INVESTIGATION OF THE EFFECT OF OCEAN ACIDIFICATION ON THE HAEMOCYTE PROTEOME OF THE SOUTH AFRICAN ABALONE, *HALIOTIS MIDAE*.

By Sarah Leigh Carroll

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Molecular and Cell Biology

Faculty of Science

University of Cape Town

January 2020

DECLARATION

I hereby declare that this thesis, submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Molecular and Cell Biology at the University of Cape Town, is the result of my own investigations, apart from the referenced work of others.

Signed by candidate

Sarah Leigh Carroll

University of Cape Town

January 2020

ACKNOWLEDGEMENTS

Throughout my life I have had many role models, but none so like Carl Sagan, Steve Irwin, Richard Feynman, Stephen Hawking and Neil De Grasse Tyson, all of whom have inspired me and nurtured my curiosity in the sciences. So, this mention of their influence in my life, as well as many others', serves as an ode to the greatest and most inspiring minds of modern Science. The world would be inexplicably otherwise without them, and I certainly wouldn't be doing what I am passionate about.

I would like to acknowledge and thank the University of Cape Town, The Harry Oppenheimer Memorial Trust and the National Research Foundation for their financial assistance throughout the duration of this research project, without whom none of this would be possible.

To those in the department of Molecular and Cell Biology, particularly my lab mates in lab 201, thank you for the encouragement you gave me from the beginning of this journey. Thank you especially to Kyle, Jarid, Lee and Julia for all the excellent banter, comical memes, motivation and support over the last few years in my endeavour to complete my research. The craziness of lab 201 ironically kept me semi-sane. To my beautiful friend Jay-Dee, thank you for your love and consistency in our time in the lab together, your strength has been a pillar for me through trying times and I am so grateful for you. I so look forward to seeing you soar in your own journey through science.

To my supervisor, Associate Professor Vernon Coyne, I want to thank you from the bottom of my heart. Thank you for the in-depth discussions and for challenging me to be a better scientist and person. You have grown to be a "surrogate father" to me over the years, and I cannot express enough how grateful I am to you for the support, advice, kindness and love you have shown me throughout this journey. Only you truly know the gruelling times I went through, compounded with the feelings associated with completing a Doctoral thesis, and through it all you have been a true confidante. Never doubt your value and the impact you have in others' lives. I hope I have made and continue to make you proud in my future endeavours. I will forever cherish the memories of our Japan trip and the bond we have now, and I look forward to the scientific adventures we have ahead of us.

To my friends, there are so many of you, thank you for the memories, love and support you have given me. I have made some strong friendships in my time in Cape Town, but to my best

friend Gemma, thank you! Thank you for loving and accepting me for who I am and for being my voice of reason. You are a friend for life, and that is truly rare in this world. To Ryan, we started this journey in Cape Town together and it was a complete learning curve for the both of us. Thank you for being my friend, my love and my champion. You have believed in me when I couldn't do so myself and have loved me more so when life has thrown its lemons. I look forward to what life has in-store for the both of us. All the much!

I wish to dedicate this dissertation and all my years of study to my parents, one of whom is no longer with us today. Mom and Dad, you both have given me so much to be thankful for in life, instilling in me the true value of perseverance and hard work. You both sacrificed so much for your children, working long hours and still being there for us at any school-related event. You both encouraged our curiosity of the world, fostering in us a fervent desire to learn and to grow up to be whatever we wanted in this world. Dad, you have been my champion since I was a little girl staring up at you in my crib, always cheering me on in everything I set my mind on excelling. This PhD, this journey, was everything we spoke about while you watched me grow and take on the world around me. Losing you at the start of it was one of the hardest things I have had to muster the courage to face and work through. But I am here now, and I hope I have made you proud. I love and miss you. Mom, you have been the rock in my life in all that we have lived through. Your strength, selflessness and unconditional love are ineffable, I would not be where I am today without you by my side. You have deep-seated bravery that I can only hope to mirror in the coming chapters of my life. So, thank you for all that you are in my life, and this world, for it would be an especially forbidding place without a soul like yours. I love you more than anything in this world.

To my brother, Wesley, this world didn't deserve you. Losing you so soon after dad was difficult, more so knowing the potential you had in life and how much more of it you still needed to live. Thank you for showing me how to face your struggles head-on, and for loving those around you so deeply. Mom and I miss you.

Finally, to my past self: Despite the hardships and loss you faced, you did it! You reached the unimaginable end. Sure, there were a lot of tears, intense frustration, but nothing worth having in life comes easy. Thank you for pushing yourself, for allowing yourself to grow in every aspect of your life in Cape Town. Thank you for acknowledging your strengths and weaknesses, for endeavouring to work on the latter. You are so much more than you allow yourself to believe, never let anyone tell you otherwise.

Table of Contents

INVESTIGATION OF THE EFFECT OF OCEAN ACIDIFICATION ON THE HAEMOCYTE PROTEOME OF THE SOUTH AFRICAN ABALONE, *HALIOTIS MIDAE*.

By Sarah Leigh Carroll

January 2020

Department of Molecular and Cell Biology, University of Cape Town, Rondebosch 7700, South Africa

Abstract

Haliotis midae is an economically important marine invertebrate that is farmed in South Africa, contributing more than half of the revenue generated by the aquaculture industry. However, the future sustainability of abalone farming in South Africa is threatened by the ongoing climate crisis. The effect of climate change is unrelenting for organisms such as abalone, which rely on a succinct balance of physico-chemical environmental properties. Indeed, the ocean environment is susceptible to these imbalances and has already witnessed changes in seawater temperatures and pH. Over the last century, global ocean surface temperatures have increased by 0.74°C and seawater pH has declined by 0.1 units, while global predictions for 2100 suggest oceans will experience a decline in pH by 0.3-0.5 units. Thus, ocean acidification (OA) is a growing cause for concern since it adversely affects marine organisms such as corals and calcareous marine invertebrates.

Research focusing on the effect of climate change on marine life has grown tremendously over the last two decades, with an emphasis on molluscs such as mussels and oysters which are considered ideal proxies for measuring environmental change and the underlying molecular effects thereof. However, research on abalone in this arena has primarily focused on larvae and the physiological effects of OA on development and shell growth. The underlying molecular mechanisms involved in the *Haliotis midae* stress response to ocean acidification have largely remained unexplored. Thus, this study sought to elucidate the effect of OA on the haemocyte

proteome of *H. midae*, as well as to gain insight into the underlying molecular mechanisms that characterize the stress response of this abalone species.

This study employed a comparative shotgun proteomics approach using isobaric tagging for relative and absolute quantification (iTRAQ) coupled with LC-MS/MS to investigate the proteomic response of *H. midae* haemocytes to reduced pH conditions representative of future predictions of ocean acidification. Four independent iTRAQ experiments were conducted where haemocytes were sampled after 12, 72 and 168 hours of exposure of *H. midae* to OA conditions (pH 7.5, represents OA predicted by 2100). Following quantitative analysis, 227 non-redundant and differentially expressed proteins were detected across the four independent experiments. Proteins of statistical significance ($p \leq 0.05$) and biological relevance (fold-change ≥ 1.2) were identified.

The 227 proteins were grouped according to their expression profile using Weighted Gene Cluster Network Analysis to gain an insight into the biological processes underlying the stress response of *H. midae* to OA conditions. Sequence similarity, Gene Ontology, data mining and network modelling based on other molluscs and well-characterised genomes were utilized for assigning putative functions to the grouped proteins. This revealed a multifaceted interplay of various biological processes and signaling pathways in *H. midae*, such as the induction of anaerobic metabolism, cytoskeletal stabilization and the induction of the ERK/MAPK signaling cascade. Notably, this analysis demonstrated a possible link between the stress and immune responses which has only been observed in other molluscs. Their complex association suggests an overlap of pathways, with putative dual functionality of proteins such as MAPK, CAMK, Serpin B2 and haemocyanin, all of which have been implicated in the stress *and* innate immune responses of other organisms. The combined data from the quantitative (Chapter 2) and functional analyses (Chapter 3) resulted in the compilation of a group of 33 candidate protein biomarkers of OA stress. A holistic and complete picture of the potential regulatory mechanisms involved in the stress response of *H. midae* was generated through protein-protein interaction network modelling and data mining.

Aquarium-based experiments were conducted to validate the candidate OA biomarker proteins, as well as a set of previously identified biomarker candidates of acute temperature stress in *H. midae* (Calder and Coyne, unpublished), using label-free protein quantification (LFQ) coupled to LC-MS/MS analysis. Furthermore, the effect of a combination of reduced pH and elevated temperature on the candidate biomarker proteins was investigated. Five candidate

biomarker proteins of OA stress were validated, while 10 candidates of acute temperature stress were detected and validated. The combined stress condition identified 7 potential biomarkers. Candidate biomarkers of OA stress were predominantly associated with the innate immune system, while those responding to temperature stress were largely associated with energetics and oxidative stress. Potential biomarkers of the combined stressors were found to be associated with signal transduction and intracellular trafficking. This component of the research project not only highlighted the importance of validation in biomarker discovery, but demonstrated the usefulness of an LFQ-proteomics approach for biomarker validation.

This is the first time high-throughput shotgun proteomics has been employed to investigate the *H. midae* haemocyte stress response. This study provides a solid foundation for elucidating the putative functional stress response of a non-model organism, and highlights the complex dynamics and interplay between the stress and immune responses of *H. midae*. On-farm experimentation will be conducted to test whether the candidate biomarkers identified in this project can reliably detect abiotically stressed *H. midae* and whether they can be integrated into a "suite" of biomarkers for a health monitoring program for farmed *H. midae*. Successful implementation of such a health monitoring program will ensure the future sustainability of the South African abalone aquaculture industry despite the severity of climate change.

Chapter 1

Literature Review

Contents

1.1 *Haliotis midae*

The South African abalone, *Haliotis midae,* belongs to the phylum Mollusca, in the class Gastropoda. It is the only commercially important member of the six endemic *Haliotis* species found along the South African coastline (Franchini *et al.*, 2011). It is distinguishable by its slug-like foot and is protected by a thick, furrowed shell, providing the animal with mechanical strength and protection in its habitat, which occurs from the low-water mark up to a depth of 35 m (Barkai and Griffiths, 1987). *H. midae* is widely distributed along the west coast of the country, extending east as far as Cape Agulhas (Barkai and Griffiths, 1986). It is within this range that the commercial abalone fishery boomed.

Following the decline in abalone yields from wild fisheries, rapid development in land-based abalone cultivation in the 1990s took place (Troell *et al.*, 2006). Abalone meat is considered a delicacy, particularly in Asian countries, being packaged in various forms such as canned goods or dried abalone, for different uses (Surtida, 2000). At the turn of the century, existing abalone farms in South Africa had an estimated production of ~500-800 tons, where abalone harvesting occurs once they reach marketable size (~10 cm), which can take up to 5 years (Sales and Britz, 2001). With abalone farming becoming widespread throughout the world in the 1990s, South Africa is currently the second largest producer of abalone, while Asia remains the leading nation in abalone cultivation. South African abalone production is primarily for export to Asia, contributing greatly to the aquaculture industry. According to a 2016 report by Britz and Venter for *World Aquaculture*, abalone exports constituted roughly 76% of the overall value generated by the aquaculture industry, as well as leading in overall product value (US$ 30-50/kg), employment and annual yields. A 32% increase in production was observed in 2013, where abalone yields reached 1470 tons, contributing to nearly half of the country's total marine aquaculture production (Britz and Venter, 2016). However, with a changing climate and an increasing need to maintain or improve annual abalone yields, the effects of abiotic stress and subsequent disease outbreaks are challenges that negatively affect abalone farming. Successful mitigation of the effects of stress and disease on farmed abalone is hampered by our limited knowledge regarding the abalone immune system response to these challenges.

1.2 Abalone farming in South Africa

In its infancy, abalone farming in South Africa proved to be successful, with *H. midae* exhibiting enhanced growth rates compared to those seen in wild abalone (Troell *et al.*, 2006). Today, most farms are reliant on a pump-ashore, flow-through system, which involves local coastal water input into header tanks and subsequent distribution to raceway tanks (Morash and Alter, 2016; Troell *et al.*, 2006). Despite seasonal and diurnal changes observed in seawater input, farms employ strict monitoring regimes in terms of stocking densities, water flow rates, nutrition and aerial exposure to minimise animal stress. However, these factors do not always occur in isolation, with an increasing number of studies highlighting the adverse effects multiple stressors have on ectothermic animals, particularly as climate change conditions become more severe (B. Griffen *et al.*, 2016; Melatunan *et al.*, 2011; Morash and Alter, 2016). As farming techniques evolve to more genetic-based programmes for improving animal growth and farm yields, there is a growing need for understanding the physiological and molecular effects abiotic stressors have on economically important marine invertebrates, especially in light of climate change.

Due to the dynamic nature of coastal areas, organisms in these ecosystems experience a range of stressors over varying periods of time. Changes in ocean water temperature, pH, dissolved organic carbon and tidal fluctuations, are just some of the parameters marine animals endure. Such physico-chemical fluctuations in seawater are likely to impact onshore farms reliant on water intake, which are already challenging environments (Morash and Alter, 2016; Stone *et al.*, 2014, 2013; Troell *et al.*, 2006). The successful establishment and growth of abalone in onshore farms is largely dictated by their metabolism and energy budgets throughout their life. Thus, increasing abiotic stress leads to diversions in energy from growth to protective mechanisms, subsequently affecting abalone production. It is, therefore, crucial that we understand the underlying mechanisms that comprise the stress response of commercially valuable marine organisms such as abalone. In doing so, not only will we garner further insight into the understanding of the stress response of an enigmatic mollusc, but also gain insight that could serve as the basis upon which health monitoring programs can be developed. This will ultimately aid in mitigating yield losses on abalone farms countrywide, especially as climate change becomes a major driving force for the success of aquaculture worldwide.

