Ana Patrícia Carreira Rodrigo

Mestre em Engenharia do Ambiente

The biotechnological value of a novel potent marine biotoxin from the polychaete worm *Eulalia viridis*: chemical and toxicological evaluation

Dissertação para obtenção do Grau de Doutor em
Ambiente e Sustentabilidade

Orientador: Prof. Doutor Pedro Manuel Broa Costa, Professor Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Co-orientador: Prof. Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes, Professora Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
Prof. Doutora Maria Helena Costa, Professora Catedrática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutora Maria Paula Antunes
Arguente(s): Prof. Doutora Maria Leonor Quintais Cancela da Fonseca
Prof. Doutor Henrique Manuel Roque Nogueira Cabral
Prof. Doutor Manu Soto
Prof. Doutor Pedro Manuel Brôa Costa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro de 2020

Ana Patrícia Carreira Rodrigo

Mestre em Engenharia do Ambiente

The biotechnological value of a novel potent marine biotoxin from the polychaete worm *Eulalia viridis*: chemical and toxicological evaluation

Dissertação para obtenção do Grau de Doutor em
Ambiente e Sustentabilidade

Orientador: Prof. Doutor Pedro Manuel Broa Costa, Professor Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Co-orientador: Prof. Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes, Professora Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa
Prof. Doutora Maria Helena Costa, Professora Catedrática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri:

Presidente: Prof. Doutora Maria Paula Antunes
Arguente(s): Prof. Doutora Maria Leonor Quintais Cancela da Fonseca
Prof. Doutor Henrique Manuel Roque Nogueira Cabral
Prof. Doutor Manu Soto
Prof. Doutor Pedro Manuel Brôa Costa

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Novembro de 2020

ACKNOWLEDGEMENTS

"Do what you love and you will never have to work a day in your life"

Even though this is true, if you are not surrounded by great people, what you love would not be sufficient.

As such, I would like to start with a special thank you to all my friends from Seatox Lab and specially Pedro, for helping and accompanying me through all the journey. To Cátia and Carla for accompanying me for more than 10 years in all the parties, late night school papers and all the hard work in the laboratory. To Carolina and Diana M. for the scientific motivation and the company in all the lunches and fun dinners. To Mariaelena for the laughs, the motivation and all the ecological and sustainable insight. To Nagore, for the ecological insight and the company during the beach life. To Diana R., despite not being part of the group, is part of the family. Also, a special thank you to all the bachelor and master students that passed through the lab that brought a fresh environment to the lab. Despite small, the Seatox team was essential to walk the walk, from sampling and laboratory work, to keeping the sugar levels up with all the pastry that able me to relax and keep working even more motivated. Also, for believing in me even in situations where I don´t and to always keep the best work environment that someone could have.

A thank you to my supervisor Alexandra and to the Human Genetics and Cancer Therapeutics lab for all the supervision, warm welcoming into their group and help in the most challenging part of my PhD. A thank you to Pedro Baptista and the whole group for giving me all the technical support and overall picture whenever needed. To Cynthia and Bruno, for making a second workplace, a second home. Also, a thank you to Ana Rita and the Computational Multi-Omics Lab for all the precious help in transcriptomics, for making all the lunches a fun part of the day and extending it to the coffee breaks.

To all my friends that accompanied me and hear whenever necessary and help me to stress down and relativize the work during the much-needed breaks in all the extra-activities. To Dulce, Sara and Andreia for all the craziness since primary school and helping me to became who I am. To Rita F. and Tânia for the great times since high school. To Rita, Teresa, Marta and Lopes for the company in all the marathons. To Hugo for the best working buddy in all the cafes and late nights. To Poggi, Tiago, Pedro A., Bruno, Durães and the aforementioned for all the waves shared. To Ana, Papa, David, Igna, Edgar, Macau, Rita, Sofia, Rasta, Katha, Mário and Teresa for being the best climbing buddies, to motivate science discussions and relaxation in the best spots.

To my parents for dealing with the hard times and celebrate the good ones and to always keep interest and proud in my work and life. To my sister for hearing me daily and always be close even living far away. To my nephew, grandpa, aunts, uncles and cousins for always remembering that the basis of life is family and family leads to happiness. And finally to Fred, for all the fresh meals, hugs, always having my back and accompany me in every situation, even when not fully understanding them.

RESUMO

O oceano contém muitos organismos por explorar com características adaptativas que os permitem prosperar no seu meio ambiente. *Eulalia viridis* é um desses organismos. Esta poliqueta verde clara é predador do intertidal rochoso sem mandibulas, equipado dum muco tóxico, permitindo assim predar invertebrados maiores extraindo pedaços de tecido através da sucção. Motivado pelo valor das toxinas como substâncias bioreactivas e pela ambição da UE de conduzir a Estratégia para o Crescimento Azul, à qual a vasta ZEE portuguesa oferece grandes perspetivas, o potencial biotecnológico desta espécie foi analisado. A análise microanatómica desta minhoca revelou adaptações cruciais na alimentação: existência de papilas sensoriais e células responsáveis pela secreção de muco e toxinas. A sequenciação completa do transcritoma e ensaios toxicológicos revelaram um complexo padrão de toxinas proteicas e enzimas com diversas funções: permeabilização de tecidos, anti-coagulação e bloqueio da atividade neuromuscular. As principais neurotoxinas da *Eulalia*, "filotoxinas", são proteínas ricas em cisteínas, que atuam como agentes imobilizadores em mexilhões e outros invertebrados, que são administradas por contacto através do muco. Por sua vez, uma maior toxicidade para com a linha celular do cancro do ovário (A2780) do que células normais, demonstraram propriedades citotóxicas e citoestáticas através da paragem do ciclo celular e indução de morte celular programada pela via extrínseca. As proteínas responsáveis por estes efeitos foram identificadas, através de transcritómica e proteómica. As secreções mucosas da minhoca contêm também complexos proteicos internalizáveis que apresentam fluorescência, cuja emissão é reversivelmente regulada pelo estado redox. Resumindo, as características encontradas nas secreções de uma espécie demonstram o vasto potencial biotecnológico de anelídeos marinhos, como uma fonte valiosa de bioreactivos, mesmo em águas temperadas. O percurso metodológico utilizado, a combinação da ecologia, toxicologia e biologia molecular, permitiu contornar dificuldades da bioprospecção marinha, revelando ser um importante contributo para exploração sustentável de novos biorecursos marinhos.

Palavras-chave: Economia Azul; Ensaios *in vitro*; Muco; Filotoxinas; Proteómica; Transcritómica.

ABSTRACT

The ocean's vastness holds still many unexplored organisms with unique adaptive features that enable them to thrive in their environment. *Eulalia viridis* is one of them. This uncanny bright green worm is a predatory marine Polychaeta of the rocky intertidal that lacks jaws, but it is equipped with a toxin-containing mucus that enables predating larger invertebrates from which soft tissue is extracted through suction. Motivated by the high-value of toxins as bioreactives and by the European Union's ambitions of leading the Blue Growth Strategy, for which Portugal's vast EEZ offers high perspectives; a case-study to explore the biotechnological potential of this species was undertaken. The worm's microanatomy revealed key adaptations for feeding such as special sensorial papillae and cells responsible for the secretion of mucus and toxins. Whole-transcriptome sequencing and toxicity assays yielded a complex pattern of proteinaceous toxins and enzymes with several functions: tissue permeabilization, coagulation impairment and blocking of neuromuscular activity. The main neurotoxins, "phyllotoxins", found to be cysteine-rich proteins, act as immobilising agents against mussels and other invertebrate prey and are delivered by contact with mucus. In turn, higher toxicity towards ovarian cancer cell line (A2780) than normal cells, involved cytotoxic and cytostatic properties through cell cycle arrest and extrinsic programmed cell death. Several proteins involved in these effects were identified by combining transcriptomics and proteomics. The worm's mucosecretions also hold internalisable proteinaceous complexes that display fluorescent properties and whose emission is reversibly switched by redox status. Altogether, the features found in the secretions of a single species disclose the immense biotechnological potential of marine annelids, and invertebrates in general, as source of valuable bio-reactives, even in temperate waters. The methodological pipeline, combining ecology, toxicology and molecular biology, circumvented many difficulties of marine bioprospecting and is an important contribution for the sustainable exploitation of novel marine bioresources.

Keywords: Blue Economy; *in vitro* assays; Mucus; Phyllotoxins; Proteomics; Transcriptomic.

LIST OF PUBLICATIONS

Rodrigo AP, Lopes A, Pereira R, Mendes V, Manadas B, Grosso AR, Baptista PV, Fernandes AR, Costa PM (2020). Endogenous fluorescent proteins in the mucus of an intertidal Polychaeta. *Submitted*.

Rodrigo AP, Mendes V, Manadas B, Grosso AR, Alves de Matos AP, Baptista PV, Costa PM, Fernandes AR (2021). Specific anti-proliferative properties of proteinaceous toxin secretions from a marine annelid onto ovarian cancer cells. *Mar Drugs* 19, 31.

Rodrigo AP, Grosso AR, Baptista PV, Fernandes AR, Costa PM (2021) A transcriptomic approach to the recruitment of venom proteins in a marine Polychaeta. *Toxins* 13, 97.

Rodrigo AP, Costa PM (2019). The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. *Environ Res* 173, 270–280.

Cuevas N, Martins M, Rodrigo AP, Martins C, Costa PM (2018) Explorations on the ecological role of toxin secretion and delivery in jawless predatory Polychaeta. *Sci Rep* 8, 7635.

Rodrigo AP, Martins C, Costa MH, Alves de Matos AP, Costa PM (2018) A morphoanatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). *J Anat* 233, 567–579.

<u>Conference Publications</u>

Rodrigo AP, Grosso A, Baptista PV, Fernandes AR, Costa PM (2020). Toxins from an unsuspected invertebrate: a poisonous cocktail from a jawless predatory polychaete. *Mar Drugs* 2020, 18(1), 40.

Rodrigo AP, Martins C, Tanoeiro L, Casaca M, Lopes AR, Parola AJ, Matos AP, Baptista PV, Fernandes AR, Costa PM (2019). New lessons from ancient life: marine invertebrates as a source of new drugs. *Ann Med*, 51(1), 45-45.

Rodrigo AP, Lopes AR, Baptista PV, Costa MH, Fernandes AR, Costa PM (2018). The great biotechnological potential of a marine polychaete: an alliance between toxin and natural fluorescence. *Front Mar Sci Conf Abstract*: IMMR'18, International Meeting on Marine Research.

LIST OF COMUNICATIONS

D'Ambrosio M, Gonçalves C, Martins C, Rodrigo AP, Madeira C, Costa PM (2019). Emerging marine bioproducts and blue biotechnology: exploring the potential of marine invertebrates. 42nd CIESM Congress, Cascais (Portugal), October 2019. *Poster communication*.

Rodrigo AP, Grosso A, Batista PV, Fernandes AR, Costa PM (2019). Toxins from an unsuspected invertebrate: a poisonous cocktail from a jawless predatory polychaete. XVI International Symposium on Marine Natural Products & XI European Conference on Marine Natural Products, Peniche (Portugal), September 2019. *Poster communication*.

Rodrigo AP, Lopes AR, Baptista PV, Costa MH, Fernandes AR, Costa PM (2018). The great biotechnological potential of a marine polychaete: an alliance between toxin and natural fluorescence. International Meeting on Marine Research, Peniche (Portugal), Jul 2018. *Oral communication.*

Rodrigo AP, Martins C, Tanoeiro L, Casaca M, Lopes AR, Parola AJ, Matos AP, Baptista PV, Fernandes AR, Costa PM (2018). New lessons from ancient life: marine invertebrates as a source of new drugs. CiiEM, Monte da Caparica (Portugal), Jun 2018. *Oral communication*.

Cuevas N, Martins M, Rodrigo AP, Martins C, Costa PM (2018). Não é preciso morder: o papel ecológico da secreção e libertação de uma toxina de uma poliqueta predatória. Primeiro Congresso de Biologia Marinha, Faro (Portugal), Jan 2018. *Oral communication.*

Rodrigo AP, Martins C, Lopes A, Alves de Matos AP, Batista PV, Fernandes AR, Costa PM (2018). Unravelling the mysteries of a marine polychaete: connecting toxicity with fluorescence. Encontro Ciência'18, Lisboa (Portugal), July 2018. *Poster communication*.

Rodrigo AP, Martins C, Cuevas N, Costa MH, Fernandes AR, Costa PM (2018). Caracterização química, molecular e toxicológica de uma nova biotoxina produzida por um poliqueta marinho: *Eulalia viridis*. Primeiro Congresso de Biologia Marinha, Faro (Portugal), January 2018. *Poster communication*.

LIST OF ABBREVIATIONS

A2780 – Ovarian carcinoma

A549 – Lung adenocarcinoma

AB – Alcian Blue

AChE – Acetylcholinesterase

ANOVA – Analysis of Variance

AO – Acridine Orange

ARSB – Arylsulfatase B

ATCC – American Type Culture Collection

AURKB – Aurora kinase B

BAG1 – BCL-2 associated athanogene family molecular chaperone regulator 1

BAX – Bcl-2-like protein 4

BCAS2 – Pre-mRNA-splicing factor SPF27

BCECF-AM – 2′,7′-Bis(2-carboxyethyl)-5(6)-carboxyfluorescein tetrakis(acetoxymethyl) ester

BCL-2 – B-cell lymphoma 2

BLASTP – Basic Local Alignment Search Tool for Proteins

BSA – Bovine serum albumin

$CaCl_2$ – Calcium chloride

CDK – Cyclin-dependent kinase

CDC5L – Cell division cycle 5-like protein

CWC15 – Spliceosome-associated protein CWC15 homolog

CHAPS – 3-[(3-Cholamidopropyl)-Dimethylammonio]-1-Propanesulfonate

CUR – Curtain gas

CWC15 – Spliceosome-associated protein CWC15 homolog

EEZ – Exclusive Economic Zone

ERK – Extracellular signal-regulated kinase

DAPI – 4',6-Diamidino-2-Phenylindole, Dihydrochloride

DMEM – Dulbecco’s modified Eagle’s medium

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic acid

DNAJB1 – DnaJ homolog subfamily B member 1

DNAJC6 – Putative tyrosine-protein phosphatase auxilin

DTT – L-dithiothreitol

EGFR – Epidermal Growth Factor Receptor

EST – Expressed sequence tag

EU – European Union

GAPDH – Glyceraldehyde 3-phosphate dehydrogenase

GA – Golgi apparatus

GEO – Gene Expression Omnibus

GFP – Green Fluorescent Protein

GUSB – Beta-glucuronidase

HCL – Hydrochloric acid

HCT116 – Human colorectal carcinoma

HIV-1 – Human Immunodeficiency Virus -1

HPLC – High Performance Liquid Chromatography

HSP90AA1 – Heat shock protein HSP 90-alpha

HSP90AB1 – Heat shock protein HSP 90-beta

HSP70 – Heat shock protein 70 kDa (kilodalton)

HSPA8 – heat shock cognate 71 kDa protein (kilodalton)

H&E – Haematoxylin and eosin staining

IDUA – Alpha-L-iduronidase

IEF – Isoelectric focusing

K562 – Myelogenic leukemic carcinoma

KRT10 – Keratin, type I cytoskeletal 10

LC-MS/MS – Liquid Chromatography with tandem mass spectrometry

MCF7 – Breast carcinoma

MEK – Mitogen-activated protein kinase

mTOR – Mechanistic Target of Rapamycin

MTS – 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2Htetrazolium

MVB – Multivesicular body

NaCl – Sodium chloride

NCBI – National Center for Biotechnology Information

NMR – Nuclear magnetic resonance

NP – Natural product

RAB3C – Ras-related protein Rab-3C

RAF1 – RAF proto-oncogene serine/threonine-protein kinase

RER – Rough endoplasmic reticulum

RNA – Ribonucleic acid

RNAseq – Ribonucleic acid sequencing

ROS – Reactive oxygen species

RPMI – Roswell Park Memorial Institute medium

RT-qPCR – Quantitative Polymerase Chain Reaction

OCT – Optimal cutting temperature medium

OsO_4 – Osmium tetroxide

PA – Picric Acid

PAS – Periodic Acid/Schiff's

PBS – Phosphate-buffered saline

PRSS1 – Trypsin-1

SEM – Scanning Electron Microscopy

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SP – Serine protease

STRING – Search Tool for the Retrieval of Interacting Genes/Proteins

SUMF1 – Sulfatase-modifying factor 1

SUMF2 – Sulfatase-modifying factor 2

TB – Toluidine Blue

TC – Tetrachrome stain

TTX – Tetrodotoxin

TEM – Transmission Electron Microscopy

UCH-L1 – ubiquitin C-terminal hydrolase-L1

UV – Ultraviolet

VG – van Gieson's trichrome

WH – Weigert's iron Haematoxylin

WHO – World Health Organization

YWHAB – 14-3-3 protein beta/alpha

YWHAZ – 14-3-3 zeta protein

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - Introduction

Part of the literature review in this chapter has been published in the following review paper:

Rodrigo AP, Costa PM (2019). The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. *Environ Res* 173, 270–280.

1.1. Introduction

1.1.1. Natural products for biotechnological applications

Natural products (NPs) can be defined as structures or compounds of biological origin that have been evolutionary perfected to interact with several key targets for defence or predation purposes (Koehn and Carter 2005). For this reason, natural products have been explored for medicinal usage for centuries by virtually all cultures, from Ancient Egypt to modern China and India (see the review of Rao et al. 2019). However, it was not until the 1950s that NPs, especially of microbial origin, have first been isolated, identified and engineered for therapeutic applications by Western medicine and biomedical research, reaching a first momentum in the 1970s and 1980s, motivated by cancer research, when the interaction of NPs in the immune response, cell cycle and cell death began to be unravelled (Koehn and Carter 2005, Rao et al. 2019, Xu et al. 2020). The search for NPs had some ups and downs, being however, difficult to measure, due to the time necessary from the discovery of NPs to their commercialization (Koehn and Carter 2005). Nevertheless, natural products remain very appealing to a variety of fields, from plague control, as insecticide, to therapeutic agents, as drugs (see, for instance, Wang et al. 2017). In the latter, their attractiveness is due to their ability to bind in the active sites of enzymes, acting as inhibitors of catalysis, bind to a variety of protein domains and folding motifs, to modulate or inhibit protein–protein interactions. As such, substances isolated from marine organisms, plants, fungus are now being used as sedatives, antimalarial, antifungal and even as cholesterol-lowering drugs (see the reviews of Koehn and Carter 2005 and Rao et al. 2019 for more details). Overall, NPs have shown to be great alternatives to the synthetic counterparts, due to better safety concerns for the environment, when using the synthetic compound and even during the synthesis. The costs of production are evidently, also a key factor for the appealing of the natural products.

Nowadays the search for natural products continues, but it requires faster screening and more detailed and efficient techniques. The major steps (extraction, concentration, fractionation and purification) are now simplified due to the vast techniques available to facilitate the process, such as HPLC-mass spectrometry (HPLC-MS), spectroscopic techniques (like NMR technologies), genomics, bioinformatic, metabolic engineering and chemical synthesis (Koehn and Carter 2005). Even though all these techniques boosted the discovery of substances, there are still several constraints mainly in the separation and purification processes and in the quantities necessary to perform all the techniques. Even though a large number of bioactive natural products have been discovered and isolated, few of their pharmacological targets have been clearly confirmed (Rao et al. 2019). In fact, several compounds have their therapeutic effects from the combined action of multiple components through their respective targets and pathways, being able to be approved as

drugs as a whole, instead of the components in separate (Koehn and Carter 2005; Rao et al. 2019). Despite all the scientific advances since the first discoveries, more molecular tools and strategies are still needed to make the search of natural product discovery more efficient (Xu et al. 2020).

1.1.1. Marine bioproducts and the Blue Growth Revolution

Marine biotechnology has gained focus over the past few years, even though the number of effective commercial applications of marine bioproducts can be, at first, somewhat disappointing, especially those from eumetazoans. In fact, motivated by the ocean's vast and ancient biodiversity, great efforts are being applied to discover new products for a wide range of potential applications, from biomedicine and foodstuff to eco-friendly pesticides, anti-foulants and cosmetics (see Molinski et al. 2009; Freitas et al. 2012; Martins et al. 2014). This endeavour has not been set astray from need to direct the exploitation of marine resources into the pathway of sustainability as an alternative to traditional invasive economic activities such as fishery. This premise, combined with the expansion of biotechnological enterprises and their demands for new bio-reactive compounds, placed the focus on that seemingly endless source of novel compounds – the seas. With such motivation at hands, the European Union (EU) introduced Blue Biotechnology as one of the pillars for the Blue Growth Agenda, whose ultimate goal is to ensure purposeful and sustainable usage of marine resources (Lillebø et al. 2017). However, more than a decade after the initial boost, the translation of research into commercial products is lagging (Mayer et al. 2010). It should not be a surprise that one of the reasons for the distance between bioprospecting and industry relates to difficulties in exploring such immense biodiversity and translating existing data into application.

1.1.2. Toxins take the lead in drug development from natural products

The search for new drugs placed the spotlight on organisms that developed chemical weapons for various purposes, since toxins have evolved to interact with specific ligands (e.g. enzymes and ion channels) and disarrange neuromuscular and cardiovascular systems, blood coagulation and general homeostasis, for instance (Calvete et al. 2009). It is thus far from unimaginable that the time-consuming, risky and expensive process of designing synthetic compounds is merely recapitulating the work already done by means of natural selection. Not surprisingly, some of the most studied marine organisms for biotechnological purposes are the ones that secrete venoms and poisons. These are usually complex proteinaceous secretions, produced in specialised glands, that contain molecules aiming at disrupting physiological processes in target organisms via binding to specific targets (Fry et al. 2009; Casewell et al. 2013). For disambiguation purposes Nelsen et al. (2014), suggested differentiating venoms from poisons based on the previous being injected into the target via an inflicted wound. The same authors introduced the term toxungen, which applies to toxin-

bearing secretions that are delivered without injury and by surface contact. These secretions are believed to hold a defensive function in most animals, to which is added an important role in predation, competition and chemical signalling as well (Calvete et al., 2009; Escoubas et al., 2008; von Reumont et al., 2014a). Altogether, the diversity and polyphyly of toxin-secreting animals suggest convergent evolution (Fry et al. 2009).

The reduced number of effective commercial applications of marine animal bioproducts, makes the painkiller Prialt (Ziconotide), developed from the venom of *Conus magus*, arguably the most prominent example (Olivera 2000; Safavi-Hemami et al. 2018). There are, however, other products from various marine animals already approved or in the queue for clinical trials and approval. Among these, the emphasis goes to sponges and anti-cancer drugs (refer to Mayer et al. 2010 for further details). Even though many of these potential pharmaceuticals are derived from poisons and venoms (von Reumont et al. 2014b; Hornbeak and Auerbach 2017), there is growing awareness that bioproducts of interest can be retrieved not only from other endogenous secretions but also from symbiotic microorganisms, encompassing a range of functions between protein inhibitors and biocides (Proksch et al. 2002; Freitas et al. 2012). Altogether, the identification of interesting bio-reactive compounds from taxa as distinct as Cephalopoda (Ueda et al. 2008), Cnidaria (Schlesinger et al. 2009) and Asteroidea (Karasudani et al. 1996) steers bioprospectors into thinking that we may be facing the sheer tip of the iceberg of novel compounds from the vast and rather unique diversity of marine invertebrates (Kastin 2006).

1.1.3. Why marine Polychaeta?

The Polychaeta are a major group among Annelida, forming a diverse, ubiquitous and ecologically-relevant group of aquatic invertebrates (see Struck et al. 2011). In the oceans, they occupy nearly every marine habitat, from the intertidal to deep-sea vents. However, research on marine Polychaeta for biotechnological purposes is relatively recent when compared to terrestrial or continental water annelids (mostly Oligochaeta). Indeed, biotechnology-oriented research on the latter already took the lead, at least with respect to anti-microbial compounds (Milochau et al. 1997; Bruhn et al. 2006) and anticoagulants from the freshwater *Hirudo* (see Basanova et al. 2002 for a compilation). Still, the first steps on biotechnological research on the Polychaeta has been motivated by incidental nefarious interactions with humans. It is the case of fireworms (Amphinomidae), a family that holds urticarial chaetae that cause injury to bathers, and bloodworms (*Glycera*) that, with their venomous jaws, are known to harm fishermen during handling as bait (Smith 2002; Hornbeak and Auerbach 2017). As more organisms from this taxon are being prospected, other types of potential bioproducts asides toxins and related come into focus, such as bioluminescent probes from *Chaetopterus* (Branchini et al. 2014), as compiled in Fig. 1.1. Altogether, these clues suggest that the diversity

and abundance of Polychaeta can make them an invaluable case-study for marine bioprospecting for a wide range of biotechnological purposes.

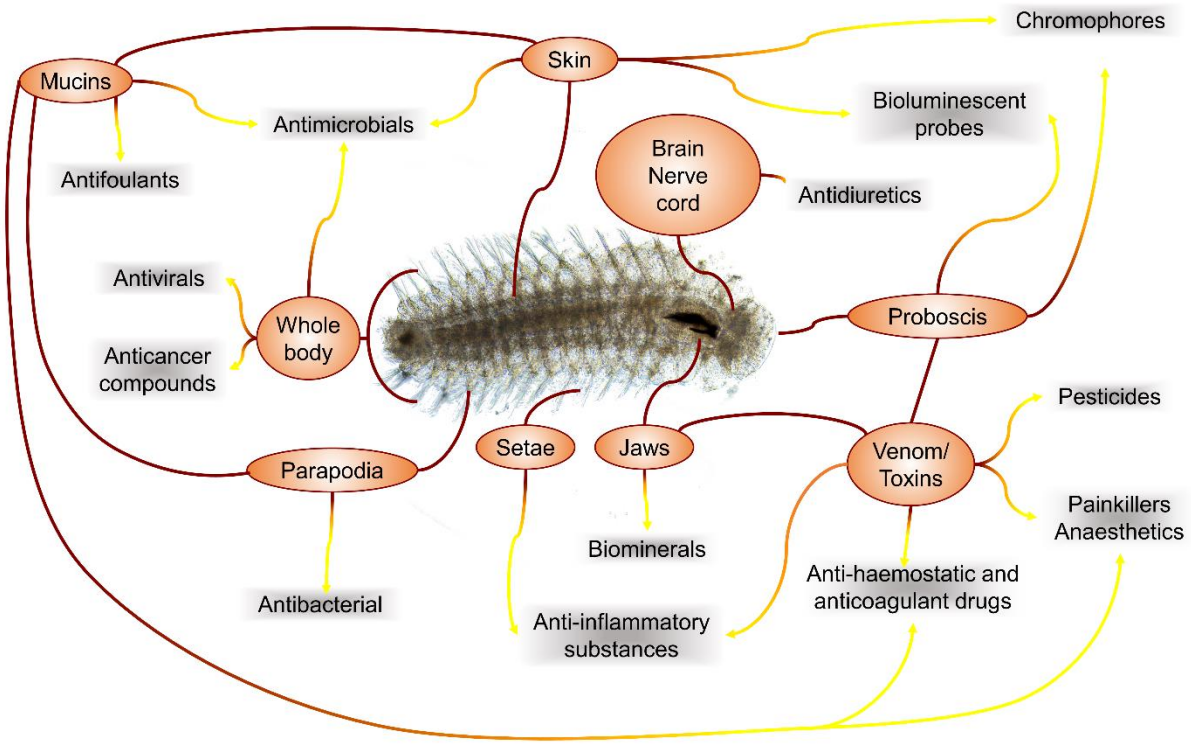


Figure 1.1. Identification of bioproducts that can be retrieved from polychaetes (red circles and arrows) and the respective applications (in grey). The example is illustrated with a Polychaeta belonging to the genus Ophryotrocha.

1.2. Bioproducts from Polychaeta: properties and potential applications

As the wide variety of bioproducts that can be retrieved from marine Polychaeta likely associate with their uncanny diversity, it is seemingly difficult to summarise their biotechnological deterrent by species or even chemical classes. On the other hand, some secretions, especially venoms, poisons and toxungens, are complex cocktails of individual substances that can, on their own, hold different functions (Fig. 1.2).

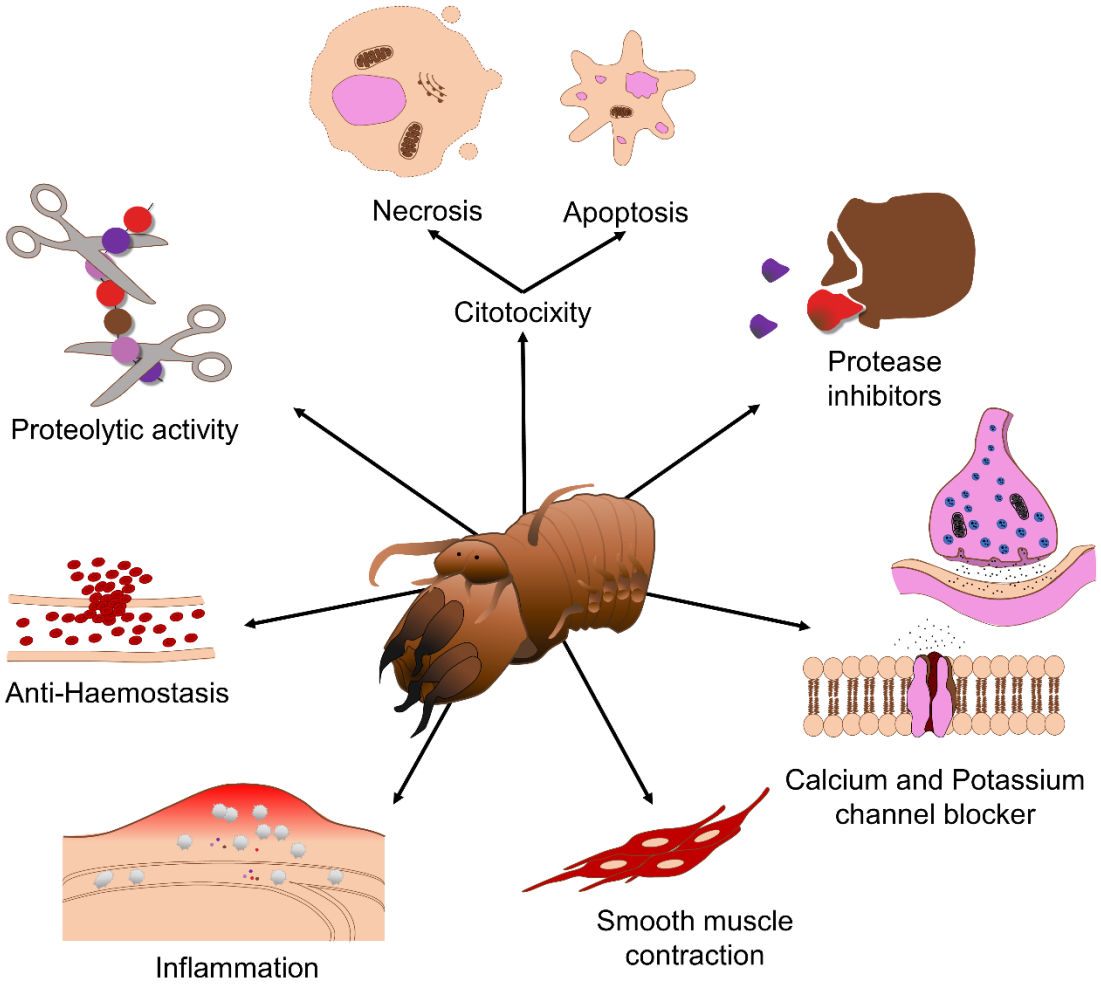


Figure 1.2. Mode-of-action of venoms from Polychaeta, identified from predator-prey interactions

There are various applications for already extracted and some synthesized products from Polychaeta sources (see Table 1.1). These applications are numerous, from pigments and anti-foulants to various type of drugs.

Table 1.1. Summary of studies focusing on the identification, characterisation and function of substances retrieved from polychaetes with biotechnological

Potential application	Species	Substance	Properties	Geographic location	References
Porphyrinoid pigments					
[Photosensitizer]	*Bonellia viridis*	Bonellin	Hematoporphyrin able to arrest embryonic development, provoke dissociation in sponges, lysis of human erythrocytes, cell blockage and flagellar motility.	Mediterranean Sea, Eastern Atlantic Ocean	Pelter et al (1976, 1978), Agius et al. (1979), Giudici (1984)
	Bonellia viridis	Neobonellin	Hematoporphyrin with blocking effect on the development of sea urchin embryos.	Mediterranean Sea, Eastern Atlantic Ocean	Cariello et al (1978)
	Chaetopterus variopedatus	Pigment	Red pigment identified as b-carotene, tough to be the main source of vitamin A.	English Channel	Kennedy & Nicol (1959), Fresneau et al (1984),
	Halla parthenopeia	Hallocrome	Anthraquinone red pigment with redox properties.	Mediterranean Sea	Prota et al (1971)
	Thelepus setosus	Thelepin	Halochromic-like pigment with antimycotic properties.	North Pacific Ocean	Higa & Scheuer (1974,1975)
	Tomopteris helgolandica	Parapodial gland	Activation of nicotinic cholinergic receptors and calcium-dependent intracellular mechanisms.	North Sea	Gouveneaux et al (2018)
Luminescence and fluorescent probes					
	Acholoe astericola	Scales	Purified fluorescent compound is the reaction product of the bioluminescent reaction, with a maximum at 520 nm.	English Channel	Fresneau et al (1984)
	Chaetopterus variopedatus	Photoprotein	Strong red fluorescent pigment	English Channel	Kennedy & Nicol (1959), Shimomura (2006) Deheyn et al (2013), Branchini et al (2014)
	Eusyllis blomstrandi	Segment trunks	Luminescent flashes of light of the posterior body part.	North Atlantic Ocean	Zörner & Fischer (2007)
	Harmothoe lunulata	Polynoidin	Photoprotein capable to detect superoxide radicals produced by activated white blood cells.	English Channel	Nicolas et al (1982), Bassot (1987), Bassot & Nicolas (1995)
	Odontosyllis enopla *Odontosyllis umdecimdonta*	Luciferin-luciferase	Identification and characterisation of luciferin and luciferase.	Sea of Japan	Shimomura et al (1963), Schultz et al (2018)
	Odontosyllis phosphorea	Photoprotein	Present of a photoprotein in the mucus, with green luminescence.	North East Pacific Ocean	Deheyn & Latz (2009)
	Swima bombiviridis	Luminescent "bombs"	Green light bioluminescent bursts.	Northeast and Western Pacific Ocean	Osborn et al (2009)
	Tomopteris helgolandica	Parapodial sites	Light emission trough activation of nicotinic cholinergic receptors and a calcium-dependent intracellular mechanism involving L-type calcium channels.	North Sea	Gouveneaux & Mallefet (2013), Gouveneaux et al (2017)
	Tomopteris sp	Aloe-emodin	Isolation of a fluorescent yellow–orange pigment.	North East Pacific Ocean	Francis et al (2014)
Eco-friendly pesticides and anti-foulants					
[Insecticide]	*Lumbriconereis heteropoda*	Nereistoxin	Neuromuscular block, central nervous system effect - the electrographic seizure discharge and behavioural convulsion.	Synthetic substance	Chiba et al (1967), Sakai (1967), Xie et al (1993)
[Anti-foulant]	*Sabella spallanzanii*	Mucus extract	Natural lysozyme-like activity and growth inhibition of *Vibrio anguillarum*, *Vibrio harveyi*, *Pseudomonas aeruginosa* and *Candida albicans*.	Mediterranean Sea	Stabili et al (2011)

Table 1.1 - Summary of studies focusing on the identification, characterisation and function of substances retrieved from polychaetes with biotechnological potential (continued)

Potential application	Species	Substance	Properties	Geographic location	References
Painkillers, anaesthetics and relaxants					
	Eulalia viridis	Phyllotoxin	Proteins with proteolytic action that able permeabilization of material and softens tissue.	North East Atlantic Ocean	Cuevas et al (2018)
	Eurythoe complanata	Extracts from the body wall	Beta-adrenoceptor-like agonist	Southwest Atlantic Ocean	Suadicani et al. (1993)
	Eurythoe complanata *Hermodice carunculata* *Paramphinome jeffreysii*	Neurotoxins turripeptide-like	Provoke prey paralysis trough neuromuscular transmission and ShKT domain that inhibits potassium channels.	North West Atlantic Coast	Verdes et al (2018)
	Glycera convoluta	Neurotoxins	Able up-regulate the activity of Cav2.2 channels (N-type).	English Channel	Meunier et al (2002)
	Glycera spp.	Neurotoxins	Pore-forming and membrane-disrupting toxins, neuromuscular transmission and inhibition of potassium channels.	English Channel	von Remount et al (2014b)
Anti-haemostatic and anticoagulant drugs					
	Eurythoe complanata *Hermodice carunculata* *Paramphinome jeffreysii*	Protease inhibitors	Several proteases known to be anti-haemostatic.	North West Atlantic Coast	Verdes et al (2018)
	Glycera spp.	Protease inhibitors, proteolytic enzymes, CAP and other proteins	Known to have anticoagulant and haemolytic properties, to provoke inflammation, vasodilation and smooth muscle contraction.	English Channel	Michel & Keil (1975), von Remount et al (2014b)
	Notomastus lobatus	Bromophenols	Used as a defence mechanism, hepatotoxic for fish, provoking hepatocellular coagulation necrosis and fatty change in the liver.	South Carolina	Casillas & Myers (1989), Lincoln et al (2005)
	Spirographis Spallanzani	Mucus extract	Thermolabile biologically active proteins, that produce haemolysis in vertebrate erythrocytes.	Mediterranean Sea	Canicatti et al (1992)
Anti-cancer					
	Nereis virens	Nereis active protease	Inhibition of the proliferation and induction of apoptosis in human lung cancer cells.	Yellow Sea	Tang et al (2017)
	Nicomache minor	Nicomicin	Peptide with selective toxicity towards cancer and normal cell lines.	Artic Sea	Panteleev et al (2018)
Anti-inflammatory substances					
[Anti-inflammatory drug]	*Eurythoe complanata*	Complanine	Acute inflammation inducer.	Oriental China Sea	Nakamura et al (2008, 2009)
Bioproducts from symbiotic bacteria					
[Ferment extract used in cosmetics]	*Alvinella pompejana (*bacteria *Alteromonas macleodii)*	Exopolysaccharides	Biopolymer with an important role in the skin immune barrier.	South Pacific Ocean	Cambon-Bonavita et al (2002), Martins et al (2014), Lelchat et al (2015)
[Ferment extract used in cosmetics]	*Alvinella pompejana (*bacteria *Vibrio diabolicus sp)*	Exopolysaccharides	Unspecified biopolymer.		Raguénès et al (1997)

Table 1.1 - Summary of studies focusing on the identification, characterisation and function of substances retrieved from polychaetes with biotechnological potential (continued).

Potential application	Species	Substance	Properties	Geographic location	References
Antimicrobial properties: biocides and antibiotics					
	Arenicola marina	Arenicin	Active against Gram-positive (*L. monocytogenes*) and Gram-negative bacteria (*E. coli*) and fungi (*C. albicans*).	White Sea	Ovchinnikova et al (2004), Shenkarev et al (2011), Maltseva et al (2014)
	Bonellia viridis	Bonellin	Hematoporphyrin with strong antibiotic and bactericidal activity on marine and terrestrial bacteria.	Mediterranean Sea	Gauthier & Giudici (1983)
	Branchiomma bairdi	Mucus extract	Active lysozyme-like enzymes.	Mediterranean Sea	Giangrande et al (2014)
	Branchiomma luctuosum			Mediterranean Sea	
	Megalomma lanigera			Mediterranean Sea	
	Myxicola infundibulum			North Pacific Ocean	
	Sabella spallanzanii			Mediterranean Sea	
	Sabellastarte spectabilis			Mediterranean Sea	
[Sewage biofilter]	*Branchiomma luctuosum*	Whole-body	Ability to host all major microbiological groups, especially autochthonous bacteria.	Mediterranean Sea	Licciano et al (2007)
	Eulalia viridis	Phyllotoxin	Proteins active against gram-negative marine bacteria.	Northeast Atlantic Ocean	Cuevas et al (2018)
	Halla parthenopeia	Mucus extract	Purple mucus with acute toxicity towards marine bacteria.	Mediterranean Sea	Iori et al (2014)
[Antibiotic drug]	*Nereis diversicolor*	Hedistin	Peptide from coelomocytes revealed activity against a large spectrum of bacteria including *Staphylococcus aureus* and *Vibrio alginolyticus*.	English Channel	Tasiemski et al (2007)
	Perinereis aibuhitensis	Perinerin	Active peptide against gram-positive and gram-negative bacteria and even fungi.	Asia	Pan et al (2004)
	Polyphysh crass	Tribromopyrrole	Active against gram-positive and -negative marine bacteria.	Baltic Sea	Emrich et al (1990)
	Saccoglossus kowalewskii	Bromophenol	Inhibition of aerobic microbial degradation of the burrow-wall mucous lining or alteration of local biogeochemistry.	North West Atlantic Coast	King (1986)
	Sabella spallanzanii	Mucus extract	Natural lysozyme-like activity and growth inhibition of *Vibrio anguillarum*, *Vibrio harveyi*, *Pseudomonas aeruginosa* and *Candida albicans*.	Mediterranean Sea	Stabili et al (2011)
[Antimycotic agents]	*Thelepus setosus*	Thelepin (dienone)	Similarities with the antimicrobial agent griseofulvin (general antifungal activity).	North Pacific Ocean	Higa & Scheuer (1974, 1975)
	Unknown species	Egg masses	Broad-spectrum bactericide.	Australia	Benkendorff et al (2001)
Antivirals					
	Chaetopterus variopedatus	β-galactose-specific lectin (CVL)	Inhibits cytopathic effect induced by HIV-1 and the production of viral p24 antigen.	Sea of Japan	Mikheyskaya et al (1995), Wang et al (2006)
Other potentialities: from detoxifiers to biominerals					
[Anti-pollution agent]	*Amphitrite ornata*	Dehalogenating Peroxidase	Haem-containing peroxidase capable of removing halogens including fluorine from haloaromatics.	Northwest Atlantic Ocean	Chen et al (1996), Roach et al (1997)
[Biomineral]	*Glycera dibranchiata*	Atacamite	Metal-containing biominerals.	Northeast Pacific Ocean	Lichtenegger et al (2002, 2003)
[Antidiuretic drug]	*Nereis diversicolor*	Peptide	Active in osmoregulatory functions.	English Channel	Fewou and Dhainaut-Courtois (1995)
[Detoxifying agent]	*Notomastus lobatus*	Chloroperoxidase	Chlorinating halogen acceptor substrates using chloride as the halogen donor.	Northwest Atlantic Ocean	Chen et al (1991), Roach et al (1997)

1.2.1. Luminescent and fluorescent probes

Despite the value of natural pigments as industrial, foodstuff and histological dyes, there is wider biotechnological potential for luminescent and fluorescent probes, part due to much enhanced signal-to-noise ratios, part to the ability to switch on/off light emission via fine control of some sort of chemical reaction or photoexcitation. Applications of fluorophores are devised not just for imaging but also the development of micro- and nanoscale bioanalytical sensors. Among these, microarrays and next-generation sequencers are included, along with state-of-the-art *in vivo* and *ex vivo* micro-imaging techniques (Roda et al. 2004). Decades ago, the search for natural luminescent and fluorescent probes began with luciferins (luminescent photoproteins) and GFP (green fluorescent protein). These, albeit mostly commercialised as recombinant forms, were retrieved from many aquatic and terrestrial organisms, with the marine cnidarian *Aequorea victoria*, becoming notorious for GFP. Nowadays, the search for novel probes accommodates a large portion of the interest in natural products.

Bioluminescence had long been investigated in marine organisms, among which the dinoflagellate *Noctiluca* and many deep-sea fish are illustrious representatives (see Haddock et al. 2010 for a review). Albeit less mediatic, luminescence in the Polychaeta has also received some attention. Even though its adaptive value remains largely unknown, intraspecific communication and even anti-predatorial defence (as the uncanny case of luminescent "bombs" dropped by *Swima bombviridis*) have been disclosed (Zörner and Fischer 2007; Osborn et al. 2009; Francis et al. 2016; Oba et al. 2017). Many Polychaeta seem to produce bioluminescence by processes somewhat different to the luciferin-luciferase system. In fact, in most cases, instead of a continuous glow, flashes of light are emitted by action of photoproteins. Photoproteins were actually first described from the mucus of a polychaete, *Chaetopterus* (Shimomura 2006; Deheyn et al. 2013; Branchini et al. 2014). Another example for photoprotein-based bioluminescence comes from the syllid polychaete *Odontosyllis phosphorea*, adequately termed fireworm (Deheyn and Latz 2009). Conversely, the luminescence of the congeneric *O. enopla* and *O. umdecimdonta* has been associated to luciferin-luciferase reactions (Shimomura et al. 1963; Schultz et al. 2018). To add further confusion, luminescent secretions can also exhibit fluorescence *per se*, as for the case of scaleworms (Aphroditidae) and *Chaetopterus* (Kennedy and Nicol 1959; Fresneau et al. 1984). Still, in most cases, the molecular structure of these compounds is not known and in many situations the distinction between luciferase- and photoprotein-based bioluminescence in the Polychaeta is not clear. Altogether, luminescence in the Polychaeta seems to be more common than anticipated, inclusively in shallow water species.

Rather comprehensive research has been conducted on polynoidins (a type of photoproteins) in the photosomes (luminescent organs) of scaleworms by Nicolas et al. (1982) and latter by Bassot and Nicolas (1995). These authors revealed that polynoidins are specifically activated by oxygen radicals and can be controlled *ex vivo* through electrical signalling, which should involve Ca^{2+} channels. This led the authors to propose application of these photoproteins in the detection of superoxide radicals in cells, discussing that they could hold more specificity then other photoproteins. Similarly, luminescence dependent of L-type calcium channels (and therefore nicotinic cholinergic receptor-controlled) was found in the deepsea planktonic polychaete *Tomopteris helgolandica*, producing precisely-controled flashes of yellow light, a colour which is not common in luminescent sea animals (Gouveneaux and Mallefet 2013).

1.2.2. Painkillers and anaesthetics

There is a nearly constant demand for safer and non-addictive painkillers and anaesthetics for biomedical use. Asides the case of Prialt mentioned earlier, investigating marine bioproducts for the purpose seems primordial. However, recent findings suggest that the Polychaeta may offer great possibilities in this area, as further neurotoxic properties of toxins and venomous secretions are unveiled. Glycerotoxin (first described from *Glycera* spp.) is a proteinaceous neurotoxin, found to be a Ca^{2+} channel agonist, whose effects were first verified in *Daphnia* (Michel and Keil 1975). Research on *Glycera* was recently complemented by "omics" that allowed a molecular overview of its full venom (Von Reumont et al., 2014b). In fact, "venomics" in *Glycera* and other few Polychaeta, namely fireworms (Amphinomidae), disclosed several neurotoxins similar to beta-adrenoreceptors-like, turripeptides and neurotoxins holding a ShKT domain (a general term for domains similar to that of a peptidic toxin from the anemone *Stichodactyla helianthus*). These neurotoxins are able to induce muscle relaxation and paralysis trough the blocking of neuromuscular transmission and K^{2+} channels (Suadicani et al., 1993; Von Reumont et al., 2014b; Verdes et al., 2018). The fact that the effects of glycerotoxin appear to be fully reversible (Meunier et al. 2002) gives it added value for biotechnology. Besides these examples, evidence for neurotoxins has been found in the poisonous mucus of *Halla parthenopeia* (Oenonidae), which also show various toxic effects against other polychaetes, rotifers, crustaceans, molluscs and even bacteria, even though the mechanisms are not fully understood in any case (Iori et al. 2014). All in all, much work needs to be done to isolate and characterise novel painkillers and anaesthetics from Polychaeta. However, as the current state-of-the art reveals a range of substances that allows predator species to immobilize their prey and enjoy their often-oversized meal, it is clear that these organisms can, in the future, offer a myriad of compounds of interest.

1.2.3. Anti-cancer therapy

As cancer is acknowledged as one of the epidemics of our century, the search of novel therapeutics, especially those less aggressive and more effective than traditional methods such as radiation therapy, is a constant (see Ruiz-Torres et al. 2017). Marine bioproducts have not been neglected for the purpose, and, in fact, much, if not one of the biggest efforts in the field has been dedicated to find new anti-cancer drugs (Mayer et al. 2010; Giordano et al. 2018). Despite the infancy in this area of marine biotechnology, there is evidence for the potential of marine bioproducts as potential anti-cancer drugs. In a recent example, Asselman et al. (2019), found that natural sea spray aerosol, which is mostly comprised of salts, also contains bio-reactive compounds, especially algal toxins, that can interact with the mechanistic Target of Rapamycin (mTOR) in human lung cancer cells, de-regulating cell cycle and promoting apoptosis. Recent research on the Polychaeta has been providing hints on this matter as well, such as proteases from *Nereis* capable of inducing apoptosis in human lung cancer cells (Tang et al. 2017). Also, the recently-discovered peptide, nicomicin, retrieved from the small artic worm *Nicomache minor*, showed differential toxicity towards cancer and normal cell lines, potentially yielding stronger effects onto the former, depending on dose (Panteleev et al. 2018). In another example, Kuzmin et al. (2018) tested anti-microbial peptides from several marine invertebrates, including the toxin arenicin from the Polychaeta *Arenicola marina*, which exhibited varying pro-apoptotic effects on various human cancer cer lines, however having haemolytic properties as well. Notheless, the exact mechanisms underneath cytotoxicity are not yet understood in any of these cases. Altogether, the available information indicates that the variety of marine products that can have differential effects on normal and neoplastic cells is just beginning to be unravelled, which seemingly applies to Polychaeta as well. New developments on *in silico* methods for predicting the targets of marine bioproducts with anti-cancer properties (Chen et al. 2017) reflect the acknowledged potential of marine life as source of novel anti-tumour drugs and can point the path for more efficient bioprospecting.

1.2.4. Anti-haemostatic and anticoagulant drugs

Natural substances that can be anti-haemorrhagic (i.e. promote haemostasis) can, conversely, prevent coagulation, which holds high relevance for biomedical applications as anti-clotting agents and related. In fact, one of the best-known pharmaceutical drugs is the anticoagulant hirudin retrieved from the leech *Hirudo* spp. (see Basanova et al. 2002 for a review). Animal venoms can have several substances whose specific function is to prevent clotting (therefore being anti-haemostatic) and healing, for better feeding efficiency, such as the serine protease inhibitors (serpins) found in the venom of some ticks and snakes (Serrano and Maroun 2005; Maritz-Olivier et al. 2007). Similarly, serine proteases retrieved from polychaetes were found to have anti-

coagulant activity (Kim et al. 2018) *in vitro* and *in vivo* in mammals (Yeon et al. 2017). In Polychaeta with already characterised venoms, these haemostatic disruptive proteases were also present, such as in bloodworms and fireworms (Von Reumont et al., 2014b; Verdes et al., 2018), which, once more, indicates strong evolutionary convergence of venoms and poisons. In their turn, mucus secretions of sabellids have been found to contain thermolabile haemolysins (Canicatti et al. 1992). Expectedly, this is an area that will mostly likely experience a boom in the next few years.

1.2.5. Anti-inflammatory substances

Several marine bioproducts with pharmacological interest as anti-inflammatory drugs have already been identified, with emphasis on those collected from a few soft corals and sponges (see, for instance Mayer et al. 2013). Also, the mucus of some echinoderms has glycoproteins that are able to block the adhesion of mammalian leukocytes (Bavington et al. 2004). Although very distant from the Polychaeta, this example shows the potential of invertebrate mucins as source of bioproducts that can interfere with the immune system. Even though no such substance from the Polychaeta is in any phase of testing or approval for commercialisation, there are a few promising case-studies, even though the direct effects of secretions of Polychaeta onto defence cells remain unknown. Curiously enough, toxins with pro-inflammatory properties (related to complanines) have been described in fireworms, presumably by interfering with TNFα-mediated pathways (Nakamura et al. 2008, 2009).

1.2.6. Antimicrobial properties: biocides and antibiotics

As major health issues resulting from the surge of multi-resistant pathogens keep making the headlines, the search for novel bactericides and antibiotics in general is given a new momentum. Part due to the characteristics of their habitats, such as sediments and rocky substrates covered with bacterial biofilms, for instance, part due to their relatively simple immune system, the Polychaeta must be very effectively able to defend against microbial pathogens. Naturally, the simplest, yet effective, adaptation for the purpose is the development of chemical warfare. Interestingly, some authors analysed the egg masses of some polychaetes and found that they possess substances that inhibit growth of a broad-spectrum of human pathogenic bacteria (Benkendorff et al. 2001). In fact, secretions such as crude mucus are known to hold anti-bacterial and anti-fungal properties, even if, in most cases, the exact substances have not been identified (Stabili et al. 2011; Iori et al. 2014). Still, it was suggested that the lysosomes in mucus secretions of sabellids are major players in the defence against pathogenic bacteria (Giangrande et al. 2014).

Proteinaceous compounds seem, so far, to form the bulk of compounds from Polychaeta revealing interesting antimicrobial properties. It is the case of a novel natural product from *Polyphysia crassa*, which has activity against both gram-positive and -negative marine bacteria (Emrich et al. 1990). Similar results were found for the peptide retrieved from *Arenicola marina*, arenicin (Ovchinnikova et al. 2004; Shenkarev et al. 2011; Maltseva et al. 2014). Nonetheless, arenicin has also been found to interfere negatively with the human complement system at high doses (Umnyakova et al. 2018). To the list is added perinerin, harvested from the clamworm *Perinereis aibuhitensis*, which, besides activity against both gram-positive and -negative, was found to have anti-fungal properties as well (Pan et al. 2004) (see Table 1.1 for further details). Another proteinaceous toxin, hedistin is another novel antimicrobial peptide, being extracted from *Hediste diversicolor* (Tasiemski et al. 2007). In most cases, the mode-of-action of these substances is not known. There are, nonetheless, efforts to provide a more comprehensive characterisation of these substances and their effects using the screening power of RNAseq. It is the example of potential pore-forming proteins and peptides from the venom of *Glycera* (Von Reumont et al., 2014b).

There are also non-proteinaceous anti-microbials from the Polychaeta. Some of the most interesting examples come from photosensitisers described above. It is the case of the effects of photoactivated chlorins and hematoporphyrins from *Bonellia viridis* onto gram-positive bacteria (Gauthier and Giudici 1983). Also, a brominated metabolite, named thelepin (produced by the Terebellidae family), was found to have similar characteristics of antimycotics from *Penicillium* (Higa and Scheuer 1974).

It must also be noted that the Polychaeta may have adaptations to cope with microorganisms that reach beyond anti-microbial substances. It is the case of *Branchiomma luctuosum*, a polychaete found in very polluted locations and known to form associations with all sorts of microorganisms, with emphasis on autochthonous bacteria, presumably as a defence against environmental chemicals, to which is added its ability to digest coliforms (Licciano et al. 2007). This unveils the potential of this worm as biofilter, providing yet an innovative approach to the possible biotechnological purpose for the Polychaeta.

1.2.7. Antivirals

So far, most animal bioproducts with antiviral properties were found in venoms (see da Mata et al., 2017, for more details). In most cases, these bioproducts were found to compete with viral particles for receptors, thus inhibiting infection (Mata et al. 2017). Antiviral properties have been widely neglected in the Polychaeta but the rising number of venomous species being discovered is likely to produce interesting outcomes in the near future. Still, lectins from *Chaetopterus variopedatus*,

which initially revealed reactivity toward ß-galactose, and agglutination of all type of human erythrocytes (Mikheyskaya et al. 1995), were also found to possess anti-HIV-1 activity, precisely by hindering the entry of viral particles into host cells (Wang et al. 2006).

1.3. The ecological value of Polychaeta

Polychaeta is a vast, seemingly polyphyletic class that encompasses species, from benthic macrofauna from the intertidal zone down to abyssal plains and deep-sea hydrothermal vents and are fundamental links in the food chain. They are in fact the most frequent and abundant marine metazoans in benthic groups (Fauchauld & Jumars 1979). Such habitat diversity is paramount for answering questions about their wide morphological and adaptive diversity, as well as clarifying evolutionary trends and their complex phylogenetic relationships that, in many cases, remain obscure and subject of constant systematic redefinition (Bartolomaeus et al. 2005; Struck et al. 2011). As long as there are large knowledge gaps on something as basal as morphology, polychaete evolution can hardly be unravelled (Bartolomaeus et al. 2005). This is the reason why the division Errantia and Sedentaria polychaetes is still in use despite its null taxonomic value. The Errantia consists of mostly vagile and free living polychaetes, while Sedentaria in semi-sessile, sessile and mostly tubicolous polychaetes (see Bartolomaeus et al. 2005, for a review). Despite the same conserved body plan, the vast polychaete biodiversity is noticeable by the diversity of sensory structures and sense organs in polychaetes (Purschke 2005), the structure of the epidermis, for instance, which precludes a wide range of supportive cells, gland cells and sensory cells (e.g. Hausen 2005a) and specially by the variety of epidermal extracellular structures, named chaetae, that shows patters specific to each species (Hausen 2005b). Indeed, annelid phylogeny is one of the largest unresolved problems within the Metazoa. Adding to the taxon's ancient radiation and diversity, discussion persists on its positioning relatively to Arthropoda, which are traditionally hypothesized to be the sister taxon to the Annelida, mostly due to being segmented (see Bartolomaeus et al. 2005).

Polychaeta class has an ancestral body plan, with a biphasic life cycle: the acoelomate/pseudocoelomate larva and the coelomate adult (reviewed by Rieger & Purschke 2005). The adult polychaete possesses chemical signals and chemoreception that mediate ecologically important behaviours, including reproduction, larval settlement, metamorphosis, recruitment and feeding (Lindsay 2009). Reproduction is also controlled by environmental factors, while spawning and mating behaviours in some species are coordinated through various chemical signals including pheromones (see the review of Lindsay 2009, for further details). Polychaetes show great diversity in fertilization mechanisms, but are fundamentally gonochoristic, being only a small number of

species hermaphrodite (Giangrande et al 1997). While some species lack a defined ovary, most have ovaries, even if location varies. The oogenesis can be extraovarian and accompanied by somatic cells (previtellogenic oocytes are released into the coelom) or intraovarian, where the oocytes undergo vitellogenesis within the ovary (Dorresteijn 2005; Eckelbarger 2005). In some groups, as the case of Phyllodocidae, the oocytes can even sprout from still unknown tissues and undergo previtellogenic and vitellogenic development in a free-floating condition within the coelomic fluid (Dorresteijn 2005). Naturally, mature ova display a wide range of egg envelope morphologies that often show some intrafamilial similarities (Eckelbarger 2005). The sperm has also great variations within form, function and maturation location, and has even proven to be of great assistant in polychaete phylogeny (Rouse 2005).

Another key aspect of Polychaeta ecology is the feeding biology. The feeding system can be grossly divided into two modes (macrophagy and microphagy), five submodes (herbivores, carnivores, filter-feeders, surface deposit-feeders, and burrowers) and a dozen of morphological subgroups (Fauchauld & Jumars 1979). The alimentary canal follows the feeding system variability, with a wide variety of food sources, structures involved in feeding and numerous specializations (Tzetlin & Purschke 2005). Besides the diversity of feeding type, most polychaetes feed on detritus, dead or moribund preys, thus giving an important contribute to the ecosystem (Checon et al 2016). They are also key food supplies for a variety of organisms, like fish and other invertebrates (Nelson & Capone 19990). Polychaeta are so diverse and abundant that being referred as a whole englobes always a variety of factors.

1.4. The *Eulalia viridis* case study

The marine annelid *Eulalia viridis* (Polychaeta: Errantia, Phyllodocidae) inhabits rocky intertidal shores, especially mussel beds (Fig. 1.3), seeking protection during the high tide underneath pebbles and clumps of mussels (Emson 1977). This species is reported to have a wide distribution, from the north Atlantic to the Mediterranean Sea and all the American Atlantic coast until central America, to which are added reports from the Alaska. This wide distribution is indicative of possible overlapping of species or subspecies. In fact, the species reported as *Eulalia viridis* in the north of France was found to be most likely *Eulalia clavigera* (Bonse et al. 1996). Like this study, there are others that indicate that the genus *Eulalia* is still poorly studied and identification down to the species level can be dubious. Whether the species present in the coast of Portugal is *E. viridis*, *E. clavigera* or a complex of both remains uncertain and thus, the name of the species in the following sections will be retained as *E. viridis*, as traditionally accepted, for the sake of simplicity. The breading season reported for this species, when collected in the United Kingdom, is from July to August. The species is gonochoristic and there are no differentiated gonads. Instead, gametocytes

derive from undifferentiated roaming coelomic cells (Olive 1975b, a). *Eulalia viridis* is believed to be an opportunist forager, since scavenger, detritivore and predator behaviours have been reported (Emson 1977; Fauchald and Jumars 1979; Morton 2011). It is acknowledged that *E. viridis* feeds mostly on barnacles, mussels and other worms, even from the same species (Emson 1977; Morton 2011). The worm, as other phyllodocids, makes use of its proboscis for collecting food items through suction, relying on strong axial musculature, instead of jaws, a typical characteristic of this family (Michel 1964; Tzetlin and Purschke 2005).

Figure 1.3. *Eulalia viridis* in natural habitat, surrounded by barnacles and limpets. The worm roams between puddles of rocky beaches at low tide. Photo obtained at Parede beach (38°41'42" N; 09°21'36"W).

1.5. Scope of the thesis

The main objective of this thesis is to contribute to the exploitation of marine invertebrate's potential as source of novel bioactives for biotechnological applications. To accomplish that, we will study a somewhat neglected but highly promising polychaete, *Eulalia viridis*. In a country like Portugal, where the exploitation of the seas is historically and culturally linked to fisheries, marine bioprospecting can offer fresh perspectives towards the sustainable exploitation of novel high-value bioresources, therefore in line with the European Union's ambitions of leading the "Blue Growth Revolution". The thesis is therefore designed as a tiered approach to comprehensively characterise the toxin-bearing secretions of *Eulalia*, from their natural origin and function to toxicity and their potential application in biomedicine. As such, the detailed objectives of the thesis are described as:

- Identify toxin-secreting organs and cells in the worm and how target substances are secreted and delivered.
- Unravel the ecological role of toxins secreted by *E. viridis*, as an adaptative feature for *Eulalia*'s natural habitat and predatory behaviour.
- Disclose the chief composition of *Eulalia*'s venomous cocktail, as well as the identification

of key components of the toxin mixture that hold biotechnological potential.

- Investigate the toxins' main effects under an ecologically-relevant perspective as a means to understand its adaptive value
- Assess the effects and mechanistics of toxins *in vitro* on human cell lines as a mean to explore potential biomedical applications, with emphasis on anti-cancer therapeutics.
- Evaluate the fluorescent properties of *Eulalia*'s secretions and their biotechnological potential as natural molecular markers.
- Contribute to develop a methodological pipeline for toxin identification and characterisation of mode-of-action based on transcriptomics and its cross-validation with proteomics.

The steps taken to achieve the proposed goals, and their association to methods and procedures are summarised in Fig.1.4.

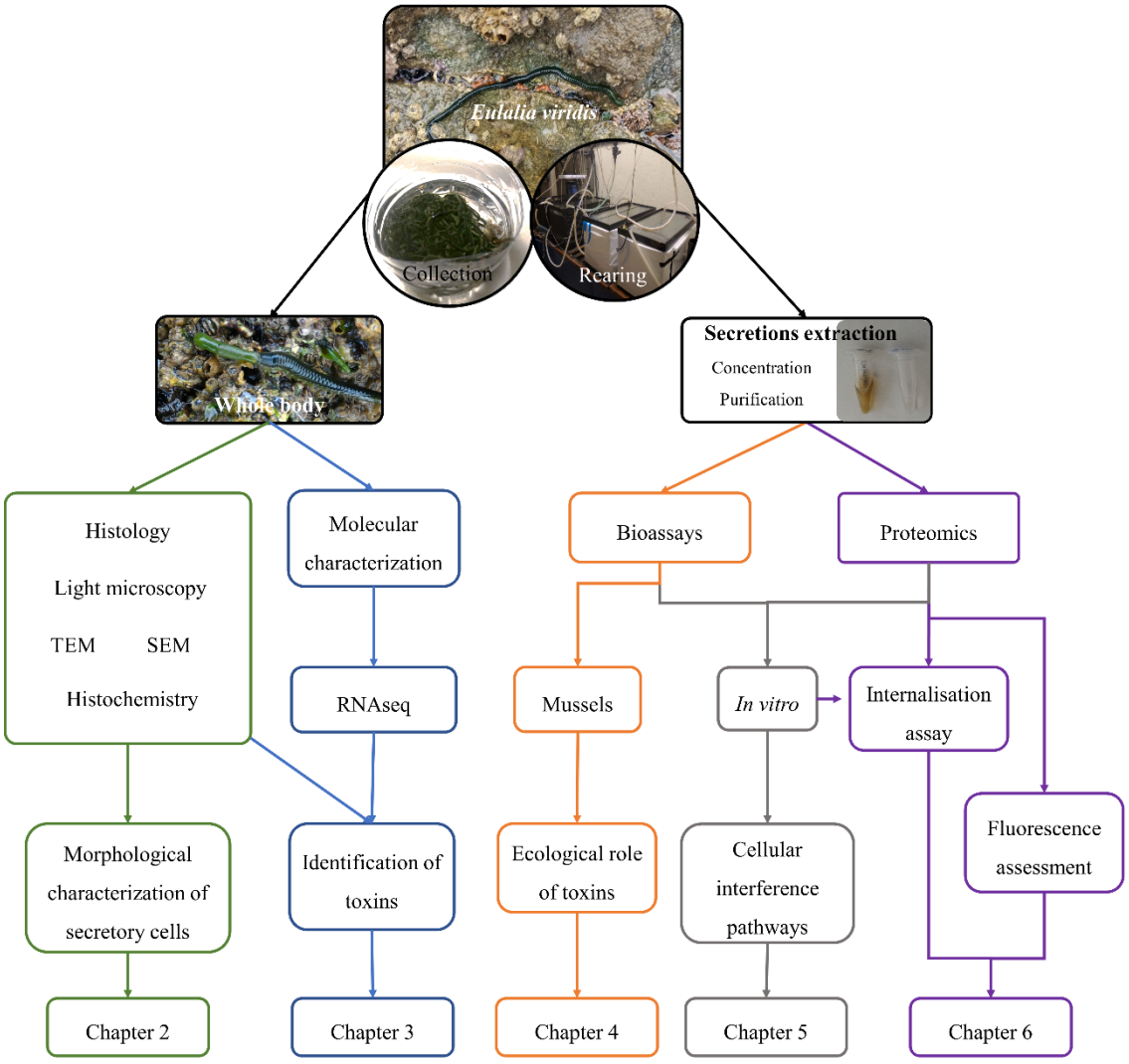


Figure 1.4. Flowchart describing the fundamental organization of the thesis.

1.6. References

Asselman J, Van Acker E, De Rijcke M, Tilleman L, Van Nieuwerburgh F, Mees J, De Schamphelaere KAC, Janssen CR (2019) Marine biogenics in sea spray aerosols interact with the mTOR signaling pathway. Sci Rep 9:1–10.

Bartolomaeus T, Purschke G, Hausen H (2005) Polychaete phylogeny based on morphological data – a comparison of current attempts. In: Bartolomaeus T, Purschke G (eds) Morphology, Molecules, Evolution and Phylogeny in Polychaeta and Related Taxa. Springer, pp 341–356

Basanova A V., Baskova IP, Zavalova LL (2002) Vascular-platelet and plasma hemostasis regulators from bloodsucking animals. Biochemistry (Mosc) 67:143–150.

Bassot J-M, Nicolas M-T (1995) Bioluminescence in scale-worm photosomes: the photoprotein polynoidin is specific for the detection of superoxide radicals. Histochem Cell Biol 104:199–210.

Bavington CD, Lever R, Mulloy B, Grundy MM, Page CP, Richardson N V., McKenzie JD (2004) Anti-adhesive glycoproteins in echinoderm mucus secretions. Comp Biochem Physiol - B 139:607–617.

Benkendorff K, Davis AR, Bremner JB (2001) 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.

Bonse S, Schmidt H, Eibye-jacobsen D, Westheide W (1996) *Eulalia viridis* (Polychaeta : Phyllodocidae) is a complex of two species in northern Europe : Results from biochemical and morphological analyses. 33–48.

Branchini BR, Behney CE, Southworth TL, Rawat R, Deheyn DD (2014) Chemical analysis of the luminous slime secreted by the marine worm Chaetopterus (Annelida, Polychaeta). Photochem Photobiol 90:247–251.

Bruhn H, Winkelmann J, Andersen C, Andrä J, Leippe M (2006) Dissection of the mechanisms of cytolytic and antibacterial activity of lysenin, a defence protein of the annelid *Eisenia fetida*. Dev Comp Immunol 30:597–606.

Calvete JJ, Sanz L, Angulo Y, Lomonte B, Gutiérrez JM (2009) Venoms, venomics, antivenomics. FEBS Lett 583:1736–1743.

Canicatti C, Ville P, Pagliara P, Roch P (1992) Hemolysins from the mucus of Spirographis spallanzani (Polychaeta: Sabellidae). Mar Biol 114:453–458.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: The evolutionary novelty of venoms. Trends Ecol Evol 28:219–229.

Checon HH, Pardo EV, Amaral ACZ (2016) Breadth and composition of polychaete diets and the importance of diatoms to species and trophic guilds. Helgol Mar Res 70:19.

Chen F, Wang Z, Wang C, Xu Q, Liang J, Xu X, Yang J, Wang C, Jiang T, Yu R (2017) Application of reverse docking for target prediction of marine compounds with anti-tumor activity. J Mol Graph Model 77:372–377.

Deheyn DD, Latz MI (2009) Internal and secreted bioluminescence of the marine polychaete *Odontosyllis phosphorea* (Syllidae). Invertebr Biol 128:31–45.

Deheyn DD, Enzor LA, Dubowitz A, Urbach JS, Blair D (2013) Optical and physicochemical characterization of the luminous mucous secreted by the marine worm *Chaetopterus sp.* Physiol Biochem Zool 86:702–705.

Dorresteijn A (2005) Cell lineage and gene expression in the development of polychaetes. Hydrobiologia 535/536: 1-22

Eckelbarger KJ (2005) Oogenesis and oocytes. Hydrobiologia 535/536: 179–198

Emrich R, Weyland H, Weber K (1990) 2,3,4-Tribromopyrrole from the marine polychaete *Polyphysia crassa*. J Nat Prod 5:703–705.

Emson RH (1977) The feeding and consequent role of *Eulalia viridis* (O. F. Muller) (Polychaeta) in intertidal communities. J Mar Biol Assoc United Kingdom 57:93–96.

Escoubas P, Quinton L, Nicholson GM (2008) Venomics: unravelling the complexity of animal venoms with mass spectrometry. J Mass Spectrom 43:279–295.

Fauchald K, Jumars P (1979) The diet of worms : A study of polychaete feeding guilds. Oceanogr Mar Biol - An Annu Rev 17:193–284.

Francis WR, Powers ML, Haddock SHD (2016) Bioluminescence spectra from three deep-sea polychaete worms. Mar Biol 163:1–7.

Freitas AC, Rodrigues D, Rocha-Santos TAP, Gomes AMP, Duarte AC (2012) Marine biotechnology advances towards applications in new functional foods. Biotechnol Adv 30:1506–1515.

Fresneau C, Arrio B, Lecuyear B, Dupaixn A, Lescure N, Vvolfin P (1984) The fluorescent product of scaleworm bioluminescent reaction: an *in vitro* study. Photochem Photobiol 39:255–261.

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RCR (2009) The toxicogenomic multiverse: Convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet 10:483–511.

Gauthier MJ, Giudici MDN (1983) Antibiotic activity of bonellin and hematoporphyrin on marine and terrestrial bacteria. Curr Microbiol 8:195–199.

Giangrande A, Geraci IS, Belmonte G (1994) Life-cycle and life-history diversity in marine invertebrates dnd the implications in community dynamics. Oceanogr Mar Biol 32. 305-333.

Giangrande A, Licciano M, Schirosi R, Musco L, Stabili L (2014) Chemical and structural defensive external strategies in six sabellid worms (Annelida). Mar Ecol 35:36–45.

