NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

SÓNIA CRISTINA DA SILVA CAMPOS

BSc in Molecular and Cellular Biology

THE BIOTECHNOLOGICAL POTENTIAL OF MARINE BIO-REACTIVES: ISOLATION AND CHARACTERIZATION OF NOVEL TOXINS FROM POLYCHAETA

MASTER IN BIOTECHNOLOGY

NOVA University Lisbon

February, 2022

NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF CHEMISTRY

SÓNIA CRISTINA DA SILVA CAMPOS

BSc in Molecular and Cellular Biology

THE BIOTECHNOLOGICAL POTENTIAL OF MARINE BIO-REACTIVES: ISOLATION AND CHARACTERIZATION OF NOVEL TOXINS FROM POLYCHAETA

MASTER IN BIOTECHNOLOGY

NOVA University Lisbon

February, 2022

DEPARTAMENT OF
CHEMISTRY

THE BIOTECHNOLOGICAL POTENTIAL OF MARINE BIO-REACTIVES: ISOLATION AND CHARACTERIZATION OF NOVEL TOXINS FROM POLYCHAETA

SÓNIA CRISTINA DA SILVA CAMPOS

BSc in Molecular and Cellular Biology

Adviser: Pedro Manuel Brôa Costa
Auxiliary Professor, NOVA University Lisbon

Co-adviser: Mariaelena D'Ambrosio
Postdoctoral Researcher, NOVA University Lisbon

Examination Committee:

Chair: Ana Cecília Afonso Roque
Associate Professor, NOVA University Lisbon

Rapporteur: Tiago Fernandes Grilo
Postdoctoral Researcher, University of Lisbon

Adviser: Pedro Manuel Brôa Costa
Auxiliary Professor, NOVA University Lisbon

MASTER IN BIOTECHNOLOGY

NOVA University Lisbon
February, 2022

The Biotechnological Potential of Marine Bio-reactives: Isolation and Characterization of Novel Toxins from Polychaeta

ACKNOWLEDGMENTS

I would like to start by thanking my advisers, Pedro Costa and Mariaelena D'Ambrosio, for the unconditional patience they had with me for the past year. Whether it was to answer my infinite questions, or to support my work, they always gave it their all to make me feel welcomed and encouraged. The quality of this work would not be possible without them, and I am very grateful.

To the rest of the members of the SeaTox Lab, Ana, Cátia, Carla, and Carolina, I would like to show my appreciation for all the help they gave me. They helped me with lab benchwork and welcomed me to the team immediately. I would also like to thank Bruno Manadas and Vera Mendes (CNC, U. Coimbra) for the LC-MS/MS analysis.

To Marta, Telma, and Inês, the ones that walked this path alongside me, I wish the best of luck for their future endeavors. Marta laughed and cried with me, making this path not only manageable but enjoyable, as the days were more fun when we were together. Telma heard my endless rambles and was always a good friend to me. Inês helped me with all the computational hardships without missing a beat, having immense patience regarding my lack of coding abilities.

I would like to thank my family as they were always supportive of my work. Specially my mom, the main recipient of my frustrations, that not only helped me deal with it but also showed me unconditional love and support, keeping me motivated and happier as hardships fell through.

To Tiago that walked this path with me, celebrating with me on the good days and cheering me up on the not so good ones. He was there for me and showered me with love and attention, making me fell cared for.

To my friends, who's presence always makes me happier.

Thank you all.

“There are two types of people in this world, those who can extrapolate from incomplete data” (No author)

ABSTRACT

As the largest and most ancient habitat on Earth, the marine environment has an immense biodiversity that can become an almost endless span of novel bioproducts. Still, only a few have already been synthesized *in vitro* and commercialised as approved drugs. Venom-derived drugs have shown to be a viable alternative to the expensive and time-consuming process that is the design of synthetic pharmaceuticals. Due to long-term coevolution between predator and prey, proteins with strong activity towards specific targets of recipient organisms were recruited into venoms and poisons. Some Polychaeta, a large class of biodiverse and ubiquitous invertebrates, are known to secrete toxins for predation and defence. The present study focused on *Glycera alba*, a common Glyceridae (Annelida: Polychaeta) from the Portuguese estuarine intertidal, whose congenerics, *G. tridactyla* and *G. dibranchiata*, are known to be venomous. As the first attempt to screen for potential proteinaceous toxins, we extracted and fractionated total protein from the *G. alba*'s proboscis and skin. The structures were compared histologically confirming the presence of specialised venom glands in the former, which are bound to reservoir-like structures and ducts connecting to hollow jaws. The SDS-PAGE analysis of protein fractions, obtained by ultrafiltration, disclosed intricate proteomes with different bands between the two organs that can be potentially associated to toxins and permeabilising agents, among other bio-reactives. Regarding the proboscis and the skin specific proteomes, LC-MS/MS-based proteomics and Gene Ontology enrichment analysis identified relevant differences related to different molecular and physiological pathways. According to toxicity assays, the higher molecular mass fraction's toxins/enzymes can exert negative effects onto their prey, with emphasis on cytotoxicity. Altogether, the encouraging findings show that venomous annelids can be important sources of novel bio-reactives, albeit illustrating the challenges of surveying organisms whose genomes and metabolism are poorly understood.

Keywords: Marine Biotechnology, Bioresources, Proteomics, Venom, Annelida, *Glycera alba*.

RESUMO

Como maior e mais antigo habitat terrestre, o ambiente marinho possui uma imensa biodiversidade com potencial de se tornar numa fonte quase infinita de novos bioprodutos. Destes, apenas alguns foram sintetizados *in vitro* e comercializados como medicamentos aprovados. Drogas derivadas de veneno têm revelado ser uma alternativa viável, ao processo caro e demorado, que é o design de produtos farmacêuticos sintéticos. Devido à longa coevolução entre predador e presa, proteínas com forte atividade contra alvos específicos de organismos recetores, foram recrutadas como venenos. Algumas Poliquetas, uma grande classe de invertebrados biodiversos e ubíquos, são conhecidas por secretar toxinas para predação e defesa. O presente estudo incidiu sobre *Glycera alba*, um Glyceridae (Annelida: Polychatea) comum do estuário intertidal português, cujos congenéricos *G. tridactyla* e *G. dibranchiata* são conhecidos produtores de veneno. Como primeira tentativa de rastrear potenciais toxinas proteicas, extraímos e fracionámos extratos de proteína da probóscide e da pele de *G. alba*. As estruturas foram comparadas histologicamente, confirmando a presença de glândulas de veneno especializadas nas primeiras, que se ligam a estruturas semelhantes a reservatórios e dutos e a mandíbulas ocas. A análise do SDS-PAGE das frações proteicas, revelaram proteomas complexos com diferentes bandas entre os dois órgãos que podem ser associadas a toxinas e agentes permeabilizantes, entre outros biorreativos. Relativamente ao proteoma específico da probóscide e pele, a proteómica baseada em LC-MS/MS e a análise de enriquecimento de ontologia genética identificaram diferenças relevantes em relação a diferentes vias moleculares e fisiológicas. De acordo com os ensaios de toxicidade, as toxinas/enzimas da fração de alta massa molecular exercem efeitos negativos nas suas presas, especialmente citotóxicos. Em suma, estes resultados encorajadores, revelam que anelídeos venenosos podem ser fontes importantes de biorreativos, contribuindo para ilustrar os desafios da investigação em organismos cujos genomas e metabolismo são mal compreendidos.

Palavras chave: Biotecnologia Marinha, Biorrecursos, Proteómica, Veneno, Anelídeo, *Glycera alba*.

CONTENTS

1 INTRODUCTION 1

2 MATERIALS AND METHODS 5

2.1 Animal collection 5

2.2 Morphoanatomical analyses of venom system 5

2.3 Protein extraction and characterisation 6

2.3.1 Protein Separation 6

2.3.2 Proteomics 6

2.3.3 Proteome fractioning 7

2.4 Toxicity testing 8

2.4.1 Histopathological analysis 9

2.4.2 Comet assay 9

2.4.3 Caspase 3 activity 10

2.5 Statistical analysis 11

3 RESULTS 13

3.1 Morphoanatomical analysis of the venom system 13

3.2 Comparative proteomics between proboscis and skin 15

3.2.1 Gel electrophoresis of crude and fractionated protein extracts 15

3.2.2 Proteomics 17

3.3 Toxicity testing 21

3.3.1 Histopathological analysis 21

3.3.2 Comet assay 23
3.3.3 Caspase 3 activity 23
4 DISCUSSION 25
5 CONCLUSION 31
A APPENDIX 37

LIST OF FIGURES

Figure 2.1 – Exemplification scheme of in series process of ultrafiltration 8
Figure 3.1 – Microanatomy of *Glycera alba*'s proboscis 14
Figure 3.2 – Body wall section of *Glycera alba* 15
Figure 3.3 – Protein signature of crude extracts from *Glycera alba* proboscis (P) and skin (S) 16
Figure 3.4 – Protein signature of four fractionated extracts from *Glycera alba* proboscis (P) and skin (S) 17
Figure 3.5 – Distribution of the identified proteins on *Glycera alba* proboscis and skin 18
Figure 3.6 – Comparison between Biological Process Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin 19
Figure 3.7 – Comparison between Molecular Function Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin 20
Figure 3.8 – Comparison between Cellular Component Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin 21
Figure 3.9 – Representative histological gill sections of *Mytilus* sp. exposed to *G. alba* proboscis and skin extracts 22
Figure 3.10 – Results from Comet assay to identify DNA damage 23
Figure 3.11 – Caspase 3 Activity in mussel gills exposed to *G. alba* proboscis and skin extracts 24

LIST OF TABLES

Table 2.1 – Reaction scheme for 96 well plate microassay method..... 11
Table 3.1 – Peptidic sequences' protein match 18

ACRONYMS

DMSO	Dimethyl sulfoxide.
DNA	Deoxyribonucleic acid.
DPX	Dibutylphthalate polystyrene xylene.
DTT	Dithiothreitol.
EDTA	Ethylenediamine tetraacetic acid.
GLTx	Glycerotoxin.
GO	Gene Ontology.
H&E	Hematoxylin and Eosin.
HMPA	High melting point agarose.
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry.
LMPA	Low melting point agarose.
PBS	Dulbecco's phosphate-buffered saline.
pNA	p-Nitroaniline.
SDS	Sodium dodecyl sulphate.
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis.
TAE	Tris-acetate-EDTA.

1

INTRODUCTION

Marine biotechnology can be defined as the combination of marine bioresources with industrial, biological, and biochemical methods to produce added-value and enable production scale-up of consumer products (see for instance Burgess, 2012). Being the largest and most ancient habitat on Earth, the marine environment benefits from an immense biodiversity that correlates with an almost endless span of novel bioproducts that result from selective pressure and adaptation to its diversified ecosystems. The oceans are thus an auspicious target for the bioprospecting of new products with potential biotechnological applications in several domains, such as agriculture, cosmetics, and pharmaceuticals (e.g., Burgess, 2012; Barbosa and Roque, 2019). Indeed, there are already a few chemically diverse marine bioproducts already marketed for various applications: from neurotoxin derived eco-friendly pesticides (e.g., Padan, as seen in Coutinho, et al., 2017) to peptides with anti-inflammatory properties, such as pseudopterosins, a group of marine diterpene glycosides isolated from the Caribbean soft coral *Pseudopterogorgia elisabethae*, used in moisturisers (Look, et al., 1986; Caplan, et al., 2016). A paradigmatic example is Ziconotide, an approved non-opioid analgesic developed from a conotoxin, a small proteinaceous neurotoxin, of the gastropod mollusc *Conus magus*, which is considered as strong as morphine (Molinski, et al., 2009; Schmidtko, et al., 2010). Additionally, Trabectedin, a synthetically-produced antineoplastic agent derived from the marine tunicate *Ecteinascidia turbinata*, was approved, in Europe, to treat patients with advanced soft tissue cancer (sarcoma) and ovarian cancer (Carter and Keam, 2007; Christinat and Leyvraz, 2009).

Despite all the progress made in discovering novel bioproducts, the oceans still remain vastly unexplored, part due to the lack of technology that is needed to access remote habitats (such as deep-sea environments) and part due to high biodiversity and insufficient genomic resources, which renders difficult the identification, extraction, and purification of compounds. The later aspect also compromises deploying DNA large-scale synthesis of recombinant technology for scaling-up synthesis of marine bioactives of peptidic nature. Nevertheless, advances in aquaculture, sampling methods, coupled with the

effort to characterise genomes and proteomes can boost marine bioprospecting, even if the oceans' immense biodiversity may itself hinder translation to industry (Molinski, et al., 2009).

For bioprospecting purposes, marine invertebrates are especially interesting. However, despite the large number of natural products added to libraries, research has only focused on few marine species (Rodrigo and Costa, 2019). Under a biotechnological point-of-view, the relevance of these animals is incited by vast molecule diversity driven by unique adaptive traits, which in turn results from adaptation to a broad range of habitats, from beaches to abyssal plains (see Leal, et al., 2012; Coutinho, et al., 2017). As an example, anti-microbial peptides from invertebrates might be used as defence against pathogens, like *Arenicola marina*, a tubicolous marine annelid that produces two peptides, arenicin-1 and arenicin-2, with activity against both Gram-positive and negative bacteria, as well as fungi (Iwanaga and Lee, 2005; Coutinho, et al., 2017). Another example is mucus secretion, a strategy used by polychaetes to minimize water loss, protect egg masses, defend against infections, and reduce friction during locomotion (Coutinho, et al., 2017; Stabili, et al., 2019).

Marine invertebrates, such as annelids, molluscs, and cnidarians, produce venomous secretions that, as higher-order animals, are constituted by a complex mixture of bioactives, from toxins to enzymes, salt, and peptides, synthesised in specialised gland cells, which is later conveyed through a wound or body surface to a target (Casewell, et al., 2013). These substances can enable or at least facilitate defence and predation, or resolve competition, by disrupting normal physiological or biochemical processes on the target animal (Fry, et al., 2009). Due to mutual interactions and evolutionary conservation, venom's proteinaceous toxins are tightly associated with humans. This long-term selective pressure, potentially leading to coevolution between prey and predator, resulted in diversified proteins with strong activity to maximise the effect of noxious substances. Thus, toxins have high specificity and huge molecular diversity towards key factors of the target organisms, such as cell receptors and membrane ion-channels (Zhang, 2015).