1.3 Climate change

According to the United Nations Population Fund (World Population Trends, UNPFA, 2017), the global population is expected to reach over 11 billion by the end of the century, highlighting the increasing urbanisation and subsequent anthropological impact expected to affect the environment. With the increasingly intensive anthropogenic activities associated with a growing population, climate change is becoming a greater reality each year, with global temperatures having increased by 0.8°C in the last century. The effects thereof on the physical conditions of the ocean are varied worldwide; however, the last 100 years has observed an increase in global ocean surface temperatures (+0.74°C) and a decline in ocean water pH by 0.1 units (Harvey *et al.*, 2013). These climate change factors are considered to be the greatest threats to marine ecosystems and their biodiversity, affecting biological processes of individuals and subsequently altering food web interactions and structures (Doney *et al.*, 2009; Harvey *et al.*, 2013).

Extensive research has been conducted on the effects the warming ocean is having on ecosystems such as coral reefs, where coral bleaching, disease outbreaks and a decline in reef growth rates have led to changes in community structure, resulting in the disappearance of corals (Hoegh-Guldberg *et al.*, 2007; Harvey *et al.*, 2013). In a comparative transcriptomic study between invasive and native blue mussels, Lockwood and Somero (2011) sought to determine the effect of thermal stress on two species of the genus *Mytilus*. The authors wanted to elucidate possible differences between the congeners, and to determine which factors confer thermal tolerance of one species over another, thus impacting community structure and invasiveness. It was found that 96 genes were species-specific in response to heat stress, providing an avenue for predicting the invasive success of *M. galloprovincialis* (Lockwood and Somero, 2011).

Further effects of ocean warming include ocean stratification, which limits the supply of nutrients to photosynthetic organisms in surface waters (Henson *et al.*, 2017). In turn, this adversely impacts the ocean environment's food webs. In an attempt to understand the effects of elevated ocean water temperature on Giant clam (*Tridacna gigas*) larvae, researchers found that at the highest temperature (33°C), larvae exhibited proportionately more developmental abnormalities than those in the low and medium water temperature treatments (Enricuso *et al.*, 2019). As a result, survival rate and settlement of larvae was notably reduced. While the vulnerability of marine organisms and ecosystems to changes in ocean temperature has been

substantially studied, the effects of declining ocean water pH on marine organisms is becoming more of an interest to researchers, especially the additive or synergistic effects these two parameters have on ocean-dwelling organisms (Wernberg *et al.*, 2011; Harvey *et al.*, 2013; Li *et al.*, 2016). The increase in global temperatures may be attributed to the increase in atmospheric CO_2 which ultimately leads to the "greenhouse effect".

As of 2018, the NOAA reported an atmospheric CO_2 level of 410 ppm. While a significant proportion of the atmospheric CO_2 contributes to the retention of solar heat, over more than half of it is known to be taken up by our ocean, which has resulted in a decline in ocean water pH. Ocean water pH is predicted to decline by 0.4 units by the turn of the century, while the prevalence of the effects of ocean acidification is becoming a worldwide cause for concern (Orr *et al.*, 2005). Ocean acidification (OA) is the dissolution of CO_2 into the ocean, resulting in an increase in hydrogen ion and carbonic acid concentrations, subsequently altering the ocean water pH (Bibby *et al.*, 2008; Stumpp *et al.*, 2011). This shift in ocean water chemistry and carbonate saturation is likely to not only impact those marine animals that rely on a succinct chemical balance for shell and skeleton formation but may also have the potential to affect animals on a physiological and molecular level.

Research in this field has focused on the effects a change in ocean water chemistry will have on calcification rates in coral, as well as the survival and growth rates of crustaceans and shell development in some marine invertebrates (Cooper *et al.*, 2008; Gomiero *et al.*, 2018). In a study looking at the effects of ocean acidification on the physiological response of the blood clam, *Tegillarca granosa*, the authors exposed the clams to three reduced seawater pH levels for a total of 40 days (Zhao *et al.*, 2017). Zhao *et al.* (2017) found that OA resulted in extracellular acidosis and decreased Ca^{2+} concentrations in the haemolymph, as well as adversely affecting the feeding activity and aerobic metabolism of the clams. In the Pacific oyster, *C. gigas*, an increase in apoptosis and reactive oxygen species (ROS) production was observed in response to elevated pCO_2 after 28 days (Wang *et al.*, 2016). Roggatz *et al.* (2016) investigated the effects of predicted ocean pH conditions on the behaviour of *Carcinus maenas* by studying the molecular impact on three signalling peptides, often involved in predator detection, settlement and reproduction. The authors found that a decline in pH is likely to increase the susceptibility of these signalling peptides to protonation, altering their overall charge and, subsequently, adversely affecting their functionality (Roggatz *et al.*, 2016). Ultimately, this observed change in signalling peptide properties, as a result of reduced pH, could lead to altered behaviour in the shore crab under future oceanic conditions (Roggatz *et*

al., 2016). Forsgren *et al.* (2013) observed the effects of ocean acidification on the embryonic development and larval phototaxis of *Gobiusculus flavescens*, a temperate marine fish. Embryonic abnormalities were pronounced and egg loss increased under experimental conditions, while the phototactic response of larvae was significantly affected (Forsgren *et al.*, 2013). The implications thereof include potentially impaired embryonic development and sensory responses, affecting fish population dynamics (Forsgren *et al.*, 2013).

These studies highlight how one aspect of climate change can have far-reaching consequences, affecting not only the physiology of marine organisms, but also their community structure, growth and development, as well as their interactions with their aquatic environment. However, environmental change is diverse, with changes in abiotic conditions rarely occurring in isolation (Gunderson *et al.*, 2016). Understanding the physiological and behavioural responses of marine organisms to ocean acidification and ocean warming is crucial for determining their level of adaptability in the future. As such, it underpins our role in finding ways to help mitigate losses and ensure extensive conservation efforts. In a study looking at the combined effects of ocean acidification and elevated seawater temperature on gilthead sea bream (a commercial fish), the authors found that mortality rates reached 50% compared to the effect of the stressors in isolation (Araújo *et al.*, 2018). The authors also investigated the muscle proteome response to ocean acidification effects in a hot ocean, as well as warming effects in an acidified ocean. They found that there was an upregulation in anaerobic metabolic processes, as well as an inhibition of proteasomal degradation, suggesting a possible threat to the survivability of sea bream (Araújo *et al.*, 2018). Macrofauna are not solely affected by changing climatic conditions. Gomiero *et al.* (2018) observed the synergistic effects of elevated water temperature and reduced seawater pH on two primary producers: a marine ciliate and a green alga. Under sub-lethal stress conditions, the authors found that there was an activation of oxidative stress and significant impairment of the lysosomal compartment, followed by major energy costs (Gomiero *et al.*, 2018). Furthermore, photosynthetic efficiency was notably impaired in the green alga.

Molluscs are the second largest phylum, inhabiting a variety of environments. Molluscs are considered suitable indicators of environmental change as a result of their diversity, abundance, extensive representation in the fossil record and their geographical distribution (Fortunato, 2016). They provide environmental information that is not directly measurable, as well as a foundation upon which inferences of changing environmental conditions can be based, in addition to identifying potential faunal adaptations. For example, a marine pteropod was

utilised as a bioindicator for ocean acidification in a study that sought to describe and quantify shell dissolution (Bednaršek *et al.*, 2012). The authors were able to classify dissolution according to three main types following incubation of animals in CO_2-rich seawater for up to 14 days. They found that the degree of shell dissolution in the pteropod specimens correlated with incubation time, concluding that the early response observed by the pteropods showed promise as proxies for future ocean acidification conditions (Bednaršek *et al.*, 2012). In essence, the extensiveness of molluscs in the fossil record serves as a record of past climatic events, which enhances our understanding of future potential climatic changes and possible effects on marine molluscs (Fortunato, 2016). For these reasons, it is unsurprising that innumerable molluscan studies have been conducted to elucidate the effect a changing climate (particularly elevated temperature) has on the biomineralization, development, physiology and immune response of several key species (Dineshram *et al.*, 2016; Gazeau *et al.*, 2013; Matozzo and Mari, 2011; Melatunan *et al.*, 2011; Navarro *et al.*, 2016; Zhao *et al.*, 2017, 2012).

There is a growing body of research investigating the synergistic effects of various climate change factors on marine invertebrates, particularly molluscs. For example, in a study investigating the effect of future climate scenarios on the pearl oyster, the overall conclusion was that reduced pH and reduced pH coupled to elevated temperature induced downregulation of genes involved in biomineralization (Liu *et al.*, 2012). Additionally, stress proteins such as HSP70 were induced in *Pinctada fucata* under reduced pH and elevated temperature conditions, cumulatively suggesting that not only will the physiology of *P. fucata* be adversely affected by predicted future climate conditions, but its adaptability to a changing ocean could be hampered (Liu *et al.*, 2012). Another study conducted on *C. gigas* focused on the effects of ocean acidification and elevated temperatures on energetics, finding that reduced pH conditions are likely to adversely affect the energy metabolism of oysters (Lannig *et al.*, 2010). This suggests vulnerability to predicted ocean conditions under a changing climate. The combined effects of reduced pH and temperature on *Argopecten purpuratus* in an upwelling-affected area were examined: low pH (7.7) and ambient temperatures (14°C) caused shell dissolution and a decline in scallop growth rates, while elevated temperature (18°C) limited the effects of reduced pH by maintaining growth rates similar to those under control conditions (Lagos *et al.*, 2016). This has implications for how scallops are farmed in the area, since upwelling brings colder waters, thus impacting the growth rates of farmed scallops. Conversely, warmer temperatures ameliorate low pH effects, but could have detrimental effects such as increased susceptibility to disease outbreaks. Since *H. midae* aquaculture in South Africa has grown

extensively, alongside continued efforts to improve health monitoring, there is a strong need for understanding the immune system and stress response of an understudied commercial asset (Kemp and Coyne, 2011). This is of particular interest not only for improving the farming of *H. midae*, but for monitoring the viability of abalone aquaculture under future climatic conditions.

1.4 Studying immunity in non-model organisms

Being invertebrates, molluscs lack an adaptive immune system, but have evolved other astute mechanisms as host defence, namely innate immunity. Should the external barriers of molluscs (i.e. calcareous shells, mucus and epithelia) fail in the first line of defence, this complex system (comprised of cellular and humoral components) is capable of recognizing and responding to cell surface proteins of potential pathogens (Iwanaga and Lee, 2005; Sokolova, 2009). Haemocytes, the main cellular component of haemolymph, are unequivocally central to the innate immune system of molluscs. They are responsible for the phagocytosis of pathogens and foreign particles, which is a complex immune response that culminates in the elimination of engulfed material (Sokolova, 2009). Furthermore, haemocytes are directly involved in the destruction of pathogens through the oxidative burst reaction, which results in reactive oxygen species (ROS). Myeloperoxidase in haemocytes converts ROS to hypochloric acid, which has strong microbiocidal properties (Sokolova, 2009). Haemocytes are not only significant components of the immune system of molluscs, but they are also involved in regulating processes such as wound and nerve repair, and biomineralization (Pila *et al.*, 2016).

Haemocytes carry out innumerable functions (such as mediating senescence or regulating immune responses), which oftentimes reduces the number of circulating haemocytes, making haematopoiesis crucial for replacing lost cells (Pila *et al.*, 2016). Although haematopoiesis is not well understood in molluscs, it is important not only in the immune system but for whole organism survival. This all highlights the central role haemocytes play in molluscan survival. Indeed, the stress and immune response of molluscs are closely linked, particularly the efficiency of the immune system under stressful conditions. For example, stressful conditions, such as elevated temperature, can elicit disease outbreaks in bivalve populations as a result of the animals being immunocompromised (Malagoli *et al.*, 2007). Abiotic stress occurs when environmental conditions deviate from species-specific optima, resulting in a decline in fitness, which can lead to reduced growth rates and increased susceptibility to infection (Sokolova,

2013). And since there are reported species-specific responses to environmental conditions (Lannig *et al.*, 2010), mussels and oysters have been the primary focus of studies of the effects of changing environmental conditions on stress and immune responses. Few studies have focused on the immune system of the genus *Haliotis*, and even fewer have investigated the impacts a changing climate has on their stress and immune responses. As such, abalone genetic data in public databases is limited. Organisms, like abalone, that have not had their genomes sequenced are considered non-model (Calvete, 2014). Conducting research on non-model organisms poses several challenges. For example, with ill-defined genomes or limited genetic data, the conundrum arises when analysing data where protein functionality is determined through inferences from evolutionarily conserved genetic data from other species (Calvete, 2014). Conversely, the short generation times, small size and the reduced-complexity of some model organisms, such as *Mus musculus* and *Drosophila* that serve as ideal representatives, are what has made working on model organisms desirable (Armengaud *et al.*, 2014). However, there is a growing interest in non-model organisms as study subjects, particularly since they fundamentally underpin comparative biology, which entails investigating and understanding the mechanisms that confer biological diversity. Furthermore, innumerable non-model organisms play crucial roles in food webs and community structures within ecosystems, highlighting the advantages of studying non-model organisms, especially in light of climate change. In doing so, changes in ecosystem dynamics could potentially be predicted, probing the knock-on effects climate change could have.

Although non-model organisms pose several challenges, continuous advances in proteomics and related tools are providing suitable avenues for pursuing such research. For example, Wang *et al.* (2010) sought to elucidate the molecular mechanism of larval settlement and development of *Bugula neritina*, a non-model marine invertebrate, that ultimately results in biofouling of ship bows and other man-made sub-surface substrata. To achieve this with limited genetic information, the authors utilised a tandem transcriptome and proteome profiling approach: a 454 pyrosequencing platform, which was able to generate 131 450 high-quality reads, ultimately resulting in 882 identified and quantified proteins (Wang *et al.*, 2010). The novel approach of coupling two high-throughput technologies proved powerful in enhancing the study of a non-model organism. Although coupling tandem mass spectrometry of peptides with database searching is the most common route for characterising molecular phenotypes, research involving non-model organisms requires a modified approach to database searching which incorporates both identity and homology in order to improve the robustness of the

analysis (Russeth *et al.*, 2006). Russeth *et al.* (2006) used four software packages for identifying proteins from mass spectrometry data generated from a non-model organism. In this way, the authors increased the probability of correct protein identification since each package utilised its own algorithm for database searching (Russeth *et al.*, 2006).