Giordano D, Costantini M, Coppola D, Lauritano C, Núñez Pons L, Ruocco N, di Prisco G, Ianora A, Verde C (2018) Biotechnological applications of bioactive peptides from marine sources. Adv Microb Physiol 73:171–220.

Gouveneaux A, Mallefet J (2013) Physiological control of bioluminescence in a deep-sea planktonic worm, *Tomopteris helgolandica*. J Exp Biol 216:4285–4289.

Haddock SHD, Moline MA, Case JF (2010) Bioluminescence in the Sea. Ann Rev Mar Sci 2:443–493.

Hausen H (2005a) Chaetae and chaetogenesis in polychaetes (Annelida). Hydrobiologia 535/536: 37-52

Hausen H (2005b) Comparative structure of the epidermis in polychaetes (Annelida). Hydrobiologia 535/536: 25–35.

Higa T, Scheuer PJ (1974) Thelepin, a new metabolite from the marine annelid *Thelepus setosus*. J Am Chem Soc 96:2246–2248.

Hornbeak KB, Auerbach PS (2017) Marine Envenomation. Emerg Med Clin North Am 35:321–337.

Iori D, Forti L, Massamba-N'Siala G, Prevedelli D, Simonini R (2014) Toxicity of the purple mucus of the polychaete *Halla parthenopeia* (Oenonidae) revealed by a battery of ecotoxicological bioassays. Sci Mar 78:589–595.

Karasudani I, Koyama T, Nakandakari S, Aniya Y (1996) Purification of anticoagulant factor from the spine venom of the crown-of-thorns starfish, *Acanthaster planci*. Toxicon 34:871–879.

Kastin AJ (2006) Handbook of biologically active peptides. Elsevier Inc.

Kennedy G, Nicol J (1959) Pigments of *Chaetopterus variopedatus* (Polychaeta). Proc R Soc London B 150:509–538.

Kim HJ, Shim KH, Yeon SJ, Shin HS (2018) A novel thrombolytic and anticoagulant serine protease from Polychaeta, *Diopatra sugokai*. J Microbiol Biotechnol 28:275–283.

Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. Nat Rev Drug

Discov 4:206–220.

Kuzmin D V., Emelianova AA, Kalashnikova MB, Panteleev P V., Balandin S V., Serebrovskaya EO, Belogurova-Ovchinnikova OY, Ovchinnikova T V. (2018) Comparative in vitro study on cytotoxicity of recombinant β-hairpin peptides. Chem Biol Drug Des 91:294–303.

Licciano M, Stabili L, Giangrande A, Cavallo RA (2007) Bacterial accumulation by *Branchiomma luctuosum* (Annelida: Polychaeta): A tool for biomonitoring marine systems and restoring polluted waters. Mar Environ Res 63:291–302.

Lillebø AI, Pita C, Garcia Rodrigues J, Ramos S, Villasante S (2017) How can marine ecosystem services support the Blue Growth agenda? Mar Policy 81:132–142.

Lindsay SM (2009) Ecology and biology of chemoreception in polychaetes. Zoosymposia 2:339-367

Maltseva AL, Kotenko ON, Kokryakov VN, Starunov V V., Krasnodembskaya AD (2014) Expression pattern of arenicins-the antimicrobial peptides of polychaete *Arenicola marina*. Front Physiol 5:1–12.

Maritz-Olivier C, Stutzer C, Jongejan F, Neitz AWH, Gaspar ARM (2007) Tick anti-hemostatics: targets for future vaccines and therapeutics. Trends Parasitol 23:397–407.

Martins A, Vieira H, Gaspar H, Santos S (2014) Marketed marine natural products in the pharmaceutical and cosmeceutical industries: Tips for success. Mar Drugs 12:1066–1101.

Martins C, Rodrigo AP, Cabrita L, Henriques P, Parola AJ, Costa PM (2019) The complexity of porphyrin-like pigments in a marine annelid sheds new light on haem metabolism in aquatic invertebrates. Sci Rep 9:1–11.

Mata ÉCG da, Mourão CBF, Rangel M, Schwartz EF (2017) Antiviral activity of animal venom peptides and related compounds. J Venom Anim Toxins Incl Trop Dis 23:1–12.

Mayer AMS, Glaser KB, Cuevas C, Jacobs RS, Kem W, Little RD, McIntosh JM, Newman DJ, Potts BC, Shuster DE (2010) The odyssey of marine pharmaceuticals: a current pipeline perspective. Trends Pharmacol Sci 31:255–265.

Mayer AMS, Rodríguez AD, Taglialatela-Scafati O, Fusetani N (2013) Marine pharmacology in 2009-2011: Marine compounds with antibacterial, antidiabetic, antifungal, anti-inflammatory, antiprotozoal, antituberculosis, and antiviral activities; affecting the immune and nervous systems, and other miscellaneous mechanisms of . Mar Drugs 11:2510–2573.

Meunier FA, Feng ZP, Molgó J, Zamponi GW, Schiavo G (2002) Glycerotoxin from *Glycera convoluta* stimulates neurosecretion by up-regulating N-type Ca2+ channel activity. EMBO J 21:6733–6743.

Michel C (1964) Histologie, histochimie et innervation de la trompe d'*Eulalia viridis* (Muller), (Polychètes Errantes Phyllodocidae). Bull Lab Mar Dinard 49–50:62–95.

Michel C, Keil B (1975) Biologically active proteins in the venomous glands of the polychaetous annelid, *Glycera convoluta* keferstein. Comp Biochem Physiol 50B:29–33.

Mikheyskaya L V., Evtushenko E V., Ovodova RG, Belogortseva NI, Ovodov YS (1995) Isolation and characterization of a new β-galactose-specific lectin from the sea worm *Chaetopterus variopedatus*. Carbohydr Res 275:193–200.

Milochau A, Lass M, Valembois P (1997) Purification, characterization and activities of two hemolytic and antibacterial proteins from coelomic fluid of the annelid *Eisenia fetida andrei*. Biochim Biophys Acta 1337:123–132.

Molinski TF, Dalisay DS, Lievens SL, Saludes JP (2009) Drug development from marine natural products. Nat Rev Drug Discov 8:69–85.

Morton B (2011) Predator-prey-scavenging interactions between *Nucella lapillus*, *Carcinus maenas* and *Eulalia viridis* all exploiting mytilus galloprovincialis on a rocky shore recovering from tributyl-tin (TBT) pollution. J Nat Hist 45:2397–2417.

Nakamura K, Tachikawa Y, Uemura D, Kitamura M, Ohno O, Suganuma M, Uemura D (2008) Complanine, an inflammation-inducing substance isolated from the marine fireworm *Eurythoe complanata*. Org Biomol Chem 6:2058–2060.

Nakamura K, Tachikawa Y, Uemura D (2009) (-)-Complanine, an inflammatory substance of marine fireworm: a synthetic study. Beilstein J Org Chem 5:4–7.

Nelsen DR, Nisani Z, Cooper AM, Fox GA, Gren ECK, Corbit AG, Hayes WK (2014) Poisons, toxungens, and venoms: Redefining and classifying toxic biological secretions and the organisms that employ them. Biol Rev 89:450–465.

Nelson WG, Capone MA (1990) Experimental studies of predation on polychaetes associated with seagrass beds. Estuaries 13(1): 51-58.

Nicolas M -T, Bassot J -M, Shimomura O (1982) Polynoidin: a membrane photoprotein isolated from the bioluminescent system of scale-worms. Photochem Photobiol 35:201–207.

Oba Y, Stevani C V., Oliveira AG, Tsarkova AS, Chepurnykh T V., Yampolsky I V. (2017) Selected least studied but not forgotten bioluminescent systems. Photochem Photobiol 93:405–415.

Olive PJW (1975a) Germinal epithelium activity in polychaetes in relation to periodicity of reproduction. Pubbl della Stn Zool di Napoli 39:267–281.

Olive PJW (1975b) Reproductive biology of *Eulalia viridis* (Müller) (Polychaeta: Phyllodocidae) In The North Eastern U.K. J Mar Biol Assoc United Kingdom 55:313–326.

Olivera BM (2000) ω-Conotoxin MVIIA : from marine snail venom to analgesic drug. In: Drugs from the Sea. pp 74–85

Osborn KJ, Haddock SHD, Pleijel F, Madin LP, Rouse GW (2009) Deep-sea, swimming worms with luminescent "bombs." Science (80-) 325:964.

Ovchinnikova T V., Aleshina GM, Balandin S V., Krasnosdembskaya AD, Markelov ML, Frolova EI, Leonova YF, Tagaev AA, Krasnodembsky EG, Kokryakov VN (2004) Purification and primary structure of two isoforms of arenicin, a novel antimicrobial peptide from marine Polychaeta *Arenicola marina*. FEBS Lett 577:209–214.

Pan W, Liu X, Ge F, Han J, Zheng T (2004) Perinerin, a novel antimicrobial peptide purified from the Clamworm *Perinereis aibuhitensis* grube and its partial characterization. J Biochem 135:297–304.

Panteleev P V, Tsarev A V, Bolosov IA, Paramonov AS, Marggraf MB, Sychev S V, Shenkarev ZO, Ovchinnikova T V (2018) Novel antimicrobial peptides from the arctic Polychaeta *Nicomache minor* provide new molecular insight into biological role of the BRICHOS domain. Mar Drugs 16:401.

Proksch P, Edrada RA, Ebel R (2002) Drugs from the seas - Current status and microbiological implications. Appl Microbiol Biotechnol 59:125–134.

Purschke G (2005) Sense organs in polychaetes (Annelida). Hydrobiologia 535/536:53–78.

Rao T, Tan Z, Peng J, Guo Y, Chen Y, Zhou H, Ouyang D (2019) The pharmacogenetics of natural products: A pharmacokinetic and pharmacodynamic perspective. Pharmacol Res 146:104283.

Rieger RM, Purschke G (2005) The coelom and the origin of the annelid body plan. Hydrobiologia 535/536:127:137.

Roda A, Pasini P, Mirasoli M, Michelini E, Guardigli M (2004) Biotechnological applications of bioluminescence and chemiluminescence. Trends Biotechnol 22:295–303.

Rouse GW (2005) Annelid sperm and fertilization biology. Hydrobiologia 535/536: 167–178

Ruiz-Torres V, Encinar JA, Herranz-López M, Pérez-Sánchez A, Galiano V, Barrajón-Catalán E, Micol V (2017) An updated review on marine anticancer compounds: The use of virtual screening for the discovery of small-molecule cancer drugs. Molecules 22:1037.

Safavi-Hemami H, Brogan SE, Olivera BM (2018) Pain therapeutics from cone snail venoms: From Ziconotide to novel non-opioid pathways. J Proteomics 0–1.

Schlesinger A, Kramarsky-Winter E, Loya Y (2009) Active nematocyst isolation via nudibranchs. Mar Biotechnol 11:441–444.

Schultz DT, Kotlobay AA, Ziganshin R, Bannikov A, Markina NM, Chepurnyh T V, Shakhova ES, Palkina K, Haddock SHD, Yampolsky I V, Oba Y (2018) Luciferase of the Japanese syllid polychaete *Odontosyllis umdecimdonta*. Biochem Biophys Res Commun 502:318–323.

Serrano SMT, Maroun RC (2005) Snake venom serine proteinases: Sequence homology vs. substrate specificity, a paradox to be solved. Toxicon 45:1115–1132.

Shenkarev ZO, Balandin S V., Trunov KI, Paramonov AS, Sukhanov S V., Barsukov LI, Arseniev AS, Ovchinnikova T V. (2011) Molecular mechanism of action of β-Hairpin antimicrobial peptide arenicin: Oligomeric structure in dodecylphosphocholine micelles and pore formation in planar lipid bilayers. Biochemistry 50:6255–6265.

Shimomura O (2006) Bioluminescence: Chemical principles and methods. World Scientific Publishing Co. Pte. Ltd.

Shimomura O, Johnson FH, Saiga Y (1963) Partial purification and properties of the Odontosyllis luminescence system. J Cell Comp Physiol 61:275–292.

Smith ML (2002) Cutaneous problems related to coastal and marine worms. Dermatol Ther 15:34–37.

Stabili L, Schirosi R, Di Benedetto A, Merendino A, Villanova L, Giangrande A (2011) First insights into the biochemistry of *Sabella spallanzanii* (Annelida: Polychaeta) mucus: A potentially unexplored resource for applicative purposes. J Mar Biol Assoc United Kingdom 91:199–208.

Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, Lieb B, Meyer A, Tiedemann R, Purschke G, Bleidorn C (2011) Phylogenomic analyses unravel annelid evolution. Nature 471:95–98.

Suadicani S0, de Freitas JC, Sawaya MI (1993) Pharmacological evidence for the presence of a beta-adrenoceptor-like agonist in the amphinomid polychaete *Eurythoe complanata*. Comp Biochem Physiol c 104:327–332.

Tang Y, Yu F, Zhang G, Yang Z, Huang F, Ding G (2017) A purified serine protease from *Nereis virens* and its impaction of apoptosis on human lung cancer cells. Molecules 22:1123.

Tasiemski A, Schikorski D, Le Marrec-Croq F, Pontoire-Van Camp C, Boidin-Wichlacz C, Sautière PE (2007) Hedistin: A novel antimicrobial peptide containing bromotryptophan constitutively expressed in the NK cells-like of the marine annelid, *Nereis diversicolor*. Dev Comp Immunol 31:749–762.

Tzetlin A, Purschke G (2005) Pharynx and intestine. Hydrobiologia 535/536:199–225.

Ueda A, Nagai H, Ishida M, Nagashima Y, Shiomi K (2008) Purification and molecular cloning of SE-cephalotoxin, a novel proteinaceous toxin from the posterior salivary gland of cuttlefish *Sepia*

esculenta. Toxicon 52:574–581.

Umnyakova ES, Gorbunov NP, Zhakhov A V., Krenev IA, Ovchinnikova T V., Kokryakov VN, Berlov MN (2018) Modulation of human complement system by antimicrobial peptide arenicin-1 from *Arenicola marina*. Mar Drugs 16:7–9.

Verdes A, Simpson D, Holford M (2018) Are fireworms venomous? Evidence for the convergent evolution of toxin homologs in three species of fireworms (Annelida, Amphinomidae). Genome Biol Evol 10:249–268.

von Reumont BM, Campbell LI, Richter S, Hering L, Sykes D, Hetmank J, Jenner RA, Bleidorn C (2014a) A polychaete's powerful punch: Venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. Genome Biol Evol 6:2406–2423.

von Reumont BM, Campbell LI, Jenner RA (2014b) Quo Vadis venomics? A roadmap to neglected venomous invertebrates. Toxins (Basel) 6:3488–3551.

Wang JH, Kong J, Li W, Molchanova V, Chikalovets I, Belogortseva N, Luk'yanov P, Zheng YT (2006) A β-galactose-specific lectin isolated from the marine worm *Chaetopterus variopedatus* possesses anti-HIV-1 activity. Comp Biochem Physiol - C Toxicol Pharmacol 142:111–117.

Wang K-L, Wu Z-H, Wang Y, Wang CY, Xu Y (2017) Mini-review: Antifouling natural products from marine microorganisms and their synthetic analogs. Mar Drugs 15, 266;

Xu W, Klumbys E, Ang EL, Zhao H (2020) Emerging molecular biology tools and strategies for engineering natural product biosynthesis. Metab Eng Commun 10:e00108.

Yeon SJ, Chung GY, Hong JS, Hwang JH, Shin HS (2017) Purification of serine protease from Polychaeta, *Lumbrineris nipponica*, and assessment of its fibrinolytic activity. Vitr Cell Dev Biol - Anim 53:494–501.

Zörner SA, Fischer A (2007) The spatial pattern of bioluminescent flashes in the polychaete *Eusyllis blomstrandi* (Annelida). Helgol Mar Res 61:55–66.

CHAPTER 2 - Morphological characterisation of the epidermis and proboscis of *Eulalia viridis*

This chapter has been published in the following paper:

Rodrigo AP, Martins C, Costa MH, Alves de Matos AP, Costa PM (2018) A morphoanatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). *J Anat* 233:567–579.

2.1. Abstract

Eulalia viridis is a marine Polychaeta of the rocky intertidal that, despite its simple anatomy, is an active predator of much larger invertebrates, from which it extracts pieces of soft tissue through suction. This uncanny feeding strategy triggered the pursuit for the morphological mechanisms that enable adaptation to its environment. The evaluation of the worm anatomy and microanatomy, combining electron and optical microscopy, revealed a series of particular adaptations in the epidermis and in the proboscis (the heavily muscled eversible pharynx). Besides its function in feeding, the proboscis is the main sensory organ, being equipped with numerous sensorial papillae holding chemoreceptors. Additionally, the proboscis possesses tentacles that become exposed when the organ is everted. These provide fast release of mucus and toxins, from mucocytes and special serous cells, respectively (the latter involving both merocrine and apocrine processes), whenever contact with a prey occurs. In its turn, the epidermis provides protection by cuticle and mucus secretion and has a sensorial function that may be associated to the worm's uncommon green pigment cells. *Eulalia viridis* presents a series of elegant adaptive tools to cope with its environment that are evolutionarily designed to counterbalance its relatively simple body plan.

2.2. Introduction

The Polychaeta are one of the most diverse and widespread group of aquatic animals, occupying ecological niches as distinct as deep-sea vents to rivers and lakes. They play a very important role in the ecosystem, virtually at all levels of the food web, from symbiosis with primary producing bacteria to nutrient recycling in sediments and predation (e.g. Hutchings 1998; Glasby and Timm 2008). Despite sharing the same relatively simple body plan, the extraordinary diversity of polychaetes implies specific adaptations to habitat and ecological niche, at both macro- and microscopical levels. The functional perspective on adaptive traits reveals several important and distinctive characteristics of these organisms. Dales (1962), for instance, even differentiated several Polychaeta families according to the proboscis (also called "trunk" or "eversible pharynx"), which shows radiated adaptation to multiple feeding and sensing strategies. Likewise, the epidermis of these animals requires special adaptations to provide protection and sensing abilities that must match the species' ecology (see Hausen 2005, for a review). In fact, the epidermis of these animals has been revealing a complex arrangement of secretory cells whose function can vary from the release of inorganic compounds to mucopolysaccharides (Dorsett and Hyde 1970a, 1970b). Other types of cells, such as specialised pigment cells, are also common (see Bandaranayake 2006).

Eulalia viridis (Linnaeus 1767) is a phyllodocid, equipped with an eversible proboscis devoid of jaws or similar structures. Yet, the species is mostly carnivore and able to predate on live and larger prey, including mussels, barnacles, gastropods and even other annelids, seemingly by extraction (via suction) of pieces of flesh (Emson 1977; Morton 2011). These odd features, together with its uncanny green coloration (provided by unknown green pigments), led to some important past works that provided the first description of its anatomy (Michel 1964, 1968, 1969, 1970). The existence of specialised serous-like cells in the proboscis was first highlighted by Michel (1964, 1968), who then hypothesized their role in digestion. These cells were thought to secrete digestive enzymes such as trypsin, similarly to other Polychaeta like the better-known predator *Nereis virens* (Michel 1970; Michel and DeVillez 1980). Similarly, despite the work that has been made to describe the epidermis of Polychaeta (Bubel 1983; Tzetlin et al. 2002), little focus has been given to *Eulalia*. Still, pigment cells in *E. viridis* were noticed in sensorial papillae and the pharynx (Michel 1964). The presence of pigments and their transformation processes in the intestine have been investigated as well (Rodrigo et al. 2015). However, the function of pigments cells and sensory organs in the epidermis of this worm remains, to date, conjectural.

The secretion of toxins by marine invertebrates is a matter of great interest, owing to the growing attention on these compounds for biotechnological purposes. The most prominent examples are conotoxins, produced by *Conus* snails (see Kumar et al. 2015 for a review). In annelids, the presence of peptidic toxins was first described in specimens that are equipped with dedicated glands, such as *Glycera* (Bon et al. 1985). However, detailed data from *Eulalia* is essentially lacking. In general, more focus has been given to mucous secretions as a defensive mechanism than toxin delivery in the Polychaeta (e.g. Prezant 1980). Altogether, the current state-of-the-art in Polychaeta morphology and microanatomy is mostly based on aged literature that still presents many gaps regarding function. The present work aims at exploring the anatomical and microanatomical features of the proboscis and epidermis of *E. viridis* in face of their function and adaptive value, enabling it to be such a successful predator of the rocky intertidal along European shorelines.

2.3. Material and methods

Adult worms were collected from the Parede beach, Western Portugal (38°41'42" N; 09°21'36"W), a rocky intertidal beach with large patches of barnacles and mussels upon which the worm is known to feed. The animals ranged between 50 and 120 mm in length and weighted about 250 mg.

The animals were fixed in 2% v/v glutaraldehyde (in 0.2 M sodium cacodylate buffer, pH 7.4) and Zenker's (2.5% m/v potassium dichromate, 3% m/v mercury chloride, 1% m/v sodium sulphate and

5% v/v glacial acetic acid). Fixation was done at room temperature for 2 h in both cases, after which worms were divided in serial sections. Fixation in glutaraldehyde was followed by washing in cacodylate buffer (3 × 15 min), overnight post-fixation in 1% m/v osmium tetroxide (OsO_4) in 0.2 M cacodylate buffer (pH 7.4), in the dark, and washing in milliQ-grade ultrapure water (3 × 10 min). A first set of these samples was dehydrated in a progressive series of ethanol (30 - 100%), intermediately infiltrated with xylene and embedded in paraffin (Paraplast) wax. Paraffin-embedded samples were sectioned (5 μm thickness) with a Jung RM2035 model rotary microtome (Leica Microsystems). A second set was embedded in Epon resin, following Luft's mixture (Sigma-Aldrich, St. Louis, MO, USA), after being dehydrated in acetone. Intermediately infiltration was done with Epon : Polypropylene oxide 1:2, 1:1 and 2:1 (30 min each). Semi-thin (200 - 500 nm) and thin (≈ 100 nm) sections were obtained using an 8800 Ultrotome ultramicrotome (LKB Bromma).

Histological and histochemical analyses, both in paraffin and resin sections, were performed using a tetrachrome (TC) technique based on Alcian Blue (AB) for acidic sugars (pH 0.5 – 2.5 to indicate sulphated and carboxylated mucins, respectively), Periodic Acid/Schiff's (PAS) for neutral polysaccharides, Weigert's iron Haematoxylin (WH) for chromatin and Picric Acid (PA) for muscle and cytoplasm, following Costa and Costa (2012), with modifications (Costa et al. 2014; Rodrigo et al. 2015). Resin sections required permeabilization using a saturated solution of potassium hydroxide in ethanol 100% during 5 min followed by 2 min of 2% m/v borax (sodium borate) just before staining. Other general staining procedures were applied to enhance specific details, such as Toluidine Blue (TB), Paragon, van Gieson's trichrome (VG), Coomassie Blue, Acridine Orange (AO) and special combinations such as PAS-TB. Details of the procedures can be found in Costa (2018). Histological analyses, focused on epidermis and proboscis, in both paraffin and resin sections, were done with a DMLB model microscope adapted for epifluorescence with an EL6000 light source for mercury short-arc reflector lamps. The microscope was equipped with A, N2.1 and I3 filters (corresponding to blue, red and green channels, respectively). All equipment was supplied by Leica Microsystems.

Thin resin sections were analysed by Transmission Electron Microscopy (TEM). For this purpose, sections were collected onto copper or nickel mesh grids and stained with 2% m/v aqueous Uranyl Acetate and Reynold's Lead Cytrate (Venable and Coggeshall 1965). Analyses were done using a JEOL 100-SX model TEM operated at 80 keV.

External structure was analysed by Scanning Electron Microscopy (SEM), following the protocol described by Inoué and Osatake (1988), with modifications. In a first approach, samples were fixed in 2% v/v glutaraldehyde (in 0.2 M sodium cacodylate buffer, pH 7.4), followed by post-fixation in

1% m/v osmium tetroxide (OsO_4) in 0.2 M cacodylate buffer (pH 7.4). The strategy yielded good structural detail, but the worm's epidermis and overlying cuticle became wrinkled. This issue was solved using Zenker's as fixative, albeit with loss of some detail. After fixation, all samples intended for SEM were dehydrated in a progressive series of ethanol (30 - 100%), infiltrated with *tert*-butanol (3 × 15 min at 40 °C) and allowed to freeze overnight at 4 °C. *Tert*-butanol was then sublimated under vacuum till complete dryness. Finally, the specimens were mounted on an aluminium disc for gold coating using a JOEL JEE-400 vacuum evaporator. Analyses were performed in a JEOL JSM-5400 Scanning Microscope operated at 15 and 35keV.

2.4. Results

2.4.1. General Anatomy

In the head, or prostomium, SEM revealed a single antenna, one pair of simple eyes, one pair of ciliated nuchal organs positioned dorsolaterally and two pairs of palps, but no jaws or similar structures (Fig. 2.1A). Four pairs of long tentacular cirri were found in the subsequent segments: one pair in the first segment, two pairs in the second and one pair in the third (Fig. 2.1A). The worm's retractile pharynx, the proboscis ("trunk"), was externally covered with sensorial papillae (Fig. 2.1B). The eyes were found to possess lenses and to be deeply embedded within the head, being covered with skin cuticle. The eyes are internally lined with a layer of melanocyte-like cells holding brown-blackish pigments (Fig. 2.1C), which is in accordance with the observations by Whittle and Goldin (1974). *E. viridis*, like most Errantia, has one pair of locomotor parapodia per segment composed by one single lobe, the neuropodium, with a central cylindrical structure supported by a central chaeta, named aciculum, surrounded by numerous small chaetae that end in a hook (Fig. 2.1D). From the centrum of the neuropodium two cirri emerge, a small ventral cirrus and a larger elongated and flattened dorsal cirrus that displayed numerous mucocyte-like cells arranged transversally (Fig. 2.1E).

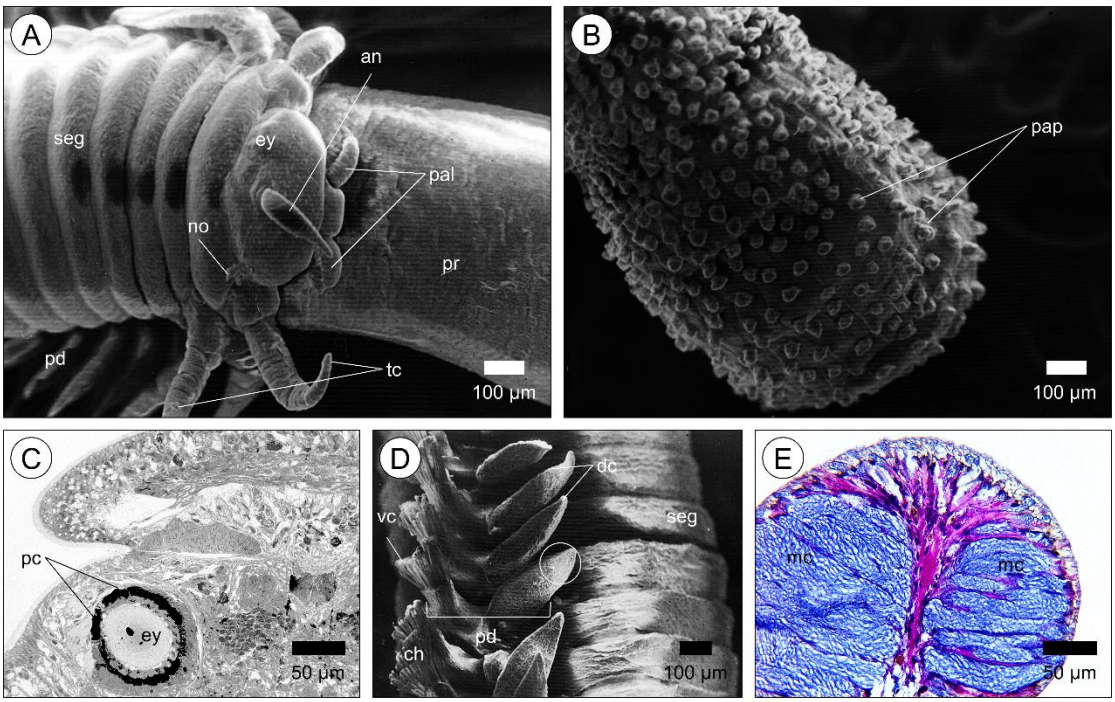


Figure 2.1. General anatomy of the Polychaeta *Eulalia viridis*. (A) First segments of the worm imaged with SEM, showing four pairs of long tentacular cirri (tc) and the parapodia (pd) in the remaining segments (seg). The worm is fitted with a single antenna (an), one pair of eyes (ey) and one pair of ciliated nuchal organs (no). Two pairs of palps (pal) are located near the mouth, through which the proboscis is extended (pr). Papillae are absent from this image due to incidental loss of proboscidial epidermis. (B) Tip of the semi-extended proboscis (pr), highlighting its numerous papillae (pap) covering its fibrous epidermis (SEM). (C) Resin section (Toluidine Blue stain) across one of the eyes (ey) of the worm, showing the densely-packed layer of melanocyte-like pigment cells (pc) underneath the photoreceptor layer. (D) SEM image of parapodia, showing a pair of parapodia (pd) per segment (seg). Note chaetae (ch) emerging from neuropodia, the small ventral cirri (vc) and the larger, elongated and flattened, dorsal cirri (dc). (E) Transversal section of the dorsal cirrus of a parapodium, highlighted with a circle in panel D (paraffin section stained with van Gieson's trichrome), exhibiting the particular distribution of mucus cells (mc).

2.4.2. Epidermis

The epidermis (Fig. 2.2) covers the entire body, with the exception of the proboscis. It is comprised of a pseudostratified epithelium formed by clearly distinct cell types. This epithelium is protected by a collagenous cuticle and an external epicuticle. Cuticle and epicuticle were found to be reactive to Periodic Acid/Schiff's (PAS-positive) and to Alcian Blue (AB-positive), respectively. Histology showed that the thickness of the epidermis is variable between body areas, being thicker in the ventral area and around parapodia. From the histochemical appraisal three main types of cells were identified: supportive cells (responsible for the secretion of cuticle and epicuticle), mucocytes

(mucous-secreting) and green pigment cells. Basal (replacement) cells could be distinguished attached to the basal lamina, easily spotted due to their pyramidal shape (Fig. 2.2A, inset).

The supportive cells were thin and elongated, had a distinctive crypt opening to the outside through a thin channel and numerous granular inclusions that shared the histochemical signature of cuticle and epicuticle, i.e., PAS-positive (strong pink) and AB-positive (blue), respectively (Fig. 2.2A). The latter case is indicative of acid sugars of mucins or the glycocalyx bound to the microvilli of epithelial cells. The fluorescence signature of supportive cells under UV light, after AO staining, also confirmed their association to and cuticle and epicuticle (Fig. 2.2B). Consequently, cuticle and inclusions associated with supportive cells were AO metachromatic (yellowish-green) whereas the epicuticle was AO orthochromatic (reddish), as shown in Fig. 2.2B (inset). The supportive cells were disseminated throughout the epidermis, albeit in larger numbers in the ventral area.

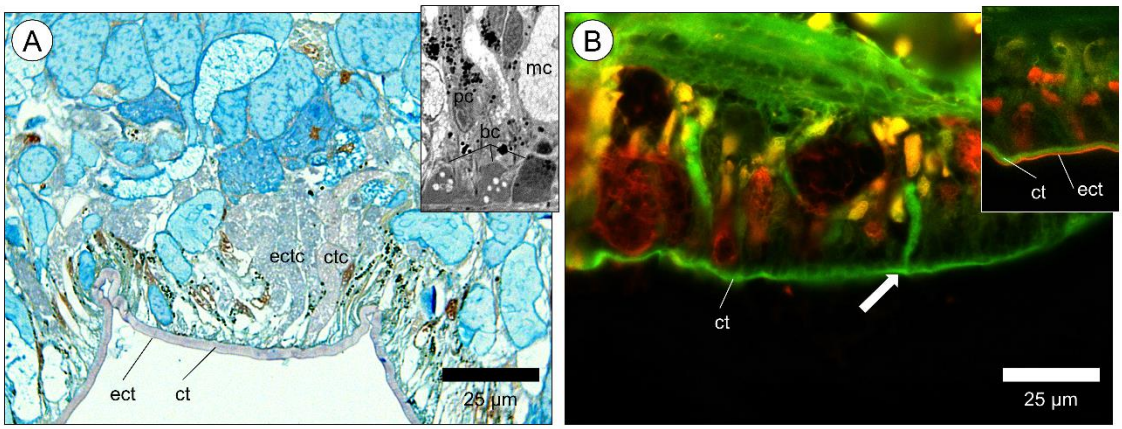


Figure 2.2. Cuticle and epicuticle secretory cells in epidermis. (A) Resin section (tetrachrome stain) across the ventral epidermis of *E. viridis*, showing the granules in supportive cells responsible for cuticle production (ctc), which presented the same PAS-positive (pink) reaction of the cuticle (ct). Similarly, granules of supportive cells responsible for epicuticle production (ectc) and epicuticle (ect) shared the same AB-positive (blue) reaction. Inset: TEM image (Toluidine Blue stain) of the epidermis showing three basal cells (bc), in which the third started already to differentiate into a pigment cell (pc). (B) Cuticle-secreting cells presented Acridine Orange fluorochrome (AO) metachromasia (yellowish-green) while epicuticle secretory cells were red-fluorescent (AO orthochromasia), seen in this image from a paraffin section. Inset: Detail of the metachromatic cuticle and orthochromatic epicuticle.

Mucocytes were present in large numbers and evenly distributed along the epidermis. The cytoplasm was mostly occupied by *saculi* whose size and histochemical signature was used to distinguish between the cells' three main maturation stages (Fig. 2.3A). At stage 1, mucopolysaccharides in *saculi* were TB metachromatic (reddish), electron-dense (Fig. 2.3B) and reactive to PAS as well. Cells at this stage were consistently closer to the basal lamina of the

epithelium. At stage 2, mucus cells had more TB-orthochromatic *saculi* (blue-purple), albeit with varying colour intensity and size. The *saculi* became then mostly electron-transparent (Fig. 2.3C) and more reactive to AB, regardless of pH, indicating likely blend of carboxylated and sulphated mucosubstances. Finally, stage 3 mucus cells were only mildly AB reactive and held large and electron-transparent *saculi* (Fig. 2.3D). Mucus cells in stages 2 and 3 were seen to possess crypts that formed narrow channels toward the surface, just as cuticle- and epicuticle-secreting cells (Fig. 2.3E). Mucin-producing cells were noted to bare an important structural function in parapodia as well. At the basis of each parapodium there was a structure holding numerous mucus cells in the three stages. Additionally, a two-layer supporting structure was observed in the dorsal cirrus of the parapodia, each layer being formed by densely-packed mucus cells at stage 2 (recall Fig. 2.1E). In either case, these mucocytes were devoid of conspicuous crypts.

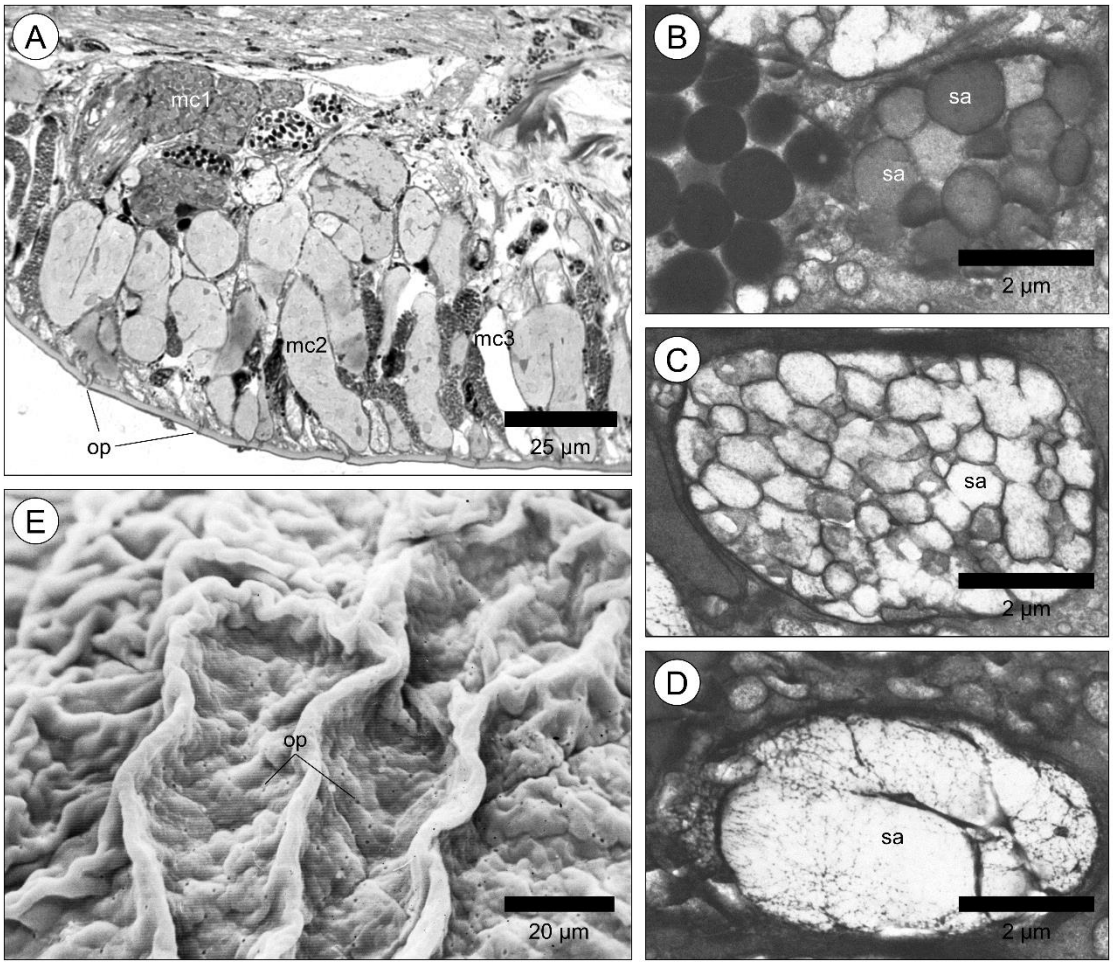


Figure 2.3. Mucus cells in the epidermis of *E. viridis*. (A) Resin section (Toluidine Blue) showing the three stages of mucus cell maturation: stage 1 (mc1), stage 2 (mc2) and stage 3 (mc3). (B) Stage 1 revealed electro-dense saculi (sa) in TEM. (C) Saculi in stage 2 were mostly electron-transparent (TEM). (D) In stage 3, the cells possessed larger and electron-transparent saculi (TEM). (E) SEM image of the epidermis, fixated with glutaraldehyde, showing openings (op), or crypts, of mucus and cuticle- or epicuticle-producing supportive cells.

Pigment cells did not appear to hold a secretory function but were seemingly responsible for the production and/or transformation and storage of pigments in granule-like structures. Pigment cells appeared in large numbers throughout the epidermis and were easily identified without the use of dyes (Fig. 2.4A). They possessed heavily electron-dense granules that remained green regardless of histological dye, in both paraffin and resin sections, as previously described by Rodrigo et al. (2015). The pigment cells were compressed between the remaining epidermis cells, to which they were bound with "cuff" links. Nuclei were positioned basally. The cells were attached to the cuticle by microvilli, forming an apical branch-like structure (Fig. 2.4A, inset). They were also found to bear several cilia that protruded through the cuticle (Fig. 2.4B).

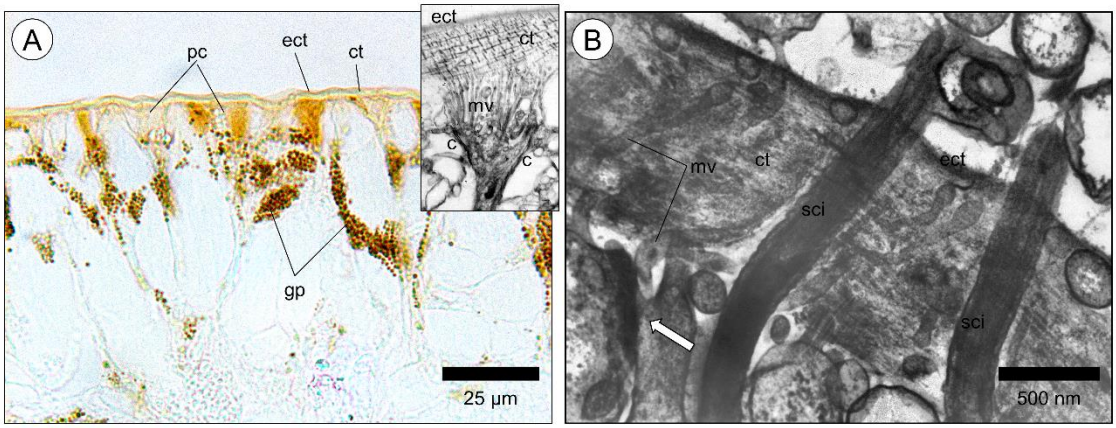


Figure 2.4. Typical green pigment cells (pc) that give the worm its distinctive coloration. (A) Green pigment granules (gp) inside pigment cells (pc) visible in an unstained paraffin section. Inset: TEM image of the apical section of a pigment cell bearing electron-dense intercellular "cuffs" (c) of adherens junctions. The cell is anchored in the cuticle with numerous microvilli (mv). ct) cuticle; ect) epicuticle. (B) TEM image showing the sensorial cilia (sci) of pigment cells protruding through the cuticle (ct) and epicuticle (ect). Evidence for a cuff (arrow) connecting a pigment cell (pc) with a mucocyte at the left. mv) microvilli.

2.4.3. Proboscis

The proboscis is a heavily-muscled eversible pharynx that was found to be lined with an integument different from the epidermis, as it was essentially fibrous and elastic and bearing numerous sensorial papillae (Fig. 2.5). When the proboscis was extended, the papillae were exposed to the environment, as well as a series of tentacles arranged in a crown around the gut opening (Fig. 2.5A). The proboscis was comprised of two main layers: inner (mostly muscular) and outer (epidermis), as shown in Fig. 2.5B. These layers were not bound by connective tissue. Instead, the space between them allowed retaining coelomic fluid during the extension process, generating hydrostatic pressure.

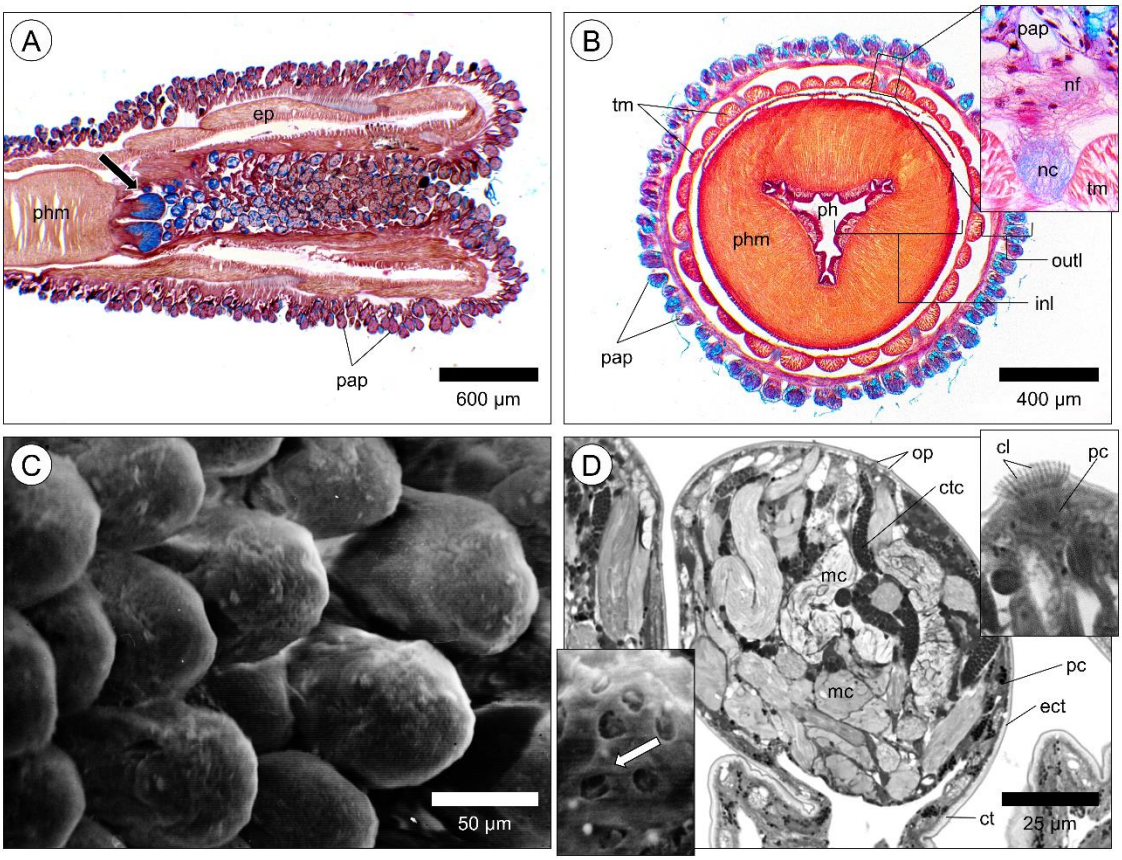


Figure 2.5. Microanatomy of the proboscis in *Eulalia viridis*. (A) Longitudinal section across the retracted organ. The outer layer is formed by a fibrous epidermis (ep), very distinct from the epithelium of the main body skin. The integument presented numerous sensorial papillae (pap) and a thin muscle layer underneath. The inner layer, inside which the specialised tentacles for toxin delivery are shown (arrow), forms a rosette attached to the strong pharynx musculature (phm). Paraffin sample, tetrachrome stain. (B) Transversal section of the proboscis. outl) outer layer; inl) inner layer where the pharynx cavity (ph) is highlighted. pap) papillae; phm) pharyngeal musculature; tm) transversal muscle bundles. Paraffin section, tetrachrome stain. Inset: Magnification of nervous fibres (nf) and nervous cord (nc) connected with all the papilla. Paraffin sample, tetrachrome stain. (C) SEM image (sample treated with Zenker's fixative) of papillae on the surface of the proboscis. Cilia and other details are not visible due to their reduced size and sample treatment with mercury chloride, which enhances structure integrity (note absent wrinkling), albeit at prejudice of some detail. (D) Transversal resin section of papillae highlighting mucus cells (mc), cuticle-producing supportive cells (ctc) and pigment cells (pc). Toluidine Blue. Upper inset: Chemoreceptor on a papilla. Note high concentration of pigment cells (pc) and the crown of cilia (cl). Resin section, tetrachrome stain. Lower inset: Large openings (arrow) in papillary chemoreceptors (SEM). ct) cuticle; ect) epicuticle; op) secretory cell openings/crypts.

The outer layer is essentially constituted by cuticle, epicuticle, fibrocytes and fibres. The papillae, which displayed a complex arrangement of different cells plus connective and nervous fibres (Fig. 2.5B, inset). They formed conspicuous rounded domes on the epidermis (Fig. 2.5C). They were also lined by epicuticle and cuticle. The major cell types in papillae could be identified as

mucocytes, pigment cells and the supportive cells responsible for the production of cuticle and epicuticle (Fig. 2.5D). Each sensorial papilla had chemoreceptors that bore pigment cells and cilia (Fig. 2.5D, upper inset) around openings (Fig. 2.5D, lower inset). The inner layer of the proboscis was lined internally by the pharynx epithelium, which consisted of a pseudostratified layer of pigment and serous cells, covered with cuticle and epicuticle as well (Fig. 2.6A, B). Special serous cells (toxin-secreting) were densely packed at the base of the tentacles mentioned earlier. The tips of the tentacles were chiefly comprised of mucus cells, mostly in stages 2 and 3, with visible crypts (Fig. 2.6C). Secretion of mucus formed domes that were easily detected by SEM (Fig. 2.6D).

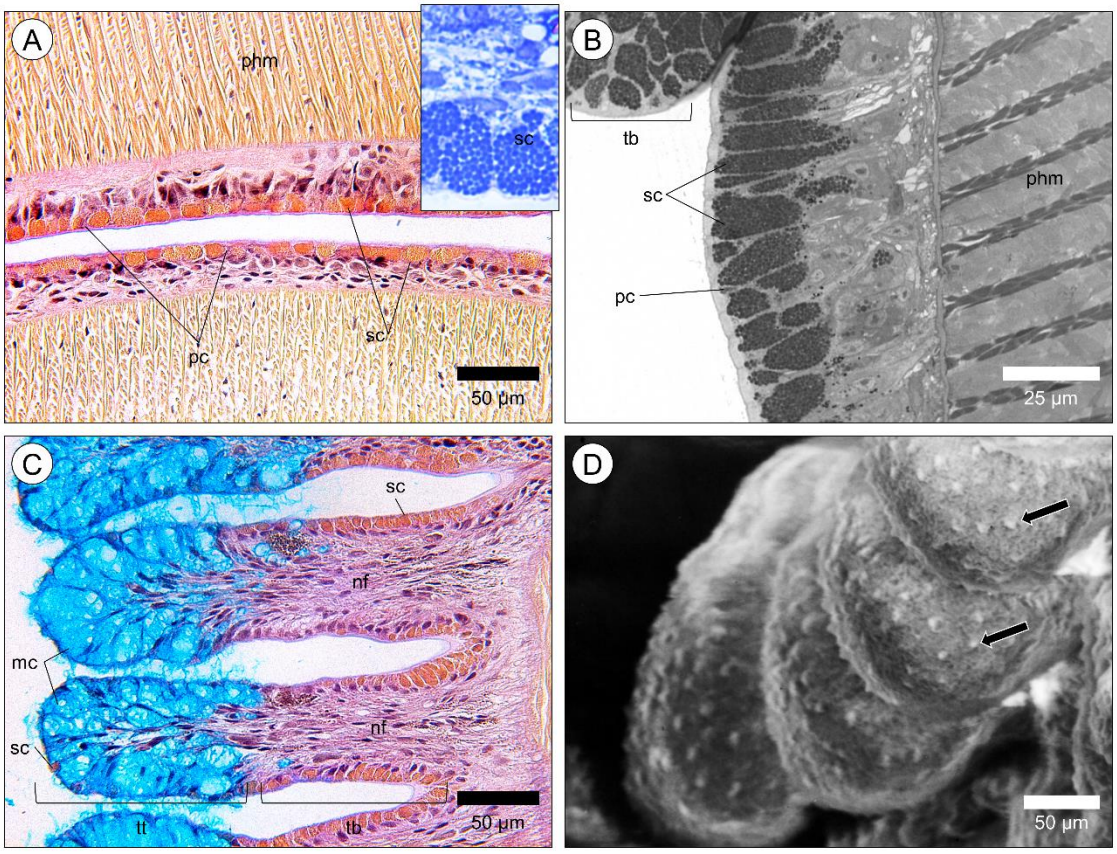


Figure 2.6. Description of the proboscidial inner layer and tentacles at the edge of the pharynx. (A) Longitudinal section (paraffin) of the pharynx (internal) epithelium with serous cells (sc), which are mildly PAS-positive and stained by Picric Acid as well. pc) pigment cells; phm) pharynx musculature. Resin section, tetrachrome stain. Inset: The vesicles of serous cells (sc) were densely stained by Coomassie Blue, indicating proteinaceous materials. (B) The base of a tentacle (tb) is characterised by densely-packed serous cells (sc) intercalated with a few pigment cells (pc). Resin section, Toluidine Blue. (C) Longitudinal section across the tentacles, which are divided in a basal area (tb) with numerous serous cells (sc) and neuronal fibres (nf), and the tentacle tip (tt), with few serous cells and numerous mucocytes (mc). Paraffin sections, tetrachrome stain. (D) SEM tentacle surface with visible domes extruding from secretory cells (arrows).

Toxin-secreting cells were easily identifiable by their idiosyncratic chalice-like shape (Fig. 2.7A). The apical part of the cell, closer to the cuticle, presented numerous electron-dense vesicles lined by a double membrane (Fig. 2.7A, inset). The nucleus was found closer to the basal lamina, as well as a well-developed rough endoplasmic reticulum (Fig. 2.7A) and numerous mitochondria. Vesicles were densely stained by Coomassie Blue, indicating proteinaceous material (Fig. 2.6A, inset). The protein vesicles were also PAS-positive (Fig. 2.6A) and TB orthochromatic (Fig. 2.6B). Different staining intensities revealed cell maturation stages and the loss of cellular material after toxin secretion (Fig. 2.7B, C). Whereas multivesicular bodies (MVBs) were a common feature in maturing toxin cells (Fig. 2.7C), mature cells held mostly electro-dense and Coomassie Blue-positive vesicles whose size prevented them from crossing the cuticle (Fig. 2.7D, E).

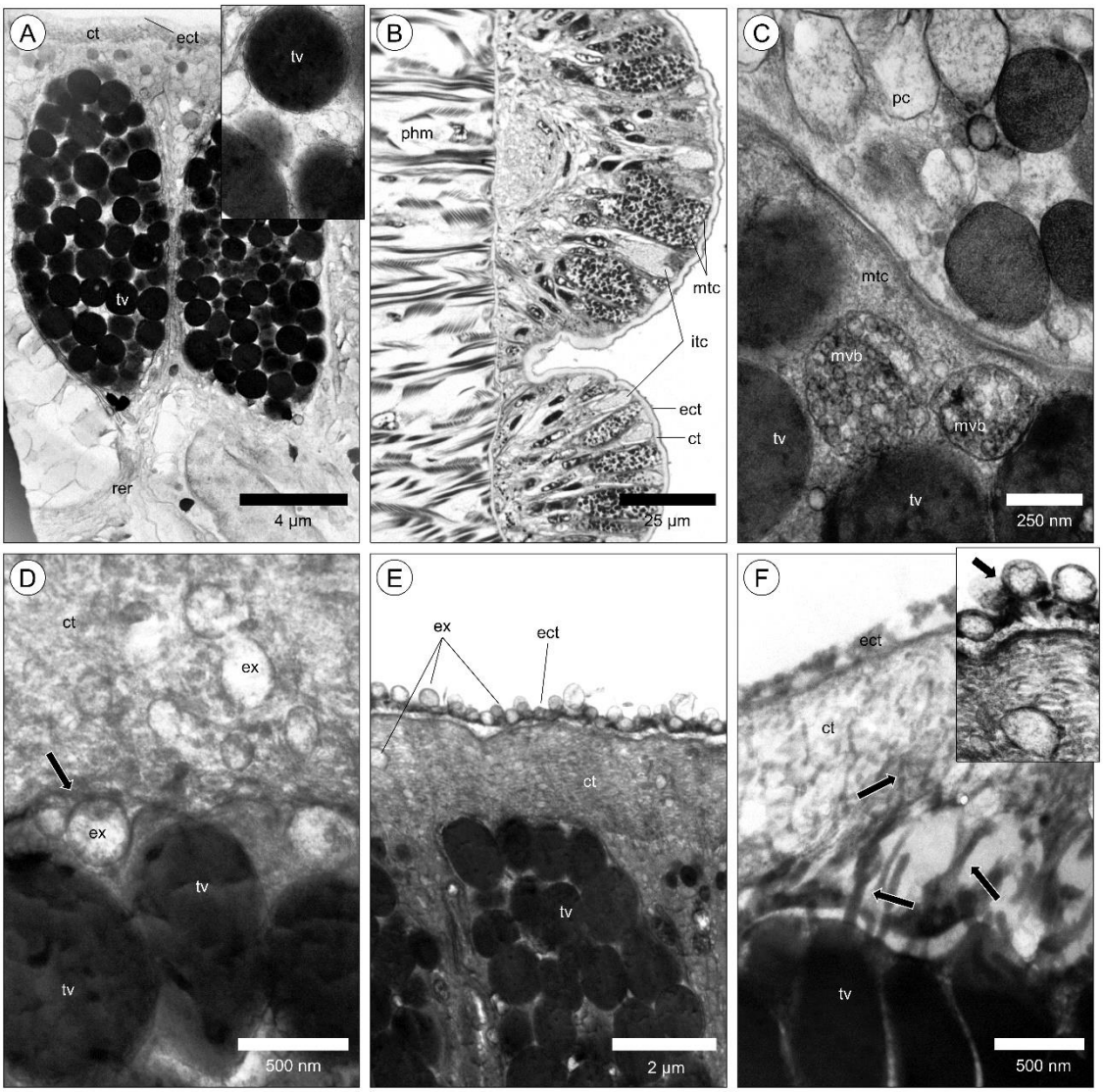


Figure 2.7. Toxin cells and toxin secretion processes. (A) Toxin cells (TEM micrograph), showing their typical chalice shape, electron-dense (mature) toxin vesicles (tv) and a well-developed rough endoplasmatic reticulum (rer). Inset: Higher-power TEM magnification of the mature toxin vesicles (tv) with double membrane. ct) cuticle; ect) epicuticle. (B) Semi-thin section (resin) of the pharynx epithelium in the proximity of the tentacles shown at Fig. 2.5A, presenting mature (mtc) and immature (itc) toxin cells (Toluidine Blue stain). ct) cuticle; ect) epicuticle; phm) pharyngeal musculature. (C) Detail of a maturing toxin cell (mtc) with multivesicular bodies (mvb) and fully mature toxin vesicles(tv), the latter holding electro-dense, highly concentrated, proteinaceous materials (TEM). Maturing vesicles are essentially multivesicular bodies that may originate exosomes or progress into mature toxin vesicles. pc) pigment cell. (D) Merocrine secretion (via exocytosis) of exosomes (ex) (TEM). ct) cuticle; tv) mature toxin vesicles. (E) Merocrine secretion in mature toxin cells (TEM). ct) cuticle; ect) epicuticle; ex) exosomes; tv) toxin vesicles. (F) Microvilli-like expansions (arrows) in the apical membrane of toxin cells formed during apocrine secretion (TEM). Inset: Detail of the apocrine and merocrine secretion processes occurring simultaneously at the surface of the epicuticle. Note profusion of exosomes and more electro-dense apocrine blebs (arrow). ct) cuticle; ect) epicuticle; tv) toxin vesicles.

Secretion from toxin cells was found to occur by two processes. Under steady-state conditions, exocytosis was a continuous process from maturing cells (merocrine secretion), with exosomes being released mostly from MVBs (Fig. 2.7D, E). However, sections from the everted proboscis revealed an apocrine-like process in mature cells, during which the apical portion of the cytoplasm and mature toxin vesicles were subjected to internal and external pressure, generating evaginations that cross the cuticle (Fig. 2.7F). These microvilli-like expansions contain the same electro-dense materials found in mature toxin vesicles (Fig. 2.7F, inset). In Fig. 2.8 the two processes are summarised, highlighting that the apocrine process is seemingly responsible for rapid secretion of toxins and associated molecules when the tip of the everted trunk contacts a target.

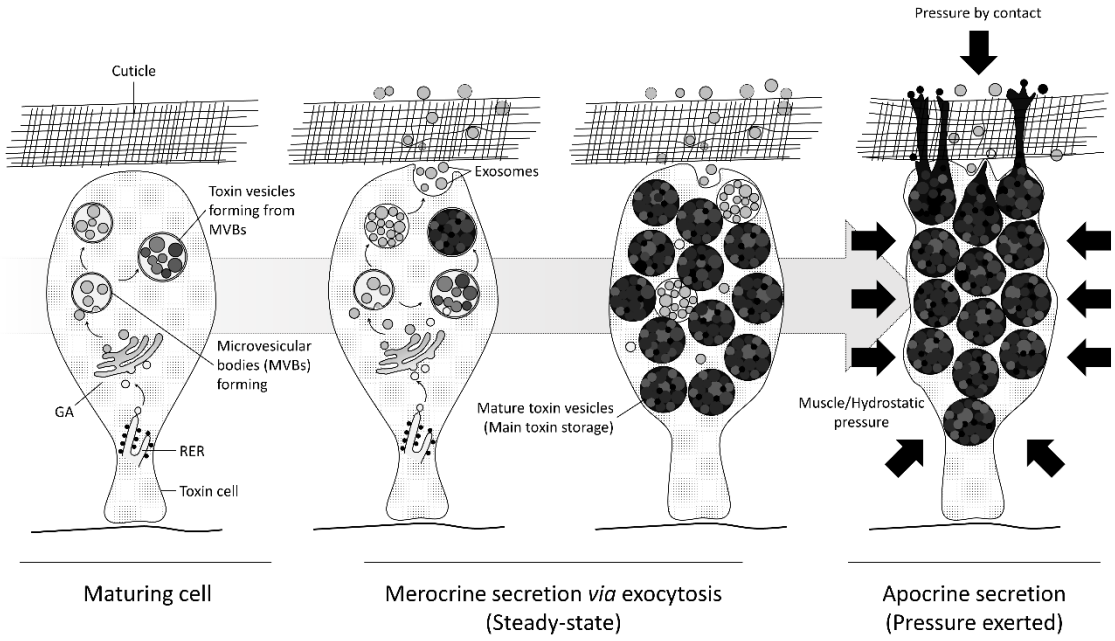


Figure 2.8. Simplified diagram illustrating the two main toxin secretion pathways. Toxins and accompanying proteinaceous materials are initially produced in the rough endoplasmic reticulum (RER), concentrated in the Golgi apparatus (GA) and allocated in multivesicular bodies (MVBs). These will either partake in the formation of exosomes or further develop into vesicles with high toxin concentration. In the first case, secretion occurs via simple exocytosis (merocrine process), when MVBs merge with the cells' apical membrane. This process is continuous, occurring in steady-state maturing and mature toxin cells. When fully mature toxin cells are subjected to pressure, they will rapidly release the stored, highly-concentrated, materials via an apocrine process. The internal pressure derives from increasing size and number of toxin vesicles. Extracellular pressure is caused by a combination of trunk muscle contraction, increased hydrostatic pressure when the proboscis is everted and by the direct contact between the tip of the proboscis (where toxin-delivery tentacles are located) and prey. The compression leads to the formation of microvilli-like expansions that rapidly protrude through cuticle and epicuticle.

2.5. Discussion

Eulalia viridis is an active predator that, like the remaining members of the family Phyllodocidae, is characterized by its strong muscular proboscis albeit being devoid of jaws or similar structures. This latter aspect is compensated by the ability to immobilise prey (such as mussels and other polychaetes) and extract a portion of soft tissue. Immobilisation requires only the direct contact between the specialised tentacles and prey as they are projected outward when the proboscis is fully everted (own observations). The process is assisted by the particular features of the proboscis that make it the species' main sensory organ. Altogether, the animal's adaptations to its feeding strategy are well reflected into its anatomy and microanatomy.

The proboscis alone is a major adaptive trait since: i) it is strongly muscular, allowing action as a suction pump; ii) it has a dense battery of sensory organs, specifically chemoreceptors, and iii) the copious production of mucins is linked to the secretion of toxins and proteolytic enzymes. The suction action of the proboscis and the secretion of mucins and enzymes have only been briefly addressed by Michel (1970). Despite these previous works, there is still little information on Polychaeta regarding their feeding strategies. Nonetheless, the basic morphoanatomy and function of the proboscis appear to be well conserved among the Polychaeta (Dales 1962). On the other hand, *Eulalia* presents the overall typical epidermis of an errant polychaete, i.e. thicker in the ventral side, with a sturdy cuticle. This arrangement is ideal to scavenge through the rocky intertidal (see the review by Hausen 2005). The epidermis that lines the proboscis is, nonetheless, morphologically very distinct from the epidermis that covers the rest of the animal's body and will be dealt with separately.

Two main functions are attributed to the skin epidermis. The first is protection and is adjuvated by the secretion of cuticle and mucus. The second is sensorial, albeit more elusive in this species. Sensing is probably associated to the same pigment cells that give the worm its characteristic green colour. Due to the absence of blood vessel networks (in parapodia, inclusively), the role of the epidermis as "gills" (Tzetlin et al. 2002) is excluded in this species. The secretion of simple substances via epidermis (acting, e.g., as pheromones), cannot be entirely dismissed. Nonetheless, the existence of multicellular glands, such as the spiral glands of *Nereis* (Dorsett and Hyde 1970b), is overruled.

Besides the usual sensory organs in the prostomium (antennae, nuchal organs, cirri and eyes) used by polychaetes to sense the environment (see Purschke, 2005 for a review), *Eulalia* uses its extended proboscis as main sensory organ, as it is used to scan the environment as the worm scouts through rocks, mussels and barnacles. When the worm encounters a prey, the proboscis is everted

via muscular action and increased hydrostatic pressure of the coelomic fluid, then externalising the specialised tentacles shown in Fig. 2.5. As these structures touch the prey, they administer a proteinaceous toxin (or toxins) accompanied by digestive enzymes (Rodrigo et al. 2014). The exact nature of these substances is still being investigated. However, they seem to immobilise prey while partially digesting the tissue at the contact area, thus enabling extraction of tissue via suction in absence of jaws or analogous structures. The serous cells in the pharynx, particularly those densely packed near the base of the tentacles, are seemingly involved in the production of the proteinaceous components of the toxin. The mucus, secreted by the mucocytes that are present throughout the animal and present in large numbers in the tips of tentacles, acts like the conveying vehicle (i.e. the carrier fluid) of toxins (recall Fig. 2.6). Michel (1968) hypothesised an "adhesive" function for the tentacles (hitherto referred to as "grosse papille") on the account of mucus secretions. However, the mucus does neither have adhesive properties, nor the existence of toxin secretions was then suspected in the aforementioned work. The function of these special tentacles is assisted by a dense complex of neuronal fibres (Fig. 2.5B, inset), which could indicate mechanoreception. It must be noted, though, that this is a sensorial function not well understood in marine invertebrates, particularly in the Annelida. Altogether, the role of the proboscis in sensing of the environment makes it a crucial feature for an active forager. The sensorial papillae (the "petite papille" referred by Michel 1964) are attached to the fibrous outer epidermis of the proboscis. The associated chemoreceptors (taste buds) have many of the species' uncanny pigment-bearing cells, which adds to the hypothesis that the green pigments may be associated to sensory functions, as discussed further below.

Toxin secretion has been hypothesized for other phyllodocids, particularly in species with specialised glands, such as in some *Glycera* (Michel 1970; Bon et al. 1985; Böggemann, et al. 2000). In *E. viridis*, however, there are no complex multicellular glands in the proboscis or epidermis. Even though serous cells lining the epithelium of the pharynx have been associated to the production of digestive enzymes (Rodrigo et al. 2015), these cells, especially those located in the vicinity of tentacles, are thus likely associated to toxin secretion.

Michel (1969, 1970) hypothesised that the secretion of proteolytic enzymes in Polychaeta is essentially merocrine, without, however, clearly identifying how the process occurs. The present findings suggest that secretion in serous toxin cells is merocrine during resting, as seen in similar vertebrate cells (Hussein et al. 2015). Nonetheless, when the worm makes contact with prey, apocrine secretion is set into motion, caused by changes in internal cell pressure, which enables faster release of highly-concentrated products (recall Fig. 2.8). For a review on the secretion of exosomes and related processes see van Niel et al. (2018). The ability to rapidly deliver the complex cocktail of toxins, permeabilizing enzymes and other substances that comprise animal poisons is an

advantage for a small predator that reaches for its targets using the proboscis as it is moved to sense the environment.

The coexistence of both apocrine and merocrine secretion has already been reported to occur in gland cells of higher-order vertebrates, from eyes to reproductive organs (Groos et al. 1999; Hussein et al. 2015). Rather than shifts in osmotic pressure, as proposed by Michel (1969), apocrine secretion in serous cells of the proboscis is trigged by the combination between: i) mechanic pressure exerted by contact with prey; ii) increased hydrostatic pressure within the proboscis, and iii) contraction of adjacent musculature (e.g. during suction). In fact, serous cell turgidity caused by water intake was not observed. Additionally, increased cell-cell and cell-cuticle spacing, which could indicate fluid retention, was noticed only when serous cells were empty. However, these features may result from fixation artefacts.

The epidermis of *E. viridis* is a well-developed organ that suits the habits of an Errantia, which implies not only defence against macro- and microorganisms, but also sensing the environment and protection during locomotion. In addition, it is possible that the epidermis is involved in chemical signalling, similarly to the alarm cells of fish (Chivers et al. 2007), as animals in captivity have been observed to react negatively to the proximity of an injured conspecific (own observation). The main cells in the epidermis of *E. viridis* are involved in the secretion of cuticle, epicuticle and mucus. The distribution and number of these cells varies between body regions. In the ventral area, more subject to abrasion with the rocky substrate, the epidermis is thicker, which is accompanied by increased numbers of supportive cells responsible for the production of cuticle and epicuticle, consistent with reports from other Phyllodocida (Tzetlin et al. 2002). In parapodia, mucus cells are abundant and hold important roles in structure and locomotion (lubrication).

The co-existence of supportive cells responsible for the production of cuticle (PAS-positive) and epicuticle (AB-positive) has been described for several polychaetes (e.g. Dorsett and Hyde 1970b, Richards 1974, Bubel 1983). Their commonness led to being simply named "epidermal cells" in the past (Krall 1968; Burke and Ross 1975). Our results are accordant with the findings by Bubel (1983) on the polychaete *Pomatoceros zamarckii*, who suggested that cuticle and epicuticle are originated from cytoplasmic granules arising from the Golgi apparatus. The collagenous nature of the cuticle of the Phyllodocida (Michel 1969; Kimura 1971; Hausen 2005) is accordant with the fibre arrangement typical of the Errantia (Westheide and Rieger 1978; Storch 1988). This pattern, which was in the past associated to body size, is now thought to be linked to mobility, as it safeguards both resistance and elasticity (see Goodman and Parrish 1971; Westheide and Rieger 1978). In addition, Bubel (1983) sustained that cuticle and epicuticle are supported by the microvilli of epidermal cells, a role that, in *Eulalia*, is seemingly secured by pigment cells. Due to the potential

role of these cells in sensing, it is possible that the function of microvilli extends to sensory as well, as discussed below. Altogether, the junction between epidermis and cuticle in annelids remains unknown, which has been pointed out since the work by Krall (1968), and it is possible that it holds significant variability between taxa.

Mucus cells, in their turn, have been extensively studied in invertebrates (Storch and Welsch 1972a), including both errant and sedentary polychaetes (Mastrodonato et al. 2005, 2006), as well as phyllodocids (Ushakov 1972). The function and processes underlying mucus secretion in Polychaeta are, however, highly variable according to species' ecology. Specifically, mucocyte function varies greatly from coastal sedentary (see Storch 1988; Hausen 2005 and Mastrodonato et al. 2005 for more details) to errant Polychaeta, especially those inhabiting intertidal areas. In this latter case, mucosubstances are involved in thermal regulation (Ushakov 1972), prevention from desiccation, lubrication for locomotion, egg protection and also as part of defence and predation mechanisms (Storch and Welsch 1972b; Prezant 1980; Storch 1988; Mastrodonato et al. 2006). All these functions can thus be justifiably applied to *E. viridis*. To these are added the structural function of mucosubstances in parapodia (recall the stage 2 mucus cells in dorsal cirri) and as carriers of toxins, which is secured by stage 3 mucus cells in the tentacles.

The ultrastructure of mucocytes in *E. viridis* agrees with previous research on aquatic invertebrates (McKenzie and Hughes 1999; Mastrodonato et al. 2005). Moreover, the present findings agree with earlier postulates for the existence of two major type of mature mucocytes in Polychaeta (Storch and Welsch 1972a; Storch 1988). Besides differences in mucus chemistry, in *Eulalia*, stage 2 cells extrude the contents of *saculi*, while in stage 3 mucins remain internalised. Our findings suggest that both stage 2 and stage 3 mucus cells arise from cells identified as stage 1 (see Fig. 2.3). This demonstrates the link and ontogeny of mucus cells proposed by some authors (Storch and Welsch 1972a; Bubel 1983; Mastrodonato et al. 2006), in contradiction with the hypothesis that these constitute, in fact, different cells (Ushakov 1972).

