorff, Dr Chantel Westley, Dr Cassandra McIver, Dr Ana Glavinic, Dr Skye Woodcock, David Rudd, Warwick Noble, Tom Stewart, Casey Campleman and Ryan Baring, your marine biology know-how (and essential skills in maintaining marine aquaria) have contributed greatly to my project. Many thanks also go to Dr Chevaun Smith, Associate Professor Kathleen Soole and Professor David Day for keeping me gainfully employed over the last year, I really appreciate it.

Thanks also to my friends Andrew Cook, Ana dela Cruz, Johan Velleman, Natasha Pietsch, Andrew Harding, Jamie Bowles and Jodi Sargent. You have no idea how much I have appreciated having a group of understanding people around me who accept me for me and can tolerate endless scientific ramblings without loosing your cool, or falling asleep with boredom. And to my gorgeous godchild Isabella, who can lighten up even the darkest mood with a giggle or a smile.

To my sisters Bridget and Siobhan Laffy, who have put up with me over the last few years. I have been moody, messy and generally disagreeable and I appreciate your support, tolerance and love.

Finally, thank you to my parents, Pat and Catherine Laffy. You have shaped me into the person I am today. In particular to Dad, who has supported me financially over the course of the last 5 years (and, well the rest of my life) and who has never questioned my need to pursue a career in research. It hasn't been easy, but I hope you are proud of the achievements I have made. I promise to get a real job as soon as I can.

Abstract

Tyrian purple is the traditional source of purple pigmentation used in the textile industry since ancient times, sourced from the Muricidae family of neogastropod molluscs. Brominated indole derivatives of tryptophan, the precursors to Tyrian purple are potent anticancer and antibacterial compounds which may have potential for pharmaceutical development. In addition to their production within the hypobranchial gland, some members of the Muricidae invest Tyrian purple precursors within their egg capsules. The first aim of this thesis was to investigate the evolution of Tyrian purple precursor investment within the egg masses of muricid molluscs using a molecular phylogenetic analysis of 18s and 28s ribosomal RNA sequences. The second aim of this thesis was to investigate the gene expression of the hypobranchial gland of the Australian muricid *Dicathais orbita*, in an effort to uncover the enzymes involved in the production of Tyrian purple.

The investigation into the evolution of Tyrian purple precursor investment within the egg capsules of muricid molluscs identified that the capacity for adults to invest these compounds in their egg capsules is a trait that was not ancestral and has arisen at least twice in the evolution of the Muricidae. Molecular analysis confirmed the monophyly of the Rapaninae and Ocenebrinae muricid subfamilies members and supports Tan's 2003 classification of a new muricid subfamily, the Haustrinae. These findings also support the use of *D. orbita* as a representative of the Rapaninae in which to study Tyrian purple synthesis and investment.

Suppressive subtractive hybridization (SSH) was used to identify genes that were up-regulated or uniquely expressed in the hypobranchial gland of *D. orbita*. A total of 438 sequences were identified to be differentially expressed in the hypobranchial gland, including an arylsulfatase gene. Arylsulfatase activity is known to be involved in the formation of Tyrian purple from precursors in muricid molluscs. The full length arylsulfatase sequence was amplified and recombinantly expressed in a mammalian expression system. No active enzyme was produced from these experiments suggesting an incompatibility between molluscan arylsulfatase and mammalian expression systems.

Initial manual sequence analysis indicated that over 65% of sequences expressed in the hypobranchial gland showed no homology to known database sequences. The subset of genes

that did show sequence matches to genes in the database showed homology to a wide variety of taxa, including chordate, molluscan and ciliate sequences. Our investigation into the gene expression of the hypobranchial gland of *D. orbita* enabled the functional assignment of 110 sequences using BLAST2GO automated sequence annotation. The hypobranchial gland plays a key role in muricid biology as a site of chemical interaction and biosynthesis. Manual sequence annotation also identified a number of sequences within our cDNA library that would only be functional if translated using an alternate codon translation system used by ciliate protozoans. Histological analysis of the hypobranchial gland identified intracellular ciliate protozoans present within the gland. Ciliate abundance varied in accordance to the reproductive condition of the host snail and 57 ciliate protein coding genes were identified within our cDNA library. Analysis of ribosomal RNA sequences from our expression library confirmed the presence of ciliate protozoans within the hypobranchial gland of *D. orbita* belonging the ciliate class Phyllopharyngea and possibly from another unidentified ciliate class. A novel use of SSH is proposed for the investigation of symbiont gene expression in other biological systems.

In summary, this thesis uses molecular techniques to explore the synthesis and evolution of Tyrian purple and hypobranchial gland gene expression in the muricid mollusc *D. orbita*. This thesis is the first study to investigate the evolution of Tyrian purple precursor investment within the egg capsules of muricid molluscs and has revealed that this is a derived trait that has arisen at least twice since the muricids diverged from other Muricoidean species. In addition, this is the first study to investigate gene expression within the hypobranchial gland of any mollusc. This study also identified one of the gene sequences involved in the enzymatic production of these bioactive compounds. Further investigations are required in order to produce active recombinant molluscan arylsulfatase enzymes. Additional investigations are also required in order to identify the other enzymes involved in the production of Tyrian purple precursors, which would then facilitate the *in vitro* synthesis of these compounds in the future. If sustainable synthesis of these compounds is established, these bioactive compounds may find use in pharmaceutical or nutraceutical treatments.

Preface

Parts of the work presented in this thesis have been published, or are currently in preparation for submission for publication. For consistency and ease of reading, all manuscripts are presented in the required formatting for the Journal of Biological Chemistry. As all references are listed in a single reference list at the end of the thesis to eliminate repetition, Harvard referencing was used in the text.

Published papers

Chapter 3. Laffy, PW, Benkendorff, K & Abbott, CA "Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae)" *Nautilus*, 123(3), 154-158.

Papers in preparation

Chapter 2. Laffy, PW, Schwarz, MP, Abbott CA & Benkendorff, K "The evolution of bioactive Tyrian purple precursors in the egg capsules of the Muricidae (Mollusca: Gastropoda)".Molluscan Research, under preparation.

Chapter 4 Laffy, PW, Benkendorff, K & Abbott, CA "Annotation and characterization of a partial transcriptome of the hypobranchial gland of *Dicathais orbita*" *Marine Biotechnology,* under preparation.

Chapter 5 Laffy, PW, Westley C, Abbott, CA & Benkendorff "Novel application of suppressive subtractive hybridization for the identification of symbionts: Discovery of ciliate protozoa in the hypobranchial gland of *Dicathais orbita* (Neogastropoda, Mollusca)" Molecular Ecology, under preparation.

Chapter 6 Laffy, PW, Chen, T, Benkendorff, K & Abbott, CA "Characterisation and expression of recombinant arylsulfatase from the marine snail *Dicathais orbita*" Comparative Biochemistry and Physiology Part B Biochemistry and Molecular Biology, under preparation.

The data published in abstract form and contributed to this thesis:

Laffy, PW, Benkendorff, K & Abbott CA (2006) "Gene expression within the hypobranchial gland of *Dicathais orbita*" Oral presentation at Molluscs 2006 meeting, University of Wollongong, Australia 6-8th December 2006.

Laffy, PW, Benkendorff, K & Abbott CA (2007) "Using molecular approaches to investigate the function of the hypobranchial gland of the marine snail, *Dicathais orbita*" Oral presentation at World Congress of Malacology, Antwerp, Belgium 15-20th July 2007

Laffy, PW, Benkendorff, K & Abbott CA (2008) "Transcriptomic analysis of the functions of the hypobranchial gland of the marine snail *Dicathais orbita*, using BLAST2GO" Poster presentation at ASMR Annual Meeting, Adelaide, Australia 4th June 2008

Laffy, PW, Benkendorff, K & Abbott CA (2008) "Annotating a partial transcriptome of the marine snail *Dicathais orbita's* hypobranchial gland using suppressive subtractive hybridization and BLAST2GO" Poster presentation at the 19th international conference on Genome Informatics, Gold Coast, Australia, 1st-3rd December 2008

Laffy, PW, Westley, C, Benkendorff, K & Abbott CA (2009) "Parallel genome identification: A novel use of Suppressive Subtractive Hybridization in systems biology" Poster presentation at the Gordon Research Conference on Evolutionary and Ecological Functional Genomics In Tilton, New Hampshire, USA, 12-17 July 2009

Abbreviations

[N-morpholino] ethanesulfonic acid	MES
Adenosine triphosphate	ATP
Amino acid	aa
Australian Genome Research Facility	AGRF
Basepair	bp
Basic local alignment search tool	BLAST
Bayesian inference	BI
Celcius	C°
Cell membrane	Cm
Complimentary DNA	cDNA
Cytochrome oxidase subunit I	COI
Cytoplasm	Cy
Daltons	Da
Deoxynucleotide triphospates	dNTPs
Deoxyribose nucleic acid	DNA
Diethyl pyrocarbonate	DEPC
Dipeptidyl peptidase	DP
Dipeptidyl peptidase IV	DPIV
Ethylenediaminetetraacetic acid	EDTA
Example	e.g.
Expressed sequence tags	ESTs
Figure	Fig
Foetal calf serum	FCS
Formylglycine Generating Enzyme	FGE
Gas chromatography-mass spectrometry	GC/MS
Gene ontology	GO
Guanosine triphosphate	GTP
H-Ala-Pro-p-nitroanilide	H-Ala-Pro-pNA
Heat shock protein	HSP
Hour	hr
Hypobranchial gland	HBG