1.5 Research objectives of this project

The South African abalone industry is considered the "powerhouse" of the aquaculture sector, with cultured *H. midae* products yielding a 76% share of the overall value of aquaculture exports, while also leading in employment numbers, production tonnage and product value (Britz and Venter, 2016). Farming techniques are continuously evolving to improve farm yields, however with climate change becoming a growing concern, the sustainability of abalone aquaculture is becoming unpredictable. This is due in-part to the pump-ashore systems abalone farms rely on for supplying header tanks, as well as to the compounding effects changing oceanic conditions will have on these onshore farm set-ups (Morash and Alter, 2016; Stone *et al.*, 2014, 2013; Troell *et al.*, 2006). It is possible that undetected farm stress in abalone under future climate conditions will be exacerbated, increasing the probability of disease outbreaks as a result of increased susceptibility to infection. Early detection of an infectious disease is challenging since there are often no clear phenotypic indications of abiotic stress in abalone. As such, investigation of the underlying mechanisms involved in the *H. midae* stress response is essential, especially with climate change being a driving force of the changing physico-chemical conditions of the ocean. In doing so, not only will the stress response of an economically valuable marine invertebrate be elucidated, but the information garnered could form the foundation for developing a health monitoring scheme for farmed abalone. Significant strides have been made in the -omics and MS-based fields, allowing for large-scale investigations at the systems level.

Thus, this research focuses on elucidating the proteomic stress response of *H. midae* haemocytes following exposure to reduced pH conditions, by utilising a high-throughput shotgun proteomics approach coupled to LC-MS/MS. With the aid of existing marine invertebrate genomic data and *de novo* sequencing, global haemocyte protein expression will be elucidated, while homology-based functional bioinformatic analyses will characterise the identified proteins. Through the identification of groups of functionally related proteins, key pathways and proteins/effectors of ocean acidification stress can be identified. This will not

only provide valuable information on the stress response of an enigmatic marine invertebrate, but it could serve as a foundation for biomarker discovery of ocean acidification (OA) stress in *H. midae*, which could ultimately be incorporated in a health monitoring scheme for farmed abalone. As such, the main aims of this project were to:

i. Quantitatively characterise the haemocyte proteome of *H. midae* exposed to reduced pH conditions representative of future oceanic pCO_2 conditions, using isobaric tagging for relative and absolute quantification (iTRAQ) and LC-MS/MS.
ii. Functionally characterise and elucidate the underlying mechanisms and associated proteins of the stress response of *H. midae*, utilising related molluscan genomic data and various bioinformatics analyses.
iii. Validate proteins that could serve as biomarkers of OA stress in *H. midae*, through the use of label-free quantification (LFQ) and LC-MS/MS. Additionally, haemocyte proteins found to be upregulated in response to acute temperature stress in a study conducted by Calder and Coyne (unpublished) will be assessed against the LFQ dataset to identify biomarkers that respond to pH and temperature stress either individually or in combination.

Chapter 2

Profiling the *Haliotis midae* haemocyte proteome in response to decreased seawater pH.

Contents

2.1 Introduction

With the advent of the "omics" revolution, the marine sciences are delving further into determining the implications a changing global climate will have on marine organisms, specifically at the systems level. Changes occurring at the gene and protein level are generally sensitive and can be suitable gauges of biological processes, though proteins have been specifically found to play a direct role in the regulation of biological pathways and physiological responses in organisms (Dowd, 2012; Timmins-Schiffman *et al.*, 2013). While online databases house a plethora of taxonomic DNA sequences, complete genome sequences are insufficient for determining biological functions, as they are not directly correlated to protein function (Boehm *et al.*, 2007; Lamond *et al.*, 2012). Proteomics has emerged as one of the most important and reliable "omics" technologies. This is due to the integral role the proteome of an animal plays in responding to environmental changes (Slattery *et al.*, 2012). In essence, the proteome is the protein complement of the genome. It is dynamic and is comprised of complex networks that respond rapidly to environmental change, providing "snap-shot" information regarding the responses of animals to their environment at any given point in time. The proteome, in many cases, works independently of the genome. This is achieved through the multi-functionality proteins exhibit as a consequence of post-translational modifications and protein-protein interactions (Tomanek, 2014). Therefore, proteins are key players in biochemical pathways that determine the phenotypic response of an organism to its environment. By studying the role proteins play, as well as their properties and interactions with members of other pathways, greater insight can be gained regarding the phenotypic response exhibited by an organism to environmental change (Tomanek, 2014).

Making its debut in the 1970s, proteomics involved the cataloguing of expressed proteins using 2-D gel electrophoresis, which has been the standard for separating and viewing protein samples. However, determining the identity and concentration of these proteins proved difficult due to the lack of sensitive and efficient analytical methods (Pandey and Mann, 2000; Westermeier and Naven, 2002). Today, proteomics research is largely aided by mass spectrometry (MS). Mainstream proteomics is dominated largely by bottom-up proteomics approaches, which involves the digestion of protein samples to their peptide constituents for protein identification and post-translational modification (PTM) determination through the use of mass spectrometry (Mayne *et al.*, 2015). The popularity of bottom-up proteomics is due to the ease and efficiency of peptide detection compared to top-down proteomics involving intact proteins (Mayne *et al.*, 2015).

Despite the growing popularity of proteomics-based research, some challenges remain such as ensuring sensitivity to compensate for the inability to amplify proteins (Cox and Mann, 2007). Furthermore, sample preparation and post-acquisition data processing, as well as the interpretation of analyses, pose some serious difficulties for researchers (Tomanek, 2014). These challenges are largely overcome through the adaptation of sample preparation protocols, such as the filter-aided sample preparation (FASP) protocol as described by Wiśniewski *et al.* (2009), as well as through comparative bioinformatics analyses that include closely related organisms (Tomanek, 2014). These homology-related searches are crucial for under-represented organisms across databases. Mass spectrometry relies on the ionization of chemical compounds to produce charged gaseous molecules under a vacuum, where the molecules' mass are determined according to their mass-to-charge (*m/z*) ratio (Piñeiro *et al.*, 2010). Mass spectrometers consist of three fundamental parts: an ion source that converts target molecules into gas-phase ions, a mass analyser that uses the *m/*z ratio of ionized molecules for separation, and a detector that observes the number of ions detected at each *m/*z value (Han *et al.*, 2008). Until recently, quantifying proteins in samples also presented a significant problem. Two main types of strategies can be employed for quantification: label-based and label-free MS-based quantification.

The label-based method relies on the introduction of stable isotopes through various means such as chemical, metabolic and enzymatic labelling (Latosinska *et al.*, 2015). Quantification in label-based methods occurs at the MS/MS level, where the number of spectra from fragmentation are associated with peptides of a particular protein. Additionally, label-based methods allow for multiplexing, thus reducing analytical variability (Latosinska *et al.*, 2015). Contrastingly, label-free methods employ spectral counting and intensity measurements for obtaining quantitative data. Intensity-based quantification involves estimating the area of the peak from the ion chromatogram and occurs at the MS1 level (Latosinska *et al.*, 2015). Both types of MS-based quantification are still widely employed, but for the purposes of this investigation, we utilised label-based quantification, employing a shotgun-proteomics approach coupled to liquid chromatography and tandem mass spectrometry (LC-MS/MS).

Shotgun proteomics is a rapidly evolving and high-throughput technology, which involves the simultaneous identification and quantification of protein members within a given system, using MS (Fuller and Morris, 2012; Martyniuk *et al.*, 2012). It allows for the evaluation and monitoring of the effects of changing environmental conditions on a biological system. Furthermore, shotgun proteomics contributes greatly to drug development as most drugs are

directed against proteins (Pandey and Mann, 2000; Westermeier and Naven, 2002). Shotgun proteomics achieves its dual functionality of identification and quantification through the use of isobaric tagging of peptides, allowing for the analysis of multiple samples (Aggarwal *et al.*, 2006). Coupling shotgun proteomics with LC-MS/MS allows for the identification of proteins, by matching the fragment spectra of peptides in a sample to theoretical spectra determined from known DNA or protein sequences deposited in databases (Boehm *et al.*, 2007; Pandey and Mann, 2000).

Isobaric tagging for relative and absolute quantification (iTRAQ) is a popular peptide-labelling tool for characterising proteins in complex samples. iTRAQ is attractive over other methods for a number of reasons; most notably, it allows for the concurrent identification and relative quantification of proteins in up to eight samples, allowing for biological replicates, time-course experiments, and dose responses (Martyniuk *et al.*, 2012). iTRAQ involves the use of isotope-encoded mass tags to label digested peptides, where the iTRAQ reagents themselves are comprised of a reactive group (mass balance group) and a reporter group (based on N-methylpoperazine). The primary amines of proteolytic peptides are linked to these groups through the formation of an amide bond (Wiese *et al.*, 2007; Ernoult *et al.*, 2008). Measurements are obtained at the peptide level following enzymatic digestion of protein samples, and then combined to generate a summarized value for the protein they comprise (Karp *et al.*, 2010; DeSouza & Siu, 2013). The enhanced labelling redundancy of these peptides strengthens the confidence in both the identification and quantification aspects of proteomics (Karp *et al.*, 2010). Additionally, due to the tags ability to label different peptides in a sample, iTRAQ is ideal for quantification of posttranslational modifications and may also be used to label whole proteins prior to protein fractionation (Unwin, 2010). However, the downside to combining labelled peptide samples is that this might introduce errors as a consequence of incongruities in individual samples processed separately prior to mixing (DeSouza and Siu, 2013). The combined differentially-labelled peptides are then separated using liquid chromatography-tandem mass spectrometry (LC-MS/MS). Due to the isobaric nature of the iTRAQ reagents, sample complexity is reduced as the same peptide from each sample appears as a single peak on the mass spectrometry chromatogram (Aggarwal *et al.*, 2006; Ernoult *et al.*, 2008). Collision-induced dissociation allows the labelled peptides to release the balance group as a neutral fragment, thus freeing tag- and sample-specific reporter ions (at 113, 114, 115, 116, 117, 118, 119, 121 *m/z*), where their peak areas are used to determine the relative

abundance of peptides, and parent proteins may be identified by searching peptide sequences against protein databases (Aggarwal *et al.*, 2006; Wiese *et al.*, 2007; Ernoult *et al.*, 2008).

To reduce sample complexity prior to MS analysis, off-gel fractionation is carried out. Here, iTRAQ-labelled peptide samples are separated in-solution through isoelectric focussing (IEF) (Ernoult and Guette, 2011). Through its high resolution, IEF is capable of separating and concentrating biomolecules according to their isoelectric points (p*I*) along a pH gradient with an applied electric force (Hörth *et al.*, 2006; Ernoult & Guette, 2011). The samples are then recovered in the liquid phase for further analysis. Not only does this technique offer efficient separation of peptides, but it also provides information regarding the isoelectric points of peptides, which serves as an independent tool for validation and eliminating false positives (Abdallah *et al.*, 2012). Additionally, this method allows for the direct introduction of fractionated peptides to LC-MS/MS analyses. Chenau *et al.* (2008) evaluated off-gel fractionation's suitability as a pre-analytical step, as well as its compatibility with iTRAQ labelling reagents. The data showed that the off-gel system is a valuable tool for fractionating complex peptide mixtures, as large amounts of sample can be loaded and separated efficiently without affecting iTRAQ-labelled peptides (Chenau *et al.*, 2008). IEF, in combination with immobilized pH gradient (IPG) gel strips, offers high reproducibility and easy recovery of peptide fractions following fractionation (Eravci *et al.*, 2014). Furthermore, the development of these gel strips has made this set-up a lot simpler to use compared to the carrier-ampholyte-driven pH gradients (Zhang *et al.*, 2013). The coupling of iTRAQ labelling and IEF in a proteomics workflow allows for efficient sample preparation prior to MS analysis, making it desirable in multi-treatment studies.

In an iTRAQ study investigating the effects of pathogenic challenge on the coelomocytes of the commercial sea cucumber *Apostichopus japonicus*, the researchers identified 288 differentially expressed proteins (Zhang *et al.*, 2014). Following functional analyses, the authors were able to identify 8 proteins involved in pathological responses in high order animals. The authors noted that iTRAQ provided them with an easy and efficient method for identifying candidate protein targets for disease screening at the micro RNA level (Zhang *et al.*, 2014). Indeed, iTRAQ has become a common proteomics tool for assessing the response of marine organisms to the changing physicochemical parameters of the ocean, especially in light of climate change. A study utilising iTRAQ identified 2379 quantifiable, differentially expressed proteins in the gills of *Crassostrea gigas*, following exposure to three different stressors (elevated temperature, low salinity and aerial exposure) (Zhang *et al.*, 2015). The

authors found that heat shock resulted in an upregulation of molecular chaperones, suggesting the increased need for protection against protein denaturation. Desiccation-responsive proteins were found to be downregulated in response to aerial exposure, suggesting possible energy conservation to manage the stress (Zhang *et al.*, 2015). The authors concluded that the validation of their proteomics data supported the suitability and reproducibility of iTRAQ and LC-MS/MS for investigating the molecular response of marine invertebrates to abiotic stressors. A study looking at the combined and individual effects of pathogenic challenge and ocean acidification on the haemocytes of *C. gigas*, utilised iTRAQ to quantitatively and functionally analyse the oyster's immune response (Cao *et al.*, 2018). Their findings highlighted that exposure to ocean acidification notably increased the oyster's risk of pathogenic infection, which would have implications for disease outbreaks in the wild, as well as in farmed populations. This would ultimately lead to adverse economic effects.

It is evident that proteomics, specifically iTRAQ and LC-MS/MS, serves as an ideal approach for assessing the genetic and phenotypic response of marine organisms to changing environmental conditions. Through elucidating the stress response of commercially important marine organisms and the implications thereof, health and stress monitoring schemes can be developed to help mitigate losses of farmed marine organisms as a consequence of climate change.