Natural selection, especially in the seas, may have resulted in a wide range of secreted bioactive proteins and peptides that may become an alternative to the expensive and time-consuming process that is the design of synthetic compounds for the purpose of drug discovery (Rodrigo and Costa, 2019). For example, cone snails (*Conus*) produce toxic secretions rich in peptides and small proteins, the most notorious of which being conotoxins and conopeptides that induce neuromuscular paralysis in their prey, which steered the interest for their potential in the management of severe and chronic pain (Schmidtko, et al., 2010). As another example, antimitotic peptide toxins have also been described in marine sponges. It is the case of hemiasterlin, a tripeptide isolated from the marine sponge *Hemiasterella minor* with high cytotoxic properties. This toxin blocks cell division during the M phase by targeting tubulin, therefore hindering the formation of the spindle, and has advance to Phase II clinical trials as an antitumour natural

product (Molinski, et al., 2009). Some nemertines (ribbon worms), which mostly live under rocks or burrows in the shallow waters, secrete toxins capable of paralyze their prey and deter predators by acting as agonist of nicotinic receptors (acetylcholine receptors) of central nervous system and cardiac ganglion. These compounds have been investigated to improve cognitive impairment in patients with Alzheimer, since these patients have shown to have a deficiency in nicotinic synaptic transmission (Kem, 1997; Göransson, et al., 2019).

The Polychaeta form a large class of aquatic annelid worms (mostly marine), with around 14000 registered species, and encompasses the gross of the Annelida diversity (Coutinho, et al., 2017). They are found in many marine habitats, from intertidal areas to deep-sea vents; therefore, can be very morphological distinct due to differences in lifestyle (Flint and Rabalais, 1980). Bioproducts retrieved from marine Polychaeta are showing promising biotechnological potential, albeit lacking effective application: from luminescent and fluorescent probes like the photoprotein-based bioluminescence of *Odontosyllis phosphorea* (Deheyn and Latz, 2009) to potential painkillers and anaesthetics, a possibility created by proteinaceous neurotoxins first described from *Glycera spp.* (von Reumont, et al., 2014).

Glycerids (Phyllodocida: Glyceridae), also known as bloodworms, are carnivorous Polychaeta that produce toxic secretions, the currently researched species are *G. tridactyla* and *G. dibranchiata* (Gibbs and Bryan, 1980; von Reumont, et al., 2014). Bloodworms typically feed on polychaetes and small crustaceans injecting a venom mixture through an eversible proboscis with four, symmetrically placed, chitinous jaws with a venom canal (Gibbs and Bryan, 1980). Studies revealed that *Glycera* venom can cause cardiac arrest, progressive paralysis, convulsions, and death to their prey; however, deleterious effects in humans are seemingly limited to a local inflammatory response (Gibbs and Bryan, 1980; von Reumont, et al., 2014). *Glycera*'s toxins, identified in *G. tridactyla* and *G. dibranchiata*, can be placed in five categories: pore-forming toxins, neurotoxins, protease inhibitors, CAP domain toxins, and other enzymes; from which a few revealed potential in bioprospecting (von Reumont, et al., 2014).

According to preceding studies, *Glycera tridactyla*'s venom can reversibly increase spontaneous neurotransmitter release due to a high-molecular weight neurotoxic glycoprotein, the glycerotoxin (GLTx) (Manaranche, et al., 1980). Having a unique mode of action, GLTx up-regulates calcium channels, in the presynaptic plasma membrane, inducing Ca^{2+} influx at resting potential. With a total weight of 300 – 320 kDa and complete polypeptide chains of around 150 kDa, GLTx functions as a dimer; however, only the large molecular weight form is active (Richter, et al., 2017). Additionally, bloodworm's venom glands also express homologs of a toxin only known from scorpaeniform fish and monotremes (SNTX-like toxin), sea anemones (gigantoxin-like neurotoxin), as well as toxins with a wide phylogenetic distribution (CAP domain, Actinoporins, Chitinases, Peptidases, etc.), (von Reumont, et al., 2014). Therefore, it is interesting to profile more species of bloodworms, such as *Glycera alba*, a common species

from the Portuguese intertidal sandy-muddy beaches that, despite having almost no published research available, was a case study with interesting results regarding its microanatomy, where specialised exocrine glands were observed (Gonçalves and Costa, 2020).

We hypothesise that *G. alba*, as other congeneric species, secretes venom, with which the worm predates on small invertebrates. As such, the current study aimed essentially to: i) characterise the venom delivery system of *Glycera alba*; ii) provide the first molecular characterisation of proteinaceous venom components of the species, and iii) test the toxicity or bioreactivity of proteinaceous toxins.

2

MATERIALS AND METHODS

2.1 Animal collection

A total of 86 *Glycera alba* (ranging from 30 – 80 mm total length and 100 – 300 mg weight) were hand-collected during low tide, from the sandy-muddy intertidal shore of Seixal Bay, Portugal (38º38'40.7"N 9º06'07.8"W) in September 2020, April 2021, and June 2021. Specimens were transported to the laboratory and kept in a mesocosm environment, an aquarium with approximately 7 cm of sand depth and 7 L of artificial saltwater in a closed circulation system with controlled temperature and photoperiod, with continuous aeration and water recirculation. Salinity, temperature, and photoperiod were restricted to 30, 17 ± 1 ºC, and 16:8 h light:dark cycle, respectively.

Mussels (*Mytilus edulis*) were hand-collected during low tide, by the end of June 2021, from a rocky intertidal area in Costa da Caparica, Western Portugal (38º38'28"N 9º14'18.9"W). The animals ranged between 35 and 45 mm total length and weighed around 10 g each. Mussels were acclimated in separate aquaria to lab conditions in the closed-circulation system described above.

2.2 Morphoanatomical analyses of venom system

To identify specialised glandular tissue, a morphoanatomical analysis based on histological methods was performed. Three whole *Glycera alba* were fixed in glutaraldehyde (2.5 % v/v glutaraldehyde, 0.1 M, pH 7.4) for 2 × 1 hour. After this, the samples were washed with 0.1 M sodium cacodylate buffer for 3 × 15 min and then microdissected to separate the proboscis from the prostomium. Post-fixation was done in the dark, overnight, with 1 % (m/v) osmium tetroxide in cacodylate buffer, 0.2 M, pH 7.4. The samples were washed, 3 × 15 min with Milli-Q water, and dehydrated through a progressive series of acetone (30 %, 50 %, 70 %, and 90 %, for 15 min each, and 100 %, for 30 min). Intermediate infiltration was done with EPON resin (Sigma-Aldrich) and propylene oxide (1:2, 1:1, and 2:1, for 30 min

each) followed by a final infiltration of EPON resin, in vacuum, for 30 minutes. Polymerization was performed at 65 ºC, overnight.

All samples were sectioned at a 2 µm thickness using a RM 2125 RTS rotary microtome (Leica Microsystems) equipped with a tungsten carbide blade. Sections were stained using Toluidine blue and the slides were mounted with DPX resin. Observations were performed using a DM 2500 model microscope (Leica Microsystem).

2.3 Protein extraction and characterisation

To collect proboscis and skin, 83 *Glycera alba* were immobilised by hypotonic shock and then microdissected, ensuring that all organs underneath stayed unscathed. Skin samples include epidermis and the underlying musculature. The target organs were homogenized separately with a pestle in cold buffer (200 mM DTT, 0.05 M Tris-HCl, pH 7.0) complemented with 1 % v/v Protease Inhibitor cocktail (Sigma-Aldrich), for protein extraction. Samples were centrifuged for 5 min, 12000 rpm, 4 ºC. The supernatants were pooled, with each pool containing extracts from about seven animals and stored at -80 ºC. Protein concentration was determined using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific).

2.3.1 Protein Separation

The discontinuous gel system (Laemmli, 1970) was used for one-dimensional separation of proteins in crude extracts under denaturing conditions by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE), as detailed by Hames (1998). Resolving and stacking gels contained 12 % (v/v) and 6 % (v/v) acrylamide, respectively. All wells were loaded with 25 µg of protein, dilutions were made using Milli-Q water. The running buffer consisted of 25 mM Tris, 192 mM glycine, and 3.5 mM SDS. The molecular standard used was NZYColour Protein Marker II (Nzytech), range 11-245 kDa. Gels were stained overnight, with Coomassie Brilliant Blue R-250, and washed with Milli-Q water.

2.3.2 Proteomics

As an attempt to screen for potential proteinaceous toxins, the crude extracts of *G. alba* proboscis and skin underwent shotgun proteomics of crude extracts through liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Protein for analysis was retrieved from SDS-PAGE gels to remove salts and other substances may interfere with analysis. The resolving and stacking gels contained 12 % (v/v) and 4 % (v/v) acrylamide, respectively. All wells were loaded with 200 µg of protein. The SDS sample buffer consisted of 70 % (v/v) Tris-HCl, pH 6.8, 30 % (v/v) glycerol, 10 % (m/v) SDS, 0.6 M

DTT, and 0.012 % (m/v) bromophenol blue. The gel was stained overnight, with Coomassie Brilliant Blue R-250, and washed with Milli-Q water. The gel bands were excised into tubes with sodium azide (0.05 %) and stored at 4 ºC.

After this, the gel bands underwent sample destaining and digestion using porcine trypsin. The analysis was performed on a Eksigent NanoLC Ultra 2D separation system (Sciex) coupled to a TripleTOF 6600 mass spectrometer (Sciex). The peptides were separated in a ChromXP C18CL column (0.3 × 150 mm, 3 µm, 120 Å, Eksigent) through chromatography by micro-LC, at 50 ºC. The parameters were set as following: flow rate (5 µL/min), mobile phase A (0.1 % v/v formic acid, 5 % v/v DMSO in water) and B (0.1 % v/v formic acid, 5 % v/v DMSO in acetonitrile). The ESI DuoSpray ionization source was conducted in the positive mode set to an ion spray voltage of 5500 V, 25 psi for both nebulizer gas 1 (GS1) and curtain gas (CUR). The rolling collision had a collision energy spread of 5. ProteinPilot 5.0.1 (Sciex) was used to achieve peptide mass fingerprinting, the parameters were: digestion by trypsin, cysteine alkylation by acrylamide, and gel-based ID as a special factor. The database used to compare the resulting aminoacid sequences, using Blast (Altschul, et al., 1990), was UniProtKB. The lowest e-values and the number of matching peptides per protein determined the accuracy of identification. The tool used to build the diagram was BioVenn.

Using R 4.1.1 (Ihaka and Gentleman, 1996), with the packages *UniprotR*, *biomaRt*, and *GenomicAlignments*, the Accession IDs, of the proteins that resulted from the previous Blast, were converted into *Homo sapiens* genes, and corresponding attributes from the BioMart database were retrieved, specifically Gene Ontology (GO) terms. Then, the ten more frequent GO terms associated to the sub-ontologies: Biological Process, Cellular Component, and Molecular function, were selected, keeping more than ten GO terms when there is more than one GO term with the same number of occurrences as the tenth place. This was made for the Accession IDs of: proteins exclusive to the proboscis and exclusive to skin, and proteins common to both. After this, a Blastp was performed, comparing all the LC-MS/MS resulting aminoacid sequences (total protein, from both organs), total aminoacid sequences of proboscis and total of skin, to customised subsets retrieved from the NCBInr database (e.g. toxin-related proteins and peptides from annelids and molluscs, and a subset specific to glycerotoxins).

2.3.3 Proteome fractioning

Protein fractions were obtained by subjecting the crude protein extracts from proboscis and skin to ultrafiltration, using 100 kDa, 10 kDa, and 3 kDa Amicon spin column filters (Merck Millipore). The three filters were used in series to obtain four different fractions: higher than 100 kDa, in between 10 and 100 kDa, in between 3 and 10 kDa, and lower than 3 kDa, as displayed by Fig. 2.1. The buffer (200 mM DTT, 0.05 M Tris-HCl, pH 7.0) complemented with 1 % v/v Protease Inhibitor cocktail (Sigma-

Aldrich) was replaced with Dulbecco's phosphate-buffered saline (PBS), pH 7.4, a physiological compatible media, through the column filters, and was then used as a vehicle. The proteinaceous nature of the fractions was assessed by SDS-PAGE, as previously described. The resolving and stacking gels contained 9 % (v/v) and 6 % (v/v) acrylamide, respectively, for the >100 kDa fraction, and 12% (v/v) and 6% (v/v) acrylamide, respectively, for the remaining fractions. Wells were loaded with extracts diluted to the same amount of protein (3 µg) in PBS. The Protein concentration of each fraction was determined using a NanoDrop 2000 apparatus (Thermo Fisher).

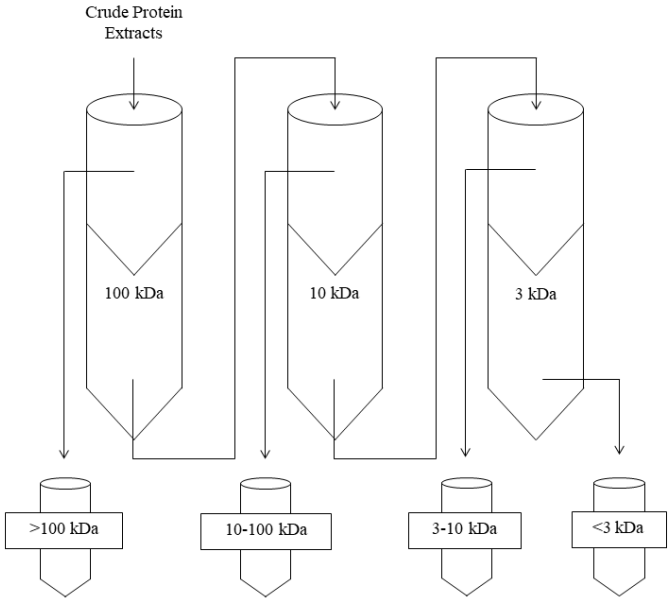

Figure 2.1 – Exemplification scheme of in series process of ultrafiltration using 100 kDa, 10 kDa, and 3 kDa Amicon column filters to obtain four different fractions: >100 kDa, 10-100 kDa, 3-10 kDa, <3 kDa.

2.4 Toxicity testing

The toxicity of proteinaceous extracts from *Glycera alba* proboscis and skin were assessed through a series of short-term *ex vivo* bioassays using gill tissue from the mussel *Mytilus edulis* as a biological model. In brief, after 48 hours of acclimatization, thirty *Mytilus edulis* were randomly assigned to the assays. Gills were harvested by separating the two valves, thus exposing demibranchs, and rinsing them with PBS (pH 7.4) to avoid desiccation. Each valve was considered a biological replicate. Tests were conducted immediately. For both organs (proboscis and skin) of *G. alba*, the protein fraction retained by 100 kDa microfilters was chosen, and five different concentrations of total protein were prepared in PBS: 1; 0.5; 0.1; and 0.05 mg/mL and only PBS (considered as biological control). The tests had six replicas, i.e., n = 6 per experimental condition (dilution and organ of extract). Tested valves were exposed to 1 mL of each test solution (including control, only exposed to cold PBS) for 10 minutes. After

exposure, valves were immediately rinsed with PBS, gill tissue carefully excised and divided in different sections for toxicopathological analyses.

