NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTMENT OF LIFE SCIENCES

CATARINA AUGUSTA FLORÊNCIO FAUSTINO

BSc in Cell and Molecular Biology

MARINE BIOPROSPECTING FOR DRUG DISCOVERY: EXPLORING NOVEL TOXIN-DERIVED BIOACTIVES FROM VENOMOUS ANNELIDS

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon
September, 2023

DEPARTMENT OF LIFE SCIENCES

MARINE BIOPROSPECTING FOR DRUG DISCOVERY: EXPLORING NOVEL TOXIN-DERIVED BIOACTIVES FROM VENOMOUS ANNELIDS

CATARINA AUGUSTA FLORÊNCIO FAUSTINO

BSc in Cell and Molecular Biology

Adviser: Pedro Manuel Brôa Costa,
Auxiliary Professor, NOVA School of Science and Technology

Co-advisers: Ana Margarida Gonçalves Carvalho Dias,
PhD, Junior Researcher, NOVA School of Science and Technology

Examination Committee:

Chair: Maria Alexandra Núncio de Carvalho Ramos Fernandes,
Associate Professor with habilitation, NOVA School of Science and Technology

Rapporteurs: Maria Filomena Andrade de Freitas,
Auxiliary Professor, NOVA School of Science and Technology

Adviser: Pedro Manuel Brôa Costa,
Auxiliary Professor, NOVA School of Science and Technology

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon
September, 2023

Marine bioprospecting for drug discovery: Exploring novel toxin-derived bioactives from venomous annelids

ACKNOWLEDGMENTS

First and foremost, I wish to express my profound gratitude to my supervisor, Professor Pedro M. Costa. Your warm welcome into this project, the opportunity to work with extraordinary teams, and the knowledge you've nurtured in me are immeasurable gifts. Thank you for your guidance, availability, patience, optimism, and encouragement throughout this past year.

A heartfelt thank you goes to my co-supervisor, Doctor Ana Margarida Dias, a constant companion in this project's development. I extend my appreciation for your unwavering attention, boundless availability, and unwavering support. Under your mentorship, I've grown both as a scientist and as an individual. Your encouragement, support, and the opportunities you've provided have been truly transformative and I will never forget them.

I must also acknowledge Professor Ana Cecília Roque, who welcomed me warmly into her team while not directly related to this project and was always available to help and guide me in whatever way possible. Furthermore, together with my supervisors, she granted me the privilege of participating in my first international scientific conference.

I am immensely grateful to both teams I've had the privilege of working with, from the Seatox Lab and the Biomolecular Engineering Lab. Thank you to all members for your warm reception, easy inclusion in the groups, and for providing such a pleasant environment of collaborative spirit in your teams. Among these teams, I wish to offer special thanks to Cátia Soares, Inês Padrão, Inês Cabral, Cátia Gonçalves, and Carla Martins. Their unwavering support, help, patience, and availability for any clarification have been instrumental.

I extend heartfelt thanks to my friends and family, who have been my steadfast companions not only this past year but throughout my journey. Your words of encouragement and unwavering support have been my pillars through both the best and the most challenging moments. Special gratitude to Maggie, Carol, and Serra for reminding me that distance can't diminish our friendship and for your tireless support. A profound thank you to Artur for consistently being there when I needed it most for your unwavering support, patience, motivation, and affection. Finally, a profound thank you to my parents. Dad, your support in providing me this opportunity and your unwavering confidence in my pursuit of dreams mean the world to me. Mom, your constant encouragement, your unshakable belief in me even when I doubted myself, and the incredible strength and courage you've demonstrated over the past few years have been my greatest inspiration.

Last but not least, I extend my gratitude to the financing entities and to NOVA School of Science and Technology (FCT NOVA), particularly UCIBIO, and the Departments of Chemistry and Life Sciences for this incredible opportunity. Fundo Azul co-financed project MARVEN (FA_05_2017_007). This work is also financed by national funds from the Portuguese Foundation for Science and Technology (FCT), in the scope of Proteios project (PTDC/CTM-CTM/3389/2021), project UIDP/04378/2020 and UIDB/04378/2020 of UCIBIO and project LA/P/0140/2020 of Associate Laboratory i4HB.

"Science is the key to unlocking the mysteries of the universe and harnessing its power for the benefit of humanity" (Vera Rubin).

ABSTRACT

The vast marine biodiversity makes the ocean an almost inexhaustible source of bioactives with significant biotechnological potential, emphasizing on marine invertebrates. Toxins, in particular, bear high value as they tend to evolve to interact with specific molecular targets as a means of protection, defense, or predation. For such reason, toxins, the main bioactive components of venoms and similar secretions, are nowadays revealing the potential for the development of new drugs. The current research is a contribution to investigate the biotechnological potential of a novel cysteine-rich secretory protein derived from the secretions of a marine Polychaeta, *Hediste diversicolor*. Heterologous expression of the recombinant protein was successful. The protein was preferentially expressed in Terrific Broth medium using E*scherichia coli* BL21(DE3) strain, induced with 1mM IPTG and incubated overnight at 18ºC, and was identified in both soluble and insoluble fractions. After optimizing the conditions for heterologous expression, the protein was characterized for its molecular weight (26kDa-31kDa) by SDS-PAGE and Western blot. The protein isoelectric point (7.7) was determined by Isoelectric Focusing, and it was similar to the theoretical value. An extract enriched in recombinant protein was obtained after Immobilized Metal Affinity Chromatography and used in *in vitro* assays to assess the protein's toxicity and potential to cause molecular damage. Microscopy analysis confirmed the quality of cells used in the specially developed cytoassays with primary mussel gill cells. Finally, the protein was further purified using Size Exclusion Chromatography, attaining ≈86% purity. The tested protein concentrations did not cause significant cytotoxicity or molecular damage to cells, which may have significant implications for further research, as the protein is believed to have an important antimicrobial role, albeit still to be ascertained. The successful cloning and expression lay out an important precedent for the testing and exploitation of novel marine proteinaceous bioactives for the purpose of drug discovery.

Keywords: marine natural products; recombinant protein; *Escherichia coli*; cysteine-rich secretory proteins; toxicity; *in vitro* assays

RESUMO

A vasta biodiversidade marinha faz do oceano uma fonte quase inesgotável de compostos bioativos com significativo potencial biotecnológico, com ênfase nos invertebrados marinhos. Particularmente, as toxinas apresentam alto valor, pois tendem a evoluir para interagir com alvos moleculares específicos como meio de proteção, defesa ou predação. Assim, as toxinas, principais componentes bioativos de venenos e secreções similares, revelam potencial para o desenvolvimento de novos fármacos. A presente tese contribuiu para a investigação do potencial biotecnológico de uma nova proteína rica em cisteínas derivada das secreções de uma Polychaeta marinha, *Hediste diversicolor*. A expressão heteróloga da proteína recombinante foi bem-sucedida. A proteína foi preferencialmente expressa em Terrific Broth utilizando a estirpe de *Escherichia coli* BL21(DE3), induzida com 1mM de IPTG e incubada a 18ºC durante a noite, tendo sido identificada na fração solúvel e insolúvel. Após otimização das condições de expressão heteróloga, a proteína foi caracterizada quanto ao seu peso molecular (26kDa-31kDa) por SDS-PAGE e Western blot. O ponto isoelétrico da proteína (7.7) foi determinado por Focagem Isoelétrica, sendo próximo do valor teórico. Um extrato enriquecido com a proteína recombinante foi obtido após Cromatografia de Afinidade com Metal Imobilizado e utilizado em ensaios *in vitro* para avaliar a toxicidade da proteína e o seu potencial para causar danos moleculares. A análise microscópica confirmou a qualidade das células utilizadas nos citoensaios especialmente desenvolvidos com células primárias de guelras de mexilhão. A proteína foi ainda purificada usando Cromatografia de Exclusão de Tamanho, atingindo ≈86% de pureza. As concentrações proteicas testadas não causaram citotoxicidade significativa ou danos moleculares, o que pode ter implicações significativas para futuros projetos, pois acredita-se que a proteína tenha um importante papel antimicrobiano, embora ainda a ser apurado. A clonagem e expressão bem-sucedidas estabelecem um precedente importante para o teste e exploração de novos bioativos proteicos marinhos para fins de descoberta de fármacos.

Palavras chave: Produtos naturais marinhos; proteína recombinante; *Escherichia coli*; proteínas secretadas ricas em cisteínas; toxicidade; ensaios *in vitro*

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	After induction.
APS	Ammonium persulphate.
BCA	Bichinchoninic acid.
BI	Before induction.
BSA	Bovine Serum Albumin.
CAP	Sequence similar between cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1.
CHAPS	3-[(3-Cholamidopropyl)-dimethyl ammonio]-1-propane sulphonate.
CRISP	Cysteine-rich secretory protein.
CV	Column volume.
DEPC	Diethyl pyrocarbonate.
DMSO	Dimethyl sulfoxide.
DNA	Deoxyribonucleic acid.
DTT	Dithiothreitol.
E4	4^{th} elution obtained during immobilized metal affinity chromatography.
EDTA	Ethylenediamine tetraacetic acid.
ExtHdiv_1	Protein extract containing Hdiv_1.
ExtP	Protein extract containing all soluble proteins intrinsic to BL21(DE3) when plasmid codifying for Hdiv_1 is absent.
GFP	Green fluorescent protein.
Hdiv_1	Recombinant protein based on the cysteine-rich secretory protein from *Hediste diversicolor*.
HRP	Horseradish peroxidase.
HSD	Honestly significance difference.
IC50	Half-maximal inhibitory concentration.
IEF	Isoelectric focusing.
IMAC	Immobilized metal affinity chromatography.

IPG	Immobilized pH gradient.
IPTG	Isopropyl β-D-1-thiogalactopyranoside.
ISF	Insoluble fraction.
LB	Luria Broth.
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry.
LMPA	Low melting point agarose.
LMW	Low molecular weight.
OD_{600}	Optical density measured at 600 nm.
ON	Overnight.
PBS	Phosphate-buffered saline.
PFA	Paraformaldehyde.
pI	Isoelectric point.
SDS	Sodium dodecyl sulphate.
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis.
SEC	Size exclusion chromatography.
SF	Soluble fraction.
SFII	Soluble fraction obtained after a second cell lysis.
SOC	Super optimal broth with catabolite repression
TAE	Tris-acetate-EDTA.
TB	Terrific Broth.
TBS	Tris-buffered saline.
TEMED	N,N,N′,N′-tetramethylethylenediamine.
Vt	Volumes of bacterial culture collected at different times for the time course analysis.

1

INTRODUCTION

1.1 Marine Bioprospecting

The interest in marine bioprospecting primarily stems from the vast and ancient biodiversity and subsequent biochemical diversity of the ocean (Sala & Knowlton, 2006), which boosts it as an almost inexhaustible source of novel bioactives. However, the full extent of marine biodiversity remains elusive. As mentioned by Sala and Knowlton (2006), many marine species still need identification, mainly due to insufficient efforts and loss of experienced taxonomists, contributing to the hindrance of marine bioprospecting. Notwithstanding, the increasing demand for novel and sustainable bioactive compounds for various applications, from biomaterials to pharmaceuticals, is steering bioprospecting toward the oceans as an unparalleled source of unique marine bioproducts with substantial potential for societal and commercial benefits (Burgess, 2012). These motivations led to the incorporation of Blue Biotechnology by the European Union as a pillar for the Blue Growth Agenda, aiming to promote a purposeful and sustainable discovery, exploitation, and application of marine resources (see, for instance, Lillebø et al., 2017). Nevertheless, despite this surge in interest and the immense diversity of marine compounds, the catalog of documented marine natural products still significantly lags behind its terrestrial counterparts (Ntie-Kang & Syozil, 2020). Figure 1.1 provides an overview of how marine biomolecules and whole-organisms offer a wide array of potential biotechnological applications, spanning eco-friendly pesticides, cosmetics, biofuels, detergents, and even food products (see Freitas et al., 2012 and Martins et al., 2014, for further details). Interestingly, a few marine-derived bioproducts already play a pivotal role in research and *in vitro* diagnostics (Burgess, 2012). Arguably, the most noteworthy example is the green fluorescent protein (GFP). Furthermore, marine biomolecules have made relevant contributions to the development of novel biomaterials and drugs (see Khrunyk et al., 2020 and Mayer et al., 2010 for reviews). Bioactive proteins and peptides from marine animal venoms, including those from Polychaeta annelids, seem particularly interesting for this last application due to their specificity but especially to their diversity (Molinski et al., 2009; Rodrigo & Costa, 2019).

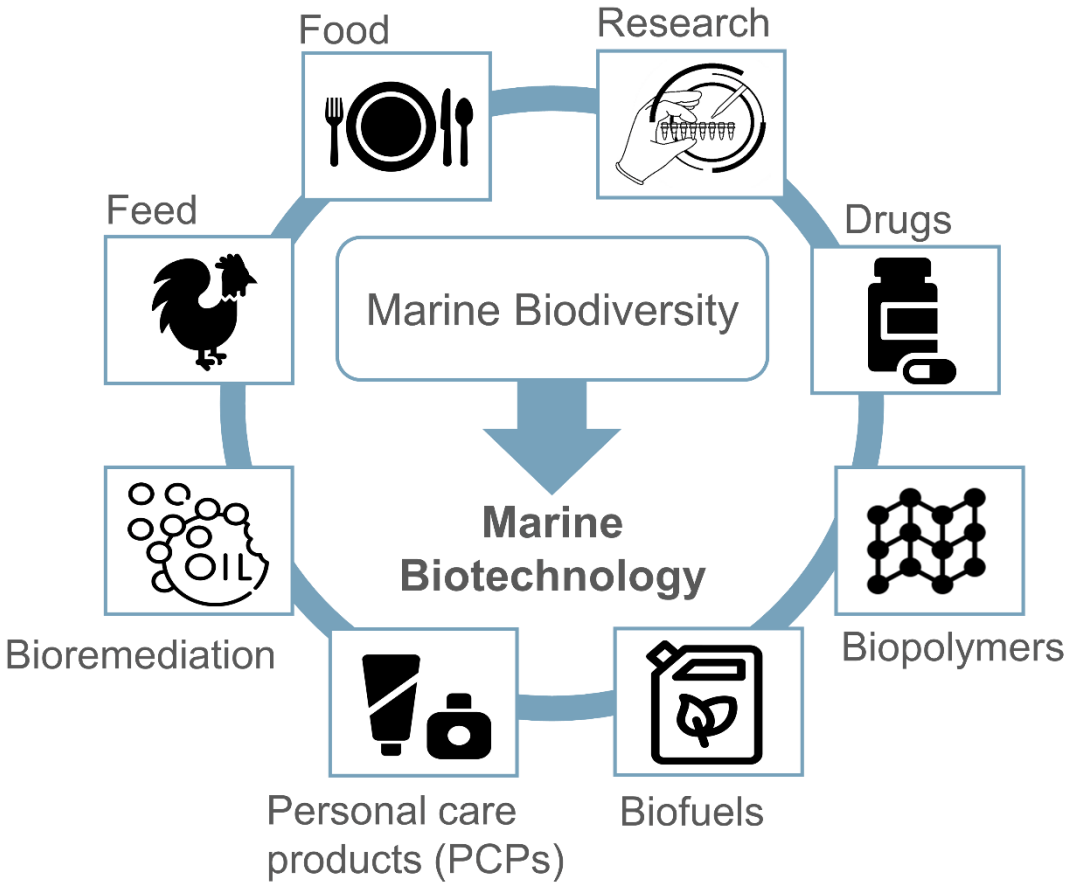

Figure 1.1 — Marine biodiversity applications. The wide marine biodiversity offers us a massive variety of chemical compounds with useful and profitable features. This makes the oceans an almost inexhaustible source of novel bioproducts, like bioactives, used in several areas. Adapted from Freitas et al. (2012), Martins et al. (2014), Khrunyk et al. (2020) and Mayer et al. (2010).

1.2 Marine Toxins

Toxins are highly prized bioactives because they have evolved to interact with specific molecular targets and physiological processes in a target organism (Wright, 2019). Proteinaceous toxins are the main components of animal venoms and poisons, which are complex cocktails that also include salts, enzymes and enzyme inhibitors, to quote a few (Fox & Serrano, 2008; Inceoglu et al., 2003). Quite naturally, among the vastness of marine biodiversity, there are many venomous animals, both vertebrate and invertebrate. Their venoms encompass a repertoire of bioactive molecules such as neurotoxins, permeabilizing agents, anticoagulants, and antimicrobial peptides (extensively revised in Fry et al., 2009). Some of these toxins exhibit a potential to act not only in prey but also to recognize mammalian receptors, making them valuable prototypes for drug development (see Molinski et al., 2009, for examples). In fact, the diversity of bioactive compounds found in natural environments, such as the ocean, made marine bioprospecting a compelling alternative to the laborious, costly, and time-consuming process of designing synthetic compounds (Rodrigo & Costa, 2019; Collins et al., 2014). Focusing on

protein-based toxins, they have several advantages over metabolites or derivatives, especially for drug development. First, proteins offer higher specificity in binding to a target and a more complex set of functions than smaller molecules, which can be justified by the higher number and stronger interactions that a protein can establish (see Dias & Roque, 2017 for a review on affinity reagents). Moreover, protein therapeutics are unlikely to interfere with other biological processes due to their high specificity. Additionally, as the human body naturally produces proteins, these molecules are typically well-tolerated by the organism. These advantages result in enhanced efficiency for drugs based on proteins, reducing the required dosage, minimizing waste, and limiting side effects. In addition, it is worth noticing that proteins can be produced nowadays by recombinant DNA technology, which could be an alternative to chemical synthesis (see Leader et al., 2008 for further details on the advantages of protein therapeutics). All these factors evidence that drugs based on proteinaceous toxins, have potential to offer superior safety, cost-effectiveness, and environmental friendliness compared to small molecules or synthetic compounds. Consequently, the interest in marine proteins with biomedical potential, such as neurotoxins, has increased. Specifically, neurotoxins hold promise as the foundation for novel painkillers and anesthetics due to their interference with the neuromuscular system. Conotoxins, a subgroup of marine neurotoxins produced by cone snails, exemplify this biomedical potential. These small proteins and peptides are mostly composed of 10 to 40 amino acids and are rich in cysteine residues (mostly 4 to 6). Consequently, conotoxins bear specialized secondary structures resulting from the formation of disulfide bonds, which are paramount for the stability and function of the proteins. These features confer resistance to proteases and are responsible for the specific and potent interactions of conotoxins with molecular targets (including ion channels, G protein-coupled receptors, transporters, and enzymes, to quote a few). In addition, their stability and compact size made possible functional expression using bacteria (see Jin et al., 2019 for further details on conotoxins). A notable success in this domain is the approved painkiller Ziconotide, derived from ω-conotoxin MVIIA present in *Conus magus* venom. This conotoxin is composed of 25 amino acids, six of which are cysteines, and can block calcium channels essential for the transduction of a neuronal signal (Molinski et al., 2009). Although the commercial version of Ziconotide is produced synthetically, the heterologous expression of the functional peptide in *Escherichia coli* has already been reported by Xia et al. (2006). Beyond Ziconotide, drugs based on non-protein marine toxins were already approved with applications in cancer and viral disease treatment, while other marine bioactives are currently in clinical trials (Mayer et al., 2010). However, the commercial success of drugs based on marine bioproducts remains limited, partly due to challenges in purifying and identifying active compounds from complex mixtures, often in small quantities (Wright, 2019). Nevertheless, the primary hurdle arises from difficulties in harnessing the vast marine biodiversity and chemical diversity for practical applications (Burgess, 2012). State-of-the-art methodologies, such as ‘omics’ and associated computational approaches, and public access databases have been assisting the discovery, characterization, and even prediction of bioactive effects in mammalians, as the case for instance of Moutinho Cabral et al., (2002), within toxins from *Glycera alba* plus *Hediste diversicolor* and their potential interactions with the human reactome.

1.3 The Polychaeta case study

The Polychaeta annelids constitute a taxonomic group of aquatic invertebrates little studied by biotechnologists, albeit representing one of the most diverse and widespread group of aquatic animals. These annelids' wide distribution has driven specific macroscopic and microscopic adaptations. Therefore, the remarkable diversity and geographical distribution of polychaetes indicate that these organisms have potential to produce bioactives compounds, which may hold biotechnological interest. Indeed, these organisms occupy a vast array of ecological niches, ranging from deep-sea vents to freshwater rivers and lakes, being virtually present in all types of marine sediments (Glasby & Timm, 2008). The distribution of these annelids is influenced by multiple environmental factors, including substrate type, sediment anoxia, pH levels, organic content, plus depth, temperature, and salinity (Hutchings, 1998). One notable example of their adaptations to the environment is the epidermis, which evolved to provide essential protection and sensory capabilities, finely tuned to match the species' ecological niches (Hausen, 2005). Furthermore, the proboscis (an eversible pharynx possessed by many species) shows diverse modifications that suit multiple feeding and sensing strategies. In addition to these adaptations, polychaetes can produce secretions containing specific bioactives (including substances such as mucins, toxins, and digestive enzymes) to compensate for a comparatively primitive immune system or to be used as defense or predation mechanisms (see Rodrigo & Costa, 2019 for a review). These bioactives may hold significant potential for the development of new drugs. As mentioned by the same authors, some of these bioactives were already reported as compounds of interest for the development of painkillers, anesthetics, anti-cancer drugs, antibiotics, and others. For instance, several neurotoxins found in fireworms (Amphinomidae) are able to induce muscle relaxation and paralysis by blocking the neuromuscular transmission and potassium channels (see Von Reumont et al., 2014 and Rodrigo & Costa, 2019 for examples).