2.1.1 Aims of this chapter

Climate change is becoming an increasing cause for concern on a global scale. Ocean acidification is predicted to decline by 0.3-0.5 pH units by the turn of the century. Understanding how different organisms respond to changing environmental conditions, such as ocean acidification, provides insight into the adaptability of different species, as well as information that could be used to help mitigate losses through specific, timeous interventions. Abalone farms naturally experience fluctuations in several abiotic factors, but with climate change predictions becoming increasingly worrisome, it is not well understood to what degree changing conditions will exacerbate the response of abalone. Shotgun proteomics seeks to identify and quantify protein populations in treated samples, using a high-throughput systems approach, allowing for the evaluation of the effects a changing environment might have on an organism.

The aim of this chapter is to quantitatively analyse the *H. midae* haemocyte proteome in response to ocean acidification (reduced pH conditions), using an aquarium-based experimental set-up. In so doing, the overall aim of the study will be to identify proteins of biological importance that could be further investigated as potential biomarkers of pH stress in farmed *H. midae*. To determine whether the applied stressor affects abalone on a physiological and immunological level, phagocytosis assays were conducted to evaluate the immune response to heat-killed *Vibrio anguillarum*, a common marine bacterial pathogen. This will shed light on whether OA conditions are likely to stimulate phagocytosis or cause immune-suppression. To evaluate the effects of reduced pH conditions on the haemocyte proteome of abalone, iTRAQ labelling of reduced peptide samples was employed, followed by LC-MS/MS analysis of fractionated peptide samples. Subsequently, MS data was processed, filtered for redundancies and statistically analysed. Functional/qualitative analysis of the identified proteins is presented in Chapter 3. To facilitate the analysis of the protein expression data obtained from iTRAQ LC-MS/MS, the following questions were posed:

i. *What is the overall proteomic response of abalone haemocytes to reduced pH conditions?*

   This will provide an overview of the general patterns observed in haemocyte protein expression, allowing for a snapshot of their response to the stressor. Utilising filtering criteria and heatmap generation, this will aid in visualising the proteomic response of *H. midae* haemocytes to reduced pH conditions, which in turn, will provide the foundation on which further analyses will be based.

ii. *Are there differentially expressed proteins that could potentially function as biomarkers of ocean acidification stress in farmed H. midae?*

   This will be determined by conducting statistical analyses and employing fold change cut-off criteria on the quantitative data. As such, this will allow us to refine the "suite" of candidate protein biomarkers of ocean acidification stress for subsequent validation using a label-free quantification proteomics approach.

2.2 Methods and materials

2.2.1 Animal maintenance

Haliotis midae (shell length ~ 7 cm), obtained from a commercial abalone farm and housed at the Department of Agriculture, Forestry and Fisheries (DAFF) Research Aquarium in Sea Point, Cape Town, South Africa, were transported to the aquarium in the Molecular and Cell Biology Department at the University of Cape Town. The animals were acclimated for a week at 14°C in circulated filtered (DoPhin C500 Canister Filter) sea water obtained from the DAFF Research Aquarium, at an ambient pH of 7.9, in 50 L glass tanks containing 24 abalone each. The tanks were aerated with ambient air using the Electro-magnetic air compressor ACQ-008A (Style King). Water was changed twice a week while animals acclimated. After the acclimation period, the abalone were divided into three groups: the control tank (pH 7.9, the prevailing pH of the holding tanks at the DAFF research aquarium) and two experimental tanks (pH 7.5, representing predicted ocean pH by 2100) maintained by bubbling CO_2 under the control of a pH probe and regulator. The CO_2 gas cylinder was set up to maintain the experimental pH at the desired level using a pH computer (Aqua Medic). The abalone were fed with fresh kelp two days prior to commencing the study and the tanks were cleaned with an aquarium siphon and topped up with ambient seawater.

2.2.2 Exploratory & final iTRAQ experiments

To identify suitable exposure times for the final iTRAQ experiments, an exploratory iTRAQ was conducted. To achieve this, we sought to investigate the early vs. late response of abalone to the stressor by utilising a sampling time course of 12 hours, 24 hours, 72 hours, 1 week (later referred to as 168 hours), 2 weeks, 3 weeks and 4 weeks. Since kelp has been found to adjust seawater pH if left for too long in the tanks (observed in previous experiments), abalone exposed for the extended times were fed fresh kelp 2 days before sampling and allowed to feed for a day. Abalone waste was removed with an aquarium siphon. Following analysis of the exploratory iTRAQ data, four time points were selected for characterising the proteomic response of abalone haemocytes to reduced pH conditions.

2.2.3 Experimental design & haemolymph sampling

Following acclimation, abalone were transferred from the holding tank to the experimental tanks that had been adjusted to pH 7.5. Abalone were exposed to reduced pH conditions for 12, 72 and 168 hours (section 2.2.2), with four abalone sampled at each time point to ensure sufficient protein could be isolated for iTRAQ labelling.

Haemolymph was sampled from the pedal sinus of the abalone using a $25G \times \frac{5}{8}$" hypodermic needle (Avacare™, Lasec) with a sterile 2.5 ml syringe. The haemolymph from abalone at the respective time points was pooled and total haemocyte counts were conducted using a haemocytometer. Haemocytes were collected through centrifugation at 8000 x *g* for 10 minutes at 4°C. The supernatant was discarded, and the pelleted haemocytes were snap frozen at -80°C.

2.2.4 Preparation of labelled bacteria for phagocytosis assay

Mortensen and Glette's (1996) method for fluorescently labelling *Vibrio* spp. was followed with some modifications. *V. anguillarum* 1989 was plated on Tryptic Soy Agar (TSA) supplemented with 2.5% (*w/v*) NaCl and incubated at room temperature for 16 hours. Single colonies were inoculated into 5 ml Tryptic Soy Broth (TSB), 2.5% (*w/v*) NaCl and incubated for 16 hours with shaking at room temperature. The 5 ml pre-culture was used to inoculate 100 ml TSB, 2.5% (*w/v*) NaCl, which was incubated for 16 hours with shaking at room temperature.

The cells were collected via centrifugation at 13 000 x *g* for 10 minutes at 4°C, after the culture had been heat inactivated for 30 minutes at 60°C. The cells were then washed twice with 10 ml 1x PBS and collected via centrifugation. The cells were resuspended in 10 ml 0.1M $NaHCO_3$ pH 9, with 0.1 mg/ml fluorescin 5-isothiocyanate (FITC) Isomer 1 (Sigma) in 10 ml 1x PBS.

The cells were then incubated with shaking overnight (~16 hours), in the dark at room temperature. Thereafter, the cells were collected by centrifugation at 13 000 x *g* for 1 minute at 4°C, and subsequently resuspended in 10 ml 1x PBS. The concentration of bacteria was determined using a 4-block cell count on a haemocytometer and calculated as follows:

Cells/ml = [(no. of cells/0.1) / 0.0025] x 1000 x dilution factor

2.2.5 Phagocytosis assay

Following total haemocyte counts at each time point, haemocytes at a concentration of 1×10^6 cells/ml were applied to acid-washed glass coverslips within a ring of silicon and allowed to adhere for 30 minutes in a dark, moist chamber. Adhered haemocytes were washed with ice cold PBS and subsequently incubated with 100 µl of FITC-labelled *V. anguillarum* 1989 (section 2.2.4), at a concentration of 1×10^8 cells/ml, for 20 minutes. The haemocytes were washed once with PBS and fixed for 10 minutes with the addition of 150 µl methanol. Methanol was removed by washing with PBS, followed by the addition of 150 µl Ethidium bromide (100 µg/ml) for 1 minute, after which the haemocytes were washed with ice cold PBS. New coverslips were placed over the silicon rings to create a chamber and the number of phagocytosing haemocytes was determined using an inverse fluorescent microscope (Nikon Inverted Microscope DIAPHOT-TMD containing a Nikon EDI-Fluorescent attachment TMD-EF) at 400x magnification using a 510 nm excitation filter. Haemocytes that contained two or more FITC-labelled *V. anguillarum* 1989 were considered as phagocytic cells. The percentage of phagocytosing haemocytes was determined from a total of 200 haemocytes:

$$\% \text{ phagocytosing} = \frac{no.\ phagocytosing\ haemocytes}{total\ haemocytes\ counted} \times 100$$

Each sample was assayed in duplicate and each duplicate was counted twice.

2.2.6 Haemocyte protein isolation & quantification

Haemocyte samples (Section 2.2.3) were thawed in lysis buffer [0.1 % molecular grade Sodium Dodecyl Sulphate (SDS) in 0.5 M Tetraethylammonium bromide (TEAB) (Sigma-Aldrich)] and subsequently lysed by vortexing (Vortex Gene 2) for 25-30 minutes at 4°C. The lysate was centrifuged (Eppendorf Centrifuge 5417R) at 8500 rpm for 30 minutes at 4°C. For protein quantitation, the bicinchoninic acid (BCA) assay was conducted on the supernatant, where a 1:10 dilution of the samples was made (10 µl sample + 90 µl MilliQ Water) to reduce the TEAB to a working concentration for quantitation. The Pierce BCA Protein Assay (Thermo Scientific) was conducted according to the manufacturer's instructions to quantify protein concentrations (Figure 1). SDS-PAGE (Bio-Rad) was conducted to check protein integrity before commencing sample preparation prior to iTRAQ labelling.

2.2.7 Filter Aided Sample Preparation for in-solution digestion (FASP) as described by Wiśniewski *et al.* (2009)

2.2.7.1 Protein Reduction

Aliquots of protein corresponding to 300 µg were placed in lo-bind Eppendorf tubes (Lasec) with 0.1 volume of 5 mM [Tris(2-chloroethyl) phosphate] TCEP to reduce cysteine disulphide bonds. The samples were vortexed, pulse spun and incubated at 60°C in a water bath for 1 hour.

The reduced protein was transferred to a 30 kDA MWCO Amicon Ultra centrifugal filter (Sigma-Aldrich) and concentrated by centrifugation at 10 000 x *g* for 20 minutes at 20°C to a final volume of 30 µl.

2.2.7.2 Alkylation

The retentates were alkylated with 100 µl of 8 M molecular grade Urea (Sigma-Aldrich) in 0.5 M TEAB (UT buffer) containing 15 mM s-methyl methanethiosulphonate (MMTS) (Sigma-Aldrich). The samples were vortexed and incubated at room temperature for 15 minutes.

UT buffer (300 µl) was added to the retentates and centrifuged at 10 000 x *g* for 30 minutes at 20°C to remove the SDS. This was carried out a total of 3 times. 60 µl of 0.5 M TEAB was added to the filter column to reduce the urea concentration. The column was centrifuged at 10 000 x *g* for 30 minutes at 20°C. This was done a total of two times, leaving a final volume of 30 µl.

2.2.7.3 Protein digestion

The protein retentate was collected by centrifugation by inverting the Amicon filter and centrifuging at 1000 x *g* for 10 minutes at 20°C. To collect any residual proteins, the filter column was rinsed with 10 µl 1 M TEAB and centrifuged at 1000 x *g*. Trypsin (Promega) digestion of the protein samples was conducted in trypsin buffer (0.5 M TEAB) at a final protein:trypsin (µg) ratio of 100:1, with a final pH of >8. The digestion reactions were conducted at 37°C for at least 18 hours (Figure 1).

2.2.8 Proteomic profiling using iTRAQ

A total of five independent iTRAQ experiments were conducted. Experiment 1 was carried out to identify time points of interest for further investigation (henceforth referred to as "initial iTRAQ"). Experiments 2 to 5 were conducted to identify and functionally analyse haemocyte proteins that were differentially expressed in abalone exposed to reduced pH conditions (henceforth referred to as "experimental iTRAQ").

2.2.8.1 iTRAQ® Labelling

In preparing the iTRAQ® reagents, 50 µl of isopropanol was used to solubilize the labelling reagents. To each sample of 300 ug (Section 2.2.7.1) of peptides, one of the 8 iTRAQ® labelling reagents was added to ensure a final concentration of 60% (*v/v*). The labelling procedure was followed as per the manufacturer's instructions (iTRAQ® Reagents-8 plex, AB Sciex Pte. Ltd.). Thereafter, the 8 samples were combined for sample clean-up, and henceforth referred to as "pooled samples" (Figure 1).

2.2.8.2 Isoelectric focussing (IEF) of peptides

Reverse-phase C18 resin columns were used to remove contaminants remaining in the biological samples to ensure efficient isoelectric focussing (IEF) and subsequent LC-MS/MS. The pooled samples (Section 2.2.8.1) were split across 8 Thermo Scientific Pierce C18 Spin Columns, following the manufacturer's instructions, to yield a total of 250 µg of peptide sample. Cleaned samples were concentrated using the Savant Speedvac® Plus SC210A (ThermoQuest). The samples were each separated into 12 fractions according to the isoelectric points of the peptides. Fractionation (Figure 1) was carried using a 12-well 3100 OFFGEL Fractionator (Agilent Technologies, USA) according to the manufacturer's instructions. In brief, pH gradient strips (Immobiline Dry-strip pH 4-7, 13 cm) (GE Healthcare, UK) were rehydrated with IPG strip Rehydration Solution (GE Healthcare) for 15 minutes. Peptide OFFGEL Stock solution (GE Healthcare) was added to the peptide samples and they were subsequently loaded into each well. Each end of the strip was hydrated with mineral oil and the wells were sealed prior to focussing at 20°C at 20 kVh with a maximum current set to 50 µA and a maximum voltage of 8000 V.