2.4.1 Histopathological analysis

Freshly harvested gill samples were fixated in Davidson's solution (9 – 10 % v/v formalin, 10 % v/v glacial acetic acid, and 30 % ethanol prepared in Milli-Q water) for 24 hours. Afterwards, samples were washed in Milli-Q water (4 × 15 min), dehydrated through a progressive series of ethanol (70 %, for 5 minutes, 95 %, for 15 min, and 100 %, for 35 min), prepared with xylene (20 min) and embedded with molten paraffin (2 × 10 min), in vacuum using a tissue processor (SHANDON, Labexchange).

All samples were sectioned at a 5 µm thickness using a RM 2125 RTS rotary microtome (Leica Microsystems). Then, sections were stained using hematoxylin and eosin as previously described by Costa (2017). All slides were dehydrated, cleared with xylenes, and mounted with DPX resin. Observations were performed using a DM 2500 model microscope (Leica Microsystem).

2.4.2 Comet assay

The alkaline Comet assay was performed as previously described by Martins and Costa (2020) for solid tissue for identifying possible DNA damage caused by exposure to toxin-containing extracts. In brief: after exposed to the *G. alba* proteinaceous extracts, fresh gill samples were placed in 1.5 mL tubes with 200 µL of PBS (pH 7.4), gently with plyers and centrifuged at 1500 rpm, 4 ºC, for 2 min to precipitate debris. The supernatant containing live cells was collected to new tubes.

The first step involved embedding 20 µL of the supernatant in 180 µL of molten (~ 40 ºC) 1 % m/v low melting point agarose (LMPA) prepared in PBS. Two 80 µL drops of LMPA with the incorporated samples were placed in slides precoated with 1.2 % m/v high melting point agarose (HMPA) prepared in TAE buffer. The precoated slides were allowed to dry for at least 48 hours. The slide was immediately covered with a coverslip and LMPA solidified in cold for 15 min. Two slides (technical replicates) were prepared per sample.

The coverslips were then gently removed, and the slides placed in cold lysis buffer (2.64 % m/v NaCl, 3.72 % m/v EDTA, and 5 mM Tris, pH 10), supplemented with 10 % v/v DMSO and 1 % v/v Triton X-100 just before use. Lysis lasted 1 hour at 4 ºC, in the dark. Following, the slides were transferred to cold electrophoresis buffer (1 mM EDTA and 300 mM NaOH) for 40 minutes. Electrophoresis was run for 30 minutes at 25 V, in cold (4 ºC) and dark conditions, using a CSL–COM20 Comet Assay Tank (Cleaver Scientific). Finally, slides were neutralized in cold 0.1 M Tris-HCL buffer (pH 7.5) for 15 minutes and dehydrated in cold methanol for 15 minutes for archiving, if not stained and analysed immediately.

Following neutralization, or rehydration of archived slides in cold Milli-Q water for 30 minutes, the preparations were stained with GreenSafe Premium (Nzytech), diluted to 1:500 in Milli-Q water. Each slide received two 40 µL drops of the diluted dye and were mounted with a coverslip. The dye was allowed to bind to DNA for 10 minutes in the dark, at room temperature, before visualization of nucleoids in a DM 2500 LED model microscope equipped with an EL6000 model ultraviolet source and a MC 190 HD camera (all from Leica Microsystems). Micrographs of Comet fields were obtained in grayscale mode at 400 x magnification.

Scoring of micrographs was done using the software CometScore 1.6 (Tritek), with 100 nucleoids being analysed per sample. Measurements were made for three main metrics: % DNA in Tail, Tail Moment, and Olive Moment. Additionally, the nucleoids were distributed into five classes of % DNA in Tail (0 – 20, 20 – 40, 40 – 60, 60 – 80, and 80 – 100), based on data retrieved from software, for statistical analyses, and quality assessment. The % DNA in Tail was considered a direct measure of DNA damage.

2.4.3 Caspase 3 activity

As an indicator of cellular death, Caspase 3 activity was determinate using Caspase 3 colorimetric assay kit (Sigma-Aldrich). The colorimetric assay performed lead to the normalised activity of the enzyme, determined from the hydrolysis of the peptide substrate acetyl-Asp-Glu-Val-Asp p-nitroanilide, resulting in the release of the p-Nitroaniline (pNA) that has a high absorbance at 405 nm. Cell lysates were obtained by adding 100 µL lysis buffer (1x, 50 mM HEPES, pH 7.4, 5 mM CHAPS, and 5 mM DTT) to approximately 30 mg of gill tissue followed by maceration with a pestle. The samples were centrifuged at 14 000 rpm for 15 min at 4 ºC. Supernatants were transferred to new tubes and stored at -80 ºC until used.

Cell lysate, Assay Buffer (1x, 20 mM HEPPES, pH 7.4, 2 mM EDTA, 0.1 % CHAPS, 5mMDTT), Caspase 3 substrate (20 mM acetyl-Asp-Glu-Val-Asp p-nitroanilide), Caspase 3 inhibitor (2 mM Acetyl-Asp-Glu-Val-Asp-al), Caspase 3 positive control (Caspase 3 in 50 mM HEPES, pH 7.4, 0.1 % CHAPES, 10 mM DTT, 100 mM NaCl, 1 mM EDTA, and 10 % sucrose) were added to a 96 well plate using the reaction scheme found on Table 2.1.

Table 2.1 – Reaction scheme for 96 well plate microassay method for the quantification of Caspase 3 activity.

	Cell lysate	Caspase 3 positive control	Assay buffer	Caspase 3 inhibitor	Caspase 3 substrate
Reagent Blank	-	-	90 µL	-	10 µL
Non-induced cells (control)	5 µL	-	85 µL	-	10 µL
Induced cells (proboscis and skin)	5 µL	-	85 µL	-	10 µL
Positive control	-	5 µL	85 µL	-	10 µL
Negative control	-	5 µL	75 µL	10 µL	10 µL

After adding the Caspase 3 substrate to each well the reaction starts, and the plate was incubated for 90 min, 180 min, and 1320 min, at room temperature. Optical density was measured at 405 nm using a Multiskan SkyHigh Microplate Spectrophotometer (Thermo Scientific).

Activity was estimated from a pNA, a compatible substrate, calibration curve. A series of pNA solutions with a concentration from 0.001 to 0.02 µmol pNA / well were made by diluting the stock solution in Assay buffer (1x, 20 mM HEPPES, pH 7.4, 2 mM EDTA, 0.1 % CHAPS, 5mMDTT). Thus, 100 µL of each dilution was added to a well, including 100 µL of Assay buffer as a blank , and the absorbance was read at 405 nm, using Multiskan SkyHigh Microplate Spectrophotometer (Thermo Scientific).

The concentration of the pNA released from the substrate was calculated from the absorbance values at 405 nm using the calibration curved prepared with defined pNA solutions. The activity of caspase 3 was calculated in µmol pNA released per min per mL of cell lysate based on the following formula:

$$Activity\ (\mu mol\ pNA\ min^{-1}\ mL^{-1}) = \frac{\mu mol\ pNA \times d}{t \times v} \quad [1]$$

Where v is the volume of sample in mL, d is the dilution factor, and t is the reaction time in minutes. After this, the results were normalised dividing the activity per concentration of protein in the cell lysates, resulting in a normalised activity in µmol pNA min^{-1} mg^{-1}.

2.5 Statistical analysis

Statistical analysis was computed with the software R 4.1.1 (Ihaka and Gentleman, 1996), with the packages *gplots*, *ggplot2*, and *dunn.test*. To determine if a set of GO terms is enriched (over-represented) in proboscis or skin, a Fisher's test (non-random associations between the GO terms) was performed. Comparisons were made between proteins exclusive to the proboscis and exclusive to skin. The FDR-corrected *p-value* was set at 0.05. A nonparametric multiple comparison between data was assessed by the Dunn's test, taking "Organ" as an independent variable and "% DNA in Tail" as a dependent variable

for comets, and normalized activity as a dependent variable for caspase 3 activity. The *p-value* was set at 0.05 for all analyses.

3

RESULTS

3.1 Morphoanatomical analysis of the venom system

The normal structure of the eversible proboscis (pharynx) of *Glycera alba* is represented in Fig. 3.1. The proboscis is an extension of the pharynx, being a chiefly muscular organ lined by a thin skin layer that bears multiple sensorial papillae (Fig. 3.1A and B). The findings revealed sac-like glandular organs lined by a pseudostratified epithelium with several types of glandular (secretory) and supporting cells (Fig. 3.1C). Calix cells (acinar-like protein vesicle-laden cells) are present in this epithelium, which is consistent with toxin secretion, while most of the remaining secretory cells in the gland are seemingly mucous-secreting (mucocytes). The inner surface of the gut opening is lined with simple columnar epithelial cells, and four cross-arranged putative venom glands were observed, each connected to a hollow tooth (venom delivery apparatus) by a duct, as highlighted in Fig. 3.1D. The proboscis musculature is longitudinal and arranged in bundles.

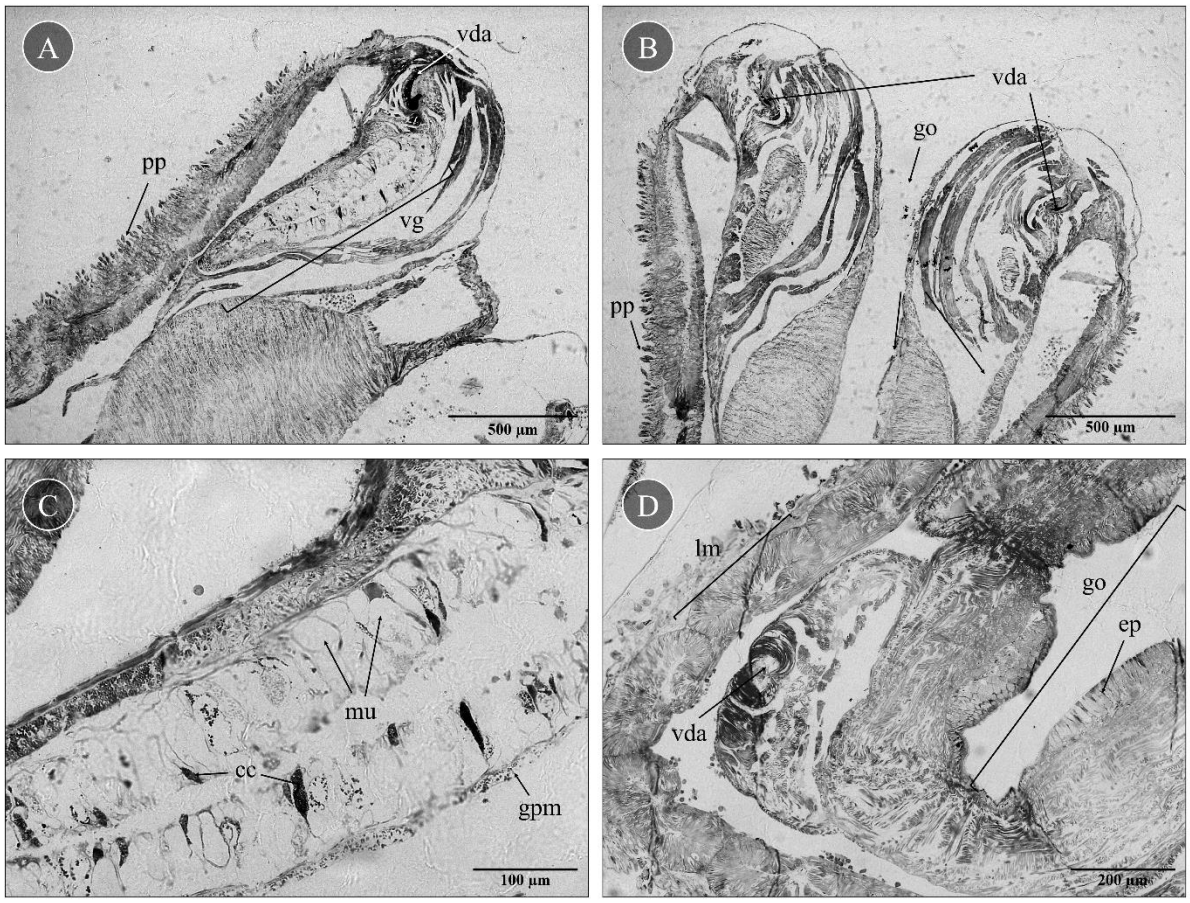

Figure 3.1 – Microanatomy of *Glycera alba*'s proboscis. (**A**) Transversal section across the proboscis, showing the skin laden with sensorial proboscidial papillae (pp) and the venom glands (vg) associated to the venom-delivery apparatus (vda) via a venom duct (not visible in the micrograph). (**B**) Tip of the proboscis evidencing the gut opening (go). It is also covered in proboscidial papillae (pp) and shows two jaws that form the venom-delivery apparatus (vda). (**C**) Higher-power magnification of a venom gland with densely packed glandular cells, calix cells (cc) and mucocytes (mu), surrounded by glandular peripheral musculature (gpm). (**D**) Longitudinal section of a proboscis where a hollow tooth (vda) can be found, as well as an inner pharynx epithelium (ep) lining the gut opening (go). Proboscis is chiefly formed by longitudinal musculature that is arranged in bundles (lm). Resin sections of *Glycera alba* stained using Toluidine blue.

Comparatively, no specialised glandular tissue, such as venom glands, was found at *Glycera alba*'s body wall, a structure, comprising skin and circular and longitudinal muscle layers, shown on Fig 3.2. Precisely, Fig. 3.2A shows a pseudo-stratified epithelium with mucus and cuticle secreting cells as well as a cuticle surrounding it. The structures responsible for locomotion are the parapodium (Fig. 3.2B). In *G. alba*, the parapodia is well-developed with two different rami supported by chaetal lobes, bristles made of chitin, covered in mucocytes. Each segment has two parapodia arranged from the body wall, divided into two branches, the dorsal notopodium (smaller rami) and the ventral neuropodium.