Internal transcribed spacer region 2	ITS2
Inter-Services Intelligence	ISI
KiloDaltons	kDa
Kyoto Encyclopedia of Genes and Genomes	KEGG
Mantle cavity	Mc
Mass spectrometry	MS
Maximum parsimony	MP
Micrograms	µg
Microlitres	µl
Micrometre	µm
Micromolar	µM
Milligram	mg
millilitre	ml
milliMolar	mM
Minutes	min
Modified Harris haematoxylin and Eosin Y	H&E
Molar	M
Nanogram	ng
Nanometres	nm
National Centre for Biotechnology Information	NCBI
Nucleus	Nu
Open reading frame	ORF
Phospho-buffered saline	PBS
Phospho-buffered saline buffered Tris	PBST
p-nitrocatechol sulfate	pNCS
Polyacrylamide gel electrophoresis	PAGE
Polymerase chain reaction	PCR
Polyvinylidene Fluoride	PVDF
Posterior probability	PP
Rectal gland	Rg
Rapid amplification of cDNA ends	RACE
Reverse jump hyperprior	rjhp
Revolutions per minute	RPM

Ribose nucleic acid	RNA
Ribosomal RNA	rRNA
Room temperature	RT
Seconds	sec
Secretion	Sc
Sodium dodecyl sulfate	SDS
South Australian Partnership for Advanced Computing	SAPAC
Standard error	S.E.
Stress-inducible protein	STI-1
Suppressive subtractive hybridization	SSH
Tetramethylethylenediamine	TEMED
units	U
Volts	V

Table of Contents

Table of Figures

Data Tables

Chapter 1. General introduction

1.1 The biomedical potential of marine natural products

The development of pharmaceutical compounds is one of the most pressing and expensive industries in current biotechnology. A survey on the development of new drugs estimated a final cost of up to US $802 million for each drug developed, taking into account the final clinical success rate of 21.5% (DiMasi *et al.* 2003). Given the astronomical cost of developing such compounds, it is logical to adopt evidence-supported methods when identifying bioactive compounds for use in disease treatment. Products identified in nature have been used for thousands of years for the treatment of a variety of ailments. Over 60% of approved pharmaceutical agents are either natural products, are derived from natural products or utilize natural products as lead compounds in their production and synthesis (Demain 2009). While history supports the use of terrestrial natural products in the development of pharmaceuticals, the marine environment is an almost untapped resource with amazing potential for future drug development. This review highlights the potential of marine natural products in the treatment and management of cancer and microbial infection and highlights the potential muricid molluscs have in the development of new pharmaceutical and nutraceutical treatments.

1.1.2 Marine-derived cancer treatments

Over 100,000 new cases of cancer were reported in Australia in the year 2005, and the incidence of cancer diagnosis was predicted to increase by 3,000 people every year in this country until the year 2010 (AIHW (Australian institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2008). In addition, 39,000 people died of cancer in Australia in 2005, with an additional 800 dying each year until 2010 (AIHW (Australian institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2008). The increases in cancer diagnoses, as well as the limitations of current treatment options, have prompted the discovery of new treatment alternatives in the clinical realm. Traditional chemotherapeutic agents act on rapidly dividing cells but are unable to differentiate between healthy and cancerous cells (Ma & Wang 2009). A more recent focus of anticancer research entails the identification of chemotherapeutic agents that specifically target cancer cells (Sawyers 2004). One of the most effective sources of target-specific anticancer compounds is the natural environment. Natural products are typically more effective in their anticancer activity while being less toxic than traditional chemotherapeutic compounds or chemically synthesized compounds (Ma & Wang 2009). While traditional naturally derived pharmaceutical compounds are almost always of terrestrial origin, a burgeoning bioprospecting

field focusing on marine organisms as a source of novel and bioactive compounds has developed (Haefner 2003).

The sea squirt-derived Trabectedin (also known as yolendis and ET-743) has been developed and marketed as an anticancer compound used to target soft tissue sarcoma (Carter & Keam 2007). The macrocyclic lactone compound Bryostatin 1 from the bryozoan *Bugula nerita* has protein kinase C inhibitory qualities and has shown potential for use in dendritic cell-based anticancer treatments (Do *et al.* 2004). The sea sponge derived calyculins, originally identified in *Discodermia calyx,* may show potential in cancer treatments due to their protein phosphatase inhibition activity (Fagerholm *et al.* 2010). Heteronemin, a sesterterpene isolated from the sponge *Hyriotis* sp. has apoptosis inducing ability against myeloid leukemia cell lines (Schumacher *et al.* 2010). Kahalahide F is a depsipeptide derived from the mollusc *Elysia rufescens,* which is currently in phase II clinical trials for the treatment of prostate cancer (Faircloth & Cuevas 2006). Dolastatin 10, a peptide from the sea hare *Dolabella auricularia,* binds to tubulin causing cell cycle arrest and has been shown to inhibit cancer growth in leukemic cells (Pettit *et al.* 1987). While dolastatin 10 proved too problematic in phase I clinical trials (Pitot *et al.* 1999), its chemical analogue TZT-1027 has shown promise in phase I solid tumor clinical trials (Schoffski *et al.* 2004). Another depsipeptide, Aplidine from the sea squirt *Aplidum albicans*, inhibits vascular endothelial growth factor, which causes cell cycle arrest and has shown potential for the treatment of breast cancer and leukemia (Erba *et al.* 2002, Cuadrado *et al.* 2003). Brominated indoles from the of the marine mollusc *Dicathais orbita* have exhibited cytotoxic activity against cancer cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011), with preclinical studies for colon cancer showing the extracts induce apoptosis *in vivo* (Westley *et al.* 2010c). The variety of compounds derived from marine organisms that have potential as cancer treatments supports the trend into surveying marine organisms for more pharmaceutically active compounds. The surveying of marine organisms for antimicrobial activity has also identified a variety of different compounds which have the potential in the treatment of microbial infections.

1.1.3 Marine-derived microbial infection treatments.

The increase in prevalence of multidrug resistant bacteria is becoming one of the most prominent medical issues in hospitals today. Nosocomial infections, also known as hospital-acquired infections, infect approximately two million people every year in the United States of America and are responsible for approximately 100,000 deaths annually (Balaban & Dell'Acqua

2005).There has been a marked decrease in antibiotic drug discovery in the pharmaceutical industry since the late 1980's (Demain 2009), caused by the low cost of recovery from their development together with a declining rate of discovery of novel chemical structures displaying antibacterial activity. This has reduced treatment options for infections and is partially responsible for the prevalence of drug resistant bacteria in our medical institutions. Hospitals around the world are faced with the financial burden of prolonged stays and additional treatment costs due to the abundance of drug resistant nosocomial infections (Croft *et al.* 2007). In addition to the increase in hospital based infections, the prevalence of community derived drug resistant bacteria place more and more of the community at risk. Without the development of new treatment methods for drug resistant bacteria, we are likely to see an increase in nosocomial infections and may start to witness the rise of multidrug resistant bacteria, resulting in increased mortality rates from infections and an inflated financial burden on healthcare systems worldwide.

Bacteria are the traditional source of the majority of antibiotics that are used in medicine (Demain 2009). This is due to the frequency of bacterial interactions and environmental competition. The marine environment is one of the most bacteria rich environments on the planet, with some studies estimating the number of bacteria per milliliter of sea water as high as 10^6-10^9 (Haug *et al.* 2004). The world's oceans span up to 70% of the planet's surface and contain an amazing diversity of organisms (Faulker 2000). Many invertebrate species in the marine environment, particularly those living in or on the benthic substrata, such as members of the Porifera, Mollusca, Cnidaria and Echinodermata, have developed potent antibacterial secondary metabolites in order to defend themselves from microorganism attack (Shaw *et al.* 1974). While traditional drug development studies rely on terrestrial species for the identification of new antibacterial compounds, the marine environment is emerging as an ideal place to identify new and novel antibacterial compounds for clinical use.

Gorgonian corals have been shown to exhibit antimicrobial activity against the marine pathogen *Aspergillus sydowii* and the human pathogen *Aspergillus flavus* (Kim *et al.* 2000). Marine steroids and sesquiterpenes isolated from these gorgonian corals were shown to also display potent antimicrobial against several human bacterial pathogens (Roussis *et al.* 2001). Other marine invertebrates including the sea urchin *Strongylocentrotus driebachiensis*, the sea star *Asterias rubens* and the sea cucumber *Cucumaria frondosa* all exhibit antibacterial activity against several Gram positive bacteria via an unidentified mechanism of action (Haug *et al.*

2002b). The polyhydroxylated fucophlorethol isolated from the marine alga *Fucus vesiculosus* exhibits antibacterial activity against Gram positive and Gram negative human pathogenic bacteria (Sandsdalen *et al.* 2003). Secondary metabolites with antibacterial activity have been identified in several marine sponges including *Didiscus oxeata* and *Hippospongia communis. Didiscus oxeata* produces at least two secondary metabolites, (+)-curcuphenol and (+)-curcudiol, which show antifungal activity against the human pathogenic fungus *Trichiphythton mentagrophytes*. *Hippospongia communis* produces (-)-untenospongin B, a C21 bisfuranoterpene with antimicrobial activity against human pathogenic fungi, bacteria and yeast (Rifai *et al.* 2004). Marine decapods have also shown antibacterial (Haug *et al.* 2002a, Haug *et al.* 2002b).