It has long been suspected that certain types of microvilli-bearing cells in the epidermis of polychaetes and other annelids hold a sensorial function, being able, inclusively, to trigger the secretion of substances (such as mucins) from adjacent cells (Lent 1973; Schlawny et al. 1991). The microvilli appear to be the key ultrastructural feature, as they transverse the cuticle and become in direct contact with the environment (revised by Hausen 2005 and Purschke 2005). It may thus be inferred that the special, microvilli-bearing pigment cells in the epidermis and proboscis of *Eulalia* possess a sensorial function. In the papilla, in particular, pigment cells form part of chemoreceptors (which was first reported by Michel 1964, albeit without a functional explanation), reinforcing the notion that they play a role in sensing the environment. Also, asides the presence of pigment

vesicles, pigment cells resemble the mechanoreceptors described for the epidermis of other species of Polychaeta, with particular respect to shape, positioning, distribution, microvilli surrounding special cilia and the electro-dense "cuffs" of adherens-like junctions (Jouin et al., 1985; Schlawny et al., 1991). It must be noted that mechanoreceptor cells in the Annelida can be extremely diversified (see Purschke, 2005). Altogether, it is most likely that the sensorial function of the epidermis is triggered by the combination between mechanical and biochemical stimulation in microvilli and cilia. However, other functions for pigments cells must not be discarded. Even though the worm's bright green colour is unlikely to act as camouflage, these uncanny green pigments may offer protection against UV light, similarly to melanins (see Riaux-Gobin et al. 2000). Even though the nature of the green substances has not yet been disclosed, they most likely result from haem breakdown (Lederer 1939; Kennedy 1975; Rodrigo et al. 2015). Another possible function of these pigments can be aposematism, i.e., the association between the animal's bright green coloration and its toxicity, therefore warning off attackers (see for instance Bandaranayake 2006, for a review).

2.6. Conclusions

Eulalia viridis compensates its simple morphoanatomy, with a series of subtle microanatomical features that provide adaptation to its environment and fit its predatorial behaviour. These include the sensorial papillae of the proboscis, which, being fitted with taste buds, are the main sensorial organs of the worm, without prejudice for eyes and cirri in the Prostomium. Special pigment cells in papillae and epidermis likely act as mechanoreceptors. This latter feature can represent a major adaptive trait, as it allows a fast, involuntary response to external stimuli. These adaptations, combined with the differentiated secretion of mucus, cuticle and toxins (the latter of which able to be discharged rapidly), render the species a very effective predator of the rocky intertidal, where it forages for invertebrate prey, which encompasses larger molluscs and even other Polychaeta, its conspecifics inclusively. It has thus been demonstrated that a large part of the evolutionary success of protostomes, in spite of their simple body plan, is enabled by subtle anatomical features that establish an inconspicuous but very efficient link between the animals' behaviour and their habitat.

2.7. References

Bandaranayake WM (2006) The nature and role of pigments of marine invertebrates. Nat Prod Rep 23, 223–255.

Böggemann M, Fiege D, Purschke G (2000) Ultrastructure of the proboscidial papillae in some *Glycera* species (Annelida: Polychaeta: Glyceridae). Cah Biol Mar 41, 143–153.

Bon C, Saliou B, Thieffry M, Manaranche R (1985) Partial purification of α-glycerotoxin, a presynaptic neurotoxin from the venom glands of the polychaete annelid *Glycera convoluta*. Neurochem Int 7, 63–75.

Bubel A (1983) A transmission and scanning electron microscopy study of the cuticle and epidermis of *Pomatoceros Zamarckii* (Polychaeta: Serpulidae). Trans Zool Soc Lond 36, 217–268.

Burke JM, Ross R (1975) A radioautographic study of collagen synthesis by earthworm epidermis. Tissue Cell 7, 631–650.

Chivers DP, Wisenden BD, Hindman CJ, Michalak TA, Kusch RC, Kaminskyj SGW, Jack KL, Ferrari COM, Pollock RJ, Halbgewachs CF, Pollock MS, Alemadi S, et al. (2007) Epidermal 'alarm substance' cells of fishes maintained by non-alarm functions: possible defence against pathogens, parasites and UVB radiation. Proc R Soc B 274, 2611–2619.

Costa PM, Costa MH (2012) Development and application of a novel histological multichrome technique on whole-body clam histopathology. J Invertebr Pathol 110, 411–414.

Costa PM, Rodrigo AP, Costa MH (2014) Microstructural and histochemical advances on the digestive gland of the common cuttlefish, *Sepia officinalis* L. Zoomorphology 133, 59–69.

Costa PM (2018) The Handbook of Histopathological Practices in Aquatic Environments: Guide to Histology for Environmental Toxicology. Academic Press, Cambridge, MA, USA, 292 pp.

Dales RP (1962) The polychaete stomodeum and the inter-relationships of the families of Polychaeta. Proc Zool Soc London 139, 389–428.

Dorsett DA, Hyde R (1970a) The Spiral Glands of *Nereis*. Z Zellforsch 110, 204–218.

Dorsett DA, Hyde R (1970b) The Epidermal Glands of *Nereis*. Z Zellforsch 110, 219–230.

Emson RH (1977) The feeding and consequent role of *Eulalia viridis* (O. F. Muller) (Polychaeta) in intertidal communities. J Mar Biol Assoc UK 57, 93–96.

Glasby CJ, Timm T (2008) Global diversity of polychaetes (Polychaeta; Annelida) in freshwater. Hydrobiologia 595, 107–115.

Goodman D, Parrish WB (1971) Ultrastructure of the epidermis in the ice worm, *Mesenchytraeus solifugus*. J Morph 135, 71–86.

Groos S, Wilhelm B, Renneberg H, Riva A, Reichelt R, Seitz J, Aumüller G (1999) Simultaneous apocrine and merocrine secretion in the rat coagulating gland. Cell Tissue Res 295, 495–504.

Hausen H (2005) Comparative structure of the epidermis in polychaetes (Annelida). Hydrobiologia 535, 25–35.

Hussein OA, Elgamal DA, Elgayar SAM (2015) Structure of the secretory cells of male and female adult guinea pigs Harderian gland. Tissue Cell 47, 323–335.

Hutchings P (1998) Biodiversity and functioning of polychaetes in benthic sediments. Biodivers Conserv 7, 1133–1145.

Inoué T, Osatake H (1988) A new drying method of biological specimens for scanning electron microscopy: the t-butyl alcohol freeze-drying method. Arch Histol Cytol 51, 53–59.

Jouin C, Tchernigovtzeff C, Baucher MF, Toulmond A (1985) Fine structure of probable mechano- and chemoreceptors in the caudal epidermis of the lugworm *Arenicola marina* (Annelida, Polychaeta). Zoomorphology 10, 76–82.

Kennedy GY (1975) Porphyrins in Invertebrates. Ann NY Acad Sci 244, 662–673.

Kimura S (1971) Studies on Marine Invertebrate Collagens-IV. Structural Difference in Collagens from Cuticle and Intestines of *Neanthes diversicolor*. B Jpn Soc Sci Fish 37, 432–440.

Krall JF (1968) The Cuticle and Epidermal Cells of *Dero obtuse* (Family Naididae). J Ultrastruct Res 25, 84–93.

Kumar PS, Kumar DS, Umamaheswari S (2015) A perspective on toxicology of *Conus* venom peptides. Asian Pac J Trop Med 337–351.

Lederer E (1939) Les pigments des invertébrés: (À l'exception des pigments respiratoires). Biol Rev 15, 273–306.

Lent CM (1973) Retzius cells: neuroeffectors controlling mucus release by the leech. Science 179, 693–696.

Mastrodonato M, Lepore E, Gherardi M, Zizza S, Sciscioli M, Ferri D (2005) Histochemical and ultrastructural analysis of the epidermal gland cells of *Branchiomma luctuosum* (Polychaeta, Sabellidae). Invertebr Biol 124, 303–309.

Mastrodonato M, Gherardi M, Todisco G, Sciscioli M, Lepore E (2006) The epidermis of *Timarete filigera* (Polychaeta, Cirratulidae): Histochemical and ultrastructural analysis of the gland cells. Tissue Cell 38, 279–284.

McKenzie JD, Hughes DJ (1999) Integument of *Maxmuelleria lankesteri* (Echiura), with notes on bacterial symbionts and possible evidence of viral activity. Invertebr Biol 118, 296–309.

Michel C (1964) Histologie, histochimie et innervation de la trompe d'*Eulalia viridis* (Müller), (Polychètes Errantes Phyllodocidae). Bull Lab Mar Dinard 49, 62–95.

Michel C (1968) Enzymes digestives de la trompe d'*Eulalia viridis* (Müller) (Phyllodocidae) et de *Glycera convoluta* (Keferstein) (Glyceridae) annélides polychètes errantes. Ann Histochim 13, 123–134.

Michel C (1969) Ultrastructure et histochimie de la cuticule pharyngienne chez *Eulalia viridis* Müller, (Annélide Polychète Errante, Phyllodocidae): Étude de ses rapports avec l'épithélium sous-jacent dans le cycle digestif. Z Zellforsch 98, 54–73.

Michel C (1970) Rôle physiologique de la trompe chez quatre annélides polychetes appartenant aux genres: *Eulalia, Phyllodoce, Glycera* et *Notomastus*. Cah Biol Mar XI, 209–228.

Michel C, DeVillez EJ (1980) Cuticle and Mucous Glands in the Oesophagus of an Annelid *(Nereis Virens).* Tissue Cell 12, 673–683.

Morton B (2011) Predator-prey-scavenging interactions between *Nucella lapillus*, *Carcinus maenas* and *Eulalia viridis* all exploiting *Mytilus galloprovincialis* on a rocky shore recovering from tributyl-tin (TBT) pollution. J Nat Hist 45, 2397–2417.

Prezant RS (1980) An antipredation mechanism of the polychaete *Phyllodoce mucosa* with notes on similar mechanism in other polychaetes. Fish Bull 77, 605–616.

Purschke G (2005) Sense organs in polychaetes (Annelida). Hydrobiologia 535, 53–78.

Riaux-Gobin C, Klein B, Duchenei J-C (2000) A pigment analysis of feeding modes of *Thelepus extensus* (Polychaeta, Terebellidae) a in relation to wave exposure at the Îles Kerguelen. Antarct Sci 72, 52–63.

Richards KS (1974) The ultrastructure of the cuticle of some British lumbricids (Annelida). J Zool 172, 303-316.

Rodrigo AP, Alves De Matos AP, Carrapiço F, Costa MH, Costa PM (2014) A physiological study of integument secretions in the marine polychaete *Eulalia viridis* and their potential biotechnological value. Front Mar Sci. Conference Abstract: IMMR | International Meeting on Marine Research 2014.

Rodrigo AP, Costa MH, Alves de Matos AP, Carrapiço F, Costa PM (2015) A study on the digestive physiology of a marine polychaete (*Eulalia viridis*) through microanatomical changes of epithelia during the digestive cycle. Microsc Microanal 21, 91–101.

Schlawny A, Grünig C, Pfannenstiel H-D (1991) Sensory and secretory cells of *Ophryotrocha puerilis* (Polychaeta). Zoomorphology 110, 209–215.

Storch V, Welsch U (1972a) The ultrastructure of epidermal mucous cells in marine invertebrates (Nemertini, Polychaeta, Prosobranchia, Opisthobranchia). Mar Biol 13, 167–175.

Storch V, Welsch U (1972b) Ultrastructure and histochemistry of the integument of air-breathing polychaetes from mangrove swamps of Sumatra. Mar Biol 17, 137–144.

Storch V (1988) The Ultrastructure of Polychaeta. I. Integument. In: Westheide, W., Hermans, C.O. (Eds.), Microfauna Marina, vol. 4. Verlag, Stuttgart, 13–36.

Tzetlin A, Dahlgren T, Purschke G (2002) Ultrastructure of the body wall, body cavity, nephridia and spermatozoa in four species of the Chrysopetalidae (Annelida, "Polychaeta"). Zool Anz 241, 37–55.

Ushakov PV (1972) Fauna of the USSR, Polychaetes. Vol. I, Polychaetes of the suborder Phyllodociformia of the polar Basiima and the northwestern part of the Pacific. Akademiya Nauk SSSR.

van Niel G, D'Angelo G, Raposo G (2018) Shedding light on the cell biology of extracellular vesicles. Nat Rev Mol Cell Bio 19, 213–228.

Venable JH, Coggeshall R (1965) A simplified lead citrate stain for use in electron microscopy. J Cell Biol 25, 407–413.

Westheide W, Rieger RM (1978) Cuticle ultrastructure of hesionid polychaetes (Annelida). Zoomorphologie 91, 1–18.

Whittle AC, Golding DW (1974) The fine structure of prostomial photoreceptors in *Eulalia viridis* (Polychaeta; Annelida). Cell Tiss Res 154, 379–398.

CHAPTER 3 - A phylogenetic approach to the recruitment of venom-like proteins in *Eulalia viridis*

This chapter has been submitted for publication:

Rodrigo AP, Grosso AR, Baptista PV, Fernandes AR, Costa PM (2020) A transcriptomic approach to the recruitment of venom proteins in a marine Polychaeta. *Toxins* 13, 97.

3.1. Abstract

The growing number of known venomous marine invertebrates indicates that chemical warfare plays a key role in adaptation to diversified ecological niches, even though it remains unclear how toxins fit into the life history of these animals. Our case study, the Polychaeta *Eulalia viridis*, is a predator that compensates lacking jaws by secreting an immobilizing cocktail proven by ecological assays. Whole-transcriptome sequencing of *E. viridis* revealed the pattern of proteinaceous toxins behind the effects previously noticed. The toxins are secreted by specialised serous cells in the proboscis and delivered by mucus. They evolved from non-toxin homologs to promote permeabilization, coagulation impairment and blocking neuromuscular activity of prey upon which the worm feeds by sucking pieces of live flesh. The main neurotoxins ("phyllotoxins") were found to be cysteine-rich proteins, a class of ubiquitous substances among venomous animals, vertebrates included. Albeit similarities with other Polychaeta, the sequences of most toxins were not immediately related to the Mollusca or more ancient groups, such as Cnidaria. The toxins of *Eulalia* and other Polychaeta may have evolved from non-toxin homologs that were recruited at different stages without the reduction in molecular mass and increased specificity of more recent substances, like conotoxins. By comparing to eumetazoan homologs, we show that the proteinaceous toxins of *Eulalia* evolved independently, as known for other organisms, e.g. from different rate of gene mutation, duplications and differential regulation of expression, to favour reduced specialisation towards prey or foe.

3.2. Introduction

The recent discovery of the first crustacean venom (von Reumont et al. 2014a) is a compelling evidence for the commonness of chemical warfare amongst marine eumetazoans. It adds to already famous taxa such as cnidarians, cone snails, cephalopods and fish, as well as a recent entry: the Polychaeta (von Reumont et al. 2014b). Since such forms of chemical warfare constitute a simple but efficient strategy for survival, it is not surprising to find venoms, which are usually complex cocktails of salts and proteinaceous compounds, in ancient groups. Indeed, addressing the evolution of venoms, involves understanding how this mixture emerged to provide adaptive leverage as a predator, prey, parasite or host. However, and despite the biotechnological potential of conotoxins (Williams et al. 2008), marine invertebrate toxinology is still lagging behind its terrestrial counterpart.

Assisted by next-generation sequencing, "venomics" enables discovery and molecular characterisation of multiple venom proteins and peptides, including in organisms with incipient

genomic annotation (Francischetti et al. 2013; Vonk et al. 2013; von Reumont et al. 2014b). It has already been offering an evolutionary insight on convergent or divergent recruitment of protein families into venoms of terrestrial animals (Fry et al. 2009; Casewell et al. 2013), providing answers to questions such as the importance of orphan genes, neofunctionalization of duplicates by amino acid substitution and the evolutionary pressures regulating the optimal use of metabolically expensive resources (Fry et al. 2009; Khalturin et al. 2009; Chang and Duda 2012; Morgenstern and King 2013). Indeed, it has been shown that snake venoms evolved from proteins holding non-toxin functions toward the "three-finger toxin" group that includes neurotoxins, anti-coagulants and cytotoxins (Fry et al. 2009; Kini and Doley 2010). Accordingly, the first insights into Polychaeta toxins are revealing a wider span of species than anticipated, including glycerids (bloodworms), and amphinomids (especially fireworms) (von Reumont et al. 2014b; Verdes et al. 2018). Even though information on toxin activity is scarce, the search for homologs suggests a combination of cytotoxins, hemotoxins, neurotoxins and venom-specific proteolytic enzymes (see Chapter 1 for specifics) that, as for other invertebrates like octopuses, should permeabilise the tissue of their prey (Whitelaw et al. 2016).

Eulalia viridis (Phyllodocidae), a toxin-secreting intertidal predator rather inconspicuous asides its bright-green colour, secretes toxic mucus whose noxious proteins (hitherto termed "phyllotoxins") assist its uncanny feeding behaviour: the worm stuns its prey and sucks a piece of flesh with its powerful proboscis, albeit devoid of jaws and complex glands. It must therefore be noted that phyllotoxins are not injected, therefore the secretions do not really constitute a venom. Also, since no wound is actually inflicted to deliver the noxious substances then, as Nelsen et al. (Nelsen et al. 2014) suggested, the novel term "toxungen" is actually more accurate, as the toxins are applied via surface contact whereas poisons, by definition, lack delivery structures altogether. However, *Eulalia* possesses specialised tentacles at the tip of the proboscis, equipped with mucocytes and a layer of calix-like cells that, when subjected to pressure, e.g. by direct contact with prey, rapidly release toxins from dense intraplasmatic vesicles (as described in Chapter 2). It must be noted that the presence of lytic enzymes in the proboscis has earlier been described decades ago (Michel 1968, 1969, 1970), although toxins were then not referred. The Polychaeta are as diverse as phylogenetically complex, seemingly not monophyletic, with many groups lacking clear evolutionary positioning and lacking support from molecular systematics (Bartolomaeus et al. 2005; Struck et al. 2011). Similarly, there is little information on the recruitment of protein-encoding genes into the composition of venoms in these protostomes (Fauchald and Jumars 1979; McHugh 2000). The present work aims at identifying the main toxin homologs secreted by *Eulalia* and compare them to representative homologs among known venomous taxa.

3.3. Material and methods

3.3.1. Animal collection

Eulalia viridis (≈ 120 mm total length and weighting ≈ 250 mg) were collected from a rocky intertidal beach in Western Portugal (38°41'42" N; 09°21'36"W). Animals were kept in a mesocosm environment recreating their natural habitat, consisting of dark-walled glass aquaria equipped with constant aeration and recirculation, fitted with natural rocks and clumps of mussels to provide shelter and feed, as set-up by Rodrigo et al. (2015). Salinity, temperature and photoperiod were restrained within 35, 16 °C and 12:12 h, respectively.

3.3.2. RNA extraction and high-throughput sequencing (RNA-seq)

To identify mRNAs coding for putative proteins in the proboscis, which is involved in toxin secretion and delivery, the body wall (which includes skin and underlining musculature) was taken as reference organ. Worms were dissected for the excision of proboscis and body wall. Extraction of total RNA was done on portions infiltrated with RNALater using the RNeasy Protect Mini Kit coupled with in-column DNA digestion using a RNAase-free DNAase set (all from Qiagen, Hilden, Germany), following manufacturer instructions. Quantification of total RNA and initial quality assessment was performed using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Samples were stored at -80 °C until further analysis. The RNA integrity number (RIN) for each sample was determined in an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), with all samples being found within suitable parameters, i.e. intact or partially degraded samples with RIN ≥ 7, input of ≥ 1 μg total RNA, free of contaminating DNA(Schroeder et al. 2006). Library preparation was done using Kapa Stranded mRNA Library Preparation Kit and the generated RNA fragments were sequenced in an Illumina HiSeq 4000 platform, using 150bp paired-end reads. One of the Proboscis (pr) and Body wall (bw) samples were sequenced with high depth for transcriptome assembly (100M reads) and normal coverage was used for two biological replicates (20M reads).

3.3.3. Transcriptome Data Analysis

The quality of RNA-seq data was assessed using FASTQC (v0.11.7, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For all libraries, TrimGalore (v0.4.4, https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) was used to: trim the 13 bp of Illumina standard adapter; remove low quality reads with “Phred” cut-off higher lower than 20; and set the minimum read length to 20bp. An average of 10% of low-quality reads were removed from

each sample. Quality filtered reads from high-depth sequenced samples were combined and assembled with Trinity v2.8.4 (Grabherr et al. 2011) using default parameters. Two strategies were used to evaluate the quality of the assembled transcriptomes by: examining the RNA-Seq read representation of the assembly (at least 80% of the reads); and computing the Ex90N50 transcript contig length (approximately around 1000 bp). Coding regions within the assembled transcriptomes were predicted using TransDecoder v5.5.0 (Haas et al. 2013), with default parameters. Proboscis-specific transcripts were selected by first mapping independently all samples to the *Eulalia* assembled transcriptome with Kallisto v0.44.0 (Bray et al. 2016), and selecting the transcripts significantly overexpressed relative to the body wall (FDR-adjusted $p < 0.05$), with expression fold-change higher than two. Statistical analysis were computed using R 3.5 (Ihaka and Gentleman 1996), through packages edgeR and limma. Finally, proboscis-specific transcripts with coding regions were functionally annotated by scanning for homology against: 1) UNIPROT cluster UniRef90 (Suzek et al. 2015) by generating a customised toxin database, with BLASTP v2.5.0 (States and Gish 1994), having set a maximum *e*-value of 10^{-5}, and protein domains from PFam (Finn et al. 2014), using HMMER v3.1b2 (Eddy 2009). Bulk data is freely accessible at Gene Expression Omnibus (GEO) DataSets database, accession number GSE143954.

3.3.4. Quality assessment and validation

The results from RNASeq were validated by RT-qPCR for the selected representative genes. In brief: cDNA was synthetized from the total RNA samples using the First-Strand cDNA Synthesis Kit (NZYTech). Primers were designed using Primer Blast and verified *in silico* with Oligo Analyser (Table 3.1) to amplify an expressed sequence tag (EST) for the selected genes (Table S3.2), namely hyaluronidase, cysteine rich venom protein (CRISP), C type lectin-like protein and metalloproteinases M12A and M12B, and GAPDH as internal control, as suggested by Thiel et al. (2017)**.** Amplification was performed in a Gradient Thermocycler96 (Biometra). Following resolving PCR products in an agarose gel, these were Sanger-sequenced, translated and matched against the Toxins database using BLASTP. All products were found to aligned with the wanted targets. The RT-qPCR was then performed in a Corbett Rotor-Gene 6000 thermal cycler (QIAGEN) using the NZY qPCR Green Master Mix (NZYTech). The programme included an initial denaturation (95 ºC, 10min), followed by 45 cycles of denaturation (94 ºC, 45s), annealing (54 ºC, 35s) and extension (72 ºC, 30s). Expression analysis was done using the ΔΔCt method (Livak and Schmittgen 2001) (Fig. 3.1). Primer melting analysis was also conducted to verify specificity of hybridisation.

Table 3.1. Primer sequences.

Target	Forward (5'-3')	Reverse (5'-3')	Amplicon size (bp)
Gap_dh	CAGGGGTGCTAAGCAGTTGG	GAGAAGGCTGGGGCTCATTT	165
Crisp	TGTAATGGCTGACGGTCTCG	AGAACATGATGTGGGGCTCG	123
C type lectin	TAGCCATGTCACTCAGCGTC	ACCTGCTCGAATAACGGTGG	130
Hyal	ACGGAGCCATTTCACCCTTT	AGATTTTGAGCAGGCCACGA	137
Reprolysin	GACACTGTGGCCTTACTCCC	TAATAGCTTCGGTCCCGTGC	132
Astacin	TGACCAGGCAATCACAAGGG	CATGATGTTTCTCGCGCCAC	134

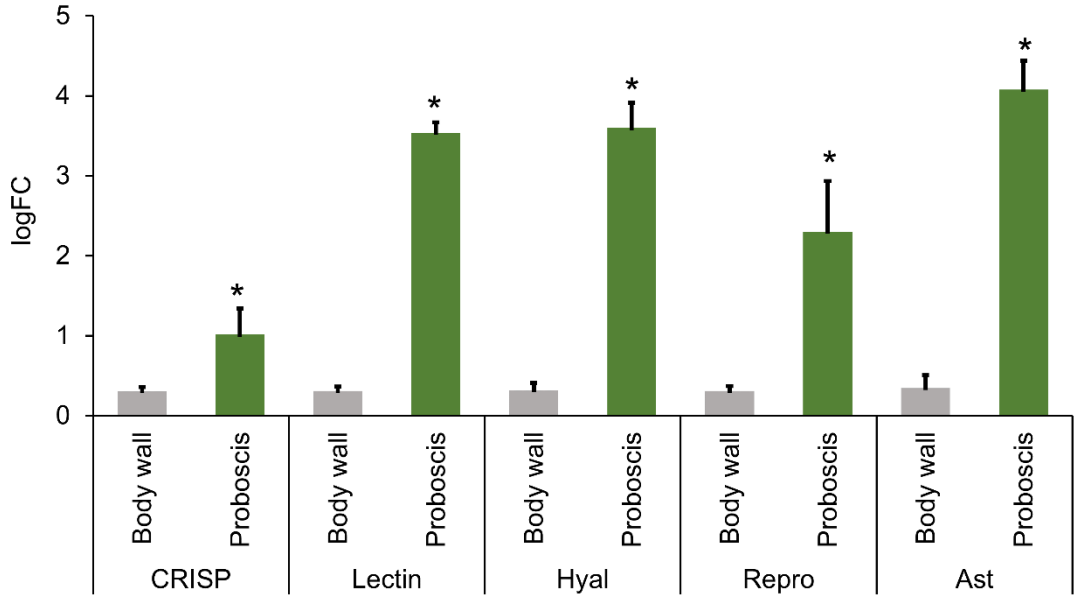


Figure 3.1. Expression analysis of key toxins by RT-qPCR, comparing the proboscis and body wall. Data are expressed as mean relative expression of Crisp (cysteine-rich venom protein), Lectin (C type lectin), Hyal (hyaluronidase), Repro (peptidase M12B) and Ast (astacin, peptidase M12A). The housekeeping gene Gap-dh was used for normalization. * Indicates significant differences to body wall for each respective target (Kruskal-Wallis H, $p < 0.05$).

3.3.5. Multi-gene phylogenetics

To compare the sequences of venom components between several clades as a proxy of functional convergence or divergence, we first produced a shortlist of translated contigs with logFCs (logarithmic fold change) above 10 from the top hits extracted from the toxin database and Pfam. A total of 38 genes were selected, corresponding to 9 different proteins (isoforms concatenated). A sequence from each different protein was chosen to scan for homology against NCBI's RefSeq, and

UNIPROT databases using Blast (Altschul et al. 1990) when necessary, with the following clade restrictions: Mammalia, Arachnida, Cephalopoda, Annelida, Hymenoptera, Bivalvia, Reptilia, Serpentes, Scorpionida, *Conus*. The best hits without restriction were also used (Tables S3.3 and S3.4). After alignment, phylogenetic trees were produced (500 bootstrap pseudoreplicates) for the sequences of each individual protein using maximum likelihood and the JTT matrix-based model, based on Tamura & Nei (1993). Sequence alignment and trees were produced with Mega X (Kumar et al. 2018).

The multi-gene phylogenetic analysis was done from amino acid sequences selected from top hits of 18 key species representative of different phyla by homology with hits for the eight genes encoding the shortlisted proteins of interest (CRISP, Hyaluronidase, C type lectin, Serine protease, Peptidase M12A and M12B, EGF domain-containing protein and endothelin converting enzyme). Genes were partitioned for analysis and the best model for each was selected according to the lowest Bayesian Information Criterion (BIC): Whelan and Goldman (WAG) model with discrete gamma distribution (G) and evolutionarily invariable (I) for Peptidase M12B, WAG+G model for CRISP, C type lectin, Serine protease and Endothelin, the Le Gascuel model (LG) plus G+I for Peptidase M12A and EGF domain-containing protein and LG+G for Hyaluronidase. The aligned and translated sequences are compiled in Supplementary Table 3.5. The Bayesian phylogenetic tree was inferred using MrBayes 3 (Tamura and Nei 1993) after 1,000,000 generations and sampling every 100 generations. The tree was rooted on the Cnidaria.

3.3.6. Microscopy

Histological samples were prepared for both light microscopy and scanning electron microscopy (SEM). The external structure was analysed by SEM for morphological characterisation, following the protocol described by Inoué (1988). The internal structure was analysed by light microscopy for phenotypic anchoring through the localisation of cells producing cysteine-rich proteins. Identification of thiol groups histochemically was performed based on the Protein Thiol Fluorescent Detection Kit (Invitrogen, ThermoFisher), as described by Gonçalves & Costa (2020). In brief: slides with samples fixed in glutaraldehyde and Bouin's and embedded in paraffin. The samples were then deparaffinated, rehydrated and stained with detection reagent, after being permeabilized (1% v/v Triton) and reduced with 1% m/v DTT to free oxidised thiols. The procedure was done using a humidity chamber. Negative control slides, without detection reagent, were included for quality assessment. The slides were mounted with DAPI for nuclei staining, photographed using a DM 2500 LED model microscope equipped with a MC 190 HD camera (both from Leica) and fluorescence images in all channels were treated with ImageJ (Schneider et al. 2012) (Fig. S3.2).

3.4. Results

3.4.1. Discovery of putative toxins

Transcriptomics was used to identify putative toxin mRNAs by running RNAseq in the proboscis (pr) and body wall (bw) to pinpoint overexpressed genes in the proboscis (Fig. 3.2A). The analysis of *Eulalia*'s proboscis and body wall whole-transcriptome yielded a total of ≈ 400,000 contigs, corresponding to ≈ 55,000 single-gene transcripts in open reading frame (ORF) (Fig. 3.2B). Of these, 2048 were significantly differentially expressed between the proboscis and body wall (adjusted $p < 0.05$), with 718 ORFs showing proboscis-enriched expression (Fig. 3.2C). Sequence matching using BLASTp by contrasting against the customised Toxins database retrieved from Uniprot resulted in 120 amino acid sequences considered to be the most significant toxin-related proteins (See Table S3.1), i.e., with lowest *e*-value, to be produced in the proboscis of the worm (Fig. 3.2B-D). Altogether, the proboscis and body wall of the worm yielded very distinct transcriptional signatures, from which 120 novel transcripts correspond to proboscis-specific and toxin-related proteins (Fig. 3.2D).

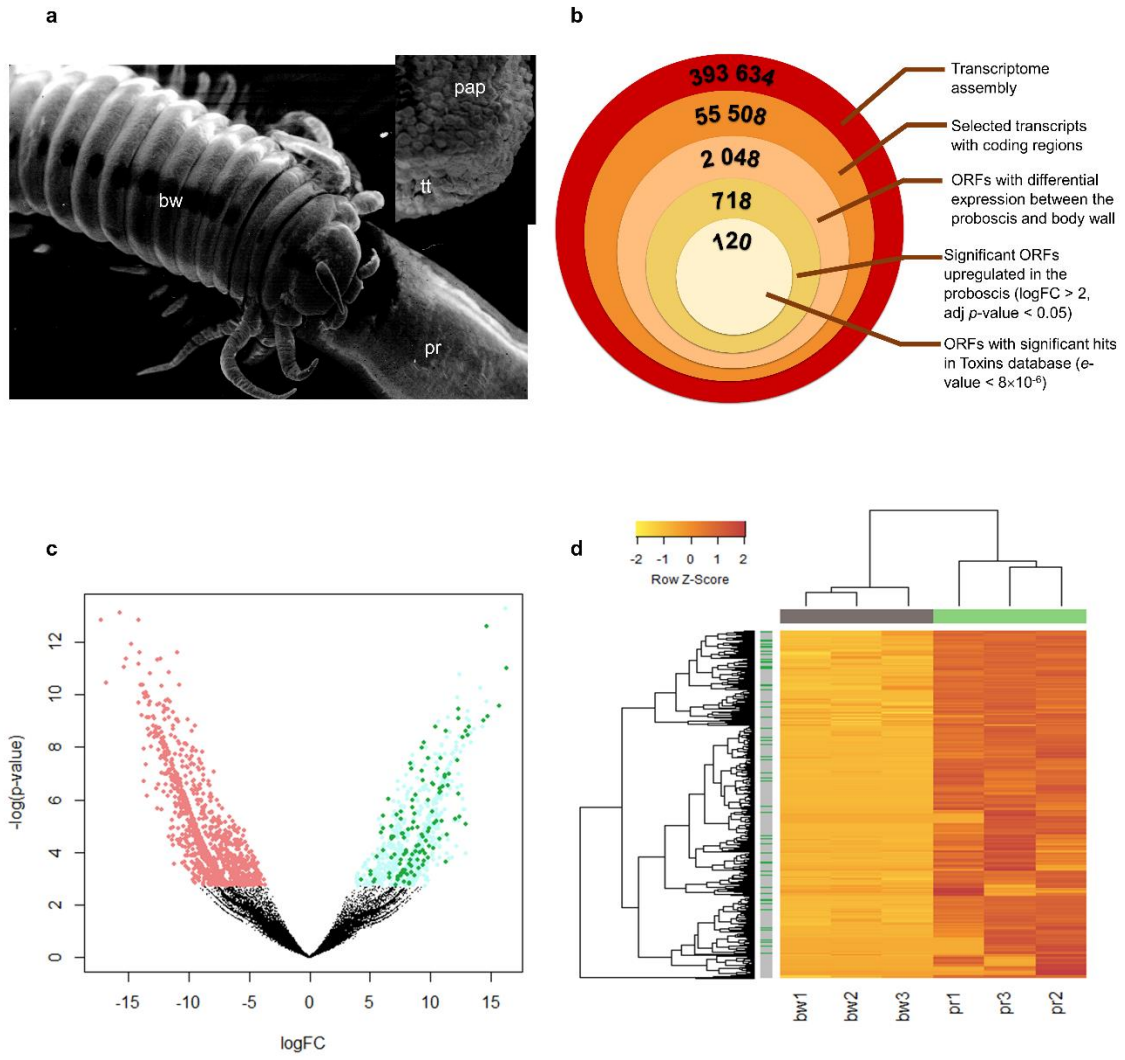


Figure 3.2. Transcriptomics revealed toxin-like proteins secreted by *Eulalia*'s proboscis. (A) Scanning electron microscopy (SEM) micrographs of the head of *Eulalia viridis*, evidencing the body wall (bw), which comprises the skin and subjacent musculature, and the everted proboscis (pr), also termed trunk of eversible pharynx. *inset*: tip of the fully everted proboscis, covered with sensorial papillae (pap) and the specialised tentacles around the mouth (tt) that the animal uses to deliver toxin-containing mucus to the surface of prey. (B) Tiered whole-transcriptome refinement to select transcripts of interest for toxin identification. (C) Volcano plot of ORFs under- (red), overexpressed (blue) in the proboscis (logFC > 2) and from those, the ones with hits in Toxins database (green). (D) Heatmap showing the 718 ORFs that yielded over-expression in the proboscis (pr), compared with the body wall (bw), for three different worms (identified as 1, 2 and 3). Side bar (grey) is indicative of genes with match in Toxins database (green). Complete linkage as employed as clustering function and Euclidian distances as metric. Data were row-normalised.

3.4.2. Conserved domains in toxin-like proteins

From the 718 ORFs with higher expression in the proboscis, those with highest expression were predominantly toxins (Fig. 3.3A). Analysis then focused on the shortlisted sequences (120 ORFs) with hits in the Toxins database, where a total of 29 types of conserved domains were found (Fig. 3.3B). Several of these domains were represented in more than eight different putative proteins. It is the case of the conserved domains: glycoside hydrolase (GH56), Peptidases M12A, M12B and M10, Adam CR 2, C type lectin, CAP, Trypsin, EGF-like, DUF1986 and CUB (for complement C1r/C1s, Uegf, Bmp1). Several domains are characteristic of peptidases that are highly expressed in the proboscis of *Eulalia*, namely peptidase M10, astacin and reprolysin (the two latter corresponding to peptidases M12A and M12B, respectively), that are zinc-dependent metallopeptidases. In fact, besides CRISPs, metalloendopeptidases and hyaluronidases were the most representative proteins matching those in the Toxins database and within the group of highly overexpressed protein homologs in the proboscis (Table S3.2).

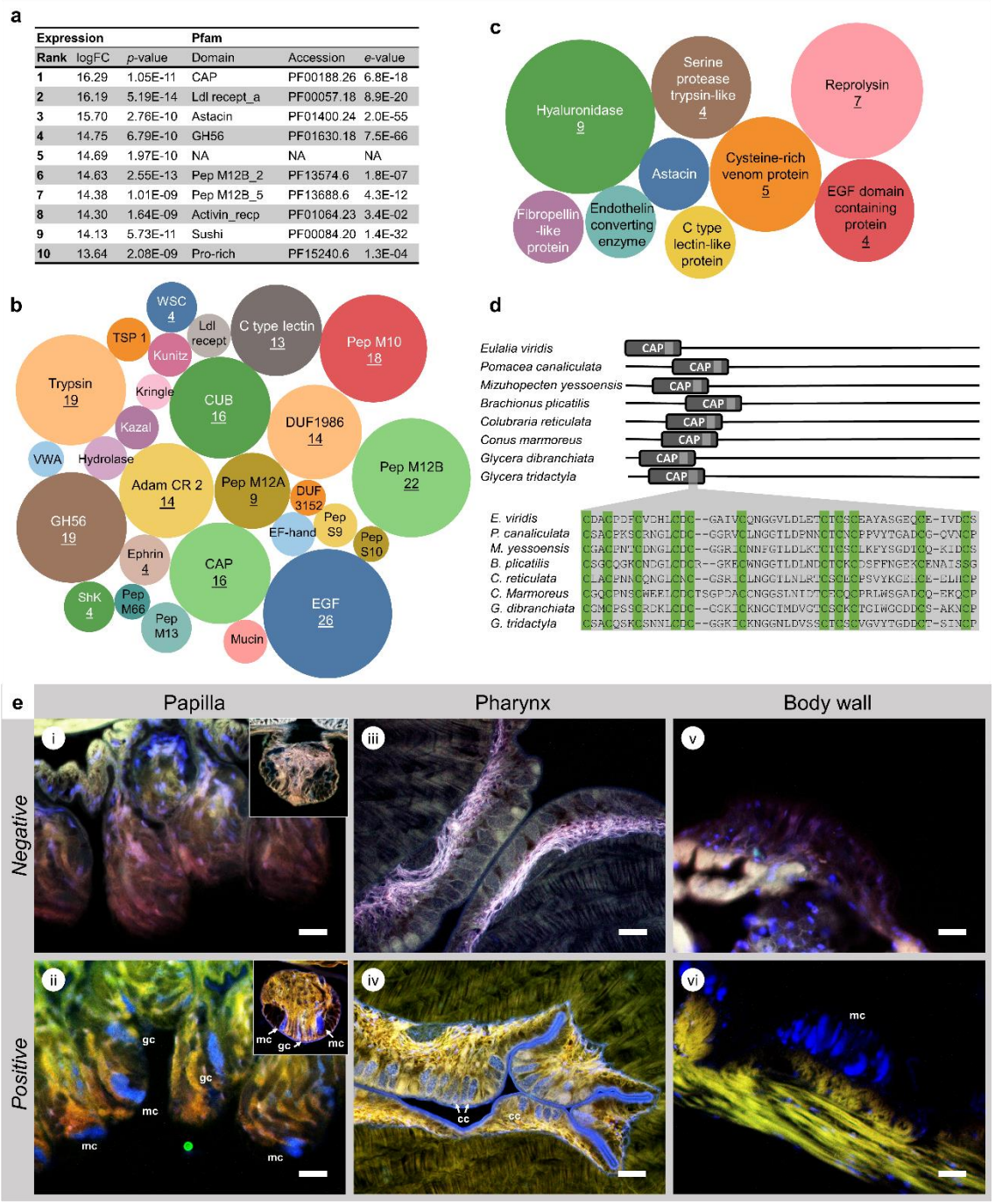


Figure 3.3. Common signatures of *Eulalia*'s toxins. (A) Top ten hits of proteinaceous substances upregulated in the proboscis, compared to the body wall and matched against Pfam. (B) Main conserved domains present in proteins with hits in Toxins database overexpressed in the worm's proboscis (logFC > 2). Figures indicate number of proteins (when ≥ 4) bearing each domain and the size of circles is representative of number of ORFs. (C) Proteins of interest upregulated in *Eulalia* proboscis with logFCs > 10, with match in the Toxins database. Figures indicate the number of different ORFs pertaining to each protein (when ≥ 4). The size of circles represents relative expression. (D) Illustrative representation of a CRISP (cysteine rich venom protein) from *E. viridis* and other organisms, highlighting the location of the well-conserved CAP domain and their cysteine alignments. (E) Localisation of toxin-secreting glandular cell in *Eulalia* through detection of CRISPs by histochemical fluorescent staining of thiols. i,ii) Negative and positive reaction of

thiols, respectively, in the sensorial papillae lining in the outer integument of the proboscis. The reaction produces a blueish fluorescent probe with a positive signal in granular cells (gc) and mucocytes (mc). The morphology of the latter is distinctive due to the presence of large mucus sacculi with homogenous positive staining for thiols due to the presence of sulphated mucins. Specimen fixated with glutaraldehyde. *Inset*: Detail of individual papillae. iii, iv) Negative and positive staining in the internal epithelium of the proboscis (pharyngeal epithelium) revealing the single-layer of calix cells (cc) that have been identified as toxin-secreting cells. The granules in these cells yield strong blue fluorescence. Specimen fixated in Bouin's. v, vi) Fluorescent labelling (negative and positive, respectively) of thiols in the skin is circumscribed to mucocytes (mc). No granular or calix cells were here detected. Specimen fixated in glutaraldehyde. Scale bars: 25 µm.

3.4.3. Cysteine-rich neurotoxins as major component of Eulalia's toxungen

Thirty-eight ORFs with significant representation in the Toxins database were found to be highly proboscis-specific, inferred from high expression ratios (logFC > 10). These ORFs represent multiple transcripts for nine different proteins (Fig. 3.3C). In terms of representativeness, hyaluronidase leads the subset, with nine sequences, followed by Reprolysin, with seven, and CRISP, with five. Indeed, cysteine-rich venom CRISP-like proteins are well-represented in this subset, but they markedly contrast with the majority of the proteins, which consist of enzymes acting on the extracellular matrix, therefore involved in tissue permeabilization. Furthermore, one of CRISP ORFs was found to be the most significantly upregulated among all matches (see Fig 3.3A), attaining high expression ratios (logFC > 16) between the proboscis and the body wall.

Since one of the main characteristics of cysteines is the presence of a highly reactive thiol side chain, CRISPs were localised histochemically, confirming the location of gland cells and providing phenotypic anchoring for CRISP differentially expressed genes (DEGs). The findings (summarised in Fig.3.3E) revealed the existence of thiol-rich granular cells in the sensorial papillae that cover the fibrous integument of the proboscis, in the calix cells within the epithelium that lines the inside of the proboscis. These two types of cells are not present in the main body wall. There are, however, mucocytes that stain positively for thiols, likely due to sulphated mucins. The cellular structure of mucocytes and granular or calix cells has already been comprehensively described in Chapter 2. The homogenous appearance of the large mucus sacculi clearly contrasts with that the dense, much smaller and regular-shaped protein granules that characterise calix cells. Mucocytes are, nonetheless, widespread throughout the surface of the animal, unlike granular cells in the proboscis.

3.4.4. Phylogenetic analysis of individual components

The sequences of shortlisted proteins (recall Fig. 3.3C) were matched against their homologs (see Table S3.3) with most significant hits in venomous and non-venomous representative eumetazoans, from cnidarians to humans, plus other organisms that produced relevant matches, such as Basiodiomycota (Fungi) and Orthonectida mesozoans, when available for comparison, were analysed based on phylogenetic models (Fig. 3.4, S3.1). The trees show different phylogenetic relationships between translated sequences, albeit a trend to group the proteins from *Eulalia* within clades that allocate homologs from venomous animals. How the clades are organised does not, however, reflect the known evolutionary history of eumetazoans. If, in one hand, the findings confirm the positioning of *Eulalia* protein homologs within common constituents of venomous secretions, on the other they evidence the intricate recruitment of proteins into venomous secretions and their separate divergence (or functional convergence) at various stages of animal evolution.

Another significant case is that of hyaluronidases, which are ubiquitous amongst animal venoms and well-conserved between eumetazoans venomous or not, that show unclear positioning, as *Eulalia* homologs were similar to those of scorpions and bees than from other annelids, even molluscs or the coral *Pocillopora damicornis*. These enzymes, combined with zinc peptidases, namely serine proteases, astacins and reprolysins, as well as the enzymes similar to endothelin-converting factor (Fig. S3.1), show the recruitment of diverse enzymes directed against the extracellular matrix of prey. In their turn, C-type lectins from *Eulalia* formed a distinct group with reptiles and scorpions, with *Glycera* being positioned farther within the same major clade.

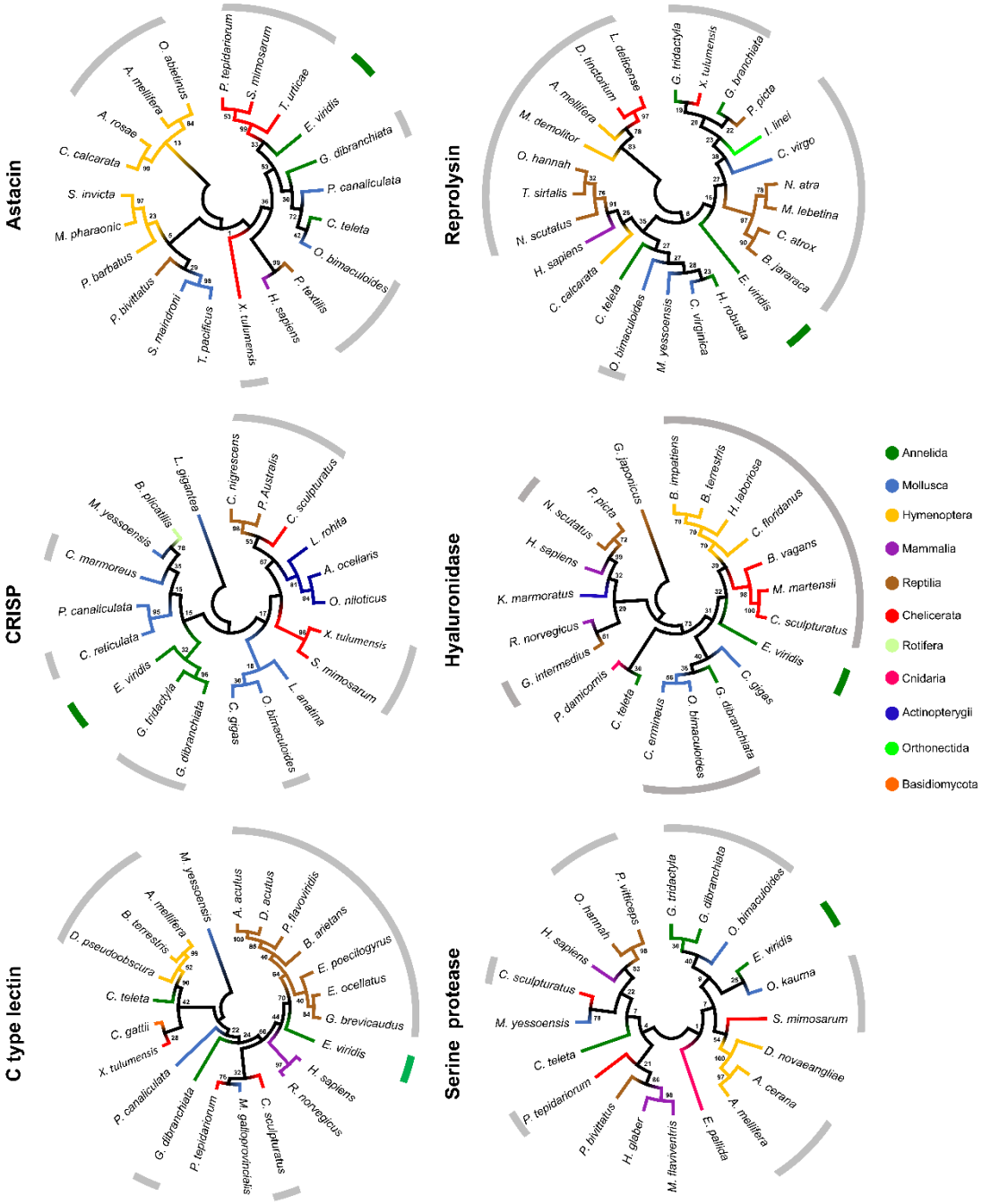


Figure 3.4. Phylogenetic trees of major toxins secreted by *Eulalia*. The model includes the most representative sequences in the worm and the best matches from venomous and non-venomous animals, for comparison (see Table S3.3). The phylogenetic reconstruction was made with MEGA X, with 500 bootstrap pseudoreplicates. Bootstrap support values are given for all nodes and clade names are indicated by coloured branches. Grey bars identify known venomous or toxins-bearing organisms and the green line identifies *E. viridis* homologs.

3.4.5. Multi-gene phylogeny

In face of the inconsistent phylogeny of individual components of *Eulalia*'s "toxungen", the composition of the mixture as a whole was evaluated to infer, not only of how the mixture evolved in *Eulalia*, but also how venoms may have functionally adapted among marine animals. After retrieving homologs for the eight shortlisted toxins in *Eulalia* (Astacin, Reprolysin, CRISP, Hyaluronidase, C type lectin, Serine protease, EGF domain-containing protein and endothelin converting enzyme), eighteen species pertaining to eight major Eumetazoan groups, from Cnidaria to Mammalia (the latter as non-toxin homologs) were included in the multi-trait model (Table S3.6). The species were chosen according to the availability of sequences and broad representativity. Cnidarians (two species) were included due to their lower positioning in the animal tree-of-life. The full phylogenetic model (Fig. 3.5) shows Cnidaria, represented here by the anemone *Exaiptasia pallida* and by the coral *Pocillopora damicornis* (both considered innocuous), forming a clearly distinct clade from all other groups. The Polychaeta, represented by *Eulalia* and *Glycera* (both toxin-secreting) plus the non-harmful *Capitella* (a small annelid whose interest as model organism is growing), and the scallop *Mizuhopecten* form a distinct group from that which irradiated into two clades, one of which holds a clade that allocates all vertebrate species in the tree, venomous or not.

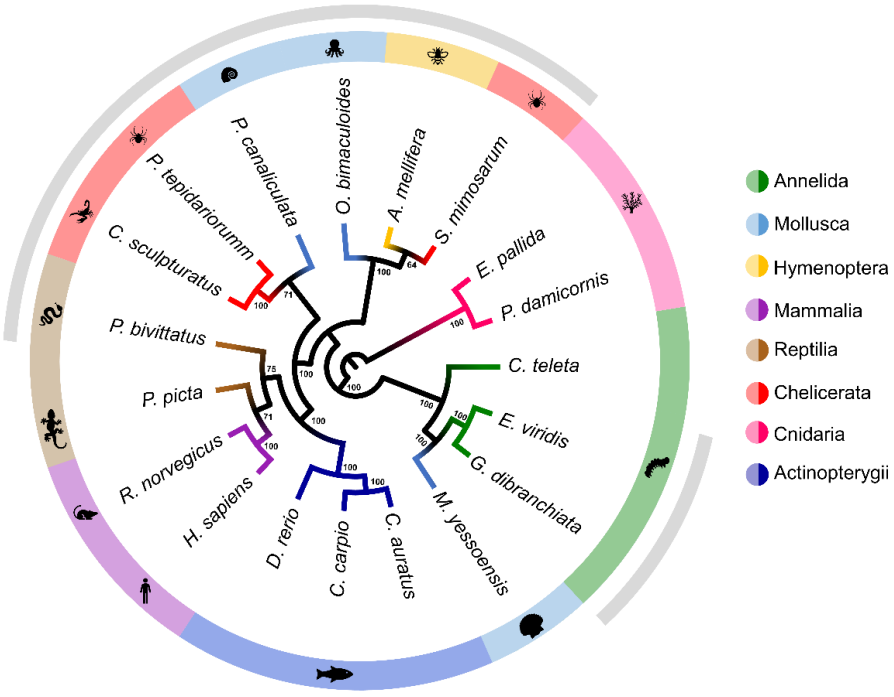


Figure 3.5. Multi-trait phylogenetic tree combining major toxin homologs: CRISP, Hyal, Astacin, Reprolysin, C type lectins C, Serine proteases, Endothelin-like enzymes and EGF domain-containing neurotoxins. The tree was produced using Bayesian inference (20,000 generations, samples recorded every 100 generations) with MrBayes 3 (see Table S3.5 for sequence information). Bootstrap support values are given for all nodes and taxonomic groups names are indicated by coloured branches. Grey bars identify known venomous organisms.

3.5. Discussion

The proboscis of *Eulalia* is an eversible pharynx with multiple roles, particularly sensing and feeding. This feeding organ, which showed the existence of specialised tentacles in its tip that are used for the delivery of noxious substances produced in specialised calix cells lining the interior epithelium of the organ. Consequently, the hypothesis that putative toxin mRNAs could be identified by running RNAseq in the proboscis (pr) and body wall (bw) to pinpoint overexpressed genes in the former was evaluated (Fig. 3.2A). This strategy has been successfully enforced by Ruder et al. (2013) and Modica et al. (2015) to identify putative toxins from the crustacean *Xibalbanus* (*Speleonectes*) *tulumensis* (Remipedia) and the vampire snail *Colubraria reticulata*, respectively.

The strategy revealed to be successful, disclosing the main constituents of *Eulalia* toxungen (recall Fig. 3.2). The overall composition is similar to what is found in other closely related organisms. For instance, the different metallopeptidases in *Eulalia*, were also found in the venom of the Polychaeta *Glycera*, believed to be responsible for the digestion of the extracellular matrix of prey, likely acting as permeabilising agents (von Reumont et al. 2014b). In its turn, Adam CR 2 is a cysteine-rich ≈70 amino acid-long domain, normally paired with the Reprolysin domain, both being typical of metallopeptidase M12B. On the other hand, proteins with EGF-like (epidermal growth factor-like) domains (which are well-conserved domains with Animalia) are ubiquitous among the Eumetazoa and hold several functions (Appella et al. 1988). Proteins with this domain are common in venomous cocktails, including astacin and C type lectin-like. One type of enzymes with conserved domain are hyaluronidases, which are characterised by possessing the long glyco hydro 56-like domains (≈ 300 amino acid residues). Hyaluronidases, whose main function is the degradation of glycosaminoglycans such as hyaluronan and chondroitin, are therefore important assets in the digestion and permeabilization of connective tissue. They are particularly well-described in scorpion and bee venoms (Bordon et al. 2015). In *Eulalia*, a domain of unknown function (DUF), namely DUF 1986 (as identified in Pfam) was one of the most frequent domains in putative toxin-like proteins. This yet uncharacterised 114 amino acid residue domain has been found in proteins of the trypsin-like serine protease superfamily, which is well-represented in venomous cocktails, as for other serine proteases. These proteases can also have CUB domains (100 amino acids) and are well-described in honeybee venoms, for instance (Winningham et al. 2004). These domains (referred to as CUB1 and 2), although ubiquitous and associated to multiple functions, are seemingly restricted to extracellular or transmembrane proteins and are primarily involved in the mediation of protein-protein interactions, including the enhancement of proteolytic activity directed against the extracellular matrix, such the case of procollagen proteinases (Bork and Beckmann 1993; Blanc et al. 2007). Overall, the most important conserved domains in toxin-related

protein secreted by the proboscis of *Eulalia* are cysteine-rich proteins and enzymes able of extensive activity against extracellular matrix, which have a primary role long acknowledged for enzymatic toxins in venomous cocktails.

Cysteine-rich proteins holding CAP domains are known to be key components in animal venoms, from terrestrial, such as snakes, to marine, including the relatively well-studied cephalopods (Fry et al. 2009; Ruder et al. 2013). In *Eulalia*, as well as in other species from different taxa, CRISPs have a well-preserved CAP domain at the N-terminus and several cysteine residues throughout the amino acid sequence (exemplified in Fig. 3.3D). These glycoproteins, which in mammals are secreted in the epididymis (albeit their function remaining unclear), have been persistently found in venoms, being well-described in snakes. Mostly, they act as neuromuscular toxins by blocking calcium channels, but they can also function as potassium channel blockers to prevent muscle contraction (Yamazaki and Morita 2004; Modica et al. 2015). The CRISP proteins in *Eulalia* are being secreted by the proboscis of *Eulalia*. Indeed, these findings are in accordance with the findings in *Eulalia*, where phyllotoxins and their neuromuscular effects were seen in bioassays and observed in the natural habitat with some of the worm's favourite prey, mussels and other Polychaeta. Being *Eulalia* devoid of jaws, it stuns its prey by repeated contact with specialised tentacles at the tip of the proboscis at the base of which, lining the interior of the proboscis, are located calix cells likely responsible for the secretion of putative neurotoxins whose delivery vehicle is mucus.

Besides CRISPs, metalloendopeptidases, hyaluronidases, trypsin-like serine protease and lectin C-like proteins were the most representative proteins overexpressed in the proboscis (recall Fig. 3.3C). Despite the reduced number of works on Polychaeta, these enzymes have already been reported to be well-represented in venomous secretions of a few, particularly *Glycera* and the Amphinomidae fireworms, with the exception of hyaluronidases and trypsin-like serine proteases (von Reumont et al. 2014b; Verdes et al. 2018). Nonetheless, even the latter two are listed as being typically present in venoms from gastropods, cephalopods and snakes (Fry et al. 2008; Ruder et al. 2013; Modica et al. 2015). In the latter, these are reported to contribute to widespread inflammation and enhanced venom diffusion (Kemparaju and Girish 2006). Endothelin-converting enzymes, another group of metalloproteases, were also found to be significantly overexpressed in the proboscis of *Eulalia*, similarly to the venomous cocktails of various animals, from scorpions to snakes (Smith et al. 2016; de Oliveira et al. 2018). As these degrade amyloid beta (whose accumulation is linked to Alzheimer's disease), it is yet another compelling factor towards the study of venom components in biomedical applications (Smith et al. 2016).

Altogether, enzymes targeting the extracellular matrix, combined with the membrane-disrupting and toxin-dissemination role of hyaluronidases described in snakes, bees and even in the intestine of humans as a facilitator of absorption (Bookbinder et al. 2006; Bordon et al. 2012, 2015), assist diffusing *Eulalia*'s cysteine-rich neurotoxin, phyllotoxin. In addition, the anticoagulant activity of C-type lectins (Fry et al. 2003) and the pro-haemorrhagic activity of metalloproteinases from the family M12 (Fry et al. 2008; Sterchi et al. 2009), hinder healing, facilitate infiltration of toxins and favour the extraction of fresh pieces of tissue via suction while the prey is chemically stunned.

The main constituents of *Eulalia* toxungen, had different resemblances with toxins from different organisms. CRISPs from *Eulalia*, for instance, were phylogenetically closer to those of *Glycera*, arguably one of the best-known venomous Polychaeta, and venomous cephalopods, respectively. With respect to CRISPs, a common origin to cysteine-rich neurotoxins may be suspected among the Phyllodocidae Polychaeta. von Reumont et al. (2014b) already discussed the presence of cysteine motif-bearing neurototoxins in *Glycera* and their potential similarity to cnidarian gigantoxins. Gigantoxins are, in their turn, members of epidermal growth factor (EGF) family and their neurotoxic activity has been described in various invertebrates, like *Pomacea* (Gastropoda) and sea anemones (Honma et al. 2003), which explains the resemblance of *Eulalia* EGF domain-containing proteins to a broad clade of organisms, spanning from *Platynereis*, another Phyllodocida, to humans (see Fig. S3.1). Altogether, these data indicate that phyllotoxins can, in fact, refer to a family of multiple cysteine-rich neurotoxins that can be ubiquitous among Phyllodocida, at least, and resulting from an ancient radiation.

Not all the main toxungens sequences from *Eulalia* were closely related with the sequences from the same phylum. C type lectin-like from *Eulalia* was closely related with the ones from Reptilia. These sequences retrieved from snakes consist of calcium-dependent anticoagulant factors, inhibiting both intrinsic and extrinsic coagulation pathways (Zhang et al. 2012). These proteins, now believed to be major components of snake venoms, are also known to hinder coagulation and promote the lysis of blood cells (Nikai et al. 2000; Havt et al. 2005; Earl et al. 2011). Interestingly, a protein similar to C type lectin was found in *Cryptococcus gattii* (Fungi: Basidiomycota), a tropical pathogenic yeast that requires interacting with the immune system of hosts, macrophages specifically (Decote-Ricardo et al. 2019). This sequence was clustered together with bees (curiously, with the innocuous model marine Polychaeta *Capitella teleta* as well), which shows not only the differential role of these proteins in venoms but also potential evolutionary convergence: anti-immune and anti-coagulant. Altogether, these data indicate that, rather than evolving as a whole, these proteins converged to form a cocktail with three major functions: immobilising, permeabilization and tissue disruption plus anti-coagulation. These properties render efficient the extraction, with powerful pharyngeal musculature, of fresh tissue from marine invertebrate prey

that do not possess competent immune systems to clear toxins but are still provided with the efficient healing well-described, e.g., in Polychaeta (Planques et al. 2019).

The close association between *Eulalia* and *Glycera* in most of the trees individualized is no surprise, as they are both predators belonging to Phyllodocida and their respective families (Phyllodocidae and Glyceridae) are seen as sister groups within the Annelida (Struck et al. 2011; Weigert and Bleidorn 2016). These two groups, together with known venomous Polychaeta, namely, fireworms (Amphinomidae) (Verdes et al. 2018), point towards the possibility that the Phyllodocida harbour more toxin-secreting species than could perhaps be anticipated. The phylogeny of these proteins in invertebrates, protostomes in the case, namely Arthropoda and Mollusca, is, however, more elusive. In fact, arachnids and molluscs are dispersed by two distinct clades that do not reflect the known tree-of-life. Nonetheless, among these two, only the snail *Pomacea canaliculata* is considered harmless.

The multi-trait tree showed that, among protostomes, venoms of Polychaeta derive from a more ancient radiation than anticipated. The presence of all eight proteins in both Cnidaria included in the multi-trait model is an indication that these proteins derive from ancient radiations (Sunagar and Moran 2015). Furthermore, the similarities between sequences are astray, with the exception of the positioning of the scallop (a non-venomous organism), from the acknowledged phylogenetic proximity between Annelida and Mollusca, which are the most important groups of the Lophotrochozoa, based on both molecular and morphological data (Peterson and Eernisse 2001). These findings indicate that toxin-bearing secretions from *Eulalia* and possibly Polychaeta in general, became functionally adapted independently of the eumetazoan life history. On the other hand, prominent molluscan toxins, like conotoxins (which are potent neurotoxins), are not only found to have evolved relatively recently (Sunagar and Moran 2015) but were also found to have a relatively distant association to the sequences retrieved from *Eulalia*, especially CRISP- and EGF domain-bearing neurotoxins (albeit within the same wide clade), which indicates convergent adaptation against neuromuscular activity. Indeed, the venom of *Conus* is nowadays considered to be a very refined specialisation, at least with respect to potency, specificity and even the relatively reduced molecular weight of conotoxins. These typically comprised of 10-20 amino acid residues, compared, for instance, with snake CRISPs with about 100 (Osipov et al. 2005; Sunagar and Moran 2015), and *Eulalia* CRISP-like proteins with at least twice as many (recall Table S3.1). These differences did not hinder clustering *Eulalia* and *Conus* CRISPs within the same major clade, as shown in Fig. 3.4, which indicates a common ancestor. In turn, cephalopod and arthropod venom proteins have arisen in distinct moments of evolution, likely by functional convergence into noxious proteins after irradiation from non-toxin homologs, potentially originating toxin and non-toxin paralogs if duplications occurred. In fact, the importance of the retention of gene duplicates that

code for toxins in the evolution of venoms has already been referred (see the review by Wong and Belov 2012), even though the subject represents little more than uncharted ground with respect to marine invertebrates. Still, the resemblance of *Eulalia* toxins to non-toxin homologs (as shown in Fig. 3.4), as well as its positioning with *Capitella* and *Mizuhopecten* in the multi-trait model, further validates this premise. Other authors pointed out, nonetheless, that post-transcriptional mechanisms, such as alternative splicing, or even post-translational cleavage can be responsible for the expression of a wide variety of toxins which, in any case, renders difficult identifying orthologues and evaluating orthologous expression of toxins (Wong and Belov 2012). It must be highlighted that similar reasoning can be applied to vertebrates, which appeared clustered in a single clade in the multi-trait model, combining venomous and non-venomous animals in a monophyletic branch that actually resembles the expected tree-of-life. It is likely that the growing interest in marine invertebrate toxins and the rapid advances in next-generation sequencing methods will roundabout important findings in the near future as new species of venomous animals are being unravelled.

Sunagar & Moran (2015) proposed the "two-speed" theory of venom evolution, according to which more ancient animals invested in the diversification of toxins as means to assure a broad range of prey whereas "younger" species invested in specialisation, potency and reduced energetic costs. Within this perspective, the multi-trait model given in Fig. 3.5, plus the wide scope of putative, high molecular weight, toxins from *Eulalia* (many of which not yet characterised) and the mild potency of its neurotoxins onto mussels, places this organism in the first group. Younger taxa, such as *Conus*, for instance, more likely pertain to the second. Given the ancient radiation of Polychaeta and the wide span of prey of *Eulalia*, it is thus reasonable to assume that the species, and likely other Phyllodocida, responded to a positive selective pressure to diversify its toxin arsenal.

3.6. Conclusions

Animal venoms and similar secretions evolved through multiple episodes of protein recruitment. Venom proteins derive from well-conserved proteins that are present in innocuous animals as well, in a process during which gene duplication likely played a key role. *Eulalia* joins the ranks of venomous marine organisms (or, better, "toxungen"-bearing), showing that toxin-secreting Polychaeta, especially phyllodocids, can be very common and evolved to recruit proteins into venomous mixtures that seemingly favour feeding diversity albeit compromising potency, specificity towards target and metabolic costs. Toxins in *Eulalia* derive from non-toxin homologs without a clear phylogenetic pattern, showing also that, in Polychaeta, venomous mixtures evolved independently from potential ancestors.

In *Eulalia*, the toxins have three major roles: i) hindering the neuromuscular activity of prey; ii) pro-haemorrhagic and anti-clotting, and iii) permeabilization and liquefaction of tissue as a mean to assist neurotoxin diffusion and assist feeding through suction. Indeed, proteinaceous animal venoms hold a common signature, with particular emphasis on blockers of neuromuscular activity and agents that assist their infiltration. Our findings suggest that, rather having a common evolution, animal toxins, particularly from marine invertebrates, evolved by functional convergence after diverging from ancestral non-toxin proteins. Likely, convergence occurred by similar selective pressures towards the same target organisms plus feeding and toxin-delivery mechanisms. This can be inferred, e.g., from the similarities between *Eulalia* astacins and those of spiders shown in Fig. 3.3. Both groups prey on invertebrates and whereas spiders feed by maceration, *Eulalia* must liquify tissue as much as possible prior to extraction by suction, as it lacks jaws. Comparatively, astacins from its sister Phyllodocida, *Glycera*, are more similar to those from *Octopus*, both organisms being equipped with venom-injection structures and preferentially targeting crustaceans either as prey or as attacker, respectively. In their turn, *Eulalia* CRISPs were closely related to those from *Glycera*. These findings may indicate evolutionary divergence, rather than convergence. Altogether, it is clear that the evolution of individual toxins is not straightforward, as proteins have been recruited in different stages of the species' life history.

Finally, it should be highlighted that Phylum Annelida is probably not monophyletic, with the recent addition of the previously considered individual phyla Sipuncula and Echiura (Struck et al. 2007) showing that much of the taxon's phylogeny and systematics remains unresolved. Additional challenges are provided by the lack of genomic annotation, despite the promises of *Capitella*, and considerable intra-phylum genomic variability (Simakov et al. 2013). Associated to increased taxon sampling as a means to tackle phylogenetic uncertainties (Hedtke et al. 2006), venomics can thus provide important clues to the origins of venom proteins, their function and their relation to the animal's milieu, either as part of its mechanisms of predation or defence against predators (Sunagar et al. 2016). In the near future, as new species join the ranks of venomous marine animals, it is expected that the eumetazoan chemical warfare increases its importance in eumetazoan phylogeny as well as their contribution to understand the role of gene duplications and supra-genomic regulation of gene expression in the evolution of species in the world's most vast and diversified ecosystems.

3.7. References

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

Appella E, Weber IT, Blasi F (1988) Structure and function of epidermal growth factor-like regions in proteins. FEBS Lett 231:1–4.

Bartolomaeus T, Purschke G, Hausen H (2005) Polychaete phylogeny based on morphological data – a comparison of current attempts. In: Bartolomaeus T, Purschke G (eds) Morphology, Molecules, Evolution and Phylogeny in Polychaeta and Related Taxa. Springer, pp 341–356

Blanc G, Font B, Eichenberger D, Moreau C, Ricard-Blum S, Hulmes DJS, Moali C (2007) Insights into how CUB domains can exert specific functions while sharing a common fold: Conserved and specific features of the CUB1 domain contribute to the molecular basis of procollagen C-proteinase enhancer-1 activity. J Biol Chem 282:16924–16933.

Bookbinder LH, Hofer A, Haller MF, Zepeda ML, Keller GA, Lim JE, Edgington TS, Shepard HM, Patton JS, Frost GI (2006) A recombinant human enzyme for enhanced interstitial transport of therapeutics. J Control Release 114:230–241.

Bordon KCF, Perino MG, Giglio JR, Arantes EC (2012) Isolation, enzymatic characterization and antiedematogenic activity of the first reported rattlesnake hyaluronidase from *Crotalus durissus* terrificus venom. Biochimie 94:2740–2748.

Bordon KCF, Wiezel GA, Amorim FG, Arantes EC (2015) Arthropod venom hyaluronidases: Biochemical properties and potential applications in medicine and biotechnology. J Venom Anim Toxins Incl Trop Dis 21:1–12.

Bork P, Beckmann G (1993) The CUB Domain. A widespread module in developmentally regulated proteins. J Mol Biol 231:539–545.

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: The evolutionary novelty of venoms. Trends Ecol Evol 28:219–229.

Chang D, Duda TF (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families. Mol Biol Evol 29:2019–2029.

De Oliveira UC, Nishiyama MY, Dos Santos MBV, De Paula Santos-Da-Silva A, De Menezes Chalkidis H, Souza-Imberg A, Candido DM, Yamanouye N, Dorce VAC, Junqueira-de-Azevedo I

de LM (2018) Proteomic endorsed transcriptomic profiles of venom glands from *Tityus obscurus* and *T. serrulatus* scorpions. PLoS One 13:1–23.

Decote-Ricardo D, LaRocque-de-Freitas IF, Rocha JDB, Nascimento DO, Nunes MP, Morrot A, Freire-de-Lima L, Previato JO, Mendonça-Previato L, Freire-de-Lima CG (2019) Immunomodulatory role of capsular polysaccharides constituents of *Cryptococcus neoformans*. Front Med 6:1–8.

Earl STH, Robson J, Trabi M, De Jersey J, Masci PP, Lavin MF (2011) Characterisation of a mannose-binding C-type lectin from *Oxyuranus scutellatus* snake venom. Biochimie 93:519–527.

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. genome informatics 205–211.

Fauchald K, Jumars P (1979) The diet of worms : A study of polychaete feeding guilds. Oceanogr Mar Biol - An Annu Rev 17:193–284.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: The protein families database. Nucleic Acids Res 42:138–141.

Francischetti IMB, Assumpção TCF, Ma D, Li Y, Vicente EC, Uieda W, Ribeiro JMC (2013) The "Vampirome": Transcriptome and proteome analysis of the principal and accessory submaxillary glands of the vampire bat *Desmodus rotundus*, a vector of human rabies. J Proteomics 82:288–319.

Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP (2003) Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J Mol Evol 57:110–129.

Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SFR, Vidal N, Poelmann RE, Norman JA (2008) Evolution of an Arsenal. Mol Cell Proteomics 7:215–246.

Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, Nevalainen TJ, Norman JA, Lewis RJ, Norton RS, Renjifo C, de la Vega RCR (2009) The toxicogenomic multiverse: Convergent recruitment of proteins into animal venoms. Annu Rev Genomics Hum Genet 10:483–511.

Gonçalves C, Costa PM (2020) Histochemical detection of free sthiols in glandular cells and tissues of different marine Polychaeta. Histochem Cell Biol.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512.