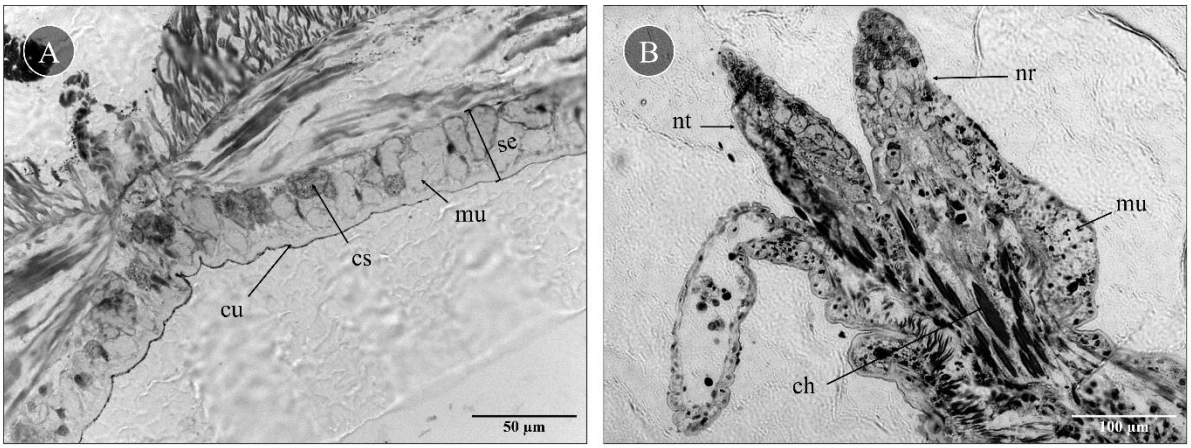

Figure 3.2 – Body wall section of *Glycera alba*. (**A**) Skin epithelium (se) with mucocytes (mu) and cuticle secreting cells (cs), and a cuticle (cu). (**B**) Parapodia divided into two branches, the notopodium (nt) and the neuropodium (nr), supported by chaetae (ch) and lined with mucocytes (mu). Resin sections stained with Toluidine blue

3.2 Comparative proteomics between proboscis and skin

3.2.1 Gel electrophoresis of crude and fractionated protein extracts

Crude protein extracts from the proboscis and skin were separated using SDS-PAGE. Separation yielded scant differences between the two organs displaying an almost identical proteome (Fig. 3.3). Apart from some differences in intensity, due to different protein quantities, both proteomes displayed the same protein bands. According to Padlan and Love (1968), the molecular weight of the single-haem haemoglobin of *G. dibranchiata* is 18 kDa, making haemoglobin a probable match for the intense band between 17 – 20 kDa, present in both organs. In addition, the earthworm *Lumbricus terrestris*'s actin is 42 kDa, which could correspond to the band between 35 – 48 kDa, in proboscis and skin (Huch, 1988). The lack of smearing indicates that the extraction and storage process did not contribute to protein degradation.

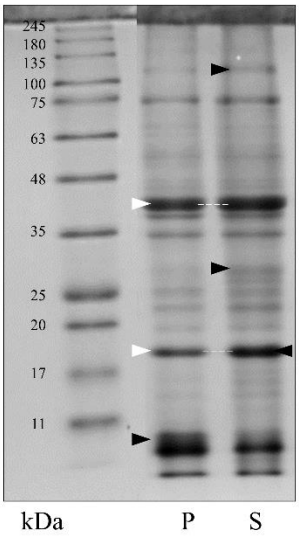

Figure 3.3 – Protein signature of crude extracts from *Glycera alba* proboscis (P) and skin (S). Protein samples were diluted to the same concentration of protein (2.5 mg/mL) in Milli-Q water before SDS-PAGE. Possible matches for the bands between 17 – 20 kDa and 35 – 45 kDa (white arrowhead) are haemoglobin and actin, respectively. The more intense bands, in comparison, are marked with a black arrowhead. The molecular standard used was NZYColour Protein Marker II (Nzytech), range 11-245 kDa.

In an attempt to simplify proteome complexity, the SDS-PAGE of four fractions of the crude extracts from proboscis and skin, were made (Fig. 3.4). Results showed how the higher (>100 kDa) and second higher molecular weight fractions (10 – 100 kDa), Fig 3.4A, and 3.4B, respectively, disclosed important differences between the structures. In these fractions, five bands were visible only in proboscis, three in between 35 – 48 kDa, one amidst 25 – 35 kDa, and one from 48 – 63 kDa. However, the two lower molecular weight fractions (3 – 10 kDa and <3 kDa, Fig. 3.4B), did not display clear differences. Despite all protein samples being diluted to the same amount of protein, 3 µg, the 3 – 10 kDa fraction is less intense than the two higher molecular weight ones, and the <3 kDa did not present a signal. The 10 – 100 kDa and 3 – 10 kDa fractions have a very similar protein signature to the crude extracts, contrasting with the >100 kDa fraction that presented more distinct differences between the organs. The possible matches for haemoglobin and actin are visible in the two middle molecular weight fractions, between 17 – 20 kDa and 35 – 45 kDa, respectively. The fraction that yielded the most crucial differences between the two organs, the one of higher molecular weight (> 100 kDa), was then tested for toxicity on mussel gills.

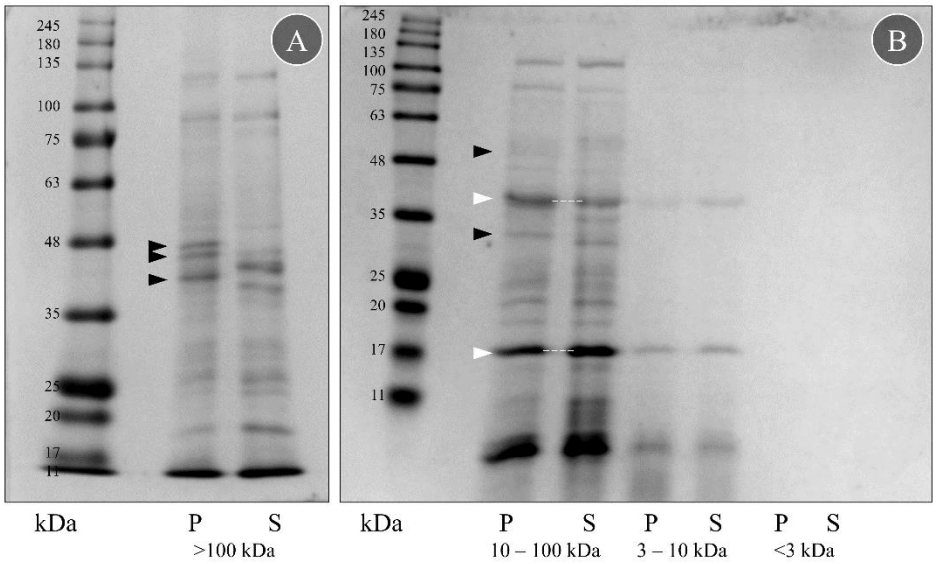

Figure 3.4 – Protein signature of four fractionated extracts from *Glycera alba* proboscis (P) and skin (S). (**A**) Fraction >100 kDa. (**B**) Fractions: 10 – 100 kDa, 3 – 10 kDa, and <3 kDa. Protein samples were diluted to the same amount of protein (3 µg) in PBS before the SDS-PAGE. Bands found only on proboscis are marked with a black arrowhead. Possible matches for the bands between 17 – 20 kDa and 35 – 45 kDa (white arrowhead) are haemoglobin and actin, respectively. The molecular standard used was NZYColour Protein Marker II (Nzytech), range 11-245 kDa.

3.2.2 Proteomics

The analysis of peptides produced by MS/MS yielded a total of 552 proteins with a positive identification against the contrasted UniProtKB database. Of these, 171 were only identified in samples from the proboscis, 121 only in skin, and 260 proteins were common to both (Fig. 3.5). The majority of peptides, and consequently the proteins against which they were matched, related to abundant structural proteins such as: actin (ACT2), myosin (MYS), and tubulin (TBA3 and TBB4B), likely indicating overshadowing, regardless of target organ. Still, metabolically- and physiologically-relevant proteins could be identified, such as heat shock 70 kDa protein (HSP70), and 14-3-3 protein zeta and epsilon (1433Z and 1433E, respectively). However, no toxins were directly identified from either proteome by contrasting against UniProtKB. As such, in an attempt to identify toxins, the homology-based search alignment against the NCBInr database identified possible proteins of interest in the proteomes of both organs (Table 3.1). A total aminoacid sequence comparison (from both organs) with three partial protein sequences of glycerotoxin from *Glycera tridactyla* (accessions AQX92142.1, AQX92141.1, and AQX92140.1), lead to 23 matches, of which three were exclusive to the proboscis. In addition, both organs' peptidic sequence matching with ToxinAnnelida_NCBI identified a synaptobrevin homolog from *Hirudo medicinalis* and five matches with Rac GTPase (*Helobdella* sp.). The last search, of both organs' aminoacid sequence against ToxinMollusca_NCBI, identified serine/threonine-protein phosphatase (*Aplysia californica*), and synaptobrevin, this time from *Aplysia californica*.

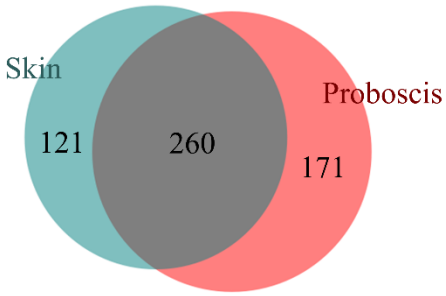

Figure 3.5 – Distribution of the identified proteins on *Glycera alba* proboscis and skin. Each section's area reflects the observed number of proteins identified.

Table 3.1 – Peptidic sequences' protein match, with respective database of alignment, number of matched peptides, and e-Value. Customised subsets retrieved from the NCBInr database: toxin-related proteins and peptides from annelids (TA) and molluscs (TM), and a subset specific to glycerotoxins (GLTx).

Protein Match	Number of matched peptides	NCBInr database	e-Value (Blastp)
Glycerotoxin partial (*Glycera tridactyla*)	23	GLTx	0.004
Rac small GTPase (*Helobdella* sp.)	5	TA	0.021
Synaptobrevin homolog (*Hirudo medicinalis*)	1	TA	2.95×10^{-06}
Synaptobrevin (*Aplysia californica*)	1	TM	5.23×10^{-04}
Serine/Threonine-protein Phosphatase (*Aplysia californica*)	5	TM	1.63×10^{-09}

The most frequent Gene Ontology (GO) terms related to Biological Process (i.e., terms that describe for instance, biological processes, functions and compartments) of genes whose products (proteins) exclusively found in the proboscis were 'protein transport', 'viral process', 'translation', 'intracellular protein transport' and 'protein deubiquitination'. On proteins exclusively to skin, the GO terms 'translation', 'post-translational protein modification', and 'viral process' were also among the more frequent, differing on GO terms such as 'phosphorylation' and 'viral transcription'. From the ten more frequent GO terms of both organs, 'phosphorylation', 'viral transcription', and 'SRP-dependent cotranslational protein targeting to membrane' where enriched (over-represented) in skin, when compared to the proboscis (Fisher's test, FDR-corrected $p < 0.05$). On the other hand, 'protein transport', 'intracellular protein transport', and 'protein deubiquitination' were GO terms enriched in the proboscis (Fisher's test, FDR-corrected $p < 0.05$), as shown in Fig. 3.6.

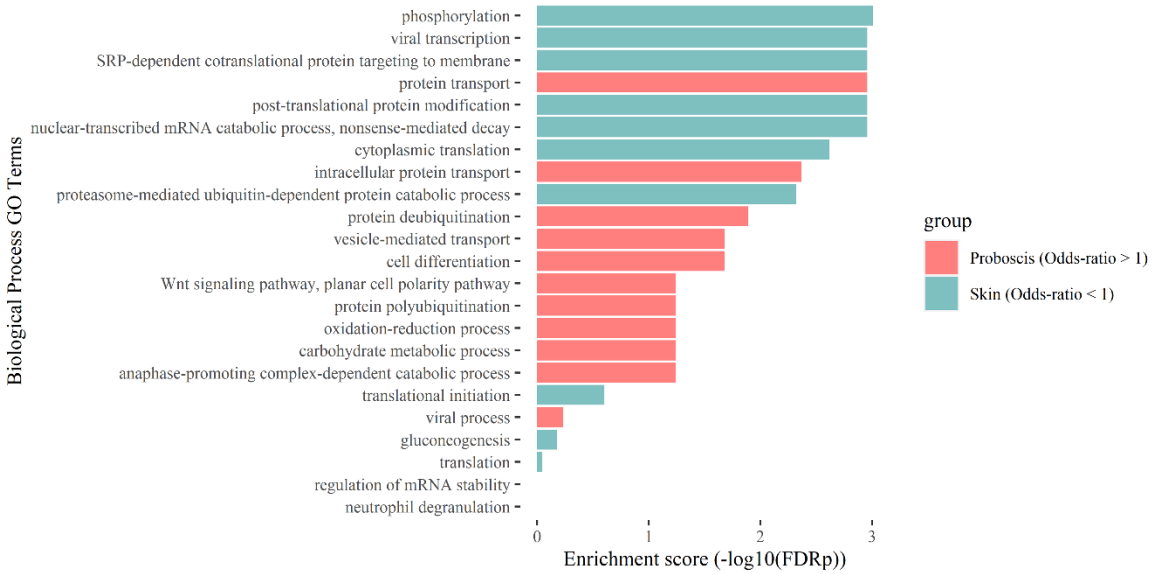

Figure 3.6 – Comparison between Biological Process Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin. When odds-ratio > 1 the GO term is enriched on proboscis, when odds-ratio < 1 the GO term is enriched on skin (Fisher's test, FDR-corrected $p < 0.05$).

The more frequent 'Molecular Function' GO term (molecular-level activities of gene products), of proteins exclusive to both the proboscis and skin, was 'protein binding', with more than twice as many occurrences as the remaining GO terms, followed by 'nucleotide binding' and 'ATP binding'. The non-random association between the GO terms (Fisher's test, FDR-corrected $p < 0.05$) revealed that 'kinase activity', 'ATPase activity', and 'structural constituent of ribosome' were GO terms enriched in skin, and 'identical protein binding', 'GTP binding', and 'GTPase activity' were molecular activities enriched on proboscis (Fig. 3.7).