2.2.9 LC-MS/MS analysis

In order to remove contaminants, in the form of residual IEF reagents, from the fractionated peptide samples and thus prevent clogging of the MS column, each fraction was desalted prior to LC-MS/MS analysis using reverse phase Pierce C18 Spin Columns (Thermo Fisher Scientific, U.S.A). The iTRAQ labelled fractionated samples were analysed using a Q-Exactive quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, U.S.A) coupled with a Dionex Ultimate 3000 nano-HPLC system (Figure 1). Extracted peptides were resuspended in loading buffer (97.5% water, 2.5% Acetonitrile, 0.1% Formic acid) and subsequently loaded on a C18 trap column (300 µm × 5mm × 5 µm). A C18 column (75 µm × 250 mm ×2 µm) was utilised for chromatographic separation. A two-solvent system was employed: Solvent A (water and 0.1% Formic acid) and solvent B (Acetonitrile and 0.1% Formic acid). The gradient for peptide separation was achieved at 250 nL/min as follows:

Time change: 77 min

Gradient change: 8 – 23% solvent B

The mass spectrometer was utilised in positive ion mode, with a capillary temperature of 250°C, while the applied electrospray voltage was 1.95 kV. Data acquisition is detailed below:

Full scan:

Resolution	70.000 (at *m/z* 200)
AGC target value	3e6
Scan range	320 – 1750 m/z
Maximal injection time (ms)	100

Data-dependent MS/MS:

Inclusion	Off
Resolution	17.500 (at *m/z* 200)
AGC target value	1e5
Maximal injection time (ms)	75
Loop count	15
Isolation window width (Da)	1
First fixed mass	100.0 m/z
NCE (%)	32

Data-dependent settings:

Underfill ratio (%)	1
Charge exclusion	Charge states 1.6-8>8
Peptide match	Preferred

Exclusion isotopes	On
Dynamic exclusion (s)	60

2.2.10 Data analysis and peptide identification using PEAKS Studio 8.5

Raw data files (.mgf) generated from LC-MS/MS analysis of each fractionated sample were converted to .mzXML format using MSConvert for processing on PEAKS Studio 8.5 (Bioinformatics Solutions Inc.) (Figure 1). The software allows for simultaneous peptide sequencing, protein identification and quantification of tandem mass spectrometry data. PEAKS utilises protein databases and *de novo* sequencing-assisted search engines to provide complete peptide sequences, as well as providing post-translational modification characterisation. Database search parameters were set as follows:

Mass error tolerance:
Precursor mass: 20.0 ppm Fragment ion: 0.1 Da
Enzyme: Trypsin
Non-specific cleavage: 1
Maximum missed cleave: 3

PTM:
Fixed: Carbamidomethylation (57.02) Variable: Oxidation (m) (15.99)
iTRAQ 8-plex (K, N term) (304.2)
Ion source: ESI (nano-spay)
Fragmentation mode: CID, CAD (y & b ions)
MS Scan mode: FT-ICR/Orbitrap
MS/MS Scan mode: Linear ion trap
FDR threshold (%): 1.0
Reporter ion type: MS2

Database: The Uniprot database was accessed on the 8th of August 2018 to retrieve all updated molluscan protein sequences.

Identified and quantified proteomics data obtained from the four biological replicates were combined and manually filtered in Excel. The criteria for filtering included removing duplicate proteins with low -10lgP values, proteins with less than two unique peptides, and proteins with a low percentage coverage. Additionally, proteins that had missing expression data were also removed from the list as these would skew the subsequent analyses.

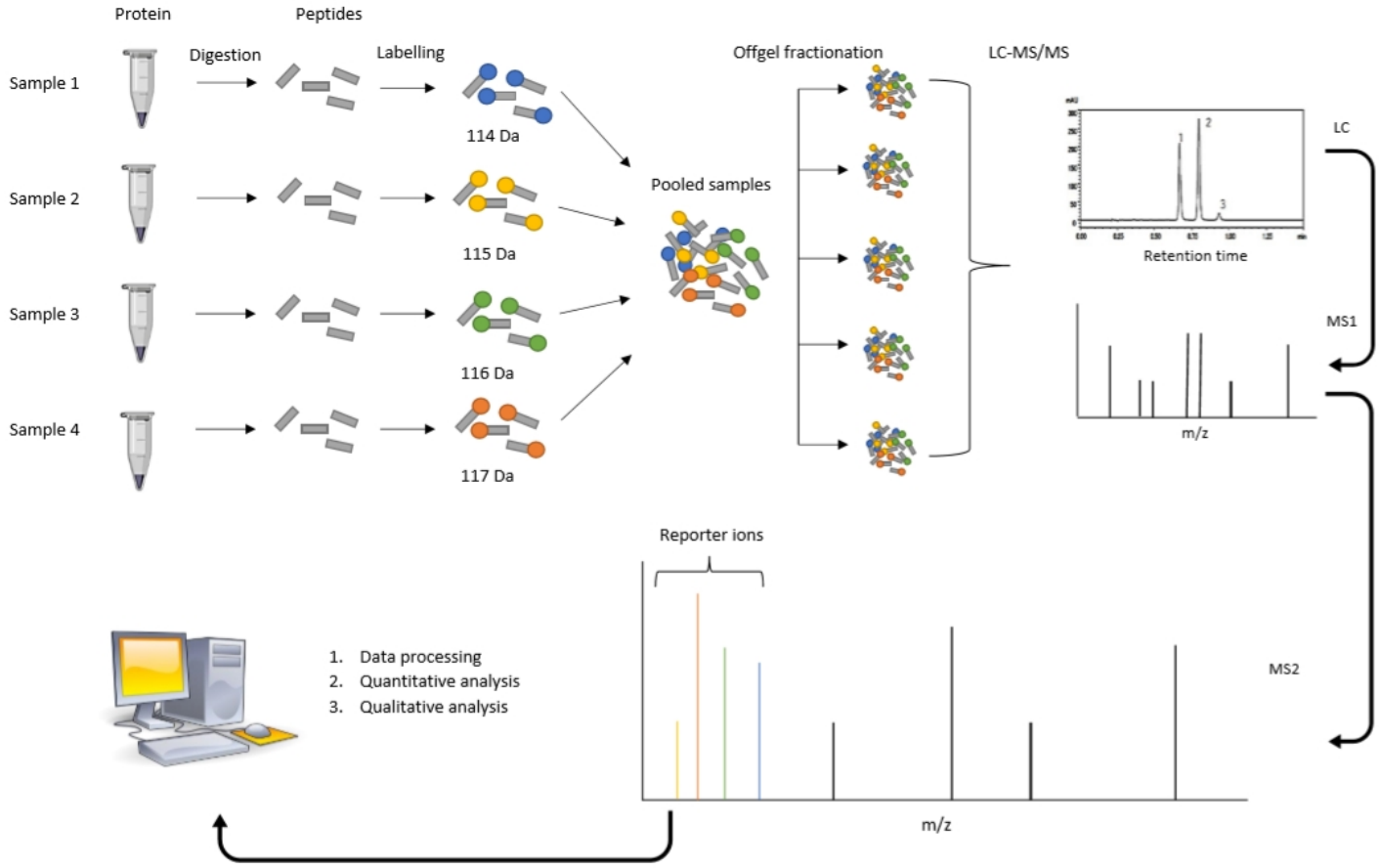

Figure 1. Schematic view of the workflow employed in this iTRAQ study to characterise the haemocyte proteome of *H. midae* following exposure to reduced pH conditions for 12, 72 and 168 hours, utilising an 8-plex iTRAQ labelling strategy. Note, only 4 of the 8 labels have been included in the schematic for simplicity. Isolated protein from haemocytes sampled from *H. midae* was prepared using FASP and subsequently trypsin digested. Peptides of samples were correspondingly labelled with iTRAQ reagents and subsequently pooled and fractionated to reduce sample complexity and ensure equal variation across samples. Following this, labelled peptides were analysed using an LC-MS/MS approach, which couples MS1 and MS2 data for protein quantification and subsequent identification in Peaks Studio.

2.2.11 Quantitative analysis of iTRAQ expression data

The filtered, collated quantitative data was imported into Perseus 1.6.5 (Tyanova *et al.*, 2016b), which is an open-source computational platform for comprehensive proteomics analysis and statistics. The imported data was further refined to ensure that the data was normally distributed. This was achieved through Log2 transformation and normalisation of all expression data to the T0 (control) values which served as the reference (Tyanova *et al.*, 2016b; Tyanova and Cox, 2018). A Two-Sample T-Test was conducted on the normalised expression data to identify proteins that were significantly ($p \leq 0.05$) differentially expressed at each sampling time point. The expression data obtained from each of the four biological replicates was averaged to facilitate further downstream analyses.

To observe the overall proteomic profile of *H. midae* haemocytes sampled from abalone exposed to sea water with a reduced pH for various time intervals, hierarchical clustering analysis was conducted in Perseus employing the Euclidean distance metric with Average linkage and pre-processing with k-means.

2.3 Results & Discussion

The South African aquaculture sector has continued to grow despite the unforgiving terrain of the coastline, with well-developed infrastructure, relatively efficient supply chains and a growing global demand for seafood (Britz and Venter, 2016). Abalone farming in the Western Cape is well established. Although animal nutrition, stocking density, aerial exposure, water flow rates and handling times are effectively established and monitored, abiotic conditions such as temperature, pH and salinity remain uncontrolled, particularly with regard to on-shore coastal farms that rely on an onshore-pump system for supplying header tanks (Morash and Alter, 2016). Multiple stressors have been found to compromise the coping ability of abalone (Morash and Alter, 2016). As such, elucidating the molecular stress response of *Haliotis midae* to various changing abiotic conditions, specifically reduced pH in this study, is critical in mitigating losses, particularly in the face of climate change.

In this study, *H. midae* were exposed to reduced pH conditions which are representative of climatic conditions predicted for 2100. In so doing, we sought to characterise the haemocyte proteome of *H. midae* in response to the stressor. Haemocytes are involved in an array of functions, such as nutrient uptake, shell and tissue repair, and transportation and distribution (Donaghy *et al.*, 2012). Furthermore, haemocytes are essential in mediating internal cellular defence mechanisms, thus being the chief drivers of the innate immune system in marine invertebrates (Barcia and Ramos-Martínez, 2011; Donaghy *et al.*, 2010). Therefore, haemocytes were selected as appropriate cells for investigating the effects of reduced ocean pH on farmed abalone.

2.3.1 Physiological immune response following OA exposure

To examine the physiological response of *H. midae* haemocytes following exposure to reduced pH conditions, phagocytic activity was assessed in duplicate samples taken from abalone at the respective exposure time points utilising heat-killed *Vibrio anguillarum,* a common marine bacterial pathogen. Phagocytosis is a cellular immune response undertaken by circulating haemocytes, that involves recognising pathogen-associated molecular patterns and the subsequent engulfment of the invading pathogen (Wang *et al.*, 2018). In some instances, haemocytes have been found to dispose of damaged tissues and cells through phagocytosing

cells and melanisation-related processes (Meseck *et al.*, 2016). As such, invading bacteria are destroyed through enzymatic activity and through the release of oxygen metabolites (Wang *et al.*, 2018). Varying environmental conditions have been found to affect phagocytosis, where a change from basal levels has been attributed to a stress response in some instances (Green *et al.*, 2017; Wu *et al.*, 2018).

The average percentage of phagocytosing haemocytes sampled from *H. midae* exposed to reduced pH conditions (pH 7.5) for 12 hours increased by 14% (Figure 2). The percentage of haemocytes able to elicit a phagocytic response subsequently increased to 40% after 72 hours of exposure to reduced pH, thereafter returning to just above the basal (control/time 0) level at 168 hours (1 week) of exposure to the stressor.

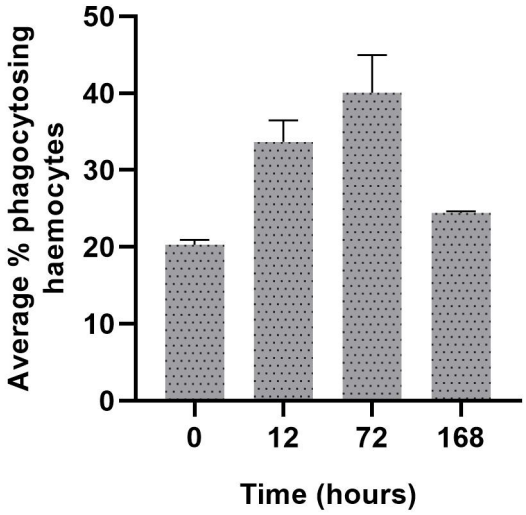

Figure 2. Mean percentage of phagocytosing *H. midae* haemocytes sampled after 0, 12, 72 and 168 hours of exposure to reduced pH conditions (pH 7.5). Error bars are the standard error of the means of two microscope slide replicate counts of phagocytosing haemocytes.

In a study investigating the effects of ocean acidification on the immune response of *Crassostrea gigas*, the authors utilised three pCO_2 levels representative of present-day conditions and conditions predicted for 2100 (~pCO_2 1000 µatm) and 2250 (~pCO_2 2000 µatm) (Wang *et al.*, 2016). The phagocytic activity of *C. gigas* haemocytes after 7 days of exposure to 2100 conditions was reduced compared to the phagocytic activity under present-day conditions, while 2250 conditions resulted in an increase in phagocytic activity, similar to that

observed in the current study (Figure 2). However, the authors found that there was no significant difference in phagocytic activity between treatments, though ROS production was significantly elevated after 28 days of exposure to 2250 CO_2 conditions. This suggested that ocean acidification might result in oxidative stress in haemocytes (Wang *et al.*, 2016). Conversely, Green *et al.* (2017) reported reduced phagocytic rates in *C. gigas* and *O. edulis* in response to elevated levels of pCO_2 after a 61-day exposure period.

The increase in phagocytic activity observed in the current study could be attributed to increased clearing and recycling of dead haemocytes due to apoptosis under reduced pH conditions. Indeed, this possibility was proposed following the observed significant increase in phagocytic activity in Tanner crab (*Chionoecetes bairdi*) haemocytes under low pH conditions (Meseck *et al.*, 2016). Since the number of circulating dead haemocytes was not measured in the current study, this is a parameter that should be investigated further in future studies. Furthermore, more biological replicates will need to be included in future phagocytosis assays to determine the statistical significance of this immune parameter, as well as investigating ROS production and phagocytic index, under this abiotic stressor.

With the discrepancies observed in phagocytic activity of haemocytes from various marine invertebrates under ocean acidification conditions, it must be taken into account that these organisms are different model systems and experience different degrees of environmental change (Figueiredo *et al.*, 2016). Varying stress and immunological responses of various marine invertebrates to ocean acidification is likely to be related to species-specific adaptations (Cunningham *et al.*, 2016). However, comparing different studies *to*, and elucidating the phagocytic response *of*, *H. midae* haemocytes to changing environmental conditions will aid in gaining further insight into the tolerance level and adaptability of this economically important marine invertebrate.