The discoveries presented above have heightened interest in investigating bioactives produced by polychaetes. This growing interest has drawn attention to abundant Polychaete groups with intriguing predatory and defensive traits present for instance on the Portuguese coast. One of these species is *Eulalia viridis,* a green worm that has been receiving particular attention recently. This species, found in the rocky intertidal zone, possesses subtle microanatomical features that serve as an adaptation to its environment and suit its predatory behavior. Among these adaptative traits, the skin of *E. viridis* bears specialized cells to secrete mucus, cuticle, and toxins, even in absence of specialized glands (for further details on *E. viridis,* see Rodrigo et al., 2018). The same authors also disclosed that some proteins present in the mucous secretion appear to specifically induce cell-cycle arrest by a mechanism specific to A2780 ovarian adenocarcinoma cells. These findings support the potential of *E. viridis* bioactives as promising candidates for novel anti-cancer treatments (Rodrigo et al., 2021a). The *E. viridis* case study showed that annelids from temperate waters, namely the Portuguese coastline, can be an important source of novel bioactives. Therefore, the biotechnological potential of Glycerids, commonly known as bloodworms, has garnered attention. These marine worms are predators of small marine invertebrates

and use their proboscis to entrap prey. The proboscis is equipped with four robust jaws connected to venom glands, each featuring channels and pores for venom release (Moses et al., 2006). Recent transcriptomic analysis suggests that the venom secreted by these annelids comprises a cocktail of marine toxins, including pore-forming toxins, neurotoxins, protease inhibitors, CAP domain toxins, and various enzymes (Von Reumont et al., 2014). The combination of these toxins may provoke progressive paralysis of small invertebrates (the typical prey of these annelids), along with cardiac arrest, convulsions, and eventually death (Michel & Keil, 1975; Bon et al., 1985). The role played by these toxins, the anatomy of the venom apparatus and field observations of feeding habits confirmed the Glycerid's predatory habits (Von Reumont et al., 2014). Within this Polychaeta group, some toxins with biotechnological potential have been identified, such as glycerotoxin, a proteinaceous neurotoxin known to act as a calcium channel agonist (Michel & Keil, 1975). In addition, the effects of glycerotoxin appear to be fully reversible, enhancing its value for biotechnology (Meunier, 2002). Conversely, *Hediste* (Nereididae), another Polychaeta group relatively abundant on the Portuguese coast, displays a distinct behavior compared with Glycerids. In particular, *Hediste diversicolor*, although possessing an eversible proboscis adorned with two jaws, its primary function is more related to grasping food (Scaps, 2002; Bryan & Gibbs, 1979). Scaps' (2020) review also mentions an alternative feeding technique involving the use of mucous secretions deposited on the worm's body surface to collect food particles. In addition, it was recently suggested that nereidids may be able to secrete unknown bioactives through their skin as a defense mechanism (Gonçalves & Costa, 2020). Another defense mechanism specific to *Hediste diversicolor* is the secretion of an identified antimicrobial peptide (hedistin) by the natural killer-like cells upon immune challenge (Tasiemski et al., 2007). As here elucidated, these two Polychaete groups, Glycerids and Nereididae, hold considerable biotechnological interest owing to their capacity to produce bioactives endowed with distinctive and useful properties. Understanding the full potential of these bioactives is, therefore, a critical tool to make these abundant animals in our coast and estuaries an important source of novel compounds for biotechnological applications.

The research performed by Rodrigo et al. (2018, 2021a) allowed the establishment of a research pipeline to explore new marine toxins. First, the focus is directed toward species exhibiting characteristics suggestive of toxin production, either in their anatomical features or behavioral traits. Upon identifying a promising candidate within a database, specimens of the species are collected from the North Atlantic Ocean. Subsequently, their proteomes and transcriptomes are analyzed *in silico* with the purpose to identify compounds of interest. The identified compounds are then produced and characterized, and finally, *in vitro* and *in vivo* experiments are conducted to assess the toxicity and bioactivity of the products. Favorable outcomes in these assays pave the way for the subsequent phase: the development of a drug. This drug would be subsequently subjected to clinical trials, as depicted in Figure 1.2.

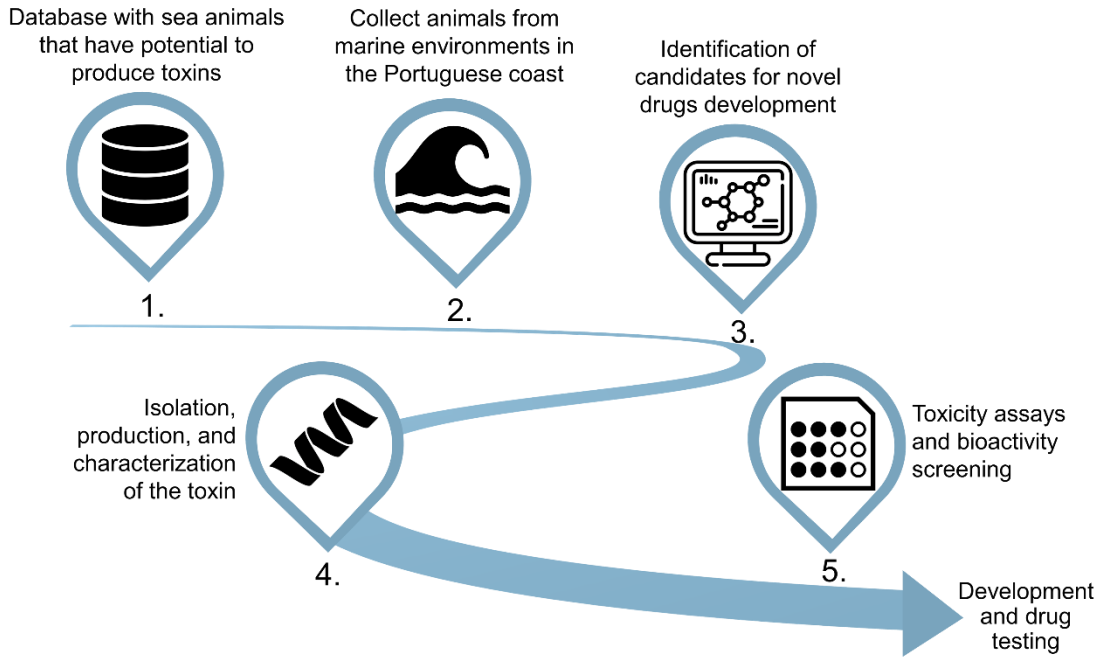

Figure 1.2 — Marine bioprospecting for drug discovery. Proposed steps to obtain a drug based on a marine toxin. Adapted from Haque and Ratemi, (2017) and Liang et al. (2019).

In line with the research strategy presented above, Moutinho Cabral et al. (2022) conducted a comparative transcriptomic analysis of secretory organs from the Polychaetes species *Glycera alba* and *Hediste diversicolor*. The same study revealed the presence of proteinaceous toxins with potential to interact with the human proteome. Based on these findings, certain proteins of particular interest, presumed to have biomedical potential, have been selected for further investigation and are detailed in Table 1.1. One of the proteins that garnered attention is a cysteine-rich venom protein, found to be overexpressed in the proboscis of *Glycera alba*. Its sequence exhibits homology to known neurotoxins from various marine invertebrates. These neurotoxins are characterized by CAP domains and their potential to inhibit potassium channels in neurons (Modica et al., 2015). While this protein demonstrated potential as a neurotoxin, it did not yield interactions in the interactome-directed analysis due to the absence of human homologs on the platform. Nonetheless, Moutinho Cabral et al. (2022) proposed that this cysteine-rich venom protein may interact with the human neuromuscular system. In contrast to *G. alba*'s neurotoxin candidate and although others cysteine-rich venom proteins appear to be related to ion channel-blocking activity (see for instance, Jin et al., 2019; Xia et al., 2006; Yamazaki et al., 2002; Rodrigo et al., 2021b), the cysteine-rich secretory protein released by *Hediste diversicolor* seems to be associated with defensive and immune functions. According to Moutinho Cabral et al. (2022), this protein resembles pathogenesis-related and glioma pathogenesis-related like proteins. Pathogenesis-related proteins are produced by plants and play a pivotal role in plant defense mechanisms, with some appearing to act as antimicrobial agents targeting molecules present in the wall of bacterial or fungal cells (van Loon, 1985; Dos Santos & Franco, 2023). Glioma pathogenesis-related like proteins, found in certain cancer types, appear to exhibit dual roles in cancer development. For instance, the pathogenesis-related like protein 1 has been identified as an oncoprotein in some cancers and as a tumor suppressor

in others, holding potential in the treatment of the last group (Wang et al., 2022). This suggests that the cysteine-rich secretory protein from *Hediste diversicolor* may possess neurotoxic and antimicrobial potential. Furthermore, the thyrostimulin beta-5 subunit, also identified in *Hediste diversicolor*, appears to be up-regulated in the proboscis' glands. This protein has potential to influence thyroid cell metabolism due to its homologous ability to interact with the G protein-coupled receptor of thyrotropin, the master regulator of thyroid gland growth and function (Nakabayashi et al., 2002). In addition to its role in thyroid regulation and according to the results of Moutinho Cabral et al. (2022) this protein may possess neuronal modulator and hormonal activities, given its non-covalent binding to the alpha-2 subunit of thyrostimulin, a glycoprotein hormone also reported in cone snails and gastropods (Heyland et al., 2012; Robinson et al., 2017). Last but not least, the serine protease inhibitor Kazal-type from *G. alba* was also identified in the venoms of other animals (including other Polychaetas' species), where it was suggested that it might have antimicrobial or anticoagulation properties (Von Reumont et al., 2014; Rodrigo et al., 2021b; Kim et al., 2013). These four proteins exhibit considerable biotechnological potential, however, not all of them will be the focus of this thesis.

Table 1.1 — Identified proteins with biotechnological potential from *Glycera alba* and *Hediste diversicolor* (proteins' properties were retrieved from Expasy Protoparam program).

Protein Name	Protein Data Bank ID	Amino Acids Number	Cysteines Number	Molecular Weight (g/mol)
***Glycera alba* cysteine-rich venom protein**	UTH78534.1	301	22	32167.59
***Hediste diversicolor* cysteine-rich secretory protein**	UTH78536.1	195	10	21135.70
***Hediste diversicolor* thyrostimulin beta-5 subunit**	UTH78537.1	151	10	16647.19
***Glycera alba* serine protease inhibitor Kazal-type**	UTH78535.1	396	40	42588.13

1.4 Objectives

This project is a contribution to explore the biotechnological potential of proteins isolated from marine Polychaeta. The main goal was to develop and test a method to produce a cysteine-rich secretory protein (CRISP), a bioactive protein secreted by the muscular proboscis region of *Hediste diversiocolor,* based on DNA recombinant technology using the *Escherichia coli* model. This protein was selected among four proteins of interest reported by Moutinho Cabral et al. (2022) due to its potential bioreactivity and lower number of cysteines, as previously shown in Table 1.1. Specifically, this thesis aimed at i) selecting and optimizing the protein expression strategy using *E. coli* as a biotechnological model; ii) developing a workflow for protein purification and biophysical characterization; and iii) preliminarily verifying its potential toxicity and bioactivity.

To achieve these goals, the research project was divided into three main tasks. The initial phase involved an analysis of the proteinaceous sequences and selection of one protein to attempt for the first-time heterologous expression in *E. coli* under different conditions. After obtaining a soluble recombinant protein, the subsequent task centered on its purification and characterization. The protein purification process encompassed the use of immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC). The characterization efforts included the determination of the molecular weight and isoelectric point of the protein through SDS-PAGE, Western blot and isoelectric focusing (IEF), plus the identification of the recombinant protein through liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). To evaluate the protein's toxicity and bioactivity, a series of cell-based assays were conducted, namely the neutral red and comet assays, which take cytotoxicity and molecular damage as main endpoints, respectively. These assays utilized primary mussel gill cells as a toxicological model for expediteness and potential sensitivity. Additionally, microscopic analysis was also performed to verify the presence of cytotoxicity signs in the cells, specifically to confirm if the cells and their organelles (e.g., nucleus) remained apparently viable after exposure to the produced protein. An overview of these sequential steps is provided in Figure 1.3.

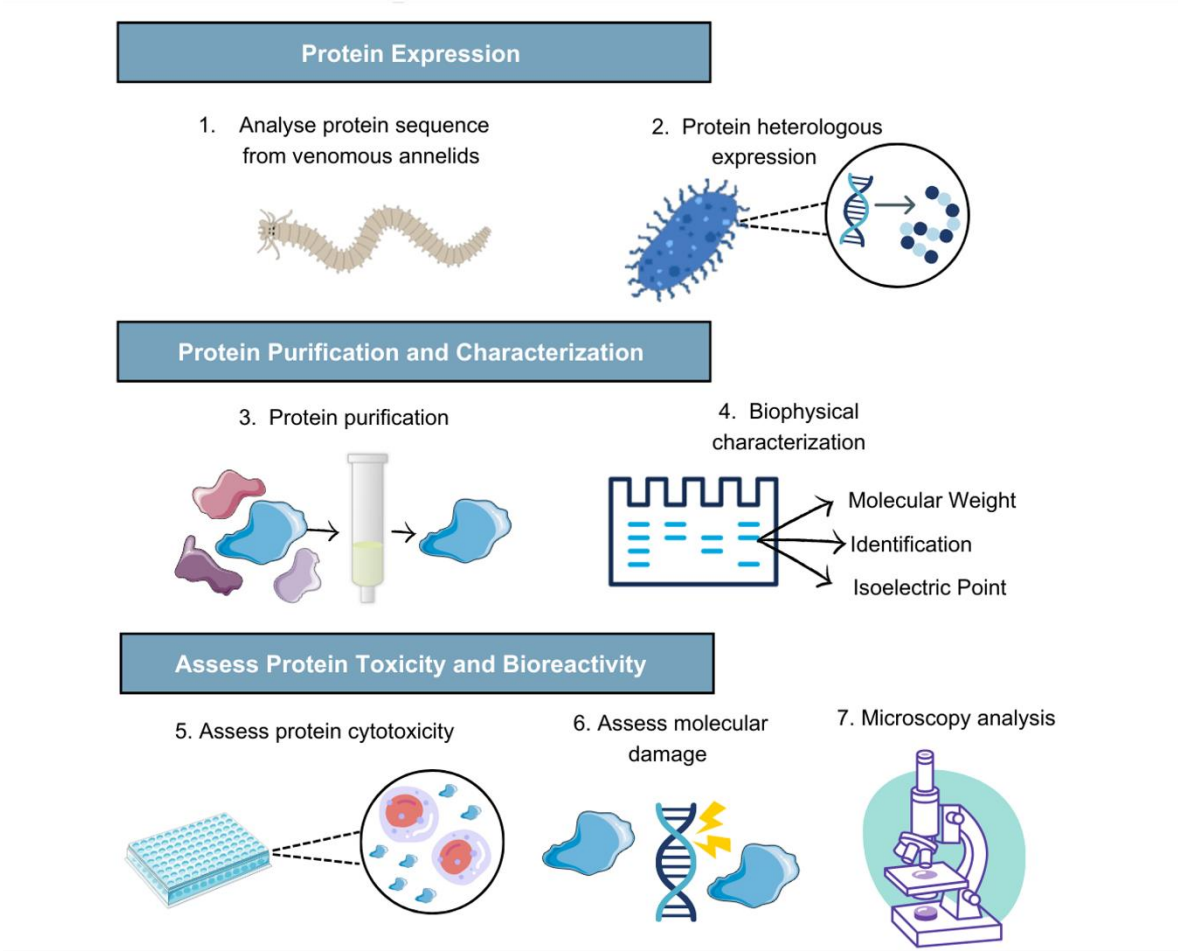

Figure 1.3 — Workflow of the research plan. Design created on Canva.com**.**

2

MATERIALS AND METHODS

2.1 Chemicals

The following chemicals were used to prepare buffer solutions and solvents: tris (CAS number 77-86-1), glycine (CAS number 56-40-6), and sodium dodecyl sulfate micropellet (SDS, CAS number 151-21-3), from NZYTech. potassium chloride (KCl, CAS number 7447-40-7), 2-mercaptoethanol (CAS number 60-24-2), sodium phosphate dibasic heptahydrate ($Na_2HPO_4 \cdot 7H_2O$, CAS number 7782-85-6), potassium phosphate dibasic (K_2HPO_4, CAS number 7758-11-4), sodium bicarbonate ($NaHCO_3$, CAS number 144-55-8), D(+)-glucose anhydrous (CAS number 50-99-7), DMSO (dimethyl sulfoxide, CAS number 67-68-5) and urea (CAS number 57-13-6) from Sigma-Aldrich. Sodium chloride (NaCl, CAS number 7647-14-5), acetic acid (CAS number 64-19-7) and dithiothreitol (DTT, CAS number 3483-12-3) from Fisher Scientific, and imidazole (CAS number 288-32-4) and glycerol (CAS number 56-81-5) from Alfa Aesar. From VWR were purchased sodium hydroxide (NaOH, CAS number 1310-73-2), tween-20 (CAS number 9005-64-5), and potassium phosphate monobasic (KH_2PO_4, CAS number 7778-77-0). Other compounds were purchased from different suppliers: bromophenol blue (CAS number 115-39-9) and ethanol absolute (≥ 99.5 %, CAS number 64-17-5) were from PanReacAppliChem, CHAPS (3-[(3-Cholamidopropyl)-dimethyl ammonio]-1-propane sulphonate, CAS number 75621-03-3) and EDTA disodium salt solution (ethylenediamine tetraacetic acid disodium, CAS number 6381-92-6) from Roth, HCl (hydrochloric acid 37 %, CAS number 7647-01-0) from Chemlab, bio-lyte (CAS number 163-2094) and mineral oil (CAS number 8042-47-5) from Bio-Rad and triton® X-100 (CAS number 9036-19-5) from Promega. The 12.5 % SDS-PAGE gels for electrophoreses were produced in-house with 30 % acrylamide: bisacrylamide 37.5:1 and 10 % sodium dodecyl sulfate (SDS) solution from Bio-Rad, ammonium persulphate (APS, ≥ 98%, CAS number 7727-54-0) and N,N,N′,N′-tetramethylethylenediamine (TEMED, CAS number 110-18-9) from NZYTech. The Coomassie blue staining solution was prepared with Coomassie brilliant blue R-250 (CAS number 6104-59-2, Sigma-Aldrich), acetic acid glacial (CAS number 64-19-7, VWR), and methanol (CAS number 67-56-1, Honeywell). Alternatively, the gels were stained with Silver Stain Plus kit from Biorad.

2.2 General Equipment

The bacterial growth and protein expression were carried out in incubator KS 4000 ic (IKA) and the optical density of bacteria cultures was measured in evolution 201/220 ultraviolet-visible (UV-Vis) spectrophotometer (Thermo Scientific). All the centrifugations made during the lysis and purification processes used the centrifuge multifuge X3 (Thermo Scientific) or the minicentrifuge Micro Star 12 (VWR). The centrifugations made during the cell-based assays used the Multi-Purpose, High-Speed Centrifuge 1248R (GYROZEN). All the solutions' pH was measured with the pHenomenal from VWR (pH 1100L). Mini-Protean Tetra System (Bio-Rad) was used for SDS-PAGE Electrophoresis and Western blots, and the results were registered using a ChemiDoc XRS+ (Bio-Rad).

2.3 Protein Production

2.3.1 Recombinant DNA construction

The gene encoding for a recombinant protein, named Hdiv_1 and formed by the cysteine-rich secretory protein sequence from *Hediste diversicolor* (Accession OL606746, see also Moutinho Cabral et al., 2022), was optimized for expression in *Escherichia coli* and synthesized as an outsourced service from GeneCust (France). The Hdiv_1 sequence, with a Histidine-tag in its N-terminal, was cloned into the pET15b vector (TB045 5/99, Novagen) using the restriction sites of NcoI and NdeI. Therefore, the Hdiv_1 sequence is fused to a His-tag in the N-terminal, as shown in Table 2.1.