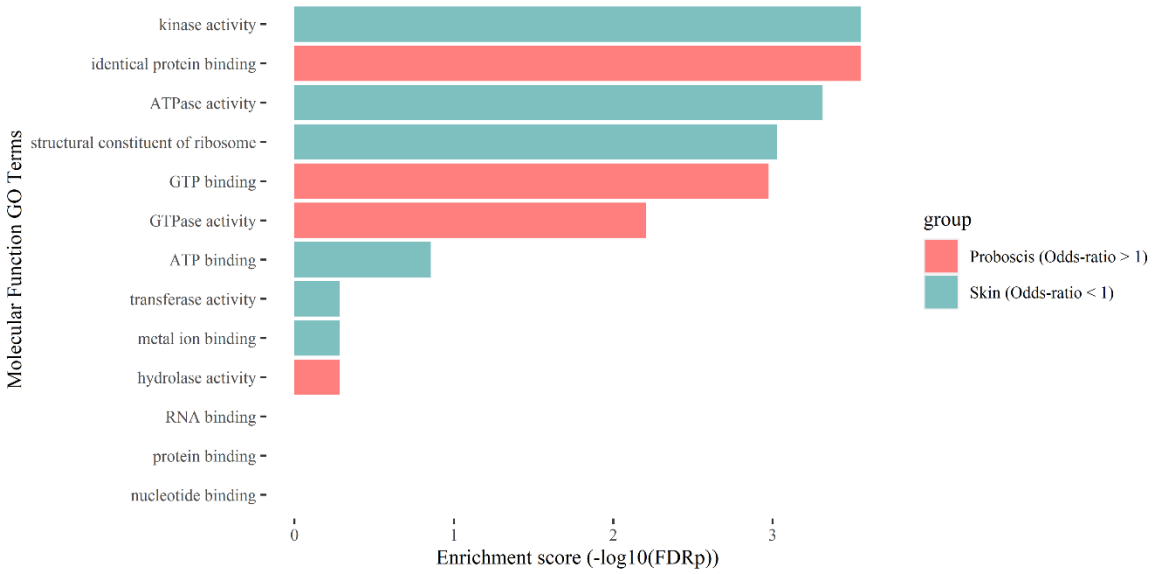

Figure 3.7 – Comparison between Molecular Function Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin. When odds-ratio > 1 the GO term is enriched on proboscis, when odds-ratio < 1 the GO term is enriched on skin (Fisher's test, FDR-corrected $p < 0.05$).

The ten more frequent locations (relative to cellular structures), in which proboscis proteins and skin proteins perform their function (‘Cellular Component’ GO term), are the same apart from ‘intracellular membrane-bounded organelle’, for proboscis, and ‘extracellular region’, for skin. Some of the more frequent GO terms were ‘cytosol’, ‘cytoplasm’, ‘extracellular exosome’, ‘membrane’, and ‘nucleus’. Fisher’s test revealed that ‘intracellular membrane-bounded organelle’ is a location over-represented in proboscis and ‘extracellular region’ a location enriched in skin (Fig. 3.8).

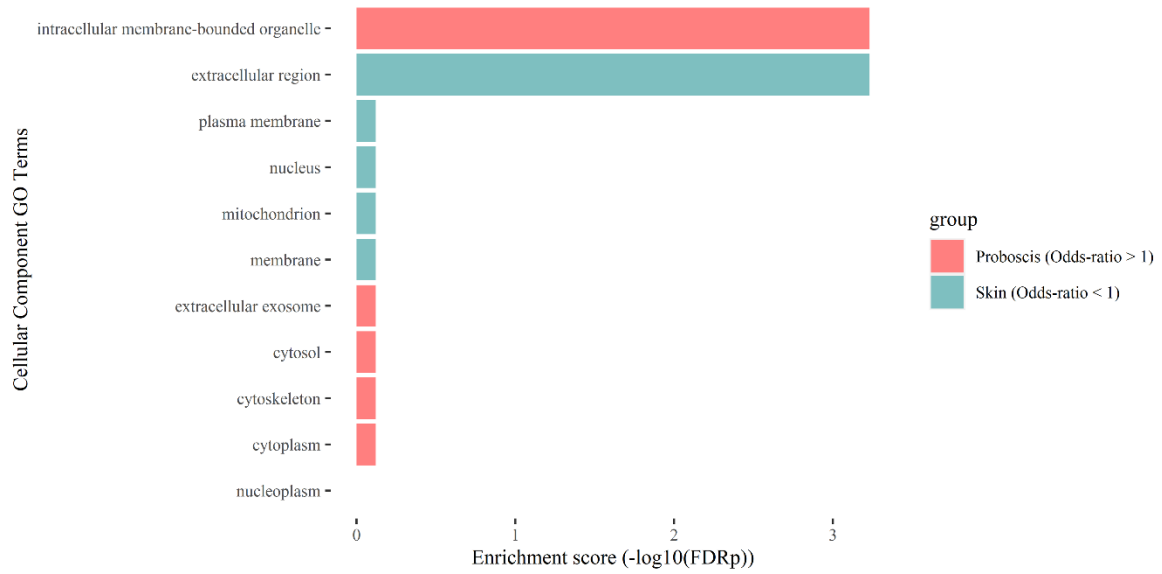

Figure 3.8 – Comparison between Cellular Component Gene Ontology (GO) terms of proteins exclusively found in proboscis and exclusively found in skin. When odds-ratio > 1 the GO term is enriched on proboscis, when odds-ratio < 1 the GO term is enriched on skin (Fisher's test, FDR-corrected $p < 0.05$).

3.3 Toxicity testing

3.3.1 Histopathological analysis

Mussel gills exposed to PBS (here considered as biological control) showed no significant lesions (Fig. 3.9A). These gills had a well preserved morphology, with two lamellar structures comprised of many filaments bound by ciliary plates, they had haemolymphatic sinus, and endothelial cells. Also, the ciliary plates were intact and few haemocytes could be found in the intertubular tissue, which should exclude significant inflammatory responses.

Comparing to the control, gills exposed to skin extracts regardless of concentration, and higher dilution proboscis extracts (0.5, 0.1, and 0.05 mg total protein/mL) evidenced no visible alterations (Fig 3.9B, C, and E). On the other hand, valves exposed to the higher concentration (1 mg/mL) extract from the *G. alba* proboscis (Fig 3.9D), yielded moderate levels of histopathological alteration to gill epithelia. Specifically, focal oedematous fluid retention in between cells was more evident, as well as the presence of haemocytes with nuclear abnormalities, mild abrasion of cilia, and brown pigment (lipofuscin-like) accumulation. Additionally, compact nucleus with darker blue cytoplasm, as opposed to pink/light purple, are an indicator of early stages of cellular death. The lack of an inflammatory response, like haemocyte migration, and intact cellular membranes indicates that cellular death occurred by apoptosis as opposed to necrosis.

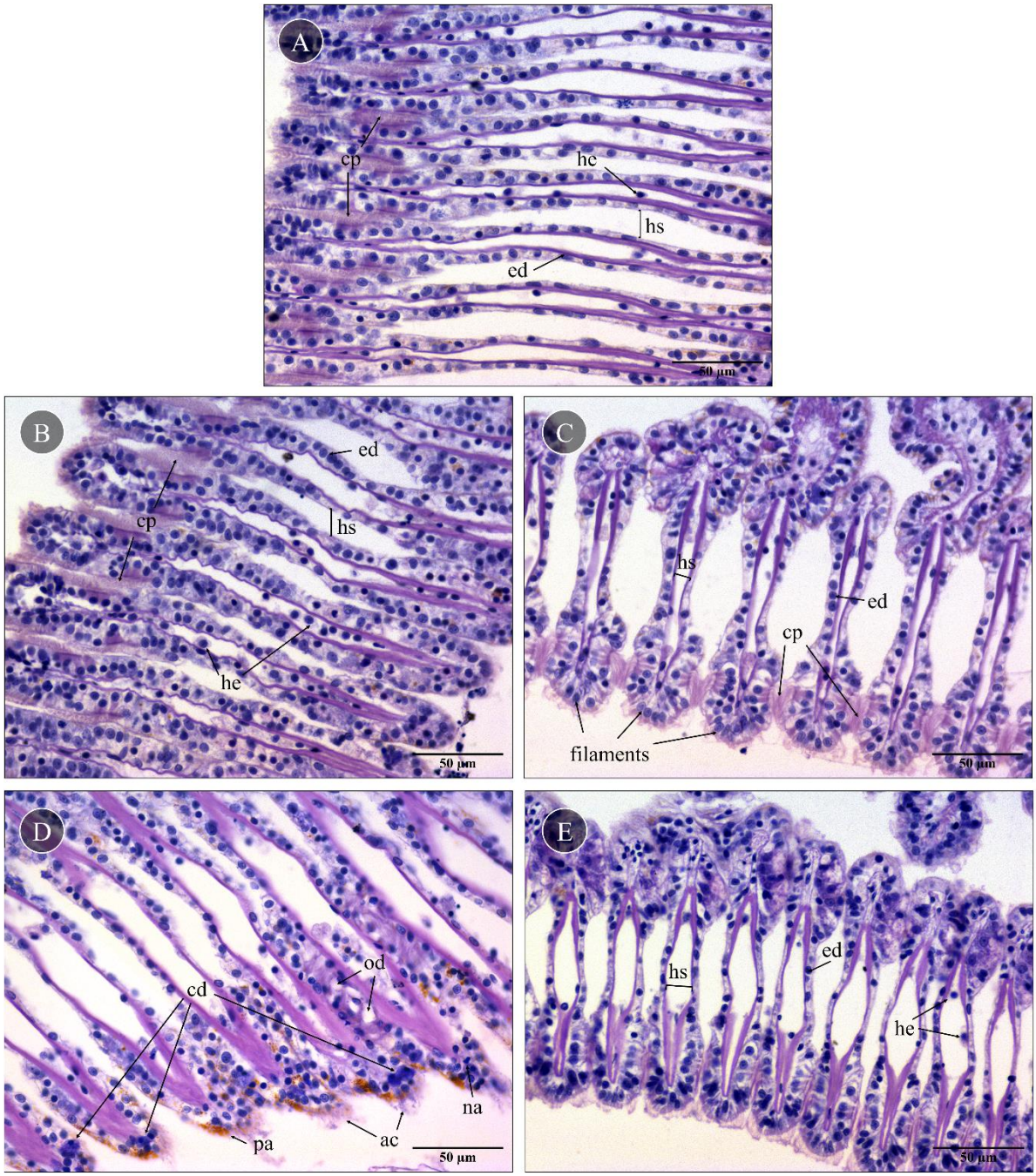

Figure 3.9 – Representative histological gill sections of *Mytilus* sp. exposed to *G. alba* proboscis and skin extracts. (**A**) Control gill, exposed to PBS. Note the presence of ciliary plates (cp), the haemolymphatic sinus (hs), endothelial cells (ed), as well as few haemocytes (he) in vessels. (**B**) and (**C**) Gills exposed to skin extracts (1 and 0.1 mg/mL, respectively). No visible alterations, comparing to the basal levels. (**D**) Gill exposed to proboscis extract of 1 mg/mL. Presence of oedema (od), early stages of cell death (cd), haemocytes with nuclear abnormalities (na), abrasion of cilia (ac), and pigment accumulation (pa). (**E**) Gill exposed to proboscis extract of 0.1 mg/mL. No visible alterations, comparing to the basal levels. Paraffin sections stained with hematoxylin and eosin (H&E).

3.3.2 Comet assay

The used method yielded quality nucleoids with a uniform dispersion on the slides (as seen in Fig. 3.10A). Each sample had a unimodal distribution of nucleoids per class, which indicates a reduced influence of the experimental procedure in generating necrotic and apoptotic cells, supporting the scoring metric reliability.

According to the scoring metric obtained, % DNA in Tail, the gill cells treated with proboscis extract had 44.3 % DNA in tail, when gill cells treated with proteins from skin and the control had 23.4 % and 20.7 % DNA in tail, respectively. The gills exposed to proboscis extracts had significantly higher DNA damage, comparatively to gills exposed to skin extracts and control (Dunn's test, $p < 0.05$), Fig 3.10B. When compared with each other, there were no significant differences between gills exposed to skin extracts and the control (Dunn's test, $p = 0.41$).

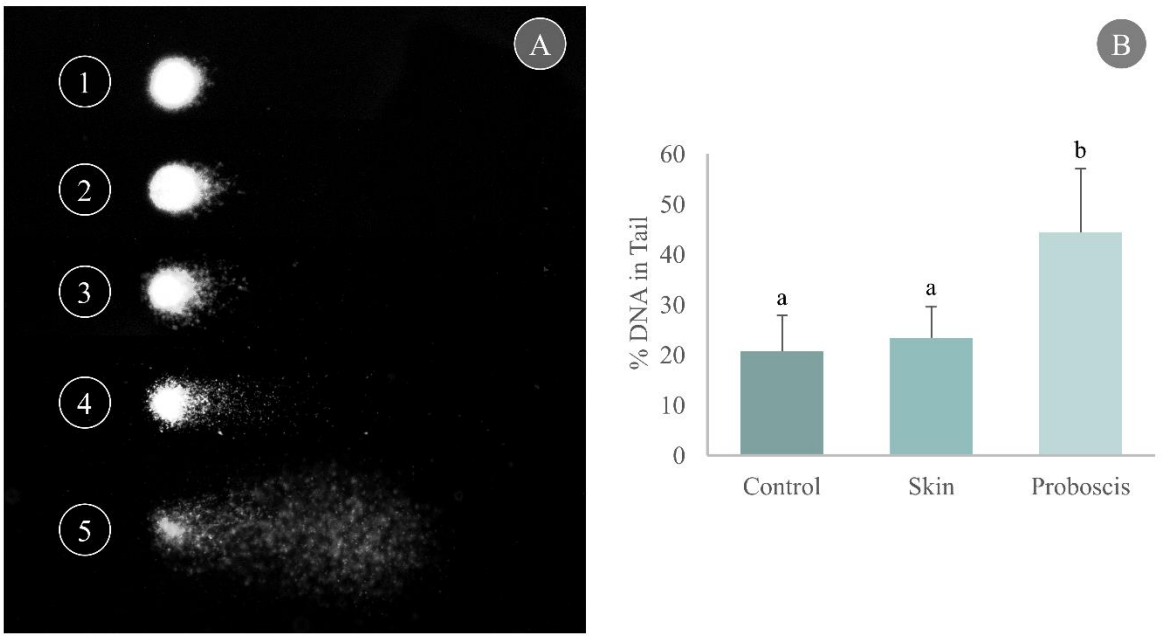

Figure 3.10 – Results from Comet assay to identify DNA damage. (**A**) Representative nucleoids of Comet classes based on % DNA in Tail, obtained using CometScore 1.6. (1) 0 – 20 %, (2) 20 – 40 %, (3) 40 – 60 %, (4) 60 – 80 %, (5) 80 – 100 %. (**B**) Comet assay results expressed as % DNA in Tail + standard deviation, from mussel gills exposed to proboscis and skin extracts, plus control (gills treated with PBS). The letters (a and b) indicate significant differences between extracts (Dunn's test, $p < 0.05$).