2.3.2 Proteomic profiling of haemocytes sampled from stressed H. midae

The initial exploratory iTRAQ experiment revealed that *H. midae* experienced an early proteomic response to reduced pH conditions rather than adverse effects from prolonged exposure (data not shown). Differential protein expression was observed in haemocytes sampled between 12 and 168 hours of exposure to the stressor, providing the basis upon which our final iTRAQ experimental time course was designed. The same aquarium-based experimental design was employed, and haemolymph was sampled after 12, 72 and 168 hours

of exposure to the stressor. The experiment was repeated to ensure a total of 4 independent biological replicates were included in the study.

Mass spectrometry data processing was conducted using PEAKS Studio 8.5 (Table 1). Since an 8-plex iTRAQ set-up was utilised, final iTRAQ experiments 1 and 2 were labelled and processed simultaneously, while trials 3 and 4 were conducted concurrently.

Table 1. Summary of the LC-MS/MS analysis of the *H. midae* haemocyte proteome following exposure to reduced pH (7.5) conditions.

	iTRAQ 1 & 2	iTRAQ 3 & 4
MS/MS Scans [a]	575335	216539
PSM [b]	8547	8135
Proteins before filtering [c]	6386	5624
Proteins after filtering [c]	227	

[a] Number of MS/MS Scans following LC-MS/MS analysis on PEAKS. [b] Number of peptide-spectrum matches following LC-MS/MS analysis. [c] Number of proteins identified before and after merging and filtering redundant proteins.

The use of -omics techniques for elucidating the molecular response of marine organisms to environmental change is becoming increasingly popular. Tomanek *et al.* (2011) utilised a proteomics workflow to gain insight into the mantle tissue proteome of *C. virginica* following exposure to reduced seawater pH conditions. Proving to be an excellent tool for elucidating uncharacterised responses to environmental changes, their proteomics approach showed that 12% of all mantle tissue proteins changed in their abundance under the stressor. While there has been limited molluscan genomic information available, the authors were able to identify 17 differentially expressed proteins, with 31% of them changing significantly (Tomanek *et al.*, 2011). Campos *et al.* (2015) employed a high-throughput proteomics approach that successfully enhanced and contributed to the understanding of the molecular immune defences of *M. edulis.* With the lack of available genetic information for marine invertebrates, the authors integrated an analysis of transcriptomic resources for their characterisation of the

mussel haemocyte proteome. This facilitated the identification and annotation of 595 proteins, and validated previously obtained genetic data (Campos *et al.*, 2015). Furthermore, the integration of various -omics approaches, such as metabolomics and proteomics in the case of Wei *et al.* (2015a, b), has proven useful in effectively annotating identified gene products, as well as providing greater insight into the effects of various environmental stressors on marine invertebrates at a molecular level. Thus, future investigations of the immune/stress response of marine invertebrates would benefit from holistic and integrative -omics approaches.

Shotgun proteomics has been identified as a versatile approach for protein identification and functional analysis of proteomes. This is particularly true for aquaculture (Rodrigues *et al.*, 2017) and climate change research, with strategies like SILAC, ICAT and iTRAQ labelling largely aiding biomarker discovery. In a comparative study of three proteomic quantitative methods (DIGE, cICAT, iTRAQ), the quantification sensitivity of iTRAQ proved most successful (Wu *et al.*, 2006). Although there are a number of constraints associated with non-model organisms, iTRAQ served as a suitable tool for conducting a global protein expression analysis on the *H. midae* haemocyte proteome in response to ocean acidification conditions.

Following data processing and filtering, a total of 227 differentially expressed proteins were identified in *H. midae* in response to varying degrees of exposure to reduced pH conditions (Table 1). Hierarchical clustering analysis provides a graphical representation of the proteomic profile of *H. midae* haemocytes under the stressor, while also grouping similarly expressed proteins (Figure 3).

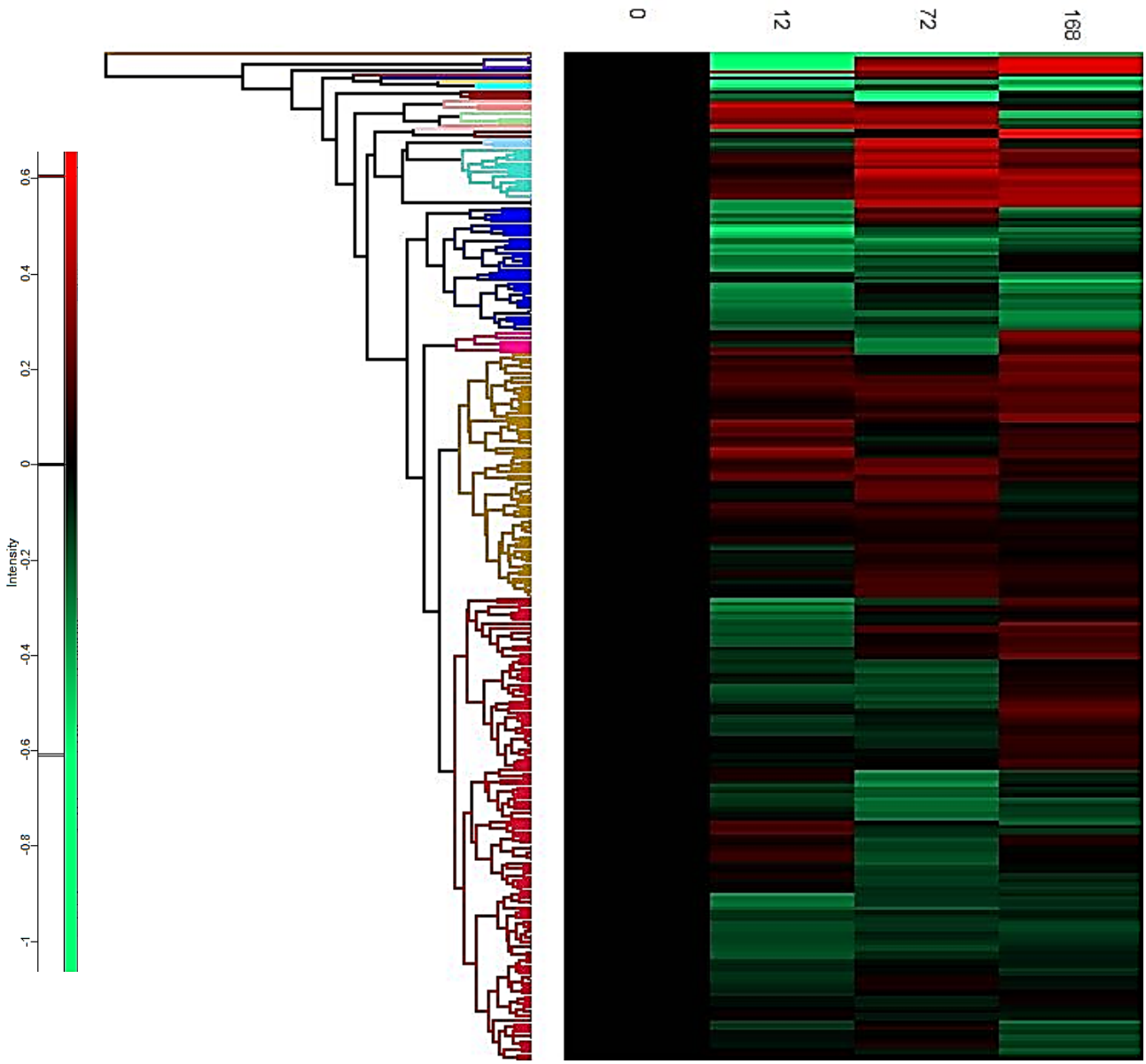

Figure 3. Hierarchical cluster analysis, using a Euclidean distance and Average linkage matrix, of the expression profiles of the 227 proteins identified in *H. midae* haemocytes following exposure to reduced pH conditions for 0, 12, 72 and 168 hours. Protein intensity (≈ expression) is indicated by the red and green coloured scale bar, where green indicates "negative" expression (downregulation), while red indicates "positive" expression (upregulation). A total of 15 clusters were identified, which are indicated by the different coloured arms of the dendrogram. A large proportion of the proteins were predominantly downregulated, while a small group of highly upregulated proteins occur at the top portion of the dendrogram.

2.3.3 Differentially expressed proteins identified in *H. midae* haemocytes

In order to identify proteins that are significantly upregulated ($p \leq 0.05$) in abalone exposed to reduced pH, a two-sample Student's T-test with a permutation-based FDR for truncation was conducted on the normalised expression data. Statistical analyses were conducted on protein expression data from *H. midae* haemocytes sampled at 12, 72 and 168 hours. To ensure that proteins that are biologically relevant (Dalman *et al.*, 2012), but not statistically significant, were not eliminated from the dataset, proteins that had a fold-change ≥ 1.2 (equivalent to an approximate 20% increase in protein expression) were retained. Tables 2-4 present proteins that were statistically significant in their expression at the respective exposure times and/or were found to be of potential biological importance in terms of their fold change. A total of 18 upregulated proteins were identified after 12 hours of exposure, 30 upregulated proteins following 72 hours of exposure, and 29 upregulated proteins following 168 hours of exposure to reduced pH (Tables 2-4).

Table 2. Upregulated proteins identified in *H. midae* haemocytes following 12 hours of exposure to reduced pH conditions (pH 7.5). A Two-sample Student's T-test with a permutation-based FDR ($p \leq 0.05$) for truncation was conducted for each protein identified at this time point, while a fold-change cut-off of ≥ 1.2 was set to include proteins that could be of biological importance.

Uniprot accession number	Protein name	P-value	Fold change
A0A2T7PHL9_POMCA	60S ribosomal protein L23a	1.00	1.3
A0A2T7P5E9_POMCA	60S ribosomal protein L24-like	1.00	1.3
A0A2T7PIY3_POMCA	60S ribosomal protein L27a	1.00	1.3
A0A2C9KD73_BIOGL	Actin, plasmodial isoform-like	1.00	1.4
K1PCW5_CRAGI	Alpha-1 3-mannosyltransferase ALG2	1.00	1.2
A0A2C9JQJ6_BIOGL	Alpha-L-fucosidase-like	0.04	1.1
A0A1I9W307_HALDH	Aminopeptidase	0.00	1.2
B3TK60_HALDV	Calmodulin-dependent protein kinase (Fragment)	0.00	1.4
E6Y9P6_BIOGL	Cdc24-like protein	0.04	1.1
Q86DH9_APLCA	Cdc42	0.04	1.1
B6RB18_HALDI	Chaperonin containing tcp1	0.05	1.2
V3YZ86_LOTGI	Deoxyribose-phosphate aldolase-like	0.00	1.1
A0A1L7H9J9_HALDI	Mitogen-activated protein kinase	0.05	1.4
B6RB86_HALDI	Mitogen-activated protein kinase (Fragment)	0.01	1.2
K1R5M2_CRAGI	Myosin-IXa	0.28	1.2
V4A8R6_LOTGI	Putative uncharacterized oxidoreductase YDR541C	0.01	1.3
O61472_APLCA	Reductase-related protein	0.02	1.2
A0A2T7P0L9_POMCA	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	0.01	1.1

Table 3. Upregulated proteins identified in *H. midae* haemocytes following 72 hours of exposure to reduced pH conditions (pH 7.5). A Two-sample Student's T-test with a permutation-based FDR ($p \leq 0.05$) for truncation was conducted for each protein identified at this time point, while a fold-change cut-off of ≥ 1.2 was set to include proteins that could be of biological importance.

Uniprot accession number	Protein name	P-value	Fold change
A0A2T7P5E9_POMCA	60S ribosomal protein L24-like	1.00	1.3
A0A2C9KD73_BIOGL	actin, plasmodial isoform-like	1.00	1.3
A0A2C9JUN2_BIOGL	actin-5C	0.45	1.2
A8DU73_9MOLL	Alpha enolase (Fragment)	0.28	1.2
A0A2T7P4L5_POMCA	Beta-parvin-like isoform X2	0.02	1.2
H9AWU2_HALDH	Chaperonin containing T-complex polypeptide subunit zeta	0.49	1.3
B6RB18_HALDI	Chaperonin containing tcp1	0.01	1.1
K1RM18_CRAGI	Dual oxidase	0.42	1.2
A0A0L8FWA8_OCTBM	Elongation factor 2-like	0.07	1.4
K1QX37_CRAGI	Enolase	0.23	1.2
K1QEZ3_CRAGI	Fascin	0.31	1.3
A0A2T7PEM0_POMCA	Hemocyanin G-type, units Oda to Odg-like	0.12	1.3
K1QNQ9_CRAGI	Hypothetical protein CGI_10024082	0.12	1.3
A0A2T7PHH2_POMCA	Major vault protein-like	0.26	1.3
A0A1L7H9J9_HALDI	Mitogen-activated protein kinase	0.04	1.1
A0A286QYA2_HALDH	Paramyosin	0.50	1.2
V4ADA9_LOTGI	Phosphoglycerate kinase	0.04	1.4
V3ZW18_LOTGI	probable methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	0.24	1.2
A0A0B6Y9I8_9EUPU	probable small nuclear ribonucleoprotein Sm D2	0.12	1.4
V4A8R6_LOTGI	Putative uncharacterized oxidoreductase YDR541C	0.41	1.2

A0A0B7BF17_9EUPU	Radixin isoform X1	0.33	1.2
A0A210QG69_MIZYE	Ras GTPase-activating-like protein IQGAP1	0.08	1.4
A0A2T7NG09_POMCA	Ras GTPase-activating-like protein IQGAP1 isoform X1	0.14	1.3
Q70MN8_CRAGI	Ribosomal protein L24 (Fragment)	0.11	1.3
K1QG65_CRAGI	rRNA 2'-O-methyltransferase fibrillarin	0.52	1.4
A0A0G2SSZ6_HALDI	Serpin B-like protein 2	0.28	1.3
Q8MUE3_9BIVA	Small nuclear ribonucleoprotein D2-like protein	0.28	1.2
A0A0B6ZMQ4_9EUPU	Solute carrier family 2. facilitated glucose transporter member 1	0.01	1.2
A0A0L8H0E1_OCTBM	Sorting nexin	0.10	1.4
A0A2C9JRW8_BIOGL	uncharacterized oxidoreductase YajO-like	0.39	1.4

Table 4. Upregulated proteins identified in *H. midae* haemocytes following 168 hours of exposure to reduced pH conditions (pH 7.5). A Two-sample Student's T-test with a permutation-based FDR ($p \leq 0.05$) for truncation was conducted for each protein identified at this time point, while a fold-change cut-off of ≥ 1.2 was set to include proteins that could be of biological importance.