Table 2.1 – Hdiv_1 sequences:

Hdiv_1 Nucleotide Sequence (5' - 3')	CCATGGGCGGCATGGGTCATCATCATCATCATCATACCAACCTGATTATTCTGTTTGCATGTTTTGCGTTT GCAGTTGCGACGCCGATTGAACTGATTAACCCGGATGCAAATAATTATGAAGAAATGATGAGCAAACGCGG TACATGTGGTGCACAGCTGGCACGTCTGAGCAGTCGTAGTGCAACCAATTTTCTGAATGCCCATAATACCA AACGCCAGCAGGAAGGCCAGGGTCTGAGTGGTCTGACCTGGGATAGTGATCTGGCGGCACGTGCTCAGGAA CTGGCAAATAAATGTGTATTTAACCACGGTCTGGCAACCGATTGCAACGGTAAAAGTTGTGGTCAGAATAT TTATTATAGTGGCGGTAGTAGCTTTAGCGCAGCAAAAGTTGTTGATAGCTGGTATAGCGAAAAGAATGATT TTACATATAGTAGCAACTCCTGCGCGAGCGGTAAAGCATGTGGTCATTATACCCAGATTGTTTGGAAAAGC ACCCAGAAAGTGGGTTGTGCGGTTGCAGATTGTACCGGCAAAGTTATGGGCTATAGCCCGGAATATATCGC AGTTTGTAATTATTTTCCGCCGGGTAATTATATTGGTCAGAAACCATATTAAGGCCATATG
Hdiv_1 Amino acid Sequence	MGGMGHHHHHHTNLIILFACFAFAVATPIELINPDANNYEEMMSKRGTCGAQLARLS SRSATNFLNAHNTKRQQEGQGLSGLTWDSDLAARAQELANKCVFNHGLATDCNGKS CGQNIYYSGGSSFSAAKVVDSWYSEKNDFTYSSNSCASGKACGHYTQIVWKSTQKVG CAVADCTGKVMGYSPEYIAVCNYFPPGNYIGQKPY

2.3.2 Bacteria Transformation

The vector provided by Genecust was transformed in *Escherichia coli* BL21(DE3) and Origami™ B(DE3)pLysS Competent Cells (Novagen), purchased from NZYTech and Sigma-Aldrich, respectively (Table 2.2). The transformation was performed by the heat shock method. In short, 1.5 μL of the plasmid (20 ng/μL) was added to 50 μL and 30 μL of BL21(DE3) and Origami™ B(DE3)pLysS bacteria, respectively. Afterward, the bacteria were left in ice for 5 minutes, heated at 42ºC for 30 seconds (Origami™ B(DE3)pLysS strain) and 40 seconds (BL21(DE3) strain), and then left in ice for 2 minutes. Finally, 250 μL of SOC medium (super optimal broth with catabolite repression) from the manufacturer kit was added to the bacteria and incubated at 37°C, 250 rpm for 60 to 90 minutes. The bacteria suspension was plated in LB agar plates (25 g/L LB Broth, 15 g/L agar granulated, both from NZYTech) with the respective antibiotics to select the correct transformants. The BL21(DE3) bacteria were plated in LB agar with ampicillin (stock 100 mg/mL; CAS number 69-52-3) and the Origami™ B(DE3)pLysS bacteria were plated in LB agar with ampicillin (stock 100 mg/mL), chloramphenicol (stock 30 mg/mL; CAS number 56-75-7), tetracycline (stock 5 mg/mL; CAS number 64-75-5), and kanamycin (stock 50 mg/mL; CAS number 25389-94-0), all from NZYTech. All plates were incubated at 37°C for overnight or until it was possible to distinguish colonies.

Table 2.2 – *Escherichia coli* strains used in this work:

***E. coli* Strains**	**Genotypes**	**Catalog Number**	**Manufacturer**
BL21(DE3)	F- *ompT gal dcm lon hsd*S_B(r_B- m_B-) λ(DE3 **lacI lacUV5-T7 gene 1 ind1 sam7 nin5*	MB00601	NZYTech
Origami™ B(DE3)pLysS	F- *ompT hsd*S_B(r_B- m_B-) *gal dcm lacY1 ahp*C (DE3) *gor522*:: *Tn10 trxB* pLysS (Cam^R, Kan^R, Tet^R)	70839	Sigma-Aldrich

2.3.3 Protein Heterologous Expression

To develop a strategy for the heterologous expression of a Cys-rich protein (Hdiv_1) in *Escherichia coli,* was decided to test different expression conditions using the BL21(DE3) strain. During the expression were used different temperatures (37°C, 30°C, and 18°C), induction times (4 and 18 hours), and two different culture media, Lysogeny Broth (LB) and Terrific Broth (TB). The LB medium was prepared with 25 g/L LB Broth (10 g/L NaCl, 10 g/L tryptone, 5 g/L yeast extract) and the Terrific Broth (TB) medium was prepared by mixing 20 g/L tryptone, 24 g/L yeast extract micro granulated and 4 mL/L glycerol (all acquired from NZYTech), with the addition of sterile potassium phosphate buffer (0.017 M KH_2PO_4, 0.072 M K_2HPO_4).

A freshly grown BL21(DE3) colony was picked from the selected transformants plate and inoculated in 6 mL of liquid LB with 6 µL of ampicillin (stock 100 mg/mL). This pre-inoculum was incubated overnight at 37ºC and 250 rpm to promote bacterial growth. The next day, the glycerol stock was prepared by mixing 500 µL of the culture with 500 µL of 50 % (v/v) glycerol and then stored at -80°C. A 5 mL sample of the overnight culture was also transferred into 50 mL of LB or TB with 55 µL of ampicillin and incubated at 37°C and 250 rpm for 3 hours in a 200 mL erlenmeyer. Afterward, 10 mL of the preculture was inoculated into 1 L of LB or TB media, and the appropriate amount of antibiotic was added. After that moment, the optical density at 600 nm (OD_{600}) was measured hourly until obtaining a value of approximately 0.8, and then the culture was induced. The induction occurred by adding to each culture 1010 µL of 1 M IPTG (to a final concentration of 1mM of Isopropyl β-D-1-thiogalactopyranoside; NZYTech) and then incubating at the different temperatures with 250 rpm for at least 18 hours. This protocol was repeated three times, using the glycerol stock for the inoculation step, to perform expression under all the different induction temperatures (at 37ºC, 30ºC, and 18ºC).

To perform a time course analysis, the OD_{600} was measured during the expression, and samples were collected before (BI) and after induction (AI: hourly and overnight). The samples were normalized overtime by collecting a specific volume determined using the following formula: $Vt = 1.2/OD_{600}$ (where

Vt is the volume collected from the culture) and the samples were stored at -20ºC. Later, all samples were centrifuged at 12300 rpm for 2 minutes and the supernatants were discarded. The pellets were resuspended in SDS-PAGE sample buffer (0.05 M tris, 5 % (v/v) SDS, 6 % (v/v) beta-mercaptoethanol, 13 % (v/v) glycerol, 0.4 mM bromophenol blue) and after incubation at 100ºC for 10 minutes, the samples were analyzed through 12.5 % acrylamide SDS-PAGE gels.

The best conditions for Hdiv_1 expression in BL21(DE3) were at 30ºC and 18ºC with TB medium, so these conditions were applied for expression in Origami™ B(DE3)pLysS strain. An Origami colony was picked from the selected transformants plate and inoculated in 1 mL of SOC medium with 1 μL of each selected antibiotic: ampicillin (stock 100 mg/mL), chloramphenicol (stock 30 mg/mL), tetracycline (stock 5 mg/mL), and kanamycin (stock 50 mg/mL). This pre-inoculum was incubated at 37ºC and 250 rpm for 2 hours to promote bacterial growth, then 1 mL of the culture was transferred to 20 mL of LB with 20 μL of each antibiotic and left growth overnight in a 200 mL erlenmeyer, maintaining the same conditions. The next day, the glycerol stock was prepared and 10 mL of the culture was transferred into 1 L of TB medium with 1010 μL of each antibiotic. The culture was incubated in the same conditions until the OD_{600} was approximately 0.8. Afterward, the culture was induced by adding 1010 μL of 1M IPTG and incubated at 30ºC and 250 rpm for at least 18 hours. This protocol was repeated one more time, using the glycerol stock for the inoculation step, to perform expression at 18ºC. The time course analysis was performed as described before. All the different protein expression conditions tested are summarized in Table 2.3.

Table 2.3 – Conditions for recombinant expression of Hdiv_1:

Expression Conditions		***E. coli* Strains**	
Temperature and Time After Induction	**Culture Medium**	**BL21(DE3)**	**Origami™ B(DE3) pLysS**
37ºC; 4 hours	LB	x	-
	TB	x	-
37ºC; 18 hours (Overnight)	LB	x	-
	TB	x	-
30ºC; 18 hours (Overnight)	LB	x	-
	TB	x	x
18ºC; 18 hours (Overnight)	LB	x	-
	TB	x	x

Note: x represents the tested conditions; - represents the conditions that were not tested

In parallel with the aforementioned procedure, we produced an *E. coli* BL21(DE3) extract without the plasmid codifying to Hdiv_1 using the best expression strategy. The resulting protein extract contained all soluble proteins intrinsic to BL21(DE3), and it was used as a control for cell-based assays. A small-scale expression was performed by inoculating 200 μL of non-transformed BL21(DE3) bacteria

into 2 mL of TB medium. The next day, 1 mL of the culture was transferred to 100 mL of TB medium and was incubated at 18ºC and 250 rpm overnight.

After expression, the bacterial cells were harvested through centrifugation at 4ºC and 4500 rpm for 20 minutes, and the pellets were stored at -80ºC. Later, the pellets were subjected to three cycles of freeze (20 minutes at -80ºC) and thaw (15 minutes at 37ºC) to fragilize the cell walls. The bacterial pellet was weighted and for each g of pellet 5 mL of lysis buffer were added. The lysis buffer contained PBS (10 mM Na phosphate pH 7,4 150 mM NaCl), lysozyme from chicken egg white (0.2 mg/mL; CAS number 12650-88-3), DNase I from bovine pancreas (5 µg/mL; CAS number 9003-98-9), and cOmplete™ EDTA-free Protease Inhibitor Cocktail, all acquired from Sigma-Aldrich. The cell pellets were incubated with lysis buffer at 23ºC and 300 rpm for 1 hour. After incubation, the suspension was centrifuged at 4ºC and 10000 rpm for 20 minutes to separate the soluble (SF) and insoluble fractions (ISF). All the fractions obtained were then stored at -80ºC.

2.4 Protein Purification and Biophysical Characterization

2.4.1 Hdiv_1 Purification

The Hdiv_1 protein was purified from the BL21(DE3) soluble fraction by immobilized metal affinity chromatography (IMAC). This purification was carried out using 2 mL of IMAC Ni-Sepharose 6 Fast Flow (CAS number 17-5318-01, Cytiva), charged with nickel ions. The resin was packed into an empty disposable PD-10 Columns (CAS number 17-0435-01, Cytiva) using polyethylene frits with 15 mm diameter from the same supplier. Sequentially, the resin was first equilibrated with a binding buffer (20 mM Na Phosphate, 500 mM NaCl, 20 mM Imidazole, pH 7.4) and the absorbance of the flowthrough was measured at 280 nm to verify that the column was clean. Then, 2 mL of the soluble fraction was diluted 4:1 in the binding buffer and loaded into the column (2.5 mL of loading with an estimated protein concentration of 14.6 mg/mL). The column was incubated with the loading sample for 2 hours at room temperature with orbital agitation of 40 rpm. After incubation, the flowthrough was collected and the resin was washed with 3 column volumes (CV) of binding buffer. Afterward, the protein was eluted using a gradient of imidazole concentrations in the elution buffers, allowing the selective elution of proteins bound to the column. First, to obtain elution 1 and 2 were used 2 CV of Elution Buffer I (20 mM Na phosphate, 500 mM NaCl and 100 mM imidazole, pH 7.4), then for elution 3 and 4 were used 2 CV of Elution Buffer II (20 mM Na phosphate, 500 mM NaCl and 200 mM imidazole, pH 7.4) and finally to obtain the elution 5 and 6 were used 2 CV of Elution Buffer III (20 mM Na phosphate, 500 mM NaCl and 300 mM imidazole, pH 7.4). The resin was used several times; therefore, a stripping method and recharging of the column was performed, as described by the resin suppliers' instructions. All protein fractions were quantified for total protein and analyzed in a 12.5 % acrylamide SDS-PAGE gel.

After IMAC purifications, all the elution fractions containing the protein of interest were pooled together. The elutions were loaded into Float-A-Lyzer G2 Dialysis Devices (Repligen), with a pore size of 3.5 kDa – 5 kDa. Some dialysis occurred against 1x PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2HPO_4, pH 7.4) for 48 hours. During the dialysis, the PBS was changed three times each day every two hours. Other dialysis were performed against Kenny's Salt Solution (400mM NaCl, 9mM KCl, 0.7 mM K_2HPO_4, 2 mM $NaHCO_3$, pH 7.5), also for 48 hours, but the buffer changes were performed with an increasing gradient of NaCl and KCl until the desired concentration was reached, to avoid protein precipitation. These two buffers were chosen for dialysis because they were used to prepare the cell suspensions in the toxicity assays, as will be explain later in this chapter. Afterward, the protein samples were concentrated using an Amicon® Ultra-15 Centrifugal Filter Unit with a pore size of 10 kDa (Millipore). The Amicon was filled with the protein sample and was centrifuged for 30-minute cycles at 4900g until reaching the desired sample volume or concentration.

Alternatively, after IMAC purification, samples' aliquots were precipitated to remove imidazole, interfering with protein quantification and in the separation of proteins by electrophoresis. Therefore, for these two cases, precipitation with ethanol was used before the procedure. First, the protein solution was mixed with cold ethanol (1:9 (v:v)) and stored for at least an hour at -20ºC, followed by centrifugation at 4ºC, 10000 rpm for 15 minutes, to separate the proteins from the salts in solution. Then, the supernatant was discarded and after the ethanol evaporation, the pellets were resuspended in 5 % (w/v) SDS in 100 mM NaOH (for total protein quantification) or in PBS (for electrophoresis).

An extract enriched in Hdiv_1 (obtained after IMAC) was further purified through size exclusion chromatography (SEC). To perform this purification was used the Superdex™ 75 Increase 10/300 GL column (Cytiva) and the manufacture instructions were followed. All the purification steps were done automatically using the ÄKTA pure™ chromatography system (Cytiva). A method was implemented in the system for the purification. The first step was the equilibration in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2HPO_4, pH 7.4) with a flow rate of 0.5 mL/min for 5 minutes. Afterward, 500 µL of the protein extract was injected into the system and finally the elution step occurred, with a flow rate of 0.5 mL/min for 70 minutes. The different fractions collected during elution were quantified and analyzed in a 12.5 % acrylamide SDS-PAGE gel.

Two methods were used for protein quantification: Bicinchoninic acid (BCA) assay and Quanti-Pro™ BCA assay. The BCA assay is a sensitive cooper-based colorimetric assay used for total protein quantification. The BCA kit from Sigma-Aldrich used in this work has a linear concentration range between 200 µg/mL and 1000 µg/mL. For the assay, 25 µL of protein solution (samples before and after purification and calibration curve solutions) were added to each Flat Transparent 96-well microplate (Sarstedt) well. To each well was also added 200 µL of the BCA working solution and the microplates were incubated at 60ºC for 15 minutes. After incubation, the absorbance was measured at 560 nm in the microplate reader (TECAN Infinite 200). The protein concentration was obtained using a calibration standard curve using a BSA (bovine serum albumin) standard solution (Sigma-Aldrich) with values

between 0 and 1000 µg/mL. The QuantiPro™ BCA assay (kit from Sigma-Aldrich) relies on the same reactions and fundamentals that the BCA assay. However, its linear concentration range occurs between 0.5 µg/mL and 30 µg/mL. In this case, 150 µL of protein solution (samples and calibration curve solutions) were added to each well of a 96-well microplate. To each well was also added 150 µL of the working solution and the microplates were incubated at 60ºC for 1 hour. Afterward, the absorbance was measured at 560 nm in the same microplate reader. The protein concentration was obtained with a BSA calibration standard curve with values between 0 and 30 µg/mL. The i-control 2.0 Microplate Reader Software from Tecan was used to read absorbances in both assays.

2.4.2 SDS-PAGE protein gels

SDS-PAGE gels with 12.5 % acrylamide were produced in-house. The running gel was prepared by mixing 750 µL of 1.5 M tris pH 8.8, 2.08 mL acrylamide: bisacrylamide 37.5:1 (v:v), 50 µL of 10% SDS, 2.1 mL of distilled water, 38 µL of 10 % (w/v) APS and 2.5 µL of TEMED. After this solution was transferred between the glasses, 1 mL of isopropanol was added to avoid bubbles. After polymerization, the isopropanol was removed and the stacking gel (5% acrylamide) was added on top. This second gel consisted in a mix of 450 µL 0.5 M tris pH 6.6, 300 µL acrylamide: bisacrylamide 37.5:1, 18 µL of 10% SDS, 940 µL of ddH20, 13.5 µL of 10 % APS and 2.0 µL of TEMED. Finally, a comb of 10 or 15 wells was placed between the glasses, and the stacking gel was left to polymerize. For electrophoresis, the gel was placed in the running module with running buffer (25 mM tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3). To each protein solution was added sample buffer (0.05 M tris, 5 % (v/v) SDS, 6 % (v/v) beta-mercaptoethanol, 13 % (v/v) glycerol, 0.4 mM bromophenol blue), and all samples were boiled at 100ºC for 10 minutes. The electrophoresis occurred at 110 V for 90 minutes and after the running, the gel was stained with Coomassie blue staining (6 mM Coomassie blue R-250, 8 % (v/v) acetic acid glacial, 45 % (v/v) methanol) for about 30 minutes. Afterward, was applied the destaining solution (7.5 % (v/v) acetic acid glacial, 45 % (v/v) methanol) and the gel and its densitometry analysis were performed using the Image Lab software (Biorad). The markers used for SDS-PAGE electrophoresis were Low Molecular Weight (LMW) Protein Marker (NZYTech) and the marker Precision Plus Protein™ Dual Xtra PreStained Protein Standards (Bio-Rad).

2.4.3 Western blot

Before and after purification, various fractions were analyzed by Western blot. This method allows the detection of histidine-tagged proteins, in this case, the detection of Hdiv_1. The protein samples were initially loaded in a 12.5 % acrylamide SDS-PAGE gel that ran at 110 V for 90 minutes. The transfer step occurred at 100 V for 30 minutes and was performed using a transfer buffer (25 mM tris, 192 mM glycine, 20 % (v/v) methanol), and a 0.45 µm nitrocellulose membrane (Bio-Rad). To confirm that the transference was successful a staining step was performed using Ponceau-S (0.1 % (w/v) Ponceau-S in 5 % (v/v) acetic acid; CAS number 6226-79-5, PanReac AppliChem). After removing the staining,

successive washes were performed with distillated water and 5% (v/v) acetic acid. Subsequently, the membrane was washed twice with TBS buffer (20 mM tris, 150 mM NaCl, pH 7.5) and blocked using a blocking solution (from the Qiagen kit or a 3 % BSA solution in TBS) for 1 hour at room temperature, after which it was washed again with TBS buffer. For the detection step was used a Penta-His HRP antibody conjugate with Horseradish Peroxidase (HRP diluted in the blocking solution in a dilution between 1:1000 to 1:1500). As blocking solution were used the solution provided by the Qiagen kit - Anti-Histidine Horseradish Peroxidase Conjugate kit, and as an alternative a 1% bovine serum albumin (BSA; CAS number 9048-46-8, Fisher Scientific) solution in TBS. The incubation was performed overnight at 4ºC. The membrane was washed twice the next day, first with 0.1 % (v/v) Tween in TBS and second with TBS buffer. To visualize the His tagged protein band, 1-Step-TMB blotting (Thermo Scientific) was added to the membrane as a colorimetric substrate for HRP. The Western blot results were registered using the Bio-Rad Image Lab software, which was also used to analyze the bands by densitometry.

2.4.4 Isoelectric Focusing

Isoelectric focusing (IEF) is a method used to separate a complex mixture of proteins according to their isoelectric point (pI) and molecular weight. In this case, IEF was used to obtain the Hdiv_1 experimental pI. For this procedure were used the PROTEAN IEF cell and 7 cm Immobilized pH gradient strips (IPG strips), more specifically Immobiline DryStrip Gels, and were followed the instructions of the manufacturer (Bio-Rad). In particular, were used rehydration buffer (8 M urea, 1 % (v/v) CHAPS, 100 mM DTT, 0.02 % (v/v) bio-lyte in nanopure water), equilibration buffer (6 M urea, 20 % (v/v) glycerol, 2 % (v/v) SDS, 130 mM DTT), mineral oil and liquid agarose. The electrophoresis occurred at 110 V for 90 minutes, using electrophoresis buffer (25 mM tris, 192 mM glycine, 0.1 % (w/v) SDS, pH 8.3) and after the running, the 12.5% SDS-PAGE gels were stained by silver staining and analyzed in the Image Lab Software.

2.4.5 Mass Spectrometry

Two bands of interest obtained after IMAC were analyzed through mass spectrometry to confirm which of the bands corresponds to Hdiv_1. Firstly, around 35 µg of the protein extract was mixed with a concentrated sample buffer (0.5 M tris.Cl, 10 % SDS, 30 % glycerol (v/v), 0.6 M DTT, 0.012 % bromophenol blue (v/v)) and the mixture was boiled during 10 minutes. When the sample reached room temperature, were added 5 µl of acrylamide solution for each 30 µl off sample (30 % acrylamide: bisacrylamide 37.5:1). After 15 minutes, the sample was loaded in a 12.5 % acrylamide SDS-PAGE gel and the conditions were set at 110 V for 90 minutes. Afterward, the gel was stained using Coomassie blue and washed several times using water treated with diethyl pyrocarbonate (DEPC; CAS number 1609-47-8; BioChemica, Panreac). As soon as it was possible to detect the bands of interest, they were excised and placed inside 1.5 mL tubes with 0.05 % sodium azide (CAS number 26628-22-8, Sigma-Aldrich) and preserved at 4ºC. The bands were sent to be performed chromatography with tandem mass

spectrometry (LC-MS/MS) analysis at the Life Sciences Life Sciences Mass Spectrometry Laboratory, Center for Neuroscience and Cell Biology of University of Coimbra.

2.5 Toxicity Testing

The toxicity or bioactivity of the recombinant product was determined through cytoassays using primary mussel gill cells, which allowed us to perform simple and practical *in vitro* assays. Specifically, the toxic effects of Hdiv_1 in mussel cells were assessed through its impact on cell viability, oxidative stress, lysosomal membrane stability, and DNA damage, using the trypan blue exclusion test, neutral red retention assay, and comet assay respectively (Curpan et al., 2022; Repetto et al., 2008; Fairbairn et al., 1995). The toxicity assays covered in this subchapter were based on the studies performed by Zhang et al. (2017), Faucet et al. (2004), and Repetto et al. (2008) and further optimized by the Seatox team.