3.3.3 Caspase 3 activity

The activity of Caspase 3 in mussel gills treated with *G. alba* proboscis and skin extracts, was measured at 90 min, 180 min, and 1320 min (the latter meaning overnight) (Fig. 3.11). All the non-parametric multiple comparisons (Dunn's test) had a $0.67 < p < 0.29$, meaning that there are no significant

differences between any experimental condition. Specifically, after 90 min, caspase 3 activity was 0.085, 0.09, and 0.11 nmol pNA min^{-1} mg^{-1} in gills exposed to proboscis and skin proteins and the control, respectively. The activity decreased over time, since the increases in absorbance of the reaction product did not emulate the increases in reaction time. Consequently, at 180 min, caspase 3 activity was determined as 0.048 nmol pNA min^{-1} mg^{-1} and at 1320 min was 0.012 nmol pNA min^{-1} mg^{-1}, in all gills.

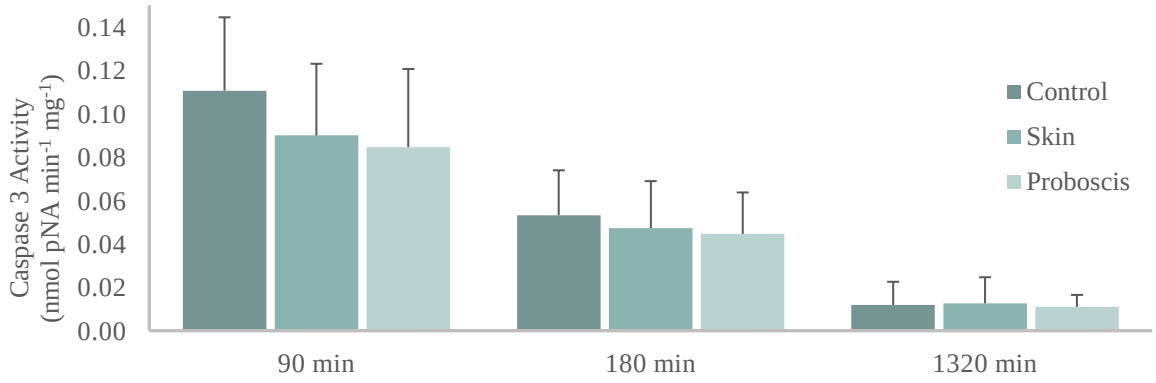

Figure 3.11 – Caspase 3 Activity in mussel gills exposed to *G. alba* proboscis and skin extracts. Normalised caspase 3 activity in mussel gills exposed to proboscis and skin extracts and control (gills treated with PBS), at 90 min, 180 min, and 1320 min. The results represent the biological replicates' mean normalised activity + standard deviation, calculated from the optical density of p-Nitroaniline (pNA) measured at 405 nm. No significant differences between the organs for all multiple comparisons (Dunn's test, $p > 0.05$).

4

DISCUSSION

The current study demonstrated that *G. alba* possesses well-differentiated venom glands located in the proboscis. These glands are the chief organ involved in toxin (venom) secretion since no venom-delivery structures/apparatus were found in the skin. The venom glands seem to have mechanisms to secrete proteinaceous toxins, a class of biomolecules that allocates important bioactives. In accordance with the morphoanatomical description, it may be inferred that the proboscis and the skin possess relevant differences regarding their specific proteomes, which is also linked to different molecular/physiological pathways. The toxicity assays indicate that *G. alba*, as for other congeneric species, likely secretes high molecular mass toxins/enzymes that can exert negative effects onto their prey, which makes them an important target for marine-based drug discovery due to the relevance of these bioactives for biomedical applications.

Glycera alba, like other congenerics such as *G. tridactyla* and *G. dibranchiata*, are ambush predators of the soft-bottom intertidal (Ockelmann and Vahl, 1970). Similarly to these other species, the current work demonstrated that *G. alba* is equipped with venom glands and perforated jaws with a venom duct linking them, which form a specialised delivery system to inject toxins into prey. Its strong muscular eversible proboscis extends quickly, allowing for a fast attack. It must also be noted that, this organ is overlaid with sensorial organs (proboscidial papillae), making it an effective structure for detecting prey and subsequent ambush predation. In addition, the existence of well-developed parapodium, arranged along the body axis, allows for fast crawling and swimming. Overall, the organism's feeding strategy is well reflected into its anatomy. Body wall sections were comprised of mainly muscle and epithelial tissue. Despite, being lined with cells of secreting nature (mucocytes), the skin epithelium mainly secretes the cuticle and the mucus layer, to supply with mechanical protection. The lack of acinar-like cells in the integument of the body wall suggests that specialised secretion of proteinaceous toxins is exclusive to proboscis, the structure of synthesis and delivery. Previous work by Gonçalves and Costa (2020), identified a venom apparatus, in *Glycera alba*'s proboscis, composed of four venom organs, each assigned to a toxin-secreting lobe fully arranged with glandular cells. The present

morphoanatomical analysis complements our existing knowledge of the *G. alba*'s venom gland through the identification of acinar like cells packed with protein vesicles, surrounded by mucocytes. The calix (acinar-like) cells laden with protein vesicles, found in the venom gland, are an indicator that the venom secretion cocktail comprises proteinaceous toxins and enzymes. The venom gland mucocytes could have a similar function to those in the anterior glands found in cephalopods, which are mainly mucus secreting organs, being responsible for the secretion of the toxin vehicle (Gennaro, et al., 1965). Overall, morphoanatomical description revealed structures that resonate with proteinaceous toxin secretion in *Glycera alba*.

Molecular characterisation revealed interesting differences regarding proboscis and skin specific proteomes and allowed for the identification of proteins of interest in *G. alba* protein extracts. Aminoacid sequence match against UniProtKB database revealed mainly the identification of abundant structural proteins, such as actin, myosin, and tubulin, regardless of target organ. The reason for that is that UniProtKB is a curated non-redundant protein sequence database, and the aminoacid sequences that resulted from LC-MS/MS were obtained through whole proteome shotgun analysis. As observed in the morphoanatomical description of *G. alba*, proboscis and skin are highly muscularised organs, so it is expectable that the approach was hindered by the overshadowing of structural proteins, not to mention that glycerids are low genomic annotated organisms. A similar result was obtained from the SDS-PAGE of crude extracts, where both proteomes displayed the same integral bands, with a possible match for actin and haemoglobin. As the structural comparison between the proboscis and skin indicates, the proboscis is the main venom-secreting organ, so differences between their proteomes are likely associated with toxins and permeabilizing agents, among other bio-reactives. This approach is imperfect but was chosen to be the first approach to an unknown proteome. To increase its reliability, a better fractioning would be required, as well as a larger sample size, and more computational work. Nonetheless, proteins with biotechnological potential were identified, such as 14-3-3 proteins and heat shock proteins 70. Heat shock protein 70 regulates several housekeeping processes, such as protein trafficking, folding, and functional maturation (Anokwuru, et al., 2021) and 14-3-3 proteins have a function in cell cycle arrest, signal transduction and apoptosis (van Hemert, et al., 2001). It must be noted that Rodrigo, et al. (2021) disclosed that the toxin mix, secreted by the intertidal annelid *Eulalia* sp., that includes these same two previous proteins, induce apoptosis and is more effective against human cancer cells than normal cells. Altogether, these findings emphasize the relevance of Polychaeta in the bioprospecting for novel proteinaceous bioactives for potential biomedical applications.

The NCBInr database search, however, identified toxins and other proteins of interest in both structures, since the data comes from various sources including GenBank and the RefSeq databases, which may or may not be manually curated. Homology matching against these subsets may thus involve biasing, but a large number of peptide matches may be expected, which implies a cautionary interpretation of data.

One of the interesting protein matches was with two, out of three, glycerotoxins (GLTx) partial protein sequences, from *G. tridactyla*. With a total of 23 *G. alba* peptide matched against these sequences, there is a probable identification of a form of neurotoxin or its precursors in *Glycera alba*'s proteome. As mentioned in the introduction, GLTx up-regulates the activity of calcium channels, which increases spontaneous neurotransmitter releases (Manaranche, et al., 1980), having possible applications regarding drug development for chronic pain. Also noteworthy is the matching against a few toxin-related proteins from other marine invertebrates, namely synaptobrevin and a synaptobrevin homolog, synaptic vesicle proteins essential in secretion (Ahnert-Hilger, et al., 1996), as well as Rac GTPases and serine/threonine-protein phosphatase. Despite the reduced quality of matching, these findings reiterate the potential of *Glycera*'s proteome.

Gene Ontology (GO) enrichment analysis was performed with annotations from matches with UniProtKB database due to better parameters for quality identification. When comparing both organs, 'protein transport' is enriched in proboscis, which is compatible with toxin/enzyme secretion since 'protein transport' is a biological process GO term defined by the directed movement of proteins into or out of a cell, as well as 'protein deubiquitination', a GO term that refers to posttranslational modifications, more specifically the removal of ubiquitin groups from a protein. As opposed to GO terms enriched in skin, such as 'phosphorylation' and 'viral transcription', the introduction of a phosphate group into a molecule, that regulates several cellular processes and contributes to signal transmission, and viral genome transcription within the host cell, respectively. Furthermore, 'vesicle mediated transport' was found to be an enriched GO term in the proboscis, which is congruent with the existence of the densely-packed protein vesicles in the calix cells of venom gland. These findings suggest high protein synthesis in the proboscis, as expected from an organ adapted for venom-secretion.

Protein separation of fractionated extracts revealed further evidence for venom secretion, the two higher molecular weight fractions (>100 and 10 – 100 kDa), from both proboscis and skin, were different from each other in the range of 48 – 63, 35 – 48, and 25 – 35 kDa. According to von Reumont, et al., (2014), *G. dibranchiata* toxins are classified into five categories: Pore-forming toxins, neurotoxins, protease inhibitors, CAP domain toxins, and other enzymes. Comparing the molecular mass of the different bands observed with the previously described toxins found at *G. dibranchiata* enabled suggesting seven possible peptide matches: serpins, chitinases, peptidase S1 and S10, hyaluronidases, lipocalins, and metalloproteinase M12, based on their molecular mass. Furthermore, the observed proteins could contain a CAP domain such as CRISPs, cysteine-rich secretory proteins, toxins found in a large range of animal species (Heyborne and Mackessy, 2016). The respective molecular mass of the previously described peptides, as well as the species in which the molecular weight has been described, can be found in Appendix 1. In Figure 3.4 are showed bands outside the range of each specific fraction, this is a consequence of the filters used. Amicon spin column filters retain around 30 % of lower molecular weight

proteins than what the filter indicates, resulting in more simplified fractions of the original extract other than defined molecular weight fractions.

The toxicity assays revealed that high molecular mass toxins, among other bio-reactives, have a cytotoxic effect in mollusc gill cells, despite the relatively low concentration of protein extracts. The higher concentration (1 mg/mL) extract from the proboscis of *G. alba* caused DNA damage that eventually led to cellular apoptosis. Additionally, oedematous fluid retention is a consequence of trouble maintaining the osmotic equilibrium, nuclear abnormalities in hemocytes reveal abnormal chromatin condensation, and brown pigment (lipofuscin) accumulation can be a result of lipid oxidation and degradation usually associated with the presence of defence cells. The lack of observable changes in the gills exposed to higher dilution proboscis extracts (0.5, 0.1, and 0.05 mg total protein/mL), when comparing to the basal levels, indicates that these concentrations are too low to trigger a toxicopathological alteration. Regarding the valves exposed to skin extracts (1, 0.5, 0.1, and 0.05 mg/mL), the results were similar, revealing that this structure does not cause a noxious effect. This resonates with both the GO enrichment analysis, that suggested that pathways involving protein secretion were more expressed on proboscis proteins, and the morphoanatomical analysis, which indicates that toxin location is exclusive to proboscis.

Another cytotoxic effect observed was DNA fragmentation, which was significantly higher in gill cells exposed to proboscis proteins, when compared to gill cells exposed to both control and skin extracts. This result is congruent with the presence of toxins in the former extract, indicating that DNA fragmentation, and consequently, molecular damage are consequences of the present cocktail of toxins, also higher percentage of DNA damage can be an indicator of cellular death. Gills exposed to skin extracts have the same % DNA in tail as the control, indicating that the structural proteins from muscle and skin do not trigger a toxicological effect. Tree comet scores metrics demonstrated effectiveness in measuring DNA damage: the percentage of the DNA in the tail (% DNA in tail), tail moment (tail length times percentage of DNA in the tail), and the Olive moment (distance between the centres of the head and tail times the rate of DNA in the tail) (Hong, et al., 2020). The previous study indicates that the two most crucial comet parameters, % DNA in tail and Olive moment, have a significantly high positive correlation, and that is why the only metric used to determine DNA damage in this study was % DNA in tail. Despite previous results, the observed apoptosis in histopathology and DNA fragmentation in comet assay, the activity of caspase 3 (biomarker used as a measurable indicator of apoptosis) was the same for all experimental conditions, suggesting that apoptosis is being equally induced in gills exposed and not exposed to toxins. Since caspase 3 is the most studied mammalian effector caspase involved in the intrinsic pathway (Inoue, et al., 2009), the lack of significant differences and the incongruence with previous results can be interpreted in two possible ways. The first explanation for this is that the early stages of apoptosis, found in the previous histopathological analysis of gill tissue, are being triggered by the extrinsic pathway as opposed to the intrinsic pathway. Considering the exposure method used to

subject the biological model to the protein extracts (submersion/absorption) and the nature of said extracts, it is plausible that apoptosis was triggered by the extrinsic pathway (activation of death cellular membrane receptors of the TNF family) and not the intrinsic pathway (which involves initial mitochondrial perturbation resulting from cellular stress or cytotoxic insults). Since caspase 3 is an effector caspase involved in the intrinsic pathway its activity would not have increased upon exposure, leading to the same activity for all organs and measurements (Inoue, et al., 2009). Another possible explanation is that since the kit used was designed for mammals and not for bivalves, the supplied substrate would not be compatible with the bivalve's caspase 3. Previous research, revealed that the bivalve's apoptotic and inflammatory pathways are different than vertebrates and suggested that there might not exist a direct homolog to certain caspases, including executioner caspases that have unique motifs in the prodomains (Vogeler, et al., 2021).

The tested protein extracts were a mixture of all types of proteins found at each respective structure, they were not comprised of purified toxins, containing structural proteins from muscle tissue and epithelium as well as other peptides. One consequence of this is that the quantities of each individual component in every extract can vary. Nevertheless, variations, in this case, seem to be neglectable, as the usage of biological replicates gave similar results. Another important aspect to note is that *G. alba*'s venom causes cardiac arrest, progressive paralysis, and convulsions, typical effects on the nervous system, and gills are mainly comprised of epithelial tissue. In addition, *Glycera alba*'s typical method of delivery is injection, not diffusion. Altogether, the observed negative effects may be subdued when compared to the *Glycera*'s natural venom injection.