Marine molluscs have also shown potential as a source of antimicrobial secondary metabolites. Antimicrobial peptides and other unidentified compounds within the horse mussel *Modiolus modiolus* inhibit the growth of bacteria including the human pathogen *Staphylococcus aureus* (Haug *et al.* 2004). Antibacterial and antiviral activity has been detected in chemical extracts from the common Cockle *Cerastoderma edule*, the common whelk *Buccinum undatum*, the Japanese carpet shell *Ruditapes philippinarum* and the European flat oyster *Ostrea edulis* (Defer *et al.* 2009). Tribromoimidazole was isolated from the egg masses of the muricid molluscs *Trunculariopsis trunculus*, *Ceratosoma erinaceum* and *Trophon geversianus* and was shown to exhibit antibacterial activity (Benkendorff *et al.* 2004a). Antimicrobial peptides have also been identified in several bivalve species (Mitta *et al.* 2000, Cellura *et al.* 2007, Zhao *et al.* 2007, Li *et al.* 2009). Brominated indoles from the egg capsules of the Muricidae also exhibit antibacterial activity (Benkendorff *et al.* 2000, Benkendorff *et al.* 2001a). A survey of antimicrobial activity from marine organisms in California in 1974 suggested that 9 out of 14 different phyla tested displayed measurable activity (Shaw *et al.* 1974), and a survey into the antimicrobial activity of molluscan egg masses identified that 18 out of 23 species tested displayed antimicrobial activity (Benkendorff *et al.* 2001b). The high incidence of bioactivity detected from marine organisms and the diversity of life in the marine environments highlights the potential of marine bioprospecting for the identification of new lead compounds which may find use in fighting infections and treating cancer in the future. Marine molluscs are one such group of organisms that show great potential for the development of new anticancer and antibacterial treatments.

1.1.4 The therapeutic potential of marine molluscs

The Mollusca is the second largest animal phylum, consisting of approximately 7% of living animals. The marine environment is home to as many as 200,000 different species of molluscs. Members of the Mollusca are essentially soft bodied invertebrates that can be found in almost every environment on the planet, with members occupying nearly every possible trophic niche (Benkendorff 2010). A summary of the current phylogeny for the Mollusca has recently been published (Fig 1.1) utilizing phylogenomic techniques and has for the first time produced a well supported topology for the Mollusca (Kocot *et al.* 2011). In addition to being one of the most speciose animal phyla in the world's oceans, marine molluscs have developed chemical strategies to effectively fight infections and bacterial attacks (Benkendorff *et al.* 2001b, Benkendorff 2010); in an environment that has as many as 10^6 microbial cells per ml of water (Whitman *et al.* 1998). The abundance and diversity of marine molluscs make them ideal candidates for marine bioprospecting investigations.

Previous investigations into the chemical ecology and secondary metabolites produced by marine molluscs have seen a major focus on the soft bodied opisthobranch gastropods (Benkendorff 2010). Most members of this family display either a reduced shell or no shell at all, implying that, as they display no physical defences against predators, they would need to develop other means of anti-predatory defence including chemical defences (Cobb & Willan 2006). The opisthobranchs belong to the molluscan class Gastropoda and subclass Orthogastropoda, which makes up to 90% of all marine molluscs (Ponder & Lindberg 2008). Gastropod molluscs can be found in terrestrial, freshwater, marine benthic, pelagic and infaunal habitats and include herbivores, scavengers and predators (Benkendorff 2010). Yet despite this diversity and abundance, it has been reported that less than one percent of gastropods have been investigated for their chemical diversity (Avila 2006, Benkendorff 2010). Over 91% of all chemical studies on orthogastropods have been performed on members of the Heterobranchia (including opistobranchs and the air-breathing pulmonates), despite the abundance of other species within the subclass. Of the chemical studies that have been performed on orthogastropods, it has been shown that members of the Caenogastropoda show the highest number of chemical compounds, but are typically overlooked for chemical investigations due to the presence of external shells in these species (Benkendorff 2010). Benkendorff (2010) suggested future chemical studies in shelled molluscs is warranted, based on previous studies,

the species diversity that is available and the historical use of shelled molluscs in natural therapies.

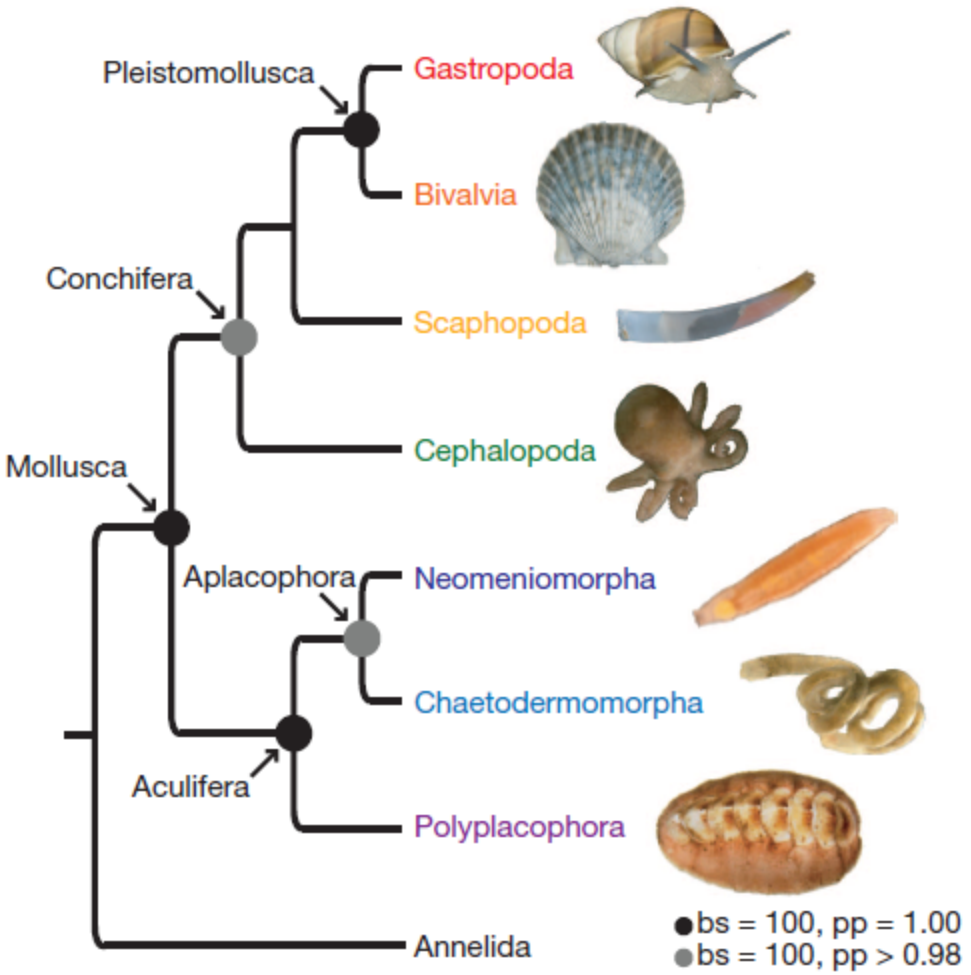

Figure 1.1 Molluscan phylogeny presented in Kocot *et al* 2011. This phylogenetic analysis was performed by compiling transcriptome data to review the evolution of the Mollusca and has been constructed from Bayesian and maximum parsimony analysis of a combined protein sequence dataset containing information from 308 protein sequences (Kocot *et al.* 2011). Reprinted by permission from Macmillan Publishers Ltd: Nature 477(7365): 452-U101.Kocot, K.M., J.T. Cannon, *et al.* (2011). "Phylogenomics reveals deep molluscan relationships."

It seems logical that members of the Caenogastropoda that are historically used in natural medicines would be an ideal source of chemical compounds, which may also find use in the pharmaceutical industry. The Murex homeopathic remedy, derived from extracts from the Muricidae, has been used since the 1800s to treat a number of ailments, including uterine and breast cancer (Boericke 1999). Brominated indoles isolated from the hypobranchial glands and egg masses of the muricid mollusc *Dicathais orbita*, have been shown to display antibacterial activity against human and marine pathogens (Benkendorff *et al.* 2000) and have exhibited

potent anticancer activity against human cancer cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011) and in a rodent model of colorectal cancer (Westley *et al.* 2010c). This historical use of murex extracts to treat illness, as well as the supportive evidence of bioactive compounds being present within the hypobranchial glands of these molluscs, prompts a deeper investigation into the medical potential of the Muricidae.

1.2 The family Muricidae

The Muricidae are a family of shelled caenogastropods, also known as murex snails. They are found all over the world and display unique shells, which are often prized by shell collectors. Muricid molluscs are predatory species that prey on other gastropods, bivalves, barnacles and ascidians (Taylor *et al.* 1980). A recent phylogenetic analysis has been performed on the Muricidae, investigating nine of the ten muricid subfamilies currently considered (Fig 1.2). The monopholy of six of these subfamilies (Ergataxinae, Rapaninae, Coralliophilinae, Haustrinae, Ocenebrinae, and Typhinae) was confirmed, as was the Monophyly of the family as a group in the Neogastropoda (Barco *et al.* 2010). Members of the Muricidae are the traditional source of the ancient pigment compound Tyrian purple.