2.5.1 Animal Collection

Adult *Mytilus edulis* (4-5 cm and 2-3 cm shell length and width, respectively) were hand-collected between April and June 2023 during the low tide from a rocky intertidal area in Costa da Caparica (West Coast of Portugal, 38°38'25.4"N 9°14'20.7"W). The mussels were acclimated in a closed-circulation system fitted with glass aquaria mesocosms, recreating their natural habitat with controlled salinity, temperature, and photoperiod (30; 19 ± 1 ºC and 12:12 h, respectively). Animals were fed each two days with *Arthrospira platensis* (spirulina from Prozis).

2.5.2 Cytotoxicity Assay

The cytotoxicity assays were optimized by preliminary tests using different buffers in the cell suspensions, as well as different exposure times to, in this case, two treatments: the buffer used in the suspension and hydrogen peroxide 30% (w/w) in H_2O (H_2O_2; CAS number 7722-84-1, Sigma-Aldrich). The cell suspensions were prepared by extracting the mussel gills through dissection, rinsing them in the sterilized tested buffer, macerating and centrifuging them, and then collecting the supernatant. The viability of the cell suspension was determined by adding trypan blue (CAS number 72-57-1, Sigma-Aldrich; 0.04 % (v/v) in the tested buffer) to 25 µL of the cell suspension (1:1). Viable and dead cells were quantified using a DM 2500 LED model microscope (Leica Microsystems) and a Neubauer chamber to determine the cell density. Cell viability was calculated by dividing the number of viable cells by the total number of cells (being expressed as percentage). In the course of this project, all prepared cell suspensions were subjected to the trypan blue assay before any treatment to verify whether cell viability was greater than 90 %. Only in these cases the suspensions were used in the assays. Afterward, cell suspensions were prepared in the same tested buffer, with a known cell density of 1×10^6 cells/mL. From this suspension, 5×10^4 cells were seeded in each exposure well of the 96 Well Cell Attachment

Treated Flat BottomTissue Culture Plates (VWR) and incubated with the treatment during the selected exposure time. Each condition had three technical replicates (three wells per each condition). Subsequently, the plate was centrifuged to increase cell adherence, and the wells were washed with the tested buffer. Finally, 100 µL of Neutral Red (PanReac; 0.04% (w/v) in the tested buffer) was added to the wells, and after 2 hours of incubation (and more washes), the dye was replaced with a destaining solution (1% (v/v) acetic acid glacial and 50% (v/v) ethanol). Cytotoxicity was determined by quantifying the dye incorporated in viable cells, measuring absorbance at 540 nm with the Multiskan SkyHigh Microplate Spectrophotometer (Thermo Scientific). The buffers tested in this optimization attempt were Dulbecco's PBS (without Mg and Ca) prepared in-house (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2HPO_4, pH 7.4), Dulbecco's PBS (without Mg and Ca) purchased from VWR, Hank's Balanced Salt Solution (138 mM NaCl, 5.33 mM KCl, 0.3 mM Na_2HPO_4; 0.44 mM KH_2PO_4, 4 mM $NaHCO_3$, 5.6 mM D(+)-glucose; pH 7.4) and Kenny's Salt Solution (400mM NaCl, 9mM KCl, 0.7 mM K_2HPO_4, 2 mM $NaHCO_3$, pH 7.5). The exposure times tested were 10, 20, 30 and 60 minutes.

After optimizing the assay, the cytotoxic potential of Hdiv_1 was assessed by subjecting the mussel gills cells to different concentrations (0 µg/mL, 1 µg/mL, 5 µg/mL, 25 µg/mL and 50 µg/mL) of the extract containing the protein of interest (ExtHdiv_1) for the estimation of the half-maximal inhibitory concentration (IC50). The same concentrations of a protein extract containing all soluble proteins intrinsic to BL21(DE3) when plasmid codifying for Hdiv_1 is absent (ExtP) were used as a control to eliminated residual interference effects caused by contaminants. The cell suspension preparation, the viability assessment and the assay itself were performed as described before, using only the Kenny's Salt Solution and 30 minutes as the exposure time. The assays were performed three times (biological replicates), and technical triplicates were included to minimize technical error. Hydrogen peroxide was used as a positive control in all assays.

2.5.3 Comet Assay

The alkaline comet assay (first described in Singh et al., 1988) was performed to determine if Hdiv_1 could potentially cause DNA damage in mussel gill cells, adapted from Martins and Costa (2020), with modifications. In brief, after obtaining a cell suspension in Kenny's Salt Solution, with viability superior to 90 % as described for the neutral red assay, the cells were exposed in tubes, to two different protein concentrations, 50 µg/mL and 5 µg/mL. The protein extracts used were ExtHdiv_1 (extract containing Hdiv_1) and ExtP (control to discard the contaminant's effects). Were also used other two controls, in one the cells were exposed to only Kenny's Salt Solution and in the other one the cells were exposed to hydrogen peroxide. After 30 minutes of exposure (at 18ºC), the suspensions were washed with Kenny's buffer as described before, and new Kenny's solution (100 µL) was added to the tubes. After resuspending the cells in the new buffer, 20 µL of the suspension was embedded in 180 µL of molten (~ 37 ºC) 1 % w/v low melting point agarose (LMPA, CAS number 39346-81-1, Sigma-Aldrich) prepared in Kenny's Salt Solution. Then, two 80 µL drops of LMPA with the incorporated samples were

placed in slides precoated with 1.2 % w/v normal melting point agarose (Eurobio Scientific) prepared in TAE buffer (40 mM tris, 1 mM EDTA; 20 mM acetic acid glacial). The slides were immediately covered with a coverslip, and LMPA solidified at 4ºC for 15 minutes. Per each sample two slides were prepared (technical replicates) and the assay was repeated three times (biological replicates). After LMPA solidification, the coverslips were gently removed, and the slides were placed in cold lysis buffer (2.64 % (w/v) NaCl, 3.72 % (w/v) EDTA, and 5 mM Tris, pH 10), supplemented with 10 % v/v DMSO and 1 % v/v Triton X-100 before use. The lysis occurred for 1 hour at 4ºC in the dark, and after this step the slides were transferred to cold electrophoresis buffer (1 mM EDTA and 300 mM NaOH) for 40 minutes. Afterward, the electrophoresis was run at 25 V and 4ºC for 30 minutes, maintaining the dark conditions. This process was performed using the CSL–COM20 Comet assay tank from Cleaver Scientific. Subsequently, 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. The slides can be directly analyzed after the neutralization step or after rehydration of archived slides in cold Milli-Q water for 30 minutes. The preparations were stained for the analysis by dropping 40 µL of GreenSafe Premium (Nzytech) diluted in Milli-Q water (1:500) in each half slide. A coverslip was placed immediately on top of the slides, and they were placed in the dark for 10 minutes at room temperature to allow the dye to bind to the DNA. Subsequently, the nucleoids were visualized in a DM 2500 LED model microscope equipped with an EL6000 model ultraviolet source and an MC 190 HD camera (all from Leica Microsystems). The micrographs of comet fields were obtained in grayscale mode at 400 x magnification and were scored using the software CometScore 1.6 (Tritek), with an analysis of 100 nucleoids per sample. The measurements were made for two main metrics: % DNA in Tail and Olive Moment. Additionally, the nucleoids were distributed into five classes based on the % DNA in Tail (0 – 20, 20 – 40, 40 – 60, 60 – 80, and 80 – 100 %), according to the data retrieved from the software. This distribution was useful for statistical analyses and quality assessment. The % DNA in Tail was considered a direct measure of DNA damage, while the derived metric Olive Moment was analyzed for comparative purposes as a metric conceived to handle heterogenous distribution of DNA in tails (Olive et al., 2012).

2.5.4 Statistical Analysis

The statistical analysis was computed using the R 4.3.1 software (Ihaka & Gentleman, 1996) and the RStudio interface with packages *car* (to evaluate assumptions for parametric analysis) and *gplots* (to plot data). The Shapiro-Wilk's and Levene's tests were used to evaluate normality and homoscedasticity, respectively. When the data met the assumptions for parametric analyses was employed ANOVA based on Tukey's HSD (honestly significance difference) test for multiple comparisons. The different protein extracts and concentrations were the two independent variables (both of which factorial). The dependent variables considered were Cytotoxicity (%) for the neutral red assays and % DNA in Tail and Olive Moment on the comet assay. The significance threshold α was set at 0.05 for all analyses.

2.5.5 Microscopy Analyses

A microscopy analysis was applied to visually assess and compare the morphology of cells subjected to various treatments. The procedures used in this subchapter were adapted from Costa (2018). The treatments studied included 50 µg/mL of ExtHdiv_1 and ExtP and the control conditions used in the comet assay. The analysis was performed using the cell suspensions left from the comet assay without the LMPA mixing. All replicates, encompassing both technical and biological replicates, were combined for fixation. The fixation process involved adding 4% (w/v) paraformaldehyde (PFA; CAS number 30525-89-4, PanReac) in Kenny's Salt Solution at 4°C to the cells. Following fixation, the cells were centrifuged at 800 g at 20ºC until visible cell pellets were formed. Subsequently, paraformaldehyde was removed, and the pellets underwent post-fixation with a solution containing 1% (w/v) osmium tetroxide (OsO_4; CAS number 20816-12-0 Sigma-Aldrich) in Kenny's Salt Solution. After two hours, the samples were centrifuged at 800 g for 7 minutes at 20ºC to re-concentrate the cells into pellets, followed by removal of the osmium tetroxide and three washes with milliQ-grade ultrapure water (each wash lasting 10 minutes). The subsequent step involved dehydration, achieved through a progressive series of acetone (ranging from 30 % to 100 %), with centrifugation between each step to maintain pellet integrity. Finally, the samples were embedded in Epon resin (Epoxy-Embedding Kit, Sigma-Aldrich). During intermediate steps, infiltration was carried out using Epon: Propylene oxide (CAS number 75-56-9, Thermo Scientific) ratios of 1:2, 1:1, and 2:1 (30 minutes each). Semi-thin sections (200 - 500 nm) were obtained using an Ultramicrotome Reichert Jung Ultracut E and stained with Toluidine Blue (CAS number 92-31-9, Thermo Scientific). Histological analyses were performed using a DM 2500 LED model microscope equipped with a MC 190 HD camera (Leica Microsystems).

3

RESULTS

3.1 Introduction

A recombinant protein named Hdiv_1, formed by the sequence of the cysteine-rich secretory protein fused to a His-tag at the N-terminal, was constructed (Figure 3.1A). The Hdiv_1 gene sequence synthesis and cloning into pET-15b expression vector was outsourced. The recombinant protein is composed of 208 amino acids, ten of which are cysteines, and has a theoretical molecular weight of ~ 22 kDa (the chemical properties were analyzed using the protein sequence and are summarized in Table 3.1). Upon arrival, the plasmid was successfully transformed into *Escherichia coli* strains – BL21(DE3) and Origami™ B(DE3)pLysS, ideal for recombinant production of Cys-rich proteins.

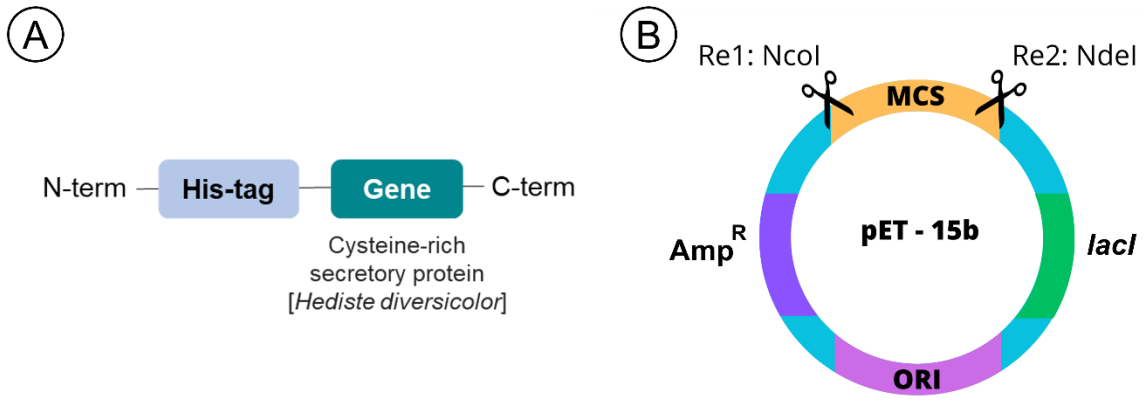

Figure 3.1 – Hdiv_1 gene and pET – 15b vector schemes. A) Hdiv_1 Construction strategy, formed by the His-tag in the N-terminal, linked to the cysteine-rich secretory protein sequence B) Vector pET – 15b Scheme. MCS: Multiple Cloning Site; AmpR: Sequence that confers resistance to Ampicillin; *lac*I: gene that codes for lac repressor protein; ORI: Replication origin; Re1: Restriction Enzyme 1 (NcoI); Re2: Restriction Enzyme 2 (NdeI).

Table 3.1 - Fusion protein properties (theorical data retrieved from Expasy Protoparam tool).

Protein Name	Number of residues	Number of Cysteines	Molecular Weight	Extinction coefficient (M^{-1} cm^{-1})	Isoelectric Point	Protein solubility
Hdiv_1	205	10	22260.9 g/mol	33515 / 32890	7.59	Poor water solubility

3.2 Protein Expression

In Figure 3.2 are presented the SDS-PAGE gels of the time course analysis for the different protein expression conditions tested (summarized before in Table 2.3). According to the SDS-PAGE gels, the band expected for Hdiv_1 (with a theoretical molecular weight of 22 kDa) was not clearly present after induction at 37ºC in both media (Figure 3.2A). On the other hand, some differences between before and after induction were observed when the expression was performed at 30ºC and 18ºC (Figure 3.2B and 3.2C). These differences were particularly noticeable when the expression was done using Terrific Broth (TB) media. The gels also showed no significant differences in the protein pool between 4 and 18 hours after induction (ON). Due to all these, the recombinant expression in *E. coli* BL21(DE3) at 30ºC and 18ºC in TB media were considered the best expression conditions and used for further studies. The SDS-PAGE gels of the time course analysis for the expression in Origami™ B(DE3)pLysS at 30ºC and 18ºC (ON) with TB medium are present in Figure 3.3. This time, the time course analysis did not show significant differences between the protein pool before and after induction during the expression in any of the conditions tested.

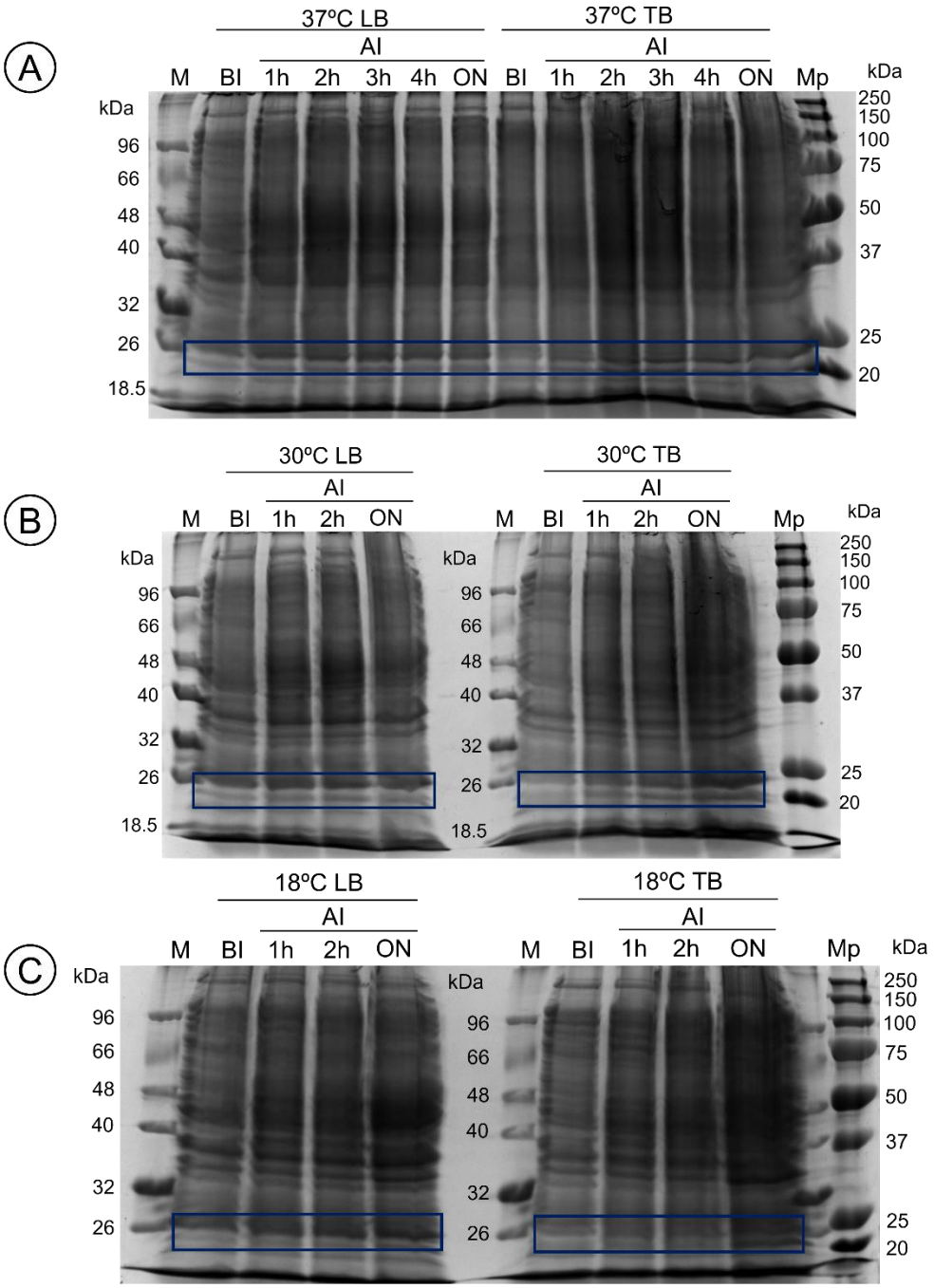

Figure 3.2 – Results of recombinant expression of Hdiv_1 under different conditions using *Escherichia coli* BL21(DE3) strain. A) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 37ºC using 1 mM IPTG and different culture media: Lysogeny Broth (LB) and Terrific Broth (TB). Samples collected (15 µL each): before induction (BI), after 1h, 2h, 3h, 4h, and 18h (ON) of induction, and before harvesting the culture (AI). B) and C) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 30ºC and 18ºC, respectively, using 1 mM IPTG and different culture media: Lysogeny Broth (LB) and Terrific Broth (TB). Samples collected (15 µL each): before induction (BI), after 1h, 2h, and 18h (ON) of induction, and before harvesting the culture (AI). Markers: Low Molecular Weight Protein Marker (M) and Precision Plus Protein™ Dual Xtra (Mp). Staining with Coomassie blue. The highlighted region indicates the expected location of Hdiv_1.

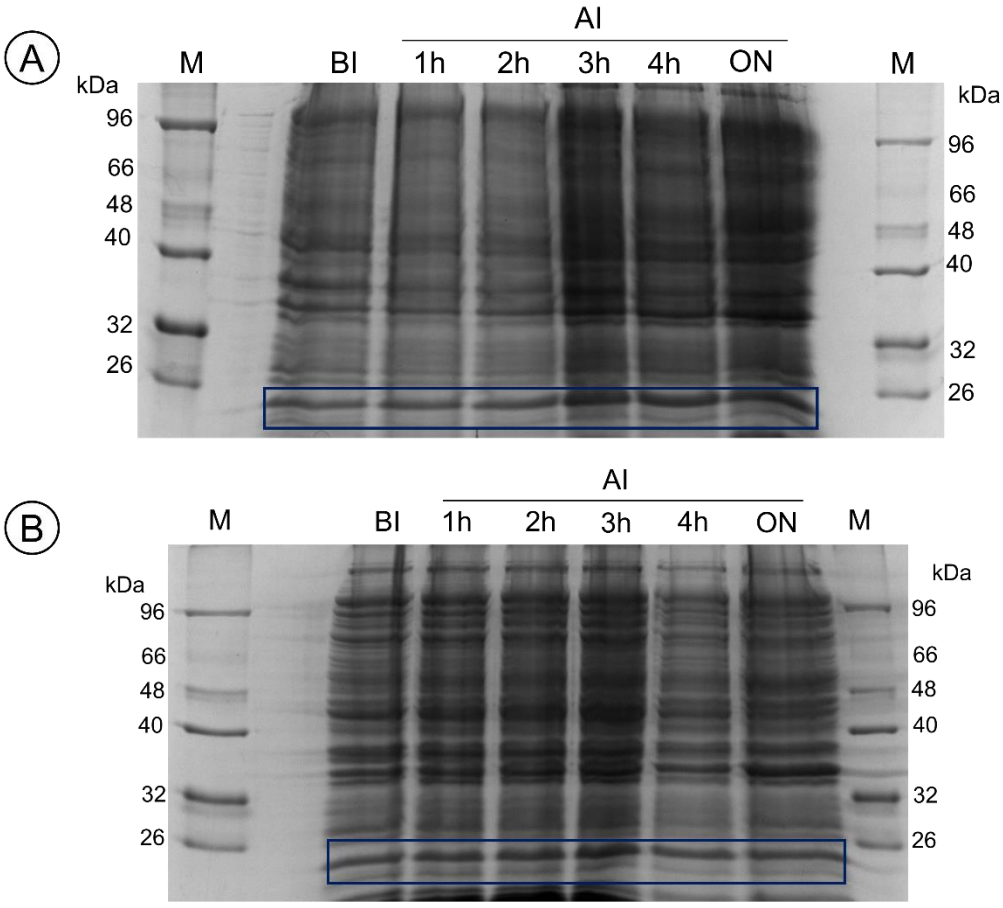

Figure 3.3 - Expression of Hdiv_1 under different conditions using *Escherichia coli* Origami™ B(DE3)pLysS strain. A) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 30ºC using 1 mM IPTG and Terrific Broth (TB) media. Samples collected (15 µL each): before induction (BI), after 1h, 2h, 3h, 4h, and 18h (ON) of induction, and before harvesting the culture (AI). B) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 18ºC using 1 mM IPTG and Terrific Broth (TB) media. Samples collected (15 µL each): before induction (BI), after 1h, 2h, 3h, 4h, and 18h (ON) of induction, and before harvesting the culture (AI). Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue. The highlighted region indicates the expected location of Hdiv_1.