5

CONCLUSION

The biotechnological employment of marine bio-reactives contributes to the development of blue economy by providing a sustainable use of novel marine molecular resources for economic growth, without deploying aggressive approaches such as fisheries. A comprehensive characterisation of these natural products is necessary for the development of possible applications and consequent sector innovation. However, marine bioprospecting encounters several hardships, as in high biodiversity, and insufficient genomic resources, which renders difficult the identification, extraction, and purification of compounds, as demonstrated by the present work. The present study contributed to the relevance of venomous invertebrates while evidencing the difficulties of the field of research.

In venomous organisms, the description of the venom-delivery apparatus is relevant to understand the ecological role of the studied toxins since the delivery system needs to be compatible with the toxin's mechanism of action and consequent symptoms. The current findings revel that *Glycera alba*, does possess a well-differentiated venom-delivery apparatus, which may suggest specialised toxins directed against specific molecular targets. Despite constraints, overshadowing caused by structural proteins and low genomic annotation for *G. alba*, several proteins of interest were identified that might, precisely exert a more specific effect onto several cellular processes. Such as, 14-3-3 protein, a biomolecule with a role in cell cycle arrest, and glycerotoxin, a potent neurotoxin. In addition, gene ontology enrichment analysis suggested that different molecular and physiological pathways are associated to different organs, with protein posttranslational modifications and transport being enriched in proboscis, which is compatible with toxin/enzyme secretion. Moreover, higher molecular weight proteins caused cytotoxic effects to molluscs' cells, such as lipid degradation, osmotic imbalance, DNA degradation, and apoptosis. A combination of the morphoanatomical assay with proteomics and toxicity testing validated the evidence of toxins and the classification of *Glycera alba* as a venomous animal such as its congenerics. Furthermore, cytotoxic effects and differentiation of venom systems can be indicative of toxin specificity and eventually even mode-of-action, which in turn leads to high biotechnological potential for the purpose of drug discovery.

In future work, since caspase 3 might not be involved in triggering the observed apoptosis, a study with caspase 8 would be of relevance, since it's an important extrinsic pathway caspase. In addition, a toxicology testing with the second higher molecular weight fraction (10 – 100 kDa), that also displayed differences between structures in the SDS-PAGE, could be appealing, to compare differences in toxicology. Moreover, a protein purification improvement would minimize the overshadowing created by structural proteins in the tested extracts. Additionally, transcriptomics could allow the screening of peptides and overcome the challenges set by mass spectrometry. Overall, the key to produce a comprehensive understanding of an organism, its proteome, and possible applications for the proteinaceous toxins seems to rely on interdisciplinary studies.

BIBLIOGRAPHY

Ahnert-Hilger, G. et al., 1996. Synaptobrevin is essential for secretion but not for the development of synaptic processes. Eur J Cell Biol. 70, 1-11. PMID: 8738414 .

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anokwuru, C., Makumire, S., Shonhai, A., 2021. Bioprospecting for Novel Heat Shock Protein Modulators: The New Frontier for Antimalarial Drug Discovery? Springer, Cham. 1340, 187-203. https://doi.org/10.1007/978-3-030-78397-6_8

Barbosa, A.J.M., Roque, A.C.A., 2019. Free Marine Natural Products Databases for Biotechnology and Bioengineering. Biotechnol. J. 14:1800607. https://doi.org/10.1002/biot.201800607.

Barth, T. et al., 2019. Biochemical and structural characterization of a protein complex containing a hyaluronidase and a CRISP-like protein isolated from the venom of the spider *Acanthoscurria natalensis*. Journal of Proteomics. 192, 102-113. https://doi.org/10.1016/j.jprot.2018.08.012.

Burgess, J.G., 2012. New and emerging analytical techniques for marine biotechnology. Curr. Opin. Biotechnol. 23, 29-33. https://doi.org/10.1016/j.copbio.2011.12.007.

Caplan, S.L., Zheng, B., Dawson-Scully, K., White, C.A., West, L.M., 2016. Pseudopterosin A: Protection of Synaptic Function and Potential as a Neuromodulatory Agent. Mar. Drugs. 14, 55-68. https://doi.org/10.3390/md14030055.

Carter, N.J., Keam, S.J., 2007. Trabectedin, A Review of its Use in the Management of Soft Tissue Sarcoma and Ovarian Cancer. Drugs. 67, 2257-2276. https://doi.org/10.2165/000034952007671-50-00009.

Casewell, N.R., Wüster, W., Vonk, F.J., Harrison, R.A., Fry, B.G., 2013. Complex cocktails: the evolutionary novelty of venoms. Trends in Ecology and Evolution. 28, 219-229. https://doi.org/-10.1016/j.tree.2012.10.020

Chew, J.S.C. et al., 2008. Cystatin C–A Paradigm of Evidence Based Laboratory Medicine. Clin Biochem Rev. 29, 47-62. PMID: 18787643.

Christinat, A., Leyvraz, S., 2009. Role of trabectedin in the treatment of soft tissue sarcoma. Onco. Targets. Ther. 2, 105-113. https://doi.org/10.2147/OTT.S4454.

Costa, P.M., 2017. The Handbook of Histopathological Practices in Aquatic Environments: Guide to Histology for Environmental Toxicology; Academic Press: Cambridge, MA, USA; ISBN 978-0-12-812032-3.

Coutinho, M.C.L., Teixeira, V.L., Santos, C.S.G., 2017. A Review of "Polychaeta" Chemicals and their Possible Ecological Role. J. Chem. Ecol. 44, 72-94. https://doi.org/10.1007/s10886-017-0915-z.

Darwiche, R. et al., 2016. Structural and functional characterization of the CAP domain of pathogen-related yeast 1 (Pry1) protein. Scientific Reports. 6, 28838. https://doi.org/10.1038/srep28838.

Deheyn, D.D., Latz, M.I., 2009. Internal and secreted bioluminescence of the marine polychaete *Odontosyllis phosphorea* (Syllidae). Invertebrate Biology. 128, 31-45. https://doi.org/10.1111/j.1744-7410.2008.00149.x.

Durani, V., Magliery, T.J., 2013. Protein Engineering and Stabilization from Sequence Statistics: Variation and Covariation Analysis. Methods in Enzymology. 523, 237-256. https://doi.org/10.-1016/B978-0-12-394292-0.00011-4.

Flint, R.W., Rabalais N.N.,1980. Polychaete Ecology and Niche Patterns: Texas Continental Shelf. Mar. Ecol. Prog. 3, 193-202.

Fry, B.G. et al., 2009. The Toxicogenomic Multiverse: Convergent Recruitment of Proteins Into Animal Venoms. Annual Review of Genomics and Human Genetics. 10, 483-511. https://doi.org/10.114-6/annurev.genom.9.081307.164356

Gennaro, J.F., Lorincz, J.A.E, Brewster, H.B., 1965. The anterior salivary gland of the octopus (*Octopus vulgaris*) and its mucous secretion. Ann N Y Acad Sci. 118, 1021–5. https://doi.org/10.1111/j.1-7496632.1965.tb40168.x.

Gibbs, P.E., Bryan G.W., 1980. Copper – The Major Meal Component of Glycerid Polychaete Jaws. Mar. Biol. Ass. U.K. 60, 205-214. https://doi.org/10.1017/S0025315400024267.

Gonçalves, C., Costa, P.M., 2020. Histochemical detection of free thiols in glandular cells and tissues of different marine Polychaeta. Histochemistry and Cell Biology. 154, 315-325. https://doi.org/10.1007/s00418-020-01889-3.

Göransson, U. et al., 2019. The Toxins of Nemertean Worms. Toxins. 11, 120. https://doi.org/ 10.3390/toxins11020120.

Hames, B.D., 1998. Gel electrophoresis of proteins: A practical approach, Third Edit. Oxford University Press, New York.

Hamid, R. et al., 2013. Chitinases: An update. J Pharm Bioallied Sci. 5, 21-29. https://doi.org/ 10.4103/0975-7406.106559.

Heyborne, W.H., Mackessy, S.P., 2016. 16 Cysteine-rich Secretory Proteins in Reptile Venoms. Handbook of Venoms and Toxins of Reptiles. 321-332.

Hong, Y. et al., 2020. Deep learning method for comet segmentation and comet assay image analysis. Scientifc Reports. 10, 18915. https://doi.org/10.1038/s41598-020-75592-7.

Huch, R., D'Haese, J., Gerday, C., 1988. A soluble calcium-binding protein from the terrestrial annelid *Lumbricus terrestris* L. Journal of Comparative Physiology B. 158, 325-334. https://doi.org/10.1007/BF00695331.

Ihaka, R., Gentleman, R., 1996. R: A Language for Data Analysis and Graphics. J. Comput Graph Stat. 5, 299-314. https://doi.org/10.1080/10618600.1996.10474713.

Inoue, S. et al., 2009. Ordering of caspases in cells undergoing apoptosis by the intrinsic pathway. Cell Death and Differentiation. 16, 1053-1061. https://doi.org/10.1038/cdd.2009.29.

Iwanaga, S., Lee, B.L., 2005. Recent Advances in the Innate Immunity of Invertebrate Animals. J. Biochem. Mol. Biol. 38, 128-150. https://doi.org/10.5483/BMBRep.2005.38.2.128

Jiang, R. et al., 2011. 93-kDa Twin-domain Serine Protease Inhibitor (Serpin) Has a Regulatory Function on the Beetle Toll Proteolytic Signaling Cascade. Journal of Biological Chemistry. 286, 35087-35095. https://doi.org/10.1074/jbc.M111.277343.

Kem, W.R., 1997. Alzheimer's drug design based upon an invertebrate toxin (anabaseine) which is a potent nicotinic receptor agonist. Invertebrate Neuroscience. 3, 251-259. https://doi.org/10-.1007/BF02480382

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 227, 680-685. https://doi.org/10.1038/227680a0.

Leal, M. et al., 2012. Bioprospecting of Marine Invertebrates for New Natural Products – A Chemical and Zoogeographical Perspective. Molecules. 17, 9842-9854. https://doi.org/10.3390/molecule-s17089842

Lee, B.R., Takeuchi, M., Kobayashi, Y., 1993. Purification and Characterization of Serine Carboxypeptidases from *Absidia zychae*. Biosci. Biotech. Biochem. 57, 618-622. https://doi.org/10.1271/bbb-.57.618.

Liscovitch, M., 2004. Receptor-Regulated Phospholipases. Encyclopedia of Endocrine Diseases. 4, 167-173.

Look, S.A., Fenical, W., Jacobst, R.S., Clardyt, J., 1986. The pseudopterosins: Anti-inflammatory and analgesic natural products from the sea whip *Pseudopterogorgia elisabethae*. Proc. Nati. Acad. Sci. USA. 83, 6238-6240. https://doi.org/10.1073/pnas.83.17.6238.

Lopes-Ferreira, M. et al., 2011. Structural and biological characterization of Nattectin, a new C-type lectin from the venomous fish *Thalassophryne nattereri.* Biochimie. 93, 971-980. https://doi.org/-10.1016/j.biochi.2011.03.001.

Manaranche, R. et al., 1980. Effect of the Venom of *Glycera convoluta* on the Spontaneous Quantal Release of Transmitter. J. Cell Biology. 85, 446-458. https://doi.org/10.1083/jcb.85.2.446.

Martins, C., Costa, P.M, 2020. Technical Updates to the Comet Assay *in vivo* for Assessing DNA Damage in Zebrafish Embryos from Fresh and Frozen Cell Suspensions. Zebrafish. 17, 220-228. https://doi.org/10.1089/zeb.2020.1857.

Molinski, T.F., Dalisay, D.S., Lievens, S.L., Saludes, J.P., 2009. Drug development from marine natural products. Nat. Ver. Drug. Discov. 8, 69-85. https://doi.org/10.1038/nrd2487.

Murayama, N. et al., 2003. The unusual high molecular mass of *Bothrops* protease A, a trypsin-like serine peptidase from the venom of *Bothrops jararaca*, is due to its high carbohydrate content. Proteins and Proteomics. 1652, 1-6. https://doi.org/10.1016/j.bbapap.2003.08.001.

Nagai, H., 2012. Marine Protein Toxins. Handbook of marine natural products. 1389-1420. https://doi.org/10.1007/978-90-481-3834-0_29.

Ockelmann, K.W., Vahl, O., 1970. On the biology of the polychaete *Glycera alba*, especially its burrowing and feeding. Ophelia. 8, 275-294. https://doi.org/10.1080/00785326.1970.10429564.

Padlan, E.A., Love, W.E., 1968. Structure of the Haemoglobin of the Marine Annelid Worm, *Glycera dibranchiata*, at 5.5 Å Resolution. Nature. 220, 376-378. https://doi.org/10.1038/220376a0.

Price, D.R.G. et al., 2009. A venom metalloproteinase from the parasitic wasp *Eulophus pennicornis* is toxic towards its host, tomato moth (*Lacanobia oleracae*). Insect Molecular Biology. 18, 195-202. https://doi.org/10.1111/j.1365-2583.2009.00864.x.

Rangaraju, S. et al., 2010. Potassium Channel Modulation by a Toxin Domain in Matrix Metalloprotease 23. Journal of Biological Chemistry. 285, 9124-9136. https://doi.org/10.1074/jbc.M109.071266.

Richter, S. et al., 2017. Comparative analyses of glycerotoxin expression unveil a novel structural organization of the bloodworm venom system. BMC Evol. Biol. 17, 1-19. https://doi.org/10.-1186/s12862-017-0904-4

Rodrigo, A.P. et al., 2021. Specific Antiproliferative Properties of Proteinaceous Toxin Secretions from the Marine Annelid *Eulalia* sp. onto Ovarian Cancer Cells. Mar. Drugs. 19, 31. https://doi.org/10.3390/md19010031.

Rodrigo, A.P., Costa, P.M., 2019. The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. Environ. Res. 173, 270-280. https://doi.org/10.1016/j.envres.2019.03.048.

Schmidtko, A. et al., 2010. Ziconotide for treatment of severe chronic pain. New Drug Class. 375, 1569-1577. https://doi.org/10.1016/S0140-6736(10)60354-6.

Shen, L. et al., 2010. Expression of a bee venom phospholipase A2 from *Apis cerana cerana* in the baculovirus-insect cell. Journal of Zhejiang University SCIENCE B. 11, 342-349. https://doi.org/10.1631/jzus.B0900254.