Havt A, Toyama MH, Do Nascimento NRF, Toyama DO, Nobre ACL, Martins AMC, Barbosa PSF, Novello JC, Boschero AC, Carneiro EM, Fonteles MC, Monteiro HSA (2005) A new C-type animal lectin isolated from *Bothrops pirajai* is responsible for the snake venom major effects in the isolated kidney. Int J Biochem Cell Biol 37:130–141.

Hedtke SM, Townsend TM, Hillis DM (2006) Resolution of phylogenetic conflict in large data sets by increased taxon sampling. Syst Biol 55:522–529.

Honma T, Nagai H, Nagashima Y, Shiomi K (2003) Molecular cloning of an epidermal growth factor-like toxin and two sodium channel toxins from the sea anemone *Stichodactyla gigantea*. Biochim Biophys Acta - Proteins Proteomics 1652:103–106.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299–314.

Inoué T OH (1988) A new drying method of biological specimens for scanning method microscopy : The t-butyl alcohol freeze-drying method. Arch Histol Cytol 51:53–59.

Kemparaju K, Girish KS (2006) Snake venom hyaluronidase: A therapeutic target. Cell Biochem Funct 24:7–12.

Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG (2009) More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet 25:404–413.

Kini RM, Doley R (2010) Structure, function and evolution of three-finger toxins: Mini proteins with multiple targets. Toxicon 56:855–867.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408.

McHugh D (2000) Molecular phylogeny of the Annelida. Can J Zool 78:1873–1884.

Michel C (1968) Enzymes digestives intestinales d' *Eulalia viridis* (Muller) (Phyllodocidae) et de *Glycera convoluta* (Keferstein) (Glyceridae) Annelides Polychetes Errantes. Ann Histochim 13:67–78.

Michel C (1969) Ultrastructure et histochimie de la cuticule pharyngienne chez *Eulalia viridis* Müller (Annélide Polychète Errant, Phyllodocidae). Z Zellforsch 98:54–73.

Michel C (1970) Rôle physiologique de la trompe chez quatre annélides polychètes appartenant aux genres: *Eulalia*, *Phyllodoce*, *Glycera* et *Notomastus*. Cariers Biol Mar 209–228.

Modica MV, Lombardo F, Franchini P, Oliverio M (2015) The venomous cocktail of the vampire snail *Colubraria reticulata* (Mollusca, Gastropoda). BMC Genomics.

Morgenstern D, King GF (2013) The venom optimization hypothesis revisited. Toxicon 63:120–128.

Nelsen DR, Nisani Z, Cooper AM, Fox GA, Gren ECK, Corbit AG, Hayes WK (2014) Poisons, toxungens, and venoms: Redefining and classifying toxic biological secretions and the organisms that employ them. Biol Rev 89:450–465.

Nikai T, Kato S, Komori Y, Sugihara H (2000) Amino acid sequence and biological properties of the lectin from the venom of *Trimeresurus okinavensis* (Himehabu). Toxicon 38:707–711.

Osipov A V., Levashov MY, Tsetlin VI, Utkin YN (2005) Cobra venom contains a pool of cysteine-rich secretory proteins. Biochem Biophys Res Commun 328:177–182.

Peterson KJ, Eernisse DJ (2001) Animal phylogeny and the ancestry of bilaterians: Inferences from morphology and 18S rDNA gene sequences. Evol Dev 3:170–205.

Planques A, Malem J, Parapar J, Vervoort M, Gazave E (2019) Morphological, cellular and molecular characterization of posterior regeneration in the marine annelid *Platynereis dumerilii*. Dev Biol 445:189–210.

Rodrigo AP, Costa MH, Alves De Matos A, Carrapiço F, Costa PM (2015) A study on the digestive physiology of a marine polychaete (*Eulalia viridis*) through microanatomical changes of epithelia during the digestive cycle. Microsc Microanal 21:91–101.

Ruder T, Sunagar K, Undheim EAB, Ali SA, Wai TC, Low DHW, Jackson TNW, King GF, Antunes A, Fry BG (2013) Molecular phylogeny and evolution of the proteins encoded by coleoid (cuttlefish, octopus, and squid) posterior venom glands. J Mol Evol 76:192–204.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675.

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: An RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol 7:1–14.

Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, Kuo DH, Larsson T, Lv J, Arendt D, Savage R, Osoegawa K, De Jong P, Grimwood J, Chapman JA, Shapiro H, Aerts

A, Otillar RP, Terry AY, Boore JL, Grigoriev I V., Lindberg DR, Seaver EC, Weisblat DA, Putnam NH, Rokhsar DS (2013) Insights into bilaterian evolution from three spiralian genomes. Nature 493:526–531.

Smith AI, Rajapakse NW, Kleifeld O, Lomonte B, Sikanyika NL, Spicer AJ, Hodgson WC, Conroy PJ, Small DH, Kaye DM, Parkington HC, Whisstock JC, Kuruppu S (2016) N-terminal domain of *Bothrops asper* myotoxin II enhances the activity of endothelin converting enzyme-1 and neprilysin. Sci Rep 6:1–11.

States DJ, Gish W (1994) Combined use of sequence similarity and codon bias for coding region identification. J Comput Biol 1:39–50.

Sterchi EE, Stöcker W, Bond JS (2009) Meprins, membrane-bound and secreted astacin metalloproteinases. Mol Aspects Med 29:309–328.

Struck TH, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych KM (2007) Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol Biol 7:1–11.

Struck TH, Paul C, Hill N, Hartmann S, Hösel C, Kube M, Lieb B, Meyer A, Tiedemann R, Purschke G, Bleidorn C (2011) Phylogenomic analyses unravel annelid evolution. Nature 471:95–98.

Sunagar K, Moran Y (2015) The rise and fall of an evolutionary innovation: Contrasting strategies of venom evolutin in ancient and young animals. PLoS Genet 11:1–20.

Sunagar K, Morgenstern D, Reitzel AM, Moran Y (2016) Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. J Proteomics 135:62–72.

Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol.

Thiel D, Hugenschutt M, Meyer H, Paululat A, Quijada-Rodriguez AR, Purschke G, Weihrauch D (2017) Ammonia excretion in the marine polychaete *Eurythoe complanata* (Annelida). J Exp Biol 220:425–436.

Verdes A, Simpson D, Holford M (2018) Are fireworms venomous? Evidence for the convergent evolution of toxin homologs in three species of fireworms (Annelida, Amphinomidae). Genome Biol Evol 10:249–268.

von Reumont BM, Blanke A, Richter S, Alvarez F, Bleidorn C, Jenner RA (2014a) The first venomous crustacean revealed by transcriptomics and functional morphology: Remipede venom

glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. Mol Biol Evol 31:48–58.

von Reumont BM, Campbell LI, Richter S, Hering L, Sykes D, Hetmank J, Jenner RA, Bleidorn C (2014b) A polychaete's powerful punch: Venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. Genome Biol Evol 6:2406–2423.

Vonk FJ, Casewell NR, Henkel C V., Heimberg AM, Jansen HJ, McCleary RJR, Kerkkamp HME, Vos RA, Guerreiro I, Calvete JJ, Wuster W, Woods AE, Logan JM, Harrison RA, Castoe TA, de Koning APJ, Pollock DD, Yandell M, Calderon D, Renjifo C, Currier RB, Salgado D, Pla D, Sanz L, Hyder AS, Ribeiro JMC, Arntzen JW, van den Thillart GEEJM, Boetzer M, Pirovano W, Dirks RP, Spaink HP, Duboule D, McGlinn E, Kini RM, Richardson MK (2013) The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc Natl Acad Sci 110:20651–20656.

Weigert A, Bleidorn C (2016) Current status of annelid phylogeny. Org Divers Evol 16:345–362.

Whitelaw BL, Strugnell JM, Faou P, Da Fonseca RR, Hall NE, Norman M, Finn J, Cooke IR (2016) Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. J Proteome Res 15:3284–3297.

Williams JA, Day M, Heavner JE (2008) Ziconotide: An update and review. Expert Opin Pharmacother 9:1575–1583.

Winningham KM, Fitch CD, Schmidt M, Hoffman DR (2004) Hymenoptera venom protease allergens. J Allergy Clin Immunol 114:928–933.

Wong ESW, Belov K (2012) Venom evolution through gene duplications. Gene 496:1–7.

Yamazaki Y, Morita T (2004) Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44:227–231.

Zhang Y, Xu X, Shen D, Song J, Guo M, Yan X (2012) Anticoagulation factor I, a snaclec (snake C-type lectin) from *Agkistrodon acutus* venom binds to FIX as well as FX: Ca^{2+} induced binding data. Toxicon 59:718–723.

Appendix Chapter 3

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases.

			Pfam					Blastp Toxins		Swissprot			
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
1	16.29	1.05E-11	CAP	PF00188.26	6.80E-18	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	2.94E-28	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	8.17E-26
2	15.70	2.76E-10	Astacin	PF01400.24	2.00E-55	Astacin-like metalloprotease toxin 2	NA	sp|C9D7R2|VMPA2_LOXIN	2.69E-31	Zinc metalloproteinase nas-36	nas-36	sp|D5FM38|NAS36_BRUMA	2.05E-38
3	14.75	6.79E-10	Glyco_hydro_56	PF01630.18	7.50E-66	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	1.52E-46	Hyaluronidase-3	HYAL3	sp|Q6RHW2|HYAL3_PIG	1.46E-47
4	14.63	2.55E-13	ADAM_CR_2	PF17771.1	2.10E-12	Zinc metalloproteinase/disintegrin	NA	sp|O57413|VM2T3_PROMU	1.55E-06	A disintegrin and metalloproteinase with thrombospondin motifs 19	Adamts19	sp|P59509|ATS19_MOUSE	2.64E-14
5	14.38	1.01E-09	Reprolysin_5	PF13688.6	4.30E-12	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	2.94E-14	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	8.16E-12
6	13.24	1.75E-09	ADAM_CR_2	PF17771.1	9.90E-12	Zinc metalloproteinase/disintegrin	NA	sp|O57413|VM2T3_PROMU	4.57E-06	Metalloprotease mig-17	mig-17	sp|Q20930|MIG17_CAEEL	5.16E-13
7	13.11	3.70E-09	WSC	PF01822.19	9.80E-20	Snaclec anticoagulant protein subunit A	NA	sp|Q9DEF9|SLAA_DEIAC	9.15E-20	Snaclec anticoagulant protein subunit A	NA	sp|Q9DEF9|SLAA_DEIAC	2.54E-17
8	12.96	2.47E-09	Glyco_hydro_56	PF01630.18	6.20E-69	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	3.06E-54	Hyaluronidase-3	HYAL3	sp|Q6RHW2|HYAL3_PIG	1.52E-55
9	12.91	8.54E-06	EGF	PF00008.27	2.10E-15	Venom prothrombin activator pseutarin-C catalytic subunit	NA	sp|Q56VR3|FAXC_PSETE	3.98E-08	Fibropellin-1	EGF1	sp|P10079|FBP1_STRPU	4.11E-45
10	12.66	4.14E-09	Glyco_hydro_56	PF01630.18	4.50E-64	Hyaluronidase-2	NA	sp|A3QVP0|HYAL2_BITAR	1.24E-45	Hyaluronidase-2	NA	sp|A3QVP0|HYAL2_BITAR	3.45E-43
11	12.41	4.73E-06	CAP	PF00188.26	3.40E-18	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	3.66E-27	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.02E-24
12	12.36	6.25E-07	Glyco_hydro_56	PF01630.18	9.20E-25	Hyaluronidase-1	NA	sp|P86100|HYAL1_MESMA	3.10E-17	Hyaluronidase-1	NA	sp|P86100|HYAL1_MESMA	8.61E-15
13	12.33	3.65E-10	CAP	PF00188.26	7.20E-19	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	2.47E-36	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	6.86E-34
14	12.32	8.42E-10	Astacin	PF01400.24	2.10E-57	Astacin-like metalloprotease toxin 2	NA	sp|C9D7R2|VMPA2_LOXIN	2.77E-27	Zinc metalloproteinase nas-13	nas-13	sp|Q20191|NAS13_CAEEL	9.88E-43
15	12.25	3.18E-08	Glyco_hydro_56	PF01630.18	4.00E-64	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	8.29E-56	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	2.30E-53
16	12.05	5.51E-06	EGF	PF00008.27	2.10E-15	Venom prothrombin activator pseutarin-C catalytic subunit	NA	sp|Q56VR3|FAXC_PSETE	3.98E-08	Fibropellin-1	EGF1	sp|P10079|FBP1_STRPU	4.11E-45
17	11.59	1.03E-07	Glyco_hydro_56	PF01630.18	2.20E-42	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	2.71E-35	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	7.53E-33
18	11.49	2.98E-07	Trypsin	PF00089.26	7.80E-66	Venom serine protease Bi-VSP	NA	sp|B5U2W0|VSP_BOMIG	2.63E-42	Plasma kallikrein	Klkb1	sp|P26262|KLKB1_MOUSE	3.97E-52
19	11.35	2.09E-05	EGF	PF00008.27	2.10E-14	Venom prothrombin activator pseutarin-C catalytic subunit	NA	sp|Q56VR3|FAXC_PSETE	6.08E-08	Fibropellin-1	EGF1	sp|P10079|FBP1_STRPU	2.35E-45
20	11.30	2.28E-04	CUB	PF00431.20	5.40E-75	Blarina toxin	BTX	sp|Q76B45|BLTX_BLABR	8.76E-42	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	6.85E-63

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases (continued).

			Pfam					Blastp Toxins		Swissprot			
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
21	11.30	2.42E-09	Glyco_hydro_56	PF01630.18	4.90E-56	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	2.34E-38	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	6.49E-36
22	11.25	1.13E-07	Ephrin_rec_like	PF07699.13	2.30E-50	Snaclec alboaggregin-B subunit beta	NA	sp|P81116|SLBB_TRIAB	4.86E-07	Signal peptide, CUB and EGF-like domain-containing protein 3	SCUB3	sp|Q66PY1|SCUB3_MOUSE	3.97E-16
23	11.24	1.44E-07	VWA	PF00092.28	1.50E-153	Venom prothrombin activator vestarin-D2	NA	sp|A6MFK8|FAXD2_DEMVE	2.26E-06	Collagen alpha-6(VI) chain	CO6A6	sp|A6NMZ7|CO6A6_HUMAN	8.17E-72
24	11.07	3.18E-07	EGF	PF00008.27	3.20E-17	Coagulation factor X isoform 2	F10	sp|Q1L658|FA102_PSETE	2.47E-07	Fibropellin-1	EGF1	sp|P10079|FBP1_STRPU	2.54E-45
25	10.94	6.29E-06	Reprolysin_3	PF13582.6	1.40E-07	Zinc metalloproteinase-disintegrin-like atrase-B	NA	sp|D6PXE8|VM3B_NAJAT	8.26E-08	Metalloprotease mig-17	mig-17	sp|Q20930|MIG17_CAEEL	1.18E-13
26	10.89	5.01E-07	ADAM_CR_2	PF17771.1	2.10E-12	Zinc metalloproteinase/disintegrin	NA	sp|O57413|VM2T3_PROMU	1.55E-06	A disintegrin and metalloproteinase with thrombospondin motifs 19	Adamts19	sp|P59509|ATS19_MOUSE	2.64E-14
27	10.83	2.76E-08	Glyco_hydro_56	PF01630.18	4.30E-52	Hyaluronidase	NA	sp|P49371|HUGA_DOLMA	3.48E-36	Hyaluronidase	NA	sp|Q9U6V9|HUGA_POLAN	1.88E-35
28	10.78	5.38E-07	Glyco_hydro_56	PF01630.18	3.30E-70	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	3.48E-58	Hyaluronidase-3	HYAL3	sp|Q6RHW2|HYAL3_PIG	2.68E-58
29	10.75	2.97E-05	CUB	PF00431.20	1.90E-50	Venom protease	NA	sp|Q7M4I3|SP4_BOMPE	1.38E-16	Ovochymase-2	OVCH2	sp|Q90WD8|OVCH2_BUFJA	2.05E-34
30	10.74	3.79E-07	Peptidase_M13_N	PF05649.13	1.90E-100	Neprilysin-1	NA	sp|W4VS99|NEP_TRILK	5.47E-129	Endothelin-converting enzyme 1	ECE1	sp|P97739|ECE1_CAVPO	8.00E-164
31	10.56	6.60E-08	CAP	PF00188.26	1.10E-13	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	3.02E-26	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	8.39E-24
32	10.53	6.53E-06	Reprolysin_5	PF13688.6	4.20E-12	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	8.26E-14	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	2.29E-11
33	10.42	1.78E-09	Ephrin_rec_like	PF07699.13	1.50E-69	Venom prothrombin activator vestarin-D1	NA	sp|A6MFK7|FAXD1_DEMVE	4.54E-06	Neurogenic locus notch homolog protein 1	NOTCH1a	sp|P46530|NOTC1_DANRE	8.05E-71
34	10.38	8.01E-06	Kringle	PF00051.18	2.80E-47	Galactose-specific lectin nattectin	NA	sp|Q66S03|LECG_THANI	1.32E-16	Plasminogen	PLG	sp|P06868|PLMN_BOVIN	2.41E-45
35	10.27	7.18E-05	Peptidase_M13_N	PF05649.13	1.10E-85	Neprilysin-1	NA	sp|W4VS99|NEP_TRILK	8.45E-119	Membrane metallo-endopeptidase-like 1	MMEL1	sp|Q495T6|MMEL1_HUMAN	2.62E-153
36	10.26	2.71E-07	Reprolysin_4	PF13583.6	2.60E-07	Zinc metalloproteinase/disintegrin	NA	sp|P0C6E4|VM2JN_PROJR	8.17E-07	NA	NA	NA	NA
37	10.20	2.93E-06	CAP	PF00188.26	1.40E-12	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	2.79E-24	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	7.74E-22
38	10.13	2.34E-07	Trypsin	PF00089.26	1.30E-66	Venom protease	NA	sp|Q7M4I3|SP4_BOMPE	4.89E-34	Trypsin-1	PRSS1	sp|P07477|TRY1_HUMAN	5.69E-47
39	9.91	1.23E-05	Glyco_hydro_56	PF01630.18	1.60E-66	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	4.15E-49	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	1.15E-46
40	9.89	6.10E-05	TSP_1	PF00090.19	5.30E-40	Kunitz-type serine protease inhibitor carinatin-1	NA	sp|Q6ITB0|VKT1_TROCA	1.51E-15	Spondin-1	Spon1	sp|Q8VCC9|SPON1_MOUSE	2.03E-44
41	9.84	1.69E-04	Lectin_C	PF00059.21	1.30E-10	Galactose-specific lectin nattectin	NA	sp|Q66S03|LECG_THANI	1.51E-11	Low affinity immunoglobulin epsilon Fc receptor	Fcer2	sp|P20693|FCER2_MOUSE	1.92E-09

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases (continued).

			Pfam					Blastp Toxins			Swissprot		
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
42	9.84	1.60E-05	DPPIV_N	PF00930.21	7.60E-91	Venom dipeptidyl peptidase 4	NA	sp|B2D0J4|VDPP4_APIME	1.80E-128	Dipeptidyl aminopeptidase-like protein 6	Dpp6	sp|Q9Z218|DPP6_MOUSE	3.27E-170
43	9.82	2.61E-08	Reprolysin_5	PF13688.6	4.80E-12	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	1.01E-14	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	2.81E-12
44	9.81	1.09E-06	Astacin	PF01400.24	1.90E-58	Astacin-like metalloprotease toxin 1	NA	sp|A0FKN6|VMPA_LOXIN	2.14E-26	Protein SpAN	SPAN	sp|P98068|SPAN_STRPU	2.01E-40
45	9.63	3.69E-05	CUB	PF00431.20	2.70E-91	Blarina toxin	BTX	sp|Q76B45|BLTX_BLABR	2.02E-41	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	8.32E-62
46	9.61	1.14E-04	Reprolysin_5	PF13688.6	4.30E-12	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	2.94E-14	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	8.16E-12
47	9.48	2.04E-04	5_nucleotid_C	PF02872.18	4.80E-41	5'-nucleotidase	NA	sp|F8S0Z7|V5NTD_CROAD	1.09E-145	5'-nucleotidase	NA	sp|P29240|5NTD_DIPOM	4.38E-157
48	9.47	6.75E-09	Ephrin_rec_like	PF07699.13	1.10E-40	Snaclec 6	NA	sp|Q6X5T2|SL6_BITAR	2.37E-10	Neurogenic locus notch homolog protein 2	Notch2	sp|O35516|NOTC2_MOUSE	1.26E-63
49	9.42	2.98E-04	Kazal_2	PF07648.15	7.70E-20	Turripeptide Pal9.2	NA	sp|P0DKT1|TU92_POLAB	1.47E-10	Four-domain proteases inhibitor	NA	sp|P82968|MCPI_MELCP	3.88E-13
50	9.41	2.65E-05	VWA	PF00092.28	1.00E-105	Venom prothrombin activator vestarin-D2	NA	sp|A6MFK8|FAXD2_DEMVE	1.67E-06	Collagen alpha-6(VI) chain	CO6A6	sp|Q8C6K9|CO6A6_MOUSE	1.34E-52
51	9.38	8.53E-05	Trypsin	PF00089.26	8.00E-46	Serine protease sp-Eoc49	NA	sp|B5U6Y3|VSP_ECHOC	1.17E-24	Tryptase	Tpsab1	sp|P27435|TRYB1_RAT	3.19E-41
52	9.35	1.77E-07	TED_complement	PF07678.14	4.70E-58	A.superbus venom factor 1	NA	sp|Q0ZZJ6|VCO31_AUSSU	1.25E-18	CD109 antigen	CD109	sp|Q6YHK3|CD109_HUMAN	6.06E-69
53	9.32	2.06E-05	Reprolysin_5	PF13688.6	1.30E-11	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	2.61E-16	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	7.24E-14
54	9.31	8.27E-08	Reprolysin_3	PF13582.6	1.70E-08	Zinc metalloproteinase-disintegrin-like atrase-B	NA	sp|D6PXE8|VM3B_NAJAT	5.68E-11	Metalloprotease mig-17	mig-17	sp|Q20930|MIG17_CAEEL	5.35E-14
55	9.28	1.04E-08	CUB	PF00431.20	2.40E-98	Serine protease harobin	NA	sp|Q5MCS0|VSPHA_HYDHA	1.40E-34	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	5.13E-54
56	9.16	1.01E-04	TSP_1	PF00090.19	1.80E-40	Kunitz-type serine protease inhibitor carinatin-1	NA	sp|Q6ITB0|VKT1_TROCA	1.56E-15	Spondin-1	Spon1	sp|Q8VCC9|SPON1_MOUSE	8.40E-43
57	9.04	5.46E-05	Ephrin_rec_like	PF07699.13	1.40E-73	Snaclec alboaggregin-B subunit beta	NA	sp|P81116|SLBB_TRIAB	5.86E-07	Signal peptide, CUB and EGF-like domain-containing protein 3	SCUB3	sp|Q66PY1|SCUB3_MOUSE	6.79E-20
58	9.02	4.22E-04	Glyco_hydro_56	PF01630.18	5.00E-69	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	1.53E-54	Hyaluronidase-3	HYAL3	sp|Q6RHW2|HYAL3_PIG	6.73E-56
59	8.98	1.93E-04	Glyco_hydro_56	PF01630.18	6.80E-76	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	1.44E-63	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	4.00E-61
60	8.94	1.17E-05	Reprolysin_5	PF13688.6	5.30E-13	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	2.97E-16	Zinc metalloproteinase/disintegrin	NA	sp|Q5XUW8|VM2IA_BOTIN	8.23E-14
61	8.88	5.89E-05	Lectin_C	PF00059.21	1.20E-14	Galactose-specific lectin nattectin	NA	sp|Q66S03|LECG_THANI	5.59E-06	NA	NA	NA	NA
62	8.80	6.75E-07	Astacin	PF01400.24	1.90E-55	Astacin-like metalloprotease toxin 2	NA	sp|C9D7R2|VMPA2_LOXIN	1.79E-31	Zinc metalloproteinase nas-36	nas-36	sp|D5FM38|NAS36_BRUMA	1.18E-38
63	8.80	9.11E-05	Glyco_hydro_56	PF01630.18	5.60E-76	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	1.08E-62	Hyaluronidase	HYAL	sp|I0CME7|HYAL_CONCN	3.01E-60

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases (continued).

			Pfam					Blastp Toxins				Swissprot	
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
64	8.72	3.53E-06	CAP	PF00188.26	1.10E-17	Cysteine-rich venom protein	NA	sp|Q7YT83|TX31_CONTE	1.02E-32	Cysteine-rich venom protein	NA	sp|Q7YT83|TX31_CONTE	2.82E-30
65	8.71	7.52E-04	Astacin	PF01400.24	5.40E-50	Nematocyst expressed protein 6	NA	sp|K7Z9Q9|VMP_NEMVE	2.06E-20	Zinc metalloproteinase dpy-31	dpy-31	sp|P98060|NAS35_CAEEL	4.09E-85
66	8.58	7.16E-06	CAP	PF00188.26	3.30E-19	Cysteine-rich venom protein	NA	sp|Q7YT83|TX31_CONTE	4.35E-37	Cysteine-rich venom protein	NA	sp|Q7YT83|TX31_CONTE	1.21E-34
67	8.56	1.23E-05	CAP	PF00188.26	1.70E-16	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	4.56E-34	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.27E-31
68	8.53	1.21E-04	CAP	PF00188.26	3.30E-18	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	2.73E-27	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	7.59E-25
69	8.49	1.01E-06	Peptidase_ M13_N	PF05649.13	1.40E-88	Neprilysin-1	NA	sp|W4VS99|NEP_TRILK	1.75E-117	Neprilysin-1	Nep1	sp|Q9W436|NEP_DROME	5.37E-158
70	8.49	2.03E-06	Trypsin	PF00089.26	8.00E-46	Serine protease sp-Eoc49	NA	sp|B5U6Y3|VSP_ECHOC	1.24E-24	Tryptase	Tpsab1	sp|P27435|TRYB1_RAT	3.30E-41
71	8.44	6.86E-08	Trypsin	PF00089.26	1.30E-69	Venom serine protease Bi-VSP	NA	sp|B5U2W0|VSP_BOMIG	5.29E-36	Transmembrane protease serine 3	TMPRSS3	sp|P57727|TMPS3_HUMAN	3.78E-54
72	8.44	5.59E-04	Ldl_recept_a	PF00057.18	3.30E-78	Gilatoxin	NA	sp|P43685|GILX_HELHO	4.45E-27	Atrial natriuretic peptide-converting enzyme	Corin	sp|Q80YN4|CORIN_RAT	1.07E-58
73	8.43	3.40E-04	Glyco_hydro_56	PF01630.18	7.00E-06	Hyaluronidase conohyal-P1	NA	sp|C0HKM3|HYAL_CONPU	1.69E-13	Hyaluronidase-1	HYAL1	sp|Q91ZJ9|HYAL1_MOUSE	8.70E-13
74	8.37	1.91E-04	CAP	PF00188.26	2.30E-18	Cysteine-rich venom protein pseudecin	NA	sp|Q8AVA3|CRVP_PSEPO	7.46E-26	Cysteine-rich secretory protein 2	CRISP2	sp|Q60477|CRIS2_CAVPO	3.54E-24
75	8.35	4.33E-08	Trypsin	PF00089.26	1.40E-66	Venom protease	NA	sp|Q7M4I3|SP4_BOMPE	7.14E-32	Trypsin-1	NA	sp|P00765|TRYP_ASTAS	2.09E-43
76	8.30	1.46E-03	Reprolysin	PF01421.19	9.00E-08	Zinc metalloproteinase-disintegrin-like VMP-III	NA	sp|C9E1R8|VM3V3_CROVV	2.52E-06	Cell surface glycoprotein	csg1	sp|Q5V7F4|CSG_HALMA	1.68E-10
77	8.28	2.87E-04	CUB	PF00431.20	5.90E-10	Snaclec mamushigin subunit alpha	NA	sp|Q9YGG9|SLA_GLOBL	3.17E-07	Cubilin	CUBN	sp|O60494|CUBN_HUMAN	9.11E-08
78	8.23	1.62E-03	Astacin	PF01400.24	5.40E-50	Nematocyst expressed protein 6	NA	sp|K7Z9Q9|VMP_NEMVE	2.06E-20	Zinc metalloproteinase dpy-31	dpy-31	sp|P98060|NAS35_CAEEL	4.09E-85
79	8.21	1.35E-05	EF-hand_7	PF13499.6	9.20E-31	Calglandulin	NA	sp|Q3SB11|CALGL_TROCA	1.35E-10	Calmodulin	NA	sp|P21251|CALM_STIJA	9.04E-20
80	8.09	6.93E-04	Glyco_hydro_56	PF01630.18	7.30E-28	Hyaluronidase-2	NA	sp|A3QVP0|HYAL2_BITAR	1.10E-22	Hyaluronidase-3	HYAL3	sp|O43820|HYAL3_HUMAN	4.49E-21
81	8.00	9.83E-05	Reprolysin_5	PF13688.6	4.20E-12	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	8.26E-14	Zinc metalloproteinase/disintegrin	NA	sp|Q98SP2|VM2J2_BOTJA	2.29E-11
82	7.99	8.46E-05	EGF	PF00008.27	6.50E-17	Coagulation factor X isoform 2	F10	sp|Q1L658|FA102_PSETE	1.31E-07	Neurogenic locus notch homolog protein 1	notch1	sp|P21783|NOTC1_XENLA	8.79E-37
83	7.89	4.96E-04	WSC	PF01822.19	1.70E-64	C-type lectin lectoxin-Lio1	NA	sp|A7X3Z4|LECM1_ERYPO	3.19E-11	WSC domain-containing protein ARB_07867	ARB_07867	sp|D4AUF1|WSCD1_ARTBC	3.95E-18
84	7.86	1.30E-03	Trypsin	PF00089.26	3.70E-58	Venom prothrombin activator porpharin-D	NA	sp|Q58L93|FAXD_PSEPO	2.60E-37	Chymotrypsinogen 2	CTRB1	sp|P04813|CTR2_CANLF	3.28E-45
85	7.86	2.92E-05	Astacin	PF01400.24	1.90E-29	Astacin-like metalloprotease toxin 3	NA	sp|C9D7R3|VMPA3_LOXIN	2.91E-09	Blastula protease 10	BP10	sp|P42674|BP10_PARLI	1.46E-19
86	7.84	6.08E-04	Ig_3	PF13927.6	6.40E-24	Turripeptide Gsp9.3	NA	sp|P0DKT3|TU93_GEMSP	4.77E-06	Follistatin-related protein 5	FSTL5	sp|Q8N475|FSTL5_HUMAN	5.92E-160

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases (continued).

			Pfam					Blastp Toxins		Swissprot			
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
87	7.84	1.26E-04	ADAM_CR_2	PF17771.1	1.70E-14	Zinc metalloproteinase-disintegrin-like VLAIP-B	NA	sp|Q4VM07|VM3VB_MACLB	8.43E-10	Metalloprotease mig-17	mig-17	sp|Q20930|MIG17_CAEEL	2.63E-25
88	7.78	4.87E-04	Glyco_hydro_56	PF01630.18	6.40E-32	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	1.08E-20			sp|Q5REQ1|HYAL3_PONAB	7.85E-22
89	7.73	1.91E-05	Kazal_2	PF07648.15	7.60E-20	Turripeptide Pal9.2	NA	sp|P0DKT1|TU92_POLAB	1.43E-10	Four-domain proteases inhibitor	NA	sp|P82968|MCPI_MELCP	4.40E-13
90	7.63	1.97E-05	CAP	PF00188.26	1.80E-16	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	5.57E-34	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.55E-31
91	7.61	6.25E-04	Lectin_C	PF00059.21	2.10E-10	Galactose-specific lectin nattectin	NA	sp|Q66S03|LECG_THANI	3.75E-06	NA	NA	NA	NA
92	7.59	2.96E-06	Peptidase_M13_N	PF05649.13	1.40E-88	Neprilysin-1	NA	sp|W4VS99|NEP_TRILK	1.79E-117	Neprilysin-1	Nep1	sp|Q9W436|NEP_DROME	1.03E-157
93	7.58	1.92E-04	Peptidase_S10	PF00450.22	1.60E-66	Venom serine carboxypeptidase	NA	sp|C9WMM5|VCP_APIME	1.72E-38	Retinoid-inducible serine carboxypeptidase	Scpep1	sp|Q920A5|RISC_MOUSE	1.98E-81
94	7.58	5.52E-06	CUB	PF00431.20	2.40E-98	Serine protease harobin	NA	sp|Q5MCS0|VSPHA_HYDHA	1.40E-34	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	5.13E-54
95	7.55	7.97E-04	Glyco_hydro_56	PF01630.18	1.80E-34	Hyaluronidase	HYAL	sp|R4J7Z9|HYAL_LOXIN	4.95E-26	Hyaluronidase-3	HYAL3	sp|O43820|HYAL3_HUMAN	1.09E-31
96	7.51	2.82E-04	Trypsin	PF00089.26	7.10E-54	Venom protease	NA	sp|Q7M4I3|SP4_BOMPE	1.13E-32	Transmembrane protease serine 5	Tmprss5	sp|Q9ER04|TMPS5_MOUSE	4.45E-46
97	7.35	8.78E-04	CAP	PF00188.26	1.40E-17	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	2.30E-44	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	6.39E-42
98	7.32	9.94E-04	Glyco_hydro_56	PF01630.18	4.00E-76	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	2.43E-62	Hyaluronidase	HYAL	sp|J3S820|HYAL_CROAD	6.75E-60
99	7.32	1.09E-03	Trypsin	PF00089.26	1.50E-52	Venom protease	NA	sp|Q7M4I3|SP4_BOMPE	1.26E-32	Anionic trypsin-2	Prss2	sp|P00763|TRY2_RAT	3.16E-42
100	7.27	3.69E-06	ADAM_CR_2	PF17771.1	3.00E-11	Zinc metalloproteinase/disintegrin	NA	sp|O57413|VM2T3_PROMU	4.97E-06	Metalloprotease mig-17	mig-17	sp|Q20930|MIG17_CAEEL	1.66E-12
101	7.17	1.79E-03	EGF	PF00008.27	1.10E-37	Venom prothrombin activator oscutarin-C catalytic subunit	NA	sp|Q58L96|FAXC_OXYSU	8.38E-06	Sushi, von Willebrand factor type A, EGF and pentraxin domain-containing protein 1	Svep1	sp|P0C6B8|SVEP1_RAT	1.40E-46
102	7.04	1.27E-04	TSP_1	PF00090.19	2.50E-40	Kunitz-type serine protease inhibitor carinatin-1	NA	sp|Q6ITB0|VKT1_TROCA	1.55E-15	Spondin-1	Spon1	sp|Q8VCC9|SPON1_MOUSE	7.83E-43
103	6.99	2.55E-05	CUB	PF00431.20	2.70E-91	Blarina toxin	BTX	sp|Q76B45|BLTX_BLABR	1.99E-41	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	9.01E-62
104	6.83	2.54E-05	Astacin	PF01400.24	5.70E-52	Nematocyst expressed protein 6	NA	sp|K7Z9Q9|VMP_NEMVE	2.25E-24	Zinc metalloproteinase nas-36	nas-36	sp|Q18206|NAS36_CAEEL	2.35E-64
105	6.79	3.20E-05	WSC	PF01822.19	1.60E-60	C-type lectin lectoxin-Lio1	NA	sp|A7X3Z4|LECM1_ERYPO	2.98E-11	WSC domain-containing protein ARB_07867	ARB_07867	sp|D4AUF1|WSCD1_ARTBC	1.72E-17
106	6.79	1.07E-05	CAP	PF00188.26	3.30E-18	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	4.61E-32	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.28E-29
107	6.77	1.63E-04	Kringle	PF00051.18	3.00E-47	Galactose-specific lectin nattectin	NA	sp|Q66S03|LECG_THANI	3.98E-16	Plasminogen	PLG	sp|P06868|PLMN_BOVIN	2.56E-45
108	6.63	6.12E-04	Trypsin	PF00089.26	8.00E-46	Serine protease sp-Eoc49	NA	sp|B5U6Y3|VSP_ECHOC	1.17E-24	Tryptase	Tpsab1	sp|P27435|TRYB1_RAT	3.19E-41
109	6.60	9.10E-04	EGF	PF00008.27	2.10E-14	Venom prothrombin activator pseutarin-C catalytic subunit	NA	sp|Q56VR3|FAXC_PSETE	6.08E-08	Fibropellin-1	EGF1	sp|P10079|FBP1_STRPU	2.35E-45

Table S3.1 - Hits of proteinaceous substances upregulated in the proboscis, compared to the body wall, with match against Toxins database. Proteins that have significantly higher expression (logFC > 2) in the proboscis and match in Toxins database, were matched against Pfam, NCBInr customised for Toxins and Swissprot databases (continued).

			Pfam					Blastp Toxins		Swissprot			
ID	**logFC**	***Adj.p*-value**	**Domain**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**	**Protein**	**Gene**	**Accession**	***e*-value**
110	6.52	9.52E-07	WSC	PF01822.19	3.00E-25	Venom prothrombin activator notecarin-D1	NA	sp|P82807|FAXD1_NOTSC	4.81E-06	WSC domain-containing protein 2	wscd2	sp|A2BGL3|WSCD2_DANRE	2.14E-17
111	6.14	8.28E-05	CAP	PF00188.26	6.10E-18	Cysteine-rich venom protein pseudecin	NA	sp|Q8AVA3|CRVP_PSEPO	1.26E-25	Cysteine-rich secretory protein 2	Crisp2	sp|P16563|CRIS2_MOUSE	9.23E-24
112	6.12	4.02E-06	Trypsin	PF00089.26	4.30E-67	Blarina toxin	BTX	sp|Q76B45|BLTX_BLABR	2.12E-36	Trypsin-1	NA	sp|P00765|TRYP_ASTAS	3.45E-52
113	6.08	1.27E-05	CAP	PF00188.26	3.40E-18	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	3.66E-27	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.02E-24
114	5.96	1.84E-05	Trypsin	PF00089.26	2.10E-69	Blarina toxin	BTX	sp|Q76B45|BLTX_BLABR	2.74E-42	Ovochymase-1	OVCH1	sp|Q7RTY7|OVCH1_HUMAN	2.94E-64
115	5.55	1.41E-04	Astacin	PF01400.24	5.60E-52	Nematocyst expressed protein 6	NA	sp|K7Z9Q9|VMP_NEMVE	2.36E-24	Zinc metalloproteinase nas-36	nas-36	sp|Q18206|NAS36_CAEEL	2.12E-64
116	5.42	1.15E-03	AbfB	PF05270.13	4.00E-14	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	4.14E-24	Cysteine-rich venom protein Mr30	NA	sp|A1BQQ5|MR30_CONMR	1.15E-21
117	5.28	1.39E-03	Peptidase_S10	PF00450.22	2.40E-67	Venom serine carboxypeptidase	NA	sp|C9WMM5|VCP_APIME	1.38E-39	Retinoid-inducible serine carboxypeptidase	Scpep1	sp|Q920A5|RISC_MOUSE	8.59E-82
118	5.10	5.27E-04	Glyco_hydro_56	PF01630.18	3.30E-70	Hyaluronidase-2	NA	sp|A3QVP0|HYAL2_BITAR	1.58E-57	Hyaluronidase-3	HYAL3	sp|Q6RHW2|HYAL3_PIG	3.27E-57
119	5.09	6.83E-04	EF-hand_7	PF13499.6	9.20E-31	Calglandulin	NA	sp|Q3SB11|CALGL_TROCA	1.35E-10	Calmodulin	NA	sp|P21251|CALM_STIJA	9.04E-20
120	4.24	1.09E-03	EGF	PF00008.27	5.20E-22	Coagulation factor X isoform 2	F10	sp|Q1L658|FA102_PSETE	5.83E-09	Protein jagged-1b	jag1b	sp|Q90Y54|JAG1B_DANRE	1.99E-20

Table S3.2. Translated aminoacid sequences from E. viridis used in phylogenetic analyses.

ID	Name	Sequence
CRISP	Cysteine-rich secretory protein	MVKMYWNDDLAASAQDWADNCEAEHDDSWDRRIATGMWINVGQNMMWGSALPDWTAVVTGWHDEVADYEYGVGSVDGRTVSHYTQVVQDVAVAVGCGYNDCDGMMLRFCNYAYAQGDFSDPYDEDVADDEDCDACPDFC VDHLCDCGAIVCQNGGVLDLETCTCSCEAYASGEQCEIVDCSEPDPSCCLDYYEGKCQEWMNVPGECPHMCGLC*
HYAL	Hyaluronidase	MRCNAESPCLRIYKPPGLYASLHPFKMALLLSACLVLFSHGVGLVAGTPAAPFEAIWNIKTQGCGRKGVNIDLAQYGIKYNSNGNDWKGDVVTIFYNHGAFWPVIPLNGSADNSEGVDGVDYSGWQSGGIPQAGDIADF EQATRDYVNERIPQEDFDGFGCLDFEEWRPLWDINFAKGEMAPYREASINKILKERPDITDEAEIEALAQEEFDVAAREFFEKHIDVAREMRPNALWCYYHFAHCFYGKRGNDYCSQISYDANKKLDWLAKKSDVLFPT IYFMPGTKDENKRGWAFGNCYSTEQWREVSGVPYMPIYPYSKLQIETKTIEDADWFYTEADLVNSIGVVADQGMDGIVFWGNAFLQADETICGKVKTYIDGKVGPYIKKVREFTQRCSDDLCNGQGRCKKNDINLDPTQ VPQTIAESHNYDEATVRAEYSCRCYEGFSGNDCSQ*
CLC	C type lectin-like protein	KCPGNNKLICGGVWNQNVYRTGPAPTEAPTPGGSAPPAPAPGLQVSPATGYLGCYNDTSARILKHQKDFKNTLTPNVCINYCKSKGYAYAAVQYTFQCFCGNDLIESRKSSEDKCTKKCPGNNKLICGGVWNQNVYRTG PAPTEAPTPVCSSGWSESEGHCYKVFKAQETWAKAEKTCSNNGGHLTSVLDQTEMDFVVSLMTNENVNQCWMGMTLSDMAIITWSDGSTVTGYEKWDKASTDPNEYGAGDACGRVYGDKWGDRPCDTKSGFICEKLKQ*
CTL	Serine protease trypsin-like	MSGCLTCGGTLSASSGTIQSLFYSDQNTDYPNLMSCEWTIETNSQQHLQLTFETLDLEVNSQCIYDSVSVYVPGSSTALATVCGNTCSPTVVVPCSYAIVNFTSDYSVTANGFKLTYALVQPGSGPSLECGCSFGQGCL GNTPTPGYTSPSQDTTSTSTTSAPTTSSTTTTTTTTSTPAPTTPTTTMTTTTMGPSTRCSGQYTASSGRILSPNYPQSYGNNENCVNEITVPSGMHIQLTFNHFELEDHSDCAGDVFEVYLGSSNQPAHELCGSHSG TLTLASNHAILILTTDDSVTATGYDVTYTAVSAGTDDDDDTGNGDDEDESGTCGGVLTSSSGAFQSPNFPANYPSGVSCTWTITVASDKLVQITLGTFHIEPNRGCTYDSLMYYDGTISSHSHQHKFCGTLFPTTMVSS GNAVTFVFTSDDFLQFTGFSLSYTSVDPEDETNPDNIACGVVPSSRIVGGSDADAHEYPWIIQLYLDGYFKCGGTLVADSWVITAAHCVDDSSIASFFTVVAGIHSTSGDESSAQTLFVSDVIWPPGYNSHTNANDIAL LKLASAVTWTNQVRPACLSSVDPVEGTSCIISGWGDTELSNPVPSVLQEAVVPILNNEDCNNSYANAGFAIGEDMMCAGYFSGGVDTCQGDSGGPLVCENSGVFTLDGVTSWGNGCADPNYPGVFTRVSRYASWISANS GGAISIGA*
ZnMc Astacin	Zinc metalloproteinase	QEDWQTPEMLFYSVAAALLSLSLVSSERIDQAITRAYENSALRHSPEQVLETGHVLEGLDMVLTADQWREKHHGNGSLSTRNGVADTKYRWTSNTIPYMYYDSFTGAEKKIIEDAMAIWTKHTCIRFRPSTNSDKNFVQ FKTRQEGCMSMLGMVGGPQRLNLQDAGCRMLPVVLHELGHAIGLDHEHARPDRDNYIVVRDQYMKPDLRFAVKKSEGTSNFGLPYDYQSIMHYGQMSFAIDESANGEPTMLTKQPEYQFLIGESQYLTFKDIKIVNVMY KCGGHCGNKQCPDGGFVDENCQCVCEGLPTRSCSEPKPDTATKCHNMNSECDYWGSIGECERNPGYMR
ZnMc Reprolysin	Zinc metalloproteinase	MLLTAVVVLAFLGQARGRVIPTEQTVFIRDSVRGDSLHKRALYGAESLPKLLPVEISSEDGFLSRLVLRANERVPNDIGITFDNEELNSRPVPQLKGVGLYQDPLNGAALLLDHSGPARKLHGTVHHDGSEYELSPMTD DEASDMMTRQGRTLTTFDHPHVLRKVERTGKYEAAGNMTHGVKQSSVSKRAQNDYFEVEYQIWLDKTEYDFWVKAKGGHDAALIQMKKIYLNQYNSIDLSYANIGNDGTGLTIRIFPLDLHILTNAADNRFITDHILGQ NDRQQWDPIRQTIVDAGKMLTDFQKYLYEWGKQNKLPDHAQIYTRYDLGGKTADGYLVGSTQGLAYVGKVCFVDHIALSYWGASLIEDIGQSAWKAAAHELGHNLGAWHDGDKCPGGYIMAPFAGYPDTYDATMRFSSC SIDYMRESLNQPDDGKEKTCVLHTDCSRRITPSGNYQTPGSEISRDELCEWYFGSGATACGRSLSGTQMCQEMECSHGGSCNYGNLASLPVDGMECDTGKECLRGWCQPKNFEPHDVLRNQCRYPPGTWKPNYGGEGPG ETTDAPDCQDDPGSHCEVRETKDCADPSSAMSYWCCATCKKVLG*
EGF	EGF domain-containing protein	QCIASWPASSVDASYSAGDNFVIIAFMNQATGKLQYYCGRYSQHTVNNVPTYTLSLNVQNANTAYCETFQTPNTAGTIQETEFIKLVFTVPTPTTTATTATTATTTASTAATITPTTQPTTTLSPATPTPPPVVCLCQ NGGACPDPNISWLCTCPSSATGTLCETLLPIGTCSAGICQSPGTKVCFDSVPTGFHLCICNPDWSGLSCEIGLSTTTPTTSMTTTPTTTTSTAPLPCECMNGGICYIPTAPRVCSCPVQATGVLCETMMLNGPCSVDAN ACMTPGTLGCMDSAPLGFHLCVCNYGWTGDRCENGPLTTATPVTTSKGVTPTTITTTTPTPSTVSTTVLPCECLNGGTCYVPNAPRQCSCLAHTTGTLCEIILMLPGTCDVADLPCDLAGTIQCVNSVPAGYHHCICHYGW TGYTCDMGPSTTSSVTASTTTVAGETTPTTTTIAPVSQCDCMNGGICYIANAPRQCSCPHETTGVLCEILMLPGTCALGEIPCDSAGTINCVDSVPTGFHFCVCLYGWTGFRCETGTTTTAITSSLPTTPSTAPSSTTT AVVPPCECLNGGVCYVPAAPRQCSCPQQTTGTLCEIRMLPGTCTLGENPCDSTGTVNCIESVPTGYHHCVCLYGWTGYRCETGVATTPTSTVSTATDSGSTAPTTAPGQCDCLNGGICYVPHAPRQCSCPYHTTGTLCE VLLLPGTCDLANNPCNFAGTTNCLDSVPAGYHQCVCQYGWTGYRCEIDVNTATTTPSTSSSTATTTTTSGTALCLCLNGGRCPNPSDPYLCICPDGYTGQVCQTVVPSASCSPNPCTHNGVCVQGTSSYTCVCPTGYGG TLCQEAVISSCNPNPCINGGTCVQGTSGFSCACPSDYDGSLCQNLLAQATTDSRLLLGLGLGLLLLIILILCLIFCCLRCQEKETEVAASQHTEHIEEASTGGKNLIVNAWYPDYTNQWARPTMALPPAEPEYVSPP PPPPPATFFAPPPPSPPPPSDHVFRVQLPAIMARGQPAKVNLHIDDLST*
ECE	Endothelin converting enzyme	MDGAGLHEMTGSYEPIDGEAPPREPRKWTRLEKLLLLLVVICICIVLFLAIFLVVQHFVPFSTPTCQSLDCVEVAGSVIAAMNWSVDPCQDFYQFSCGNWIHTTMAPPSEPKWSQFSVVGERNNAIVKRLLEEPGDGYL SKRSSAVGKAKQFYKACMNTTHIALRGITPLLEFINRLGSWTVTNDATSGMWSAETWDLVESLVTVFPFATPFFGMGVDIDSKNNTIYEIAFDQSGLLFGDPTFYVGNSSERYRAAEVKLFTTFAGLLGSNAEVAKIAN DIWQLEYRLAKIMTPAEDRVDPSATYNPMTLEELQTLTGTWIDLSAYLAAVFGRTIPPSERVMVYDPTYLSKIGEVVTNTSKETLANYLVWSALPPLWMVLPQDFQEAQRVYEQTLFGIDALKEQWKSCVITTDLAFPY VTGGMFVNSVFSAEQRQKVDQILTTIQEQFYARLPKTSWMDETTKINAMEKVQALRRRIGYPNITQSPTKLDKYYHNVSTTDDAFHNYISLHQFKTDLMLAHLGSPVDLSIWDMSAMEVNAYYDPLFNEMVFPAGILQR VFYDSVYPMAMNYGAIGMVMGHELTHGFDNTGRHFDKYGNMVDWWTNQSSDAFQERAQCLQDQYSSYQIGDNMHINGLTTLGENIADNGGVKTAYYAYQNWLKANDYHENRAYLPGLNMTEDQLFFVSFAQVWCLHETE EYMEQAITTDVHTYGPYRVRGVLSNMKEFSTAFSCAPNTPMNPTHKCTVW*
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase	MGKVKVGVNGFGRIGRLVTRAAFNSGKVDIVAINDPFIDLNYMVYMFQYDSTHGKFHGTVKAENGKLVINGNPITIFQERDPSKIKWGDAGAEYVVESTGVFTTMEKAGAHLQGGAKRVIISAPSADAPMFVMGVNHEK YDNSLKIISNASCTTNCLAPLAKVIHDNFGIVEGLMTTVHAITATQKTVDGPSGKLWRDGRGALQNIIPASTGAAKAVGKVIPELNGKLTGMAFRVPTANVSVVDLTCRLEKPAKYDDIKKVVKQASEGPLKGILGYTE HQVVSSDFNSDTHSSTFDAGAGIALNDHFVKLISWYDNEFGYSNRVVDLMAHMASKE*

Table S3.3. Accession numbers and IDs for sequences used in phylogenetic analyses. Sequences are ordered by *e*-value (lowest - highest). See Fig. 3.2 for details.

Protein	ID	Species	Max Score	Total Score	Query Cover	e- value	Per. ident.
Hyaluronidase	XP 017263634.1	*Kryptolebias marmoratus*	214	214	95%	3E-64	30%
	GASB01020098.1	*Glycera dibranchiata*	228	228	89%	3E-64	30%
	XP 027044971.1	*Pocillopora damicornis*	209	209	95%	6E-63	32%
	ALB06111.1	*Gloydius intermedius*	192	192	88%	4E-59	31%
	GCF53585.1	*Paroedura picta*	194	194	94%	6E-57	31%
	XP 026529072.1	*Notechis scutatus*	192	192	88%	9E-56	31%
	AFS33217.1	*Brachypelma vagans*	180	180	88%	1E-55	31%
	EKC31491.1	*Crassostrea gigas*	191	191	94%	1E-55	30%
	XP 015273339.1	*Gekko japonicus*	189	189	90%	3E-55	29%
	XP 011255949.1	*Camponotus floridanus*	177	177	76%	4E-51	30%
	AXL95480.1	*Conus ermineus*	169	169	88%	7E-50	29%
	AAC70915.1	*Homo sapiens*	177	177	94%	1E-49	30%
	ACY69673.1	*Mesobuthus martensii*	166	166	88%	2E-49	29%
	XP 023226974.1	*Centruroides sculpturatus*	163	163	78%	1E-45	31%
	XP 012242409.1	*Bombus impatiens*	162	162	87%	2E-45	27%
	XP 012171522.1	*Bombus terrestris*	162	162	87%	2E-45	27%
	ELT97073.1	*Capitella teleta*	158	158	73%	9E-45	28%
	KOC68535.1	*Habropoda laboriosa*	158	158	79%	3E-44	26%
	NP 001094250.1	*Rattus norvegicus*	157	157	88%	2E-42	27%
	XP 014771328.1	*Octopus bimaculoides*	69	135	50%	1E-12	27%
Lectin	XP 025083615.1	*Pomacea canaliculata*	97	97	51%	5E-22	34%
	XP 021341992.1	*Mizuhopecten yessoensis*	92	92	49%	9E-22	34%
	1Y17_A	*Agkistrodon Acutus*	79	79	44%	5E-21	38%
	Q9DEF9.1	*Deinagkistrodon acutus*	80	80	44%	6E-21	38%
	KGB76302.1	*Cryptococcus gattii*	82	252	74%	1E-21	34%
	A7X3Z4.1	*Erythrolamprus poecilogyrus*	76	76	45%	2E-20	36%
	Q6X5S5.1	*Echis ocellatus*	75	75	44%	3E-19	38%
	Q7LZK5.1	*Bitis arietans*	72	72	45%	2E-18	37%
	Q9YGN4.1	*Gloydius brevicaudus*	73	73	44%	2E-18	36%
	XP_001354039.1	*Drosophila pseudoobscura*	84	120	49%	4E-18	44%
	Q7LZ71.1	*Protobothrops flavoviridis*	72	72	44%	7E-18	38%
	AJQ21525.1	*Mytilus galloprovincialis*	78	330	71%	4E-17	33%
	ELT92401.1	*Capitella teleta*	81	81	35%	5E-17	40%
	AAA41836.1	*Rattus norvegicus*	76	76	47%	2E-15	34%
	AAA35726.1	*Homo sapiens*	75	75	45%	5E-14	35%
	XP 021000063.1	*Parasteatoda tepidariorum*	70	70	48%	3E-13	28%
	XP 397293.3	*Apis mellifera*	70	70	42%	4E-13	35%
	XP 003400499.1	*Bombus terrestris*	68	96	44%	7E-13	36%
	XP 023224705.1	*Centruroides sculpturatus*	59	159	63%	1E-09	28%
	GASB01040406.1	*Glycera dibranchiata*	55	92	80%	2E-05	27%
	DMPC15203781	*Xibalbanus tulumensis*	19	19	24%	4E-02	25%

Table S3.3. Accession numbers and IDs for sequences used in phylogenetic analyses. Sequences are ordered by *e*-value (lowest - highest). See Fig. 3.2 for details (continued).

Protein	ID	Species	Max Score	Total Score	Query Cover	e-value	Per. ident.
Astacin	GASB01021908.1	*Glycera dibranchiata*	241	241	91%	3E-72	40%
	GARU01002669.1	*Glycera tridactyla*					
	XP 025092948.1	*Pomacea canaliculata*	218	218	98%	2E-67	34%
	KOF87573.1	*Octopus bimaculoides*	199	199	71%	1E-59	39%
	ELT93756.1	*Capitella teleta*	176	250	96%	8E-51	33%
	XP 012287980.1	*Orussus abietinus*	155	155	58%	6E-45	40%
	XP 020710873.1	*Athalia rosae*	152	152	67%	2E-43	38%
	XP 012526641.1	*Monomorium pharaonis*	149	149	66%	1E-42	37%
	XP 027055506.1	*Pocillopora damicornis*	152	152	80%	2E-42	33%
	XP 025988894.1	*Solenopsis invicta*	147	147	66%	4E-42	36%
	XP 015781958.1	*Tetranychus urticae*	152	152	85%	5E-42	32%
	XP 011632701.1	*Pogonomyrmex barbatus*	143	143	75%	1E-40	34%
	XP 015905753.1	*Parasteatoda tepidariorum*	147	147	85%	2E-40	32%
	XP 022804319.1	*Stylophora pistillata*	147	147	71%	2E-40	35%
	KFM60032.1	*Stegodyphus mimosarum*	145	145	77%	4E-40	33%
	XP 017890270.1	*Ceratina calcarata*	135	135	57%	3E-37	36%
	XP 003250231.1	*Apis mellifera*	132	132	57%	3E-36	34%
	NP 036597.1	*Homo sapiens*	120	120	57%	9E-29	34%
	XP 026555731.1	*Pseudonaja textilis*	115	115	58%	8E-28	32%
	XP 007429967.1	*Python bivittatus*	113	113	58%	8E-28	35%
	BAC16239.1	*Todarodes pacificus*	104	104	58%	1E-27	34%
	AGF33810.1	*Sepiella maindroni*	89	89	57%	2E-22	32%
	DMPC15181880	*Xibalbanus tulumensis*	64	77	33%	1E-17	39%
CRISP	GART01002098.1	*Glycera tridactyla*	174	174	100%	3E-50	42%
	QBH70090.1	*Colubraria reticulata*	157	211	100%	1E-48	39%
	XP 021366565.1	*Mizuhopecten yessoensis*	151	151	100%	1E-44	39%
	XP 025083739.1	*Pomacea canaliculata*	155	224	100%	2E-44	41%
	GASB01046798.1	*Glycera dibranchiata*	150	150	100%	2E-41	38%
	XP 013418137.1	*Lingula anatina*	142	142	100%	9E-41	37%
	XP 014790560.1	*Octopus bimaculoides*	142	142	100%	1E-40	39%
	XP 009053966.1	*Lottia gigantea*	111	111	67%	4E-31	42%
	EKC42042.1	*Crassostrea gigas*	109	109	100%	2E-28	33%
	ABG36845.1	*Conus marmoreus*	99	99	100%	8E-28	36%
	XP 023138666.1	*Amphiprion ocellaris*	105	105	68%	1E-27	40%
	RXN24579.1	*Labeo rohita*	105	105	68%	1E-27	40%
	RNA25572.1	*Brachionus plicatilis*	106	106	82%	1E-26	36%
	XP 005472920.1	*Oreochromis niloticus*	101	101	68%	6E-26	36%
	ABK63576.1	*Cryptophis nigrescens*	86	86	67%	1E-23	34%
	2DDA_A	*Pseudechis australis*	84	84	67%	6E-23	32%
	XP 023221739.1	*Centruroides sculpturatus*	75	75	65%	4E-16	32%
	KFM73511.1	*Stegodyphus mimosarum*	72	72	63%	5E-15	33%
	DMPC15197540	*Xibalbanus tulumensis*	24	67	25%	8E-04	44%

Table S3.3. Accession numbers and IDs for sequences used in phylogenetic analyses. Sequences are ordered by *e*-value (lowest - highest). See Fig. 3.2 for details (continued).

Protein	ID	Species	Max Score	Total Score	Query Cover	e- value	Per. ident.
Reprolysin	GASB01030463.1	*Glycera dibranchiata*	147	147	82%	9E-35	26%
	OAF71358.1	*Intoshia linei*	116	116	92%	2E-27	26%
	XP 021341707.1	*Mizuhopecten yessoensis*	113	113	96%	6E-26	25%
	XP 022324387.1	*Crassostrea virginica*	109	109	80%	2E-24	25%
	XP 008546259.2	*Microplitis demolitor*	105	105	74%	7E-24	27%
	XP 009023307.1	*Helobdella robusta*	98	98	65%	3E-21	25%
	XP 026669057.1	*Ceratina calcarata*	98	98	82%	4E-21	24%
	RWS29889.1	*Leptotrombidium deliense*	94	94	62%	3E-20	27%
	RWS17879.1	*Dinothrombium tinctorium*	94	94	62%	2E-20	26%
	XP 006562833.2	*Apis mellifera*	95	95	84%	5E-20	26%
	ELT91965.1	*Capitella teleta*	78	78	61%	7E-15	24%
	XP 014768561.1	*Octopus bimaculoides*	75	75	57%	1E-13	25%
	GCF58184.1	*Paroedura picta*	72	72	35%	6E-13	31%
	ETE65286.1	*Ophiophagus hannah*	69	69	39%	4E-12	28%
	D3TTC1.1	*Naja atra*	61	61	70%	5E-12	22%
	XP 013916470.1	*Thamnophis sirtalis*	68	68	38%	9E-12	27%
	XP 026529750.1	*Notechis scutatus*	67	67	39%	3E-11	28%
	EAW62382.1	*Homo sapiens*	66	66	49%	4E-10	24%
	Q7T046.1	*Macrovipera lebetina*	45	45	35%	2E-07	26%
	Q90282.1	*Crotalus atrox*	44	78	41%	1E-08	26%
	P30431.1	*Bothrops jararaca*	42	42	38%	2E-06	25%
	ADZ74134.1	*Conus virgo*	24	24	3%	2E-02	42%
	GARU01002825.1	*Glycera tridactyla*	15	15	3%	2E+00	33%
	DMPC15156935	*Xibalbanus tulumensis*	14	14	1%	6E+00	57%
Serine	XP 020916782.1	*Exaiptasia pallida*	284	409	91%	3E-86	33%
	XP 021351138.1	*Mizuhopecten yessoensis*	283	538	97%	4E-83	29%
	XP 020652847.1	*Pogona vitticeps*	273	551	88%	2E-80	32%
	ETE72753.1	*Ophiophagus hannah*	256	453	88%	2E-73	30%
	XP 011518940.1	*Homo sapiens*	244	450	88%	1E-67	31%
	XP 023218618.1	*Centruroides sculpturatus*	225	269	64%	4E-66	31%
	ELU12131.1	*Capitella teleta*	209	209	34%	1E-62	43%
	XP 394832.1	*Apis mellifera*	198	246	50%	3E-62	43%
	PBC27019.1	*Apis cerana*	199	199	35%	6E-58	42%
	XP 027804158.1	*Marmota flaviventris*	197	197	37%	4E-57	39%
	KZC09429.1	*Dufourea novaeangliae*	195	195	34%	1E-56	39%
	EHB14284.1	*Heterocephalus glaber*	196	196	34%	2E-56	41%
	XP 025029175.1	*Python bivittatus*	192	192	33%	3E-55	41%
	KFM57227.1	*Stegodyphus mimosarum*	195	303	69%	4E-53	30%
	XP 015908189.2	*Parasteatoda tepidariorum*	184	184	34%	2E-51	38%
	GASB01005460.1	*Glycera dibranchiata*	178	178	33%	1E-47	42%
	KOF65524.1	*Octopus bimaculoides*	158	348	64%	1E-39	28%
	ACJ22649.1	*Octopus kaurna*	130	130	33%	3E-37	36%
	GARU01001067.1	*Glycera tridactyla*	152	152	33%	3E-35	36%

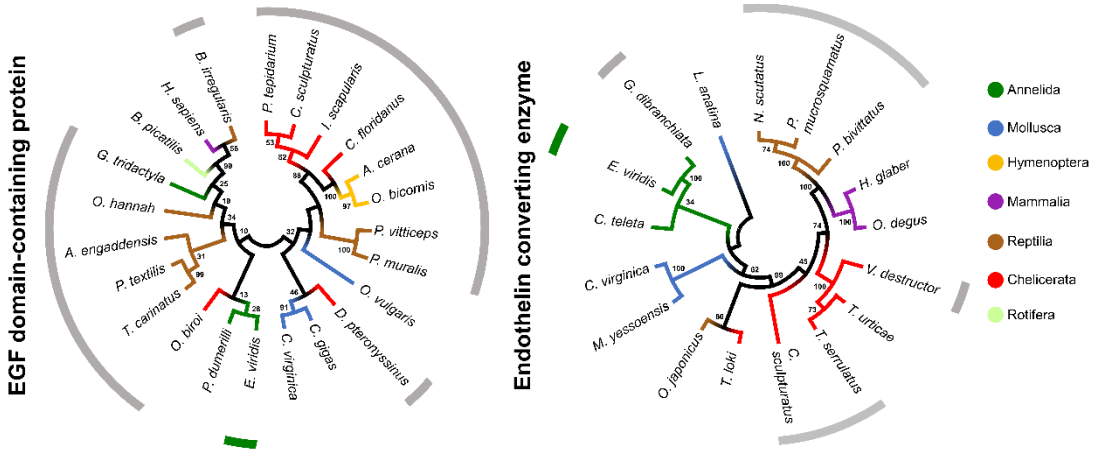


Figure S3.1. Phylogenetic trees of EGF domain-containing protein and endothelin converting enzyme. These proteins were shortlisted from ORFs with higher relative expression in the proboscis and best-matched against the Toxins database. Phylogenetic reconstruction was made with MEGA X, with 500 bootstrap pseudoreplicates. Bootstrap support values are given for all nodes and clade names are indicated by coloured branches. Grey bars identify known venomous or toxins-bearing organisms and green line identifies *E. viridis* homologs.

Table S3.4. Accession numbers and IDs of Endothelin converting enzyme and EGF domain-containing protein.

Protein	Accession	Species	Max Score	Total Score	Query Cover	e-value	Per. ident.
Endothelin converting enzyme	XP_021353180.1	*Mizuhopecten yessoensis*	552	552	99%	0E+00	38%
	XP_013417588.1	*Lingula anatina*	628	628	98%	0E+00	42%
	XP_022312294.1	*Crassostrea virginica*	575	575	95%	0E+00	41%
	ELT90095.1	*Capitella teleta*	595	595	93%	0E+00	42%
	GASB01026619.1	*Glycera dibranchiata*	701	701	91%	0E+00	49%
	EHB08923.1	*Heterocephalus glaber*	496	496	97%	1E-165	37%
	XP_023578223.1	*Octodon degus*	496	496	97%	1E-165	37%
	XP_015788241.1	*Tetranychus urticae*	494	494	97%	3E-164	36%
	AMO02504.1	*Tityus serrulatus*	484	484	91%	2E-163	38%
	XP_022666791.1	*Varroa destructor*	481	481	94%	3E-159	37%
	XP_026543448.1	*Notechis scutatus*	474	474	95%	4E-157	36%
	XP_007441074.1	*Python bivittatus*	473	473	95%	9E-157	36%
	XP_015679750.1	*Protobothrops mucrosquamatus*	473	473	96%	1E-156	36%
	XP_023240703.1	*Centruroides sculpturatus*	456	456	89%	3E-151	36%
	BAT62444.1	*Osedax japonicus*	286	286	86%	1E-89	29%
EGF domain-containing protein	XP_023215803.1	*Centruroides sculpturatus*	202	836	76%	1E-52	27%
	XP_015911895.1	*Parasteatoda tepidariorum*	198	723	75%	1E-51	28%
	XP_022346053.1	*Crassostrea virginica*	187	2332	74%	1E-47	28%
	XP_011445318.1	*Crassostrea gigas*	182	1265	73%	3E-46	27%
	XP_028571703.1	*Podarcis muralis*	173	759	75%	3E-43	27%
	XP_020650495.1	*Pogona vitticeps*	172	893	73%	5E-43	27%
	NP_060087.3	*Homo sapiens*	174	1136	73%	5E-43	28%
	ETE70770.1	*Ophiophagus hannah*	170	977	73%	7E-43	27%
	XP_002403554.1	*Ixodes scapularis*	155	1074	77%	8E-42	26%
	XP_025268967.1	*Camponotus floridanus*	165	731	72%	6E-41	27%
	AWT22989.1	*Octopus vulgaris*	164	666	74%	1E-40	26%
	XP_027195277.1	*Dermatophagoides pteronyssinus*	159	486	72%	1E-39	28%
	XP_028522734.1	*Apis cerana*	160	842	72%	2E-39	27%
	XP_026828708.1	*Ooceraea biroi*	158	839	78%	7E-39	26%
	XP_029033765.1	*Osmia bicornis bicornis*	156	893	73%	2E-38	27%
	RNA00791.1	*Brachionus plicatilis*	145	550	81%	4E-35	25%
	AKP06505.1	*Platynereis dumerilii*	126	343	73%	6E-31	25%
	Q4QXT9.1	*Tropidechis carinatus*	48.5	73.9	11%	9E-09	32%
	Q56VR3.2	*Pseudonaja textilis*	50.1	84.7	9%	8E-06	32%
	AQX92140.1	*Glycera tridactyla*	25	25	4%	4E-01	37%
	JAG68058.1	*Boiga irregularis*	17.3	50.4	4%	4.7	28%
	P13208.2	*Atractaspis engaddensis*	15	324	0.46	15	26%

Table S3.5. Accession numbers of sequences employed in multi-trait phylogenetics. The model combined CRISP, Hyaluronidase, Astacin, Reprolysin, C type lectin, Serine, Endothelin converting enzyme and EGF domain-containing homologs from representative species of key Eumetazoan phyla (venomous and non-venomous organisms). Refer to Fig. 3.4 for the full phylogenetic model.