Figure 3.4 shows the SDS-PAGE gel and Western blot with the soluble (SF and SFII) and insoluble fractions (ISF) from the *E. coli* BL21(DE3) extracts obtained at 18ºC and 30ºC (Figure 3.4A and 3.4B, respectively). As shown in the Western blot, it was possible to correlate a band with Hdiv_1 in all fractions (soluble and insoluble) resulting from the expression at 18ºC in TB medium (Figure 3.4B). The identified band appeared to have a molecular weight of 30 kDa. Using gel densitometric analysis, it was noticed that this band corresponds to a 30 kDa band in the normalized gel shown in Figure 3.4A. Even though Western blot did not detect this band in the 30ºC fractions, a 30 kDa band was present in the 30ºC lanes in the normalized gel. The fractions from the expression at 18ºC and 30ºC using the Origami™ B(DE3)pLysS strain were also analyzed through SDS-PAGE gel (Figure 3.4C). A band with 30 kDa was also present in all fractions (soluble and insoluble). However, the total protein concentration of the soluble fractions, especially the one from the expression at 18ºC, is considerably higher in BL21(DE3) expression (20.30 ± 3.65 mg/mL) when compared to that obtained with Origami™ B(DE3)pLysS expression (4.52 ± 0.59 mg/mL).

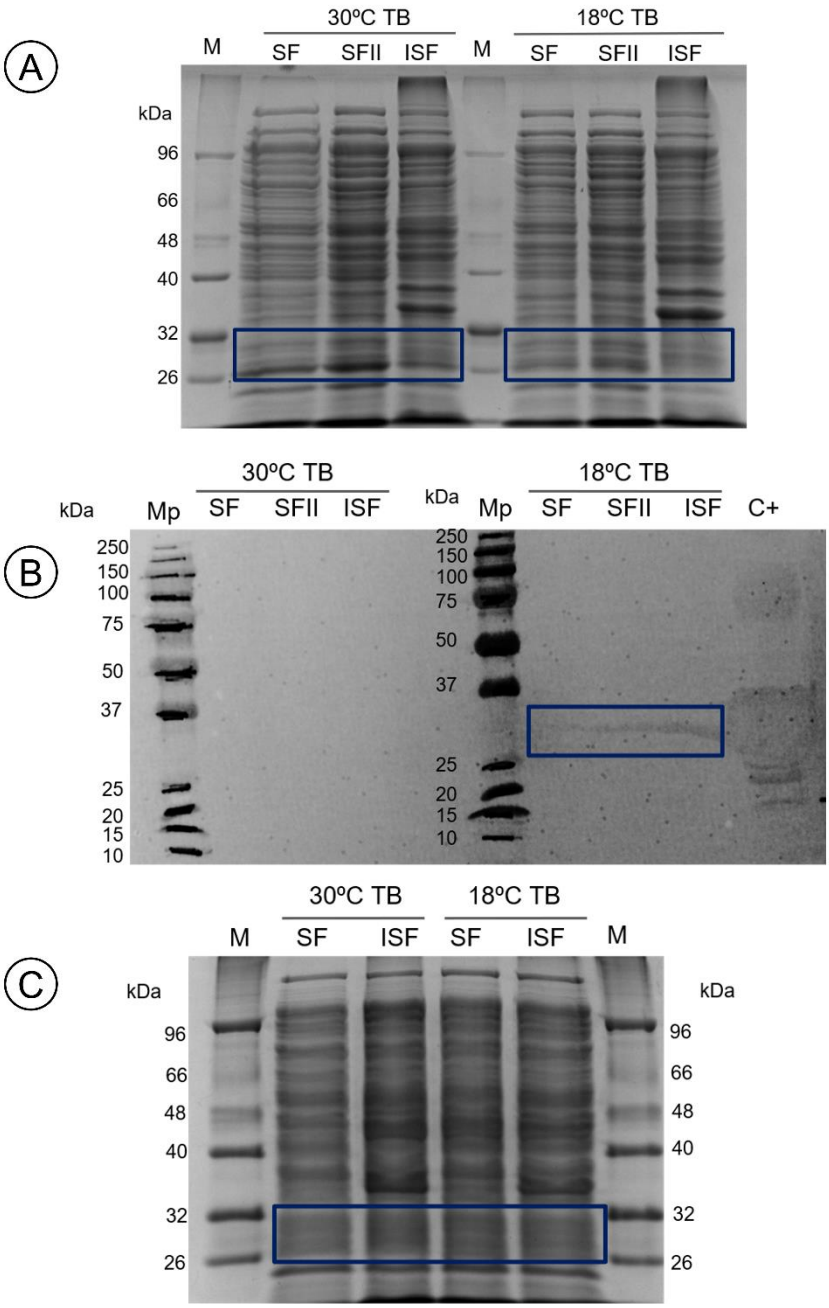

Figure 3.4 – Analysis of the different protein fractions from expression in *Escherichia coli* BL21(DE3) and Origami™ B(DE3)pLysS after cell lysis. A) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 30ºC and 18ºC in BL21(DE3) using Terrific Broth (TB) medium and cell lysis. Normalized gel (35 µg of total protein per lane) with samples from the different fractions: Soluble fractions (SF and SFII, obtained after the first and second lysis, respectively) and Insoluble fractions (ISF). Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue. B) Western blot using a 0.45 µm nitrocellulose membrane and revealed using a colorimetric detection method. In each lane, 35 µg of total protein per lane from the different BL21(DE3) expression fractions and R1b as the positive control (C+). Marker: Precision Plus Protein™ Dual Xtra (Mp). C) 12.5 % SDS-PAGE gel of crude extracts obtained after expression at 30ºC and 18ºC in Origami™ B(DE3)pLysS using Terrific Broth (TB) medium and cell lysis. Normalized gel (35 µg of total protein per lane) with samples from the different fractions:

Soluble fractions (SF) and Insoluble fractions (ISF). Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue. The highlighted region indicates the expected location of Hdiv_1.

3.3 Purification of Hdiv_1

After comparing the total protein and the estimated Hdiv_1 concentration obtained in all fractions (from the two different *E. coli* strains), the soluble fraction (SF) from the BL21(DE3) expression at 18ºC was selected to attempt Hdiv_1 purification by immobilized metal affinity chromatography (IMAC). From this purification, was obtained a normalized SDS-PAGE and a Western Blot (Figure 3.5A and 3.5B). After IMAC, a band with 26.5 kDa was detected by Western blot, which should correspond to Hdiv_1, showing that this protein was eluted in elution E4. The same band was not detected by Western blot in other fractions (loading and other elutions), notwithstanding in the normalized gel (Figure 3.5A), bands with molecular weights close to the band of interest were observed in other lanes. The band detected at this moment had a lower molecular weight (26.5 kDa) compared with the band detected before purification (30 kDa), as shown before in Figure 3.4B. The densitometry analysis of the SDS-PAGE gel (Figure 3.5A) allowed to characterize the two bands that appeared in the elution E4: the one detected before purification with a molecular weight around 30 kDa and an intensity of 36.6 % and the band detected after IMAC with a molecular weight around 27 kDa and an intensity of 18.9 %. Thereby, in subsequent analysis of SDS-PAGE gels these two bands were taken into account. Figures 3.5C and 3.5D show the SDS-PAGE gels obtained after the dialysis and concentration of the elution E4 in two different buffers: Dulbecco's PBS and Kenny's Salt Solution. In these gels was observed the presence of the bands of interest after dialysis, showing a similar pattern to E4. The same happened in the samples after concentration, suggesting that significant protein loss did not occur during these steps.

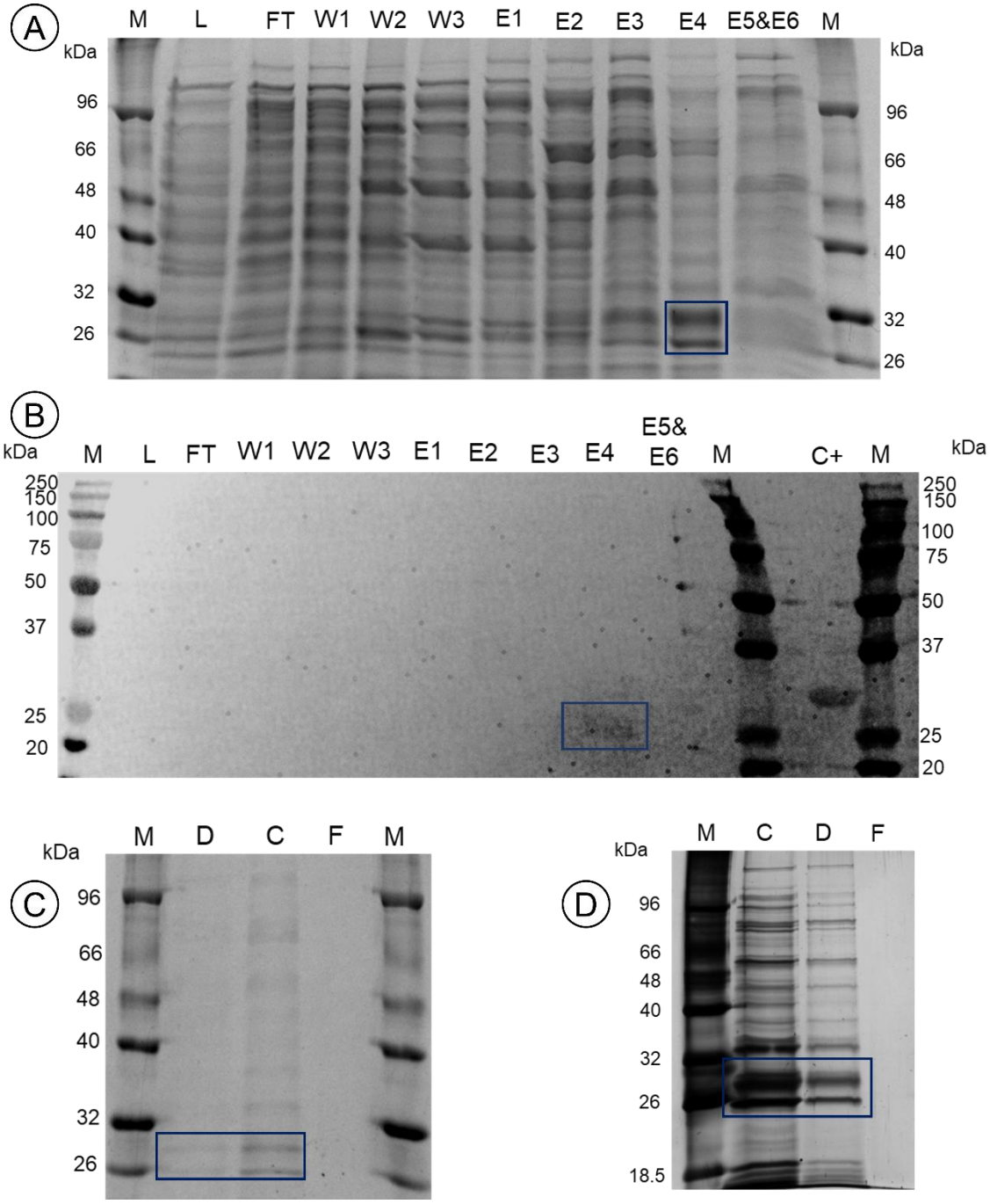

Figure 3.5 - Hdiv_1 purification by immobilized metal affinity chromatography and dialysis. A) 12.5 % SDS-PAGE gel of Hdiv_1 purification. Normalized gel (10 ug of total protein per lane) with samples from the different fractions. Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue. B) Western blot using a nitrocellulose membrane 0.45 µm and revealed using a colorimetric detection method.10µg of total protein per lane and R1b as the positive control (C+). Marker: Precision Plus Protein™ Dual Xtra (M). C) 12.5 % SDS-PAGE gel of elution E4 after dialysis to PBS and concentration (Load of 15 µL). Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue. D) 12.5 % SDS-PAGE gel of E4 dialysis to Kenny's Salt Solution and concentration (Load of 15 µL). Marker: Low Molecular Weight Protein Marker (M). Staining with silver stain. D: E4 after dialysis; C: E4 concentrated; F: filtrated during concentration. The highlighted region indicates the expected location of Hdiv_1.

As observed in Figure 3.5A, many other protein bands were present in the elution E4, corresponding to contaminants. Figure 3.6 exhibits an SDS-PAGE gel obtained after size exclusion chromatography (SEC), performed as an attempt to improve Hdiv_1 purification. The fractions collected over time from SEC were analyzed in the gel, and fraction C appeared to be the purest, showing a band at 31 kDa with an intensity of 85.8 %. However, the other band of interest (with a molecular weight around 27 kDa) did not appear to be present in this fraction. The total protein concentration of fraction C was 2.23 ± 0.07 µg/mL. Summarizing, it was possible to obtain a significant estimated amount of both bands of interest in the soluble fraction after expression, based on 1 L of culture (792 mg and 2355 mg). However, these quantities decreased significantly after purification by IMAC and SEC (in which only one band was detected). As a consequence, and presuming that the band detected after SEC corresponded in fact to Hdiv_1, it was obtained a yield of 0.24 %. The quantities of proteins obtained along the expression and purification process are summarized in Table 3.3.

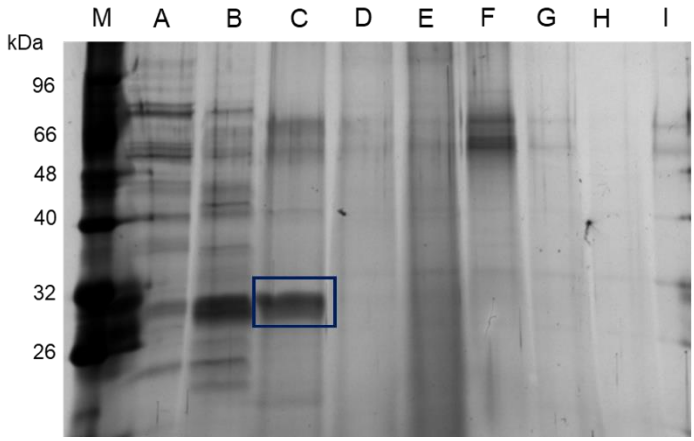

Figure 3.6 – Hdiv_1 purification by size exclusion chromatography. 12.5 % SDS-PAGE gel of Hdiv_1 purification (approximately 30 kDa). A to I correspond to fractions collected over time (Load of 15 µL) Marker: Low Molecular Weight Protein Marker (M). Staining with silver stain. The highlighted region indicates the expected location of Hdiv_1.

Table 3.2 – Summary of Hdiv_1 production yield after expression in BL21(DE3):

	***Escherichia coli* Strain**	**BL21(DE3)**		
	Sample Fractions	SF	SFII	ISF
Crude	Total protein mass (mg/L)	20 300 ± 3 650	5 380 ± 2 400	53 890 ± 2 530
	Initial Hdiv_1 content (%)	3.9[A] 11.6[B]	4.2[A] 12.2[B]	2.0 [A] 6.8[B]
	Estimated Hdiv_1 mass (mg)	792[A] 2 355[B]	226[A] 656[B]	1 078[A] 3 665[B]
Purified	Total protein mass after IMAC (mg)	89 ± 11	-	-
	Estimated Hdiv_1 content after IMAC (%)	36.6[A] 18.9[B]	-	-
	Total protein mass after SEC (mg)	2.23 ± 0.07	-	-
	Estimated Hdiv_1 content after SEC (%)	85.8[A]	-	-
	Estimated Hdiv_1 mass (mg)	1.91 [A]	-	-
	Yield (%)	0.24 [A]	-	-

Notes:
SF – Soluble Fraction
ISF – Insoluble Fraction
[A] – Values considering the SDS-PAGE band with 30-31 kDa
[B] - Values considering the SDS-PAGE band with 26-27 kDa
Crude:
Total protein mass – Total mass of protein produced in one liter of culture quantified by BCA assay;
Initial Hdiv_1 content – Hdiv_1 content estimated by SDS-PAGE gel densitometry (Figure 3.4A).
Estimated Hdiv_1 mass – Hdiv_1 mass in one liter of culture calculated from estimated content x total protein mass
Purified:
Total protein mass after IMAC – Total mass of protein in E4 elution obtained by IMAC quantified by QuantiPro™ BCA Assay, estimation was based on 1 L of culture;
Estimated Hdiv_1 content after IMAC – Hdiv_1 content estimated by SDS-PAGE gel densitometry (Figure 3.5A)
Total protein mass after SEC – Total mass of protein in Elution C obtained after SEC quantified by QuantiPro™ BCA Assay, estimation was based on 1 L of culture;
Estimated Hdiv_1 content after SEC – Hdiv_1 content estimated by SDS-PAGE gel densitometry (Figure 3.6) is also used as Purity Ratio
Estimated Hdiv_1 mass – Hdiv_1 mass in one liter of culture after purification calculated from estimated content x total protein mass
Yield - Estimated purified Hdiv_1 Mass / Estimated crude Hdiv_1 Mass x 100

3.4 Biophysical Characterization

In Figure 3.7 are present the results of isoelectric focusing (IEF), using soluble and insoluble fractions from the expression at 18ºC with TB media (*E. coli* BL21(DE3) strain). In both fractions we observed the presence of proteins with molecular weights compatible with those of our protein-of interest, furthermore exhibiting a pI similar to the theoretical (7.6 obtained from Expasy Protoparam tool), around 7.7. However, excising the respective bands from the gel was cumbersome and the subsequent mass spectrometry results (by LC-MS/MS) obtained at the time of the thesis submission were inconclusive, suggesting that contamination with proteins intrinsic to *E. coli* occurred.

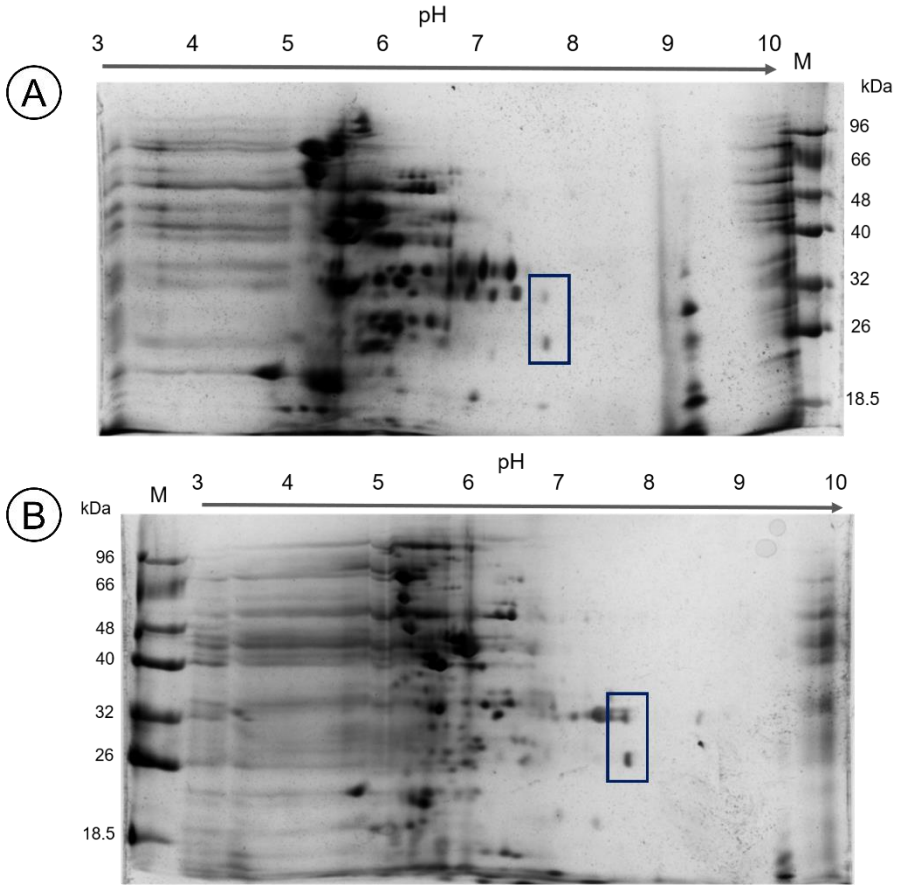

Figure 3.7 – Protein characterization by isoelectric focusing. 12.5 % SDS-PAGE gel with samples after cell lysis from expression at 18ºC using Terrific Broth (TB) medium. Normalized gel with 169 µg of total protein. A) Soluble fraction; B) Insoluble fraction. Marker: Low Molecular Weight Protein Marker (M). Staining with Coomassie blue.

3.5 Cytotoxicity Assays

Preliminary assays using PBS as cell medium and protein vehicle resulted in high cytotoxicity (similar or even higher than the H_2O_2 control) in 10-min bioassays, which led to the assay optimization by performing tests with different buffers and exposure times. Of the four buffers tested, the one that demonstrated a more regular curve between cytotoxicity and exposure time during the assays was Kenny's Salt Solution. As represented in Figure 3.8, this buffer showed a cytotoxicity of around 0 % until 30 minutes of exposure, while after 42.64 minutes of exposure, the cytotoxicity was half the maximum value. According to these results, the cytotoxicity of Hdiv_1 was then tested using cell suspensions in Kennys' Salt Solution and an exposure time of 30 minutes.