1.2.1 Tyrian purple production

Tyrian purple, also known as royal purple, shellfish purple or purple of the ancients describes a pigmented compound that has been used since ancient times to dye garments and fabrics purple (Baker 1974). This compound was produced from the extracts and secretions of muricid molluscs and evidence of this textile industry dates back as far as the 13th century B.C in the Mediterranean (Naegel & Alvarez 2005). The biosynthesis of Tyrian purple precursor compounds occurs in the hypobranchial gland of muricid molluscs, although extracts from the gland only develop the intense purple colouring under sunlight-stimulated oxidation (Cooksey & Sinclair 2005). The chemical structure of Tyrian purple was first identified as 6,6'-dibromoindigo in 1909 by Friedlander from *Murex brandaris* (Friedlander 1909). Four Tyrian purple precursors or prochromagens were identified in *M. trunculus* and a fifth was also identified in *D. orbita* (Baker & Sutherland 1968, Baker 1974). The combination of precursors produced by each snail varies between species (Baker 1974), however all are derived from the tryptophan derivative indoxyl sulfate (Cooksey 2001).

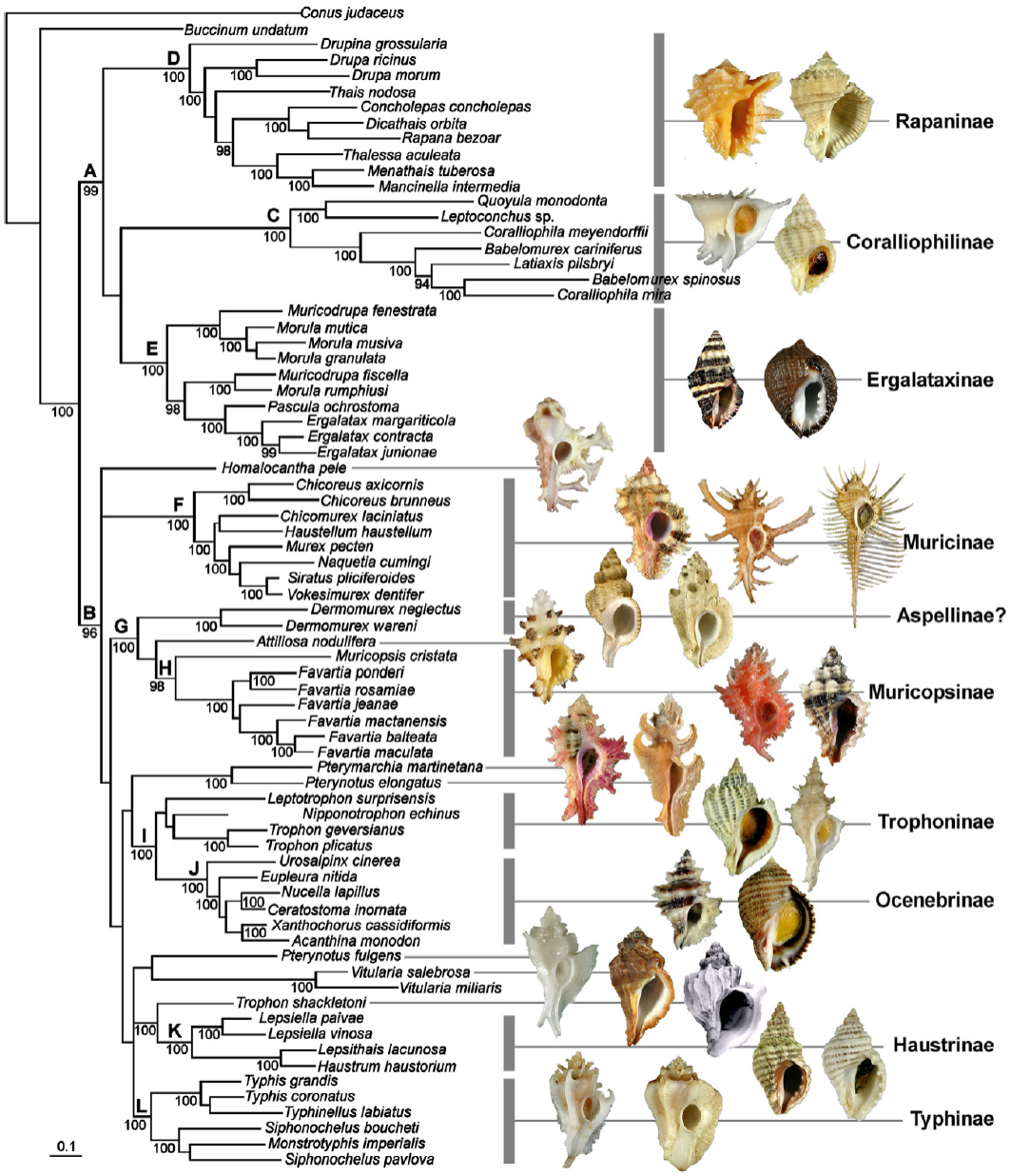

Figure 1.2 Bayesian inference of the Muricidae as reported by Barco *et al.* 2010. Phylogenetic analysis was constructed from a combined dataset of 12s, 16s and 28s ribosomal RNA sequences and Cytochrome oxidase subunit I (Barco *et al.* 2010). Bayesian analysis confirms the monophyly of subfamilies Coralliophilinae (C), Rapaninae (D), Ergalataxinae (E), Muricinae (F), Muricopsinae (H) and Haustrinae (K). Reprinted from Molecular Phylogenetics and Evolution, Volume 56(3), Barco, A., M. Claremont, *et al.* "A molecular phylogenetic framework for the Muricidae, a diverse family of carnivorous gastropods." 1025-1039 (2010).

The chemical formation of Tyrian purple in *D. orbita* is summarized in Figure 1.3. In order to produce the initial precursor tyrindoxyl sulfate, it has been suggested that different enzymes are required to convert the amino acid tryptophan into a brominated indole. Westley *et al.* (2006) suggested the most likely method of indole production is the *de novo* conversion of