Species	Phylum	CRISP	Hyaluronidase	C type lectin	Serine protease	Astacin	Reprolysin	EGF domain-containing protein	Endothelin converting enzyme
Glycera dibranchiata	Annelida	GASB01046798.1	GASB01020098.1	GASB01040405.1	GASB01005460.1	GASB01021908.1	GASB01030463.1	GASB01017903.1	GASB01026619.1
Capitella teleta	Annelida	ELU09214.1	ELT97073.1	ELT92401.1	ELU12131.1	ELT93756.1	ELT98077.1	ELU11161.1	ELT90095.1
Parasteatoda tepidariorum	Chelicerata	XP_015904653.1	XP_015922490.1	XP_021000063.1	XP_015908189.2	XP_015905753.1	XP_015921582.1	XP_015911895.1	XP_015927713.1
Stegodyphus mimosarum	Chelicerata	KFM73511.1	KFM61642.1	KFM65020.1	KFM57227.1	KFM60032.1	KFM61983.1	KFM70600.1	KFM77382.1
Centruroides sculpturatus	Chelicerata	XP_023221739.1	XP_023226974.1	XP_023222872.1	XP_023218618.1	XP_023222430.1	XP_023240235.1	XP_023215803.1	XP_023211934.1
Exaiptasia pallida	Cnidaria	XP_020904389.1	XP_020913987.1	XP_028518136.1	XP_020916782.1	XP_020907018.2	XP_020916248.1	XP_020899053.1	XP_020899185.1
Pocillopora damicornis	Cnidaria	XP_027048055.1	XP_027044971.1	RMX49478.1	RMX51212.1	XP_027055506.1	XP_027050425.1	XP_027051833.1	XP_027046906.1
Apis mellifera	Hymenoptera	XP_016766808.1	NP_001011619.1	XP_397293.3	XP_016769420.1	XP_003250231.1	XP_006562833.2	XP_026295156.1	XP_392043.3
Homo sapiens	Mammalia	AAI01540.1	AAH05896.1	AAA35726.1	XP_011518940.1	AAH13871.1	EAW62382.1	NP_060087.3	NP_001106819.1
Rattus norvegicus	Mammalia	AAD48090.1	NP_997482.2	1TDQ_B	EDL83068.1	EDL94198.1	NP_001101903.1	NP_001099191.1	BAA06152.1
Octopus bimaculoides	Mollusca	XP_014790560.1	XP_014771328.1	KOF89271.1	XP_014787085.1	KOF87573.1	KOF93529.1	XP_014781344.1	XP_014790282.1
Mizuhopecten yessoensis	Mollusca	XP_021366565.1	XP_021353240.1	XP_021341992.1	XP_021351138.1	XP_021365700.1	XP_021341707.1	XP_021342460.1	XP_021353180.1
Pomacea canaliculata	Mollusca	XP_025083739.1	XP_025098740.1	XP_025083615.1	XP_025114564.1	XP_025092025.1	XP_025104189.1	PVD22447.1	XP_025083320.1
Python bivittatus	Reptilia	XP_025028116.1	XP_007420722.1	XP_007422481.1	XP_025032198.1	XP_007434871.1	XP_007422351.1	XP_007426224.1	XP_007441074.1
Paroedura picta	Reptilia	GCF55372.1	GCF53585.1	GCF42045.1	GCF56440.1	GCF44322.1	GCF58184.1	GCF55798.1	GCF49036.1
Danio rerio	Teleostei	XP_021322709.1	NP_001121765.1	XP_002660413.1	NP_001099071.1	AAH95288.1	XP_700384.3	XP_005168706.1	NP_001071260.1
Cyprinus carpio	Teleostei	XP_018974179.1	KTF93957.1	KTG02892.1	KTF77069.1	XP_018918842.1	KTG21912.1	AOW71522.1	KTF75647.1
Carassius auratus	Teleostei	XP_026109449.1	XP_026115374.1	XP_026061246.1	XP_026069014.1	XP_026090900.1	XP_026090266.1	XP_026067988.1	XP_026131351.1

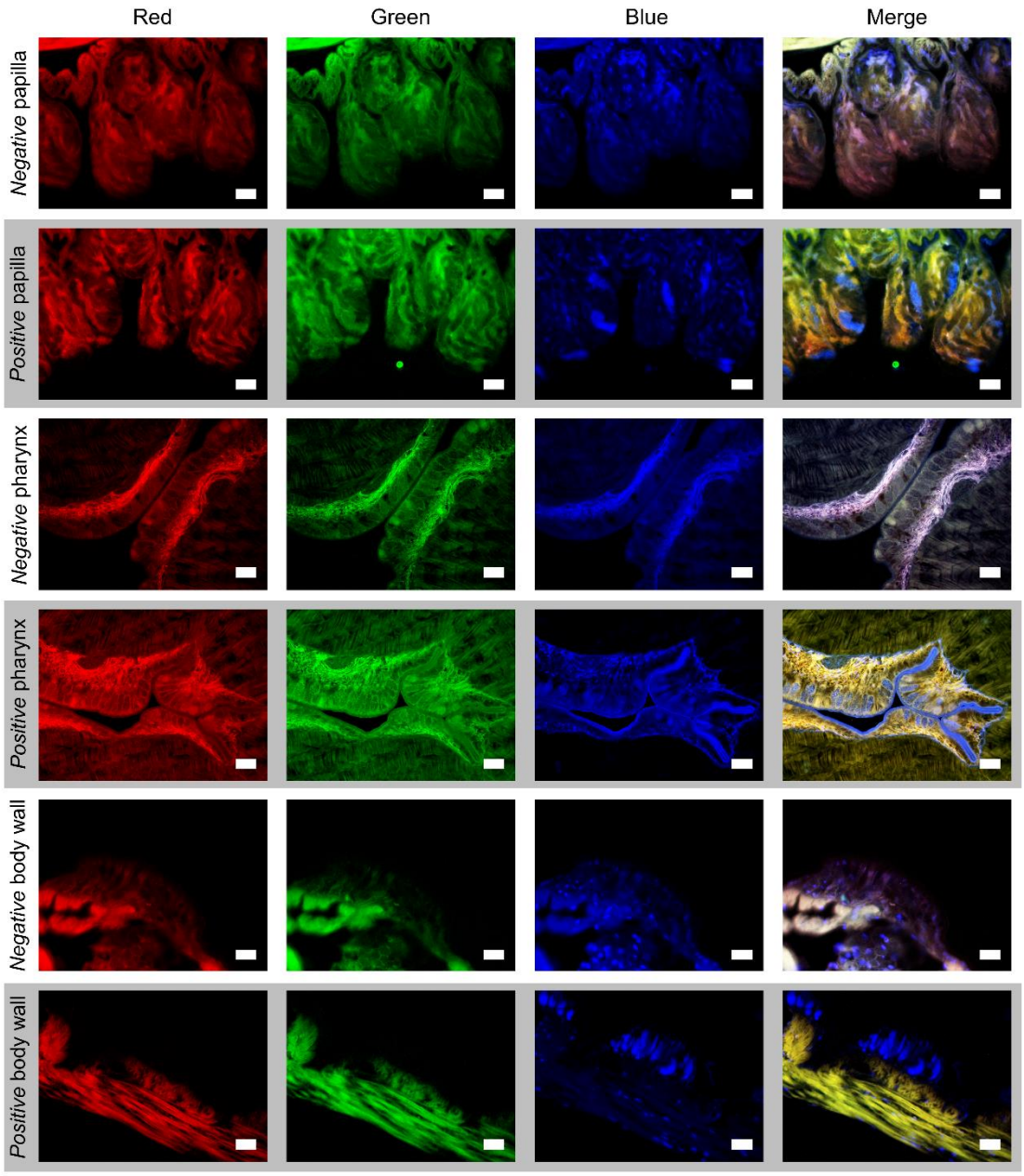


Figure S3.2. Channel-split fluorescent-labelled of thiols in various tissues of Eulalia as a marker for CRISPs. The histological (paraffin) sections were stained with a fluorescent histochemical marker for thiols after treatment with a reducing agent (DTT), producing a bluish labelling. Nuclear counterstaining was produced with DAPI. Negative controls were obtained by treating with DTT but no marker. Here are shown the Individual RGB channels plus respective composite image. Papilla and body wall samples were fixated in glutaraldehyde and pharynx samples in Bouin's solution. Papilla and pharynx present in the proboscis stained positive for thiol groups (granular blue cells) and mucocytes (due to the presence of sulphated mucins). Conversely, sections from the sensorial papillae and body wall stained positive only for mucocytes, which explains the differential expression of CRISPs (higher in the proboscis). Scale bars: 25 µm.

CHAPTER 4 - Ecological role of toxins in *Eulalia viridis*

This chapter has been published in the following paper:

Cuevas N, Martins M, Rodrigo AP, Martins C, Costa PM (2018) Explorations on the ecological role of toxin secretion and delivery in jawless predatory Polychaeta. Sci Rep 8:7635.

4.1. Abstract

Motivated by biotechnological prospects, there is increasing evidence that we may just be scraping the tip of the iceberg of poisonous marine invertebrates, among which the Polychaeta are promising candidates for bioprospecting. Here we show that an inconspicuous phyllodocid uses toxins in its uncanny feeding strategy. The worm, a jawless active predator characterised by its bright green colour, preys on larger invertebrates (including conspecifics) by extracting tissue portions with its powerful proboscis through suction. The animal is even able to penetrate through the valves and plates of live molluscs and barnacles. Observations *in situ* and a series of experiments demonstrated that the worm compensates its simple anatomy with secretion of a novel toxin, or mixture of toxins, referred to by us as "phyllotoxins". These are carried by mucus and delivered via repeated contact with the tip of the proboscis until the prey is relaxed or immobilised (reversibly). Proteolytic action permeabilises material to toxins and softens tissue to enable extraction by suction. The findings show that toxins are a major ecological trait and therefore play a key role in evolutionary success and diversification of Polychaeta, demonstrating also that understanding adaptative features may become the best showcase for novel animal toxins.

4.2. Introduction

Chemical warfare is one of the most cost-effective strategies adopted by animals to defend against parasites, predators or to become predators themselves (Casewell et al. 2013). Biological toxins (biotoxins) can thus play an important ecological role and be regarded as adaptative features. Given the vastness of oceans and the ancient radiation of marine life, it is not surprising that the diversity of toxins may correlate with their immense biodiversity. Indeed, there has been a big effort to describe and catalogue novel toxins from marine eumetazoans. This enterprise is mostly motivated by biotechnological implications, which usually implies the very challenging endeavour to characterise the complex mixtures of proteins, small peptides and salts that comprise poisons and venoms (Casewell et al. 2013). Indeed, and despite much early promise, biotechnological applications of marine toxins seem disappointing (Fusetani and Kem 2009; Burgess 2012). At least in part, this issue results from the failure to understand the combined effect of the various elements in these mixtures. To this are added the difficulties in characterising toxins at the molecular level that derive from poor genomic annotation, a problem that hinders marine animal research in general. The present work is thus set upon the hypothesis that understanding the ecological role of biotoxins is the first step to understand the function and evolution of marine animal chemical weaponry.

Recent descriptions of novel toxins from Polychaeta and even the very first crustacean venom indicate that we may be merely facing a small part of the vast diversity of poisonous marine invertebrates (van der Ham and Felgenhauer 2007; Von Reumont et al. 2013). In line with the trend to find novel biotoxins fortuitously, we recently came across a novel unknown proteinaceous toxin (≈ 40 kDa) secreted by a hitherto inconspicuous annelid, *Eulalia viridis* (Phyllodocidae), whose mucous secretions had a strong inhibitory reaction against the marine bacterium *Vibrio fischeri* (Rodrigo et al. 2014). The species inhabits rocky intertidal shores and has a clear preference for mussel beds (Morton 2011). One of the most interesting aspects of the animal's ecology is that it is an active predator of much larger prey, particularly live mussels, barnacles and even Polychaeta (including other *E. viridis*). However, the species, as other members of the order, is devoid of jaws, relying solely on its powerful muscular proboscis for feeding (Tzetlin and Purschke 2005). We conjectured, then, that the worm uses toxins, referred to by us as phyllotoxins, as part of its preying strategy, enabling it to extract a portion of the prey's soft body via suction.

The rapidly-expanding literature on marine animal toxins is making use of high-content screening molecular approaches to ascertain the nature of the proteinaceous materials in toxin mixtures (Sunagar et al. 2016). Still, only in a few instances' homology-based analyses have been able to produce convincing clues on the role of these secretions in interspecific interactions. Among these, Whitelaw et al. (2016) related the presence of chitinases in the scarcely known toxin mixtures from some cephalopods that predate on crustaceans ("cephalotoxins"). The authors argued that, considering the elevated contents of chitin in arthropod tissue (a glucose-derived polymer with analogous function to vertebrate collagen), this enzyme increases permeability to facilitate infiltration of neurotoxins. Understanding the composition of toxin mixtures is a challenge beyond the problem of genomic annotation, as it may be a function of environmental parameters such as diet (Gao et al. 2011; Casewell et al. 2013; von Reumont et al. 2014). It has been discovered that some animals can even produce different venoms for predation and defence, such as some cone snails and a few arthropods, like scorpions, for instance (Casewell et al. 2013; Dutertre et al. 2014; Nisani and Hayes 2015). In any case, adaptative traits offer a solid ground to steer research and such is the motto of the present work. By combining ecological and toxicological endpoints, we aim at understanding the ecological role of toxins in the very particular behaviour of *E. viridis*, a discreet but resourceful organism that revealed itself to be a fierce predator.

4.3. Material and methods

4.3.1. Mucotoxin harvesting and characterisation

Adult worms (*c.a.* 5 – 10 cm length) were sampled by hand ($n \approx 300$) during low tide at rocky beach in W Portugal (38°41'42" N; 09°21'36"W) and maintained in the laboratory in a microcosm environment. Crude mucous secretions were harvested by gentle mechanical stimulation using blunt-tipped plastic tweezers. The mucus samples were pooled and centrifuged to remove solid deposits (5000 × *g*, 4 ºC, 5 min) and stored at -80 ºC until further analyses. To verify the proteinaceous nature of toxins, an aliquot of the mucus was subjected to ultrafiltration with using 3kDa Amicon spin column filters (Merck Millipore) after preliminary filtration through cellulose acetate filter (0.22 µm). Dulbecco's phosphate-buffered saline (PBS), pH 7.4 was employed vehicle to assess reactivity in physiologically-compatible media. The proteinaceous nature and toxicity of purified and crude secretions was assessed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and the standardised Microtox test, respectively (Rodrigo et al. 2014). Total protein, determined with a NanoDrop 2000 apparatus (Thermo Fisher), was used to indicate concentration of toxins.

4.3.2. Experimental assessment of toxin mode-of-action in vivo

Mussels, *Mytilus* sp. (4.5 – 5.5 cm shell length), were hand-collected between February and April 2017 from a clean rocky intertidal area in W Portugal as well. Crude secretions for testing were diluted in filtered and autoclaved seawater. The doses (concentrations) are designated as Concen. 3 (highest – 1 mg. mL^{-1} total protein, determined as above), Concen. 2 (0.5 mg. mL^{-1}) and Concen. 1 (lowest – 0.25 mg. mL^{-1}). The animals were subjected to intravalvar injection once with each dose (0.6 mL). Controls (seawater only) were included, as well as blanks (no injection). Several independent assays were conducted in order to address multiple endpoints. The animals were collected at several time-points between 5 and 60 min after injection, depending on endpoint. The frequency of valve movements (opening or closure) was quantified from video analysis. Six biological replicates were analysed for behavioural endpoints (n = 6). Latency time was determined from stimulus (tapping) to valve re-opening. Alterations to physiology (n = 3) were determined from oxygen consumption using a Multiline 340i/SET electrode (WTW, Germany), and microalga cell (specially cultured *Tetraselmis suecica*) removal using a Multisizer 3 Counter (Beckman Coulter). Acetylcholine esterase (AChE) activity was determined in the adductor muscle of mussels according to the method developed by Ellman et al. (1961), modified for microplates. These results are provided as nmol hydrolysed substrate min^{-1}. mg. $protein^{-1}$. Histopathological alterations were determined in whole soft-body of mussels (n = 6), fixated in Davidson's solution and embedded in Paraplast. Sections (5 µm) were stained with haematoxylin and eosin (H&E) and the tetrachrome procedure described by Costa and Costa (2012). Semi-quantitative histopathological condition indexes were obtained according to the method described by Costa et al. (2013), adapted to mussels by Cuevas et al. (2015). Briefly, the procedure is based on the product between dissemination (from

0 – absent to 6 – diffuse) and biological significance (1 – lowest severity to 3 – highest). The histopathological alterations selected for the estimation of indexes (following preliminary observations) were lipofuscin aggregates, haemocytic infiltration and diffusion of brown cells, all of which have a biological significance of 1 (Costa et al. 2013; Cuevas et al. 2015). Accuracy was checked by blind reviews. The methodology produces an integrated histopathological condition index that ranges between 1 (maximum predicted histopathological condition) and 0. Indexes were obtained per individual and integrated measurements from visceral mass, gills, nephridium and gonad.

4.3.3. Toxicopathological effects under natural conditions

In order to simulate direct contact with the toxic secretions, freshly-collected whole mucus was applied directly onto soft tissues live mussels or freshly excised organs (*ex vivo* assessment) (n = 2). Tissue samples were then processed for histopathological analyses, to which was added analyses of prey collected from the natural habitat after being preyed by *E. viridis*.

4.3.4. Statistical analyses

The normality and homoscedasticity of data were analysed through Kolmogorov–Smirnov's and Levene's tests, respectively. Considering the invalidation of at least one of the assumptions, non-parametric statistics were employed, namely the Kruskal-Wallis ANOVA-by-ranks *H* was applied for multiple comparisons (testing of effects) and the Mann-Whitney *U*-test for comparisons between experimental treatments and controls. Cluster analyses were also carried out for grouping tested variables. Statistics were performed with R 3.3.x (Ihaka and Gentleman 1996) and the significance level was set at 0.05 for all analyses.

4.4. Results

4.4.1. Eulalia viridis feeding behaviour

In its preferential habitat, *i.e.* rocky intertidal mussel beds, in western Portugal, *E. viridis* was observed to be an opportunistic but active predator of a wide range of other invertebrates, with a preference for live mussels, barnacles and even other Polychaeta. The worm used the proboscis for sensing the environment during foraging and inserted it between the valves of mussels and plates of barnacles (Fig. 4.1). A clearer picture of the use of this organ was seen when attacking other annelids. The proboscis enters in contact with the target organism and its repeats the contact via

extension of the proboscis, accompanied by copious secretion of mucus. The prey tries to escape while becoming immobilised until shrivelling. At this point, *E. viridis* attempts to extract a piece of its flesh through the wound created at the contact area through suction movements with its proboscis.

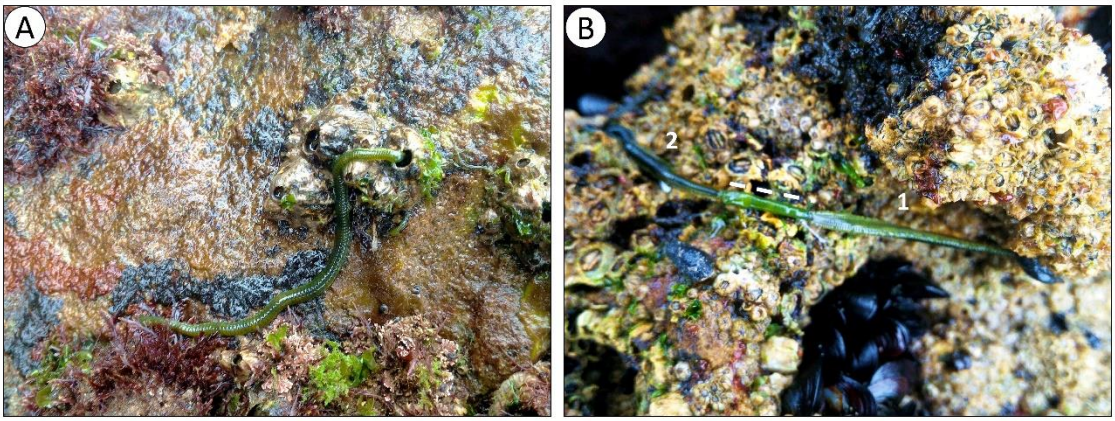


Figure 4.1. Feeding behaviour of *E. viridis* in its natural environment. (A) A worm inserting the proboscis between the plates of a barnacle. (B) A worm (1) preying on another individual from the same species (2) using its proboscis (dashed line).

4.4.2. Toxin reactivity

The mucus was found to be moderately viscous, little adhesive and rapidly dissolved in natural seawater and other aqueous media. Purification by ultrafiltration isolated molecules larger than 3 kDa from secretions, yielding a multi-protein/peptide signature similar to that of crude mucus after removal of salts and other small constituents from secretions, with major bands between c.a. 6 and 40 kDa, which is compatible with toxins from marine invertebrates, such as conotoxins (Fig. 4.2). The bioreactivity of the crude secretions and purified peptides was then asserted using the standardised Microtox test, which determines toxicity from the inhibition of the luminescence of the marine bacterium *Vibrio fischeri*. The effect to bacteria was dose-dependent and the EC_{50} threshold at 5 min (the half-maximal reduction in bioluminescence) to the bacteria was 98 µg total protein. mL^{-1} (95% confidence interval: 74–129) in seawater. The IC_{50} for purified peptides in PBS, at 5 min also, was higher (due to re-concentration of toxins as smaller compounds were removed) but within the same magnitude: 47 µg total protein. mL^{-1} (38–56). The results confirm that the toxicity is chiefly conferred by the proteinaceous component of the mixture.

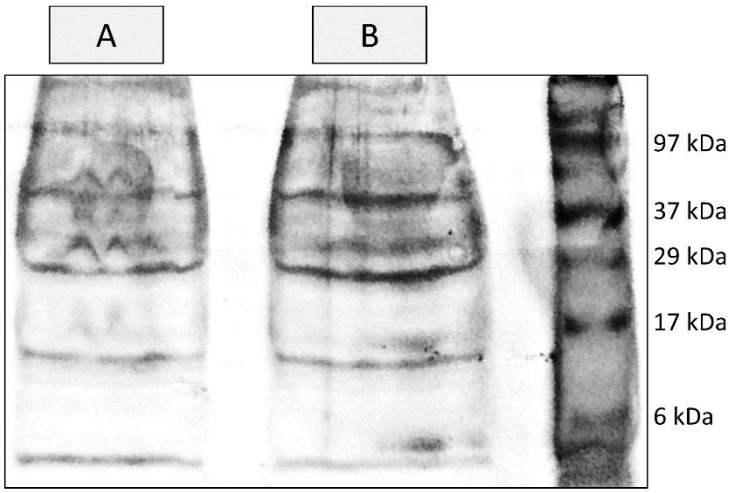


Figure 4.2. Protein signature of crude (A) and purified extracts from mucosecretions (B), diluted to the same amount of total protein (1 mg. mL^{-1}) in PBS and sterilised seawater, respectively, as visualised through SDS-PAGE (silver staining). Purification was done by ultrafiltration using a 3 kDa membrane, therefore removing salts and other small molecules.

4.4.3. Toxin mode-of-action in vivo

Using live mussels as convenient model, the toxin's mode-of-action was inferred through a series of bioassays conducted with crude mucous secretions diluted in sterilised seawater (whose reactivity was checked through the Microtox assay, as stated above) and applied via intravalvar injection to mimic administration of the toxin-containing secretions as the proboscis of the worm inserts itself in the mantle cavity. The frequency of valvar movements (opening or closure), determined during one hour after injection, was significantly reduced to about half in mussels exposed to the crude secretions, albeit without any evident dose-response (Fig. 4.3A). Also as a behavioural biomarker, increased latency time (elapsed time between touch and valve re-opening) was evident in treated animals, relatively to controls, particularly after 10 min, then showing recovery (Fig. 4.3B). These data agree with the reduction, followed by recovery, of two physiological biomarkers in exposed mussels, namely filtration rate (Fig. 4.4A), determined by microalgae removal from water, and oxygen consumption (Fig. 4.4B). An interesting effect was noted regarding the previous, though, as the mussels clearly over-responded one hour elapsed after injection with the most concentrated form of the crude toxin, suggesting hormesis. The recovery is reflected by the steeper rates (slopes) of the two parameters in exposed animals resulting from augmented removal of algae and oxygen from water 60 min after injection. Similarly, no significant reduction in acetylcholinesterase (AChE) activity was observed in the adductor muscle of exposed mussels, but rather an increase, more obvious one hour after administration of crude secretion, once more without an evident dose-response (Fig. 4.5A).

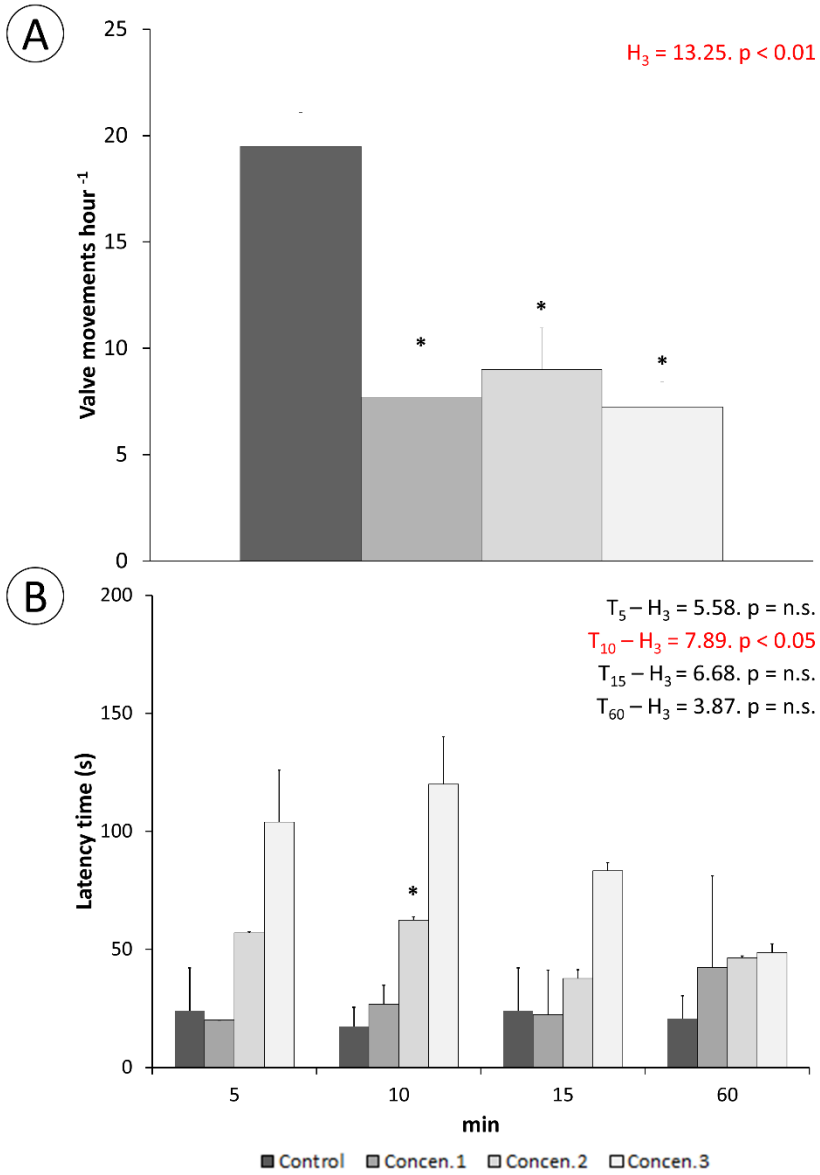


Figure 4.3. Behavioural responses in mussels exposed to toxic secretions (three concentrations) via intravalvar injection. (A) Valvar movement frequency. (B) Latency time to response (valvar re-opening) following stimulus (touch) in tested mussels treated with increasing concentrations of toxic secretions. The results are provided as means ± SEM. Statistical analyses were obtained with the Kruskal-Wallis ANOVA by ranks *H* for multiple comparisons. * Indicates significant differences to respective controls (Mann-Whitney *U*-test, $p < 0.05$).

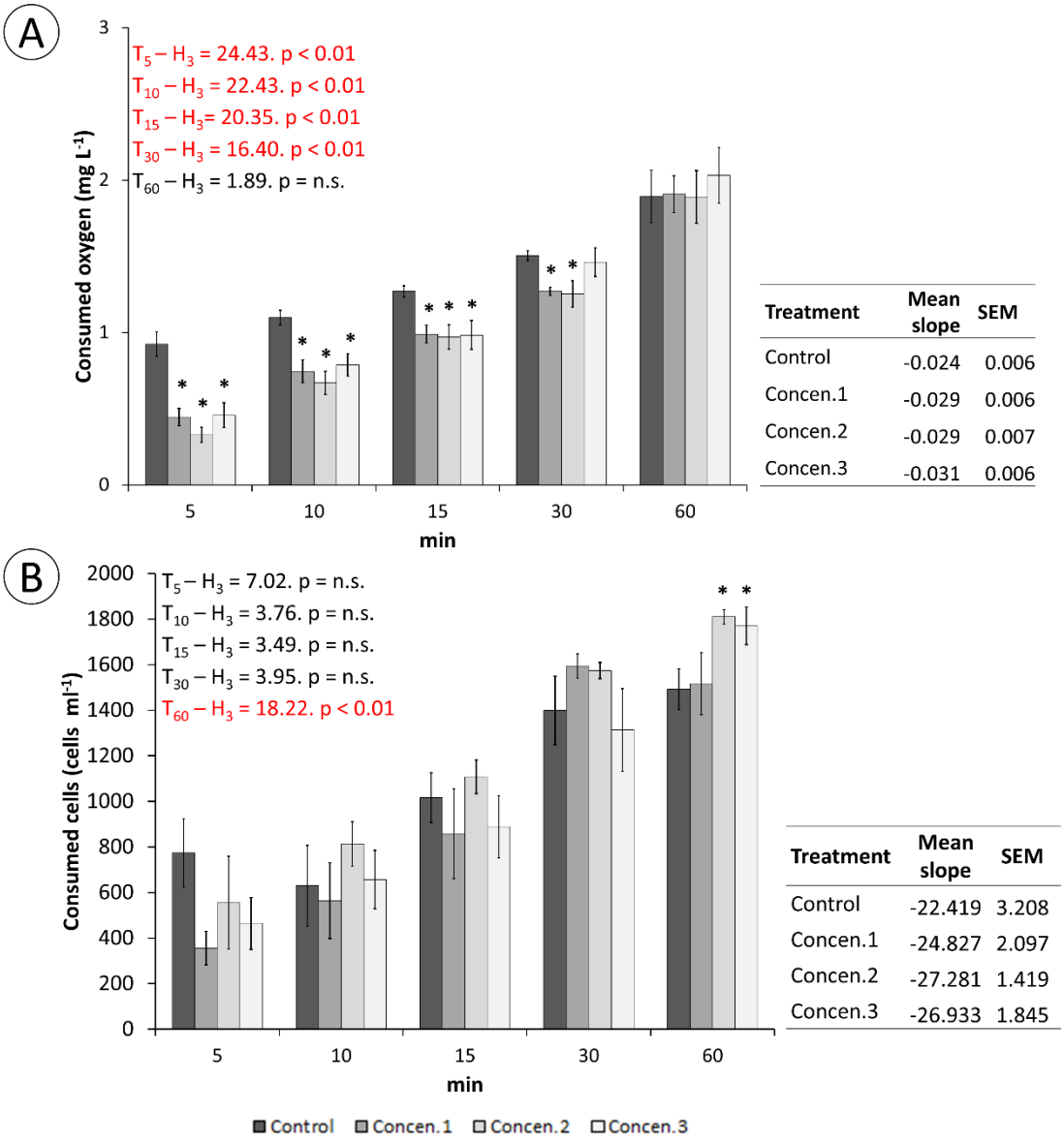


Figure 4.4. Physiological parameters in mussels exposed to increasing concentrations (Concen. 1 to Concen. 3) of toxic secretion. (A) Consumed oxygen. (B) Microalgae removal from water. Side panels show rates of removal (oxygen and alga cells) from water, per experimental treatment. The results are provided as means ± SEM. Statistical analyses were obtained with the Kruskal-Wallis ANOVA by ranks *H* for multiple comparisons. * Indicates significant differences to respective controls (Mann-Whitney *U*-test, $p < 0.05$).

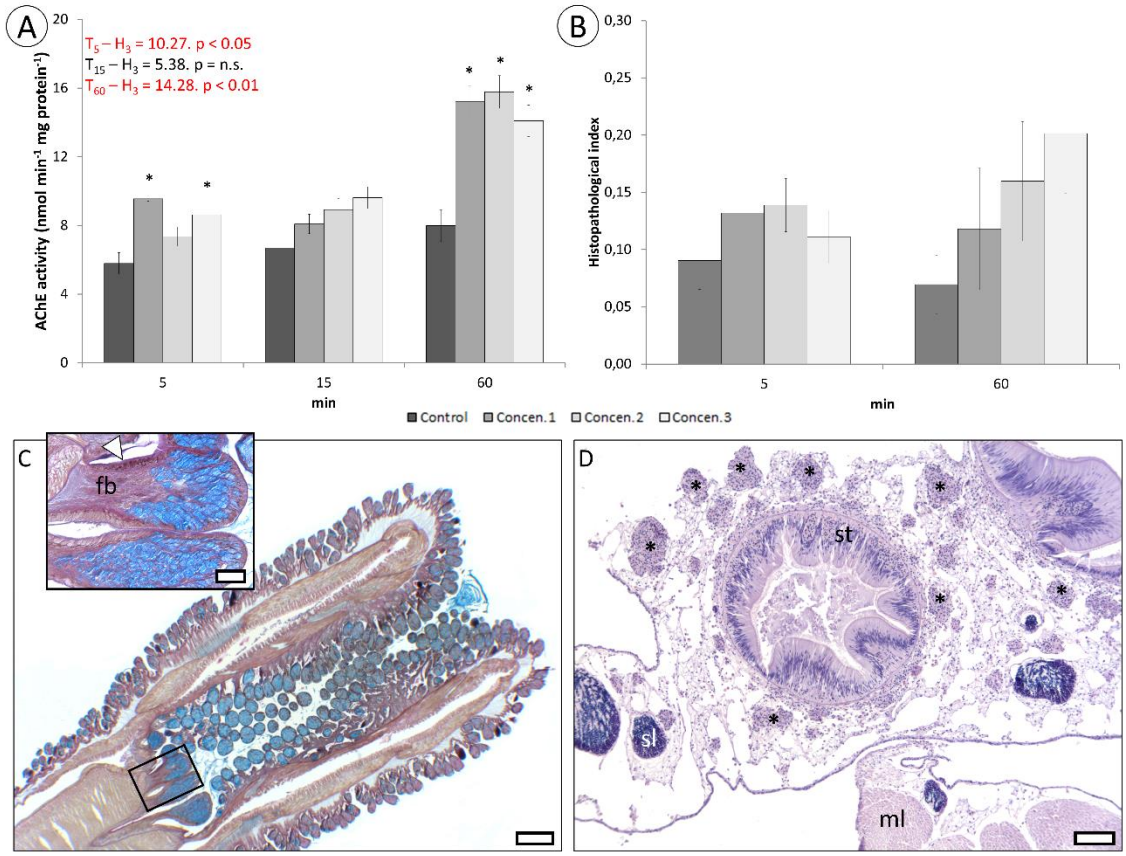


Figure 4.5. Neurochemical and toxicopathological effects in mussels exposed to increasing concentrations of toxin secretions (Concen. 1 to Concen. 3) at different timepoints. (A) Acetylcholine esterase (AChE) activity in adductor muscle. (B) Integrated (multi-organ) histopathological condition index. The results are provided as means ± SEM. Statistics were obtained with the Kruskal-Wallis ANOVA by ranks *H* for multiple comparisons. * Indicates significant differences to respective controls (Mann-Whitney *U*-test, $p < 0.05$). (C) Longitudinal section of the proboscis of *E. viridis* (not fully everted) highlighting toxin-delivery tentacles (box). Blueish cells are mucocytes, revealed by Alcian Blue dye in the tetrachrome stain. Inset: high-power magnification of toxin-delivery tentacles, highlighting serous cells were toxins and enzymes are produced (arrowhead) and bundles of connective and nervous fibres (fb). (D) Histological section across the visceral mass of mussels one hour elapsed after injection of the highest concentration of toxin (Concen. 3). Several foci of defence cells agglomerates (*) indicate mild inflammation, close to stomach (st) (H&E stain). sl: seminiferous lobe; ml: muscle. Scale bars: (C) 200 µm, inset 50 µm; (D) 200 µm.

Histopathological alterations in exposed animals, translated into a condition index, were scarce and chiefly related to increased inflammatory response in exposed animals. These alterations refer to focal infiltration of haemocytes in several tissues, with emphasis on the external portion of the digestion gland and adjacent mantle (*i.e.*, the more exposed areas). Despite the trend to increase 1 h after exposure, the dissemination of alterations was variable, which made determining statistical significance difficult (Fig. 4.5B). Nonetheless, significant differences after 1 h of injection between the condition index of animals exposed to the highest dose (Concen. 3) and controls were dismissed

with $p = 0.06$ (Mann-Whitney *U*-test). For guidance, in Fig. 4.5C is shown the location of the toxin delivery tentacles at the tip of the proboscis, which, in this case, is partially inverted. The dense agglomerates of mucous (blue) and serous (toxin-secreting) cells are evident, the latter of which line the entire base of tentacles. In Fig. 4.5D is exemplified the formation of focal agglomerates of defence cells (haemocytes) in an area of the digestive gland near the mantle of an exposed mussel (highest concentration of the toxin secretions).

Hierarchical clustering of measured responses and effects during these bioassays (Fig. 4.6) allowed segregating experimental treatments in two major clusters. The fist comprised controls and blanks (*i.e.* animals injected with seawater only), the second all exposure treatments. Among the latter, exposure to the highest concentration (*i.e.* lowest dilution) stand out from the two preceding, with particular respect to delayed responsive behaviour (latency time) and histopathological alterations, especially after one hour of injection. In their turn, variables are segregated in two major clusters (upmost hierarchical tree), the first of which including physiologically-related variables, namely algae filtration, O_2 consumption and valvar movement. These parameters were rapidly reduced by exposure to toxin secretions. The remaining responses correspond to effects that were enhanced by exposure, from AChE activity to histopathology and latency time, which thus showed an inverse pattern to that of the preceding cluster. The positioning of AChE parted from valve movements and related responses indicates that phyllotoxins are unlikely to have an inhibitory effect on this post-synaptic enzyme.

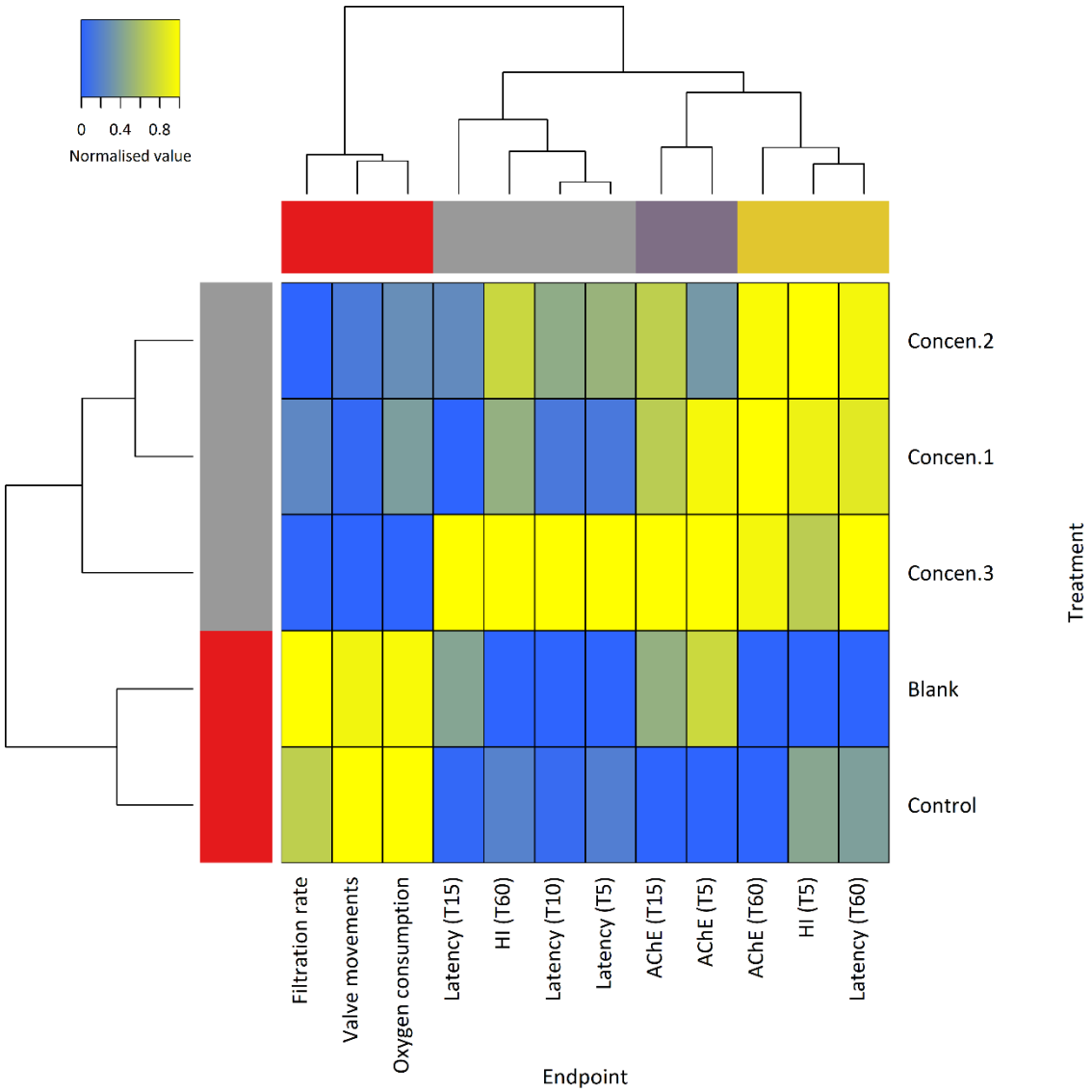


Figure 4.6. Heatmap and hierarchical clustering of normalised results obtained with complete linkage and Euclidean distances, aggregating all measured responses from mussels treated with toxins (separated by time after exposure). Hierarchical trees and colour bars indicate association between variables (up) and experimental conditions (side). Note the clustering between control and blanks, clearly separated from the treatments with toxins. The association between filtration, oxygen consumption and valve movements is also clear, which makes physiological sense, all of which being negatively affected by the toxin.

4.4.4. Pathological aspects of direct contact with mucous secretions

As intravalvar injections likely dispersed of mucus and toxins, there was the need to ascertain the realistic effects resulting from the direct contact with mucous secretions. In order to mimic direct contact, we applied freshly collected mucus onto the soft tissue of whole live mussels or through *ex vivo* contact with freshly excised organs. The results showed more prominent histopathological alterations, comparatively to the previous assays. In Fig. 4.7A-C are shown sections from foot

muscle, showing digestion of muscle fibres evidenced by myocyte hyalinisation, accompanied of disorganisation of connective fibres and apoptotic cells. The effects were stronger in whole mussels than in *ex vivo* assays. Alterations in the glandular epithelium lining the mantle were also evident in the contact area (Fig. 4.7D-E). These alterations are chiefly related to increased secretory activity, as seen from the proliferation of granular cells suggests detoxification and increased production of protective glycoproteins, mucins included.

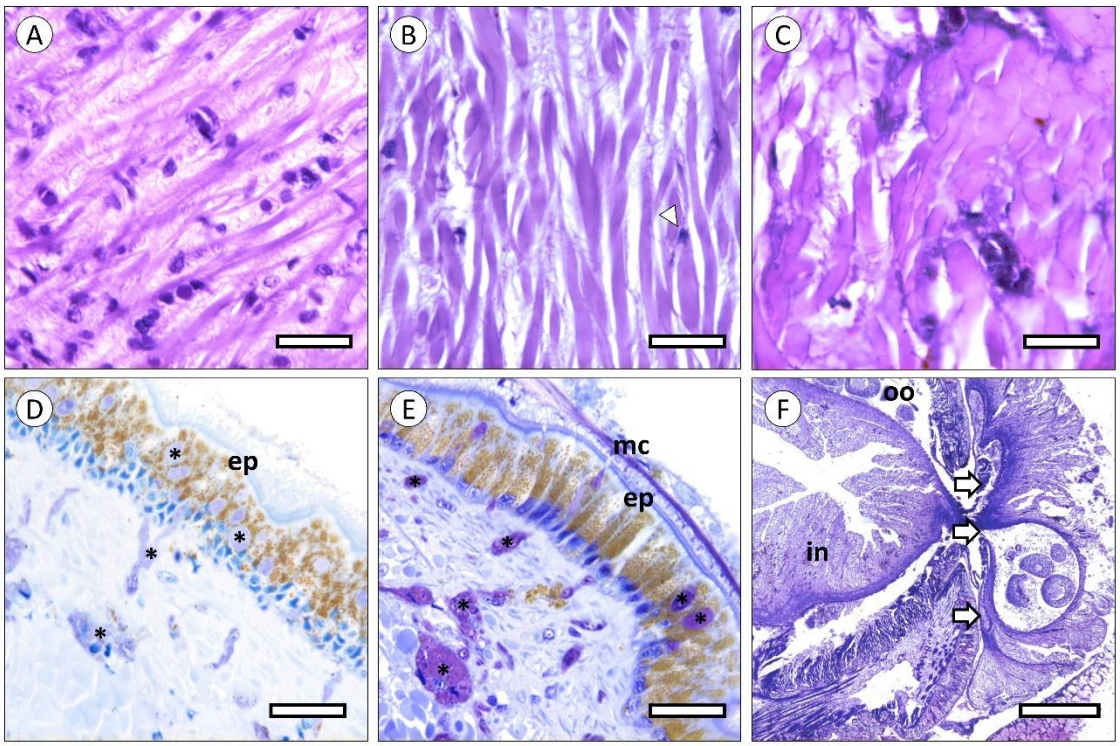


Figure 4.7. Histopathological evaluation of the effects of toxin in *E. viridis* mucus at the contact area with the crude secretions, in natural prey. (A-C) Muscle tissue from the foot of mussels (H&E). (A) Normal aspects of tissue. (B) Moderately affected muscle in a foot exposed *ex vivo* to the mucous secretions, showing disorganization of connective fibres and hyalinisation (digestion) of muscle fibres, which become more eosinophilic (pink). Note the abnormal aspect of nuclei (arrowhead), indicating early cell death. (C) Foot muscle of a live mussel exposed to crude mucus, revealing severe hyalinisation of muscle fibres. (D-E) Resin (semi-thin) sections of the edge of the foot (PAS-Toluidine). (D) Normal secretory epithelia and underlying connective tissue (unexposed animal). (E) Proliferation of secretory cells (*) and increased secretion. Note the layer of mucus (mc) atop the ciliated epithelium of the mussel (ep). Resin section (semi-thin) of the wound caused by an *E. viridis* onto a conspecific, in its habitat. (F) The toxic secretion partially digested the body wall of the prey, enabling suction of contents (arrows), up to the point where oocytes (oo), which mature in the coelom, can be observed being pulled out, together with a portion of the intestine (in). Scale bars: (A-E) 20 µm; (F) 400 µm.

4.5. Discussion

As *E. viridis* joins the ranks of toxin-secreting Polychaeta, it is shown that small body size and low level of organ differentiation, features shared by most extant protostomes, can be circumvented by chemical warfare as strategies for predation and defence. The aggregate effect of the cocktail of proteinaceous toxins, here referred as "phyllotoxins", present in the mucous secretions of the worm are applied to the prey by repeated contact with the tip of the proboscis, where specialised tentacles are located, eventually causing (reversible) immobilisation while partially digesting the soft tissue. The absence of massive tissue digestion (recall Figs. 4.5 and 4.7), indicates that the main function of enzymes in mucosecretions is not extracorporal digestion but a combination between permeabilization to toxins and tissue softening to facilitate extraction by suction, as shown in Fig. 4.7F. It must be noticed that *E. viridis* is not only able to extract and ingest large portions of soft tissue but also entire prey, including conspecifics, as shown in our previous work on the species' digestive function (Rodrigo et al. 2015). This ability reiterates the convenience of secreting immobilising or relaxing toxins. It must be noted, however, that the complexity of secretions containing toxins, plus the impossibility of neutralising noxious substances only, hinders establishing direct cause-effect relationships. Despite the aggregate evidence for *Eulalia* toxins, as for other Phyllodocida, which is based on multiple endpoints, there are recent indications that bivalves can respond to undisclosed waterborne chemical clues from potential predators. Sensing these clues can lead to various defensive changes, from the reduction of metabolic and filtration rates to the thickening of shells in a predator-laden habitat (Freeman and Byers 2006; Naddafi et al. 2007; Antoł et al. 2018). It is thus possible that these chemical clues may have contributed to the lowering of physiological and related behavioural parameters such as gas exchange, filtration and valvar movements, even though they cannot explain *per se* the observed toxicopathological effects and how *Eulalia* can penetrate through the valves. In any case, the complexity of toxin-bearing animal secretions mandates some caution when inferring specific effects.

Despite the little work done, so far, on Polychaeta, some authors have already reported toxins from these animals. The toxicology and the ecological role of these substances remains, nonetheless, little understood. Among these, the best-known cases are nereistoxin, arenicin and glycerotoxin (Okaichi and Hashimoto 1962; Ovchinnikova et al. 2004; Richter et al. 2017). Nereistoxin, in particular, is a non-proteinaceous neurotoxic alkaloid that originates from the salivary gland of *Lumbrinereis heteropoda*. The role of this toxin remains elusive, as it is known by its insecticide properties with neuromuscular effects that have been associated to its metabolites and not to the parental compound or compounds (Deguchi et al. 1971; Xie et al. 1996). Nereistoxin is nowadays considered one of the most promising basis to develop "ecologically-sustainable" insecticides. Glycerotoxin, in its turn, refers to a well-known high molecular weight (c.a. 320 kDa) protein from

the venom glands of bloodworms (*Glycera tridactyla*), which belong to Phyllodocida just as *Eulalia*. Even though recent works suggest the neurotoxicity of glycerotoxin based on molecular data (Richter et al. 2017), the first descriptions of *Glycera* venom glands plus the successful isolation and testing of glycerotoxin isoforms as neurotoxins already date from the 1960s to the 1980s (see Bon et al. 1985; Schenning et al. 2006 and references therein). Another example, arenicin, however, pertains to antimicrobial peptides produced in the coelomocytes of the marine Polychaeta *Arenicola marina* (lugworm), a burrower worm that, unlike the former, is not a predator (Ovchinnikova et al. 2004).

Being able to immobilise prey, even if partially or temporarily, is always an advantage for a predator. However, it is of particular importance for those that cannot physically out-compete their targets, such as *E. viridis*. There are several uncanny adaptations for the purpose. It is the case, for instance, of the modification of one of the claws of the snapping shrimp *Alpheus heterochaelis*, whose clacking creates an air bubble that pops violently, producing a strong sound that stuns prey (or keeps attackers and competitors at bay), including much larger fish (Versluis et al. 2000). Chemical stunning seems to be more common, though, and likely more cost-effective. Existing data on marine invertebrate neurotoxins, particularly from *Conus*, support this premise but attempts to validate these occurrences as adaptative features, which involves understanding the mode-of-action of whole venoms onto ecologically-relevant targets, are scarce. In *E. viridis*, the immobilising effect of the toxin is reversible and not immediate. In fact, *in situ* observations and the endpoints illustrated in Figs. 4.3 to 4.5 suggest a peak of effects between 5 and 10 min, followed not only by recovery and even over-response, most likely a hormetic effect to compensate non-lethal challenge, *i.e.* a beneficial over-response to reduced toxicological stress (see Calabrese, 2008). It must be noticed that the mussel bioassays may underestimate the full potency of phyllotoxins, as they do not mimic the repeated, direct contact, promoted by the worm *in situ*. This explains lower histopathological effects in tested mussels, comparatively to the *ex vivo* assessment (Fig. 4.7A-E), which was done by direct contact of mucus onto tissue rather than intravalvar diffusion, and the consequences to the captured annelid shown at Fig. 4.7F. However, in either case, histological alterations pertain mostly to fibrous tissue, which indicates the need to soften and permeabilise the material prior to extraction of large pieces and not to uphold extra-corporal digestion, as in maceration.

Comparatively, conotoxins are fast-acting and many are lethal. Among well-known lethal toxins, tetrodotoxin (TTX), which has been detected in several marine organisms such as blue-ringed octopus and pufferfish, is one of the most powerful neurotoxins (Lago et al. 2015). Additionally, potent and fast-acting toxins are usually injected, which allocates mixtures of conopeptides and other neurotoxins into the category of "venoms" (Nelsen et al. 2014). However, there is a huge variety of conotoxins, by far the best-studied natural marine toxic compounds, the vast majority of

which is studied from recombinant forms and not in its native forms, which also means that the effects on prey are not understood (see Akondi et al. 2013, for a review). Altogether, our data suggest that phyllotoxins have a mode-of-action *in vivo* distinct of most fast-acting neurotoxic conopeptides and similar, many of which target voltage-gated ion channels directly. In addition, the failure to correlate "behavioural" parameters such as valve movements and latency time with AChE suggests that the toxin does not interfere with this serine hydrolase directly and that its increase is indeed the consequence of the over-response mentioned earlier. The absence of any form of complex gland of venom-injecting apparatus in *E. viridis* is in accordance with the animal's behaviour and toxin administration via mucous secretion from the tip of the proboscis, where specialised tentacles are located. Thus, as the toxins are not delivered through a wound, they fall within the proposed category of "toxungen" (Nelsen et al. 2014). This form of delivery is coupled with repeated contact with indiscriminate soft-bodied prey, which may range from mussels to other Polychaeta. Still, while contact with annelids and even with some gastropods (like *Patella* spp.) can be relatively conspicuous and appears to be a relatively simple process, the way how the worm penetrates valves of mussels and plates of barnacles is more difficult to record and to explain. Nonetheless, the existence of the toxin *per se* sheds light on the process. Rovero et al. (1999) suggested that dogwhelks (*Nucella* spp.) can penetrate barnacles and even between the valves of bivalves using its proboscis as an alternative behaviour to the more common process of drilling through the shells of mussels, using its radula, by secreting an unknown immobiliser ("relaxant") to facilitate insertion. Given the form and function of the proboscis, it seems evident that *E. viridis* operates in a similar way, especially considering that the animal is not equipped with structures able to force against the powerful adductor muscles of bivalves or drill through shells and plates, for instance.

4.6. Conclusion

As in the case of cephalotoxins, it has been hypothesised that the presence of specific enzymes in toxin mixtures may facilitate the infiltration of neurotoxins (Whitelaw et al. 2016). In *E. viridis*, these enzymes seem to have the function of partially digesting tissue to allow extraction. The effects of the secreted substances appear to be the digestion of muscle fibres and connective tissue. Although some evidence for cell death (likely apoptosis due to condensation and blebbing of nuclei) having been registered (Fig. 4.7B-C), it is not possible to ascertain whether this is a direct effect of some specific pro-apoptotic compound or due to action of proteinases, many of which are known to favour, if not trigger, programmed cell death. Still, this issue is not yet well understood in invertebrates. In any case, the effect is sufficiently potent to perforate the body wall of other annelids and pull, through suction, the contents from the coelomic cavity, as well as the remnants

from the partially digested musculature. The combination between powerful suction, immobilising toxins and proteolytic activity thus maximises the predatorial abilities of the worm, in spite of the reduced complexity and differentiation so common in protostomes. These features make it a highly efficient predator in its environment, not just against immobile prey like barnacles and bivalves, but also against its immediate rivals, such as other predatorial Polychaeta, including its conspecifics.

4.7. References

Akondi KB, Muttenthaler M, Kaas Q, Craik DJ, Lewis RJ, Alewood PF (2013) Discovery , Synthesis , and Structure − Activity Relationships of Conotoxins.

Antoł A, Kierat J, Czarnoleski M (2018) Sedentary prey facing an acute predation risk: testing the hypothesis of inducible metabolite emission suppression in zebra mussels, *Dreissena polymorpha*. Hydrobiologia 810:109–117.

Bon C, Saliou B, Thieffry M, Manaranche R (1985) Partial purification of α-glycerotoxin, a presynaptic neurotoxin from the venom glands of the polychaete annelid *Glycera convoluta*. Neurochem Int 7:63–75.

Burgess JG (2012) New and emerging analytical techniques for marine biotechnology. Curr Opin Biotechnol 23:29–33.

Calabrese EJ (2008). Hormesis: Why it is important to toxicology and toxicologists. Environ. Toxicol. Chem. 27, 1451-1474.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG (2013) Complex cocktails: The evolutionary novelty of venoms. Trends Ecol Evol 28:219–229.

Costa PM, Costa MH (2012) Development and application of a novel histological multichrome technique for clam histopathology. J Invertebr Pathol 110:411–414.

Costa PM, Carreira S, Costa MH, Caeiro S (2013) Development of histopathological indices in a commercial marine bivalve (*Ruditapes decussatus*) to determine environmental quality. Aquat Toxicol 126:442–454.

Cuevas N, Zorita I, Costa PM, Franco J, Larreta J (2015) Development of histopathological indices in the digestive gland and gonad of mussels: Integration with contamination levels and effects of confounding factors. Aquat Toxicol 162:152–164.

Deguchi T, Narahashi T, Haas HG (1971) Mode of Action of Nereistoxin Transmission on the in the Frog. Pestic Biochem Physiol 1:196–204.

Dutertre S, Jin AH, Vetter I, Hamilton B, Sunagar K, Lavergne V, Dutertre V, Fry BG, Antunes A, Venter DJ, Alewood PF, Lewis RJ (2014) Evolution of separate predation-and defence-evoked venoms in carnivorous cone snails. Nat Commun.

Ellman GL, Courtney KD, Andres V, Featherstone RM (1961) A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem Pharmacol 7:88–95.

Freeman AS, Byers JE (2006) Divergent induced responses to an invasive predator in marine mussel populations. Science 313:831–833.

Fusetani N, Kem W (2009) Marine Toxins as Research Tools.

Gao B, Peigneur S, Dalziel J, Tytgat J, Zhu S (2011) Molecular divergence of two orthologous scorpion toxins affecting potassium channels. Comp Biochem Physiol - A Mol Integr Physiol 159:313–321.

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. J Comput Graph Stat 5:299–314.

Lago J, Rodriguez LP, Blanco L, Vieites JM, Cabado AG (2015) Tetrodotoxin, an extremely potent marine neurotoxin: Distribution, toxicity, origin and therapeutical uses. Mar Drugs 13:6384–6406.

Morton B (2011) Predator-prey-scavenging interactions between *Nucella lapillus*, *Carcinus maenas* and *Eulalia viridis* all exploiting mytilus galloprovincialis on a rocky shore recovering from tributyl-tin (TBT) pollution. J Nat Hist 45:2397–2417.

Naddafi R, Eklöv P, Pettersson K (2007) Non-lethal predator effects on the feeding rate and prey selection of the exotic zebra mussel *Dreissena polymorpha*. Oikos 116:1289–1298.

Nelsen DR, Nisani Z, Cooper AM, Fox GA, Gren ECK, Corbit AG, Hayes WK (2014) Poisons, toxungens, and venoms: Redefining and classifying toxic biological secretions and the organisms that employ them. Biol Rev 89:450–465.

Nisani Z, Hayes WK (2015) Venom-spraying behavior of the scorpion parabuthus transvaalicus (arachnida: Buthidae). Behav Processes 115:46–52.

Okaichi T, Hashimoto Y (1962) The structure of nereistoxin. Agric Biol Chem 26:224–227.

Ovchinnikova T V., Aleshina GM, Balandin S V., Krasnosdembskaya AD, Markelov ML, Frolova EI, Leonova YF, Tagaev AA, Krasnodembsky EG, Kokryakov VN (2004) Purification and primary structure of two isoforms of arenicin, a novel antimicrobial peptide from marine Polychaeta *Arenicola marina*. FEBS Lett 577:209–214.

Richter S, Helm C, Meunier FA, Hering L, Campbell LI, Drukewitz SH, Undheim EAB, Jenner RA, Schiavo G, Bleidorn C (2017) Comparative analyses of glycerotoxin expression unveil a novel structural organization of the bloodworm venom system. BMC Evol Biol 17:1–19.

Rodrigo AP, Alves de Matos AP, Carrapiço F, Costa MH, Costa PM (2014). A physiological study of integument secretions in the marine polychaete *Eulalia viridis* and their potential biotechnological value. Front. Mar. Sci. Conference.

Rodrigo AP, Costa MH, Alves De Matos AP, Carrapiço F, Costa PM (2015) A study on the digestive physiology of a marine polychaete (*Eulalia viridis*) through microanatomical changes of epithelia during the digestive cycle. Microsc Microanal 21:91–101.

Rovero F, Hughes RN, Chelazzi G (1999) Cardiac and behavioural responses of mussels to risk of predation by dogwhelks. Anim Behav 58:707–714.

Schenning M, Proctor DT, Ragnarsson L, Barbier J, Lavidis NA, Molgó JJ, Zamponi GW, Schiavo G, Meunier FA (2006) Glycerotoxin stimulates neurotransmitter release from N-type Ca^{2+} channel expressing neurons. J Neurochem 98:894–904.

Sunagar K, Morgenstern D, Reitzel AM, Moran Y (2016) Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. J Proteomics 135:62–72.

Tzetlin A, Purschke G (2005) Pharynx and intestine. Hydrobiologia 535/536:199–225.

van der Ham JL, Felgenhauer BE (2007) The Functional Morphology of the Putative Injecting Apparatus of Speleonectes Tanumekes (Remipedia). J Crustac Biol 27:1–9.

Versluis M, Schmit B, Heydt A von der, Lohse D (2000) How Snapping Shrimp Snap: Through Cavitating Bubbles. Science 289:2114–2117.

von Reumont BM, Campbell LI, Richter S, Hering L, Sykes D, Hetmank J, Jenner RA, Bleidorn C (2014) A polychaete's powerful punch: Venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. Genome Biol Evol 6:2406–2423.

Von Reumont BM, Blanke A, Richter S, Alvarez F, Bleidorn C, Jenner RA (2013) The first venomous crustacean revealed by transcriptomics and functional morphology: Remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. Mol Biol Evol 31:48–58.

Xie Y, McHugh T, McKay J, Jones GS, Loring, RH (1996). Evidence that a nereistoxin metabolite, and not nereistoxin itself, reduces neuronal nicotinic receptors: studies in the whole chick ciliary ganglion, on isolated neurons and immunoprecipitated receptors. J. Pharmacol. Exp. Ther. 276, 169-177.

Whitelaw BL, Strugnell JM, Faou P, Da Fonseca RR, Hall NE, Norman M, Finn J, Cooke IR (2016) Combined transcriptomic and proteomic analysis of the posterior salivary gland from the southern blue-ringed octopus and the southern sand octopus. J Proteome Res 15:3284–3297.

CHAPTER 5 - How *Eulalia viridis* can give a small step into cancer fighting: targeting ovarian carcinoma

This chapter has been submitted for publication:

Rodrigo AP, Mendes V, Manadas B, Grosso AR, Alves de Matos AP, Baptista PV, Costa PM, Fernandes AR (2020). Specific anti-proliferative properties of proteinaceous toxin secretions from the marine annelid *Eulalia* sp. onto ovarian cancer cells. *Mar Drugs* 19, 31

5.1. Abstract

As Yondelis joins the ranks of approved anti-cancer drugs, the benefit from exploring oceans' biodiversity becomes clear. From marine toxins, relevant bioproducts can be obtained due to their potential to interfere with specific pathways. We explored the cytotoxicity of toxin-bearing secretions of the polychaete *Eulalia viridis* onto a battery of normal and cancer human cell lines and discovered that the cocktail of proteins is more toxic towards an ovarian cancer cell line (A2780). The secretions' main proteins were identified by proteomics and transcriptomics: 14-3-3 protein, Hsp70, Rab3, Arylsulfatase B and serine protease, the latter two being known toxins. This mixture of toxins induces cell cycle arrest at G2/M phase after 3h exposure in A2780 cells and extrinsic programmed cell death. These findings indicate that partial re-activation of the G2/M checkpoint, which is inactivated in many cancer cells, can be partly reversed by the toxic mixture. Protein-protein interaction networks partake in two cytotoxic effects: cell cycle arrest with a link to RAB3C and RAF1; and lytic activity of arylsulfatases. The discovery of both mechanisms indicates that venomous mixtures may affect proliferating cells in a specific manner, highlighting the cocktails' potential in the fine-tuning of anti-cancer therapeutics targeting cell cycle and protein homeostasis.

5.2. Introduction

According to the World Health Organization (WHO), cancer is the second leading cause of death globally. The numbers of cases are expected to rise every year despite research on the subject being given top priority (reviewed by Das et al. 2018). From all the types of cancers, breast and lung cancer are the leading causes of death, especially in Europe, with prostate cancer following the rank (Ferlay et al. 2018). Ovarian cancer is the deadliest gynaecological cancer in the world (see, for instance, Beaufort et al. 2014), in large part due to the late diagnosis (Lalwani et al. 2011). This type of cancer is in fact a range of neoplasic diseases that can have different origins (Anglesio et al. 2013). Research on new therapeutics for this disease is motivated by the combined effect between genomic heterogeneity and cell resistance (Hallas-Potts et al. 2019). Moreover, finding a suitable therapy with reduced effects on healthy tissue remains a challenge, leading to the screening for novel anti-cancer compounds that offer more specific and less deleterious strategies than traditional treatments.

Anti-cancer drugs developed from natural products such as toxins and alkaloids nowadays represent 50% of all drugs that are used in chemotherapy (Mann 2002). After plants, fungi and bacteria have been leading the search for anti-cancer compounds, in addition to novel antibiotics

(Mann 2002). Some of the main targets of the quest for anti-cancer drugs are signalling pathways that regulate cell survival and proliferation. Indeed, most research is seemingly focused on the inhibition of specific protein kinases, such as Ras/Raf mitogen-activated protein kinase (MEK) and extracellular signal-regulated kinase (ERK) signalling pathways (Roberts and Der 2007).

There is a growing awareness for the potential of the seas as a source of novel therapeutic agents, in face of their immense biodiversity (Mann 2002; Leal et al. 2020). In fact, the number of marine natural products explored for various biotechnological purposes has been increasing over the past few years (see the review of Hu et al. 2015) for a statistical appraisal on marine bioprospecting). Accordingly, there is growing interest in the search for novel anti-cancer compounds in marine life, benefitting from the constant advances in omics methods from metabolomics to proteomics and transcriptomics (Khalifa et al. 2019). With the exception of Yondelis (trabectedin), an approved drug developed from a secondary metabolite (a quinoline) of the tropical tunicate *Ecteinascidia turbinata* that competes with enzyme and transcription factors that bind to DNA (Cuevas and Francesch 2009), most research has yet to be materialized into effect applications. Still, there are promising indications, from various marine animals as varied as the bivalves, the Briozoa and the Porifera, on the effects of novel compounds on the inhibition of DNA synthesis inhibition, RNA polymerase activity, activation of apoptosis and autophagy and cell cycle arrest in cancer cells (Chernikov et al. 2017; do Nascimento-Neto et al. 2018; Figuerola and Avila 2019; Khalifa et al. 2019). In most cases, though, cancer cells are merely used as a toxicological model, without a clear indication of mode-of-action. From the Polychaeta, a little-explored but most promising group of marine animals for the bioprospecting of novel bioproducts, there are growing indications for the existence of cytotoxic toxins. Among the most significant examples, we may find the toxin arenicin (from *Arenicola marina*), nicomicin (from *Nicomache minor*), both peptidic in nature, and proteases from *Nereis*, all presenting cytotoxicity towards cancer cell lines, seemingly with pro-apoptotic effects (see Chapter 1 for further details). The mechanisms by which these substances exert noxious effects are not, however, fully understood. In large part, this is due to constraints in handling the complex mixtures of which poisons and venoms are comprised, particularly in marine invertebrates, for whom genomic resources are much reduced.

Recently, we showed the secretion of toxins by an intertidal Polychaeta, *Eulalia viridis*, whose toxins, secreted by specialized cells in the eversible pharynx, are delivered to its prey using mucus as a vehicle with the purpose of immobilization and cytolysis by means yet unknown. In the aftermath of these early findings, the current work aims primarily at disclosing the mechanisms by which proteins in *Eulalia*'s secretions become cytotoxic to normal and cancer human cells, with the ultimate goal of asserting their biotechnological potential. To meet our objectives, we isolated proteins from secretions and combined proteomics with transcriptomics for their

identification as a mean to circumvent the challenges of investigating a complex mixture of proteinaceous compounds and provide inter-validation in a scenario of low genomic resources. The purified extracts were then tested onto both normal and cancer humans cell lines.

5.3. Material and methods

5.3.1. Protein extraction and identification

Animal collection

Adult *Eulalia viridis* (≈ 120 mm total length and weighting ≈ 250 mg each) were collected from the West Coast of Portugal (38°41′42″N, 09°21′36″W), in an intertidal rocky beach. Animals were reared in the laboratory in a mesocosm environment recreating their natural habitat, with controlled salinity, temperature and photoperiod (35 ± 1, 16 ± 1 ºC and 10:16 h, respectively). Animals were fed with live mussels, one of the species' favourite prey.

Protein collection and purification

Protein was collected from mucous secretions harvested non-invasively by placing the worms in microfuge tubes with filtered sterile seawater, followed by mechanical stimulation of the proboscis, leading to increased secretion. Crude protein extracts were obtained from approximately fifty animals per pool. The crude extracts were immediately diluted ≈ 1:1 in 0.05 M Tris-HCl containing 10% (w/v) L-dithiothreitol (DTT) and 1% (v/v) protease inhibitor cocktail (Sigma-Aldrich). Samples were then filtered through a syringe cellulose acetate filter (0.22 µm). Purification and concentration of extracts was done by dialysis (ultrafiltration) using 3kDa Amicon Ultra centrifugal filters (Merck), with Dulbecco's phosphate-buffered saline (DPBS, pH 7.4) as the vehicle. Total protein content was determined using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific). Purified extracts were stored at -80 ºC until further analyses.

Protein separation

Proteomics was employed to analyse the purified protein extracts using either, one- or two-dimensional gel electrophoresis (1DE and 2DE, respectively) for initial separation (Gomes et al. 2018). The Laemmli discontinuous gel of Laemmli (1970) was used for 1DE separation by sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE), as detailed by Hames (1998). Stacking and running gels (8×9 cm by 0.75 mm thick) contained 12% (v/v) and 6% (v/v) acrylamide, respectively. The migration buffer consisted of 25 mM Tris, 192 mM glycine, pH 8.5. Molecular standards (NZYColour Protein Marker II range 11-245 kDa) was purchased from

Nzytech. The 2-D separation was obtained by isoelectric focusing (IEF) followed by SDS-PAGE. Briefly, the sample consisted of 100 µL of purified protein extract treated with 2-D Clean-Up Kit (GE Healthcare), according to manufacturer's instructions. Precipitated proteins were resuspended in 100 µL of rehydration solution containing 4 % (w/v) CHAPS, 7 M urea, 2 M thiourea, 1 % (w/v) DTT GE Healthcare, 0.5 % (v/v) Immobilized pH Gradient buffer (GE Healthcare) and 0.002 % (w/v) bromophenol blue, and incubated at room temperature for 36 h. After protein quantification using the Pierce 660 nm protein assay reagent (Thermo Scientific), 125 µL of rehydration buffer containing 200 µg of protein were applied onto Immobiline Drystrip IPG strips (GE Healthcare) with 7 cm and 3 ± 10 non-linear pH. After rehydration, Isoelectric focusing (IEF) was done in an Ettan IPGphor instrument (GE Healthcare) at a maximum voltage of 5000 V. The IPG strips were afterwards incubated for 15 min in equilibration buffer 1 (50 mM Tris-HCl pH 8.8, 2 % (w/v) SDS, 6 M urea, 30 % (v/v) glycerol, 1 % (w/v) DTT, followed by incubation for another 15 min in equilibration buffer 2 (50 mM Tris-HCl pH 8.8, 2 % (w/v) SDS, 6 M urea, 30% (v/v) glycerol, 2.5% (w/v) iodoacetamide). Strips were then placed on top of a 10% SDS-PAGE gel for separation by mass. Both gels were stained with Coomassie Brilliant Blue G-250 to identify most representative spots and bands (Fig. 5.1) to be picked for liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analyses. Excised gel bands and spots were stored at -80 ºC until further processing.

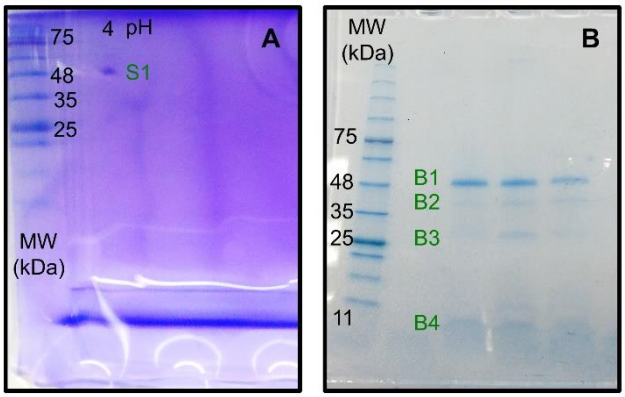


Figure 5.1. Image of proteins identified by MS/MS. (A) Spot obtained from 2-DE gel (S1 – Arylsulfatase B). (B) Bands retrieved from SDS-PAGE gel (B1 – Heat shock protein 70, B2 - 14-3-3 protein, B3 - Rab3, B4 – Serine protease).

Protein identification by LC-MS/MS

After gel sample distaining and digestion with porcine trypsin, peptides were analysed on a NanoLC Ultra 2D separation system (Eksigent) coupled to a TripleTOF 6600 mass spectrometer (Sciex). The chromatographic separation by micro-LC was achieved in a ChromXP C18CL column (0.3 × 150 mm, 3 μm, 120 Å, from Eksigent), at 50 °C. The flow rate was set to 5 μL. min^{-1} and mobile phase A and B were 0.1% (v/v) formic acid plus 5% (v/v) DMSO in water and 0.1% (v/v) formic acid plus 5% (v/v) DMSO in acetonitrile, respectively. The ionization source (ESI DuoSprayTM Source) was operated in the positive mode set to an ion spray voltage of 5500 V, 25 psi for nebulizer gas 1 (GS1), 25 psi for the curtain gas (CUR). The rolling collision was used with a collision energy spread of 5. Peptide mass fingerprinting was performed using ProteinPilot 5.0.1 (Sciex) considering the following parameters: cysteine alkylation by acrylamide, digestion by trypsin, and gel-based ID as a special factor. The resulting amino acid sequences were contrasted against UniProtKB and NCBInr databases using Blast (Altschul et al. 1990). Accuracy of identification was determined by the lowest *e*-values and number of matching peptides per protein.