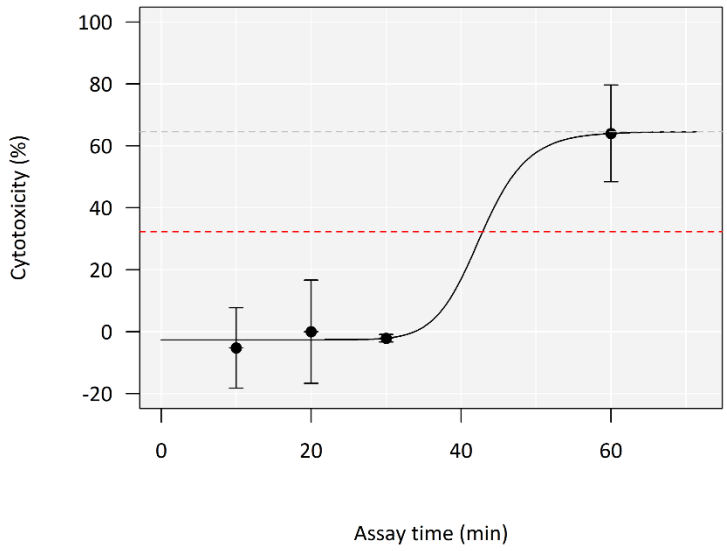

Figure 3.8 – Dose-response curve of Kenny's Salt Solution cytotoxicity over time. Results from the neutral red assay optimization attempt. Cytotoxicity of Kenny's Salt Solution in mussel gills cells over exposure time. The half-maximal value for cytotoxicity was estimated at approximately 43 min. Results were obtained from independent experiments. Values were given in mean +- standard deviation. Curve was produced by non-linear estimation through a log-linear fit.

As it was unfeasible to acquire a substantial quantity of highly purified Hdiv_1 for conducting *in vitro* assays, the E4 elution resulting from IMAC was used, henceforth referred to as ExtHdiv_1 (protein extract containing Hdiv_1). The neutral red assay results showed that as the protein concentration of ExtHdiv_1 increased, the cytotoxicity percentage did not show a corresponding rise, remaining below 6 % in all assays. A similar trend was observed with the control ExtP, in which cytotoxicity also remained below 6 % for the different concentrations tested. When comparing all these results and the negative control (cell suspension treated only with Kenny's Salt Solution, revealing around 4 % cytotoxicity), no significant differences were found (Tukey's HSD test, $p = 0.932$) unlike when comparing to the positive control, which showed a cytotoxicity of 86 % (Tukey's HSD test, $p = 5 \times 10^{-11}$).

3.6 Comet Assay

Regardless of the metric analyzed (% DNA in tail or Olive Moment) no significant differences in DNA damage were detected between the treatments ExtHdiv_1 and ExtP and the different concentrations tested (Tukey's HSD test, $p = 0.823$ when comparing % DNA in tail and $p = 0.779$ when comparing Olive Moment), as shown in Figure 3.9.

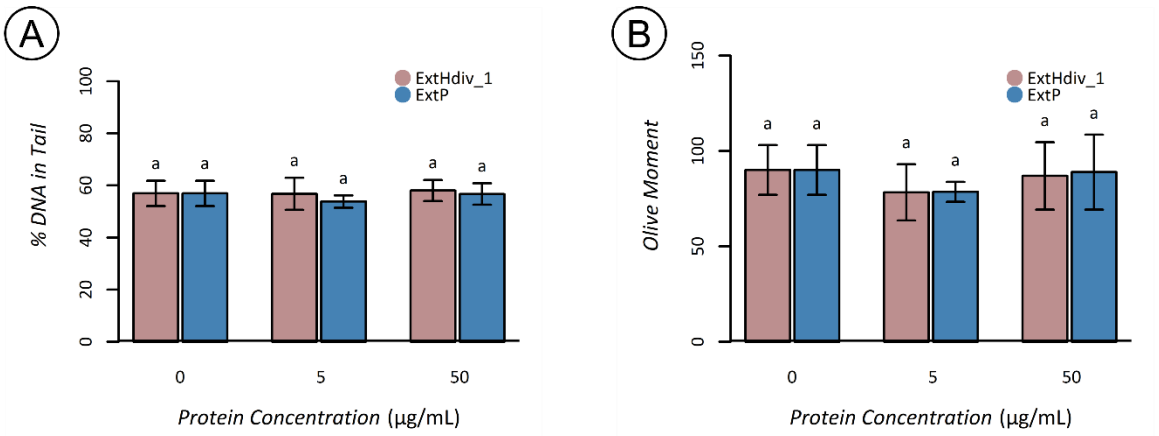

Figure 3.9 – Comet assay results to identify DNA damage. A) Results from Comet assay expressed as % DNA in Tail plus standard deviation, from mussel gills cells exposed to different concentrations of ExtHdiv_1 (extract containing Hdiv_1) and ExtP (extract without Hdiv_1). The control (cells only exposed to Kenny's Salt Solution) is represented by the concentration 0 µg/mL. The letter "a" indicates there are no significant differences between extracts (Tukey's HSD test, $p > 0.05$). B) Results from Comet assay expressed as Olive Moment plus standard deviation, from mussel gills cells exposed to different concentrations of ExtHdiv_1 (extract containing Hdiv_1) and ExtP (extract without Hdiv_1). The control (cells only exposed to Kenny's Salt Solution) is represented by the concentration 0 µg/mL. The letter "a" indicates that there are no significant differences between extracts (Tukey's HSD test, $p > 0.05$). All results were obtained from independent experiments (n=3). Values were given in mean +- standard deviation.

3.7 Microscopy

The results of optical microscopy illustrated in Figure 3.10 revealed the success of cell fixation and impregnation in the resin, even in the presence of a considerable quantity of osmium residues within the samples. Upon microscopy analysis, we discerned what appeared to be viable cells in the samples treated with 50 µg/mL of ExtHdiv_1, 50 µg/mL of ExtP, and with only Kenny's Salt Solution (control) (Figure 3.10A-C). In these cases, it was observed that the cell nuclei and membranes appeared to remain relatively undamaged, with no significant disparities observed among the three treatments. Additionally, these samples seemed heterogeneous with respect to cell and nucleus morphology, as expected, as they should contain different cell types, potentially including hemocytes (cells from invertebrate's immune system) and epithelial cells. Conversely, the sections of the suspension treated with hydrogen peroxide exhibited a substantial accumulation of what seemed to be cellular debris (Figure 3.10D). Moreover, the individual cells identified in this last condition appeared to have nuclear or membrane damage. Overall, cells were seemingly intact (in the first three conditions), which showed that the harvesting procedure and sub-acute toxicity testing was adequate.

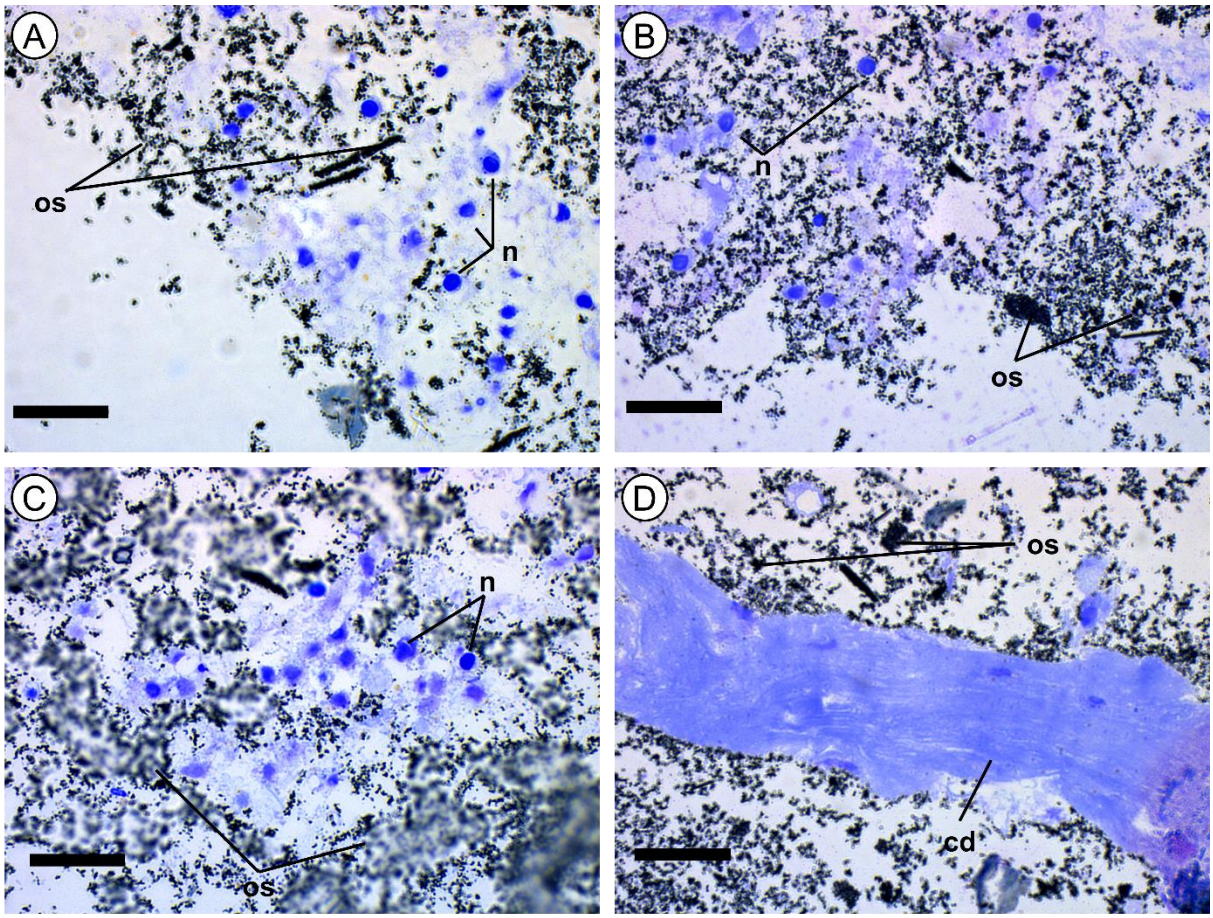

Figure 3.10 – Mussel gill cells exposed to different treatments. A) Resin (EPON) sections of primary mussel gill cells in suspension exposed to 50 µg/mL of ExtHdiv_1. B) Resin (EPON) sections of primary mussel gill cells in suspension exposed to 50 µg/mL of ExtP. C) Resin (EPON) sections of primary mussel gill cells only exposed to Kenny's Salt Solution (control). D) Resin (EPON) sections of primary mussel gill cells in suspension exposed to hydrogen peroxide. Scale corresponds to 25 µm. All sections stained with toluidine blue. n: nuclei; os: osmium residues; cd: cell debris.

4

DISCUSSION

In the last years, multiple successful cases involving the expression of animal toxins using different models, from *Escherichia coli* to mammalian cells, have been reported (see Rivera-de-Torre et al., 2022, for a review). However, the documented cases concerning heterologous expression of marine proteinaceous toxins apart from conotoxins are relatively scarce. Conotoxins have been successfully expressed in *E. coli* (e.g., Yu et al., 2018; Nielsen et al., 2019); nevertheless, these toxins and other marine peptides and small proteins are typically synthesized by solid-phase chemistry (for further details, see Papon et al., 2022). Only recently have been documented the successful heterologous expression of proteins derived from other marine invertebrates, for instance, nicomycins derived from a Polychaeta (Panteleev et al., 2018). Some challenges emerge when dealing with larger and more complex proteins, particularly in attempting to obtain properly folded proteins with a precise disulfide pattern, vital for the activity of a Cys-rich toxin (Zhang et al., 2011). Even in this context, we managed to successfully produce a recombinant protein derived from the cysteine-rich secretory protein from *Hediste diversicolor* (Hdiv_1, 10 cysteines), and additionally, we optimized the conditions for its efficient expression. The experimental conditions selected for protein expression were meticulously chosen. Firstly, both culture media used, Lysogeny Broth (LB) and Terrific Broth (TB), were already reported in the successful expression of other marine proteins (e.g., conotoxins and reflectins, respectively) (see examples in Yu et al., 2018; Nielsen et al., 2019; Phan et al., 2016; Lychko et al., 2023). During the current work, the TB medium emerged as the most adequate for Hdiv_1 expression, probably due to its richer composition compared with LB medium. The TB medium has a higher content of yeast extract than LB and an additional carbon source (glycerol). In contrast, LB medium provides amino acids as the primary source of carbon and energy, which could increase the media pH during catabolism, as previously reported in Kram and Finkel (2015). Moreover, TB medium does not have NaCl in its composition unlike LB, and instead contains potassium phosphate salts that, according to the same authors, may delay the medium alkalization. This TB buffering effect prevents a rapid rise in the pH of the medium, which is critical, as medium alkalization inhibits bacterial growth and accelerates the entry into death phase (Farrell & Finkel, 2003). Furthermore, the induction temperatures used for protein expression were also reported in the literature. The expression at 37ºC and 30ºC was already performed in the heterologous expression of conotoxins, as described in Yu et al. (2018) and Liu et al. (2020). However, it was also

reported that lower temperatures (15ºC – 23ºC) favor the expression of soluble and functional proteins (for further explanation, see Sahdev et al., 2007). This effect can be related for instance to increased stability and correct protein folding patterns at lower temperatures, due to the temperature-dependent nature of the hydrophobic interactions responsible for the formation of inclusion bodies (Strandberg & Enfors, 1991). Taking this information into account, we added a lower expression temperature (18ºC) into our approach, which showed to be the most adequate temperature for Hdiv_1 expression.

The Hdiv_1 product is based on a eukaryotic protein, however, it was expected that a functional protein could be produced using *Escherichia coli* strains, as Hdiv_1 does not appear to require post-translational modifications. The results indicate that Hdiv_1 was expressed using both *E. coli* strains (BL21(DE3) and Origami™ B(DE3)pLysS), but with the BL21(DE3) strain was possible to obtain around four times more soluble protein than using the Origami™ B(DE3)pLysS strain. The BL21(DE3) strain is used as a standard for protein expression, and it was already used in the successful expression of some functional conotoxins (se for instance Yu et al., 2018 and Xia et al., 2006). However, these conotoxins have less cysteines in their composition compared to Hdiv_1, so, unlike them, Hdiv_1 could be non-functional when expressed in the BL21 cytosol due to challenges in its folding, what is in accordance with Stewart, (1998) (e.g., the conotoxins studied in Yu, et al., 2018 and Xia et al., 2006 have 4-6 cysteines, while Hdiv_1 has 10 cysteines). For the production of functional cysteine-rich proteins, their folding must be precise, and it requires the correct formation of disulfide bonds between cysteines. These bonds are formed through oxidation, thus for their establishment in bacteria is necessary to change the reducing environment of the cytosol or send the protein to the periplasm, where the environment is more oxidative (Ma et al., 2020; De Marco, 2009). In this context, the Origami™ B(DE3)pLysS strain has a genetic advantage since it has mutations on reductase genes, contributing to oxidations in the bacterial cytosol. Thereby, in theory, this strain should enhance the expression of soluble and functional disulfide bond-dependent proteins in the bacteria cytosol, as reported in some cases (e.g., Rabhi-Essafi et al., 2007). However, the results presented here do not show evidence of improved expression. In some of their studies, López-Cano et al. (2022) and Nozach et al. (2013) achieved similar results, showing that probably the efficient expression of a soluble protein using the Origami™ B(DE3)pLysS strain is dependent on the expressed protein composition and fusion tag, as hypothesized by the last author.

Western blots enabled us to detect two bands that potentially correspond to Hdiv_1 - at 27 kDa and 30 kDa. Both bands have a higher molecular weight compared with the theoretical value of the fusion protein (22 kDa). This effect also occurred with other authors when studying Cys-rich proteins (Iakoucheva et al., 2001). Iakoucheva and colleagues observed more than one band with close molecular weights in SDS-PAGE gels, all corresponding to the recombinant protein. They further explained that, although denaturation of the protein was induced by boiling the samples with a buffer containing b-mercaptoethanol, the highly oxidative environment in the SDS-PAGE gel (due to the addition of APS) may have led to the formation of disulfide bonds during electrophoresis. On the other hand, they hypothesized that the differences between the theoretical and the experimental molecular weight could

exist due to disordered C-terminal and N-terminal domains. After Hdiv_1 purification attempt by IMAC, the same behavior reported by Iakoucheva and colleagues was observed in SDS-PAGE gel (Figure 3.5) – two bands in the elution E4 (later called ExtHdiv_1). Therefore, exists a possibility of both bands detected during this thesis corresponding to Hdiv_1. To verify this hypothesis and to confirm the production of Hdiv_1, LC-MS/MS analysis was performed. However, the inconclusive results obtained after all the experimental work suggested contamination with proteins intrinsic to *E. coli*. This may be related to the collection of incorrect bands on the 12.5% SDS-PAGE gel due to difficulties in analyzing the gel. In any case, considering the previous Western blot results and the proximity between the pI of the detected bands and the theoretical Hdiv_1 pI, the possibility of one of these bands (or both) corresponding to Hdiv_1 should not be disregarded. Despite the contaminants (other proteins) present in E4 fraction, IMAC allowed an enrichment of 33 % of the band detected around 30 kDa and 7 % of the one detected around 27 kDa. In any case, this method was not as efficient as expected for Hdiv_1 purification, probably related to the inaccessibility of the protein tag to the immobilized metal due to the occlusion of the tag in the protein. To optimize this purification step, the His-tag could be extended to increase its affinity for the resin, allowing washes at high imidazole concentrations to remove non-specifically bound contaminants (Grisshammer & Tucker, 1997). However, a longer tag could cause perturbation in the protein function (Bornhorst & Falke, 2000). Therefore, it would need to be removed after purification, and for that, it would be necessary to include a sequence recognized by a protease between the protein of interest and the tag, as well as an additional purification step to separate the tag from the protein, which can be cumbersome (Pina et al., 2021). Alternatively, another purification step for polishing could be added to achieve a higher purification (e.g., ion exchange or size exclusion chromatography) common strategy for the purification of therapeutic proteins (Sánchez-Trasviña et al., 2021). In fact, we tested SEC, which resulted in one of the bands (the one around 30 kDa) with a purity of 86 %, so including this strategy in the purification workflow could be crucial. However, for functional characterization of Hdiv_1, we used the samples from E4 (sample ExtHdiv_1), since it contained the two bands of interest and the extract obtained after SEC had a lower protein concentration. As a control, we used a crude extract containing all soluble proteins intrinsic to *E. coli* BL21(DE3) when plasmid codifying for Hdiv_1 is absent (sample ExtP). The ExtP control contains all the proteins that may represent contaminants in the ExtHdiv_1, so consequently the sample ExtHdiv_1 was considered of sufficient quality to be used in the *in vitro* cell-based assays.

The cytotoxic potential of Hdiv_1 was first evaluated through the neutral red assay, in which the ability of viable cells to incorporate the supravital dye neutral red in the lysosomes is used to estimate the percentage of viable cells in the culture (Repetto et al., 2008). The results showed that any of the extracts (ExtHdiv_1 and ExtP) or different protein concentrations tested (1 µg/mL, 5 µg/mL, 25 µg/mL, and 50 µg/mL) seemed to have a significant cytotoxic effect on mussel gill cells under the assessment circumstances. In addition, the comet assay also showed that both extracts did not seem to cause DNA damage. The metrics analyzed correlated with DNA damage: the percentage of DNA in the tail (% DNA in tail) and Olive moment (distance between the centers of the head and tail times the rate of DNA in

the tail) showed the same pattern when comparing any extract with the negative control (cells treated only with the suspension buffer). For more information on the comet metrics, see Hong et al., (2020). In the other control, where the cells were treated with hydrogen peroxide, comets were not detected, possibly due to cell death. The results of the cytotoxicity assays were corroborated by microscopic analysis, in which Hdiv_1 (50 µg/mL) did not appear to have more harmful effects on cells than the negative control used (Kenny's Salt Solution). All these results seem to indicate that Hdiv_1 does not cause cytotoxic effects, which could have several possible explanations. Firstly, it is worth considering that the effect of Hdiv_1 might be obscured by contaminants present in the extract (ExtHdiv_1). The presence of contaminants in the extract causes the dilution of Hdiv_1, and consequently, the Hdiv_1 effect may have been reduced. Maybe distinct results could have been obtained if we had used the extract resulting from size exclusion chromatography since this sample boasts an approximate 86 % purity for one of the bands of interest. On the other hand, the function of this protein may not be associated with a damaging effect on cells. Although cytotoxic proteins have recently been identified in the venom of another polychaeta (*E. viridis*), reported in Rodrigo et al. (2021a), Hdiv_1 may have other role. For example, as a CRISP, Hdiv_1 has some potential to act as a neurotoxin since other CRISPs have already been reported as such (see for instance Yamazaki et al., 2002). In this way, the effects of Hdiv_1 would only be visible when evaluated in whole organisms, analyzing the impact on the nervous system. Besides its neurotoxic potential, this protein may be more related to a defensive role, according to Moutinho Cabral et al., (2022). As mentioned by Moutinho Cabral and colleagues, the *Hediste diversicolor* itself appears to produce more bioactives associated with defense mechanisms than predation. In this context, the Hdiv_1 potential to act as an antimicrobial should be considered and studied in future approaches. If indeed this protein proves to be valuable as a neurotoxin or antimicrobial compound, it is advantageous not to cause death, mutations, or other types of damage in eukaryotic cells. Another hypothesis that must be considered is the possibility that the produced Hdiv_1 may not be active. The actual folding of the produced protein was not studied in this project (due to time constraints), so although the protein has a three-dimensional structure (since it was obtained in the soluble fraction), this does not prove that it holds its native folding, which it is essential for Cys-rich proteins activity (see for instance, Nielsen et al., 2019). This is particularly relevant due to the large number of cysteines in the composition of Hdiv_1. The disulfide bonds between these cysteines are essential for the correct folding and consequently activity of the protein (Zhang et al., 2011). Knowing that there are some struggles for these bonds formation in the BL21(DE3) cytosol (Stewart, 1998), the ones that were formed in Hdiv_1 could be incorrect, leading to the expression of a dysfunctional protein. In addition, it is not yet known whether any cofactor or specific protein interaction, naturally present in the annelid venom, would be necessary for this protein activity and stability (see an example of a cofactor essential for protein activity in Gierula and Ahnström, 2020). If this is the case, this factor must be present in the same extract as Hdiv_1 to induce its activity. Thus, further studies will be necessary to test these hypotheses, namely characterization of secondary structure by circular dichroism. However, regardless of which hypothesis will be verified in the future, the potential for new drug development of the cysteine-rich secretory protein of *H. diversicolor* or its derived Hdiv_1 should not be discarded.