Singer, E. et al., 2013. Neutrophil gelatinase-associated lipocalin: pathophysiology and clinical applications. Acta. Physiol. 207, 663-672. https://doi.org/10.1111/apha.12054.

Stabili, L. et al., 2019. First insight on the mucus of the annelid *Myxicola infundibulum* (Polychaeta, Sabellidae) as a potential prospect for drug discovery. Mar. Drugs. 17, 396-411. https://doi.org/10.3390/md17070396.

van Hemert, M.J., Steensma, H.Y., van Heusden, G.P., 2001. 14-3-3 proteins: key regulators of cell division, signalling and apoptosis. Bioessays. 23, 936-46. https://doi.org/10.1002/bies.1134.

Vogeler, S., Carboni, S., Li, X., Joyce, A., 2021. Phylogenetic analysis of the caspase family in bivalves: implications for programmed cell death, immune response and development. BMC Genomics. 22, 80. https://doi.org/10.1186/s12864-021-07380-0.

von Reumont, B.M. et al., 2014. A Polychaete's Powerful Punch: Venom Gland Transcriptomics of *Glycera* Reveals a Complex Cocktail of Toxin Homologs. Genome Biol. Evol. 6, 2406-2423. https://doi.org/10.1093/gbe/evu190.

Zhang, Y., 2015. Why do we study animal toxins? Zool. Res. 36, 183-222. https://doi.org/10.13918/j.issn.2095-8137.2015.4.183.

A

APPENDIX

Appendix 1 – Identified putative *G. dibranchiata* toxins, with respective species in which the molecular weight has been described and molecular weight in kDa. The potential toxins are classified into five functional categories, according to von Reumont et al., 2014.

Peptide	Molecular weight (kDa)	Category	Species	Reference
Actinoporin	20	Pore-forming	*G. dibranchiata*	von Reumont, et al., (2014)
SNTx-like toxin	150	Pore-forming	*S. horrida*	Nagai, (2012)
Sphingomyelinase	70-75	Pore-forming	Mammalian cells	Liscovitch, (2004)
Glycerotoxin	300	Neurotoxin	*G. tridactyla*	Richter, et al., (2017)
ShKT Domain	4	Neurotoxin	*S. helianthus*	Rangaraju, et al., (2010)
Cystatin	13	Protease inhibitors	All nucleated cells	Chew, et al., (2008)
Kunitz Domain	6	Protease inhibitors	*B. taurus*	Durani and Magliery, (2013)
Lipocalin	25	Protease inhibitors	*H. sapiens*	Singer, et al., (2013)
Serpin	35-50	Protease inhibitors	*T. molitor*	Jiang, et al., (2011)
CAP Domain	~15	CAP Domain	*S. cerevisiae*	Darwiche, et al., (2016)
C-type Lectin	15	Other enzymes	*T. nattereri*	Lopes-Ferreira, et al., (2011)
Chitinase	20-90	Other enzymes	Bacteria, fungi, plants, arthropods, and humans	Hamid, et al., (2013)
Hyaluronidase	53	Other enzymes	*A. natalensis*	Barth, et al., (2018)
Phospholipase	16-18	Other enzymes	*A. cerana*	Shen, et al., (2010)
Peptidase S1	25-40	Other enzymes	*Viperidae*	Murayama, et al., (2013)
Peptidase S10	48	Other enzymes	*A. zychae*	Lee, et al., (1993)
Metalloproteinase M12	20-30	Other enzymes	*E. pennicornis*	Price, et al., (2009)

Appendix 2 – R script for the Dunn's test (nonparametric multiple comparison between data), taking "Organ" as an independent variable and "% DNA in Tail" as a dependent variable for comets, and normalized activity as a dependent variable for caspase 3 activity. The p-value was set at 0.05 for all analyses.

```
##Dunn's test
##Loading required packages
library(dunn.test)

##Import Data
Filename <- "C:\\Users\\Asus UX303UB-C4063T\\Desktop\\Tese Mestrado\\Exceis\\R
(BM+C)\\data.csv"
Data <- read.table(filename,sep=";",header=TRUE)
summary(Data)
rownames(Data) <- Data[,1]
Data <- Data[,-1]

##% DNA in Tail (pDNA) for Comets
##Normalized activity for caspase 3 activity
for(i in 5:ncol(Data)){
  meanData <- aggregate(data=Data,Data[,i]~Organ,mean,na.rm=TRUE)
        meanData <- meanData[c(1,3,2),]
        sdData <- aggregate(data=Data,Data[,i]~Organ,sd,na.rm=TRUE)
  sdData <- sdData[c(1,3,2),]

##Dunn's Test
print(colnames(Data[i]))
print(dunn.test(Data[,i],g=Data$Organ,method="bh"))
}
```

Appendix 3 – R script to convert the list of proteins from LC-MS/MS in *Homo sapiens* genes from UniProt, with specific attributes, such as GO terms.

```
##Glycera proteomics
##Loading required packages
library(UniprotR)
library(biomaRt)
library(GenomicAlignments)

##Import Data
setwd("C:\\Users\\Asus UX303UB-C4063T\\Desktop\\Tese Mestrado\\Exceis\\MS\\R\\")
filename<-"R.csv"
Data<-read.table(filename,sep=",",header=TRUE)

##Retrieve Gene names from UniProt for a list of proteins accessions
##Where: P is protein only found at proboscis
##       S is protein only found at skin
##       C is protein common to proboscis and skin
NameTaxa_P<-GetNamesTaxa(Data$AccN_P)
head(NameTaxa_P)
Gene_P<-NameTaxa_P$Gene.names...primary..

NameTaxa_S<-GetNamesTaxa(Data$AccN_S)
head(NameTaxa_S)
```

```
Gene_S<-NameTaxa_S$Gene.names...primary..

NameTaxa_C<-GetNamesTaxa(Data$AccN_C)
head(NameTaxa_C)
Gene_C<-NameTaxa_C$Gene.names...primary..

##Loading Database
geneSet = useMart(biomart="ensembl", dataset="hsapiens_gene_ensembl")

##Retrieves specified attributes from the BioMart database
GeneList_P<-getBM(
  filters="entrezgene_accession",
  attributes = c(
   "description",
    "entrezgene_accession",
    "entrezgene_id",
    "go_id",
    "name_1006",
    "namespace_1003"
  ),
  values=as.factor(Gene_P),
  uniqueRows=TRUE,
  mart=geneSet
)
GeneList_S<-getBM(
  filters="entrezgene_accession",
  attributes = c(
    "description",
    "entrezgene_accession",
    "entrezgene_id",
    "go_id",
    "name_1006",
    "namespace_1003"
  ),
  values=as.factor(Gene_S),
  uniqueRows=TRUE,
  mart=geneSet
)
GeneList_C<-getBM(
  filters="entrezgene_accession",
  attributes = c(
    "description",
    "entrezgene_accession",
    "entrezgene_id",
    "go_id",
    "name_1006",
    "namespace_1003"
  ),
  values=as.factor(Gene_C),
  uniqueRows=TRUE,
  mart=geneSet
)
```

Appendix 4 – R script to determine the ten more frequent GO Terms for each organ and each sub-ontology. 'X' was replaced with 'P' (for proteins only found at proboscis), 'S' (for proteins only found at skin), and 'C' (for proteins common to both), for each organ analysis, and 'Y' was replaced with 'biological_process', 'cellular_component', and 'molecular_function', for each sub-ontology analysis.

```
##Creating a subset of genes with GO Terms associated with the sub-ontology
SO_X <- subset(GeneList_X, GeneList_X$namespace_1003 == 'Y')

##Counting the number of occurrences of the GO Terms
GoT_X <- unique(SO_X$name_1006)
GO_X <- sapply(GoT_X, function(x) nrow(SO_X[which(GeneList_X$name_1006 %in% x),]))
GO_X <- GO_X[order(-GO_X)]

##Selecting the 10 more frequent GO Terms
##We kept more than 10 when more than one Go term had the same number of occurrences as
place number 10, N is the number of GO terms kept
GOsame <- which(GO_X == GO_X[10])
GOF_X <- GO_X[1:N]

##Remove duplicates
GeneN_X<-unique(SO_X$entrezgene_accession)
```

Appendix 5 – R script to determine enrichment scores between each sub-ontology GO terms of proteins exclusive to the proboscis and proteins exclusive to skin. 'Y' was replaced with 'Biological Process', 'Cellular Component', and 'Molecular Function', for each sub-ontology analysis.

```
##Proboscis vs Skin
##Loading required packages
library(ggplot2)
library(extrafont)
loadfonts(device = "win")

##Merge two data frames by row names
GOF_P.GOF_S<-merge(GOF_P,GOF_S, by = "row.names", all = TRUE)
GOF_P.GOF_S[which(is.na(GOF_P.GOF_S[,2])),2]<-0
GOF_P.GOF_S[which(is.na(GOF_P.GOF_S[,3])),3]<-0

##Fisher's test
Table_P_S<-data.frame(GO_Terms=NA,Only_Proboscis=NA, Only_Skin=NA, odds_ratio=NA,
pvalue=NA)
for(i in 1:nrow(GOF_P.GOF_S)){
  contTable<-rbind(GOF_P.GOF_S[i,2:3],c(length(GeneN_P)-GOF_P.GOF_S[i,2],length(GeneN_S)-
GOF_P.GOF_S[i,3]))
  fisherTest<-fisher.test(contTable)
  Table_P_S[i,]<-c(GOF_P.GOF_S[i,1], GOF_P.GOF_S[i,2], GOF_P.GOF_S[i,3], fisherTest$esti-
mate, fisherTest$p.value)
}
Table_P_S$Only_Proboscis<-as.numeric(Table_P_S$Only_Proboscis)
Table_P_S$Only_Skin<-as.numeric(Table_P_S$Only_Skin)
Table_P_S$odds_ratio<-as.numeric(Table_P_S$odds_ratio)
Table_P_S$pvalue<-as.numeric(Table_P_S$pvalue)
```

```
##Create adjusted p-value (FDRp)
Table_P_S$FDRp<-p.adjust(Table_P_S$pvalue, method = "fdr")
Table_P_S<-Table_P_S[order(Table_P_S$FDRp, -Table_P_S$odds_ratio),]

##To show in which organ is the GO term enriched
Table_P_S$Graph[which(Table_P_S$odds_ratio>1)]<-"Proboscis (Odds-ratio > 1)"
Table_P_S$Graph[which(Table_P_S$odds_ratio<1)]<-"Skin (Odds-ratio < 1)"

##Plot a graphic
FisherGraph_P_S<-data.frame(GOTerm = Table_P_S$GO_Terms,
                            Numbers = -log10(Table_P_S$FDRp),
                            group = Table_P_S$Graph)
ggplot(FisherGraph_P_S, aes(reorder(GOTerm, Numbers), Numbers, fill = group))+
  geom_col()+
  coord_flip()+
  xlab("Y GO Terms")+
  ylab("Enrichment score (-log10(FDRp))")+
  scale_fill_manual(values = c("#ff7f7f", "#7ebfc0"),
                    breaks = c("Proboscis (Odds-ratio > 1)", "Skin (Odds-ratio < 1)"))+
  theme(panel.grid = element_blank(),
        panel.background = element_blank(),
        text = element_text(colour = "black", size = 10, family = "serif"))
dev.print(
  tiff,
  "C:\\Users\\Asus UX303UB-C4063T\\Desktop\\Tese Mestrado\\Exceis\\MS\\R\\P_S_Y.tif",
  width=20,
  height=10,
  unit="cm",
  type="windows",
  res=600)
```

Appendix 6 – Aminoacid sequence matching against subsets retrieved from the NCBInr database. 'ToxinAnnelida_NCBI' is the NCBInr subset with toxin-related proteins and peptides from annelids. 'ToxinMollusca_NCBI' is the NCBInr subset with toxin-related proteins and peptides from molluscs. 'Glycerotoxin_NCBI' is the NCBInr subset specific to glycerotoxins.

```
##R script to convert the aminoacid sequence file to FASTA
##Loading required packages
library(seqinr)

##Import Data
setwd("C:\\Users\\Asus UX303UB-C4063T\\Desktop\\Tese Mestrado\\Exceis\\MS\\R\\BLAST")
file<-"Sequence.csv"
x<-read.table(file,sep=",",header = T)

##Total Aminoacid sequences, from both organs
write.fasta(sequences=as.list(x[,1]),names=row.names(x), file.out="BLAST_T.fasta")
##Total Aminoacid sequences from proboscis
write.fasta(sequences=as.list(x[,2]),names=row.names(x), file.out="BLAST_P.fasta")
##Total Aminoacid sequences from skin
write.fasta(sequences=as.list(x[,3]),names=row.names(x), file.out="BLAST_S.fasta")
```

```
##Determine workspace
cd C:\\Users\\Asus UX303UB-C4063T\\Desktop\\Tese Mestrado\\Exceis\\MS\\R\\BLAST

##Create ToxinAnnelida_NCBI database for Blastp
makeblastdb -in ToxinAnnelida_NCBI.fasta -dbtype prot -out BLAST.fa

##Contrast previous database with the total aminoacid sequences
blastp -query BLAST_T.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTTA_T.csv
##Contrast previous database with the total proboscis aminoacid sequences
blastp -query BLAST_PT.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTTA_PT.csv
##Contrast previous database with the total skin aminoacid sequences
blastp -query BLAST_ST.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTTA_ST.csv

##Create ToxinMollusca_NCBI database for Blastp
makeblastdb -in ToxinMollusca_NCBI.fasta -dbtype prot -out BLAST.fa

##Contrast previous database with the total aminoacid sequences
blastp -query BLAST_T.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTTM_T.csv
##Contrast previous database with the total proboscis aminoacid sequences
blastp -query BLAST_PT.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTTM_PT.csv
##Contrast previous database with the total skin aminoacid sequences
blastp -query BLAST_ST.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLAST2_ST.csv

##Create Glycerotoxin_NCBI database for Blastp
makeblastdb -in Glycerotoxin_NCBI.fasta -dbtype prot -out BLAST.fa

##Contrast previous database with the total aminoacid sequences
blastp -query BLAST_T.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 > BLASTG_T.csv
##Contrast previous database with the total proboscis aminoacid sequences
blastp -query BLAST_PT.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTG_PT.csv
##Contrast previous database with the total skin aminoacid sequences
blastp -query BLAST_ST.fasta -db BLAST.fa -max_target_seqs 1 -outfmt "6 delim=; qseqid sse-
qid ssequid pident nident length evalue stitle" -evalue 1e-1 -num_threads 20 >
BLASTG_ST.csv
```