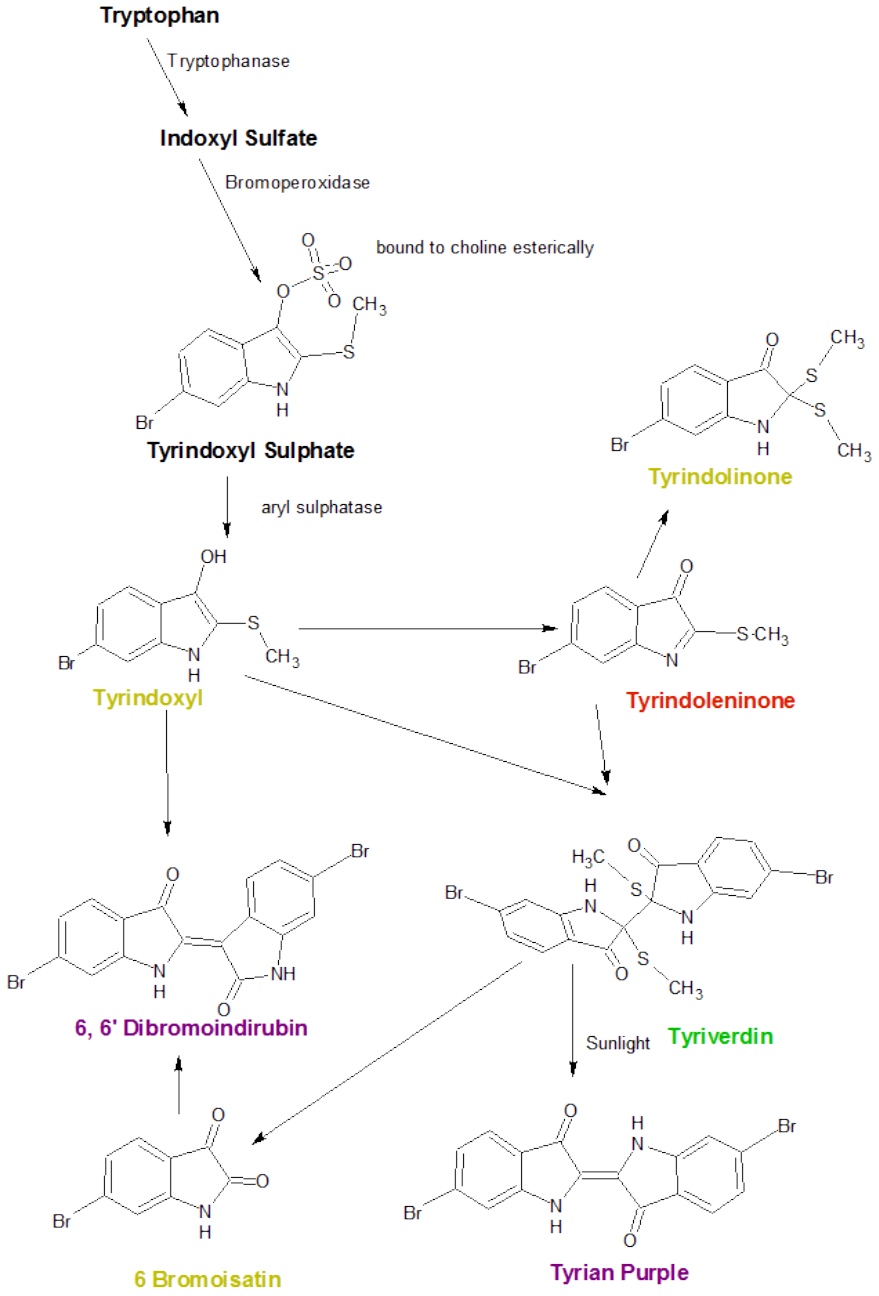

Figure 1.3 Chemical pathway involved in the formation of Tyrian purple.
This figure shows the various colour changes seen in the formation of Tyrian purple. The amino acid tryptophan undergoes degradation from a (hypothetical) tryptophanase enzyme forming indoxyl sulfate, and bromoperoxidase then reacts with the indole ring, forming tyrindoxyl sulfate (1), with the addition of methane thiol from unknown enzymes. The white compound tyrindoxyl sulfate (1), found in the hypobranchial gland is cleaved by an arylsulfatase enzyme, forming the yellow compound tyrindoxyl (2). Tyrindoxyl (2) forms the orange tyrindoleninone (3) which either reacts with a methane thiol group to form the green/yellow tyrindolinone (5) or reacts with tyrindoxyl (2) to form the ultimate precursor of Tyrian purple (6), tyriverdin (4). The green compound tyriverdin (4) is cleaved or degraded under sunlight to form Tyrian purple (6). Tyriverdin (4), tyrindoleninone and tyrindoxyl can also oxidize to form the compound 6-bromoisatin (7), a compound yellow in colour. The reaction of 6-bromoisatin with tyrindoleninone generates 6.6' dibromoindirubin (8), a structural isomer of Tyrian purple, 6,6, dibromoindigo. (Adapted from Cooksey, 2001 & Westley *et al.* , 2006)

dietary derived tryptophan to indole or indoxyl sulfate using a tryptophanase enzyme. Additional investigations identified significant concentrations of tryptophan in the hypobranchial gland region based on histochemical examination (Westley & Benkendorff 2008, Westley & Benkendorff 2009). The incorporation of the bromine in the 6-position of the indole ring of the prochromagens (1-4) implies that a bromoperoxidase or other haloperoxidase is also involved in the processing of tryptophan or indole. Bromoperoxidase activity has been identified in hypobranchial gland homogenates in *M. trunculus* (Jannun & Coe 1987) and histological sections of the hypobranchial gland of *D.orbita* (Westley & Benkendorff 2009), but the protein responsible for this activity has yet to be isolated. The ultimate precursor to Tyrian purple (6), tyrindoxyl sulfate (1), an indoxyl sulfate which is esterically bound to choline esters in the hypobranchial gland (Baker & Duke 1976), is oxidized by an arylsulfatase enzyme to form tyrindoxyl (2) (Baker & Sutherland 1968). The resulting prochromagen, which is in turn oxidized to form tyrindoleninone (3) was isolated from diethyl ether extracts of the gland identified, along with small amounts of the methanethiol adduct tyrindolinone (5) (Baker & Duke 1973b, Baker 1974). The prochromagens tyrindoleninone (3) and tyrindoxyl (2) combine to form the intermediate prochromagen tyriverdin (4) (Christopherson *et al.* 1978), which is externally oxidized to form the commercially important Tyrian purple (6). An oxidative artifact to Tyrian purple, 6-bromoisatin (7), has also been identified as a degradation product of tyriverdin (4) (Baker 1974), tyrindoleninone or tyindoxyl (2) (Cooksey, 2001).

1.2.2 Muricid bioactive compounds,

As discussed previously, several of the precursors of Tyrian purple have been found to be biologically active, exhibiting anticancer and antibacterial activity. The precursors tyrindoleninone, tyriverdin and the artefact 6-bromoisatin, isolated from the muricid *D. orbita*, have all been shown to exhibit varying levels of antibacterial activity (Benkendorff *et al.* 2000). These compounds are produced in the hypobranchial gland (Baker & Sutherland 1968, Baker & Duke 1976), but have also been reported from the egg capsules and reproductive organs (Benkendorff *et al.* 2000, Westley & Benkendorff 2008). It has been suggested that *D. orbita* invests tyrindoxyl sulfate inside egg capsules where the bioactive compounds are generated, protecting the developing embryos from bacterial infections (Benkendorff *et al.* 2000, Westley & Benkendorff 2008). Furthermore, the artefact 6-bromoisatin and the precursor tyrindoleninone have been shown to be cytotoxic to colon and breast carcinoma and lymphoma cell lines (Vine *et al.* 2007, Benkendorff *et al.* 2011), as well as inducing apoptosis in response to a carcinogen

in an *in vivo* rodent model for colorectal cancer (Westley *et al.* 2010c). Additional brominated indoles form as minor pigments of Tyrian purple, 6- dibromoindirubin and 6,6'-dibromoindirubin. These were isolated from the hypobranchial glands of *D. orbita* (Baker & Duke 1973a, Baker 1974) and in *Hexaplex trunculus,* where its specific glycogen-synthase-kinase-3 (GSK3) inhibitory activity has been observed (Meijer *et al.* 2003). The inhibition of GSK3 in humans reduces the rate of apoptotic cell death and increases neuronal survival by facilitating the regulation of Tau and β-Catenin molecules, implicated in neurodegeneration (Culbert *et al.* 2001). A traditional Chinese remedy, Ganggui Luhui Wan, that is used to treat chronic myelogenous leukemia in China, contains plant derived indirubin as an active ingredient, responsible for the treatments' antiproliferative and apoptosis inducing mechanism (Xiao *et al.* 2002). In addition to the bioactivity that several Tyrian purple precursors and pigments display, the ultimate precursor to Tyrian purple is a salt of a choline ester, that displays potent neuromuscular blocking action and muscle relaxing properties (Baker & Duke 1976, Roseghini *et al.* 1995). Tyrian purple precursor synthesis varies between muricid species (Baker 1974), and a better understanding of muricid evolution is required in order to understand the mechanisms involved in the synthesis of these bioactive compounds.

1.2.3 Muricid taxonomy and phylogeny

Australia has close to 150 different species of muricids and approximately 1600 species exist internationally (Tan 2003, Barco *et al.* 2010). However, the phylogenetic classification of species within this family is a complex field, with limited studies dedicated to resolving muricid classifications at the subfamily level. Molluscan taxonomic classifications have historically been presented based on shell morphology and anatomical characters (Kocot *et al.* 2011), however disparity between morphological characters and of family members has made Muricidae taxonomy difficult to resolve. While the application of molecular phylogenetics in the field of systematic classification was initially met with significant trepidation (Lipscomb *et al.* 2003, Seberg *et al.* 2003, Will & Rubinoff 2004), molecular taxonomy is an informative and useful tool in systematic classifications, particularly in areas where traditional morphological data is inconclusive (Barco *et al.* 2010, Kocot *et al.* 2011).

The taxonomic analysis of the Rapaninae subfamily by Kool (1993a), based on gross anatomical, radular, opercular and protoconch morphology combined with shell ultrastructure, reported that an updated systematic classification was required (Kool 1993a). Species that were previously classified as belonging to the subfamily Thaididae/nae were reclassified due to

the morphologically derived paraphyletic grouping of the clade, and it was proposed that the species be reclassified to belong to either the newly declared Rapaninae subfamily or the Ocenebrinae subfamily (Kool 1993a). An additional study by Kool (1993b) was performed to investigate the classification of the muricid genera *Nucella*, *Trophon* and *Ocenebra* based on anatomy, radular, protoconch, operculum and shell ultrastructure. Kool (1993b) concluded that these three genera appeared more closely related to each other than to representatives of the Rapaninae subfamily, and suggested that members of the *Nucella* genus be classified in the subfamily Ocenebrinae and that members of the *Trophon* genus be classified in the subfamily Trophoninae (Kool 1993b). Vermeij and Carlson (2000) reviewed the subfamily Rapaninae classification performing a phylogenetic analysis on the subfamily, looking at shell characteristics incorporating fossil evidence as well as gross morphological characters (Vermeij & Carlson 2000). Their results conclude that shell characteristics alone are not effective in resolving phylogenetic relationships, but they do have their place in the phylogenetic analysis, particularly in supporting lower-level relationships (Vermeij & Carlson 2000). The revision of taxonomy of Australian and New Zealand muricid species by Tan (2003) supported the classification of subfamilies Ocenebrinae and Rapaninae, and proposed a new subfamily Haustrinae (Tan 2003).

In addition to resolving the taxonomic classification of species, a robust phylogeny can be used to investigate the evolution of specific character traits within this family. A study into the development of labral spines in Ocenebrinae muricids utilized cytochrome oxidase and 12s ribosomal RNA sequences to construct the molecular phylogeny and found that the development of labral spines from marine gastropods is a trait that has developed more than once in evolutionary history (Marko & Vermeij 1999). In 2001, Oliverio and Mariottini investigated the molecular framework and phylogeny of the Corallilophila subfamily of muricids, which confirmed the monophyletic relationship observed in this subfamily (Oliverio & Mariottini 2001). Oliverio *et al.* (2002) investigated the molecular phylogeny of Muricidae family members utilizing ITS2 (internal transcribed spacer region 2) and identified the Rapaninae as a closely related sister clade to the Corallilophilae subfamily, but the bootstrap values from maximum parsimony analysis showed weak support for several clade formations (Oliverio *et al.* 2002). The largest molecular investigation into muricid phylogeny was published by Barco *et al* (2010) and resolved the monophyly of six of the ten currently classified muricid subfamilies, as well as confirming the monophyly of the Muricidae family of Neogastropods (Barco *et al.* 2010). This phylogenetic analysis of the Muricidae was only possible by performing a multigenetic study

using three mitochondrial sequences (12s, 28s ribosomal RNA and cytochrome oxidase subunit I), as well as one nuclear sequence (16s ribosomal RNA). While the relationship between all muricid subfamilies was not resolved using this dataset, Barco *et al.* (2010)'s study is, to date, the most thorough investigation into muricid taxonomy that has been presented and will greatly influence future classification of this cosmopolitan family of Neogastropods.