The conditions used in the cytotoxicity assays may not have been the most adequate to obtain significant cytotoxicity, which could also explain the cytotoxicity lack of Hdiv_1. In the first attempts of the neutral red assay, the results did not seem very consistent since appeared to exist less cytotoxicity when the cells were treated with hydrogen peroxide compared to all other treatments, including the negative control. This makes no sense considering that H_2O_2 was being used as a positive control, since it was already reported as potentially cytotoxic (Eastman & Guilarte, 1990). Consequently, the assay was re-optimized before testing Hdiv_1-containing extracts. Thereby, different buffers were tested in the preparation of the suspension, as well as different exposure times of the cells to the treatment. In fact, the difficulties previously experienced with this assay seemed to be related to the suspension buffer. Among the buffers tested was Dulbecco's PBS (calcium and magnesium-free), the one prepared in the laboratory that had previously been used in this assay and also a commercial version. Both had similar compositions, with only a slight change in sodium phosphate concentration. Hank's Salt Solution has the same constituents as PBS with the addition of glucose and sodium bicarbonate. The last buffer tested, Kenny's Salt Solution, has a higher NaCl and KCl concentration than the previous buffers and does not have sodium phosphates in its composition. Both Dulbecco's PBS and Hank's Salt Solution are commonly used in mammalian cell cultures (Repetto et al., 2008; Osterberg & Roussa, 2009). Kenny's Salt Solution is not as widely-used as the former, but it has also been referenced in comet assays using fish liver cells, for instance (Amado et al., 2006). The results of this optimization attempt showed that Kenny's Salt Solution seemed to be the most suitable buffer for these cell cultures. Additionally, it is perhaps the most interesting buffer for cultures of marine organisms with an inefficient osmotic balance maintenance system since its composition is similar to the saline composition of salty water due to its large proportion of NaCl. The remaining buffers have a lower salt concentration, which could lead to osmotic stress in the mussel cells or even cause cytolysis during the assay (lysis due to osmotic disbalance). On the other hand, some potential neutral red precipitates were detected when Hank's Salt Solution was used, so some compounds present in this buffer may cause neutral red precipitation during the assay, leading to inconsistent results. Optimizing this protocol also allowed to increase the exposure time of the cells to the treatments compared to what was initially being used. This also makes the assay more sensitive in detecting cytotoxic effects, as proteins have more time to exert their effect on cells. However, we also have to consider that the different concentrations of total protein used and the time of cells' exposure to the protein could not have been enough to trigger Hdiv_1 effect. Notwithstanding, it was possible to successfully optimize this strategy, converting it into a quick and practical tool for cytotoxicity screening.

The culture of primary mussel gill cells used in the assays may also have influenced the cytotoxicity results. Mollusks such as mussels are an attractive model organism for toxicity studies. They are particularly suitable for ecotoxicological studies of pharmaceuticals and other contaminants since they can be used to monitor the presence and changes in concentration of compounds in the ocean (Świacka et al., 2019). In fact, these marine animals have some characteristics that make them useful for toxicological studies, for instance, their capability to bio-accumulate different xenobiotics like pesticides, drugs,

heavy metals, and toxins (Ozkoc et al., 2007; Solla et al., 2016; Blanco, 2018; Pempkowiak et al., 1999). In addition, *Mytilus edulis* (the species used in this project), has a sedentary behavior, a high distribution on the Portuguese coast, and a great resistance and adaptation capacity (Sukhotin et al., 2003). Moreover, the collection, maintenance, and manipulation of these mollusks are easy and low-cost, and toxicity assays on these animals (*in vivo*) or on their cells (*in vitro*) are also easy to perform, cost-effective, and relatively fast (Curpan et al., 2022; Ladhar-Chaabouni & Hamza-Chaffai, 2016). In this specific project, cultures of primary cells derived from mussel gills were used to assess Hdiv_1 toxicity. This type of cells was selected after some histological studies performed by the Seatox team showed that the gills of the mussels collected had fewer parasites and less cell damage when extracted compared to other tissues (e.g., digestive glands). It is important to note that the primary cell population used is heterogeneous, presenting different cell types, from epithelial cells to defense cells, as reported in Gómez-Mendikute et al. (2005). The microscopic analyses performed show this same panorama, as several cells with different appearances were detected, many of which we believe to be defense cells. Therefore, the heterogeneity of the culture and the presence of numerous defense cells in it may also have influenced the effects detected on the cytotoxic assays. Nonetheless, the cytopathological alterations caused by exposure to the recombinant protein could not be assessed in detail during the present work and will in the near future require further sectioning and transmission electron microscopy to determine, e.g., cell death type and alterations to the plasma membrane and mitochondria.

5

CONCLUSIONS AND FUTURE PRESPECTIVES

Notwithstanding the encountered limitations, this research successfully yielded a recombinant marine Polychaeta protein from its coding sequence using a highly cost-effective and practical model for such endeavors. Additionally, this thesis made strides in assessing the bioactivity of the recombinant protein, revealing that it does not exhibit a significant level of cytotoxicity toward animal cells, although its effects on vertebrates remain uncharted territory. This information will be valuable for guiding future investigations into the bioactivity of Hdiv_1. Furthermore, the work conducted in this project made a considerable technical contribution, as the employed strategy holds the potential to pave the way for various applications. An efficient and cost-effective cytoassay was also refined throughout this endeavor, offering a valuable tool for initial bioactivity screenings of diverse compounds of interest. In summary, this thesis plays a pivotal role in surmounting multiple challenges within blue biotechnology.

In brief, marine bioactives have exhibited great biotechnological potential in recent years, serving as a secure and cost-effective resource for developing novel products with advantageous and economically viable properties across various domains. Therefore, marine toxins have garnered considerable attention from the scientific community, stemming from their potential to yield new, safer, environmentally friendly, and efficacious drugs. The Polychaeta, a class of marine invertebrates, have revealed a great interest in this matter. Their wide distribution has compelled the species within this class to develop different compensatory mechanisms to address inefficient predation, defense and immune systems, such as the secretion of venoms, complex mixtures of toxins. However, more information about these annelids or the properties of the toxins they secrete is still needed to understand their biotechnological potential. To investigate the biotechnological potential of marine proteins, the recombinant protein known as Hdiv_1, derived from the cysteine-rich secretory protein secreted by *Hediste diversicolor*, was expressed for the first time. The protein was partially characterized by its experimental molecular weight and isoelectric point. Although some uncertainty persists regarding which of the two detected bands corresponds to Hdiv_1 in SDS-PAGE gels, both bands were successfully concentrated in the same extract, which was obtained in enough amounts to perform the cell-based assays. Efforts were also made to obtain a purer form of Hdiv_1, resulting in high purity for one of the bands of interest but in limited quantities for cell-based assays. Furthermore, Hdiv_1 did not exhibit cytotoxic effects, nor did it

induce DNA damage on primary cultures of cells from mussel gills. This observation could be attributed to various factors, including the possibility that the protein has not evolved to cause significant cytotoxicity in eukaryote cells, especially mollusks cells, which are not its intended target. Nevertheless, the potential of the cysteine-rich secretory protein for developing new drugs should not be underestimated.

In the future, the mass spectrometry analysis should be repeated and the Hdiv_1 folding should be studied through circular dichroism to elucidate whether the recombinant protein produced assumes a specific secondary structure pattern. Alternatively, we could study the folding *in silico* with structures of other Cys-rich proteins and conotoxins, already available in public databases, or explore the potential of AlphaFold as an AI tool to predict the folding of Hdiv_1. If the expressed protein is indeed folded, additional assays should be conducted to assess the bioactivity of the recombinant protein and, consequently, its biotechnological potential, which may involve repeating cytotoxicity assays with increased protein concentrations and further cytopathological analysis. Furthermore, it would be prudent to explore the potential neurotoxic and antimicrobial function of the protein through assays involving animal models and bacterial cells, respectively. Conversely, if the produced protein is found to be dysfunctional, modifications in the production and purification workflow will be necessary. Regarding expression, an alternative *E. coli* strain, such as SHuffle® Express Competent *E. coli*, could be considered. Like Origami™ B(DE3)pLysS, this strain possesses a genotype favorable to forming disulfide bonds in *E. coli* cytosol. Specifically, SHuffle stain constitutively expresses an isomerase capable of facilitating the formation of correct disulfide bonds between cysteine residues. Another approach could involve isolating the protein from the insoluble fraction rather than the soluble one. This strategy is noteworthy as successful cases where functional proteins were obtained from inclusion bodies have been reported (see an example in Maksum et al., 2022). To obtain a functional protein the inclusion bodies would need to be treated with different buffers. First, it would be necessary to solubilize them and then induce protein refolding. This method is more time-consuming but may lead to the correct protein conformation. Overall, the key to a comprehensive understanding of the cysteine-rich secretory protein hinges on further investigations into its folding and potential functions. Finally, Hdiv_1 purification could be optimized to obtain a higher yield by attempting other methodologies, e.g., ion exchange chromatography. Moreover, although the bioactivity or potential toxicity of Hdiv_1 remains uncertain, this project successfully produced a CRISP in both soluble and insoluble forms (in inclusion bodies) via a rapid and cost-effective methodology: recombinant DNA technology. Demonstrating how the feasibility of this production approach for complex marine proteins, such as CRISPs, holds significance in the future of blue biotechnology.

REFERENCES

Amado, L. L., Rosa, C. E. D., Leite, A. M., Moraes, L., Pires, W. V., Pinho, G. L. L., Martins, C. M. G., Robaldo, R. B., Nery, L. E. M., Monserrat, J. M., Bianchini, A., Martinez, P. E., & Geracitano, L. A. (2006). Biomarkers in croakers *Micropogonias furnieri* (Teleostei: Sciaenidae) from polluted and non-polluted areas from the Patos Lagoon estuary (Southern Brazil): Evidences of genotoxic and immunological effects. *Marine Pollution Bulletin*, 52(2), 199–206. https://doi.org/10.1016/j.marpolbul.2005.11.006

Blanco J. (2018). Accumulation of dinophysis toxins in bivalve molluscs. *Toxins*, 10(11), 453. https://doi.org/10.3390/toxins10110453

Bon, C., Saliou, B., Thieffry, M., Manaranche, R. (1985). Partial purification of α-glycerotoxin, a presynaptic neurotoxin from the venom glands of the polychaete annelid *Glycera convoluta*, *Neurochemistry International,* 7(1), 63–75. https://doi.org/10.1016/0197-0186(85)90009-9

Bornhorst, J. A., & Falke, J. J. (2000). Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology*, 326, 245–254. https://doi.org/10.1016/s0076-6879(00)26058-8

Bryan, G. W., & Gibbs, P. E. (1979). Zinc—A major inorganic component of nereid polychaete jaws. *Journal of the Marine Biological Association of the United Kingdom*, 59(4), 969–973. https://doi.org/10.1017/S0025315400036961

Burgess, J. G. (2012). New and emerging analytical techniques for marine biotechnology. *Current Opinion in Biotechnology*, 23(1), 29–33. https://doi.org/10.1016/j.copbio.2011.12.007

Collins, K. D., Gensch, T., & Glorius, F. (2014). Contemporary screening approaches to reaction discovery and development. *Nature Chemistry*, 6(10), 859–871. https://doi.org/10.1038/nchem.2062

Curpan, A.-S., Impellitteri, F., Plavan, G., Ciobica, A., & Faggio, C. (2022). Review: *Mytilus galloprovincialis*: An essential, low-cost model organism for the impact of xenobiotics on oxidative stress and public health. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 256, 109302. https://doi.org/10.1016/j.cbpc.2022.109302

Costa, P. M. (2018). *The Handbook of Histopathological Practices in Aquatic Environments: Guide to Histology for Environmental Toxicology*. New York: Academic Press. https://doi.org/10.1016/C2016-0-01962-9

De Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microbial Cell Factories*, 8(1), 26. https://doi.org/10.1186/1475-2859-8-26

Dias, A.M.G.C. and Roque, A.C.A. (2017), The future of protein scaffolds as affinity reagents for purification. *Biotechnol. Bioeng*., 114: 481-491. https://doi.org/10.1002/bit.26090

Dos Santos, C., & Franco, O. L. (2023). Pathogenesis related proteins (PRs) with enzyme activity activating plant defense responses*. Plants*, 12(11), 2226. https://doi.org/10.3390/plants12112226

Eastman, C. L., & Guilarte, T. R. (1990). The role of hydrogen peroxide in the *in vitro* cytotoxicity of 3-hydroxykynurenine. *Neurochemical Research*, 15(11), 1101–1107. https://doi.org/10.1007/BF01101711

Fairbairn, D. W., Olive, P. L., & O'Neill, K. L. (1995). The comet assay: A comprehensive review. *Mutation Research/Reviews in Genetic Toxicology*, 339(1), 37–59. https://doi.org/10.1016/0165-1110(94)00013-3

Farrell, M. J., & Finkel, S. E. (2003). The growth advantage in stationary-phase phenotype conferred by rpoS mutations is dependent on the pH and nutrient environment. *Journal of Bacteriology*, 185(24), 7044–7052. https://doi.org/10.1128/JB.185.24.7044-7052.2003

Faucet, J., Maurice, M., Gagnaire, B., Renault, T., & Burgeot, T. (2004). Isolation and primary culture of gill and digestive gland cells from the common mussel *Mytilus edulis*. *Methods in Cell Science*, 25(3–4), 177–184. https://doi.org/10.1007/s11022-004-8227-4

Fox, J. W., & Serrano, S. M. (2008). Exploring snake venom proteomes: multifaceted analyses for complex toxin mixtures. *Proteomics*, 8(4), 909–920. https://doi.org/10.1002/pmic.200700777

Freitas, A. C., Rodrigues, D., Rocha-Santos, T. A. P., Gomes, A. M. P., & Duarte, A. C. (2012). Marine biotechnology advances toward applications in new functional foods. *Biotechnology Advances*, 30(6), 1506–1515. https://doi.org/10.1016/j.biotechadv.2012.03.006

Fry, B. G., Roelants, K., Champagne, D. E., Scheib, H., Tyndall, J. D. A., King, G. F., Nevalainen, T. J., Norman, J. A., Lewis, R. J., Norton, R. S., Renjifo, C., & De La Vega, R. C. R. (2009). The toxicogenomic multiverse: Convergent recruitment of proteins into animal venoms. *Annual Review of Genomics and Human Genetics*, 10(1), 483–511. https://doi.org/10.1146/annurev.genom.9.081307.164356

Gierula, M., & Ahnström, J. (2020). Anticoagulant protein S-New insights on interactions and functions. *Journal of Thrombosis and Haemostasis*: JTH, 18(11), 2801–2811. https://doi.org/10.1111/jth.15025

Glasby, C. J., & Timm, T. (2008). Global diversity of polychaetes (Polychaeta; Annelida) in freshwater. *Hydrobiologia*, 595(1), 107–115. https://doi.org/10.1007/s10750-007-9008-2

Gómez-Mendikute, A., Elizondo, M., Venier, P., & Cajaraville, M. P. (2005). Characterization of mussel gill cells *in vivo* and *in vitro*. *Cell and Tissue Research*, 321(1), 131–140. https://doi.org/10.1007/s00441-005-1093-9

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(3), 315–325. https://doi.org/10.1007/s00418-020-01889-3

Grisshammer, R., & Tucker, J. (1997). Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein Expression and Purification*, 11(1), 53–60. https://doi.org/10.1006/prep.1997.0766

Haque, S. M., & Ratemi, E. S. (2017). Drug development and analysis review. *Pharmaceutical Chemistry Journal*, *50*(12), 837–850. https://doi.org/10.1007/s11094-017-1543-1

Hausen, H. (2005). Comparative structure of the epidermis in polychaetes (Annelida). *Hydrobiologia* 535, 25–35. https://doi.org/10.1007/s10750-004-4442-x

Heyland, A., Plachetzki, D., Donelly, E., Gunaratne, D., Bobkova, Y., Jacobson, J., Kohn, A. B., & Moroz, L. L. (2012). Distinct expression patterns of glycoprotein hormone subunits in the lophotrochozoan *Aplysia*: implications for the evolution of neuroendocrine systems in animals. *Endocrinology*, 153(11), 5440–5451. https://doi.org/10.1210/en.2012-1677

Hong, Y., Han, H.-J., Lee, H., Lee, D., Ko, J., Hong, Z., Lee, J.-Y., Seok, J.-H., Lim, H. S., Son, W.-C., & Sohn, I. (2020). Deep learning method for comet segmentation and comet assay image analysis. *Scientific Reports*, 10(1), 18915. https://doi.org/10.1038/s41598-020-75592-7

Hutchings, P. (1998). Biodiversity and functioning of polychaetes in benthic sediments. *Biodiversity and Conservation*, 7(9), 1133–1145. https://doi.org/10.1023/A:1008871430178

Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Smith, R. D., Dunker, A. K., & Ackerman, E. J. (2001). Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Science: A Publication of the Protein Society*, 10(7), 1353–1362. https://doi.org/10.1110/ps.ps.40101

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

Inceoglu, B., Lango, J., Jing, J., Chen, L., Doymaz, F., Pessah, I. N., & Hammock, B. D. (2003). One scorpion, two venoms: prevenom of *Parabuthus transvaalicus* acts as an alternative type of venom with distinct mechanism of action. *Proceedings of the National Academy of Sciences of the United States of America*, 100(3), 922–927. https://doi.org/10.1073/pnas.242735499

Jin, A.-H., Muttenthaler, M., Dutertre, S., Himaya, S. W. A., Kaas, Q., Craik, D. J., Lewis, R. J., & Alewood, P. F. (2019). Conotoxins: Chemistry and biology. *Chemical Reviews*, 119(21), 11510–11549. https://doi.org/10.1021/acs.chemrev.9b00207

Khrunyk, Y., Lach, S., Petrenko, I., & Ehrlich, H. (2020). Progress in modern marine biomaterials research. *Marine Drugs*, *18*(12). https://doi.org/10.3390/md18120589

Kim, B. Y., Lee, K. S., Zou, F. M., Wan, H., Choi, Y. S., Yoon, H. J., Kwon, H. W., Je, Y. H., & Jin, B. R. (2013). Antimicrobial activity of a honeybee (*Apis cerana*) venom Kazal-type serine protease inhibitor. *Toxicon: Official Journal of the International Society on Toxinology*, 76, 110–117. https://doi.org/10.1016/j.toxicon.2013.09.017

Kram, K. E., & Finkel, S. E. (2015). Rich medium composition affects *Escherichia coli* survival, glycation, and mutation frequency during long-term batch culture. *Applied and Environmental Microbiology*, 81(13), 4442–4450. https://doi.org/10.1128/AEM.00722-15

Ladhar-Chaabouni, R., & Hamza-Chaffai, A. (2016). The cell cultures and the use of haemocytes from marine molluscs for ecotoxicology assessment. *Cytotechnology*, 68(5), 1669–1685. https://doi.org/10.1007/s10616-015-9932-3

Leader, B., Baca, Q. J., & Golan, D. E. (2008). Protein therapeutics: A summary and pharmacological classification. *Nature Reviews Drug Discovery*, *7*(1), 21–39. https://doi.org/10.1038/nrd2399

Liang, X., Luo, D., & Luesch, H. (2019). Advances in exploring the therapeutic potential of marine natural products. *Pharmacological Research*, *147*, 104373. https://doi.org/10.1016/j.phrs.2019.104373

Lillebø, A. I., Pita, C., Garcia Rodrigues, J., Ramos, S., & Villasante, S. (2017). How can marine ecosystem services support the Blue Growth agenda? *Marine Policy*, 81, 132–142. https://doi.org/10.1016/j.marpol.2017.03.008

Liu, Y., Yin, Y., Song, Y., Wang, K., Wu, F., & Jiang, H. (2020). α-Conotoxin as potential to α7-nAChR recombinant expressed in *Escherichia coli*. *Marine Drugs*, 18(8), 422. https://doi.org/10.3390/md18080422

López-Cano, A., Martínez-Miguel, M., Guasch, J., Ratera, I., Arís, A., & Garcia-Fruitós, E. (2022). Exploring the impact of the recombinant *Escherichia coli* strain on defensins antimicrobial activity: BL21 versus Origami strain. *Microbial Cell Factories*, 21(1), 77. https://doi.org/10.1186/s12934-022-01803-7