Uniprot accession number	Protein name	P-value	Fold change
K1RH70_CRAGI	6-phosphogluconate dehydrogenase decarboxylating	0.03	1.2
V4BUJ1_LOTGI	78 kDa glucose-regulated protein	0.05	1.2
K1RG10_CRAGI	AP complex subunit sigma	0.04	1.1
B3TK60_HALDV	Calmodulin-dependent protein kinase (Fragment)	0.02	1.5
B6RB18_HALDI	Chaperonin containing tcp1	0.01	1.1
A0A0L8FL54_OCTBM	charged multivesicular body protein 1a-like	0.06	1.6
A0A2C9JEU7_BIOGL	cytoplasmic dynein 1 heavy chain 1-like isoform X2	1.00	1.2
A0A0L8FWA8_OCTBM	Elongation factor 2-like	1.00	1.3
K1QX37_CRAGI	Enolase	0.04	1.3
B6RB30_HALDI	Glyceraldehyde-3-phosphate dehydrogenase	0.01	1.2
A0A2T7PEM0_POMCA	Haemocyanin G-type, units Oda to Odg-like	0.55	1.4
K0J2U1_9BIVA	Histone 3 (Fragment)	0.35	1.2
A0A2T7P1P2_POMCA	Kelch-like protein 24 isoform X1	0.01	1.2
V4C6I3_LOTGI	Major vault protein-like	0.04	1.3
B6RB86_HALDI	Mitogen-activated protein kinase (Fragment)	0.04	1.1
A0A2C9JK44_BIOGL	Obg-like ATPase 1	0.05	1.2
A0A286QYA2_HALDH	Paramyosin	0.07	1.3
V3ZW18_LOTGI	Probable methylmalonate-semialdehyde dehydrogenase [acylating]. Mitochondrial	0.02	1.3
A0A0B6Y9I8_9EUPU	Probable small nuclear ribonucleoprotein Sm D2	0.07	1.4
A0A0B7BF17_9EUPU	Radixin isoform X1	0.30	1.2
A0A210QG69_MIZYE	Ras GTPase-activating-like protein IQGAP1	0.17	1.2
O61472_APLCA	Reductase-related protein	0.06	1.2

A0A210Q7D8_MIZYE	Retinal dehydrogenase 2	0.00	1.2
Q8MUE3_9BIVA	Small nuclear ribonucleoprotein D2-like protein	0.52	1.5
A0A0B6ZMQ4_9EUPU	Solute carrier family 2, facilitated glucose transporter member 1	0.29	1.3
A0A0L8G346_OCTBM	SWI/SNF complex subunit SMARCC2-like	0.00	1.1
A0A0G2R2A9_9BIVA	Tubulin beta chain (Fragment)	0.17	1.2
A0A2C9LJS7_BIOGL	uncharacterized oxidoreductase YajO-like	0.42	1.3
A0A210PT48_MIZYE	Vacuolar protein sorting-associated protein 13A	0.14	1.6

A number of metabolic proteins (i.e. aminopeptidase, alpha-L-fucosidase, reductase-related protein, serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform, enolase, glyceraldehyde-3-phosphate dehydrogenase, Obg-like ATPase 1), signal transduction proteins (tubulin beta chain, mitogen-activated protein kinase, calmodulin-dependent protein kinase, major vault protein, chaperonin containing tcp1, serpin B, Ras GTPase-activating IQGAP1) and cytoskeletal proteins (i.e. radixin, fascin, myosin IXa) were found to be either statistically significantly upregulated, biologically relevant (fold-change ≥1.2) or both in haemocytes sampled from *H. midae* exposed to decreased seawater pH for specific time periods (Tables 2-4).

In a 4-week study investigating the effects of elevated pCO_2 on the proteome of *C. gigas*, Timmins-Schiffman *et al.* (2014) identified 48 differentially expressed proteins across four treatment conditions, which served to represent the general proteomic stress response of the Pacific oyster to elevated pCO_2. Functional analysis revealed a range of biological processes associated with several of the identified proteins. As in the current study, the authors identified proteins such as α-L-fucosidase (carbohydrate metabolism), mitogen-activated protein kinase and calmodulin dependent protein kinase (signalling), aminopeptidases (polypeptide and protein degradation), heat shock protein 90 and co-chaperones Cdc24 and Cdc42 (stress response), as well as several ribosomal proteins (transcription and translation) (Timmins-Schiffman *et al.*, 2014). Changes observed in the expression of proteins involved in carbohydrate metabolism suggest shifts in energetic costs when exposed to elevated pCO_2, particularly where glucose production is concerned (Timmins-Schiffman *et al.*, 2014).

Ocean acidification has been reported to affect the metabolism and energetics of marine invertebrates, especially in the larval stages. In the current study, a number of metabolically-related proteins were found to be significantly ($p < 0.05$) upregulated or biologically relevant (fold-change ≥ 1.2) following exposure to reduced pH conditions (Tables 2-4). Proteins such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 6-phosphogluconate dehydrogenase, solute carrier family 2, retinal dehydrogenase and phosphoglycerate kinase are among the proteins that are largely involved in ATP synthesis and carbohydrate metabolism. Interestingly, Chang *et al.* (2016) found GAPDH to be downregulated in brine shrimp (*Artemia sinica*) under seawater acidification. While some metabolic proteins were upregulated in their study, the authors suggested that the observed downregulation of proteins such as GAPDH and cytochrome *c* oxidase, which are key players in energy production and anti-apoptosis mechanisms, could be due to a shift in metabolism towards the glycolysis pathway (Chang *et al.*, 2016). The upregulated metabolic proteins identified in this study are largely associated with anaerobic metabolism, suggesting a possible shift in metabolic pathways in *H. midae* under reduced pH conditions.

In a study investigating the energetics of *Littorina littorea* exposed to elevated temperature and reduced pH conditions (both in isolation and in combination), the authors observed an increased reliance on anaerobic metabolism (Melatunan *et al.*, 2011). As such, the observed chronic imbalance between energy production and energetic costs in *L. littorea* is postulated to make them more vulnerable under changing environmental conditions in the future. Hüning *et al.* (2013) investigated the effects of ocean acidification on mantle tissue gene expression in the Baltic Sea blue mussel, where they found a depression in aerobic metabolism with a predicted shift towards anaerobic metabolism to support ATP demand under stress conditions. While it has been suggested that sustained metabolic depression and induction of anaerobic metabolism, under environmental stressors, can lead to reduced growth rates and mortality in some cases (Stevens and Gobler, 2018), no mortalities were observed in the current study. This might be attributed to a greater hypercapnia tolerance range in *H. midae*, or alternatively, the period of exposure to reduced pH conditions wasn't long enough to elicit total tissue damage and subsequent mortality.

In the proteomics study by Chang *et al.* (2016), of the 67 differentially expressed proteins identified, several were found to be involved in stress and immune responses such as peroxiredoxin, thioredoxin peroxidase, 70 kDa heat shock protein, prophenoloxidase and ferritin. These were likely induced as part of an oxidative stress response, as well as in the

induction of melanisation in response to damaged tissue (Chang *et al.*, 2016). Although most of these proteins were not observed in the current study, other stress and immune related proteins were found to be upregulated in *H. midae* haemocytes (Tables 2-4). Proteins such as ras GTPase-activating-like protein IQGAP1 isoform X1, fascin, serpin B-like protein 2, histone H3, 78 kDa glucose-regulated protein, haemocyanin G-type and major vault protein are associated with stress and immune related responses. Melanisation is a phagocytic activity that is part of the proPO system which is induced upon pathogenic infection or following tissue damage, and is a major component of the innate immune system of marine invertebrates (Duan *et al.*, 2013; Meseck *et al.*, 2016). Phenoloxidase and phenoloxidase-like proteins (such as histone H3), as well as serine protease inhibitors, are some of the key players involved in the melanisation cascade. Indeed, some studies have reported varying levels of phenoloxidase and serine protease inhibitors activity under different pH and temperature treatments. For example, Mukherjee *et al.* (2016) found significantly higher phenoloxidase activity following exposure of *Eunapius carteri*, a freshwater sponge, at pH 7.4 and pH 8.4, compared to those maintained at pH 6.4. A similar observation was made in the present study (i.e. histone H3 and serpin B2) which employed an experimental pH of 7.5, thus making it comparable to the mid-level pH treatment (7.4) in the study by Mukherjee *et al.* (2016). Conversely, phagocytic and melanisation activity was noticeably less effective in removing dead haemocytes in Tanner crabs maintained at the low pH treatment of pH 7.5 (Meseck *et al.*, 2016). It has been reported that the internal pH of haemocytes involved in carrying out the proPO cascade affects the success of this component of the innate immune system, where the pH range can be between pH 6.0 and 8.0, and is largely dependent on the species (Meseck *et al.*, 2016). This might explain the possible species-specific results observed in *H. midae* exposed to the experimental pH (pH 7.5) since OA causes extracellular acidosis that adversely affects intracellular pH (pH_i) (Salameh *et al.*, 2014; Zhao *et al.*, 2017), which could induce the upregulation of proPO-like proteins if pH_i falls within this range. This hypothesis requires validation through further investigation.

A combined metabolome and proteome analysis of the effects of reduced seawater pH on the mantle tissue of *C. gigas* found differentially expressed proteins predominantly involved in energy metabolism and cytoskeleton functions (Wei *et al.*, 2015a, b). Besides the microtubule-related proteins, such as myosin essential light chain and myosin regulatory light chain A, calponin-2 and tropomyosins, other upregulated cytoskeletal proteins involved in cytoskeletal organisation were identified in the Pacific oyster. In a multi-stressor study investigating the

effects of ocean acidification on the immune function of the Pacific oyster, cytoskeletal proteins were among the upregulated proteins identified in response to single and combined stressors (Cao *et al.*, 2018b). In the present study, upregulated cytoskeletal-related proteins identified in *H. midae* under stress conditions included ras GTPase-activating-like protein IQGAP1 isoform X1, myosin-IXa, fascin, chaperonin containing T-complex polypeptide subunit zeta, radixin isoform X1 and tubulin beta chain (Tables 2-4). The cytoskeleton plays a crucial role in mediating cell motility and shape, intracellular movement, endocytosis and exocytosis, cellular component organisation and intercellular junction formation (Wei *et al.*, 2015a). Cytoskeletal-related proteins, such as those comprising the actin system, have been identified as being highly sensitive to oxidative stress, suggesting that increased cytoskeletal protein induction could be a compensatory mechanism adopted by haemocytes in response to increased protein turnover incurred by oxidative damage, which attempts to establish cellular homeostasis (Cao *et al.*, 2018b). Thus, the significant upregulation of the cytoskeletal proteins in *H midae* haemocytes may be attributed to the need to maintain homeostasis and remove damaged cells by phagocytosis.

Signal transduction and related functions have also been found to be affected by ocean acidification in marine invertebrates such as *C. gigas* (Timmins-Schiffman *et al.*, 2014). Similarly, several signalling-related proteins were identified as significantly upregulated in *H. midae* in the current study. Mitogen-activated protein kinase, calmodulin-dependent protein kinase, ras GTPase-activating-like protein IQGAP1 isoform X1, sorting nexin and major vault protein are involved in intracellular signalling and related signalling activities (Tables 2-4). Furthermore, a number of these proteins are involved in signalling related to cell proliferation, apoptosis, inflammation, endocytosis and phagocytic-related activities. Two signal transduction proteins, found to be significantly upregulated in *Artemia sinica* in response to ocean acidification, were found to be involved in the PI3K/AKT/mTOR and calcium-protein kinase C signalling pathways, respectively (Chang *et al.*, 2016). Several intracellular cell signalling proteins, involved in cell proliferation and cytoskeletal reorganisation, were significantly upregulated in the arctic pteropod, *Limacina helicina*, under reduced pH conditions (Koh *et al.*, 2015). It has been suggested that with changing water chemistry under ocean acidification conditions, shell dissolution might lead to increased Ca^{2+} concentration in the haemolymph of bivalves, which affects cellular metabolism and signal transduction pathways (Matozzo and Mari, 2011). This is particularly true for calcium-dependent pathways, many of which are involved in immune and stress-related functions.

In conclusion, shotgun proteomics utilising an isobaric labelling approach proved to be an invaluable tool for analysis of global protein expression in *H. midae* haemocytes sampled from abalone exposed to reduced pH conditions. A large proportion of the significantly upregulated and biologically relevant proteins identified amongst the 227 proteins are involved in metabolic-related processes, phagocytosis and related activities, as well as cytoskeleton functions and signal transduction. Proteomic profiling of *H. midae* haemocytes has provided a foundation on which functional analyses of key groups of proteins could be conducted, further elucidating the stress response of abalone to changing environmental conditions, particularly ocean acidification.

Chapter 3

Towards biomarker discovery in *Haliotis midae* haemocytes exposed to ocean acidification conditions, following iTRAQ analysis.

Contents

3.1 Introduction

Through advancements made in the -omics era, high-throughput sequencing and protein identification technologies have comprehensively transformed biological research at the systems level (Khatri *et al.*, 2012). Consequently, large-scale studies investigating the quantitative and functional responses of an organism's proteome are made possible. Such research produces large volumes of data that characteristically result in extensive "lists" of differentially expressed genes or proteins, which can prove useful in further elucidating their roles under a particular environmental condition. However, these lengthy "lists" can become cumbersome and daunting, as they lack contextualised functional information. Furthermore, there is no established methodical approach to dealing with large datasets, making analysis difficult from the outset.