1.2.4 *Dicathais orbita* as a model for Tyrian purple biosynthesis

Dicathais orbita (Fig 1.4) is a marine predator found on shallow subtidal and intertidal rocky reefs across the southern coasts of Australia and New Zealand (Woodcock & Benkendorff 2008). *D. orbita* is an ideal candidate for gene expression studies due to the extensive work that has been performed on Tyrian purple production of this species. The Tyrian purple indole precursors were first identified in *D. orbita* (Baker & Duke 1976). The bioactivity of these indole compounds was also first identified from *D. orbita* extracts (Benkendorff *et al.* 2000, Vine *et al.* 2007, Benkendorff *et al.* 2011). Preliminary investigations into the aquaculture of the snail (e.g. Woodcock and Benkendorff, 2008; Noble *et al.* , 2009), as well as a detailed anatomical and histochemical investigation into the enzymatic synthesis of bioactive Tyrian purple precursors has been performed (Westley 2008, Westley & Benkendorff 2008, Westley & Benkendorff 2009, Westley *et al.* 2010a, Westley *et al.* 2010b). The strong supportive information regarding the activity of enzymes involved in the formation of these bioactive compounds makes *D. orbita* the ideal model organism in which to investigate Tyrian purple biogenesis from a gene expression standpoint.

1.3 Sustainable supply of marine bioactives

While marine compounds may hold potential for the development of new pharmaceuticals, nutraceuticals and medical treatments, without a sustainable supply development will cease.

When trying to obtain commercial quantities of marine bioactive compounds, several strategies can be employed, including aquaculture, fisheries development and synthetic production (Benkendorff 2009). In order to harvest commercial quantities of bioactive compounds directly from wild harvested marine organisms, you must ensure that fishing of the target species is a sustainable industry that can withstand the commercial demand. Alternatively, the development or implementation of aquaculture industry for the target organism could be employed; ensuring adequate commercial quantities are produced. Finally, it may also be possible to produce bioactive compounds using a synthetic route, using current advances and knowledge of chemistry and biochemistry to synthesize compounds of interest.

Figure 1.4 *Dicathais orbita* laying egg capsules in the marine aquaria.
Two female *D. orbita* laying egg capsules on rock substrate housed within the marine aquaria at Flinders University, South Australia. Egg capsules are visible on the substrate in varying stages of development. The purple pigmentation of older capsules is due to the presence of Tyrian purple.

1.3.1 Chemical synthesis

Chemical synthesis is always the preferred option for large scale supply for the pharmaceutical industry (Benkendorff 2009). The chemical structure of several marine natural compounds allows chemical synthesis in order to obtain commercial quantities. Ara-A and Ara-C are antiviral and antileukemic compounds that were the first marine-derived pharmaceuticals on the drug market and are synthesized via microbial fermentation of a chemical analogue and additional chemical synthesis (Sipkema *et al.* 2005). A synthetically produced analogue of the cone shell toxin ω-conotoxin MVIIA, also known as Ziconotide, is marketed as a potent painkiller (Prialt), due to its potent neuromuscular blocking action (Garber 2005). Ecteinascidin 743, a drug with anticancer activity derived from the Caribbean tunicate *Ecteinascidia turbinata* was only further developed as a viable pharmaceutical once a large scale semi-synthetic production was developed. Eribulin mesylate, a synthetic isomer of the marine sponge anti-tumor agent Halichondrin B is currently involved in phase III clinical trials in the treatment of breast cancer in Europe and USA (Molinski *et al.* 2009). If chemical synthesis is not possible,

other techniques, such as harvesting from the wild, may be implemented in order to supply the pharmaceutical industry with the required bioactive compounds.

1.3.2 Wild harvest

The development of a sustainable fisheries industry may provide marine bioactives for the pharmaceutical or nutraceutical industry, depending on the target species in question. Sufficient quantities of the marine sponge derived compounds manoalide, bryostatin-1 and avarol were isolated from sponges obtained from their natural environments for preclinical trials (Schaufelberger *et al.* 1991, Sipkema *et al.* 2005). However, it is unlikely that any of these compounds could be sustainably harvested from the wild should the pharmaceutical industry require commercial quantities. It has been reported that for preclinical and clinical trials on natural products derived from marine organisms, up to 1000kg of the source organism is required and thousands of metric tonnes would be required per annum to maximise investment returns from from a patented product, to cover licensing costs (Benkendorff 2009). With such large quantities required for a sustainable marketable agent, it is clear that wild harvesting could have long-term impacts on the sustainability of target populations and therefore the biodiversity of the oceans. One notable exception may be the northern pacific sea star, *Asterias amurensis* which exhibits anti-inflammatory and anti-cancer activity (Fernandez *et al.* 2005). It has previously been shown that 6.5kg of the starfish *A. amurensis* are required to produce anywhere from five to 12mg of bioactive asterosaponins from chemical extracts (Hwang *et al.* 2011). This marine star is considered a pest in the waters off the coast of Melbourne in Australia, and due to its prolific nature and abundance in coastal Australian waters it may actually sustain viable commercial harvesting of bioactives if a market arises in the future. Glucosamine, used in the treatment of joint inflammation and rheumatoid arthritis (Towheed *et al.* 2005), is almost entirely harvested from crustaceans and shark cartilage, taking advantage of the chitin found in the exoskeletal and cartilage waste products involved in the fishing industry (Maria *et al.* 2008). However, some shark fisheries are now considered overfished and the blackmarket practice of removing fins from live sharks is ethically questionable. Aquaculture may be an alternative that will facilitate the adequate production of marine bioactive compounds if wild harvesting and chemical synthesis fails to provide adequate supplies for the pharmaceutical industry.

1.3.3 Aquaculture

As the global population gets larger and larger, human dependency on the fisheries industry to provide adequate nutrition is resulting in the worldwide collapse of fish populations, due to overharvesting and ineffective management (Zeller & Pauly 2005). The aquaculture industry has been established in order to sustain not only the growing seafood demand, but in order to supply other marine based products that are required in the healthcare industry. The nutraceutical Lyprinol®, produced from the farmed green-lip mussel, is a stable lipid extract that is used to minimize arthritis symptoms and severity (Brien *et al.* 2008) and has the potential to reduce ameliorating symptoms of inflammatory bowel disease (Tenikoff *et al.* 2005). Microalgal culture is also being employed to produce omega-3 fatty acids due to the stability, quality and low cost of production in comparison to marine animal lipids (Lebeau & Robert 2003). Sea-based or land-based aquaculture may also be used to obtain sponge-based bioactives in the future, with several sponge species being successfully grown in aquaculture systems, although not all sponge species can be commercially produced in aquaculture systems (de Caralt *et al.* 2003, van Treeck *et al.* 2003).

1.3.4 Microbial biosynthesis

It is interesting to note that several bioactive compounds from marine sponges are produced by bacterial and fungal symbionts living within the sponge (Thomas *et al.* 2010). The aquaculture of host species is then used as a culture media, supporting the proliferation of symbionts and increasing the synthesis of target compounds (Hentschel *et al.* 2006). Although there are problems with current methods for culturing marine sponges for the purposes of harvesting bioactive secondary metabolites (Proksch *et al.* 2003), our current capacity to culture these bioactive producing microbes *ex situ* has been estimated to have a success rate of <1% (Fortman & Sherman 2005). If we are unable to obtain these secondary metabolites by direct microbial culture methods, the culturing of these secondary metabolites within the host sponge is the only alternative method of ensuring adequate sustainable supply (Thomas *et al.* 2010).

Standard microbial techniques are successful at culturing marine bacterium *Micromonospora sp*. from Indonesian marine sponges leading to the large-scale vat fermentation of manzamine-producing cultures (Taylor *et al.* 2007). Manzamine compounds from *Micromonospora* sp. exhibit antimalarial activity (Ang *et al.* 2000). Actinomycete symbionts of the marine sponge *Craniella australiensis*, are responsible for the production of broad spectrum antimicrobial

agents within their sponge hosts, and have been successfully isolated and cultured in the laboratory in order to maintain sustainable supplies (Li & Liu 2006).

Recent studies have focused on metagenomic experiments, profiling the microbial communities present within marine sponges, identifying the bacterial enzymes responsible for the production of bioactive secondary metabolites, then using recombinant expression systems to facilitate synthesis using freely available source metabolites (Taylor *et al.* 2007). The antibiotic Cyanosafracin B, is produced via a large scale fermentation of bacteria *Pseudomonas fluorescens* and is utilized as a starting material to produce commercial quantities of the marine-derived antitumor agent Ecteinascidin ET-743 (Cuevas *et al.* 2000). The partial or complete synthesis of biologically active marine compounds is crucial in the further development of these compounds as pharmaceutical agents, highlighting how important sustainable supply is in the successful application of marine natural products for medical applications. While it is not currently known whether Tyrian purple precursor synthesis is influenced by microbial symbiosis, further studies are required to fully understand the synthesis of these compounds in the Muricidae if a sustainable supply of these compounds is to be achieved.