Lychko, I., Soares, C. L., Dias, A. M. G. C., & Roque, A. C. A. (2023). A scalable method to purify reflectins from inclusion bodies. *Separation and Purification Technology*, 315, 123736. https://doi.org/10.1016/j.seppur.2023.123736

Ma, Y., Lee, C.-J., & Park, J.-S. (2020). Strategies for optimizing the production of proteins and peptides with multiple disulfide bonds. *Antibiotics*, 9(9), 541. https://doi.org/10.3390/antibiotics9090541

Maksum, I. P., Yosua, Y., Nabiel, A., Pratiwi, R. D., Sriwidodo, S., & Soedjanaatmadja, U. M. S. (2022). Refolding of bioactive human epidermal growth factor from *E. coli* BL21(DE3) inclusion bodies & evaluations on its *in vitro* & *in vivo* bioactivity. *Heliyon*, *8*(4), e09306. https://doi.org/10.1016/j.heliyon.2022.e09306

Martins, A., Vieira, H., Gaspar, H., & Santos, S. (2014). Marketed marine natural products in the pharmaceutical and cosmeceutical industries: Tips for success. *Marine Drugs*, 12(2), 1066–1101. https://doi.org/10.3390/md12021066

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(3), 220–228. https://doi.org/10.1089/zeb.2020.1857

Mayer, A. M. S., Glaser, K. B., Cuevas, C., Jacobs, R. S., Kem, W., Little, R. D., McIntosh, J. M., Newman, D. J., Potts, B. C., & Shuster, D. E. (2010). The odyssey of marine pharmaceuticals: A current pipeline perspective. *Trends in Pharmacological Sciences*, 31(6), 255–265. https://doi.org/10.1016/j.tips.2010.02.005

Meunier, F. A. (2002). Glycerotoxin from *Glycera convoluta* stimulates neurosecretion by up-regulating N-type Ca^{2+} channel activity. *The EMBO Journal*, 21(24), 6733–6743. https://doi.org/10.1093/emboj/cdf677

Michel, C., & Keil, B. (1975). Biologically active proteins in the venomous glands of the polychaetous annelid, *Glycera convoluta* keferstein. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 50(1), 29–33. https://doi.org/10.1016/0305-0491(75)90294-1

Modica, M. V., Lombardo, F., Franchini, P., & Oliverio, M. (2015). The venomous cocktail of the vampire snail *Colubraria reticulata* (*Mollusca, Gastropoda*). *BMC Genomics*, 16(1), 441. https://doi.org/10.1186/s12864-015-1648-4

Molinski, T. F., Dalisay, D. S., Lievens, S. L., & Saludes, J. P. (2009). Drug development from marine natural products. *Nature Reviews Drug Discovery*, 8(1), 69–85. https://doi.org/10.1038/nrd2487

Moses, D. N., Mattoni, M. A., Slack, N. L., Waite, J. H., & Zok, F. W. (2006). Role of melanin in mechanical properties of *Glycera* jaws. *Acta Biomaterialia*, 2(5), 521–530. https://doi.org/10.1016/j.actbio.2006.05.002

Moutinho Cabral, I., Madeira, C., Grosso, A. R., & Costa, P. M. (2022). A drug discovery approach based on comparative transcriptomics between two toxin-secreting marine annelids: *Glycera alba* and *Hediste diversicolor. Molecular Omics*, 18(8), 731–744. https://doi.org/10.1039/d2mo00138a

Nakabayashi, K., Matsumi, H., Bhalla, A., Bae, J., Mosselman, S., Hsu, S. Y., & Hsueh, A. J. W. (2002). Thyrostimulin, a heterodimer of two new human glycoprotein hormone subunits, activates the thyroid-stimulating hormone receptor. *Journal of Clinical Investigation*, 109(11), 1445–1452. https://doi.org/10.1172/JCI0214340

Nielsen, L. D., Foged, M. M., Albert, A., Bertelsen, A. B., Søltoft, C. L., Robinson, S. D., Petersen, S. V., Purcell, A. W., Olivera, B. M., Norton, R. S., Vasskog, T., Safavi-Hemami, H., Teilum, K., & Ellgaard, L. (2019). The three-dimensional structure of an H-superfamily conotoxin reveals a granulin fold arising from a common ICK cysteine framework. *The Journal of Biological Chemistry*, 294(22), 8745-8759. https://doi.org/10.1074/jbc.RA119.007491

Nozach, H., Fruchart-Gaillard, C., Fenaille, F., Beau, F., Ramos, O. H. P., Douzi, B., Saez, N. J., Moutiez, M., Servent, D., Gondry, M., Thaï, R., Cuniasse, P., Vincentelli, R., & Dive, V. (2013). High throughput screening identifies disulfide isomerase DsbC as a very efficient partner for recombinant expression of small disulfide-rich proteins in *E. coli*. *Microbial Cell Factories*, 12(1), 37. https://doi.org/10.1186/1475-2859-12-37

Ntie-Kang, F., & Svozil, D. (2020). An enumeration of natural products from microbial, marine and terrestrial sources. *Physical Sciences Reviews*, 5(8), 20180121. https://doi.org/doi:10.1515/psr-2018-0121

Olive, P. L., Banáth, J. P., & Durand, R. E. (2012). Heterogeneity in radiation-induced DNA damage and repair in tumor and normal cells measured using the "comet" assay. 1990. *Radiation Research*, 178(2), AV35–AV42. https://doi.org/10.1667/rrav04.1

Osterberg, N., & Roussa, E. (2009). Characterization of primary neurospheres generated from mouse ventral rostral hindbrain. *Cell and Tissue Research*, 336(1), 11–20. https://doi.org/10.1007/s00441-008-0743-0

Ozkoc, H. B., Bakan, G., & Ariman, S. (2007). Distribution and bioaccumulation of organochlorine pesticides along the Black Sea coast. *Environmental Geochemistry and Health*, 29(1), 59–68. https://doi.org/10.1007/s10653-006-9064-y

Panteleev, P. V., Tsarev, A. V., Bolosov, I. A., Paramonov, A. S., Marggraf, M. B., Sychev, S. V., Shenkarev, Z. O., & Ovchinnikova, T. V. (2018). Novel antimicrobial peptides from the arctic Polychaeta *Nicomache minor* provide new molecular insight into biological role of the BRICHOS domain. *Marine drugs*, 16(11), 401. https://doi.org/10.3390/md16110401

Papon, N., Copp, B. R., & Courdavault, V. (2022). Marine drugs: Biology, pipelines, current and future prospects for production. *Biotechnology Advances*, 54, 107871. https://doi.org/10.1016/j.biotechadv.2021.107871

Pempkowiak, J., Sikora, A., & Biernacka, E. (1999). Speciation of heavy metals in marine sediments vs their bioaccumulation by mussels. *Chemosphere*, 39(2), 313–321. https://doi.org/10.1016/S0045-6535(99)00112-5

Phan, L., Kautz, R., Arulmoli, J., Kim, I. H., Le, D. T., Shenk, M. A., Pathak, M. M., Flanagan, L. A., Tombola, F., & Gorodetsky, A. A. (2016). Reflectin as a material for neural stem cell growth. *ACS Applied Materials & Interfaces*, 8(1), 278–284. https://doi.org/10.1021/acsami.5b08717

Pina, A. S., Batalha, Í. L., Dias, A. M. G. C., & Roque, A. C. A. (2021). Affinity tags in protein purification and peptide enrichment: An overview. *Methods in Molecular Biology* (Clifton, N.J.), 2178, 107—132. https://doi.org/10.1007/978-1-0716-0775-6_10

Rabhi-Essafi, I., Sadok, A., Khalaf, N., & Fathallah, D. M. (2007). A strategy for high-level expression of soluble and functional human interferon alpha as a GST-fusion protein in *E. coli*. *Protein Engineering, Design & Selection*: PEDS, 20(5), 201–209. https://doi.org/10.1093/protein/gzm012

Repetto, G., Del Peso, A., & Zurita, J. L. (2008). Neutral red uptake assay for the estimation of cell viability/cytotoxicity. *Nature Protocols*, 3(7), 1125–1131. https://doi.org/10.1038/nprot.2008.75

Rivera-de-Torre, E., Rimbault, C., Jenkins, T. P., Sørensen, C. V., Damsbo, A., Saez, N. J., Duhoo, Y., Hackney, C. M., Ellgaard, L., & Laustsen, A. H. (2022). Strategies for heterologous expression, synthesis, and purification of animal venom toxins. *Frontiers in Bioengineering and Biotechnology*, 9, 811905. https://doi.org/10.3389/fbioe.2021.811905

Robinson, S. D., Li, Q., Bandyopadhyay, P. K., Gajewiak, J., Yandell, M., Papenfuss, A. T., Purcell, A. W., Norton, R. S., & Safavi-Hemami, H. (2017). Hormone-like peptides in the venoms of marine cone snails. *General and Comparative Endocrinology*, 244, 11–18. https://doi.org/10.1016/j.ygcen.2015.07.012

Rodrigo, A. P., & Costa, P. M. (2019). The hidden biotechnological potential of marine invertebrates: The Polychaeta case study. *Environmental Research*, 173, 270–280. https://doi.org/10.1016/j.envres.2019.03.048

Rodrigo, A. P., Grosso, A. R., Baptista, P. V., Fernandes, A. R., & Costa, P. M. (2021b). A transcriptomic approach to the recruitment of venom proteins in a marine annelid. *Toxins*, 13(2), 97. https://doi.org/10.3390/toxins13020097

Rodrigo, A. P., Martins, C., Costa, M. H., Alves De Matos, A. P., & Costa, P. M. (2018). A morpho-anatomical approach to the adaptive features of the epidermis and proboscis of a marine Polychaeta: *Eulalia viridis* (Phyllodocida: Phyllodocidae). *Journal of Anatomy*, 233(5), 567–579. https://doi.org/10.1111/joa.12870

Rodrigo, A. P., Mendes, V. M., Manadas, B., Grosso, A. R., Alves De Matos, A. P., Baptista, P. V., Costa, P. M., & Fernandes, A. R. (2021a). Specific antiproliferative properties of proteinaceous toxin secretions from the marine annelid *Eulalia* sp. onto ovarian cancer cells. *Marine Drugs*, 19(1), 31. https://doi.org/10.3390/md19010031

Sahdev, S., Khattar, S. K., & Saini, K. S. (2007). Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, 307(1–2), 249–264. https://doi.org/10.1007/s11010-007-9603-6

Sala, E., & Knowlton, N. (2006). Global marine biodiversity trends. *Annual Review of Environment and Resources*, 31(1), 93–122. https://doi.org/10.1146/annurev.energy.31.020105.100235

Sánchez-Trasviña, C., Flores-Gatica, M., Enriquez-Ochoa, D., Rito-Palomares, M., & Mayolo-Deloisa, K. (2021). Purification of modified therapeutic proteins available on the market: An analysis of chromatography-based strategies. *Frontiers in Bioengineering and Biotechnology*, 9, 717326. https://doi.org/10.3389/fbioe.2021.717326

Scaps, P. (2002). A review of the biology, ecology and potential use of the common ragworm *Hediste diversicolor* (O.F. Müller) (Annelida: Polychaeta). *Hydrobiologia* 470, 203–218. https://doi.org/10.1023/A:1015681605656

Singh, N. P., McCoy, M. T., Tice, R. R., & Schneider, E. L. (1988). A simple technique for quantitation of low levels of DNA damage in individual cells. *Experimental Cell Research*, 175(1), 184–191. https://doi.org/10.1016/0014-4827(88)90265-0

Solla, S. R. de, Gilroy, È. A. M., Klinck, J. S., King, L. E., McInnis, R., Struger, J., Backus, S. M., & Gillis, P. L. (2016). Bioaccumulation of pharmaceuticals and personal care products in the unionid

mussel *Lasmigona costata* in a river receiving wastewater effluent. *Chemosphere*, 146, 486–496. https://doi.org/10.1016/j.chemosphere.2015.12.022

Stewart, E. J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: An *in vivo* role reversal for the thioredoxins. *The EMBO Journal*, 17(19), 5543–5550. https://doi.org/10.1093/emboj/17.19.5543

Strandberg, L., & Enfors, S. O. (1991). Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. *Applied and Environmental Microbiology*, 57(6), 1669–1674. https://doi.org/10.1128/aem.57.6.1669-1674.1991

Sukhotin, A. A., Lajus, D. L., & Lesin, P. A. (2003). Influence of age and size on pumping activity and stress resistance in the marine bivalve *Mytilus edulis* L. *Journal of Experimental Marine Biology and Ecology,* 284(1), 129–144. https://doi.org/10.1016/S0022-0981(02)00497-5

Świacka, K., Maculewicz, J., Smolarz, K., Szaniawska, A., & Caban, M. (2019). Mytilidae as model organisms in the marine ecotoxicology of pharmaceuticals - A review. *Environmental Pollution* (Barking, Essex: 1987), 254, 113082. https://doi.org/10.1016/j.envpol.2019.113082

Tasiemski, A., Schikorski, D., Le Marrec-Croq, F., Pontoire-Van Camp, C., Boidin-Wichlacz, C., & Sautière, P.-E. (2007). Hedistin: A novel antimicrobial peptide containing bromotryptophan constitutively expressed in the NK cells-like of the marine annelid, *Nereis diversicolor*. *Developmental & Comparative Immunology*, 31(8), 749–762. https://doi.org/10.1016/j.dci.2006.11.003

van Loon L. C. (1985). Pathogenesis-related proteins. *Plant Molecular Biology*, 4(2-3), 111–116. https://doi.org/10.1007/BF02418757

Von Reumont, B. M., Campbell, L. I., Richter, S., Hering, L., Sykes, D., Hetmank, J., Jenner, R. A., & Bleidorn, C. (2014). A Polychaete's powerful punch: venom gland transcriptomics of *Glycera* reveals a complex cocktail of toxin homologs. *Genome Biology and Evolution*, 6(9), 2406–2423. https://doi.org/10.1093/gbe/evu190

Wang, J., Li, Z., Yin, F., Zhang, R., Zhang, Y., Wang, Z., & Sheng, X. (2022). Glioma pathogenesis-related protein 1 performs dual functions in tumor cells. *Cancer Gene Therapy,* 29(3-4), 253–263. https://doi.org/10.1038/s41417-021-00321-9

Wright, G. D. (2019). Unlocking the potential of natural products in drug discovery. *Microbial Biotechnology*, 12(1), 55–57. https://doi.org/10.1111/1751-7915.13351

Xia, Z., Chen, Y., Zhu, Y., Wang, F., Xu, X., & Zhan, J. (2006). Recombinant omega-conotoxin MVIIA possesses strong analgesic activity. *BioDrugs : Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy*, 20(5), 275–281. https://doi.org/10.2165/00063030-200620050-00003

Yamazaki, Y., Koike, H., Sugiyama, Y., Motoyoshi, K., Wada, T., Hishinuma, S., Mita, M., & Morita, T. (2002). Cloning and characterization of novel snake venom proteins that block smooth muscle contraction. *European Journal of Biochemistry*, 269(11), 2708–2715. https://doi.org/10.1046/j.1432-1033.2002.02940.x

Yu, J., Zhu, X., Yang, Y., Luo, S., & Zhangsun, D. (2018). Expression in *Escherichia coli* of fusion protein comprising α-conotoxin TxIB and preservation of selectivity to nicotinic acetylcholine receptors in the purified product. *Chemical Biology & Drug Design*, 91(2), 349–358. https://doi.org/10.1111/cbdd.13104

Zhang, L., Chou, C. P., & Moo-Young, M. (2011). Disulfide bond formation and its impact on the biological activity and stability of recombinant therapeutic proteins produced by *Escherichia coli* expression system. *Biotechnology Advances*, 29(6), 923–929. https://doi.org/10.1016/j.biotechadv.2011.07.013

Zhang, Y. F., Chen, S. Y., Qu, M. J., Adeleye, A. O., & Di, Y. N. (2017). Utilization of isolated marine mussel cells as an *in vitro* model to assess xenobiotics induced genotoxicity. *Toxicology in Vitro*, 44, 219–229. https://doi.org/10.1016/j.tiv.2017.05.018

A

APPENDIX

A.1 R script for assumption checking

```
      # First, the assumptions of the parametric tests must be tested (the variables are inde-
pendent, the normality and homogeneity of the variables will be tested):
      library(carData)
      library(car)
      #Open and organize data. This code was adapted for each variable analyzed in this thesis:
cytotoxicity from the cytotoxicity assays and % DNA in Tail and Olive Moment from the Comet
assay
      Dados<-read.csv("C:\\Users\\Utilizador\\OneDrive - FCT NOVA\\Ambiente de Trabalho\\Cata-
rinaFaustino\\Análise Comet\\DNA in Tail.txt",sep=";",header=TRUE,row.names=1)
      Dados$Protein.Concentration<-as.factor(Dados$Protein.Concentration)
      nFactors=nlevels(Dados[,1])
      Factors=levels(Dados[,1])
      nVariables=ncol(Dados)
      Variables=colnames(Dados)
      ## Check normality for each sample (concentrations vs variable), using shapiro test--> if
p>0.05 then the data follows a normal distribution:
      Dados$X..DNA.in.Tail<-as.numeric(Dados$X..DNA.in.Tail)
      for(i in 2:nVariables)
      {
        for(j in 1:nFactors)
        {
          print(paste(sep=" ","Variable:",colnames(Dados)[i],";","Factor:",Factors[j]))
          print(shapiro.test(Dados[Dados$Protein.Concentration==Factors[j],i]))
        }
      }
      ## Results: The data of variables "Cytotoxicity", "% DNA_in_Tail" and "Olive_Moment"
follow a normal distribution (p>0.05).
      ##Then the homogeneity of variances was tested: if p>0.05 then the data are homogeneous:
      for(i in 2:nVariables)
      {
        print(paste(sep=" ","Variable:",colnames(Dados)[i]))
```

```
  print(leveneTest(Dados[,i]~as.factor(Dados[,1]),center=mean))
}
# Results - The variances from the variables "Cytotoxicity", "% DNA_in_Tail" and "Olive_Moment" present homogeneity
# According to the results, the variables data "Cytotoxicity", "% DNA_in_Tail" and "Olive_Moment" fulfill the assumptions of the parametric tests.
```

A.2 R script for the Tukey's test (parametric multiple comparisons)

```
# Multiple comparisons were made by the F-test (Anova). Again, this code was adapted for each variable ("Cytotoxicity", "% DNA in Tail" and "Olive Moment") and was run after running the code in A.1

for(i in 2:nVariables)
{
  print(paste(sep=" ","Variable:",colnames(Dados)[i]))
  print(summary(aov(Dados[,i]~Dados[,1])))
  print(TukeyHSD(aov(Dados[,i]~Dados[,1])))
}
```

A.3 R script for plotting the results (Comet Assay)

```
library(gplots)
# Mean and deviation calculation (in this case the % DNA in tail results):
DNAMeans<-aggregate(Dados, X..DNA.in.Tail~Protein.Concentration, mean, na.rm=TRUE)
DNASD<-aggregate(Dados, X..DNA.in.Tail~Protein.Concentration, sd, na.rm=TRUE)
# Defining data order
DNAMeans<-DNAMeans[c(1, 1, 4, 2, 5, 3),]
DNASD<-DNASD[c(1, 1, 4, 2, 5, 3),]

# Defining errors
plotLimits<-as.data.frame(
  cbind(
    (DNAMeans[,2]+DNASD[,2]),
    (DNAMeans[,2]-DNASD[,2])
  )
)
colnames(plotLimits)<-c("upper", "lower")
#ANOVA result (multiple comparison from Tukey's test)
statsText<-c("a", "a", "a","a", "a", "a")

windowsFonts(Calibri = windowsFont("Calibri"))
windows(6, 5)
par(lwd = 2, mar = c(5, 5, 1, 1), family = "Calibri")
```

```
# Defining custom names for each pair of bars
custom_names <- c("0", "5", "50")  # One name for each pair of bars

# Specifying the positions of the custom names
midpoints <- c(1.15, 4.3, 7.4) # values adjusted as needed

DNAPlot <- barplot2(
  DNAMeans$X..DNA.in.Tail,            # y values
  plot.ci = TRUE,                     # plot error bars
  ci.u = plotLimits$upper,            # upper limit of error bars
  ci.l = plotLimits$lower,            # lower limit of error bars

  col = c("rosybrown", "steelblue", "rosybrown", "steelblue", "rosybrown", "steelblue"),

  space = c(0.1, 0.1, 1, 0.1, 1, 0.1), # Create space between bars
  xlab = expression(italic("Protein Concentration") * ' (µg/mL)'),
  ylab = expression(italic("% DNA in Tail")),
  ylim = c(0, 1.9 * round(max(plotLimits$upper, 1))), ## change the y limit
  lwd = 2,
  ci.lwd = "2",
  ci.width = 0.4,
  cex.lab = 1.2,
  cex.axis = 1,
  names.arg = rep("", length(DNAMeans$X..DNA.in.Tail))
)
# Adjust the graph
abline(h = 0, lwd = 4)
text(
  DNAPlot[, 1],
  plotLimits$upper + 0.1 * max(plotLimits$upper),
  statsText
)
# Add a legend
legend("topright", legend = c("ExtHdiv_1", "ExtP"),
       col = c("rosybrown", "steelblue"),
       bty = "n",
       pch = 20,
       pt.cex = 2.5,
       cex = 1,
       horiz = FALSE,
       inset = c(0.1, 0.1))

# Add custom names for pairs of bars using the specified midpoints
axis(1,
     at = midpoints,
     labels = custom_names,
     tick = FALSE)

# The same code was adapted for plotting the Olive Moment results
```