RNA-Seq

cDNA was synthesized from the total RNA samples using the First-Strand cDNA Synthesis Kit (NZYTech). Primers were designed using Primer Blast and verified *in silico* with Oligo Analyser (Table 5.1) to amplify an expressed sequence tag (EST) for mRNAs coding for proteins identified by MS/MS and RNA-Seq (Fig. 5.2). The calibrator was GAPDH, as suggested (Thiel et al. 2017). Amplification was done in a Gradient Thermocycler96 (Biometra). Following resolving PCR products in an agarose gel, these were Sanger-sequenced (as a service), translated and matched against 1) UNIPROT cluster UniRef90 (Suzek et al. 2015) by generating a customized toxin database, with BLASTP v2.5.0 (States and Gish 1994), having set a maximum *e*-value of 10^{-5}; 2) NCBInr using BlastP. The results from RNASeq were validated by RT-qPCR. The RT-qPCR was then done in a Rotor-Gene 6000 thermal cycler (Corbett Research) using the NZY qPCR Green Master Mix (NZYTech), according to manufacturer instructions. The programme was set as: denaturation 94 ºC, 45 seconds (s), annealing 55 ºC, 25 s and extension 72 ºC, 30 s, with 45 cycles per run. Primer melting analysis was also conducted to verify the specificity of hybridization. Relative expression was determined by the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen 2001). The peptide sequences were contrasted against *E. viridis* proboscis transcriptome products (as described in Chapter 3), deposited in Gene Expression Omnibus (GEO) under the accession number GSE143954 (Fig. S5.1).

Table 5.1. Sequences of primers used in RT-qPCR. Primers were designed using primer blast (www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi, National Center for Biotechnology Information) and

their quality assessed using OligoAnalyzer Tool (www.idtdna.com/pages/tools/oligoanalyzer, Integrated DA Technologies) and optimized by PCR.

Target	Forward primer (5'-3')	Tm (°C)	Reverse primer (5'-3')	Tm (°C)	Amplicon
Gap_dh	CAGGGGTGCTAAGCAGTTGG	60.96	GAGAAGGCTGGGGCTCATTT	60.03	165
Rab3	CAGTTGCTCTCATCCACGGT	60.04	AGCCTGGAATGGAAGTCACG	60.04	136
14-3-3	GTCCGTTGTTTGACACGCTC	60.04	GACCCACAACTCCTACTGGC	60.04	155
ArsB	GACACTGGCGACCTCTACTG	59.83	GGTAGCTCTCCTTTGTGCGT	60.04	155
SP	ACGGGGTTGCTTAATTCGGT	59.96	AGCCACACTAATGCGAACGA	60.04	158
Hsp70	TCCCGTACTCTTGTCCACAG	58.75	AGCGTGCCATGACAAAGGAT	60.32	152

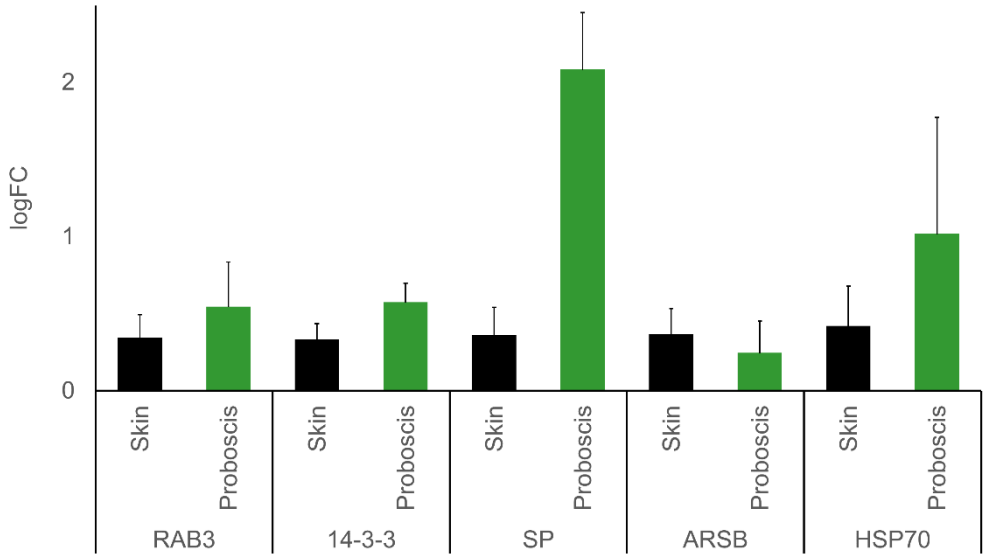


Figure 5.2. Expression analysis of key proteins in the mucus extract by RT-qPCR, comparing the proboscis and body wall. Data are expressed as mean relative expression of *Rab3*, *14-3-3* protein, *ArsB* (Arylsulfatase B), *SP* (Serine protease) and *Hsp70* (Heat-shock protein 70 kDa). The housekeeping gene Gap-dh was used as calibrator.

5.3.2. Network analysis

After the identification of the main proteins by combining proteomics and transcriptomics, *Homo sapiens* homologues were retrieved using BLASTP. Gene network analyses were conducted using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) by setting the confidence cut-off for interaction links between proteins at 0.400 (Szklarczyk et al. 2019).

5.3.3. Cell-based assays

HCT116 (human colorectal carcinoma), A549 (lung adenocarcinoma), Fibroblasts, MCF7 (breast carcinoma) and K562 (myelogenic leukemic carcinoma) cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Grand Island, NY, USA) supplemented with 10% (v/v) fetal bovine serum and 1% (v/v) antibiotic/antimycotic solution (Invitrogen), while A2780 (ovarian carcinoma) cells were grown in modified McCoy's medium: Roswell Park Memorial Institute medium (RPMI) 1630 (Ma et al. 2016; Czerwińska et al. 2017). All cell lines were purchased from ATCC (http://www.atcc.org), except for A2780 that was purchased from Sigma-Aldrich (http://www.sigmaaldrich.com). All cell cultures were maintained at 37 °C in a humidified atmosphere containing 5% (v/v) CO_2 and incubated in the same atmosphere when necessary. Only purified extracts stored for less than 30 days (-80 ºC) were used in biological assays and all dilutions (and controls) were prepared in PBS. The same volume of PBS was added to the respective medium of each assay (controls). Total protein content was determined prior to biological assays using a Nanodrop spectrophotometer. The assays were conducted with cells seeded for 24 h and in the range of 10^5 cells. mL^{-1}, in the conditions described before, in 6, 24 or 96-well plates, depending on the endpoint.

Cell viability

Cell viability was determined by subjecting all cell lineages to successive dilutions of the purified protein extract (2×10^{-3} to 1.6 µg. μL^{-1}) for the estimation of and the half-maximal inhibitory concentration (IC_{50}). Cells were incubated for 48 h with each diluted extract. Viability was determined with the CellTiter 96 AQueous Non-Radioactive Cell Proliferation Assay (Promega, Madison, WI, USA), using 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2Htetrazolium, inner salt (MTS). Quantification of the formazan dye formed in viable cells was done spectrophotometrically by measuring absorbance at 490 nm with a Bio-Rad microplate reader Model 680 (Bio-Rad, Hercules, CA, USA) (Ma et al. 2016; Czerwińska et al. 2017; Almeida et al. 2019). The assays were performed in triplicate and technical duplicates were included, to minimize technical error. Doxorubicin (0.4 mM) was used as a positive control in all assays.

Apoptosis

Apoptosis was determined based on the Hoechst 33258 labelling assay. Cells were seeded on sterilized coverslips and incubated for 24 h to obtain an adherent monolayer. Cells were then incubated with the purified extract for 12 h (IC_{50}), 24 h (IC_{50}) and 48 h (1/10 IC_{50}, 1/5 IC_{50} and IC_{50}). After incubation, cells were washed with PBS and fixed with fresh 4% (m/v) formaldehyde (reconstituted from paraformaldehyde), for 10 min in the dark. Cells were then incubated with the dye (provided by Sigma, Missouri, USA), prepared as 5 µg. mL^{-1} in PBS. After 10 min, slides were washed with PBS and mounted. Fluorescent nuclei (100 per sample) were sorted in a DM

2500 LED model microscope, adapted for epifluorescence with an external EL6000 light source (Leica Microsystems). The assay was performed in triplicate for each timeline. The results are expressed as the mean percentage of apoptotic nuclei.

Autophagy

Cells were exposed to the extracts for 12 h, 24 h and 48 h, at the same concentrations than the previous assay. Rapamycin (50 mM) was used as a positive control (Almeida et al. 2019; Choroba et al. 2019). By the end of each assay, the medium was removed, and cells were stained with the CYTO-ID Autophagy Detection Kit (ENZO, New York, NY, USA), following manufacturer instructions. Stained cells were imaged with the aforementioned epifluorescence microscope following counterstaining with DAPI. The assay was performed in triplicate for each timeline.

Caspase 8 assay

The activity of Caspase 8 was determined to assess the extrinsic apoptotic pathway using the Caspase 8 assay kit from Abcam, following manufacturer instructions. Briefly, cells were plated on T-flasks with an area of 25 cm^2 with 6×10^5 cells. mL^{-1} for 24 h and exposed to the purified extract (IC_{50} dose), plus PBS (control) and incubated for 48 h. Cells were then scraped, washed in PBS, resuspended in cell lysis buffer and incubated on ice for 30 min. Total protein in lysates was quantified as previous and adjusted to 200 µg. 50 μL^{-1} buffer. Each microplate well was loaded with 200 µg protein. 50 μL^{-1} 50 µL of 2× Reaction Buffer with DTT (100 mM final concentration) and 5 µL of IETD-pNA substrate (4 M). After incubation (2 h), absorption at 400 nm was quantified on a 680 Microplate Reader (Bio-Rad, Hercules, CA, USA).

Detection of Reactive Oxygen Species

Intracellular reactive oxygen species (ROS) were determined using 2,7-dichlorodihydrofluorescein diacetate (H_2DCF-DA, from Sigma-Aldrich), as previously described (Das et al. 2018). Briefly, cells were seeded and incubated for 24 h, after which the culture medium was removed and replaced by fresh medium containing 1/10 IC_{50}, 1/5 IC_{50} and IC_{50} (0.08 µg. μL^{-1}), plus controls. Hydrogen peroxide (50 µM) was used as a positive control. Cells were harvested after 48 h, washed twice with PBS before resuspension in pre-warmed (37 °C) PBS containing 100 µM H_2DCF-DA and incubation at 37 °C for 20 min, in the dark. Fluorescence intensity was measured on an Eclipse Ti inverted microscope equipped with a DS-QiMc camera and adapted for epifluorescence (all from Nikon), with fixed exposure time for all samples. The green fluorescent signal was normalized and analysed using ImageJ (Schneider et al. 2012).

Mitochondrial transmembrane potential

The mitochondrial transmembrane potential was analysed using the bis(5,6-dichloro-1,3-diethyl-2-benzimidazole)trimethinecyanine iodide (JC-1) dye (Abnova Corporation, Walnut, CA, USA). After seeding and incubation cells were treated with the protein extract at the concentrations of 1/10 IC_{50}, 1/5 IC_{50} and IC_{50} diluted in fresh medium and exposed for 48 h. Cells were then stained with the JC-1 staining solution for 20 min at 37 °C in the dark. Image acquisition and analysis was done as above, based the ration between green and red fluorescence.

Plasmatic transmembrane potential

The plasmatic transmembrane potential was analysed using 2′,7′-Bis(2-carboxyethyl)-5(6)-carboxyfluorescein tetrakis(acetoxymethyl) ester (BCECF-AM) dye (Invitrogen). After seeding and incubation, cells were treated with the purified protein extract at the concentration of IC_{50}, diluted in fresh medium (RPMI), and incubated for 1 h and 3 h. Cells were then trypsinised and washed with PBS and RPMI medium (without phenol red) before incubation with BCECF-AM (2 µM in the same medium) for 20 min, at 37 ºC. After washing in the culture medium, data was collected using an Attune acoustic focusing flow cytometer (ThermoFisher Scientific) and analysed using FCS Express 6 Flow Cytometry (De Novo Software). The results were normalized for cell counts and the respective control. The assays were performed in duplicate and technical duplicates were included to minimize technical error. Doxorubicin (0.4 mM) was used as a positive control in all assays.

Interference of cellular pH or membrane integrity

After seeding and incubation, cells were treated with the protein extract at the concentration of IC_{50} diluted in fresh RPMI medium for 1 h and 3 h. After washing with RPMI (without phenol red), cells were incubated with BCECF-AM (µM) for 20 min, washed again and incubated with propidium iodide (2.5 µg. mL^{-1}) for 10 min. Cells were then washed and visualized in RPMI (without phenol red) under an Eclipse Ti fluorescence microscope, equipped with a DS-QiMc camera (Nikon Corporation).

Western-Blot

Cells were plated on T-flasks with an area of 75 cm^2 with 4×10^5 cells. mL^{-1}. After incubation for 24 h to allow adhesion, cells were treated with fresh medium supplemented with purified extract at the concentration of IC_{50} or PBS (control) and incubated for 48 h. The cell monolayer was washed three times with PBS and the cells were detached by scraping. The cells were pelletised by centrifugation at 500×g for 5 min (4 °C). The following procedure was adapted from Vinhas et al. (2017), using a PVDF membrane (GE Healthcare). Signal acquisition was obtained with Hyperfilm ECL (GE Healthcare). The protein band intensities relative to controls was quantified using ImageJ (Schindelin et al. 2009) following normalization to β-actin values.

Cell cycle analysis

Cells were seeded in a 6-well plate at 1×10^5 cells per well and synchronized in early S-phase using a thymidine double block (Almeida et al. 2019). After the second block, the medium was replaced with fresh medium containing the extract or control (PBS only). The cells were collected by trypsinization after 1 h, 2 h, 3 h, 4 h, 9 h and 24 h. After centrifugation at 650×g for 5 min at 4 °C, pelleted cells were washed with 1 mL cold PBS and centrifuged at 3000×g for 5 min at 4 °C. The cells were then suspended in 100 µL cold PBS and completely disaggregated by soft pipetting before gently adding 1 mL ice-cold ethanol (80% (v/v)). After at least 16 h at 4 °C, the cells were pelletized by centrifuging at 5000×g for 10 min at 4 °C and the supernatant discarded. The cells were treated for 30 min with RNase A 50 µg. mL^{-1} at 37 °C and propidium iodide was added to a final concentration of 25 µg. mL^{-1}. Data was collected using an Attune acoustic focusing cytometer (ThermoFisher Scientific) and analysed using FCS Express 6 Flow Cytometry (De Novo Software).

Determination of cytoskeletal alterations

Cells were seeded for 24 h in 24-well plates at 1×10^5 cells. mL^{-1} for 3 h, 6 h and 12 h and at 0.375×10^5 cells. mL^{-1} for 48 h. Cells were then fixed in 4% (v/v) formaldehyde, permeabilized with 0.1% (v/v) Triton X-100 and blocked with 1% (w/v) bovine serum albumin (BSA). The cells were probed with Alexa Fluor 488 Phalloidin (2.4% (v/v) in 1% (w/v) BSA, 20 min) to stain actin filaments, followed by Hoechst 33258 (Sigma, Missouri, USA) at 5 µg. mL^{-1} in PBS (15 min) as the nuclear counterstain. Alpha tubulin was tagged with the DM1A anti-α-tubulin monoclonal antibody (Sigma), diluted to 1:200 in 1% (w/v) BSA. After incubation o/n at 4 ºC, the primary antibody was tagged with anti-mouse TRITC-conjugated polyclonal secondary antibody (Sigma), diluted to 1:64 in 1% (w/v) BSA (1 h). Cells were then washed with Tris-buffered saline (pH 7.6) with Tween 20 (TBST) and mounted with PBS. The cells were then examined with an Eclipse Ti fluorescence microscope (Nikon), as described previously.

Cytology

Cells, seeded in 24-well plates with a density of 10^6 cells. mL^{-1}, were exposed to different concentrations of the purified extract (IC_{50}, 1/5 IC_{50}, 1/10 IC_{50}), plus controls (PBS), for 48 h and additionally to IC_{50} for 1 h and 3 h. After exposure, cells were scraped and processed for transmission electron microscopy (TEM) as described in Vogt et al. (Vogt et al. 2015), with modifications. Briefly, cells were fixed in 2.5 % (v/v) glutaraldehyde in 0.1 M phosphate buffer and post-fixed in 2% (w/v) osmium tetroxide in 0.1M phosphate buffer, overnight. Pellets were washed in ultrapure water and dehydrated in a progressive series of acetone. Pelletized cells were embedded in Epon resin (Sigma-Aldrich) using propylene oxide as intermediate. Thin sections

(50-60 nm) were obtained with an Ultracut E ultramicrotome Leica. Sections were collected onto copper and nickel mesh grids and contrasted with 2% (w/v) aqueous Uranyl Acetate and Reynold's Lead Citrate (Venable and Coggeshall 1965). Grids were analysed using a JEOL 100-SX model TEM operated at 80 kV.

Statistical analysis

Normality and homoscedasticity were evaluated using Shapiro-Wilk's and Levene's tests, respectively. When assumption were met for parametric analyses, we employed ANOVA based on *F* and Dunn's tests for multiple comparisons, otherwise the non-parametric Kruskal-Wallis *H* statistic was used, followed by Student's *T*-test and Mann-Whitney's *U* (parametric and non-parametric tests, respectively) for paired-sample tests. Statistics were computed using R 3.5 (Ihaka and Gentleman 1996). The significant level was set at 0.05 for all the analyses.

5.4. Results

5.4.1. Protein identification

Five major proteins were conclusively identified in *E. viridis* mucosecretions by superimposing proteomics and transcriptomics (Table 5.2 and Fig. 5.1). The results identified Arylsulfatase B (ARSB), Heat shock protein 70 KDa (HSP70), 14-3-3 protein, RAB3 and serine protease (SP) as the main proteins in extracts. When comparing the relative expression of these proteins in the proboscis and body wall, serine protease is the protein with higher expression in the proboscis, followed by ARSB, RAB3 and lastly HSP70 and 14-3-3 protein. The peptide sequences had a 100% match with the translated mRNAs, with the exception of SP, which yielded multiple transcriptional variants (Fig. S5.1).

Table 5.2. Matched peptidic sequences and translated mRNAs upregulated in the proboscis and respective contrasting against Pfam and Toxins databases.

Protein	Method	logFC	Conserved Domain (Pfam)	Protein match	e-value (blastp)
SP	RNAseq	11.85	Trypsin-like serine protease	Serine protease 33-like [Orbicella faveolata]	6×10^{-54}
	MS/MS		-	Serine protease [Protobothrops flavoviridis]	6×10^{-5}
	MS/MS		-	Transmembrane protease serine 7 [Ophiophagus hannah]	7.8
ARSB	RNAseq	5.08	ALP P like superfamily	Arylsulfatase B [Penaeus vannamei]	2×10^{-83}
	MS/MS		-	N-acetylgalactosamine 6-sulfate sulfatase [Candidatus Moranbacteria bacterium]	19
RAB3	RNAseq	3.96	Rab GTPase family 3	Rab3 [Doryteuthis pealeii]	2.00×10^{-147}
	MS/MS		Rab GTPase family 3	Ras family protein [Trichuris suis]	1.00×10^{-06}
	MS/MS		-	Ras-related protein Rab-3 [Melipona quadrifasciata]	0.45
14-3-3	RNAseq	1.53	14-3-3 superfamily	14-3-3 zeta [Pristionchus pacificus] (38kDa)	2.00×10^{-121}
	MS/MS		-	14-3-3-like protein 4 [Pseudotsuga menziesii]	0.21
HSP70	RNAseq	1.09	Heat shock 70 kDa protein	HSPA8 protein [Homo sapiens]	0
	MS/MS		HSP70	Heat shock cognate 71 kDa protein [Cricetulus griseus]	3.00×10^{-15}

5.4.2. Cytotoxic effects

The purified extract caused higher antiproliferative effect towards several human cancer cell lines, compared to normal human primary fibroblasts (Fig. 5.3, Table 5.3). Most cancer cells presented a loss of cell viability when exposed to the concentration of the extract within the range of 0.19 to 0.42 µg. μL^{-1}, with the exception of the ovarian carcinoma cell line (A2780), which yielded an IC_{50} value one order of magnitude lower (0.08 µg. μL^{-1}). Consequently, the ovarian carcinoma cancer cell line was then chosen for further biological characterisation towards addressing the mechanisms underlying the loss of A2780 cell viability.

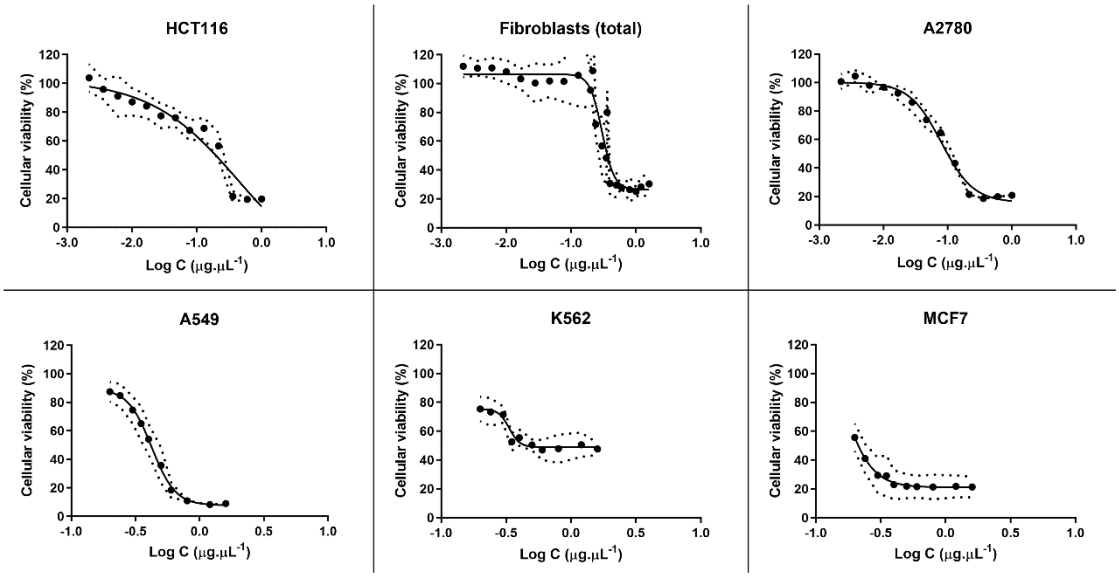


Figure 5.3. Cell viability determined by the MTS assay. Cells were exposed for 48 h to purified proteins extracts corresponding to concentrations between 0.0022 and 1.6 μg. μL^{-1}, plus controls (PBS only). Measurements are indicated as means ± SEE of three independent replicates. A positive control (doxorubicin) was used in all assays.

Table 5.3. Half maximal inhibitory concentration (IC_{50}) determined through the cell viability assay (MTS assay). Values for the IC_{50} determined for each cell line are indicated as means ± SEE of three independent replicates.

Cells	IC_{50} ± SEE
A2780	0.08 ± 0.01
HCT116	0.24 ± 4.00
A549	0.46 ± 0.05
MCF7	0.15 ± 0.04
K562	0.37 ± 0.04
Fibroblasts	0.29 ± 0.02

The frequency of apoptotic A2780 cells approximately doubled at IC_{50} concentrations, for both 24 h and 48 h (Fig. 5.4). However, after 12 h, no significant differences were found when the cells were subjected to lower concentrations. Since apoptosis accounted only for a maximum of 35% cell death, an additional mechanism of programmed cell death, namely via autophagy, was assessed (Fig. 5.5). Interestingly, a significant increase in autophagic cell death was observed after 12 h, 24 h and 48 h of exposure. To further confirm the induction of programmed cell death TEM analyses were performed (Fig. 5.6). Apoptotic bodies were easily noticed as early as 1 h and 3 h of exposure, as well as chromatin condensation and nuclear fragmentation by blebbing, in accordance with Hoechst labelling (Fig. 5.4). Moreover, autophagolysosomes, associated with the autophagic process, were noticed, especially after 24 h, also supporting the previous

fluorescence microscopy results (Fig. 5.5). Exposure to higher concentrations (IC_{50}), revealed further cellular alterations, namely mitochondrial shape changes, nuclear pleomorphisms and an increase in lysosome number.

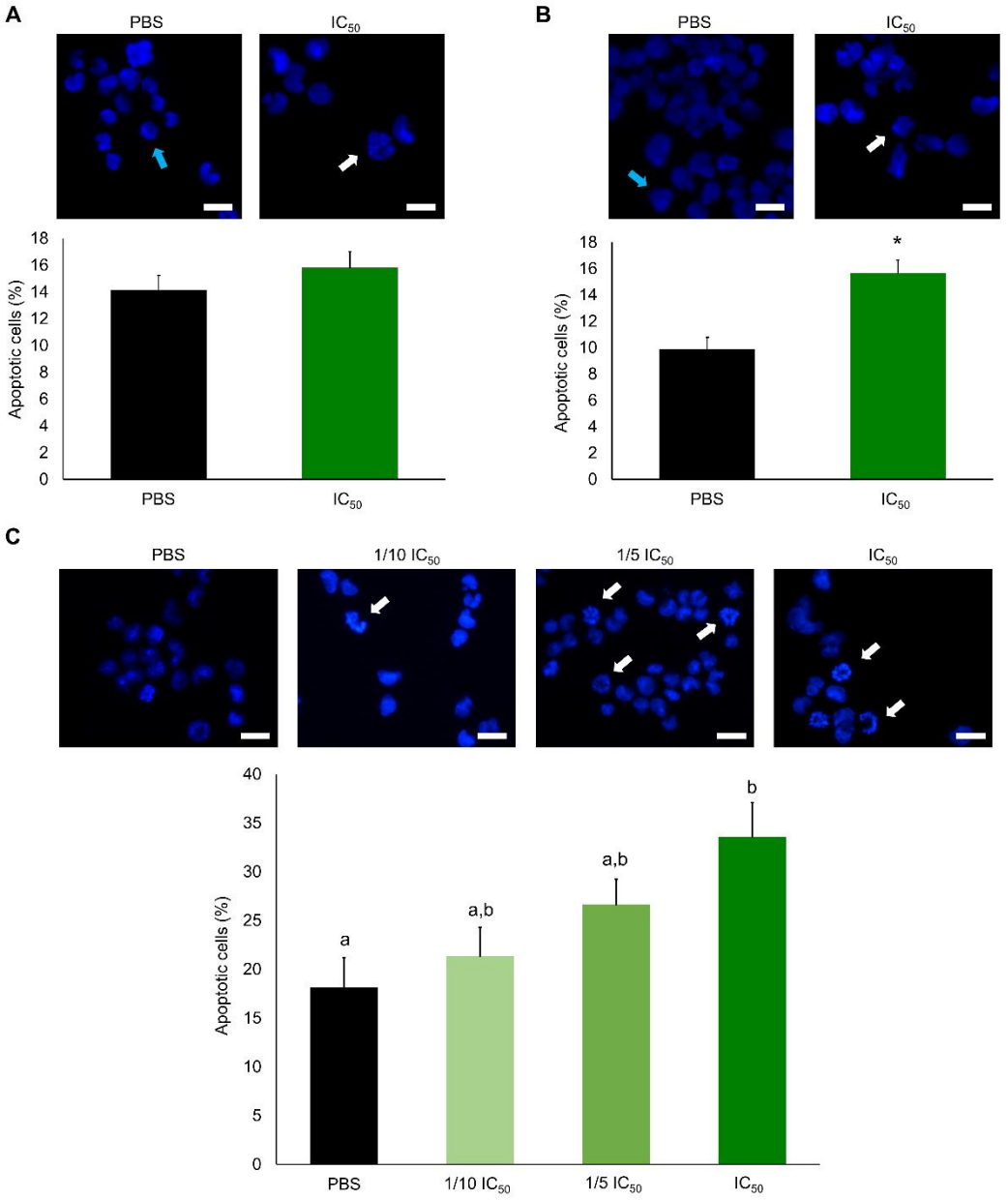


Figure 5.4. Determination of apoptosis in the A2780 cell line exposed for 12 h (A) and 24 h (B) to purified extracts corresponding to the IC_{50} threshold (0.08 µg. μL^{-1}) and exposed for 48 h (C) between 1/10 IC_{50} (0.08 µg. μL^{-1}) and IC_{50}. Blue arrows are indicative of normal nuclei and white arrows of apoptotic nuclei. *Indicates significant differences towards control (Student's *t* test, $p<0.01$). Different letters are indicative of significant differences (Dunn's test, $p<0.05$).

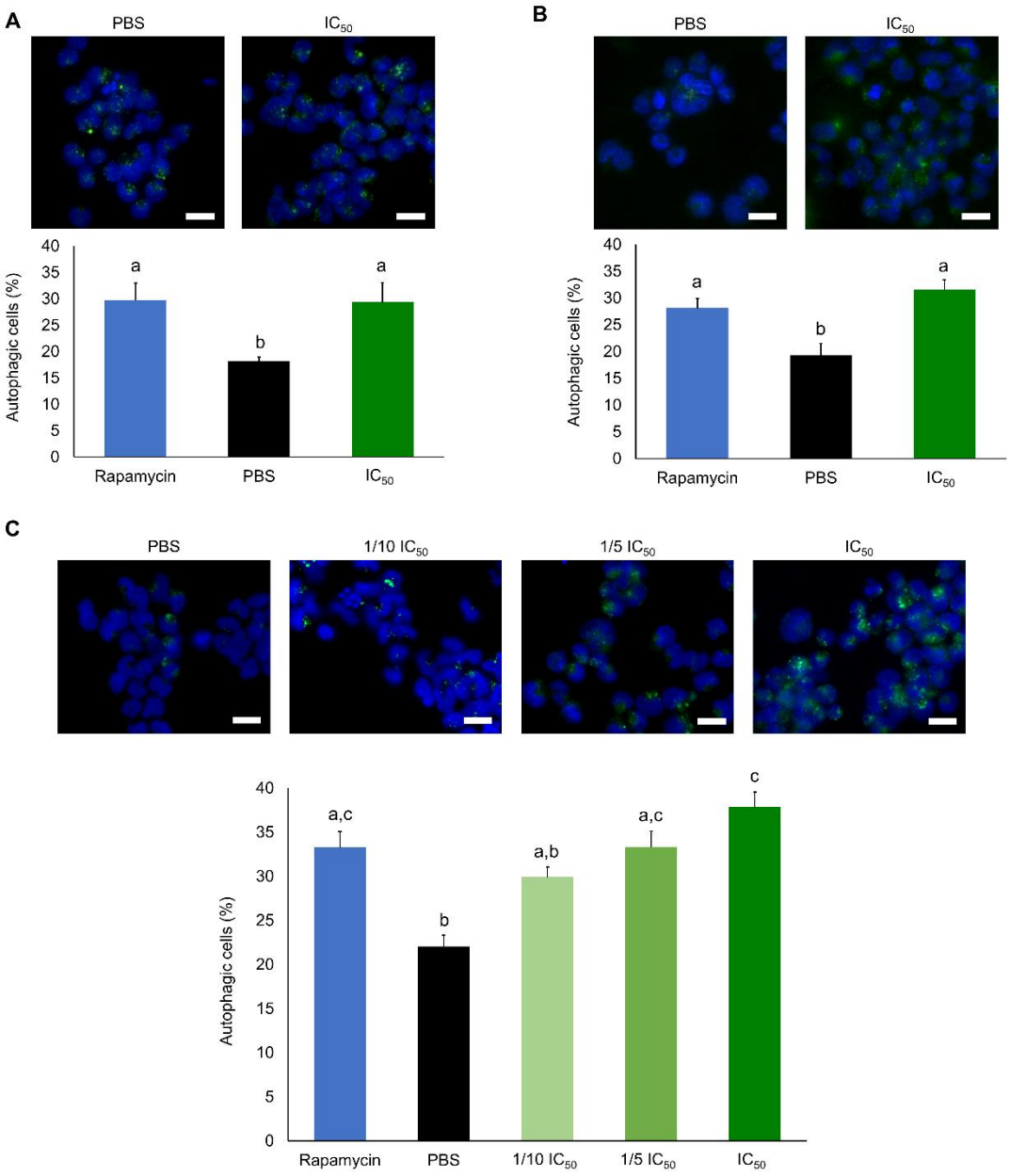


Figure 5.5. Determination of autophagy in A2780 cell line exposed for 12 h (A) and 24 h (B) to purified extracts at IC_{50} (0.08 µg. $µL^{-1}$) and exposed for 48 h (C) to purified extract between 1/10 IC_{50} (0.08 µg. $µL^{-1}$) and IC_{50}. A positive control was used (rapamycin) in all assays. Different letters are indicative of significant differences (Dunn's test, $p<0.05$).

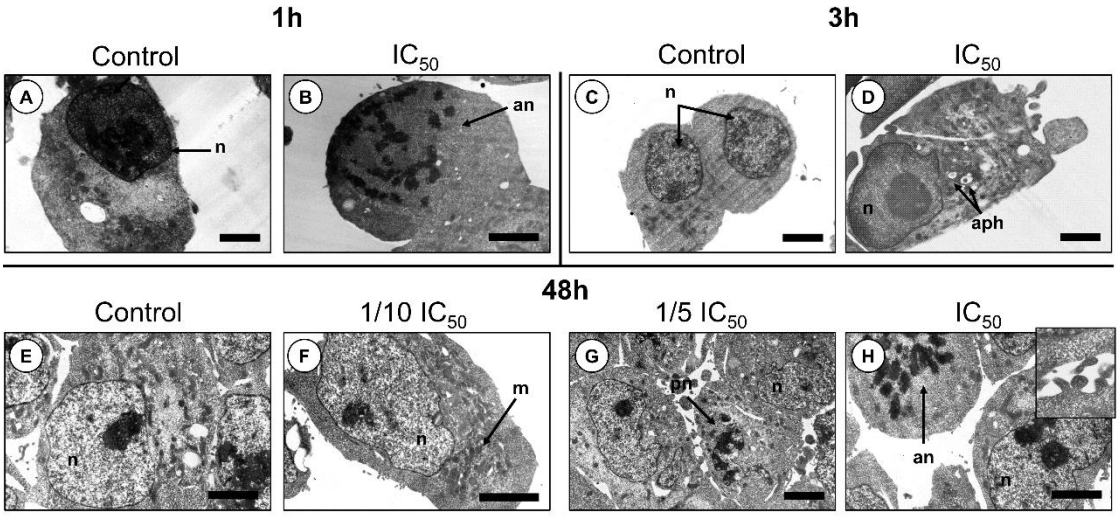


Figure 5.6. Changes in cell ultrastructure revealed by transmission electron microscopy (TEM) applied to A2780 cells. Cells were exposed for 1 h and 3 h to the IC_{50} concentration of purified extracts and respective control (PBS) and 48 h to 1/10 IC_{50}, 1/5 IC_{50} and IC_{50}). Apoptotic nuclei (an) were visible after 1 h of exposure, whereas autophagosomes (aph) became evident after 3 h. After 48 h of exposure, nuclear pleomorphisms (pn) were more obvious, as well as an increase in the number of mitochondria, regardless of apoptotic or and autophagic cells, when compared with controls. *Inset*: membrane detail of cells exposed to IC_{50} dosage during 48 h. Nuclei (n). Scale bars: 2 µm.

For further clarification of the mechanism underlying apoptosis induction, the expression of BAX and BCL-2, the production of ROS, the mitochondrial membrane potential ($\Delta\Psi m$) and the induction caspase 8 were analysed in A2780 cells exposed to the IC_{50} of extracts. The low BAX/BCL-2 expression ratio (Fig. 5.7A, B), the absence of ROS production and the unchangeable $\Delta\Psi m$ in A2780 cells exposed to the IC_{50} concentration of the extracts (Fig. S5.2), and the significant induction of caspase 8 activity (Fig. 5.7C) clearly suggests the activation of the extrinsic apoptotic pathway via activation of death receptors. However, alterations to plasma membrane potential ($\Delta\Psi p$) were noticeable after 1 h and 3 h of exposure, as determined by flow cytometry and fluorescence microscopy (Fig. 5.8A, B).

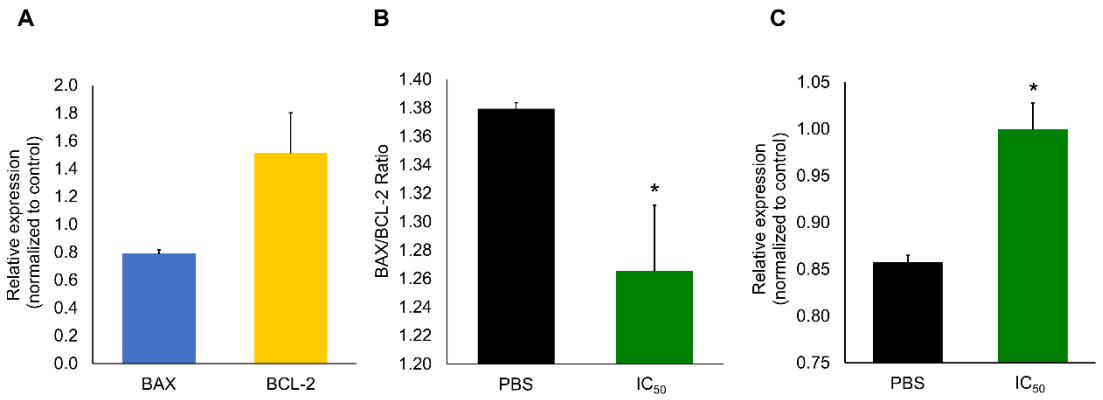


Figure 5.7. Expression of BAX, BCL-2, their ratio and Caspase 8 activity in A2780 cells exposed to the IC_{50} of purified extracts for 48 h (0.08 µg. µL^{-1}). (A) Relative expression of BAX and BCL-2, normalized to control. (B) Apoptotic index estimated from the BAX/BCL-2 ratio, after incubation with PBS (control) and extract (IC_{50}). (C) Caspase 8 activity normalized to control. * Indicates significant differences to control (Kruskal-Wallis *H*, $p<0.05$).

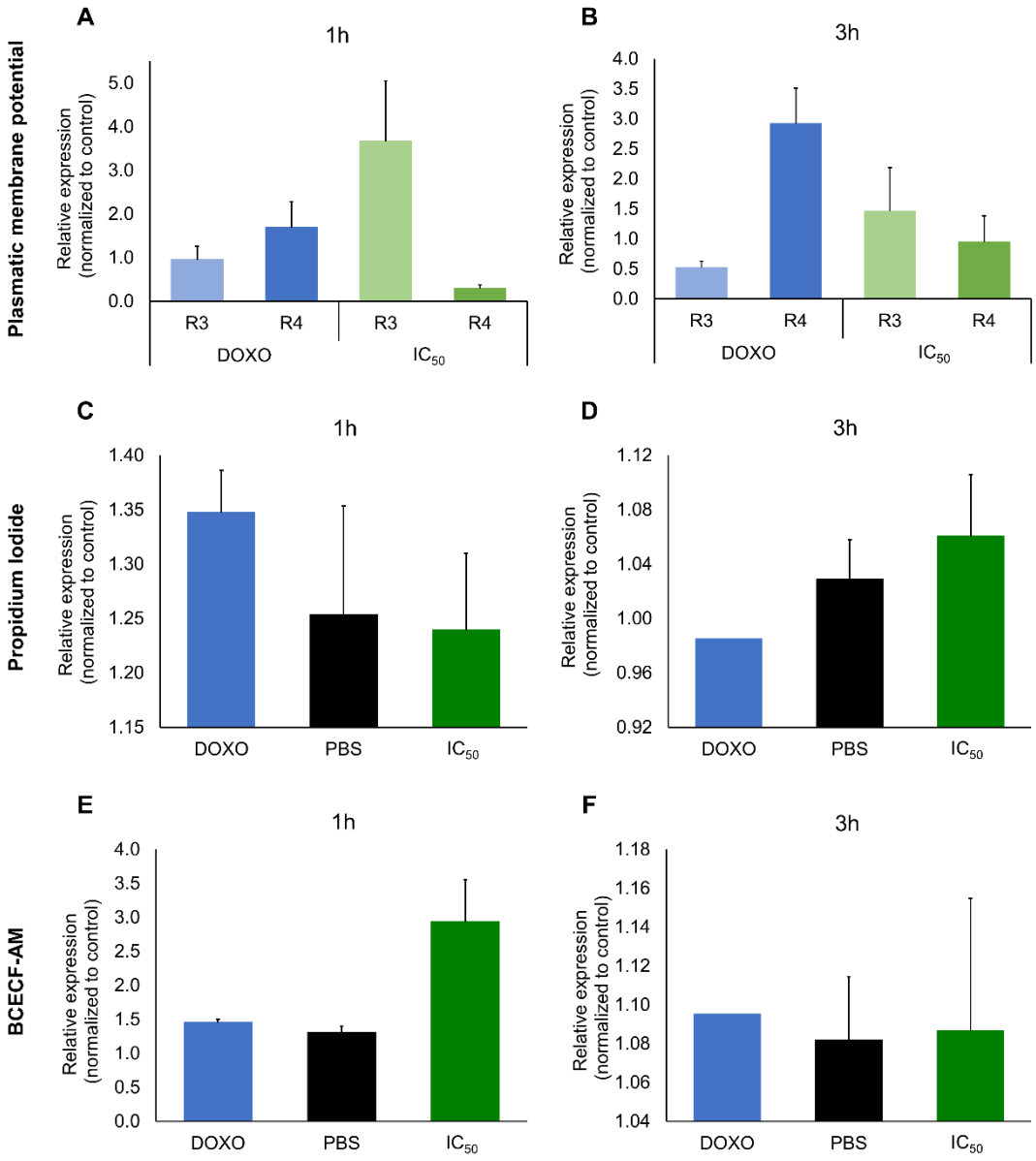


Figure 5.8. Plasmatic membrane potential (A, B) and fluorescent staining assays (propidium iodide for nucleic acid staining and BCECF-AM for plasmatic membrane staining), (C-F) in A2780 cells exposed to IC_{50} extract concentration (0.08 µg. $µL^{-1}$), and the respective control (PBS) for 1 h and 3 h.

The effect of the extracts in A2780 cell cycle progression was evaluated over time of exposure (Fig. 5.9). Cell cycle arrest at phase S was observed 2 h after exposure (Fig. 5.9B), becoming significant after 3 h and 4 h exposure (Fig. 5.9C, D). A slight delay in the percentage of cells entering G2/M phase was also verified after 2 h exposure and extended until 4 h exposure. Alterations in tubulin were also seen with immunohistochemical staining after 3 h exposure. After 4 h exposure, an overcompensation of cell in the phase G2/M started to be noticed, specially at 12 h exposure (Fig. 5.9F). The changes in the early hours were consistent with quantification of both β-actin at 1 h and α-tubulin at 3 h (Fig. 5.8C-F).

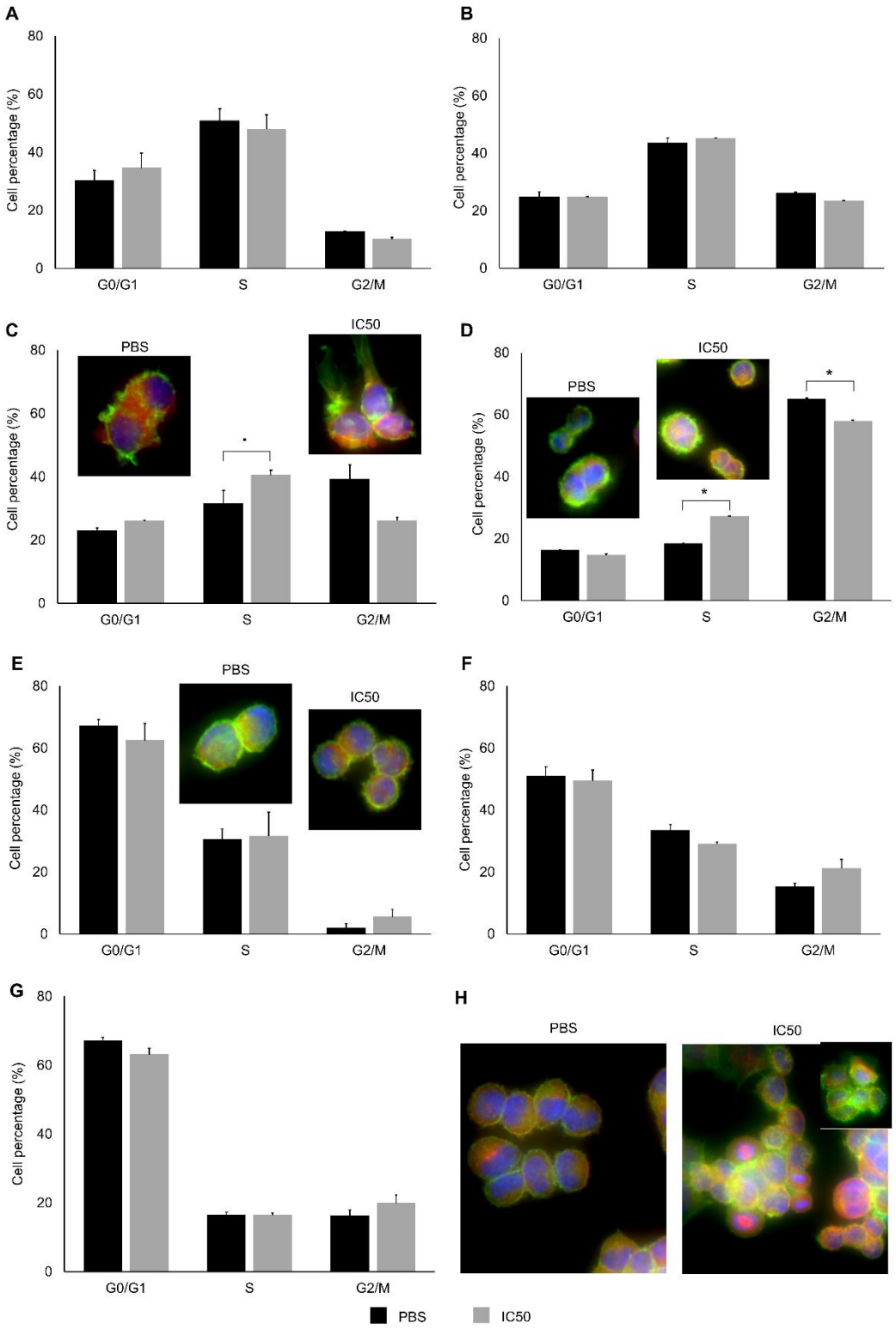
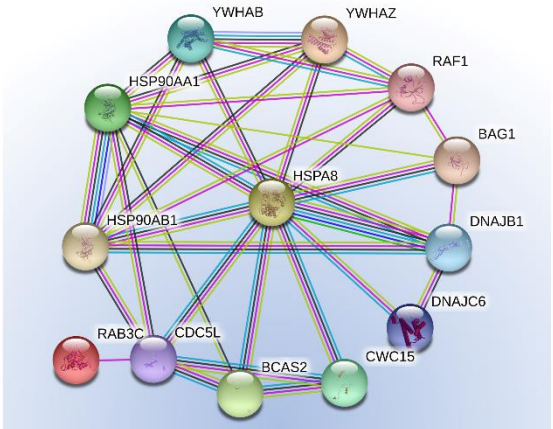


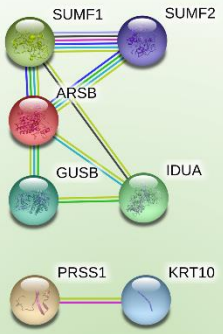
Figure 5.9. Cell cycle analysis and immunofluorescence assay at 1 h (A), 2 h (B), 3 h (C), 4 h (D), 9 h (E), 12 h (F), 24 h (G) and 48 h (H), for A2780 cell line exposed to the extract (IC_{50}, 0.08 µg. $µL^{-1}$) and the respective control (PBS). The cell cycle was analysed by flow cytometry. The cells for the immunofluorescent assay were probed with F-actin (green) and α-tubulin (red) and counterstained with Hoechst dye 33258 (blue). * Indicates significant differences to control (Student's *t*-test, $p<0.05$). ˙ Indicates differences to control ($p<0.1$).

5.4.3. Network analysis

The interaction between the human homologs of the main proteins present in extracts is depicted in Fig. 5.10. The first network yielded interaction between 14-3-3 zeta protein (YWHAZ), heat shock cognate 71 kDa protein (HSPA8) and Ras-related protein Rab-3C (RAB3C). YWHAZ and HSPA8 yielded the strongest connection with reports of co-expression in humans and other mammals. HSPA8 and RAB3C were also interacting indirectly, trough cell division cycle 5-like protein (CDC5L), a DNA-binding protein involved in cell cycle control. The connection of HSPA8 and RAB3C and CDC5L have been both supported by experimental data. HSPA8 and CDC5L were also connected with the spliceosome-associated protein CWC15 homolog (CWC15) and Pre-mRNA-splicing factor SPF27 (BCAS2), important proteins that are involved in the spliceosome as well. Due to the lack of direct interactions between PRSS1 and ARSB with the remaining proteins and each other, the radius of the network was expanded to ten proteins. Individually, Arylsulphatase B (ARSB) is interacting mostly with glycoside hydrolase domain-containing proteins (IDUA, GUSB) and with enzymes responsible for catalysing the hydrolase of sulphate esters (SUMF1 and 2), while PRSS1 only interacts with keratin (KRT10).



	GO-term	Description	Count in gene set	False discovery rate
Biological function	GO:1900034	Regulation of cellular response to heat	5 of 45	9.21E-08
	GO:0051085	Chaperone cofactor-dependent protein refolding	3 of 16	8.23E-05
	GO:0006457	Protein folding	5 of 214	8.23E-05
Molecular function	GO:0019904	Protein domain specific binding	7 of 706	1.75E-05
	GO:0051082	Unfolded protein binding	4 of 106	4.86E-05
	GO:0019899	Enzyme binding	9 of 2197	6.14E-05
Cellular component	GO:0042470	Melanosome	5 of 105	1.02E-06
	GO:0000974	Prp19 complex	3 of 12	5.56E-06
	GO:0071007	U2-type catalytic step 2 spliceosome	3 of 29	4.51E-05
KEGG pathway	hsa04141	Protein processing in endoplasmic reticulum	5 of 161	4.47E-06
	hsa04151	PI3K-Akt signalling pathway	5 of 348	6.36E-05
	hsa03040	Spliceosome	4 of 130	6.36E-05



	GO-term	Description	Count in gene set	False discovery rate
Biological function	GO:0006027	Glycosaminoglycan catabolic process	3 of 62	0.00024
	GO:1903510	Mucopolysaccharide metabolic process	3 of 109	0.00042
	GO:0030207	Chondroitin sulfate catabolic process	2 of 14	0.00065
Molecular function	GO:0004553	Hydrolase activity, hydrolysing O-glycosyl compounds	2 of 99	0.0191
Cellular component	GO:0043202	Lysosomal lumen	3 of 94	0.00025
KEGG pathway	Hsa04142	Lysosome	4 of 123	5.16E-07
	Hsa00531	Glycosaminoglycan degradation	3 of 19	5.16E-07

Figure 5.10. A network analysis between human protein homologs (HSPA8, RAB3C, YWHAZ, ARSB and PRSS1) using STRING. The two autonomous protein interaction networks were expanded to allocate additional ten proteins to increase the radius of interactions beyond the proteins identified in the present work. The confidence cut-off for showing interaction links has been set to medium (0.400). ARSB - Arylsulfatase B; BAG1 - BCL-2 associated athanogene family molecular chaperone regulator 1; BCAS2 - Pre-mRNA-splicing factor SPF27; CDC5L - Cell division cycle 5-like protein; CWC15 - Spliceosome-associated protein CWC15 homolog; DNAJB1 - DnaJ homolog subfamily B member 1; DNAJC6 - Putative tyrosine-protein phosphatase auxilin; GUSB - Beta-glucuronidase; HSPA8 - Heat shock cognate 71 kDa protein; HSP90AA1 - Heat shock protein HSP 90-alpha; HSP90AB1 - Heat shock protein HSP 90-beta; IDUA - Alpha-L-iduronidase; KRT10 - Keratin, type I cytoskeletal 10; PRSS1 - Trypsin-1; RAB3C - Ras-related protein Rab-3C; RAF1 – RAF proto-oncogene serine/threonine-protein kinase; SUMF1 - Sulfatase-modifying factor 1; SUMF2 - Sulfatase-modifying factor 2; YWHAB - 14-3-3 protein beta/alpha; YWHAZ - 14-3-3 protein zeta.

5.5. Discussion

The mucus secreted by the marine worm *Eulalia* contains a mixture of toxin and non-toxin proteins involved in multiple functions for attacking prey and defence against predators and environmental variables (as described in this chapter and chapter 3). This cocktail of proteins revealed higher toxicity towards the A2780 ovarian cancer cell line, compared to other human cancer cells and even normal fibroblasts, with a strong indication for a dose- and time-dependent responsiveness. We hypothesized that combining proteomics with transcriptomics would shed light on the main proteins involved in cytotoxicity and offer insights on their mode-of-action against A2780 cells, which is seemingly associated to cell cycle arrest and the induction of programmed cell via extrinsic apoptosis and autophagy. Five proteins were cross-validated by RNA-Seq, whose function, isolated or combined, are responsible for the specific effects in A2780 cells.

The cytolytic properties of animal venoms and poisons toward cancer cells has been investigated with promising results, albeit scant comparisons between the differential effects against normal and cancer cells. Bee venom has arguably become one of the most studied and promising cases, with melittin (a 26-amino acid lytic peptide responsible for the painful sting) and phospholipase A being, so far, pinpointed as key substances involved in the promotion of apoptosis (Oršolić 2012). Venoms from snakes also received attention, in research with cancer cell lines inclusively (Marcinkiewicz 2013; Calderon et al. 2014; Simões-Silva et al. 2018). For example, crude venom samples from *Naja ashei* induced cell cycle arrest and promoted apoptosis with the activation of Caspase 3 and 9 in colorectal carcinoma (HCT116) cells (Antolikova et al. 2019). Additionally, venoms from spiders showed high cytotoxic activity towards K^+ channels expressing cancer cells (Okada et al. 2019).

From the five proteins identified from *Eulalia* mucus secretions, arylsulfatase B (ARSB) and serine proteases (SP) are commonly found in venoms (Matsui et al. 2000; Nok et al. 2003; Pesentseva et al. 2012). Sulfatases are known to play a key role in the control of physiological processes like cell signalling, hormone regulation and degradation of complex macromolecules such as glycosaminoglycans (Fricke and Hartmann 1974; Diez-Roux and Ballabio 2005; de Graaf et al. 2010). Arylsulfatase B, specifically, is known to regulate the kininogen–bradykinin axis trough the effect on chondroitin-4-sulfation and the interaction of C4S with kininogen in rat kidney epithelial cells, influencing blood pressure in mammals (Bhattacharyya et al. 2010). In lymphocytes, arylsulfatase may function in the degradation of cerebroside sulphate ester components of the target cell membrane to initiate the lytic event, having an important role in the immune response (Zucker-franklin et al. 1983). In molluscs, it can function as a sulphur-

scavenging enzyme, which may play a role in the breakdown of sulphated saponins (Pesentseva et al. 2012). The exact function of arylsulfatases in poisonous and venomous secretions are still unknown, even though this enzyme exists at high quantities in a broad range of organisms, from snakes to predatorial gastropods (Nok et al. 2003; Bose et al. 2017). In turn, serine proteases form one of the largest group of endopeptidases. They are involved in a wide range of physiological processes, from digestion to blood coagulation, fibrinolysis, development, fertilization, apoptosis and play an immunity role as well (Page and Di Cera 2008). In venoms, serine proteases are present in a wide span of animals as well, being well described in snakes and bees. These proteases have thrombin-like effects, promote fibrinogen degradation in prey and activate the phenoloxidase cascade that induces a lethal melanization response in insects attacking bees and their hives (Choo et al. 2010; Amorim et al. 2018; Carone et al. 2018).

Other proteins in the venomous cocktail of *E. viridis* are not commonly associated with venoms and poisons. However, they play paramount roles in cell homeostasis. It is the case of heat shock proteins (HSP). Heat shock protein 70 KDa (HSP70) is usually involved in stress response, helping protein folding (assembly and refolding) in the endoplasmic reticulum (Jee 2016). In turn, RAB3C is located in synaptic and secretory vesicles and regulates exocytosis Ca^{2+}-dependent secretion and vesicle docking, thus being essential in exocytic and endocytic membrane trafficking (Schlüter et al. 2002; Schwartz et al. 2008). The RAB3C protein can also be linked with ZW10-interacting protein 1 (Zwint-1), a protein implicated in the regulation of chromosomal segregation through binding to the kinetochore (van Vlijmen et al. 2008). The gene encoding this protein, *ZWINT*, was discovered to be overexpressed in ovarian cancer, among other genes (such as *CCNB1*, *CENPF*, *KIF11*) related with cell cycle, nuclear division and oocyte meiosis (Xu et al. 2016). The 14-3-3 protein plays a role in various cellular processes like signal transduction, cell cycle regulation, apoptosis, stress response, cytoskeleton organization and malignant transformation (Van Hemert et al. 2001). It must be noted, though, that other proteins may either have not been identified by cross-validation between MS/MS on mucus secretions and skin whole-transcriptome analyses. We must, at this stage, highlight difficulties in homology-based identification of peptides from organisms with reduced genomic annotation. Even though the purification process can, itself, be responsible for the loss of proteins, the toxicological appraisal of the purified extracts does indicate that the active components that yielded effects against A2780 cells were not lost.

The identified proteins, when integrated within protein interaction networks (Fig. 5.10), suggest a potential combined effect. Nonetheless, there are seemingly two pathways: one involving HSPs, 14-3-3 protein and RAB3C, and another where ARSB is pivotal. In the first pathway, 14-3-3 can be the key component, as these proteins are known to play a part in apoptosis and cell cycle arrest.

However, different 14-3-3 isoforms may produce distinct outcomes (Van Hemert et al. 2001). Still, the protein here identified holds higher similarities to 14-3-3 protein zeta (YWHAZ), which suggests promotion of programmed cell death. However, some isoforms may have indirect association with BAG1, which may in turn bind to BCL-2 and promote cell survival. It is also able to bind to the serine/threonine-specific protein kinase RAF1 and stimulate RAF1 kinase activity. Moreover, BAG1 also binds to the chaperone HSP70 (HSPA8), being, in fact, a possible chaperone inhibitor (Takayama et al. 1997). Speculations have been made to whether BCL-2 forms or not channels to transport proteins through membranes and if HSP70 unfolds proteins during such transport (Reed 1997; Takayama et al. 1997). In any case, the formation of these channels may contribute to explain BCL-2 overexpression (Fig. 5.7A). At the same time membrane integrity may be compromised by other proteins in the cocktail, explaining the changes in plasmatic membrane potential from 1 h of exposure (recall Fig. 5.8A, B), then likely triggering a series of events that may ultimately lead to cell cycle arrest and cell death.

The RAF proto-oncogene identified in the network is responsible for the regulation of many fundamental cell processes, including proliferation, differentiation, apoptosis, motility and metabolism (Maurer et al. 2011). The activation of this protein and the mitogen-activated protein kinase (MAPK) pathway is dependant of proteins present in the network such as HSPA8, HSP90 and CDC protein. These proteins were found to be crucial for the maturation and activation of RAF1 both in *Drosophila* and human cell lines (Grammatikakis et al. 1999; Huang et al. 2017). RAF1 activity is also regulated by 14-3-3 proteins, trough phosphorylation, to control its enzymatic activity (Jaumot and Hancock 2001). The CDC protein is directly connected with CWC15 and BCAS2, both involved in the regulation of gene expression, and is an essential regulator that activates cyclin-dependent kinases (CDKs) at critical stages of the cell cycle (Mils et al. 2000). The disruption of the regulatory function can lead to cell cycle arrest and autophagy, which are the mechanisms by which A2780 cells were affected. More specifically, the 14-3-3 protein present in the extract may well be, directly or indirectly, keeping the cyclin-dependent kinase CDC2 (cell division control protein 2) inactive, thereby preventing G2 to M transition (Ferguson et al. 2000; Van Hemert et al. 2001), as seen in A2780 cells between 2 h to 4 h of exposure (Fig. 5.8). The RAB3C protein is also responsible for cell cycle arrest through binding to ZW10-interacting protein 1 (van Vlijmen et al. 2008; Xu et al. 2016). Additionally, cell cycle arrest in the S phase is also connected with the depolarization of the resting membrane potential (MacFarlane and Sontheimer 2000), which is in accordance with our results after 3 and 4 h of exposure. After 9 h exposure to the cocktail, cell cycle progression was promoted when the assays were extended until 48 h. This effect was not, however, sufficient to avoid loss of cell viability.

Heat shock proteins 70 controls many housekeeping processes (Goloubinoff 2017), which is probably the reason why HSPA8 one of their forms, is linked directly with almost all proteins in the network (except RAB3C) described in Fig. 5.10. Two of these interplays occur with other chaperones, namely HSP90 alpha and beta. Besides interfering with cell cycle progression, HPSs can also be involved in the activation of matrix metalloproteinases. In fact, it has been seen that extracellular HSP90 alpha in fibrosarcoma and breast cancer cells promote MMP2 activation, which is critical for tumour invasiveness (Eustace et al. 2004). In turn, nuclear HSP70 promotes aurora kinase B (AURKB) activity, which is normally dysregulated in cancer cell lines (Cho et al. 2012).

Other homologs from known toxins (PRSS1 and ARSB) appeared, at first sight, to have a smaller contribution to the effects seen in A2780. However, ARSB is known to have its expression reduced in most carcinomas when compared with normal tissues (Kovacs et al. 2019). This reduction was found to be related to prostate cancer aggressiveness, due to the association between ARSB and the epidermal growth factor receptor (EGFR), commonly overexpressed when ARSB is reduced (Feferman et al. 2013; Bhattacharyya et al. 2018). The EGFR has a crucial role in cell differentiation and proliferation and cancer, being ARSB regulation an important target for cancer therapy (Bhattacharyya et al. 2018).

Finally, it must be noted that the co-occurrence of apoptosis and cell death by autophagy has already been reported in A2780 cells exposed, e.g. to metal-based compounds (Lenis-Rojas et al. 2016; Das et al. 2018; Choroba et al. 2019). Both pathways, occurring simultaneously, were also reported in a cervical cancer cell line, where the accumulation of autophagic vacuoles preceded the apoptotic cell death (González-Polo et al. 2005). However, all the studies have been performed with isolated compounds, while the present study dealt with a mixture of proteins, leading to complex interactions and likely intricate downstream effects on cell homeostasis. Even though the combined effect of venom components from marine invertebrates has received little attention, there are indications that venomous mixtures are more effective against human cancer cells than individualised components, such as in a recent study with glycoconjugates extracted from a starfish tested on colorectal carcinoma cells (Malyarenko et al. 2018).

5.6. Conclusion

In large part due to the number and variety of compounds therein, toxin-bearing secretions of marine invertebrates may have cytotoxic properties beyond the functions acquired by means of natural selection. The present study showed that proteins present in the mucus secreted by a

venomous marine worm, five of which were positively identified, can specifically cause cell cycle arrest by a mechanism that is specific to A2780 ovarian adenocarcinoma cells among a panel of six human cell lines, normal included. Even though the precise mechanisms still need further exploration, the effects on A2780 cells seem to derive from the collision of two interaction networks: one centred in cell cycle arrest as a consequence of dysregulation of protein expression and folding, with a link to RAB3C oncoprotein function; another centred in the lytic activity of arylsulfatases, common enzymes in animal venoms and poisons. Even though the mechanisms may seem complex, the findings support the potential of mixed proteins in venoms against aggressively-proliferating cells, as seen, e.g., for been venoms, steering toward a novel approach to anti-cancer treatments while highlighting the immense potential of the oceans as a source of novel therapeutics.

5.7. References

Almeida J, Roma-Rodrigues C, Mahmoud AG, Guedes da Silva MFC, Pombeiro AJL, Martins LMDRS, Baptista P V., Fernandes AR (2019) Structural characterization and biological properties of silver(I) tris(pyrazolyl)methane sulfonate. J Inorg Biochem 199:110789.

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

Amorim FG, Menaldo DL, Carone SEI, Silva TA, Sartim MA, De Pauw E, Quinton L, Sampaio S V. (2018) New insights on moojase, a thrombin-like serine protease from *Bothrops moojeni* snake venom. Toxins (Basel) 10:1–19.

Anglesio MS, Wiegand KC, Melnyk N, Chow C, Salamanca C, Prentice LM, Senz J, Yang W, Spillman MA, Cochrane DR, Shumansky K, Shah SP, Kalloger SE, Huntsman DG (2013) Type-specific cell line models for type-specific ovarian cancer research. PLoS One.

Antolikova NR, Kello M, Zigova M, Tischlerova V, Petrilla V, Pirnik Z, Mojzisova G, Mojzis J (2019) *Naja asheiz* venom induces mitochondria-mediated apoptosis in human colorectal cancer cells. Acta Biochim Pol 66:207–213.

Beaufort CM, Helmijr JCA, Piskorz AM, Hoogstraat M, Ruigrok-Ritstier K, Besselink N, Murtaza M, Van IJcken WFJ, Heine AAJ, Smid M, Koudijs MJ, Brenton JD, Berns EMJJ, Helleman J (2014) Ovarian cancer cell line panel (OCCP): Clinical importance of *in vitro* morphological subtypes. PLoS One.

Bhattacharyya S, Kotlo K, Danziger R, Tobacman JK (2010) Arylsulfatase B regulates interaction of chondroitin-4-sulfate and kininogen in renal epithelial cells. Biochim Biophys Acta - Mol Basis

Dis 1802:472–477.

Bhattacharyya S, Feferman L, Han X, Ouyang Y, Zhang F, Linhardt RJ, Tobacman JK (2018) Decline in arylsulfatase B expression increases EGFR expression by inhibiting the protein-tyrosine phosphatase SHP2 and activating JNK in prostate cells. J Biol Chem 293:11076–11087.

Bose U, Wang T, Zhao M, Motti CA, Hall MR, Cummins SF (2017) Multiomics analysis of the giant triton snail salivary gland, a crown-of-thorns starfish predator. Sci Rep 7:1–14.

Calderon LA, Sobrinho JC, Zaqueo KD, Moura AA De, Grabner AN, Mazzi M V, Marcussi S, Nomizo A, Fernandes CFC, Zuliani JP, Carvalho BMA, Silva SL, Stábeli RG, Soares AM (2014) Antitumaral activities of snake venom proteins: new trends in cancer therapy.

Carone SEI, Menaldo DL, Sartim MA, Bernardes CP, Caetano RC, da Silva RR, Cabral H, Barraviera B, Ferreira Junior RS, Sampaio S V. (2018) BjSP, a novel serine protease from *Bothrops jararaca* snake venom that degrades fibrinogen without forming fibrin clots. Toxicol Appl Pharmacol 357:50–61.

Chernikov O, Kuzmich A, Chikalovets I, Molchanova V, Hua KF (2017) Lectin CGL from the sea mussel *Crenomytilus grayanus* induces Burkitt's lymphoma cells death via interaction with surface glycan. Int J Biol Macromol 104:508–514.

Cho HS, Shimazu T, Toyokawa G, Daigo Y, Maehara Y, Hayami S, Ito A, Masuda K, Ikawa N, Field HI, Tsuchiya E, Ohnuma SI, Ponder BAJ, Yoshida M, Nakamura Y, Hamamoto R (2012) Enhanced HSP70 lysine methylation promotes proliferation of cancer cells through activation of Aurora kinase B. Nat Commun.

Choo YM, Lee KS, Yoon HJ, Kim BY, Sohn MR, Roh JY, Je YH, Kim NJ, Kim I, Woo SD, Sohn HD, Jin BR (2010) Dual function of a bee venom serine protease: Prophenoloxidase-activating factor in arthropods and fibrin(ogen)olytic enzyme in mammals. PLoS One.

Choroba K, MacHura B, Raposo LR, Małecki JG, Kula S, Pajak M, Erfurt K, Maroń AM, Fernandes AR (2019) Platinum(II) complexes showing high cytotoxicity toward A2780 ovarian carcinoma cells. Dalt Trans 48:13081–13093.

Cuevas C, Francesch A (2009) Development of Yondelis® (trabectedin, ET-743). A semisynthetic process solves the supply problem. Nat Prod Rep 26:322–337.

Czerwińska K, Machura B, Kula S, Krompiec S, Erfurt K, Roma-Rodrigues C, Fernandes AR, Shul'Pina LS, Ikonnikov NS, Shul'Pin GB (2017) Copper(II) complexes of functionalized 2,2′:6′,2″-terpyridines and 2,6-di(thiazol-2-yl)pyridine: Structure, spectroscopy, cytotoxicity and catalytic activity. Dalt Trans 46:9591–9604.

Das K, Beyene BB, Datta A, Garribba E, Roma-Rodrigues C, Silva A, Fernandes AR, Hung CH

(2018) EPR and electrochemical interpretation of bispyrazolylacetate anchored Ni(ii) and Mn(ii) complexes: Cytotoxicity and anti-proliferative activity towards human cancer cell lines. New J Chem 42:9126–9139. doi: 10.1039/c8nj01033a

de Graaf DC, Aerts M, Brunain M, Desjardins CA, Jacobs FJ, Werren JH, Devreese B (2010) Insights into the venom composition of the ectoparasitoid wasp *Nasonia vitripennis* from bioinformatic and proteomic studies. Insect Mol Biol 19:11–26. doi: 10.1111/j.1365-2583.2009.00914.x

Diez-Roux G, Ballabio A (2005) Sulfatases and human disease. Annu Rev Genomics Hum Genet 6:355–379. doi: 10.1146/annurev.genom.6.080604.162334

do Nascimento-Neto LG, Cabral MG, Carneiro RF, Silva Z, Arruda FVS, Nagano CS, Fernandes AR, Sampaio AH, Teixeira EH, Videira PA (2018) Halilectin-3, a lectin from the marine sponge *Haliclona caerulea*, induces apoptosis and autophagy in human breast cancer MCF7 cells through caspase-9 pathway and LC3-II protein expression. Anticancer Agents Med Chem 18:521–528. doi: 10.2174/1871520617666171114094847

Eustace BK, Sakurai T, Stewart JK, Yimlamai D, Unger C, Zehetmeier C, Lain B, Torella C, Henning SW, Beste G, Scroggins BT, Neckers L, Ilag LL, Jay DG (2004) Functional proteomic screens reveal an essential extracellular role for hsp90α in cancer cell invasiveness. Nat Cell Biol 6:507–514. doi: 10.1038/ncb1131

Feferman L, Bhattacharyya S, Deaton R, Gann P, Guzman G, Kajdacsy-Balla A, Tobacman JK (2013) Arylsulfatase B (N-acetylgalactosamine-4-sulfatase): Potential role as a biomarker in prostate cancer. Prostate Cancer Prostatic Dis 16:277–284. doi: 10.1038/pcan.2013.18

Ferguson AT, Evron E, Umbricht CB, Pandita TK, Chan TA, Hermeking H, Marks JR, Lambers AR, Futreal PA, Stampfer MR, Sukumar S (2000) High frequency of hypermethylation at the 14-3-3 σ locus leads to gene silencing in breast cancer. Proc Natl Acad Sci U S A 97:6049–6054. doi: 10.1073/pnas.100566997

Ferlay J, Colombet M, Soerjomataram I, Dyba T, Randi G, Bettio M, Gavin A, Visser O, Bray F (2018) Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. Eur J Cancer 103:356–387. doi: 10.1016/j.ejca.2018.07.005

Figuerola B, Avila C (2019) The phylum bryozoa as a promising source of anticancer drugs. Mar Drugs 17:1–23. doi: 10.3390/md17080477

Fricke ER, Hartmann F (1974) Connective Tissues Biochemistry and Pathophysiology.

Goloubinoff P (2017) Editorial: The HSP70 molecular chaperone machines. Front Mol Biosci 4:1–4. doi: 10.3389/fmolb.2017.00001

Gomes SE, Pereira DM, Roma-rodrigues C, Rodrigues MP, Fernandes R, Borralho PM (2018) Convergence of miR-143 overexpression , oxidative stress and cell death in HCT116 human colon cancer cells. PLoS One 13:1–19.

González-Polo RA, Boya P, Pauleau AL, Jalil A, Larochette N, Souquère S, Eskelinen EL, Pierron G, Saftig P, Kroemer G (2005) The apoptosis/autophagy paradox: Autophagic vacuolization before apoptotic death. J Cell Sci 118:3091–3102.

Grammatikakis N, Lin J-H, Grammatikakis A, Tsichlis PN, Cochran BH (1999) $p50^{cdc37}$ acting in concert with Hsp90 is Required for Raf-1 function. Mol Cell Biol 19:1661–1672.