1.3.5 The development of muricid bioactives as pharmaceutical or nutraceutical agents.

When attempting to develop indole compounds from muricid molluscs into pharmaceutical or nutraceutical products it is imperative that commercial quantities of the compounds are produced. It is unlikely that fishing of target species would provide enough bioactive compounds for the health care industry without decimating wild populations. Populations of the Southern American Mucicidae *Concholepas concholepas* have been decimated in Chile resulting in fisheries closure (Disalvo 1994). Abalone farms and oyster leases in South Australia provide adequate numbers of the muricid *D. orbita* in order to perform preclinical trials investigating the anti-cancer effects of muricid bioactive indole extracts (Benkendorff 2009). However, the pharmaceutical or nutraceutical market would require substantially more source materials than current suppliers could provide. In order to farm muricid species for use in medical treatments, the breeding cycle of muricids needs to be replicated in an aquaculture system. Pilot hatching and rearing techniques for culturing the South American *Concholepas concholepas* (Manriquez *et al.* 2008) and *Hexaplex trunculus* (Vasconcelos *et al.* 2004) have been performed, but an economically viable muricid aquaculture industry has yet to be

developed. Some of the bioactive precursors of Tyrian purple are commercially available (e.g. indole and 6' bromoisatin via TCI chemicals and API chemicals) whereas others cannot be produced synthetically in commercial quantities (i.e. tyrindoleninone (Vine *et al.* 2007)). One possible solution for the commercial production may be a semi-synthetic process, where mollusc-derived enzymes are recombinantly produced and used to facilitate synthetic chemical production. By identifying the individual genes responsible for the production of muricid bioactive indoles we may be able to facilitate large-scale production of these compounds for a commercial market.

1.4 Molluscan genomics and bioinformatics

While traditional biological experiments such as histology, biochemical analysis, enzyme kinetics and field studies allow us to gain an insight into how biological systems function, there are few studies that can produce the volume of information that is generated from large scale gene sequence analysis. Since the completion of the human genome project in 2001 (Venter *et al.* 2001), the field of genomics has grown exponentially and there are currently (as of October 2011) 1786 prokaryotic and 698 eukaryotic genome sequence projects in progress, under annotation or completely sequenced in the National Centre of Biotechnology Information (NCBI) public database. Furthermore, the field of transcriptomics is a burgeoning area of research, allowing researchers to not only identify species-specific global gene expression, but to compare global expression levels within different tissues or under different environmental conditions or disease states. Molluscan sequencing projects have been fairly under-represented in the current genomics climate, with only one genome in draft assembly stage (Moroz *et al.* 2004), and three genome projects in progress (www.newswise.com 2005, Raghavan & Knight 2006). Nevertheless, there have been several studies investigating the transcriptomes of specific molluscs, largely focussing on three areas of research: Biomineralization (Jackson *et al.* 2006, Joubert *et al.* 2010), investigations into molluscan interactions with the Schistosome parasite (Oliveira *et al.* 2004, Lockyer *et al.* 2007) and the use of molluscs as model organisms in neurological research (Moroz *et al.* 2006, Feng *et al.* 2009). There have also been a handful of transcriptome studies investigating environmental and immune responses in molluscs, as well as developmental processes and specific cellular and physiological mechanisms. A summary of molluscan transcriptome information has been listed in Table 1.1.

Table 1.1 Summary and location of Molluscan transcriptome data. Total number of EST sequences are listed, or where next generation sequencing was used, the total number of assembled contigs is listed. Locations are defined as follows; NCBI EST pertains to the EST databse in Genbank (http://www.ncbi.nlm.nih.gov/nucest), NCBI SRI pertains to the short read archive database (http://www.ncbi.nlm.nih.gov/sra). DDBJ pertains to the DNA database of Japan short read archive (http://www.ddbj.nig.ac.jp/). MG RAST pertains to the metagenomics RAST server (http://metagenomics.anl.gov/).

Species	# of ESTs/ Contigs	Location	Reference
Aplysia californica	267411	NCBI EST	(Moroz *et al.* 2006, Walters & Moroz 2009, York *et al.* 2010)
Lottia gigantea	252093	NCBI EST	http://genome.jgi.doe.gov/Lotgi1/Lotgi1.home.html
Biomphalaria glabrata	86936	NCBI EST	(Guillou *et al.* 2007, Lockyer *et al.* 2007)
Plakobranchus cellatus	77648	NCBI SRA	(Waegele *et al.* 2011)
Mytilus galloprivincialis	67942	NCBI EST	(Pantzartzi *et al.* 2010)
Radix balthica	54450	published as additional material in journal	(Feldmeyer *et al.* 2011)
Crepidula fornicata	39897	NCBI SRA	(Henry *et al.* 2010)
	29682	DDBJ SRA	(Kinoshita *et al.* 2011)
Elysia timida	24200	NCBI SRA	(Waegele *et al.* 2011)
Concholepas concholepas	19219	MG RAST	(Cardenas *et al.* 2011)
Laternula elliptica	18290	NCBI SRA	(Clark *et al.* 2010)
Pinctada margaritera	15606	NCBI EST	(Berland *et al.* 2011)
Lymnaea stagnalis	12287	NCBI EST	(Feng *et al.* 2009)
Hyriopsis cumingii	10156	NCBI EST	(Bai *et al.* 2010)
Mizuhopecten yessoensis	9100	NCBI EST	(Meng *et al.* 2010)
Haliotis asinina	8341	NCBI EST	(Jackson *et al.* 2005, Jackson & Degnan 2006, Jackson *et al.* 2006, York *et al.* 2010)
Pinctada maxima	7099	NCBI EST	(Jackson *et al.* 2010)
Ruditapes decussatus	4646	NCBI EST	(Gestal *et al.* 2007)
Chlyamys farreri	3537	NCBI EST	(Wang *et al.* 2009)
Pinctada fucata	1374	NCBI EST	(Fang *et al.* 2011)

The development of nacre and shell formation in molluscs holds particular importance due to the commercial pearl industry, and the potential of molluscan biomineralization in dental and bone regeneration studies (Atlan *et al.* 1997, Westbroek & Marin 1998), as well as its application in materials sciences (Lin & Meyers 2005). Transcriptomics was first used in 2006 by Jackson *et al.* , in order to investigate the shell secretome of the tropical abalone *Haliotis asinina* (Jackson *et al.* 2006). Since this pioneering study, there have been further investigations into pearl biomineralization in several bivalve species, including the black-lipped

pearl oyster *Pinctada margaritifera* (Joubert *et al.* 2010, Berland *et al.* 2011) the freshwater pearl mussel *Hyriopsis cumingii* (Bai *et al.* 2010), the Antarctic bivalve *Laternula elliptica* (Clark *et al.* 2010) and the pearl oyster *Pinctada fucata* (Fang *et al.* 2011). Interestingly, it has been shown that transcript information varies greatly between gastropods and bivalves. The analysis of the transcriptomes from *H. asinina* and *Pinctada maxima* showed that nacre biomineralization is a process that has arisen via convergent evolution in these two molluscan classes (Jackson *et al.* 2010). It has even been shown that transcripts involved in pearl formation and mother-of-pearl formation in the same animal are vastly different, when significant differences in gene expression in nacre-secreting mantle tissue and the pearl sac of *P. fucata* was reported (Kinoshita *et al.* 2011). It is likely that the application of transcriptomics will be further utilised in other mollusc species in the future, and this may facilitate a much deeper understanding of how nacre and shell structures are synthesised and maintained.

Schistosomiasis, a disease caused by trematode parasites, is prevalent in over 76 countries, affecting over 200,000 individuals and causing an average of 20,000 deaths each year (Oliveira *et al.* 2004). Trematode parasites responsible for this disease, *Schistosoma mansoni* and *Schistosoma japonicum*, spend a proportion of their lifecycle in their secondary host, freshwater snails belonging to the *Biomphalaria* genus, which plays a significant role in the spread of the disease (Reeves *et al.* 2008). A genome sequencing project is currently under way for the host snail *Biomphalaria glabrata*, with the aim of gaining a better understanding of the genetic processes involved in parasite lifecycles within this snail (Raghavan & Knight 2006). In order to further our understanding of snail-parasite interactions, several transcriptome projects have been performed on *B. glabrata*, investigating cell-signalling and transcriptional control (Lockyer *et al.* 2007) as well as anti-parasitic response mechanisms (Guillou *et al.* 2007). It is hoped that these studies will lead to new control mechanisms being developed that will reduce the spread of this disease, and future transcriptome investigations into this and other molluscan hosts will be vital in eradicating this problem for future generations.

The large size of some molluscan neurons, combined with the relative simplicity of mollusc nervous systems has made it an ideal model system for investigating neuronal interactions and neurodegenerative disease conditions (Sattelle & Buckingham 2006). It is primarily because of these implications in neuroscience that the *Aplysia* genome project was initiated in 2004 (Moroz *et al.* 2004). While still in the assembly stage, considerable sequence data is available from *Aplysia californica*, with 5165 nucleotide and 267,411 EST entries submitted to NCBI (as

of October 2011). The neuronal transcriptome of *A. californica* was first published in 2006 and provided functional data about genes expression in the central nervous system and in individual neurons, as well as the gene expression involved in sensory response mechanisms in this mollusc (Moroz *et al.* 2006). A recent investigation into genes expressed in the nervous system of *A. californica* identified that the molluscan nervous system shares the greatest sequence similarity with human neuronal sequences compared to other invertebrate model systems, and as such, makes this model organism the ideal species in which to investigate gene signalling in neurological disorders (Walters & Moroz 2009). The freshwater snail *Lymnaea stagnalis* is another model species used to investigate neuron signalling, and has undergone transcriptome sequencing, which identified a surprising amount of gene expression variation to *A. californica* (Feng *et al.* 2009). The study of molluscan neuronal systems is not limited to *A. californica* and *L. stagnalis*, as the transcriptome of the neuronal ganglia of *H. asinina* has been investigated in order to better understand growth regulation mechanisms (York *et al.* 2010). Because molluscs are a good model for neurological studies, it is likely that we will see a greater focus on large scale molluscan transcriptome and genome analysis in the field of neuroscience in the future.