Hallas-Potts A, Dawson JC, Herrington CS (2019) Ovarian cancer cell lines derived from non-serous carcinomas migrate and invade more aggressively than those derived from high-grade serous carcinomas. Sci Rep 9:1–10.

Hames BD (1998) Gel electrophoresis of proteins: A practical approach, Third Edit. Oxford University Press, New York

Hu Y, Chen J, Hu G, Yu J, Zhu X, Lin Y, Chen S, Yuan J (2015) Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. Mar Drugs 13:202–221.

Huang Y, Guo X-X, Han B, Zhang X-M, An S, Zhang X-Y, Yang Y, Liu Y, Hao Q, Xu T-R (2017) Decoding the full picture of Raf1 function based on its interacting proteins. Oncotarget 8:68329–68337.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat 5:299–314.

Jaumot M, Hancock JF (2001) Protein phosphatases 1 and 2A promote Raf-1 activation by regulating 14-3-3 interactions. Oncogene 20:3949–3958.

Jee H (2016) Size dependent classification of heat shock proteins: a mini-review. J Exerc Rehabil 12:255–259.

Khalifa SAM, Elias N, Farag MA, Chen L, Saeed A, Hegazy MEF, Moustafa MS, El-Wahed AA, Al-Mousawi SM, Musharraf SG, Chang FR, Iwasaki A, Suenaga K, Alajlani M, Göransson U, El-Seedi HR (2019) Marine natural products: A source of novel anticancer drugs. Mar Drugs.

Kovacs Z, Jung I, Gurzu S (2019) Arylsulfatases A and B: From normal tissues to malignant tumors. Pathol Res Pract 215:152516.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

Lalwani N, Prasad SR, Vikram R, Shanbhogue AK, Huettner PC, Fasih N (2011) Histologic,

molecular, and cytogenetic features of ovarian cancers: Implications for diagnosis and treatment. Radiographics 31:625–646.

Leal MC, Anaya-Rojas JM, Munro MHG, Blunt JW, Melian CJ, Calado R, Lürig MD (2020) Fifty years of capacity building in the search for new marine natural products. Proc Natl Acad Sci 117:202007610.

Lenis-Rojas OA, Fernandes AR, Roma-Rodrigues C, Baptista P V., Marques F, Pérez-Fernández D, Guerra-Varela J, Sánchez L, Vázquez-García D, Torres ML, Fernández A, Fernández JJ (2016) Heteroleptic mononuclear compounds of ruthenium(II): synthesis, structural analyses, *in vitro* antitumor activity and *in vivo* toxicity on zebrafish embryos. Dalt Trans 45:19127–19140.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402–408.

Ma Z, Zhang B, Guedes Da Silva MFC, Silva J, Mendo AS, Baptista PV, Fernandes AR, Pombeiro AJL (2016) Synthesis, characterization, thermal properties and antiproliferative potential of copper(II) 4′-phenyl-terpyridine compounds. Dalt Trans 45:5339–5355.

MacFarlane SN, Sontheimer H (2000) Changes in ion channel expression accompany cell cycle progression of spinal cord astrocytes. Glia 30:39–48.

Malyarenko T V., Malyarenko OS, Kicha AA, Ivanchina N V., Kalinovsky AI, Dmitrenok PS, Ermakova SP, Stonik VA (2018) In vitro anticancer and proapoptotic activities of steroidal glycosides from the starfish *Anthenea aspera*. Mar Drugs.

Mann J (2002) Natural products in cancer chemotherapy: Past, present and future. Nat Rev Cancer 2:143–148.

Marcinkiewicz C (2013) Applications of snake venom components to modulate integrin activities in cell-matrix interactions. Int J Biochem Cell Biol 45:1974–1986.

Matsui T, Fujimura Y, Titani K (2000) Snake venom proteases affecting hemostasis and thrombosis. Biochim Biophys Acta - Protein Struct Mol Enzymol 1477:146–156.

Maurer G, Tarkowski B, Baccarini M (2011) Raf kinases in cancer-roles and therapeutic opportunities. Oncogene 30:3477–3488.

Mils V, Baldin V, Goubin F, Pinta I, Papin C, Waye M, Eychene A, Ducommun B (2000) Specific interaction between 14-3-3 isoforms and the human CDC25B phosphatase. Oncogene 19:1257–1265.

Nok AJ, Abubakar MS, Adaudi A, Balogun E (2003) Aryl sulfatase from *Naja nigricolis* venom: characterization and possible contribution in the pathology of snake poisoning. J Biochem Mol Toxicol 17:59–66.

Okada M, Ortiz E, Corzo G, Possani LD (2019) Pore-forming spider venom peptides show cytotoxicity to hyperpolarized cancer cells expressing K^+ channels: A lentiviral vector approach. PLoS One 14:1–16.

Oršolić N (2012) Bee venom in cancer therapy. Cancer Metastasis Rev 31:173–194.

Page MJ, Di Cera E (2008) Evolution of peptidase diversity. J Biol Chem 283:30010–30014.

Pesentseva MS, Sova V V., Silchenko AS, Kicha AA, Silchenko AS, Haertle T, Zvyagintseva TN (2012) A new arylsulfatase from the marine mollusk *Turbo chrysostomus*. Chem Nat Compd 48:853–859.

Reed JC (1997) Double identity for proteins of the Bcl-2 family. Nature 387:773–776.

Roberts PJ, Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 26:3291–3310.

Schindelin J, Arganda-Carrera I, Frise E, Verena K, Mark L, Tobias P, Stephan P, Curtis R, Stephan S, Benjamin S, Jean-Yves T, Daniel JW, Volker H, Kevin E, Pavel T, Albert C (2009) Fiji - an open platform for biological image analysis. Nat Methods 9:241.

Schlüter OM, Khvotchev M, Jahn R, Südhof TC (2002) Localization *versus* function of Rab3 proteins: Evidence for a common regulatory role in controlling fusion. J Biol Chem 277:40919–40929.

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675.

Schwartz SL, Cao C, Pylypenko O, Rak A, Wandinger-Ness A (2008) Rab GTPases at a glance. J Cell Sci 120:3905–3910.

Simões-Silva R, Alfonso J, Gomez A, Holanda RJ, Sobrinho JC, Zaqueo KD, Moreira-Dill LS, Kayano AM, Grabner FP, da Silva SL, Almeida JR, Stabeli RG, Zuliani JP, Soares AM (2018) Snake venom, a natural library of new potential therapeutic molecules: challenges and current perspectives. Curr Pharm Biotechnol 19:308–335.

States DJ, Gish W (1994) Combined use of sequence similarity and codon bias for coding region identification. J Comput Biol 1:39–50.

Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH (2015) UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics 31:926–932.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Von Mering C (2019) STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613.

Takayama S, Bimston DN, Matsuzawa SI, Freeman BC, Aime-Sempe C, Xie Z, Morimoto RI, Reed JC (1997) BAG-1 modulates the chaperone activity of Hsp70/Hsc70. EMBO J 16:4887–4896.

Thiel D, Hugenschutt M, Meyer H, Paululat A, Quijada-Rodriguez AR, Purschke G, Weihrauch D (2017) Ammonia excretion in the marine polychaete *Eurythoe complanata* (Annelida). J Exp Biol 220:425–436.

Van Hemert MJ, Yde Steensma H, Van Paul G (2001) 14-3-3 Proteins: Key regulators of cell division, signalling and apoptosis. BioEssays 23:936–946.

van Vlijmen T, Vleugel M, Evers M, Mohammed S, Wulf PS, Heck AJR, Hoogenraad CC, van der Sluijs P (2008) A unique residue in rab3c determines the interaction with novel binding protein Zwint-1. FEBS Lett 582:2838–2842.

Venable JH, Coggeshall R (1965) A simplified lead citrate stain for use in electron microscopy. J Cell Biol 25:407–408.

Vinhas R, Fernandes AR, Baptista P V. (2017) Gold nanoparticles for BCR-ABL1 gene silencing: improving tyrosine kinase inhibitor efficacy in chronic myeloid leukemia. Mol Ther - Nucleic Acids 7:408–416.

Vogt C, Pernemalm M, Kohonen P, Laurent S, Hultenby K, Vahter M, Lehtiö J, Toprak MS, Fadeel B (2015) Proteomics analysis reveals distinct corona composition on magnetic nanoparticles with different surface coatings: Implications for interactions with primary human macrophages. PLoS One 10:1–20.

Xu Z, Zhou Y, Cao Y, Dinh TLA, Wan J, Zhao M (2016) Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis. Med Oncol.

Zucker-franklin D, Grusky G, Yang J (1983) Arylsulfatase in natural killer cells: Its possible role in cytotoxicity. Proc Natl Acad Sci USA 80:6977–6981.

Appendix Chapter 5

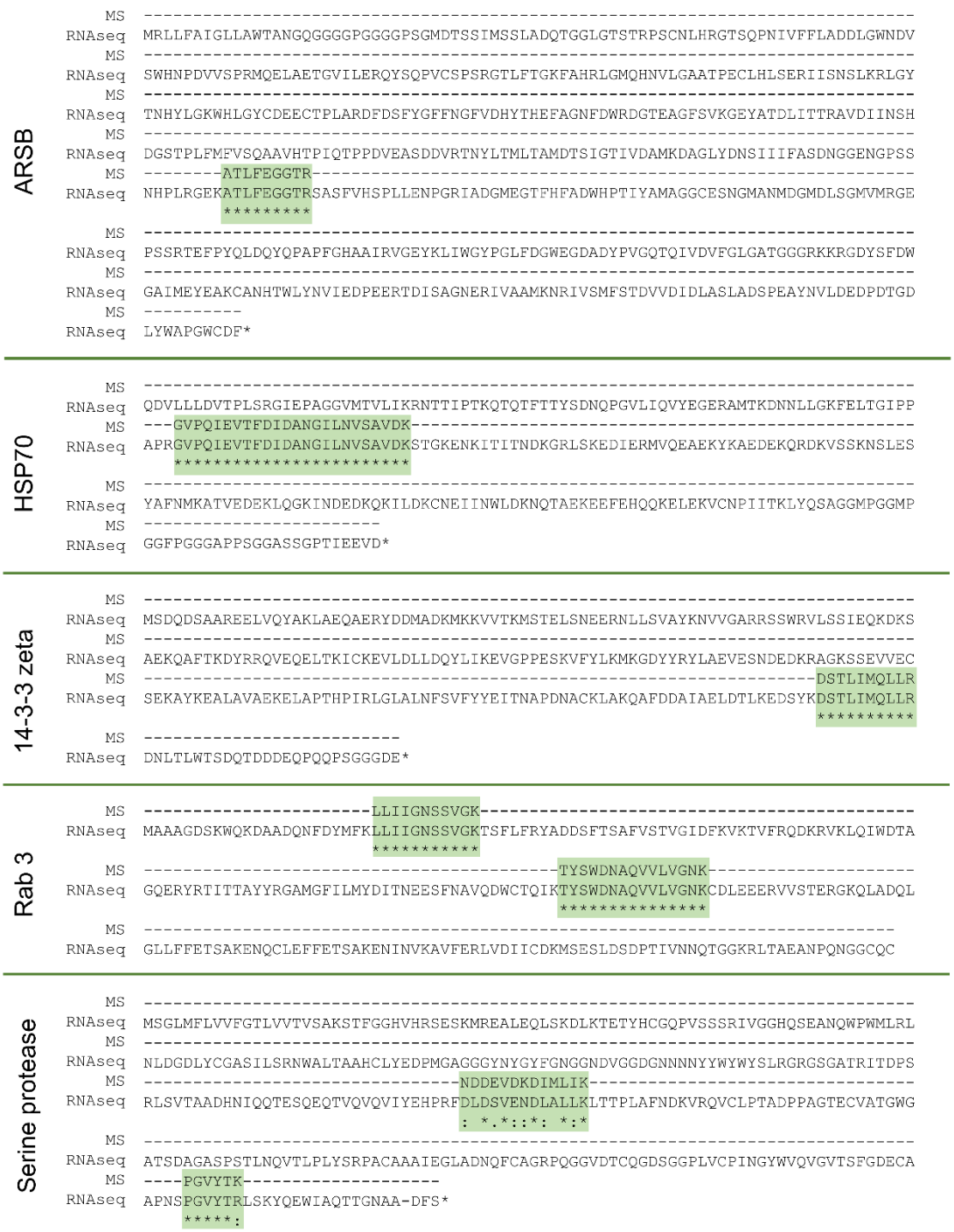


Figure S5.1. Alignment of the peptides obtained from MS/MS and RNAseq. Similarities are indicated as follow: "*" - Indicative of fully conserved residue, ":" - Indicative of strongly similar properties - scoring > 0.5 in the Gonnet PAM 250 matrix, "." - Indicative of weakly similar properties - scoring ≤ 0.5 in the Gonnet PAM 250 matrix.

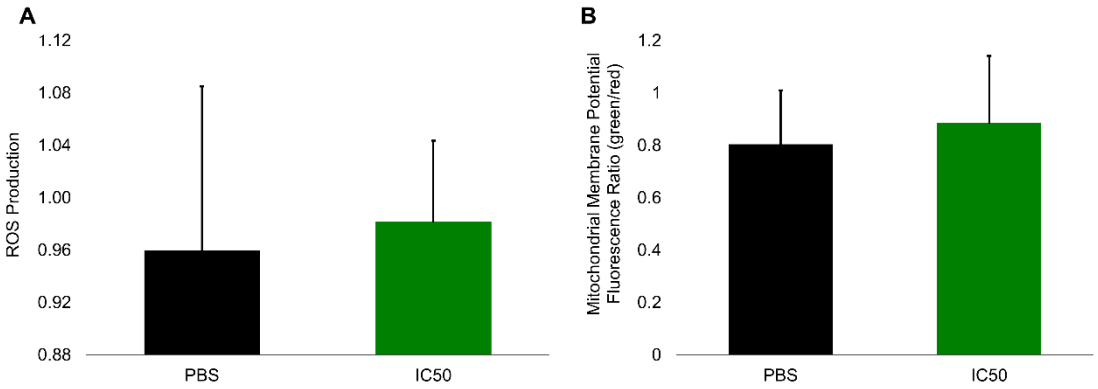


Figure S5.2. Production of Reactive Oxygen Species (A) and Mitochondrial Membrane Potential (B) in A2780 cells exposed during 48 hours to the extract (IC_{50}, 0.08 µg. µL^{-1}) and the control (PBS). No significant statistical differences were found between treatments.

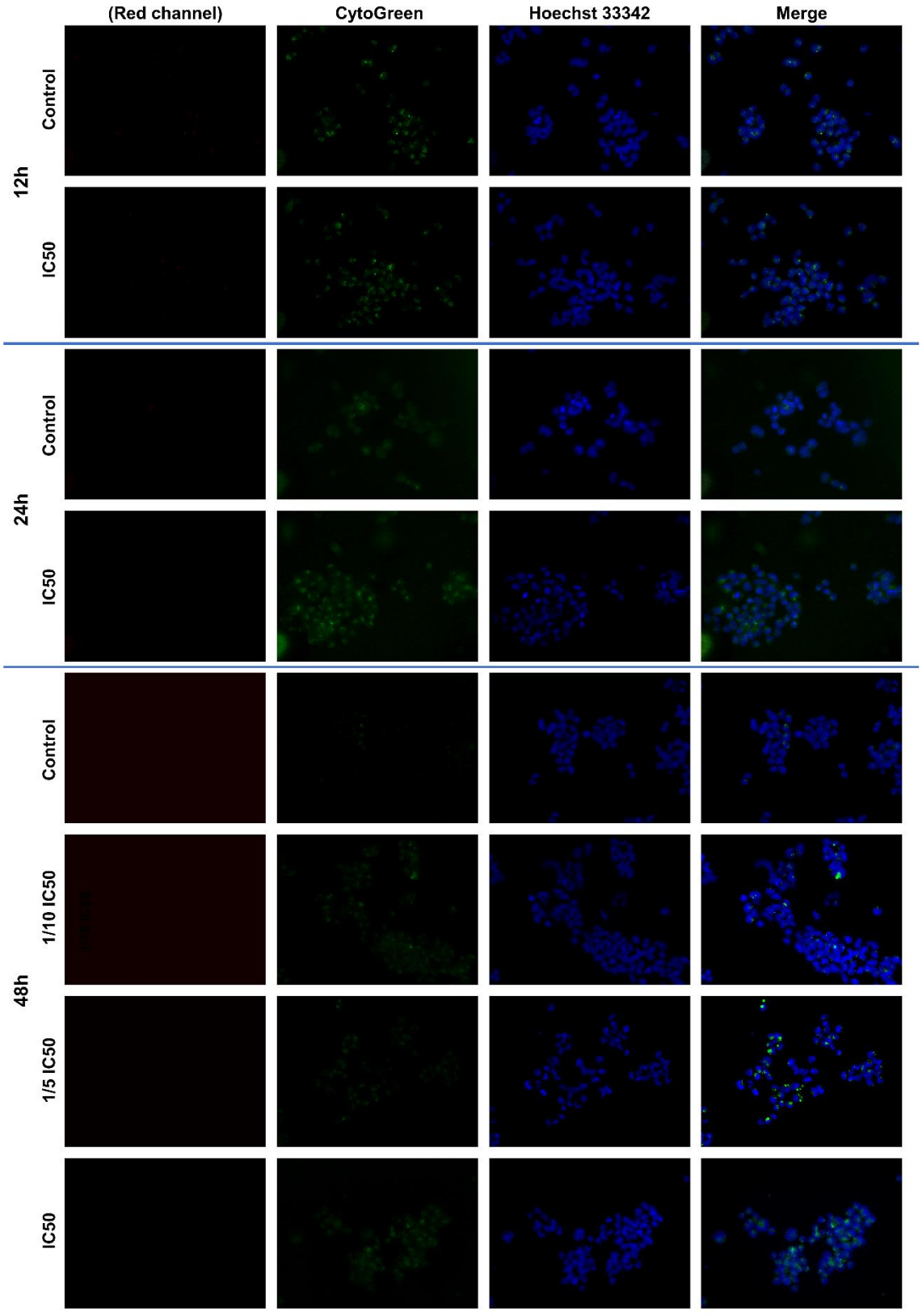


Figure S5.3. Autophagy assay of cells exposed to the extract and the control (PBS) during 12 h, 24 h and 48 h to an extract concentration of IC_{50} (0.08 µg. $µL^{-1}$) and during 48 h to a concentration of 1/10 IC_{50} and 1/5 IC_{50}. A2780 cell lines stained with CytoGreen (green) for autophagy and counterstain with hoechst 33342 (blue).

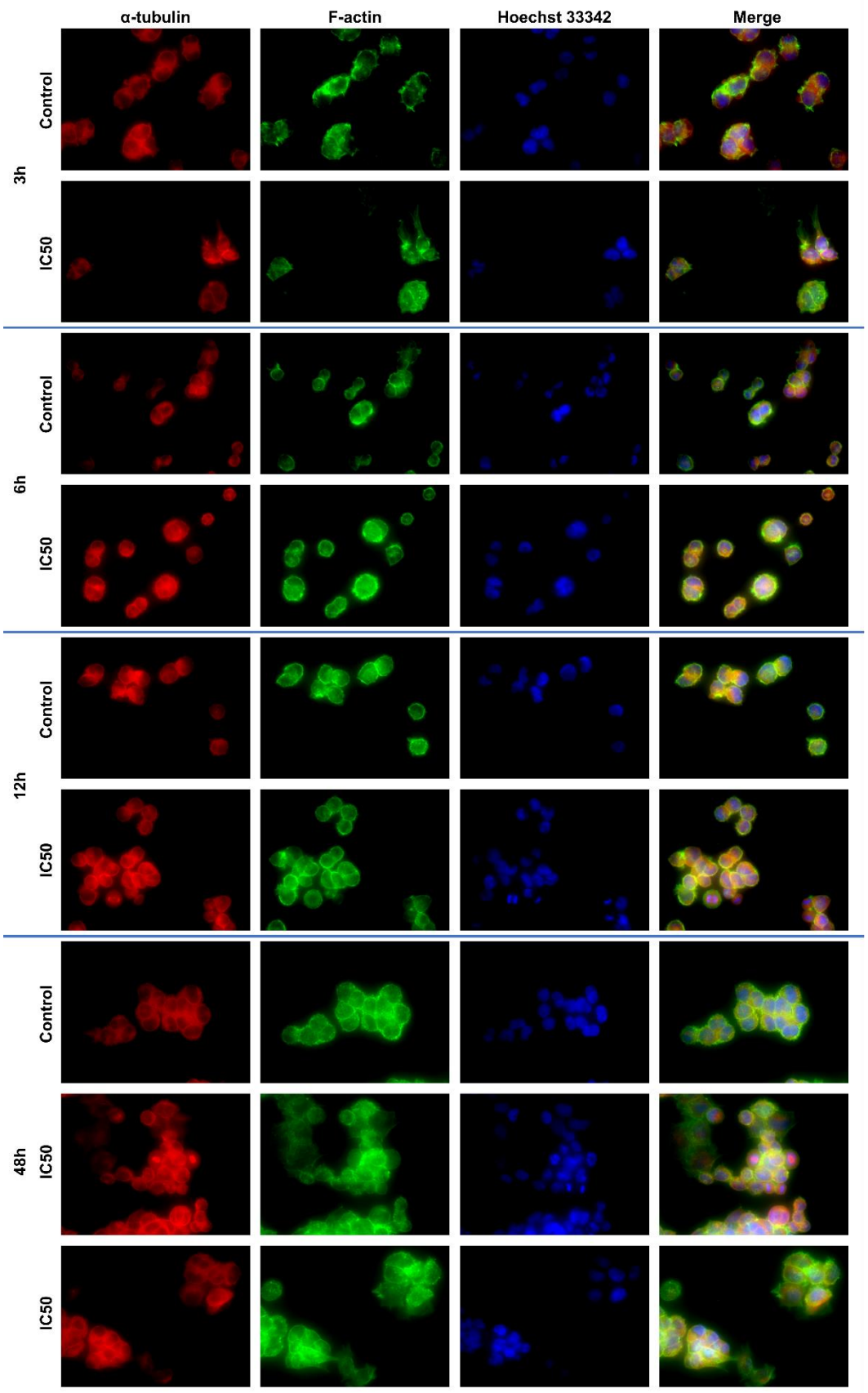


Figure S5.4. Immunofluorescent assay used in A2780 cells exposed to the extract (IC_{50}, 0.08 µg. $µL^{-1}$) and the control (PBS) during 3 h, 6 h, 12 h and 48 h. The cells were probed with F-actin (green) and α-tubulin (red) and counterstained with Hoechst dye 33258 (blue).

Table S5.1. Individual results from STRING for each of the homologous in *Homo sapiens* (ARSB - arylsulfatase B, PRSS1 - trypsin-1 serine protease, YWHAZ - 14-3-3 protein zeta, RAB3C - Ras-related protein Rab-3C, HSPA8 - Heat shock protein 70 kDa), for the proteins identified in *Eulalia viridis*. The network for each protein has been expanded to additional ten proteins and the confidence cut-off for showing interaction links has been set to medium (0.400).

Protein	Biological process				Molecular function				Cellular component				KEGG pathway			
	GO-term	description	count in gene set	false discovery rate	GO-term	description	count in gene set	false discovery rate	GO-term	description	count in gene set	false discovery rate	GO-term	description	count in gene set	false discovery rate
ARSB	GO:0006027	Glycosaminoglycan catabolic process	3 of 62	1.1E-4	GO:0004065	Arylsulfatase activity	2 of 12	0.00018	GO:0005788	Endoplasmic reticulum lumen	4 of 299	3.87E-05	hsa04142	Lysosome	4 of 123	1.48E-07
	GO:1903510	Mucopolysaccharide metabolic process	3 of 109	1.9E-4	GO:0004553	Hydrolase activity, hydrolysing O-glycosyl compounds	2 of 99	0.0033	GO:0043202	Lysosomal lumen	3 of 94	5.49E-05	hsa00531	Glycosamino-glycan degradation	3 of 19	1.48E-07
	GO:0030207	Chondroitin sulphate catabolic process	2 of 14	3.8E-4												
PRSS1	GO:0022617	Extracellular matrix disassembly	6 of 58	1.84E-10	GO:0004252	Serine-type endopeptidase activity	7 of 180	9.11E-11	GO:0031012	Extracellular matrix	5 of 283	1.52E-05	hsa04657	IL-17 signalling pathway	3 of 92	3.90E-04
	GO:0030574	Collagen catabolic process	5 of 37	3.42E-09	GO:0004222	Metalloendo-peptidase activity	5 of 110	2.53E-08	GO:0005576	Extracellular region	9 of 2505	1.52E-05	hsa04926	Relaxin signalling pathway	3 of 130	5.30E-04
YWHAZ	GO:1900740	Positive regulation of protein insertion into mitochondrial membrane	5 of 25	2.45E-09	GO:0019904	Protein domain specific binding	6 of 706	1.20E-04	GO:0005739	Mitochondrion	7 of 1531	4.80E-04	hsa04151	PI3K-Akt signalling pathway	8 of 348	1.57E-10
	GO:0043065	Positive regulation of apoptotic process	7 of 604	6.17E-07	GO:0019899	Enzyme binding	8 of 2197	2.00E-04	GO:0042470	Melanosome	3 of 105	1.40E-03	hsa04114	Oocyte meiosis	6 of 116	1.05E-09
													hsa04110	Cell cycle	6 of 123	1.05E-09
RAB3C	GO:0015031	protein transport	9 of 1391	6.81E-07	GO:0003924	GTPase activity	9 of 283	1.56E-13	GO:0030659	Cytoplasmic vesicle membrane	7 of 724	1.83E-06	HSA-8873719	Rab geranylgeranylation	9 of 63	7.61E-20
	GO:0016192	vesicle-mediated transport	8 of 1699	1.40E-05	GO:0005525	GTP binding	9 of 366	7.55E-13	GO:0008021	Synaptic vesicle	5 of 155	1.83E-06	HSA-9007101	Rab regulation of trafficking	5 of 118	2.56E-08
HSPA8	GO:1900034	regulation of cellular response to heat	3 of 45	1.40E-04	GO:0051082	Unfolded protein binding	3 of 106	2.80E-04	GO:0000974	Prp19 complex	2 of 12	9.40E-04	hsa04141	Protein processing in endoplasmic reticulum	3 of 161	1.30E-04
	GO:0061684	Chaperone-mediated autophagy	2 of 4	3.00E-04	GO:0051087	Chaperone binding	2 of 100	1.65E-02	GO:0071007	U2-type catalytic step 2 spliceosome	3 of 187	1.20E-03	hsa03040	Spliceosome	3 of 130	1.30E-04

CHAPTER 6 - Fluorescence in *Eulalia viridis* extract

This chapter has been submitted for publication:

Rodrigo AP, Lopes A, Pereira R, Mendes V, Manadas B, Grosso AR, Baptista PV, Fernandes AR, Costa PM (2020). Endogenous fluorescent proteins in the mucus of an intertidal Polychaeta. *Submitted*.

6.1. Abstract

The vast ocean holds many unexplored organisms with unique adaptive features that enable them to thrive in their environment. The secretion of fluorescent proteins, such as green fluorescent protein (GFP) first isolated from several Cnidaria is one of them. Reports on the presence of such compounds in Polychaeta are scarce but highlight marine invertebrates as source of novel bioproducts. The intertidal *Eulalia viridis* is a recent example. The worm secretes copious amounts of mucus whose purified and concentrated extracts yielded strong fluorescence under UV light. Emission has two main maxima, one at 400 nm and other at 500 nm, the latter responsible for the blue-greenish fluorescence. Combining proteomics and transcriptomics techniques, we identified ubiquitin, peroxiredoxin and 14-3-3 protein as key elements in the mucus. Fluorescence was found to be modulated by redox status and pH, more significantly than buffer, salinity or the presence of calcium, being consistently upheld in extracts prepared in Tris-HCl buffer complemented with DTT (reducing agent) at pH 7 and excited at 330 nm. The proteins associated with the fluorescent signal were localized in secretory cells in the pharynx. The fluorescent complexes were internalised by A2780 ovarian cancer cells, albeit emission shifting towards reddish. The results indicate that the secretion of fluorescent proteinaceous compounds or complexes can be an important defence against UV for this dweller of the rocky intertidal. Additionally, internalisation and modulation of fluorescence by simple modulation of redox status bears important considerations for biotechnological application of mucus components as markers and reports.

6.2. Introduction

The first report of fluorescence in organisms dates back from the 16th century when Spanish explorers of the West Indies noticed that water would turn blue when infused with the wood from *Lignum nephriticum* (see Lagorio et al. 2015 for further details). Nonetheless, only when chlorophyll and quinone fluorescence were reported from plants in the beginning of the 1800's from *Prunus laurocerasus* and *Cinchona officinalis*, respectively, was fluorescence perceived as it is in our days. Fluorescence was first reported in animals in the 19th century, more specifically in the feathers of some species of as a mean to lure females during mating season. From this point onward, reports of luminescence sprang up from the whole animal kingdom, a concept that now precludes fluorescence, bioluminescence and chemiluminescence (see Lagorio et al. 2015, for further details on historical aspects). Bioluminescence has arguably received most of the attention (recall the luciferin-luciferase system), since the discovery of the green fluorescent protein (GFP), a photoprotein that yields both bioluminescence and fluorescence was first isolated from the hydrozoan medusa *Aequorea victoria* (Shimomura 2006). Since then, the exploitation of

luminescent substances of marine origins gained momentum, in large part pushed by the biotechnological applications of GFP in its natural and recombinant forms (see Tsien 1998 for a review).

Animal bioluminescence, chemiluminescence or fluorescence, invariably relates with feeding, mating or defence mechanisms (Widder 2010). Due to its diversity of habitats and concomitant biodiversity, the marine environment has been revealing an extraordinary span of luminescent organisms. Deep-sea fish, in particular, are well-known to explore endogenous or symbiotic-driven bioluminescence, whereas those in shallower waters exploit the restricted spectra of sunlight penetrating the depths by producing fluorescent patterns to reduce or enhance visual contrast (see Sparks et al. 2014). Although less explored, marine invertebrates may hold even more diverse luminescent adaptations to their environment. Indeed, among the Cnidaria alone, several groups besides hydrozoans were found to have GFP analogues, such as the Anthozoa (Lagorio et al. 2015). Organisms from this class possess fluorescent proteins that regulate the availability of sunlight to polyps (Salih et al. 2000). Fluorescence was also detected in siphonophores, which has been attributed to the presence of porphyrins in their tentacles that are bioluminescent and red fluorescent, a feature that is believed to lure fish (Bonnett et al. 1979; Haddock et al. 2005). Among marine invertebrates Cnidaria is not only the phylum with the most identified fluorescent substances, but it is also the best-studied invertebrate group for biotechnological applications, from food to insecticides (Santhanam 2016). Still, in Arthropoda, copepods from the family Pontellidae are known to possess GFP-like proteins for recognition (Shagin et al. 2004). In turn, cephalochordate GFP-like proteins are thought to play a role in photoreception and photoprotection against ultra violet (UV) and blue light (Deheyn et al. 2007). Fluorescence has also been attributed to microorganisms, such as bacteria and algae, living in symbiosis with marine organisms. For example, algae living in symbiosis with corals facilitate light availability and may have a photoprotective role (see Lagorio et al. 2015). Fluorescence resulting from symbiosis in marine invertebrates has been used to understand their interactions in novel ways and use their fluorescence in techniques such as live cell analysis (Ainsworth et al. 2008).

Information on fluorescent substances in Polychaeta are scarce (see Rodrigo and Costa 2019) despite being one of the most abundant and diversified taxa, that occupy every habitat between deep-sea hydrothermal vents to the intertidal (Bartolomaeus et al. 2005). The first reports began with *Chaetopterus*, whose chaetopterin pigments was found to be fluorescent (see, for instance, Nicol, 1952; Kennedy and Nicol, 1959). Fluorescent photoproteins have also been found in the scales of scale worms, Polynoidae (Fresneau et al. 1984). *Eulalia viridis*, a green-coloured predator polychaete of the rocky intertidal, has recently been investigated for its porphyrinoid pigments, seemingly associated to UV absorption as protective measure while foraging (Martins et al. 2019).

Even though these pigments are not fluorescent, the toxin-bearing mucus secreted by the worm (recall chapter 3) possesses unknown fluorescent proteinaceous substances that strongly emit blue-green light when excited with UV light (Rodrigo et al. 2019). We hypothesize that the fluorescent compounds in the mucus is an adaptation to provide the worm with an additional defence against exposure to daylight. The aim of this work is thus a contribution to the identification of the proteinaceous compounds responsible for the fluorescence of the mucus secreted by *E. viridis* and the main exogenous variables that modulate emission.

6.3. Material and methods

6.3.1. Protein extraction

Eulalia (≈120 mm total length and weighting ≈250 mg each) were hand-collected collected at rocky intertidal beaches from the west coast of Portugal. Animals were reared in laboratory in a microcosm environment fitted with natural pebbles and clumps of mussels to provide shelter and feed. Crude mucus samples were harvested non-invasively through mechanic perturbation in sterile and filtered seawater. Each crude sample consisted of a pool of mucus harvested from 50-80 worms. Buffer was immediately added (1:1) to each sample: 0.05M Tris-HCl pH 7, containing 10 % m/v L-dithiothreitol (DTT) and 1 % v/v protease inhibitor cocktail (Sigma-Aldrich). Samples were then filtered through a cellulose acetate filter (0.22 µm) to remove solid materials and polymerised mucins. Extracts were then subjected to dialysis using 3 kDa Amicon ultrafiltraton tubes (Merck) to replace the buffer with experimental media: sterilised seawater, phosphate-buffered saline or 0.05M Tris-HCl. Total protein in purified extracts was quantified spectrophotometrically using a Nanodrop 1000 (Thermo Scientific). Samples were stored at -80 ºC until further analyses.

6.3.2. Experimental design

A tiered battery of experiments was performed to assess the effects of medium and natural external (environmental) factors on absorption and fluorescence spectra of crude and purified protein extracts from mucus, namely buffer, redox status, pH, calcium concentration and salinity. Samples were first dialysed to seawater, PBS pH 7.4 or 0.05 M Tris-HCl pH 7. The buffer that offered highest fluorescence and reduced inter-sample variability was selected as vehicle to test the effect of hydrogen peroxide (H_2O_2, 2 % (v/v)) and L-dithiothreitol (20 mM DTT) as oxidising and reducing agents, respectively. The pH of the buffer was then modulated (pH 4, 7 and 9) to determine the most adequate pH following the same criteria. Samples were then subjected to 10 mM and 100 mM calcium chloride ($CaCl_2$) and 120 mM and 600 mM sodium chloride (NaCl), corresponding to

physiological and seawater conditions, respectively. All samples were normalized for the same protein concentration (1 mg. mL^{-1}) before assays, which were always performed in triplicate. Replicates consisted of independent mucus samples. All analyses were conducted at room temperature.

Absorption spectra were obtained with an Evolution 300 Spectrophotometer (Thermo Scientific) with 2 nm bandwidth. Emission spectra were analysed on a Varian Cary Eclipse Fluorescence Spectrophotometer (Agilent) with 5 nm bandwidth excitation and emission slits in a 3 mm optical path quartz cuvette (HELLMA, Germany). Emission spectra and maxima were determined following excitation at wavelengths determined from the absorption spectra maxima.

6.3.3. Molecular characterization

Protein separation

Purified mucus samples 0.05 M Tris-HCl pH 7 were analysed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) to separate peptides by molecular mass and isolate the fluorescent band for tandem mass spectrometry analyses (MS/MS). Gels (8×9 cm by 0.75 mm thick) were based on the discontinuous system developed by Laemmli (1970), as detailed by Hames (1998). In brief: samples were mixed with sampling buffer supplemented with β-mercaptoethanol (1:1), boiled for 2 min (for protein denaturation) and loaded to an acrylamide gel composed of stacking and running gels containing 6 % (0.5M Tris HCl pH 6.8) and 12 % (1.5M Tris-HCl pH 8.8) acrylamide, respectively. The electrophoresis ran between 90-110 in a denaturing buffer containing 25 mM Tris, 192 mM glycine, pH 8.5. The Protein Marker II 11-245 kDa (NZYtech) was used as molecular standard. After identification of the fluorescent bands in a Gel-Doc 2000 documenting system equipped with a UV transilluminator (BIO-RAD), the bands were excised and stored at -80 ºC until further analysis.

Protein identification (LC-MS-MS)

After gel digestion with porcine trypsin, the band was analysed on a NanoLC Ultra 2D separation system (Eksigent) coupled to a Triple TOFTM 6600 System mass spectrometer (Sciex). The chromatographic separation was performed by the Column ChromXP C18CL (0.3 x 150 mm, 3 μm, 120 Å, Eksigent) at 50 °C. The flow rate was set to 5 μL. min^{-1} and mobile phase A and B were 0.1 % formic acid plus 5 % (v/v) DMSO in water and 0.1 % (v/v) formic acid plus 5 % (v/v) DMSO in acetonitrile, respectively. The ionization source (ESI DuoSprayTM Source) was operated in the positive mode set to an ion spray voltage of 5500 V, 25 psi for nebulizer gas 1 (GS1), 25 psi for the curtain gas (CUR). Rolling collision was used with a collision energy spread of 5. Protein and peptide identification were performed using ProteinPilot (Sciex) considering the following

parameters: cysteine alkylation by acrylamide, digestion by trypsin, and gel-based ID as a special factor. Data acquired for each sample were submitted to a database containing the UniProtKB databases, all manually reviewed: venom proteins and toxins, Annelida, Conotoxin, Mollusca and Polychaeta.

Protein identification (RNA-Seq)

The peptide sequences obtained by MS-MS were then contrasted with *Eulalia*'s proboscis transcriptome described in chapter 3, deposited in Gene Expression Omnibus (GEO) under the accession number GSE143954. The translated cDNAs were contrasted against NCBInr database using Blast (Altschul et al. 1990). Best hits were selected by number of matching peptides, coverage, and lowest *e*-value.

6.3.4. Histology

Specimens were fixed in 2 % (v/v) glutaraldehyde (in 0.1 mM sodium cacodylate buffer, pH 7.4) during 2 h or in Zenker's solution (2.5 % (w/v) potassium dichromate, 3 % (w/v) mercury chloride, 1 % (w/v) sodium sulphate and 5 % (v/v) glacial acetic acid) overnight. Worms were then divided in serial sections, washed in cacodylate buffer (3 × 15 min), dehydrated in a progressive series of ethanol (30–100 %), intermediately infiltrated with xylene and embedded in paraffin (Paraplast) wax. Paraffin-embedded samples were sectioned (5 μm thickness) with a Jung RM2035 model rotary microtome (Leica Microsystems). Histological sections were stained using a tetrachrome stain (TC) and Haematoxylin-Eosin (HE). The tetrachrome protocol is based on Alcian Blue pH 2.5 (AB) for acidic sugars, Periodic Acid/ Schiff's (PAS) for neutral polysaccharides, Weigert's iron Haematoxylin (WH) for chromatin, and Picric Acid (PA) as counterstain for muscle and cytoplasm. The procedures follow the protocols described by Costa (2018).

Other specimens were snap-frozen in liquid nitrogen, divided in six sections and placed in optimal cutting temperature (OCT) medium in an appropriate tissue mould. The OCT medium with the tissue was then frozen and longitudinal sections with 5-15 μm thick were cut in a CM3600 XP cryomacrotome (Leica Biosystems) at -20 °C. Sections were transferred to pre-adhesivated slides (Thermo Scientific Superfrost Ultra Plus) and stored at -80 °C until analyses. Slides were then treated with H_2O_2 (2 % v/v) and DTT (20 mM) to assess changes to tissue fluorescence. All slides were visualised in a DM 2500 LED model microscope adapted for epifluorescence with an EL6000 light source for mercury short-arc reflector lamps. The microscope was equipped with A, N2.1 and I3 filters (corresponding to blue, red and green channels, respectively). All equipment was supplied by Leica Microsystems.

6.3.5. Immunohistochemistry

Paraffin and cryopreserved sections were used to analyse immunohistochemically ubiquitin B in *E. viridis*, according to Costa (2018). In brief, after deparaffination, both sections were rehydrated in PBS and permeabilized for 15 min with 0.1 % Triton X-100 in PBS. Antigen blocking was done for during 30 min, in the dark at room temperature, with 2 % bovine serum albumin (BSA) in PBS with 0.1 % (v/v) Triton X-100. Slides were then incubated with rabbit anti-human ubiquitin B polyclonal antibody (Invitrogen #PA5-95195) at 1 µg. mL^{-1} in 2 % (w/v) BSA in PBS with 0.1 % (v/v) Triton X-100. Incubation was done overnight at 4 °C in a humidity chamber. Slides were afterwards incubated with the secondary antibody incubation for 2 h, the goat anti-rabbit IGg Alexa Fluor 568 (Invitrogen), with 10 µg. L^{-1} in 2 % (w/v) BSA in PBS with 0.1 % (v/v) Triton X-100. Nuclear staining was done with DAPI. Slides were visualized as previously described.

6.3.6. Fluorochrome internalisation assay

Internalisation of the purified mucus was assessed through an *in vitro* assay using the ovarian carcinoma cell line A2780 (purchased from Sigma-Aldrich) taking advantage of the natural fluorescence of the target mucosubstances. Briefly, cells were grown in McCoy's medium modified as Roswell Park Memorial Institute medium (RPMI 1640) and maintained at 37 °C in a humidified atmosphere with 5 % (v/v) CO_2. Cells were exposed to the purified extract (1 mg. mL^{-1} in PBS) supplemented and with redox modulators (2 % (v/v) H_2O_2 and 20 mM DTT) and controls during 1 h, 3 h and 6 h, at 37 °C. After the incubation period, cells were washed with PBS and visualised in an Eclipse Ti microscope equipped with a DS-QiMc camera and adapted for epifluorescence (Nikon Corporation).

6.4. Results

6.4.1. Molecular characterisation

Purified protein extracts from mucus in unstained SDS-PAGE gels yielded strong blue-green fluorescence when subjected to UV transillumination (Fig. 6.1A). Gels revealed two almost joint main fluorescent bands (with 48 kDa, approximately) with a smear of dimmer bands corresponding to lower molecular weight peptides. The excised main bands showed, by matching MS-MS results with sequences obtained from analysing *Eulalia´*s transcriptome, including proboscis and epidermis (Fig. 6.1B), the presence of three peptides and proteins (Fig. 6.1C, Table S6.1): polyubiquitin, peroxiredoxin and 14-3-3. Both polyubiquitin and 14-3-3 proteins have very characteristic domains (ubiquitin-like domain and 14-3-3-beta-zeta domain). Peroxiredoxin (Prx) was cross-validated through the identification of a conserved domain belonging to the 2-Cys Prx subfamily. Polyubiquitin was present in four distinct isoforms (Fig. 6.1C).

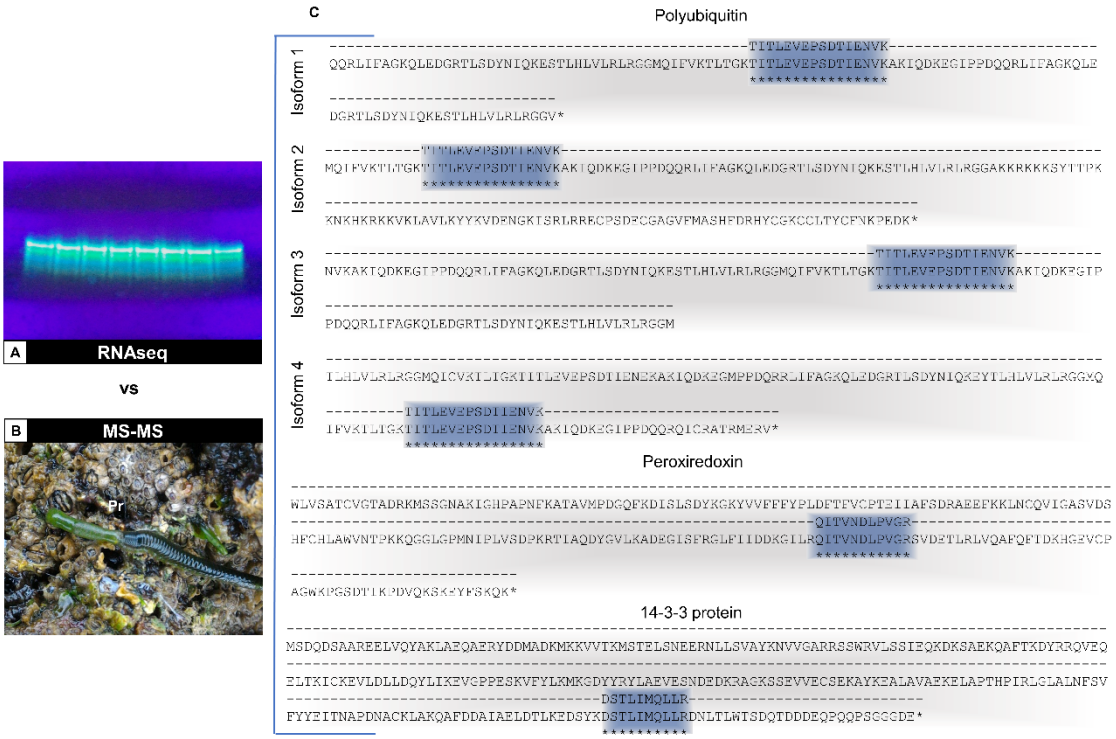
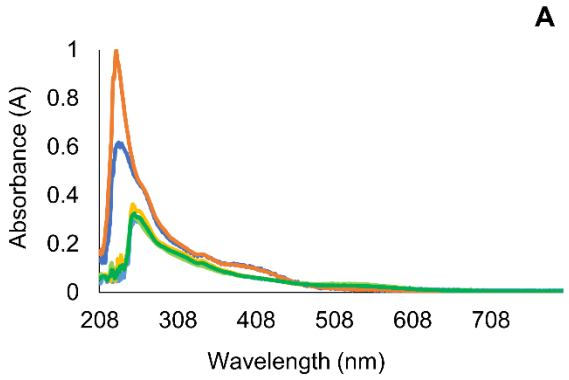


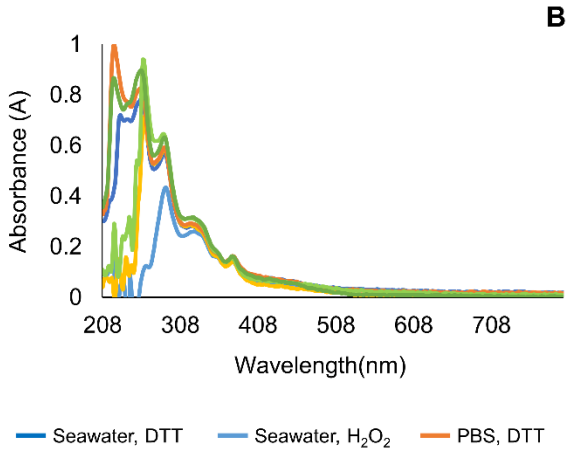
Figure 6.1. Identification of proteins responsible for the fluorescence of the mucus secreted by the marine Polychaeta *E. viridis*. (A) *Eulalia* in the wild with the proboscis (Pr) everted. (B) Photograph of an SDS-PAGE gel visualised trough trans-UV, evidencing the fluorescent bands of the purified protein extract. (C) Alignment of the peptides and translated mRNAs obtained from MS-MS and RNA-seq, respectively. Similarities are indicated by *.

6.4.2. Modulation of absorption and fluorescence

Differences were found between the absorption spectra of crude and purified extracts excited with light within the UV-A range (200-800 nm) (Fig. 6.2). Crude mucus samples yielded a single maximum at 230-260 nm depending on the buffer whereas purified mucus yielded multiple absorption peaks, namely at 225, 260 and 286 nm. The absorption spectra of crude samples were similar between treatments with different buffers/media and redox agents (H_2O_2 or DTT), while the spectra of purified samples had great variability at shorter wavelengths (200 to 300 nm). This variation was mostly redox agent-dependant.



Buffer	Peaks (nm)				
Seawater, H_2O_2	-	255	320	340	-
Seawater, DTT	230	255	320	340	390
PBS, H_2O_2	250	260	-	340	-
PBS, DTT	230	260	290	330	390
Tris-HCl, H_2O_2	250	-	-	340	400
Tris-HCl, DTT	250	260	-	330	390



Buffer	Peaks (nm)					
Seawater, H_2O_2	-	260	286	320	355	374
Seawater, DTT	225	260	286	320	-	374
PBS, H_2O_2	-	260	286	320	355	374
PBS, DTT	225	260	286	320	355	374
Tris-HCl, H_2O_2	-	260	286	320	355	374
Tris-HCl, DTT	225	260	286	320	355	374

Figure 6.2. Averaged absorbance spectra of mucus samples in different buffers (seawater, PBS and Tris-HCl pH 7) and redox treatments (H_2O_2 and DTT) and the respective peaks. (A) Crude mucus samples. (B) Purified mucus samples. Experiments were done in triplicate with independent extracts. Concentration of extracts was normalised to 1 mg total protein. mL^{-1}.

Each sample was then excited with light corresponding to the respective absorption maxima. The highest emission intensity in crude mucus samples was obtained in oxidised spectra excited at 260 nm (Table S6.2), but at around 330 nm the spectra was more consistent for both crude (Fig. S6.1) and purified mucus samples (Table S6.3, Fig. S6.2). Overall, the emission spectra were more stable

at around 330 nm than other excitation wavelengths for crude and purified samples (Fig. 6.3). The emission maxima ≈ 400 nm was consistent, whereas maxima at 505-510 nm, responsible for blue-greenish fluorescence, was only present in purified extracts (Fig. 6.3B, indicated by the arrow). Moreover, in purified samples, the redox status was a key factor for the presence of the maxima at ≈500 nm in samples treated with DTT (Fig. S6.3). This peak was consistent in all the replicates (Fig. S6.2). Overall, reduced samples in Tris-HCl yielded higher fluorescence (Fig. S6.3, Fig. 6.3B).

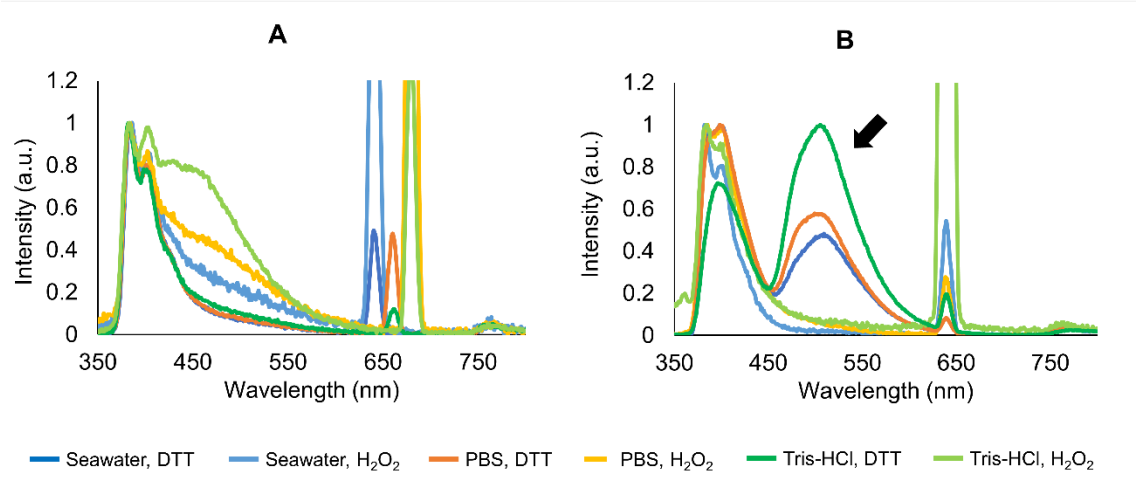
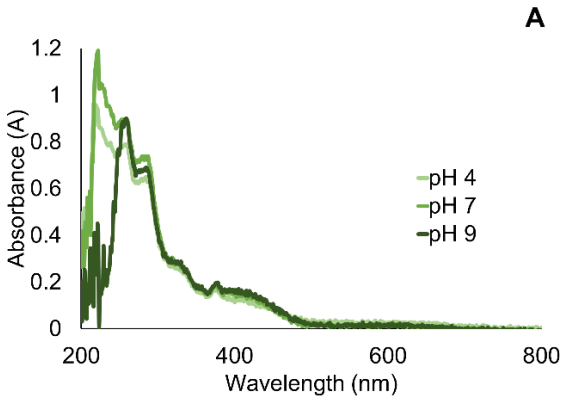
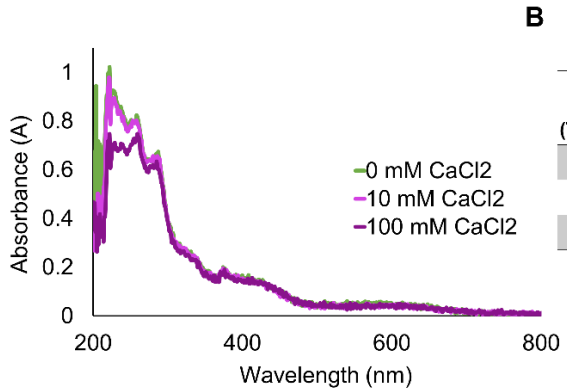


Figure 6.3. Averaged and normalised emission spectra of crude (A) and purified by ultrafiltration (B) mucus samples in different media buffers (seawater, PBS and Tris-HCl buffers, both pH 7) and redox treatments (H_2O_2 and DTT). Samples were excited at 320-340 nm, corresponding to absorbance maxima. Arrow is indicative of visible fluorescent maxima. Experiments were done in triplicate (three independent purified extracts). Concentration of extracts was normalised to 1 mg total protein. mL^{-1}.

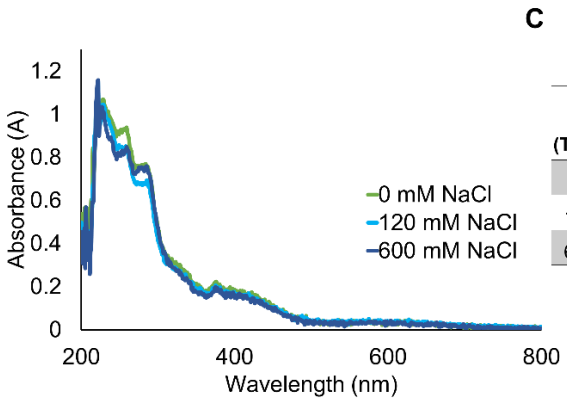
The combined effects of modulating pH, calcium and salinity on absorption spectra are present in Fig. 6.4. The absorption spectra of the purified extracts in Tris-HCl buffer (supplemented with DTT) with different pH showed little changes, albeit a trend to increase absorption at lower wavelengths with decreasing pH, which led to the absence of a peak at ≈ 220 nm at pH 9 (Fig. 6.4A). As such, due to the more consistent results provided by Tris-HCl pH 7, this medium was elected for subsequent analyses with different concentrations of $CaCl_2$ and NaCl. Still, purified extracts in Tris-HCl buffer pH 7 with different concentrations of Ca (as $CaCl_2$) and salt (as NaCl) yielded negligible differences between absorption spectra (Figs. 6.4B and 6.4C, respectively).



Buffer (Tris-HCl)	Peaks (nm)					
pH 4	219	257	284	320*	355*	377
pH 7	222	257	287	320*	355*	376
pH 9	-	259	285	320*	355*	378



Buffer (Tris-HCl, pH 7)	Peaks (nm)					
0 mM $CaCl_2$	220	258	287	320*	355*	374
10 mM $CaCl_2$	221	258	286	320*	355*	374
100 mM $CaCl_2$	222	259	285	320*	355*	376



Buffer (Tris-HCl, pH 7)	Peaks (nm)					
0 mM NaCl	222	259	284	320*	355*	376
120 mM NaCl	222	258	285	320*	355*	375
600 mM NaCl	222	259	285	320*	355*	375

Figure 6.4. Averaged absorbance spectra of purified mucus samples in Tris-HCl supplemented with 20 mM DTT buffer. (A) Different pH (4, 7 and 9). (B), (C) Samples in Tris-HCl pH 7 complemented with different concentrations of $CaCl_2$ and NaCl, respectively. Experiments were done in triplicate (three independent purified extracts). Concentration of extracts was normalised to 1 mg total protein mL^{-1}.

Increasing buffer pH led to an overall increase in fluorescence intensity, with a peak at 496 nm corresponding to green-blueish light, higher after excitation at 330 nm (Figs. 6.5A, B and Fig. S6.4 for absolute values). However, despite strongest emission, extracts in Tris-HCl pH 9 yielded higher inter-replicate variation regardless of excitation wavelength (Fig. S6.5). For these reasons, as previous, Tris-HCl pH 7 was selected as medium for the subsequent tests. In accordance with the reduced variation in absorption spectra, the emission spectra of samples treated with different concentrations of Ca and NaCl yielded reduced variation. Calcium caused a small decrease of

emission intensity regardless of concentration, compared to control (Figs. 6.5C-F). Similarly, salinity did not produce an obvious effect in emission (Figs. 6.5E, F). Both treatments yielded, however, noticeable inter-replicate variation (Figs. S6.6 and S6.7) that was verified by SDS-PAGE as well (Fig. 6.6), revealing then a variation in emission intensity down to 24% among replicates normalised to 1 mg total protein. mL^{-1}).

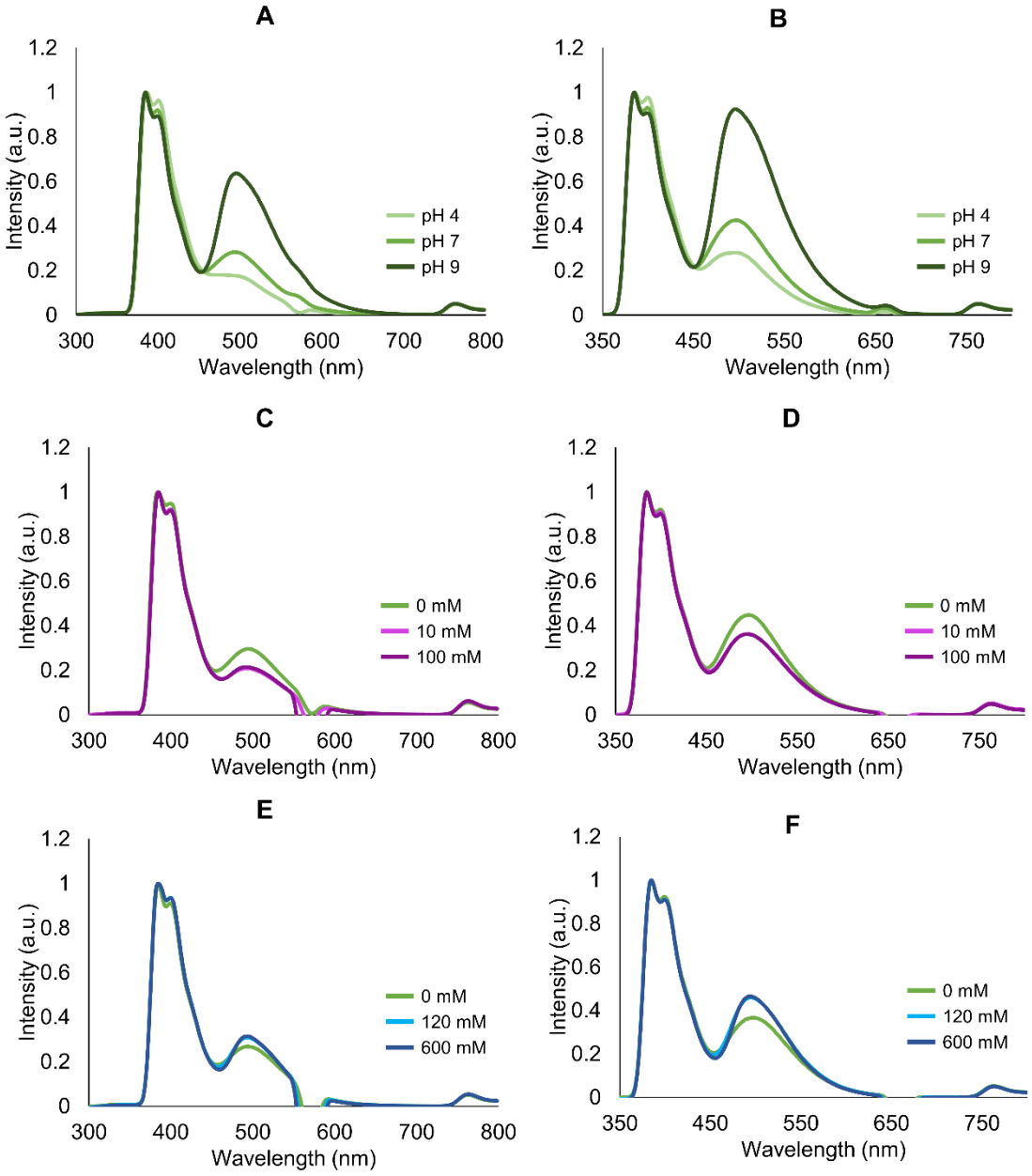


Figure 6.5. Averaged and normalised emission spectra of purified mucus samples excited at 285 nm and 330 nm. Samples comprised of protein extracts in Tris-HCl buffer supplemented with 20 mM DTT, tested with different pH, salinity and Ca concentration. (A), (B) pH 4, 7 and 9, excited at 285 and 330 nm, respectively. (B), (C) calcium (as $CaCl_2$), 0, 10 and 100 mM, excited at 285 and 330 nm, respectively (pH 7). (E), (F) NaCl, 0, 120 and 600 mM, excited at 285 and 330 nm, respectively (pH 7). Experiments were done in triplicate (three independent purified extracts). Concentration of extracts was normalised to 1 mg total protein mL^{-1}.

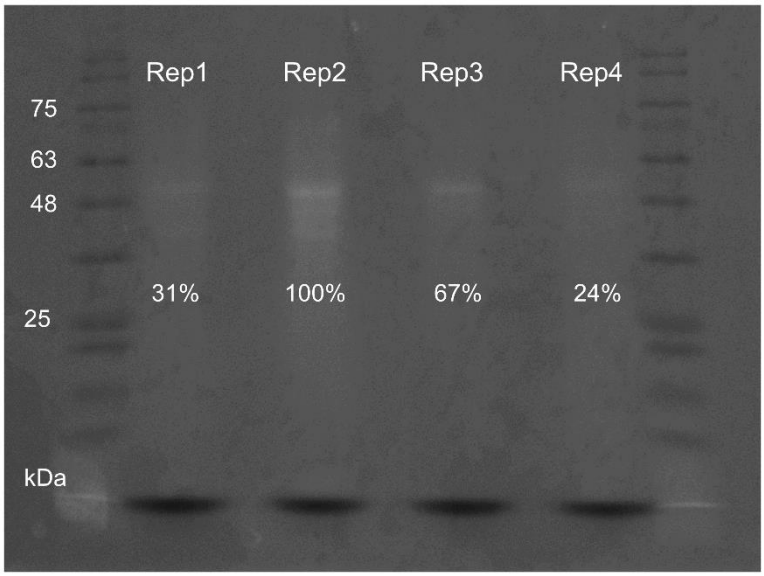


Figure 6.6. SDS-PAGE of purified protein extracts from *Eulalia*'s mucosecretions, visualised by trans-UV (unstained). Four different samples (identified as replicates 1 through 4), were prepared in Tris-HCl buffer (pH 7), supplemented with 20 mM DTT. The amount of total protein was normalised to 1 mg. mL^{-1}. Relative fluorescence intensity is expressed as percentage.

6.4.3. Histological localisation of secretion

Fluorescent secretions were histologically localised in secretory cells in the pharynx at the proboscis after treatment with DTT (unstained) and excitation with UV light (Figs. 6.7A, B). Fluorescent secretions were also revealed in cryosections treated with DTT and left unstained (Figs. 6.7C, D). In this case, fluorescent signal was detected in cells as well as in the adjacent mucus layer. The secretion of one of the isoforms (isoform 3, in Fig. 6.8A) of polyubiquitin was localized immunohistochemically (Fig. 6.8). The presence of this isoform was noticed mainly in parapodia and in some defence cells, i.e., coelomocytes (Figs. 6.8B, C). Additionally, polyubiquitin was also detected in initial stage oocytes in the oocyte cuticle (Figs. 6.8E to J).

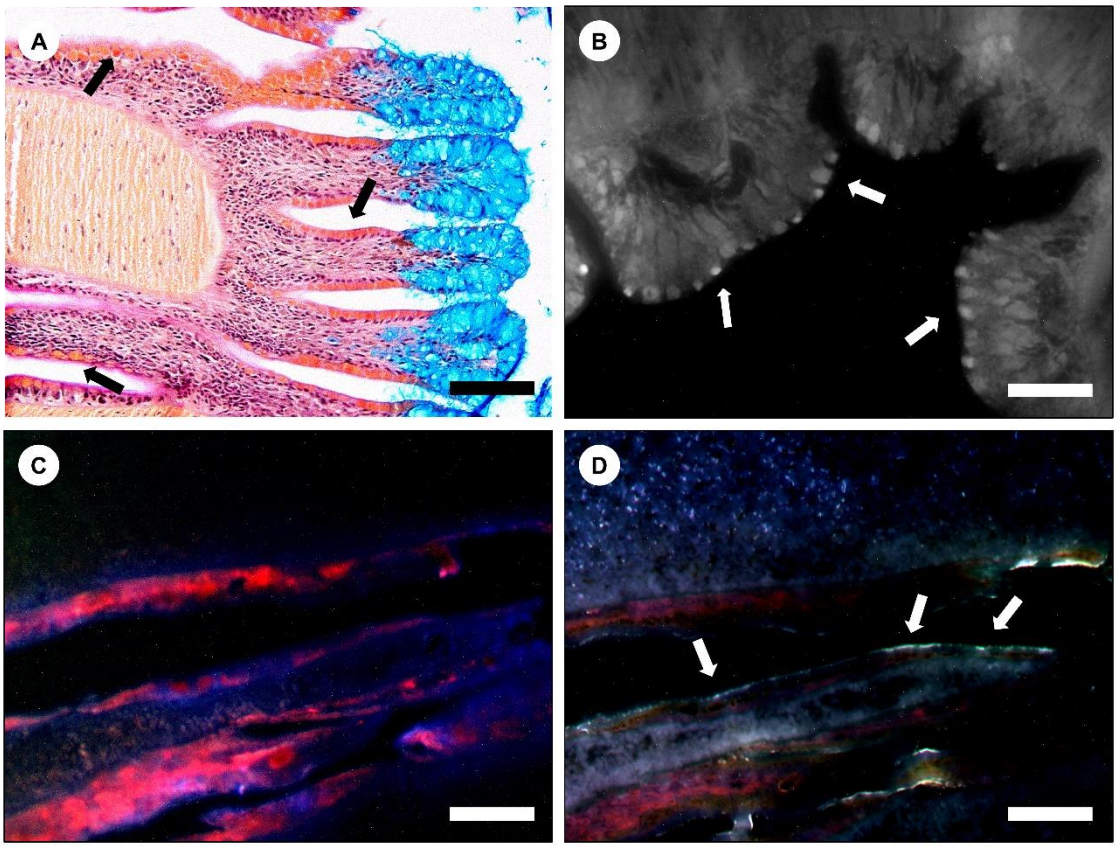


Figure 6.7. Localisation of mucosecretions in histological sections of *Eulalias*'s proboscis through their fluorescent signal. (A) Histological section (fixed with glutaraldehyde) stained with a tetrachrome stain, showing the tentacles in the anterior section of the pharynx. Arrows are indicative of secretory calyx (serous) cells. Scale bar: 50 µm. (B) Histological section fixed with glutaraldehyde and treated with DTT (20 mM). Arrows show fluorescent secretory calyx cells. Scale bar: 25 µm. (C) Longitudinal cryosection treated with 2 % H_2O_2 (v/v). Scale bar: 50 µm. (D) Longitudinal cryosection treated with DTT (20 mM). Arrows are indicative of fluorescence signal lining the inner epithelium of the pharynx. Scale bars: 50 µm.

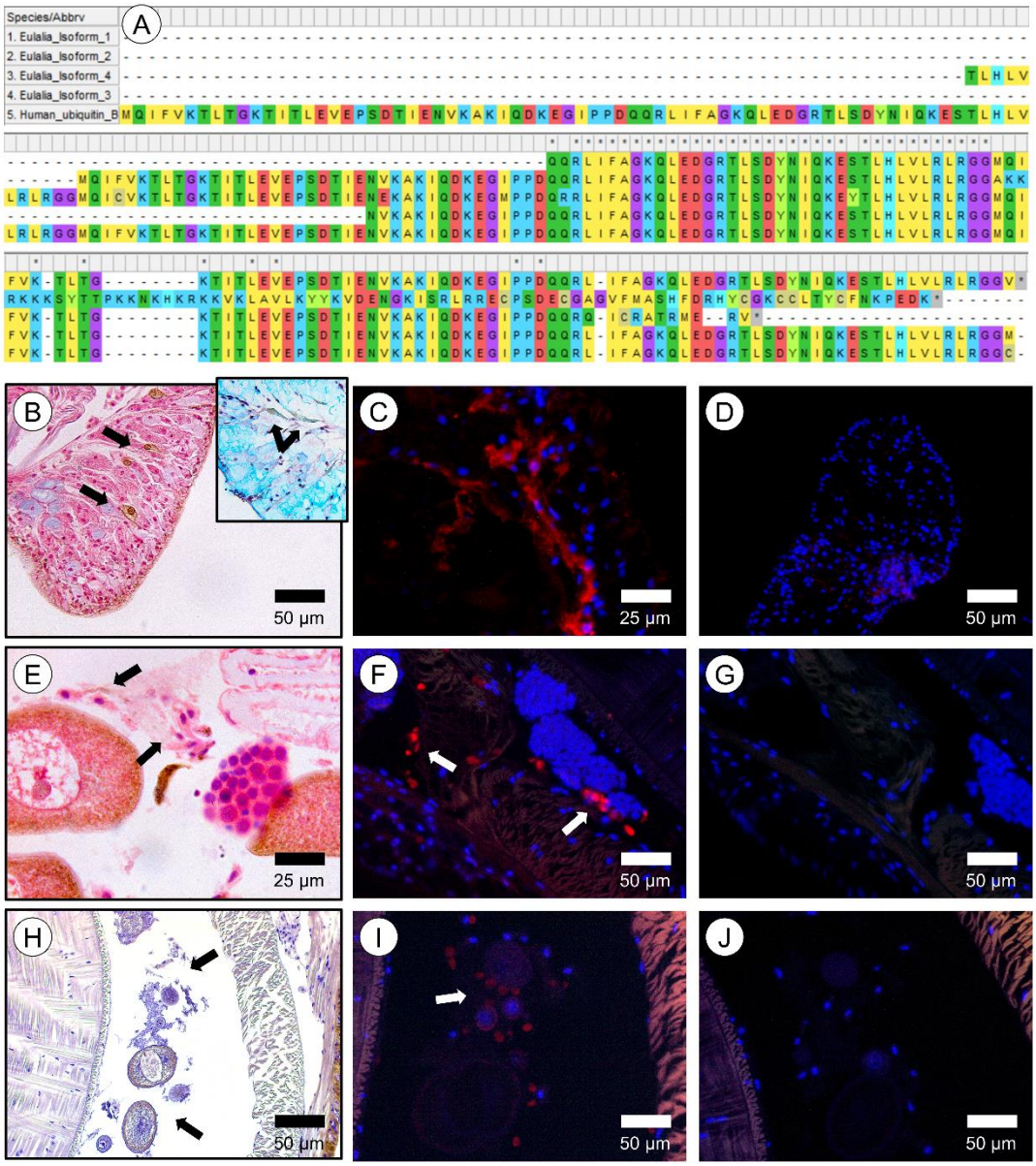


Figure 6.8. Immunohistochemical localisation of ubiquitin B in *Eulalia*. (A) Alignment of peptides from MS-MS on fluorescent SDS-PAGE gel bands, RNAseq sequences with positive matches to ubiquitin and the human ubiquitin B sequence against which the polyclonal antibody was used. The antibody was chosen based on similarity with *Eulalia* ubiquitin isoform 3. (B) Paraffin section across a parapodium of the worm fixed with Zenker and stained with HE (haematoxylin and eosin). Haemocytes are identified by the arrows. Inset: Paraffin section fixed with glutaraldehyde and stained with TC (tetrachrome stain) showing haematocytes naturally-coloured by green pigments. (C) Compositive of a cryopreserved section of a parapodia marked for ubiquitin B. (D) Negative control of cryopreserved section of a parapodia. (E) Paraffin section fixed with Zenker and stained with HE. Arrows are indicative of stem cells. (F) Composite of a paraffin section of the ventral lateral section marked for ubiquitin B. Arrows indicate the stem cells marked by the antibody. (G) Negative control of cryopreserved section of the ventral lateral area, close to the parapodia. (H) Paraffin section fixed with glutaraldehyde and stained with HE, showing the female gametes in the celomic cavity. (I) Composite of celomic cavity with positive signal for ubiquitin B (arrow). (J) Negative control of cryopreserved section of the celomic cavity.

6.4.4. Internalisation by A2780 cells

After 3 h exposure, we observed increased fluorescent signal within A2780 cells, more intense in cells treated with DTT (Fig. 6.9). Residual fluorescence could be noticed following treatment with H_2O_2 and PBS. The fluorescent signal in cells treated with DTT increased in a time-responsive manner as early as 1 h and after 6 h was evident in all cells. Conversely, only after 3 h of exposure the signal could be found in cells treated with H_2O_2 (Fig. S6.8). The mucus gave a reddish fluorescence to cells after internalisation, whereas in controls the cells presented a bluish fluorescence. The mucus cytotoxicity was assessed in the same cells used for cytotoxic assays (chapter 5).

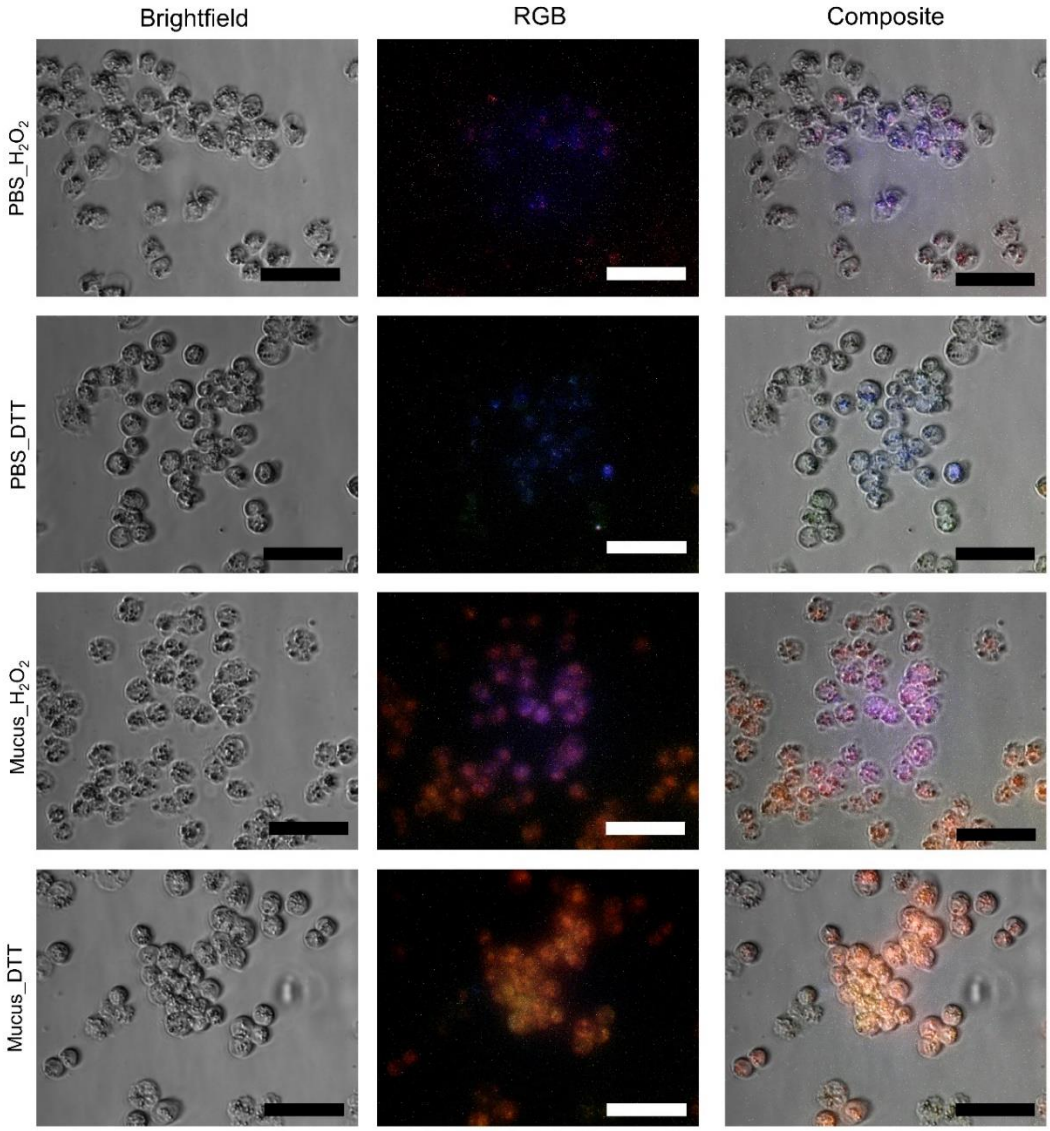


Figure 6.9. Internalisation assay for purified mucosubstances onto the human ovarian cancer cell line A2780. Cells were incubated during 3 h with purified mucus extract prepared with PBS (pH 7.4) and the respective control (PBS only). Either experimental condition was subjected to reduced or oxidising agents (DTT and H_2O_2). Composite images were produced by overlapping red, green and blue channels with brightfield images.

6.5. Discussion

The mucus secreted by *Eulalia* bears a proteinaceous complex that yields strong fluorescence when excited with UV light. Emission, within the range of the blue-green spectrum, is most significantly modulated by redox status, increasing when the mixture is reduced and decreasing upon oxidation, reversibly, therefore acting like a biochemical switch. Another important factor is pH, as emission tends to increase with alkalinity. However, normal physiological pH seems to offer a good compromise between emission intensity and inter-sample consistency. The results also indicate that, aside shift in redox status, vehicle (buffer) is versatile, with the compounds retaining their basic properties both seawater and physiological media, as well as in Tris-HCl, one of the buffers of choice to extract proteinaceous substances in the laboratory which may indicate potential biotechnological interest. Altogether, these findings suggest that the worm actively secretes these compounds to the mucus, remaining fluorescent in seawater until oxidation, with salinity and Ca concentrations rendering little or null impact on emission. The fluorescent complex likely contributes to protection against UV light by absorbing UV radiation and releasing energy in the form of visible light, holding antioxidant properties as well. These properties confer advantages to an intertidal diurnal forager like *Eulalia*, who actively roams the rocky intertidal in search for prey. Other functions for the fluorescence, such as communication and warning remain speculative at this stage. It must be noted, though, that most reports on luminescence in marine invertebrates focus on bioluminescence. Indeed, there is scant information on the secretion of fluorescent substances by Polychaeta or the Annelida in general (see Rodrigo and Costa 2019 for further details). Nevertheless, the ecological purposes of bioluminescence and fluorescence, such as warning or defence, can be similar (Widder 2010). In the particular case of *Eulalia*, the uncanny bright green pigmentation provided by the accumulation of porphyrinoid pigments, which absorb UV light, already indicated that these animals are particularly well-fit to rove through the rocky intertidal (Martins et al. 2019). On the other hand, the copious secretion of mucus by the worm has multiple roles, namely lubrication during locomotion, protection against desiccation and toxin delivery (Rodrigo et al. 2018), to which may now be added important functions against UV and oxidation.

The emission spectra of purified *Eulalia* mucus extracts presents some similarities with other organisms secreting green fluorescent proteinaceous compounds, such as a maximum at around 500 nm, from the flowers of *Achillea millefolium* and the roots of *Raphanus sativus* (Roshchina et al. 2011) to wild-type GFP from *Aequorea* (see Tsien 1998). Such similarities may indeed confirm the proteinaceous nature of *Eulalia* fluorochromes. Tsien (1998) also noticed variations of the emission peak with the excitation wavelength, indicative of, at least, two chemically-distinct species interfering with fluorescence (see Table S6.7 for averaged maxima). We recorded a similar occurrent in *Eulalia*, as the fluorescent peak (around 500 nm) shows small variations that depend

mostly on excitation wavelength. This is no surprise in *Eulalia*, due to the presence of several proteins in the extract (recall chapter 3), three of which were positively identified by mRNA sequencing and protein MS-MS, namely ubiquitin, 14-3-3 protein and peroxiredoxin, even though the exact nature of the fluorescent proteins or peptides (or eventual complexes) remains elusive.

Ubiquitin is a small c.a. 8.5 kDa regulatory protein (comprised approximately of 76 amino acids) well-conserved among eukaryotes. Ubiquitin can be found conjugated with other proteins or in free form as polyubiquitin. When conjugated, it can be linked to lysin residues of target proteins and in rare cases, conjugated to cysteine or serine residues (Catic and Ploegh 2005). There are several ubiquitin-binding domains characterized by the type of proteins they bind to, having associations ranging from zinc finger domain to domains typical of endoplasmic reticulum degradation (see Hicke et al. 2005 for a review). Ubiquitin is mostly associated with the function of the proteasome by tagging misfolded or damaged proteins for lysis (Wang and Maldonado 2006). Even though there is scant information on the function of ubiquitin in animal secretions, it is found in the venom of some species, such as snakes, as a non-toxic protein with unknown function (Melani et al. 2016). In *Eulalia* an isoform homolog to human and mouse ubiquitin B was chiefly identified in defence cells, coelomocytes or static cells in the parapodia and in membrane of female gametes (recall Fig. 6.8). This isoform (isoform 3) is the most conserved of the isoforms identified in the fluorescent band. This form is part of the innate immune response, being involved in related cell signalling cascades in humans (Ebner et al. 2017) and other organisms, which explains its presence in *Eulalia* defence cells (recall Fig. 6.8B, C). Even though this isoform was not immunolocalized in the proboscis, its secretion is notorious in the parapodia, which are major players in mucus secretion. We may, therefore, infer a relevant immune function of the mucus and the role of ubiquitin. If indeed associated to fluorescence, polyubiquitins may explain the multiple, slightly different molecular weight bands observed in gels (Fig. 6.1B).

Peroxiredoxins (22-27 kDa) are important antioxidant proteins. These proteins reduce a range of hydroperoxides to protect against reactive oxygen species (ROS). Peroxiredoxins can be found in almost every animal tissue where they can act as peroxide scavenging enzymes, as well as being key players in immunity and as regulators of cell death (see Abbas et al. 2019 for a review). They can also be found in the secretion of parasitic nematodes to avoid the host immune response (Price et al. 2019). Their presence in *Eulalia*'s mucus confirm its antioxidant properties and may offer a link to the modulation of fluorescence by changes in redox potential. In their turn, 14-3-3 (28–30 kDa) are conserved in most eukaryotic organisms that can bind to several different proteins and thus be involved in important cellular processes such as signal transduction, cell-cycle control, apoptosis and stress response (see Van Hemert et al. 2001, for further details). When found in mucus, 14-3-3 proteins are linked to the immune response of fish and thought to be related with

anti-microbial resistance in fish (such as the zebrafish and lumpsucker) and insects (Ulvila et al. 2011; Beck and Peatman 2015; Patel and Brinchmann 2017).

The association between these three proteins in mucosecretions of Polychaeta as well as their role in fluorescence has not been reported before. However, their presence in mucus secretions has been described in fish, cephalopods and even in flatworms, as part of complex cocktails of substances (Rajan et al. 2011; Bocchinfuso et al. 2012; Caruana et al. 2016; Patel and Brinchmann 2017). Additionally, there are reports of the presence of ubiquitin C and 14-3-3 protein in the integument of the parasitic flatworm *Schistosoma japonicum*, presumably with a role in maintenance of tegument integrity as part of the defence mechanisms against the host response (Liu et al. 2019). Curiously, the interactions between ubiquitin and 14-3-3 protein has been identified in tomato plant as being involved in cell wall-related metabolism, more specifically, cell wall building and resistance, degradation of xyloglucan, among other functions (Lu et al. 2016). In fact, the connection between 14-3-3 and ubiquitin may have the same protective functions in the cuticle that outlines the inner epithelium of the pharynx of *Eulalia* (recall Figs. 6.2 and 6.8). If considering only these two proteins (ubiquitin and 14-3-3 protein), their size (48 kDa) is approximate the same than the fluorescent band in gels (≈ 50 kDa). If indeed forming complexes, these thus exclude Prx (approximately 22 kDa). It must be noted, though, that ubiquitin can be a chaperon of peroxiredoxins, as demonstrated by the biding abilities of the de-ubiquitinating enzyme ubiquitin C-terminal hydrolase-L1 (UCH-L1) to 2-Cys Prx in *in vitro* (Lee et al. 2018). Nonetheless, their interaction in mucosecretions is unknown. The fact that Prx belongs to a family of peroxidases that bind either thermally or oxidatively to free radicals (Abbas et al. 2019), may explain its influence on redox status of mucus and in its fluorescence. In fact, Prx is ubiquitous in Eumetazoa, including in Polychaeta. In *Arenicola marina*, for instance, Prx is mostly present in tissues directly exposed to the external environment, enhancing its importance in the protection against oxidative stress (Loumaye et al. 2008). In fish, their role in mucus as antioxidative agent has been reported with their part in protection against infections (Valero et al. 2015).

The fluorescence of ubiquitin and its modulation with redox state and pH has been reported elsewhere. For instance, Noronha et al. (2004), found that bovine ubiquitin increase its fluorescence with pH. Curiously, the same pH effect was seen in a GFP-based calcium sensor, a fluorophore very different from ubiquitin, generating two major emission peaks, one around 400 nm and the other around 500 nm (Zhuo et al. 2015). This variation was thus similar to the variation with pH seen in *Eulalia*. In turn, the interference of Ca and Na ions was seemingly negligible on the fluorescence of mucosecretions. There are, however, contradictory reports on the influence of salts on fluorochromes from marine animals. As an example, the fluorescence of hyalin, a fluorescent protein from some echinoderms, has been found to be stabilized by increasing Ca concentrations

up to 1 mM $CaCl_2$ at least (Robinson et al. 1988). On the other hand, the fluorescence of some recombinant GFP is decreased by Ca binding (Romoser et al. 1997). By increasing the concentrations of NaCl (to 100 mM), the fluorescence of hyalin has been increased by almost 45 % (Robinson et al. 1988). In *Eulalia*, concentrations in the same order of magnitude (120 mM) and even higher (600 mM) did not interfere with fluorescence.

The fluorescent peptides and proteins in the mucus are not secreted by mucocytes but rather by specialised calyx cells lining the interior of the pharynx that are also responsible for the secretion of toxins, being then conveyed by the mucus (see Chapter 2 for more details). These results indicate that the secretion of fluorescent substances may be linked to toxin secretion and, therefore, to feeding and defence mechanisms (see chapter 4). However, the secretion of the fluorescent protein in the mucus extract is not proportional to the total amount of protein, which highlights that the mucus is a far more complex mixture of proteinaceous materials than expected (recall Fig. 6.6). When a physiologically-compatible buffer (PBS) is employed as vehicle, these proteins present in the extract can be internalized by human cells and retain their fluorescence, especially if previously treated with a reducing agent. However, internalization shifted the fluorescence signal from blue-greenish to yellow-reddish by undisclosed factors that likely reflect the internal milieu of the target cells (A2780). Some studies try to achieve the same results chemically, obtaining a fluorescent signal both temperate- and time-dependant to study intracellular dynamics (Subach et al. 2009). These changes may be caused by increased temperature (37 ºC), the cells' internal pH or by cells metabolic activity. In any case, internalisation by living cells breaks ground towards application, as it indicates bioreactivity without loss of function, to which we added the uniqueness of modulating redox status as switch.

6.6. Conclusion

The marine Polychaeta *Eulalia viridis* secretes copious amounts of mucus that serve multiple purposes, from lubrication and protection against desiccation to toxin delivery. The secretion of a protein mixture that holds fluorescent properties when exposed to UV light and responds to oxidation suggests that the mucus also contributes to protect this intertidal worm during foraging. Even though the exact chemical nature of the mixture of proteins that make the fluorescent secretion, ubiquitin is seemingly a key player, as well as a peroxiredoxin. Likely, fluorescence is caused by ubiquitin-conjugated complexes rather than specialised fluorescent proteins such as GFP. One of the most interesting properties of this fluorescent proteinaceous complex is the ability to be switched on or off reversibly by reduction or oxidation under physiologically-compatible

conditions, which, together with cellular uptake and the retention of fluorescence in various buffers and in the presence of various salts, may indicate biotechnological potential.

6.7. References

Abbas MN, Kausar S, Cui H (2019) The biological role of peroxiredoxins in innate immune responses of aquatic invertebrates. Fish Shellfish Immunol 89:91–97.

Ainsworth TD, Hoegh-Guldberg O, Leggat W (2008) Imaging the fluorescence of marine invertebrates and their associated flora. J Microsc 232:197–199.

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

Armstrong E, Yan L, Boyd KG, Wright PC, Burgess JG (2001) The symbiotic role of marine microbes on living surfaces. Hydrobiologia 461:37–40.

Bartolomaeus T, Purschke G, Hausen H (2005) Polychaete phylogeny based on morphological data – a comparison of current attempts. In: Bartolomaeus T, Purschke G (eds) Morphology, Molecules, Evolution and Phylogeny in Polychaeta and Related Taxa. Springer, pp 341–356

Beck BH, Peatman E (2015) Mucosal health in aquaculture. Elsevier

Bocchinfuso DG, Taylor P, Ross E, Ignatchenko A, Ignatchenko V, Kislinger T, Pearson BJ, Moran MF (2012) Proteomic profiling of the planarian Schmidtea mediterranea and its mucous reveals similarities with human secretions and those predicted for parasitic flatworms. Mol Cell Proteomics 11:681–691.

Bonnett R, Head EJ, Herring PJ (1979) Porphyrin pigments of some deep-sea medusae. J Mar Biol Assoc United Kingdom 59:565–573.

Caruana NJ, Cooke IR, Faou P, Finn J, Hall NE, Norman M, Pineda SS, Strugnell JM (2016) A combined proteomic and transcriptomic analysis of slime secreted by the southern bottletail squid, *Sepiadarium austrinum* (Cephalopoda). J Proteomics 148:170–182.

Catic A, Ploegh HL (2005) Ubiquitin - Conserved protein or selfish gene? Trends Biochem Sci 30:600–604.

Costa PM (2018) The handbook of histopathological practices in aquatic environments. Academic Press, Cambridge

Deheyn DD, Kubokawa K, Mccarthy JK, Murakami A, Porrachia M, Rouse GW, Holland ND (2007) Endogenous green fluorescent protein (GFP) in amphioxus. Biol Bull 213:95–100.

Ebner P, Versteeg GA, Ikeda F (2017) Ubiquitin enzymes in the regulation of immune responses. Crit Rev Biochem Mol Biol 52:425–460.

Fresneau C, Arrio B, Lecuyear B, Dupaixn A, Lescure N, Vvolfin P (1984) The fluorescent product of scaleworm bioluminescent reaction: an *in vitro* study. Photochem Photobiol 39:255–261.

Haddock SHD, Dunn CW, Pugh PR, Schnitzler CE (2005) Ecology: Bioluminescent and red-fluorescent lures in a deep-sea siphonophore. Science 309:263.

Hames BD (1998) Gel electrophoresis of proteins: A practical approach, Third Edit. Oxford University Press, New York

Hicke L, Schubert HL, Hill CP (2005) Ubiquitin-binding domains. Nat Rev Mol Cell Biol 6:610–621.

Kennedy G, Nicol J (1959) Pigments of *Chaetopterus variopedatus* (Polychaeta). Proc R Soc London B 150:509–538.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

Lagorio MG, Cordon GB, Iriel A (2015) Reviewing the relevance of fluorescence in biological systems. Photochem Photobiol Sci 14:1538–1559.

Lee SP, Park CM, Kim KS, Kim E, Jeong M, Shin JY, Yun CH, Kim K, Chock PB, Chae HZ (2018) Structural and biochemical analyses reveal ubiquitin C-terminal hydrolase-L1 as a specific client of the peroxiredoxin II chaperone. Arch Biochem Biophys 640:61–74.

Liu J, Giri BR, Chen Y, Cheng G (2019) 14-3-3 protein and ubiquitin C acting as SjIAP interaction partners facilitate tegumental integrity in *Schistosoma japonicum*. Int J Parasitol 49:355–364.

Loumaye E, Andersen AC, Clippe A, Degand H, Dubuisson M, Zal F, Morsomme P, Rees JF, Knoops B (2008) Cloning and characterization of *Arenicola marina* peroxiredoxin 6, an annelid two-cysteine peroxiredoxin highly homologous to mammalian one-cysteine peroxiredoxins. Free Radic Biol Med 45:482–493.

Lu Y, Yasuda S, Li X, Fukao Y, Tohge T, Fernie AR, Matsukura C, Ezura H, Sato T, Yamaguchi J, To- T, Fernie AR, Matsukura C, Ezura H, Sato T, Shigetaka Y, Xingwen L, Yoichiro F, R FA, Chiaki M, Hiroshi E, Takeo S (2016) Characterization of ubiquitin ligase SlATL31 and proteomic analysis of 14-3-3 targets in tomato fruit tissue (*Solanum lycopersicum* L.). J Proteomics 143:254–264.

Martins C, Rodrigo AP, Cabrita L, Henriques P, Parola AJ, Costa PM (2019) The complexity of porphyrin-like pigments in a marine annelid sheds new light on haem metabolism in aquatic invertebrates. Sci Rep 9:1–11.

Melani RD, Skinner OS, Fornelli L, Domont GB, Compton PD, Kelleher NL (2016) Mapping proteoforms and protein complexes from king cobra venom using both denaturing and native top-down proteomics. Mol Cell Proteomics 15:2423–2434.

Nicol JAC (1952) Studies on Chaetopterus variopedatus (Renier). I. The light-producing glands. J Mar Biol Assoc United Kingdom 30:417–430.

Noronha M, Lima JC, Bastos M, Santos H, Maçanita AL (2004) Unfolding of ubiquitin studied by picosecond time-resolved fluorescence of the tyrosine residue. Biophys J 87:2609–2620.

Patel DM, Brinchmann MF (2017) Skin mucus proteins of lumpsucker (*Cyclopterus lumpus*). Biochem Biophys Reports 9:217–225.

Price DRG, Nisbet AJ, Frew D, Bartley Y, Oliver EM, McLean K, Inglis NF, Watson E, Corripio-Miyar Y, McNeilly TN (2019) Characterisation of a niche-specific excretory-secretory peroxiredoxin from the parasitic nematode *Teladorsagia circumcincta*. Parasite Vector 12:1–14.

Rajan B, Fernandes JMO, Caipang CMA, Kiron V, Rombout JHWM, Brinchmann MF (2011) Proteome reference map of the skin mucus of Atlantic cod (*Gadus morhua*) revealing immune competent molecules. Fish Shellfish Immunol 31:224–231.

Robinson JJ, Taylor L, Ananthanarayanan VS (1988) Role of calcium in stabilizing the structure of hyalin, a major protein component of the sea urchin extraembryonic hyaline layer. Biochem Biophys Res Commun 152:830–836.

Rodrigo A, Lopes A, Baptista P, Costa M, Fernandes A, Costa P (2019) The great biotechnological potential of a marine polychaete: An alliance between toxin and natural fluorescence. Front Mar Sci Conf Abstr IMMR'18.

Rodrigo AP, Costa PM (2019) The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. Environ Res 173:270–280.

Rodrigo AP, Martins C, Costa MH, Alves de Matos AP, Costa PM (2018) A morphoanatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). J Anat 233:567–579.

Romoser VA, Hinkle PM, Persechini A (1997) Detection in living cells of Ca^{2+}-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. J Biol Chem 272:13270–13274.

Salih A, Larkum A, Cox G, Kühl M, Hoegh-Guldberg O (2000) Fluorescent pigments in corals are photoprotective. Nature 408:850–853.

Santhanam R (2016) Biology and ecology of venomous marine cnidarians. Springer

Shagin DA, Barsova E V., Yanushevich YG, Fradkov AF, Lukyanov KA, Labas YA, Semenova TN, Ugalde JA, Meyers A, Nunez JM, Widder EA, Lukyanov SA, Matz M V. (2004) GFP-like proteins as ubiquitous metazoan superfamily: Evolution of functional features and structural complexity. Mol Biol Evol 21:841–850.

Shimomura O (2006) Bioluminescence: Chemical principles and methods. World Scientific Publishing Co. Pte. Ltd.

Sparks JS, Schelly RC, Smith WL, Davis MP, Tchernov D, Pieribone VA, Gruber DF (2014) The covert world of fish biofluorescence: A phylogenetically widespread and phenotypically variable phenomenon. PLoS One.

Subach FV, Subach OM, Gundorov IS, Morozova KS, Kiryl D, Cuervo AM, Verkhusha VV (2009) Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. Nat Chem Biol 5:118–126.

Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67:509–544.

Ulvila J, Vanha-aho L-M, Kleino A, Vaha-Makila M, Vuoksio M, Eskelinen S, Hultmark D, Kocks C, Hallman M, Parikka M, Ramet M (2011) Cofilin regulator 14-3-3 is an evolutionarily conserved protein required for phagocytosis and microbial resistance. J Leukoc Biol 89:649–659.

Valero Y, Martínez-Morcillo FJ, Esteban MÁ, Chaves-Pozo E, Cuesta A (2015) Fish peroxiredoxins and their role in immunity. Biology (Basel) 4:860–880.

Van Hemert MJ, Yde Steensma H, Van Paul G (2001) 14-3-3 Proteins: Key regulators of cell division, signalling and apoptosis. BioEssays 23:936–946.

Wang J, Maldonado MA (2006) The ubiquitin-proteasome system and its role in inflammatory and autoimmune diseases. Cell Mol Immunol 3:255–261.

Widder EA (2010) Bioluminescence in the ocean: Origins of biological, chemical, and ecological diversity. Science 328:704–708.

Zhuo Y, Solntsev KM, Reddish F, Tang S, Yang JJ (2015) Effect of Ca^{2+} on the steady-state and time-resolved emission properties of the genetically encoded fluorescent sensor catcher. J Phys Chem B 119:2103–2111.

Appendix Chapter 6

Table S6.1. Protein identification by homology using BLASTP.

Protein	Origin	Conserved domain (Pfam)	Protein match	Query cover	e-value (Blastp)
Polyubiquitin	RNA-seq	Ubiquitin-like domain	Polyubiquitin-B (*Nematostella vectensis*)	100%	1E-88
	MS-MS	Ubiquitin-like domain	Polyubiquitin-C (*Nematostella vectensis*)	100%	9E-10
Peroxiredoxin	RNA-seq	Peroxiredoxin family, 2-Cys PRX subfamily	Peroxiredoxin-1 (*Theropithecus gelada*)	94%	2E-147
	MS-MS	Tryparedoxin peroxidase	Peroxiredoxin-1-like (*Dendronephthya gigantea*)	100%	1E-04
14-3-3	RNA-seq	14-3-3 superfamily	14-3-3 zeta (*Pristionchus pacificus*)	95%	3E-121
	MS-MS	-	14-3-3 protein beta/alpha-A (*Stylophora pistillata*)	100%	6E-04

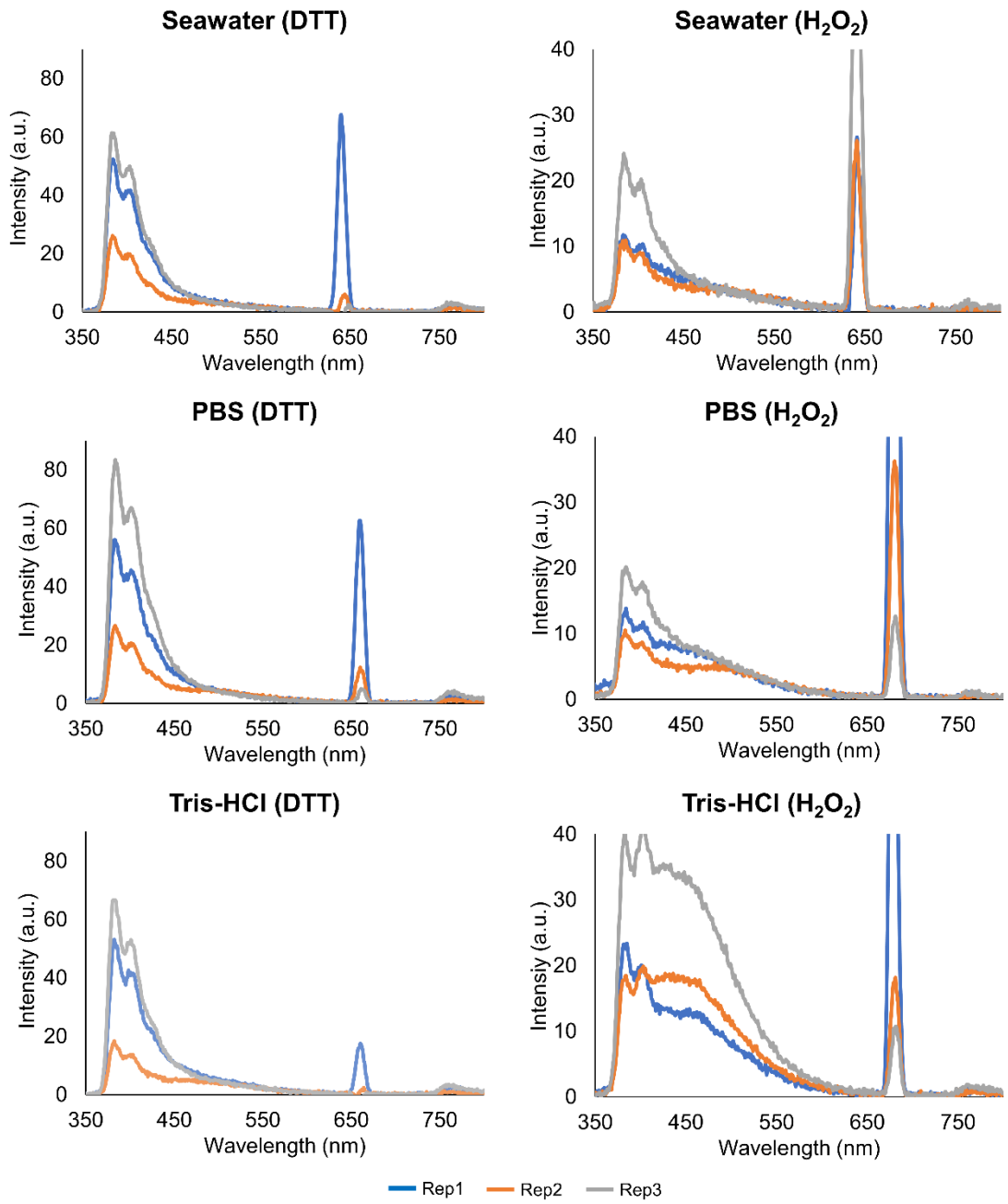


Figure S6.1. Emission spectra of crude mucus samples in different buffers (Seawater, PBS and Tris-HCl, pH 7) and redox treatments (H_2O_2 and DTT). Samples were excited at 320-340 nm. Rep1-3 indicate independent replicates. Concentration of extracts was normalised to 1 mg total protein. mL^{-1}.

Table S6.2. Averaged peak maximum intensities and respective wavelengths of the emission spectra in crude mucus samples. Samples were excited at 260 and 320-340 nm for oxidised (H_2O_2) and reduced samples (DTT). Experiments were done in triplicate with independent extracts.

Buffer	Excitation (nm)	Emission max (nm)	Intensity max (a.u.)	SD (a.u.)
Seawater + H_2O_2	255	384	8.36	2.66
		400	7.37	2.11
	340	383	15.08	4.71
		403	13.07	3.95
PBS + H_2O_2	250	385	7.2	2.9
		401	6.18	2.16
	340	384	15.25	3.76
		402	12.97	3.73
Tris-HCl + H_2O_2	250	384	10.10	3.96
		401	8.63	3.44
	340	383	27.23	9.33
		403	26.66	9.83
Seawater + DTT	255	384	55.74	15.49
		402	44.72	12.76
	340	385	44.06	15.67
		402	35.84	12.89
PBS + DTT	260	384	70.36	26.65
		404	55.89	21.75
	330	384	57.04	23.1
		403	45.76	18.86
Tris-HCl + DTT	260	384	50.21	21.31
		403	39.09	17.02
	330	383	47.83	20.99
		402	37.69	17.11

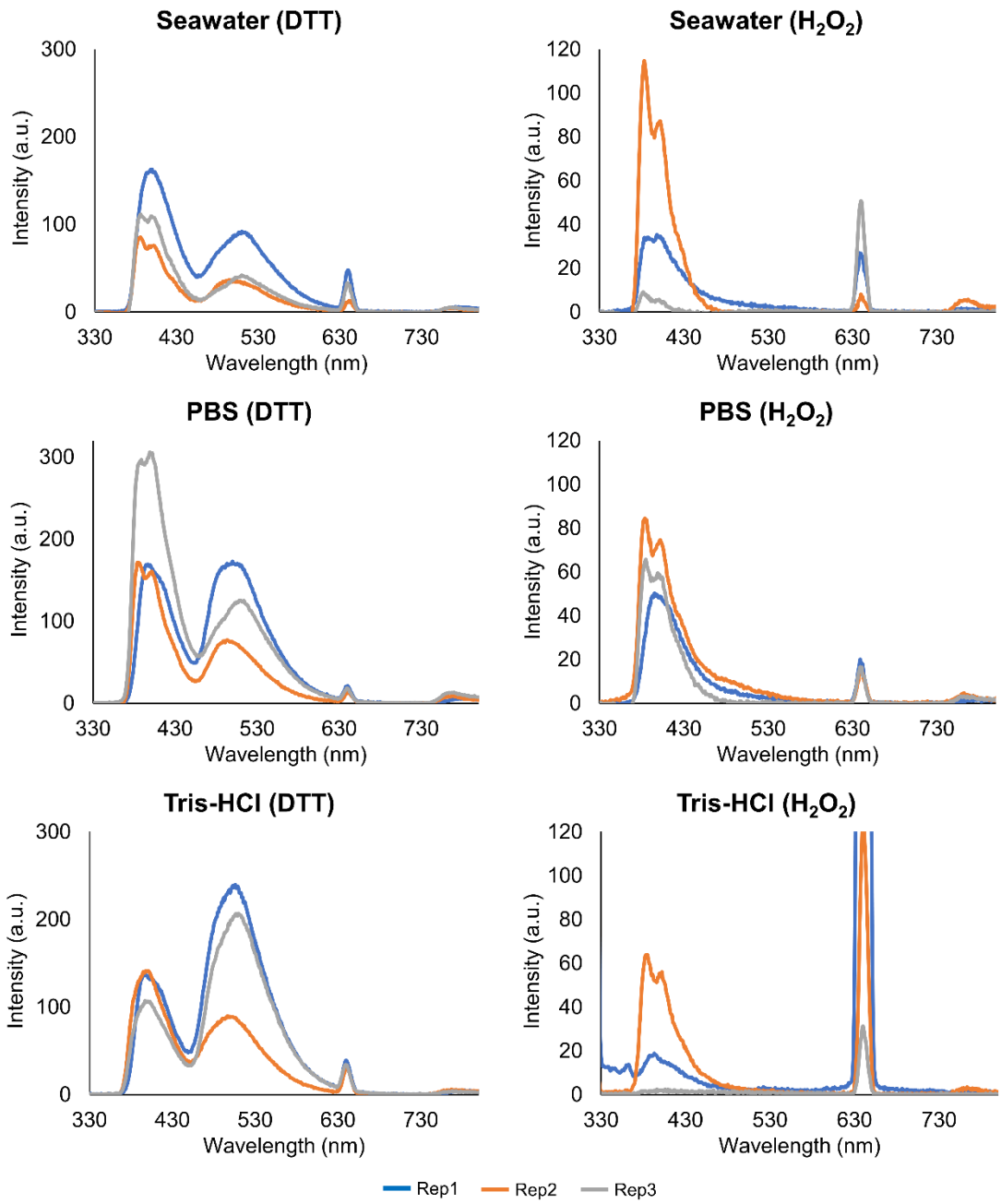


Figure S6.2. Emission spectra of purified mucus samples in different buffers (Seawater, PBS and Tris-HCl, pH 7) and redox treatments (H_2O_2 and DTT). Samples were excited at 320 nm. Rep1-3 indicate independent replicates. Concentration of extracts was normalised to 1 mg total protein. mL^{-1}.

Table S6.3. Averaged peak maximum intensities and respective wavelengths of the emission spectra in purified mucus samples. Samples were excited at 286 and 320 nm for oxidised samples (H_2O_2) and at 260, 286 and 320 nm for reduced samples (DTT). Experiments were done in triplicate with independent extracts.

Buffer	Excitation (nm)	Emission max (nm)	Intensity max (a.u.)	SD (a.u.)
Seawater + H_2O_2	286	382	67.69	50.74
		400	55.53	38.22
	320	382	52.36	45.31
		400	42.15	33.65
PBS + H_2O_2	286	386	60.59	13.23
		399	59.18	6.44
	320	385	61.43	18.97
		400	60.56	10.44
Tris-HCl + H_2O_2	286	383	31.94	32.79
		398	28.08	26.79
	320	384	26.94	26.78
		399	24.50	22.48
Seawater + DTT	260	402	199.93	69.09
		509	54.71	24.31
	286	399	170.84	58.94
		510	44.34	18.19
	320	398	115.92	35.74
		509	55.99	25.76
PBS + DTT	260	399	328.62	128.22
		511	100.54	32.74
	286	399	290.00	108.44
		510	77.64	22.00
	320	398	210.76	67.13
		509	120.46	40.89
Tris-HCl + DTT	260	398	204.66	18.78
		508	174.26	29.23
	286	398	185.33	20.28
		509	133.98	25.82
	320	396	128.31	15.14
		505	177.49	65.43

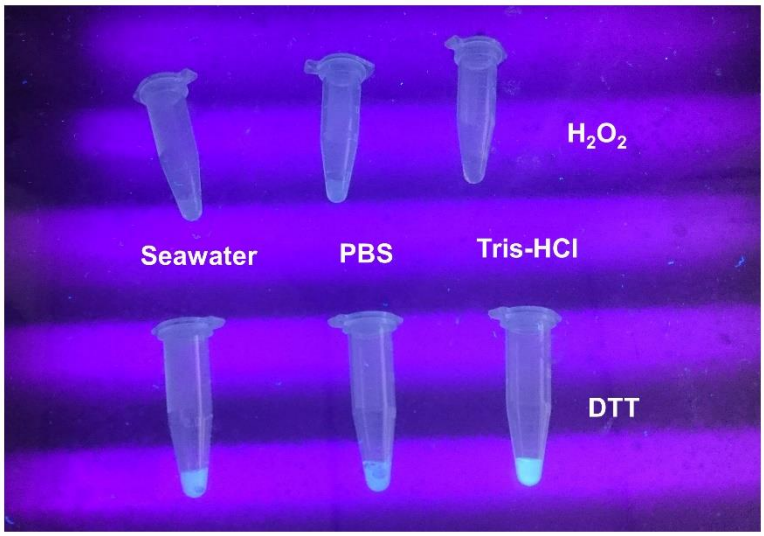


Figure S6.3. Fluorescence of purified mucus samples in different buffers (Seawater, PBS and Tris-HCl pH 7) and treated with redox agents (H_2O_2 and DTT), under UV-light. The blue-greenish fluorescence is notorious in all samples treated with DTT.

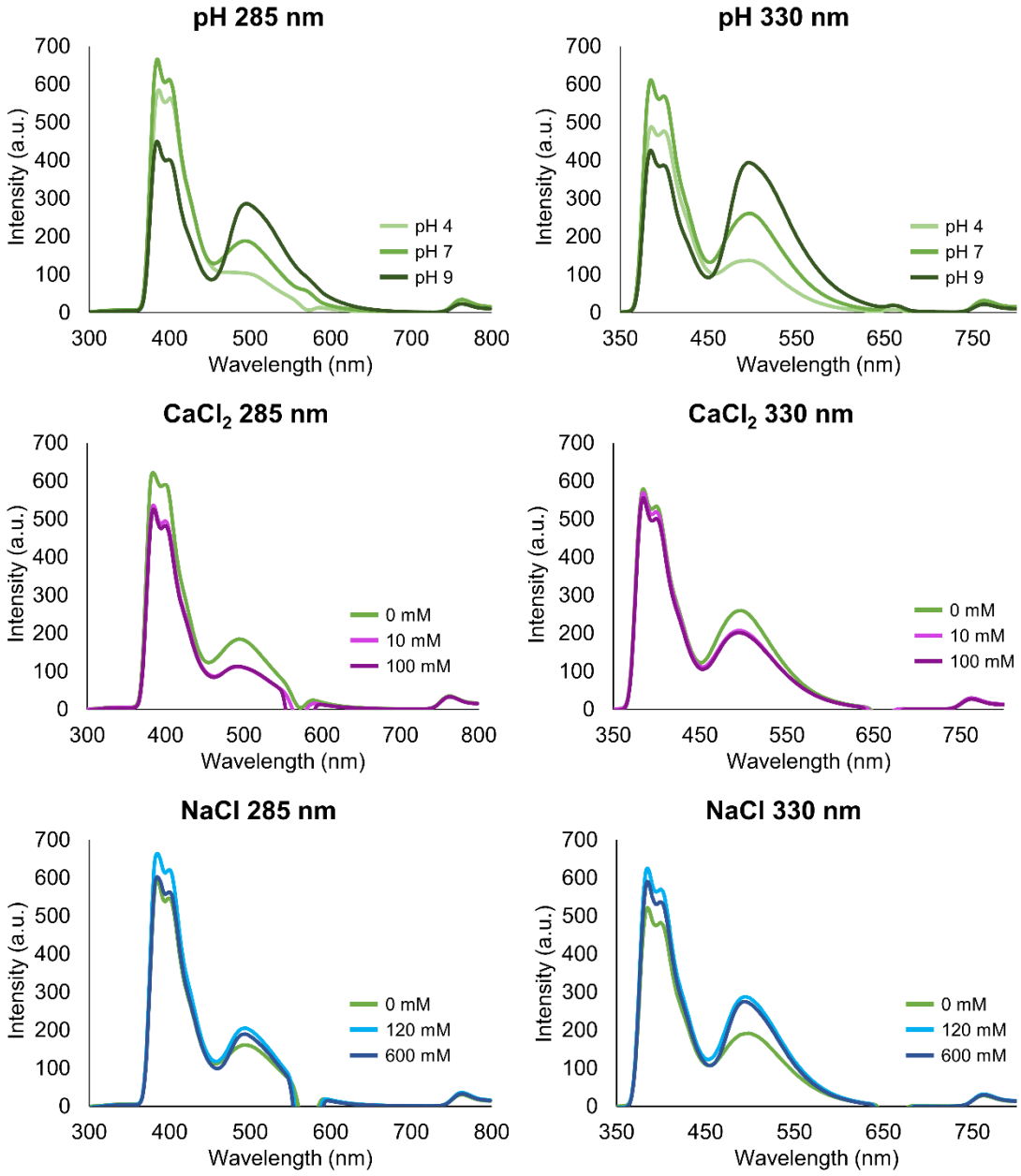


Figure S6.4. Averaged absolute intensity emission spectra of purified mucus samples excited at 285 nm and 330 nm. Samples were kept in Tris-HCl, 20 mM DTT buffer and modulated according to: pH (4, 7 and 9), calcium chloride (0, 10 and 100 mM $CaCl_2$), sodium chloride (0, 120 and 600 mM NaCl). Experiments were done in triplicate with independent extracts. Concentration of extracts was normalised to 1 mg total protein. mL^{-1}.

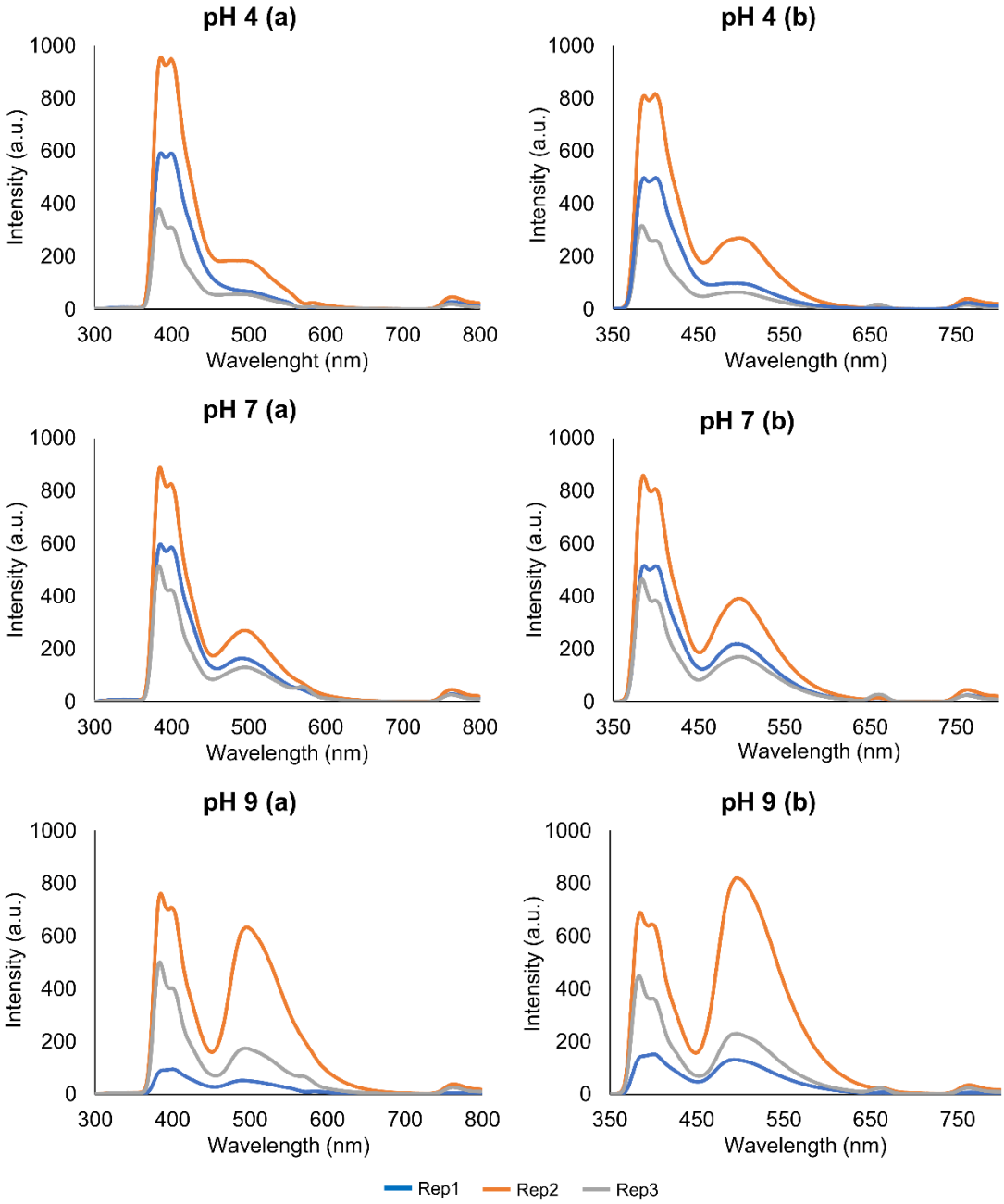


Figure S6.5. Emission spectra of reduced purified mucus samples (Tris-HCl plus DTT buffer), at different pHs (pH 4, 7 and 9). Samples were excited at 285 (a) and 330 nm (b). Rep1-3 indicate independent replicates.

Table S6.4. Averaged maximum intensities and respective wavelengths of purified mucus samples (Tris-HCl plus DTT buffer) in the peaks identified in the emission spectra excited at 285 and 330 nm, at different pH's. Experiments were done in triplicate with independent extracts.

Buffer	Excitation (nm)	Emission max (nm)	Intensity max (a.u.)	SD (a.u.)
Tris-HCl pH 4	285	385	584.32	261.50
		399	563.43	277.27
		491	103.90	57.76
	330	385	487.95	225.89
		399	476.81	243.76
		496	137.17	94.42
Tris-HCl pH 7	285	384	665.07	159.75
		399	563.43	277.27
		493	188.51	59.35
	330	385	611.53	176.25
		399	569.30	176.88
		496	260.59	94.94
Tris-HCl pH 9	285	384	449.54	276.42
		399	563.43	277.27
		495	286.66	250.73
	330	385	425.88	224.15
		399	386.75	201.95
		496	394.28	304.50

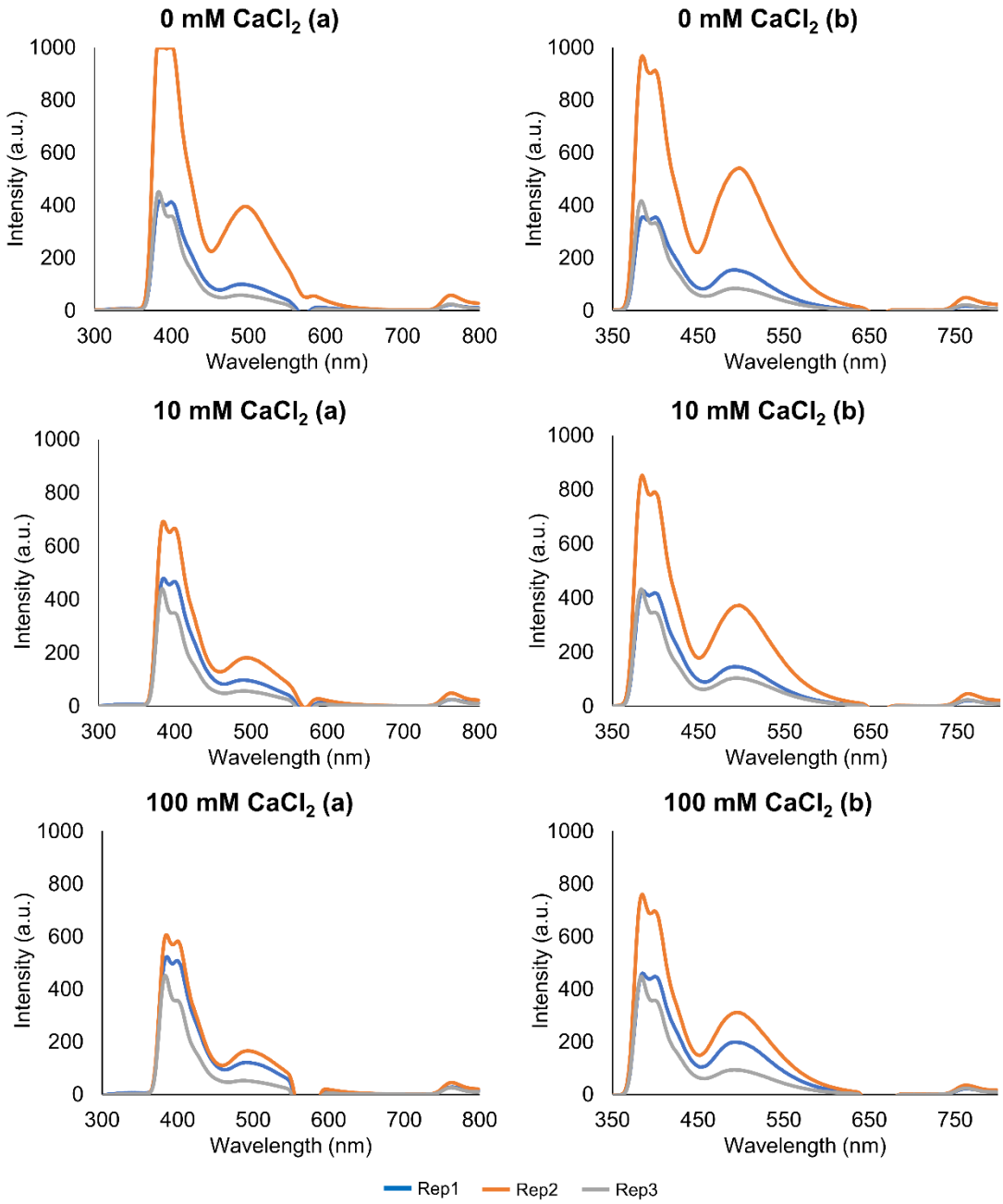


Figure S6.6. Emission spectra of purified mucus samples (Tris-HCl pH 7 plus DTT buffer), with different concentrations of calcium chloride ($CaCl_2$, 0, 10 and 100 mM). Rep1-3 indicate independent replicates. Samples were excited 285 (a) and 330 nm (b).

Table S6.5. Averaged maximum intensities and respective wavelengths of purified mucus samples (Tris-HCl pH 7 buffer) with different $CaCl_2$ concentrations (0, 10 and 100 mM). Samples were excited at 285 and 330 nm. Experiments were done in triplicate with independent extracts.

Buffer	Excitation (nm)	Emission max (nm)	Intensity max (a.u.)	SD (a.u.)
$CaCl_2$ 0 mM	285	384	621.73	267.38
		399	590.24	290.15
		494	184.65	150.16
	330	385	578.66	276.09
		400	533.81	268.05
		497	259.66	201.36
$CaCl_2$ 10 mM	285	384	535.60	110.46
		399	494.40	130.45
		491	111.97	51.41
	330	385	568.89	200.12
		399	519.08	194.72
		497	206.77	118.04
$CaCl_2$ 100 mM	285	385	525.30	64.87
		400	482.79	93.96
		492	113.14	46.99
	330	385	555.29	145.34
		399	501.10	143.44
		494	201.26	89.01

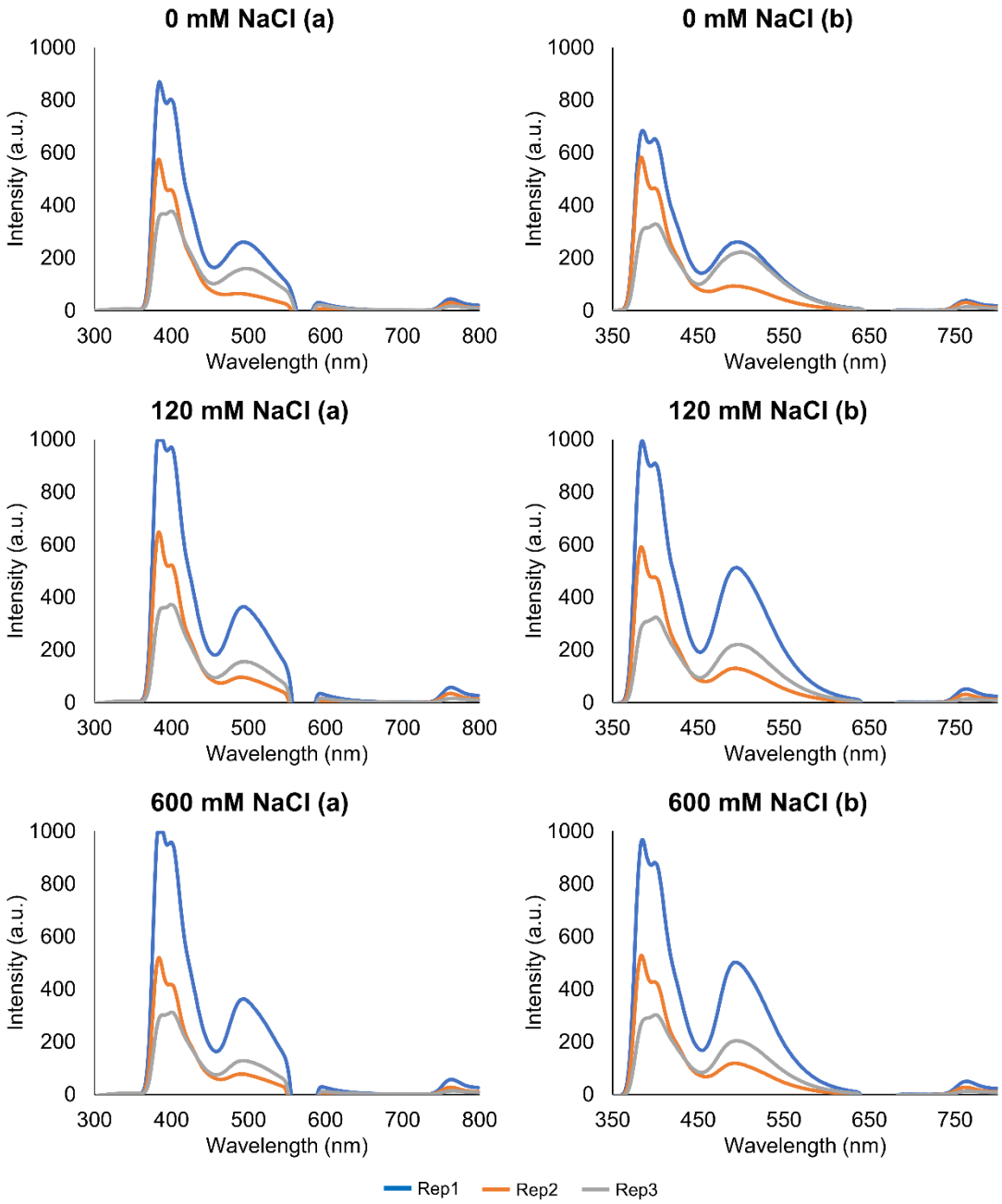


Figure S6.7. Emission spectra of purified mucus samples (Tris-HCl pH 7 plus DTT buffer), with different concentrations of sodium chloride (NaCl, 0, 120 and 600 mM). Rep1-3 indicate independent replicates. Samples were excited at 285 (a) and 330 nm (b).

Table S6.6. Averaged maximum intensities and respective peaks wavelengths for purified mucus samples (Tris-HCl pH 7 plus DTT buffer), with different NaCl concentrations (0, 120 and 600 mM). Samples were excited at 285 and 330 nm. Experiments were done in triplicate with independent extracts.

Buffer	Excitation (nm)	Emission max (nm)	Intensity max (a.u.)	SD (a.u.)
NaCl 0 mM	285	384	600.12	209.76
		399	546.57	184.32
		493	161.31	80.28
	330	385	521.70	158.95
		399	482.19	132.77
		498	191.62	71.40
NaCl 120 mM	285	384	663.70	266.21
		399	621.27	254.07
		493	205.23	115.14
	330	385	624.94	286.40
		400	569.84	247.67
		495	287.85	163.43
NaCl 600 mM	285	384	602.35	294.37
		399	563.08	283.23
		494	189.59	124.79
	330	385	590.16	285.23
		399	536.71	248.75
		495	274.86	164.49

Table S6.7. Averaged emission maximum peaks wavelengths, excited at different wavelengths (285, 330 and 374) for purified mucus samples. Samples were in different buffers treated with DTT (Seawater, PBS and Tris-HCl pH 7), different pHs (pH 4, 7 and 9 in Tris-HCl 20 mM DTT buffer), different $CaCl_2$ concentrations (0, 10 and 100 mM in Tris-HCl 20 mM DTT pH 7) and different NaCl concentrations (0, 120 and 600 mM in Tris-HCl 20 mM DTT pH 7). Experiments were done in triplicate with independent extracts.

Fluorescence emission maxima wavelengths (nm)

		Excitation Wavelength		
		285 nm	**330 nm**	**374 nm**
Buffer (DTT)	Seawater	513	507	509
	PBS	510	504	507
	Tris-HCl	509	505	507
pH	4	492	496	486
	7	493	496	493
	9	495	496	495
$CaCl_2$	0 mM	494	497	495
	10 mM	491	497	494
	100 mM	493	496	490
NaCl	0 mM	493	499	493
	10 mM	493	495	494
	100 mM	494	495	494

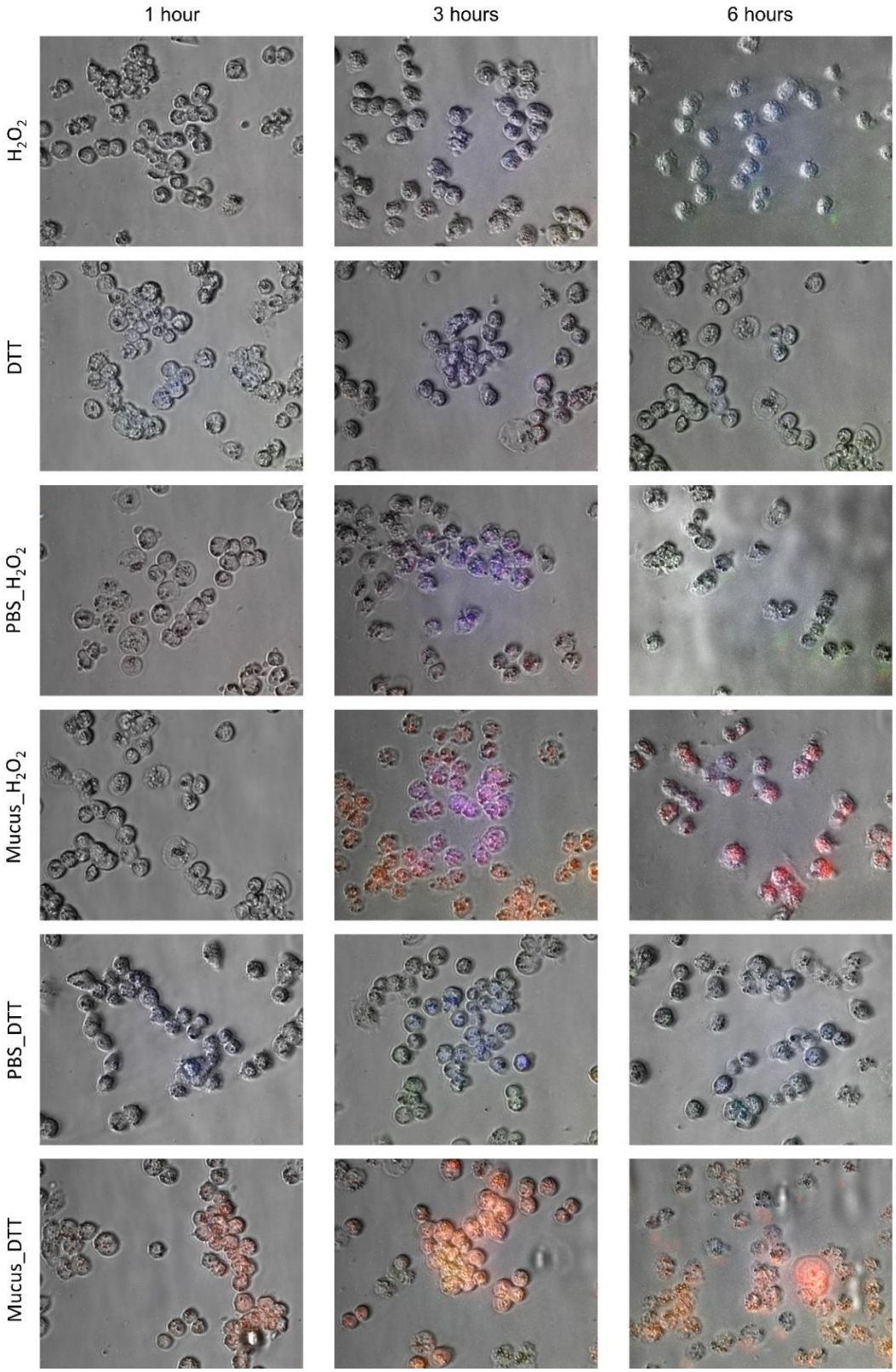


Figure S6.8. Internalisation assay for purified mucosubstances onto the ovarian cancer cell line A2780. Cells were incubated during 1 h, 3 h and 6 h with purified mucus extract prepared with PBS (pH 7.4) and the respective control (PBS). Either experimental condition was subjected to reduced or oxidizing agents (DTT and H_2O_2). Composite images were produced by overlapping red, green and blue channels with brightfield images.

CHAPTER 7 – Final conclusions

7.1. Final considerations

The high biotechnological potential of marine bioactives is long acknowledged for a wide range of applications, from eco-friendly antifoulants and pesticides to food and biomass plus novel therapeutical agents. These novel resources can have a major impact in sustainable "blue" bioeconomy" by contributing to shift exploitation of the seas towards more sustainable practices than fisheries or intensive aquaculture such as recombinant synthesis of marine bioproducts and inshore culturing. Such success is largely dependent on comprehensive characterisation of target bioproducts, from gene to mode-of-action and safety. Marine bioprospecting is, nonetheless, a more challenging enterprise than its terrestrial counterpart due to factors such as habitat accessibility, persistent knowledge gaps in ecophysiology, lack of genomic resources and immense biodiversity. Even though the present study highlighted the promises of marine invertebrates, by demonstrating that *Eulalia* secretes a complex cocktail of bioactives, especially toxins, it also demonstrated the complexity and difficulties of this field of research.

In toxin bearing organisms, the description of the toxin-delivery apparatus is notoriously relevant to understand the evolution and the ecological and adaptive value of toxins; the natural efficacy of venomous mixtures and devise strategies for substance harvesting and isolation. In *Eulalia*, these questions could be answered addressing the role of the proboscis and the dynamics of the interaction between the worms and one of their prey of choice: mussels. Understanding the relationship between those two species yielded fundamental hints on the existence of paralysing substances and thus the importance of toxins in predation mechanisms. Moreover, the integration of ecological assays with omics techniques enabled the discovery of a cocktail of toxins and the classification of *Eulalia* as a toxungenous organism. This allowed the Phyllodocidae family to join the short list of the so far known toxin-bearing polychaetes. Also of importance, the recruitment of individual toxins at different stages of the Polychaeta history-of-life indicates divergent evolution from non-toxin-secreting ancestors. Such evolution resulted in a relatively reduced level of toxin specialisation that confers *Eulalia* and other Polychaeta plasticity against a wide variety of prey. Altogether, the findings demonstrate the importance of addressing eumetazoan phylogeny and processes such as gene duplication and supra-genomic regulation of gene expression to fully understand the evolution of venomous secretions.

Even though this study was mostly dedicated to the identifications of novel bioactives with biotechnological potential from *Eulalia*, it revealed the importance of understanding the ecophysiology of marine invertebrates. Indeed, ecological traits such as feeding behaviour and habitat can offer vital clue to find species of interest. It is the case of *Eulalia* stunning live prey using its toxic mucosecretions or its use of UV-absorbing mucus as protection during foraging

through the intertidal. As such, marine biotechnology and marine bioprospecting should not focus only on the pure search for bioproducts, but also on the importance of a species' ecology and of the role its adaptive to features in the natural environment.

Unravelling the toxungenous cocktail secreted by *Eulalia* provided important clues for the development of "venomics" strategies in marine invertebrates, Polychaeta in particular, even in face of reduced genomic annotation for these organisms, as for marine fauna in general. With this respect, transcriptomics allowed the screening for a vast span of proteins and peptides that, by quantifying the differential expression between skin and proboscis, permitted the identification of toxin candidates. This success, however, could not have been achieved without the comprehensive morphoanatomical appraisal that often lacks in marine invertebrates. The combination with proteomics validated the findings and further shortlisting of proteins of interest without relying on fortuitous outcomes of targeted approaches.

Even though the proteins, peptides and other substances comprising *Eulalia*'s toxungen are naturally directed towards molluscs and other marine invertebrates, they caused very significant cytotoxic and cytostatic effects to a human cancer cell line, more than e.g., normal cells. This effect, which seems specifically linked to interference with cell cycle and programme cell death, breaks ground to potential applications. Even though the mechanisms by which the cocktail of proteins' mode-of-action, the findings support further research of venom proteins or their mixtures as therapeutical agents against highly proliferating cells.

Eulalia viridis secretions seem also to be equipped with fluorescent proteins that strongly absorb UV radiation and whose emission can be easily switched on or off by reduction or oxidation, respectively. The novel fluorochromes, albeit further research being needed to clarify its chemical nature, was found to be compatible with physiological media and being able to endure cellular uptake. This unexpected outcome point towards yet another application of bioproducts from *Eulalia*, for instance as molecular probes.

Altogether, this work exemplifies the potential that a single marine species can offer in bioreactives with high-value biotechnological applications, from anti-cancer therapeutics to molecular probes. Even though research must still be expanded with respect to isolation and characterisation of individual toxins, fundamental methodological pipelines to address knowledge marine toxinology under a biotechnological perspective have been successfully implemented. Altogether, the outcomes of the research plan can contribute to raise awareness for Blue Biotechnology and the sustainable exploitation of novel marine resources, especially in Portugal, where the vast Exclusive

Economic Zone (EEZ), together with high biodiversity and productivity at the Atlantic-Mediterranean confluence envisage high prospects in the field.

7.2. Future perspectives

The polychaete *Eulalia viridis* yielded multiple bioproducts of interest, which means that we are likely just facing the tip of the iceberg with respect to natural bioreactives from marine annelids. The isolation and characterisation of the exact chemical structure of key molecules in this worm secretions are naturally the next steps to be undertaken (such as the exact nature of the fluorescent proteinaceous complexes). This is a needed step to continue testing for effects, safety and mechanisms (*in vivo*, in particular) and potentially even synthetising proteins and peptides *in vitro* using, e.g., recombinant methods, which is particularly relevant to translate knowledge to industry. It must also be noted, future investigations of *Eulalia* transcriptome and proteome are going to, most likely, unravel new proteins specific for this species, many of which can hold biotechnological potential, toxins or not, or even bioactive secondary metabolites, a subject that has yet to the thoroughly explored in the species of annelids in general despite the growing interest in these molecules in drug developments (recall, for instance, the example of Yondelis). Even though the presence of holobionts in the species could not, at this stage, be verified, since even little evidence for symbiotic bacteria was found (results not shown), the search for microorganisms interactions can also be a promising line of research in this and other species, as several symbionts were found in other marine species with great biotechnological potential (see section 1.2.6 for examples). Also, some basic ecological features are yet needed to be explored, especially those related to reproduction and life cycle. Indeed, aspects such as life cycle and aquaculture may be needed if this worm becomes a constant in bioproduct discovery. Finally, is must also be mentioned that the correct species classification is lagging, in line with the current constraints in the systematics of Polychaeta, a group that is acknowledged to by polyphyletic and therefore of scant taxonomic value.

As the catalogue of potential bioproducts from marine invertebrates extends, it becomes clearer that marine biodiversity is an asset that must be nurtured and protected. Also, in comparison to traditional marine economic activities, bioprospecting-based biotechnology has far less deleterious consequences for the environment, since rearing invertebrates and novel cloning methods do not imply invasive procedures onto ecosystems. Consequently, bioprospecting the seas is adequately positioned to be the single most important showcase of just how important safeguarding biodiversity is and how it can shift the exploitation of living marine resources toward sustainability. With this respect, Polychaeta can offer a much wider diversity of novel bioproducts than could, perhaps, be anticipated. Nonetheless, it is clear that we are just beginning to unravel their potential.

Biodiversity can become a challenge as much as it is an asset, particularly in the case of such a diverse group of organisms with extremely reduced genomic annotation. Still, the current state-of-the art shows the advantages of high-throughput “omics”, for the finding and characterisation of bioactive proteinaceous substances via homology search, with emphasis on proteomics and RNAseq or their combination. On the other hand, the molecular mechanisms of interaction between the compounds and potential target receptors is scarcely studied in these animals, albeit being one of the most critical aspects in drug discovery. Once again, novel molecular tools and state-of-the art bioinformatics can provide invaluable data on issues such as receptor orthosteric or allosteric docking sites and even predict interactions. However, this information must still be complemented with biochemistry, toxicology and even with the ecology and physiology of the Polychaeta in order to map taxa of interest and identify species that can offer purposeful bioproducts. Species whose physiology and ecology are still largely unknown, such as deep-sea vent worms and other extremophiles, are revealing the most extraordinary adaptations to some of the harshest habitats on Earth, many of which are likely linked to endogenous or symbiosis-derived bioproducts. These results should steer marine biotechnologists to look beyond coastal habitats. Altogether, multidisciplinarity is a key concept to produce the comprehensive set of information that is needed to hinge between the extraordinary biodiversity of marine Polychaeta and the immediate needs of industry.