While biomineralization, Schistisome parasite-interactions and neurological modelling have been the dominant areas of molluscan transcriptome research, there have been a variety of other studies which have used these techniques in order to better understand molluscan physiology, processes and populations. The immune system of molluscs has been investigated under parasite-infection conditions in the Shistosome intermediate-host *Biomphalaria glabrata* (Guillou *et al.* 2007), the Zhikong scallop *Chlamys farreri* (Wang *et al.* 2009) and in the hemocytes of the carpet-shell clam *Ruditapes decussatus* (Gestal *et al.* 2007). The transcriptomic effect of environmental stressors in *Mytilus galloprivincialis* has been investigated, where genes that are overexpressed with heat shock proteins were identified (Pantzartzi *et al.* 2010). In an effort to understand what genetic factors effect flesh quality in the Japanese scallop *Mizuhopecten yessoensis*, the transcriptome of the adductor muscle was sequenced and annotated (Meng *et al.* 2010). A considerable study investigating the transcriptome of early development in the common slipper shell *Crepidula fornicata*, was performed in order to better understand asymmetrical cell division (Henry *et al.* 2010). Transcriptomic developmental studies have also been performed on *H. asinina* both at the fundamental level (Jackson & Degnan 2006), as well as under starvation conditions (York *et al.* 2010). The opportunistic sequestering of photosynthetic plastids from algae in sacoglossan sea

slugs has been investigated using transcriptomics techniques and has identified that both *Elysia timida* and *Plakobranchus cellatus* do not import nuclear information from algal sources to maintain plastid activity, but partial digestion of plastids, as well as signal peptide interactions are likely responsible for the longevity of these photosynthetically active plastids in slug physiology (Waegele *et al.* 2011). A lot of these recent sequencing projects have been possible due to the use of next generation sequencing techniques, and to that end, the transcriptome of a non-model snail species (*Radix balthica*) was sequenced using Illumina technology, in order to compare sequence assembly software, and describe that a combined analysis using several software tools to create a meta-assembly was the best method to use when assembling sequence information from non-model organisms (Feldmeyer *et al.* 2011). Transcriptomics sequencing has also been a useful tool in determining molecular markers, for use in population genetics and genotyping studies, including in the Chilean muricid species *Concholepas concholepas* (Cardenas *et al.* 2011) and several bivalve species (Tanguy *et al.* 2008). To date the second largest molluscan transcriptome that is available is from the gastropod *Lottia gigantia*, which was produced in an effort to obtain a full length sequence, however, there has been no publication regarding the transcriptom of *L. gigantea* however EST sequences are available in the NCBI EST database and its genome is available through the JGI portal (http://genome.jgi.doe.gov/Lotgi1/Lotgi1.home.html).

Current molluscan sequencing projects typically focus on the commercially important species (Hedgecock *et al.* 2005, Jackson *et al.* 2005, Jackson & Degnan 2006, Saavedra & Bachere 2006) or those with an application or relevance to human health (Raghavan & Knight 2006, Gestal *et al.* 2007, Zhao *et al.* 2007). The identification of functionally expressed genes is not only beneficial to increase the general pool of knowledge on molluscs, but in the case of *D. orbita*, may allow us to identify key enzymes involved in the formation of bioactive Tyrian purple precursors. The potential of *D. orbita* as both a new aquaculture species (Woodcock & Benkendorff 2008) and in the development of bioactive compounds (Westley *et al.* 2010c, Benkendorff *et al.* 2011) makes it an ideal candidate for genomics research.

1. 5 Thesis aims, significance, structure and objectives

1.5.1 Thesis aims and significance

Through the use of molecular biology techniques and bioinformatics applications, this thesis aims to investigate and identify enzymes involved in the production of bioactive Tyrian purple precursors in the marine snail *D. orbita*. An investigation into Muricidae phylogenetics was undertaken in order to determine whether the presence of Tyrian purple precursors within the egg capsules of muricid molluscs is an ancestral or derived trait. In order to identify the enzymes involved in Tyrian purple biosynthesis, a transcritome approach was used to identify genes that are upregulated or uniquely expressed in the hypobranchial gland of *D. orbita* using suppressive subtractive hybridization. In order to best interpret our findings, a survey into the molluscan sequence homology trends was investigated, before thorough analysis and classification of genes was performed. The arylsulfatase enzyme that was identified from our transcriptome analysis was further studied, to obtain full length expressed sequences and studies were performed to investigate their potential recombinant expression. In terms of molluscan biology, it is hoped that the information gained from this study will further explain the role of the hypobranchial gland in molluscan biology, identify key enzymes involved in Tyrian purple production and determine whether indole transfer to egg capsules has occurred within a single lineage of the Muricidae, when mapped to a phylogenetic classification of Muricidae subfamilies. Ultimately it is hoped that the results of this thesis will assist in the development of Tyrian purple precursors as pharmaceutical agents, gaining both a better understanding of how these compounds are produced and facilitating the future production of bioactive compounds on a commercial scale.

1.5.2 Thesis structure

This thesis is presented in manuscript format. Although each chapter is intended for independent publication, or in the case of chapter three has already been published in a peer reviewed journal, the underlying concepts comprise a progressive body of research. To maintain continuity in presentation, all chapters with the exception of chapter three have been formatted in a consistent manner and have adopted the referencing format outlined in the Journal of Biological Chemistry. Chapter three is presented as published in Volume 123, issue 3 of *The Nautilus*. Whilst I am the first author on all chapters and personally responsible for the experimental design, conducting the research and preparing the manuscripts, the contribution

of additional authors on published manuscripts or those currently under review are outlined in the acknowledgements of each chapter. The following outlines the objective of each chapter and indicates the publication status.

1.5.3 Chapter objectives

Chapter two. The evolution of bioactive Tyrian purple precursors in the egg capsules of the Muricidae (Mollusca: Gastropoda).

Objective: Chapter two aims to identify whether the presence of brominated indoles in the egg capsules of muricid molluscs is a trait that is specific to individual subfamilies or is an ancestral trait that has been subsequently gained and/or lost in some lineages.

Chapter three. Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae). *Nautilus (2009) 123(3): 154-158.*

Objective: Chapter three investigates the trends observed from BLAST sequence homology analysis from a molluscan expressed sequence library. Patterns are reported based on e-values of the pairwise sequence matches. The taxa of the most significant blast sequence match for each sequence are investigated.

Chapter four. Annotation and characterization of a partial transcriptome of the hypobranchial gland of *Dicathais orbita*.

Objective: Chapter four uses suppressive subtractive hybridization to create a cDNA library that contains genes that are expressed within the hypobranchial gland of *D. orbita*. Genes are annotated based on sequence homology results using BLAST2GO and the identified biological processes are further discussed. A subset of sequences identified from this study display an expression profile similar to ciliate protozoans which is further investigated in Chapter five. A partial sequence for an arylsulfatase enzyme was also identified from this study which is further investigated in Chapter six.

Chapter five. Novel application of suppressive subtractive hybridization for the identification of symbionts: Discovery of ciliate protozoa in the hypobranchial gland of *Dicathais orbita* (Neogastropoda, Mollusca).

Objective: Chapter five investigates the surprise finding of ciliate protozoa present within the hypobranchial gland of *D. orbita* based on gene expression patterns, phylogenetic analysis of 18s ribosomal RNA sequences and histological analysis of the hypobranchial gland. This investigation, while not an expected result of our sequence analysis, is important in understanding the biology of the hypobranchial gland of *D. orbita* and in explaining complex symbiotic relationships that exist within the gland.

Chapter six. Characterisation and expression of recombinant arylsulfatase from the marine snail *Dicathais orbita.*

Objective: Chapter six investigates the identification of a full length arylsulfatase gene isolated from the hypobranchial gland and our attempts to express this in HEK293t mammalian tissue culture expression systems. This investigation aims to clone one of the enzymes involved in the production of bioactive Tyrian purple precursors, allowing synthesis of these bioactive compounds *in vitro*.

Chapter seven. The use of molecular techniques for the investigation of Tyrian purple precursor synthesis in the Muricidae: Outcomes and future directions.

Objective: This chapter provides a summary of the research findings of this thesis and expands on individual chapter discussions by investigating gene expression as a whole in the hypobranchial gland of *D. orbita*. This chapter also highlights future directions for the investigation of remaining Tyrian purple enzymes, the production of bioactive precursors and how our findings relate to other muricid species.

Three Appendices are included in this thesis; Appendix I is supplementary to Chapter two and lists all additional phylogenetic analyses performed; Appendix II is supplementary to Chapter six and details recombinant arylsulfatase purification and peptide sequencing; Appendix III details additional investigations performed in order to amplify additional arylsulfatase, bromoperoxidase and tryptophanase gene sequences from *D. orbita*.