This challenge has largely been addressed through grouping of long lists of genes/proteins into smaller, potentially related groups. This has led to the development and ongoing curation of knowledge resources and related computational platforms to facilitate this task (Huang *et al.*, 2009; Khatri *et al.*, 2012). Functional analysis of proteins into related groups underscores a large part of any proteomics-based study: the understanding of protein complexes, protein interactions and signalling pathways, as well as the interconnectedness of proteins; all within the context of a given environmental condition (Wu *et al.*, 2014). A plethora of bioinformatics solutions exist, based on the public databases available, which enable systems level and biological enrichment analyses of proteomics data. For instance, the Gene Ontology Consortium (GOC) was developed as part of a collaborative effort to provide evidence-based functional ontological information about gene products ('Gene Ontology Consortium: going forward', 2015). Through the curation of evidence-based biological databases, the GOC provides annotations of gene products in a biological context, aiding functional analysis of proteomics data. Additionally, KEGG (Kyoto Encyclopedia of Genes and Genomes) was initiated in 1995, as part of the Japanese Human Genome Program, following the increasing need for a reference knowledge source for the interpretation of genetic data. This venture led to the development of numerous function-specific resources, such as the KEGG PATHWAY database, which allows for the assignment of complex functions to queried genes in the form of pathway maps (Kanehisa *et al.*, 2016). It goes without saying that measurable strides have been made in the -omics arena, as well as in the curation of methodologies and genetic databases to facilitate biological interpretation of proteomics data in model organisms. This is further reinforced by the sheer number of bioinformatics tools that are based on already-

sequenced genomes, as well as prior annotation of their corresponding transcriptomes and proteomes (Kultz *et al.*, 2007).

Despite the advancements in model organisms, working on non-model organisms (e.g. *H. midae*) poses several challenges, most notably their low representation in protein databases (Campos *et al.*, 2016b). This has led to an increasing need for more research in non-model organisms, which would ultimately elucidate the mechanisms that underpin biological diversity (Armengaud *et al.*, 2014). As a result of the limitations incurred by the low representation of non-model organisms in public databases, the identification and characterisation of proteins can be hampered. As such, an increasingly strong emphasis is being placed on homology-based identification and biological inference to overcome this conundrum (Campos *et al.*, 2016b; Dowd, 2012). And in recent years, high-throughput technologies are facilitating transcriptomic and proteomic analyses of non-model organisms, ensuring the robustness of protein identification and functional analyses thereof from organisms with little known sequence data (Kultz *et al.*, 2007). Ultimately, proteomic studies of non-model organisms require a multifaceted approach that combines MS/MS analysis with *de novo* sequencing, in addition to cross-species homology-based protein identification (Russeth *et al.*, 2006; Vertommen *et al.*, 2011). Due to the hurdles experienced in -omics investigation of non-model organisms, bioinformatics tools are continuously improving their search algorithms and false discovery rate parameters; thus, highlighting the dynamic nature of these fields of research.

Following LC-MS/MS analysis, peptide identification from MS spectra is a crucial step in the proteomics pipeline. Several software programs have been developed for the algorithmic identification of proteins from spectra. However, the accuracy and robustness of these algorithms inherently affect the success of the identification process (Zhang *et al.*, 2012). The most common approach for this task relies on peptide matches from experimental MS/MS data with online databases. This is achieved through the calculation and matching of unknown peptides to observed peptides, following which the masses of the expected fragment ions are determined and compared with experimental masses (Cottrell, 2011). Naturally, there are several drawbacks with such an approach, especially when dealing with non-model candidates (Ma *et al.*, 2003; Zhang *et al.*, 2012). PEAKS studio, utilised in this study, employs *de novo* sequencing, in addition to database searches, for peptide identification. Based on a sophisticated algorithm, *de novo* sequencing entails peptide sequence determination directly from MS/MS spectra, which improves the filtering and scoring method for peptide matches

(Ma *et al.*, 2003; Zhang *et al.*, 2012). This facilitates and improves accurate peptide identification in studies involving non-model organisms.

MS/MS analysis with PEAKS generates curated "lists" of identified proteins with corresponding mass values, quantitative data, and other attributes. Processing and statistical analysis of this data constitutes the quantitative component (i.e. Chapter 2) of any proteomics study. Qualitative/functional analysis of proteomics data typically entails the integration of annotation data from Gene Ontology analyses with pathway databases, such as KEGG and Reactome (Kumar and Mann, 2009). These analyses provide functional insights into the identified groups of proteins, which are usually achievable with an assortment of bioinformatics tools and computational software. For example, Zhang *et al.* (2019) investigated the genes involved in an immune response to hypoxia/thermal stress in *Haliotis diversicolor*, using an integrative transcriptomic approach. They annotated their transcriptomic data using a combination of Swissprot and NCBI, while functional characterisation was achieved through the Gene Ontology Consortium and Clusters of Orthologous Groups (COG) analysis accessible through the National Center for Biotechnology Information (NCBI) database (Zhang *et al.*, 2019). To facilitate pathway analysis of their data, the authors utilised the KEGG PATHWAY database. They were able to identify potential key signalling pathways involved in immune regulation in response to the stressors, highlighting the importance of homology-related searches in discovery -omics studies (Zhang *et al.*, 2019).

In the age of an increasing number of studies utilising non-model organisms, elucidating the exact function of deduced proteins is becoming more reliant on homology-based classification, all in an attempt to predict biochemical and mechanistic functions (Tatusov *et al.*, 2000). COG achieves this through sequence homology based on identifying orthologs. Orthologs are conserved genes/proteins across taxa that have evolved from a common ancestral sequence, inferring a conserved mechanistic function (Kaufmann, 2006). What began as a means of genome-wide clustering of orthologs by Tatusov *et al.*, ultimately culminated in the curation of the COG central database, which is a critical tool in comparative genomics (Huerta-Cepas *et al.*, 2019; Kaufmann, 2006; Tatusov *et al.*, 2000). Wei *et al.* (2015a) conducted a comprehensive metabolomic and proteomic analysis of *Crassostrea gigas* mantle tissue in response to elevated pCO_2. The 11 differentially expressed proteins they identified were found to be predominantly involved in energy metabolism and cytoskeletal function, following COG protein classification (Wei *et al.*, 2015a). The distinct advantage COG protein classification has is that orthologs are predisposed to retaining their ancestral properties, with little variation,

unlike paralogs whose evolution arises from divergence from an ancestor (Huerta-Cepas *et al.*, 2019). Several databases have been developed to facilitate this approach to protein functional classification, such as eggNOG (Huerta-Cepas *et al.*, 2019), which can be utilised as a standalone resource or is implemented as part of other functional annotation software programs such as Blast2GO (Conesa *et al.*, 2005; Götz *et al.*, 2008).

Since functional annotation and interpretation have become crucial steps in the proteomics analysis pipeline, particularly with regards to novel and uncharacterised sequence data, more emphasis is being placed on the efficiency and reliability of functional annotation platforms and resources. Manual curation is immensely time consuming and often tedious, especially when one considers the speed at which novel genetic data is being obtained, resulting in increasing desirability of automatic functional annotation (Gotz *et al.*, 2008). Automating functional annotation relies on the sequence, structure, phylogenetic or co-expression parallels between sequenced and novel genetic data (Gotz *et al.*, 2008). Satisfying the criteria, Blast2GO was initially released in 2005 and has since undergone numerous updates to incorporate improved data mining and comprehensive functional annotation algorithms. Blast2GO serves as a comprehensive tool for functional annotation of high-throughput sequence data, utilising the Gene Ontology vocabulary and interactive GO graphs/charts to achieve this (Götz *et al.*, 2008). The initial step in Blast2GO functional annotation involves a homolog search between the queried sequence data and the NCBI-BLAST database. Functional analysis of large data sets using Blast2GO has proved useful for identifying and understanding the physiological meaning of experimental sequence data.

Functional analysis of proteins usually entails the elucidation of protein-protein interactions (PPI) within/between groups of proteins, as well as identifying key biochemical pathways and mechanisms. Investigating these aspects of the proteomics data is important for identifying proteins involved in key pathways, especially when studying temporally and spatially diverse tissues. Protein interactions describe how proteins associate with one another and other macromolecules to form complex metabolic/signalling pathways, to carry out particular functions (Keskin *et al.*, 2016). Numerous *in silico* bioinformatics tools have been developed and resources curated for the purpose of identifying and predicting PPIs. PPIs are identified using an increasing proportion of data curated from sequences, as well as studies investigating co-evolution, co-expression, structures, text-mining or even sub-cellular co-localisation (Hamp and Rost, 2015). PPI databases are generally characterised according to their curation method, for example: pathway databases are curated through consensus knowledge and

experimentation, such as Reactome, KEGG and IntAct; while some interaction databases are curated through computational predictions and text-mining (Szklarczyk and Jensen, 2015). Fortunately, for the latter type of PPI database, confidence scores accompany the PPI analysis, allowing the user to focus entirely on those interactions of statistical significance. The STRING database relies on the curation of primary and predicted interactions for PPIs at a broader scale, as well as integrating annotated pathway information, metadata mining, and inter-species transfer (Szklarczyk *et al.*, 2019). As such, the STRING database provides increased coverage and a reliable scoring system, as well as network visualisation tools that offer an exportable format for visualisation tools such as Cytoscape (Szklarczyk and Jensen, 2015). Currently, STRING hosts the largest number of organisms (5090) and proteins (24.6 million), as well as having protein-sequence-searching capabilities, allowing for homology-based searches (Szklarczyk *et al.*, 2019). This makes it a suitable tool for the prediction of PPI networks, as well as serving as a hub of associated accessory data.

Pathway analysis (PA) tools seek to identify groups of related genes/proteins involved in a functional response to a specific environmental condition, allowing for the functional enrichment of large data sets (García-Campos *et al.*, 2015). The premise that groups of genes/proteins are involved in complex biochemical reactions, and not individual macromolecules alone, is what fortifies any PA approach to the functional analysis of -omics data (García-Campos *et al.*, 2015). The Reactome Knowledgebase is an open-source, manual curation of peer-reviewed and evidence-based data of biochemical/biomolecular pathways (Fabregat *et al.*, 2017). It serves as a tool for the discovery of potential pathways, or macromolecules, that might be of biological significance in the context of the research at hand. Utilising the over-representation analysis method, Reactome identifies applicable pathways if the proportion of differentially expressed genes/proteins, in a given pathway, exceeds the number of genes/proteins that could be randomly anticipated (García-Campos *et al.*, 2015; Fabregat *et al.*, 2017). In this way, the relevant pathways and associated molecules provide biological context, compared to other analytical methods that are entirely data-driven (García-Campos *et al.*, 2015). Although the Reactome database focuses on the mapping of a subset of taxonomic data (i.e. *Mus musculus*, *Rattus norvegicus*) to human proteins, it offers detailed information on signalling and biochemical pathways to generate an ordered map of processes (Fabregat *et al.*, 2018). Thus, it provides a basis upon which further investigation and analyses can be conducted, particularly with regard to non-model organism research.

The trajectory of molecular research is such that comprehensive bioinformatics tools are becoming increasingly important in functionally analysing high-throughput data. As such, the -omics fields are not only advancing, but are also becoming more desirable avenues for answering complex biological questions. All this has led to the constant evolution of a range of bioinformatics solutions, each with its own strengths and weaknesses. Nonetheless, all these tools serve as a solid foundation upon which further comprehensive research and analysis can be developed.

3.1.1 Aims of this chapter

Functional analysis of differentially expressed proteins is a crucial step to elucidating the underlying mechanisms and biochemical pathways that are over- or under-represented in an organism in response to changing environmental conditions. Proteomics and related bioinformatics tools have witnessed immense technological strides that have facilitated more holistic analyses of proteomic data. In delving into the mechanistic properties that govern an organism's cellular phenotypic response through functional analysis, scientists are able to identify candidate gene products of particular interest to their respective fields of research. In doing so, novel proteins, particularly in non-model organisms, are discovered that could serve as proxies for detecting environmental change, such as ocean acidification. Functional analysis of proteomics data is a multifaceted feat that entails pathway analyses, characterising protein-protein interactions within a network, as well as homology-based analyses such as Clusters of Orthologous Groups. As such, the qualitative analysis aspect of proteomics research is a holistic approach to functionally characterising groups of differentially expressed proteins.

The aim of this chapter is to functionally analyse the *H. midae* haemocyte proteome in response to reduced pH conditions, following MS/MS-based quantitative analysis of iTRAQ labelled peptides. Furthermore, this chapter seeks to consolidate the quantitative and functional analysis of the iTRAQ data. In doing so, the overall aim is to provide a holistic analysis of the haemocyte stress response in abalone subjected to reduced pH conditions. In turn, this will highlight key biochemical pathways and biological functions necessary for maintaining homeostasis under unfavourable conditions, particularly ocean acidification. To evaluate the effects of reduced pH conditions on the functional stress response of the *H. midae* proteome, several bioinformatics tools were utilised to functionally analyse the haemocyte proteome. Specifically, Blast2GO and QuickGO (Binns *et al.*, 2009) were used to annotate the identified differentially expressed proteins with their corresponding Gene Ontological information, providing a homology-based overview of the biological processes represented by the dataset. Weighted Gene Correlation Network Analysis (WGCNA) was conducted to identify groups of highly correlated proteins based on co-expression, thereby further simplifying the analysis. Following this, STRING database and Cytoscape were utilised for determining protein-protein interactions and generating network maps of WGCNA-defined eigengene modules. Reactome was employed for pathway analysis of each eigengene module to identify statistically significant pathways represented by members of the eigengene modules.

To facilitate the functional analysis process and ensure that a focused approach was executed, the following questions were posed:

i. *What is the overall functional response of abalone haemocytes to reduced pH conditions?*

This will provide an initial glimpse of the overall functional response of abalone haemocytes to reduced pH conditions through Gene Ontology (GO) and Clusters of Orthologous Groups (COGs) analyses. This will in turn provide the foundation from which further analyses will be conducted, allowing us to focus the analysis on specific groups of proteins. The GO and COGs of the significantly differentially expressed proteins (Ch. 2) will clarify the functional responses that are activated in response to the stressor.

ii. *Are there groups of highly correlated proteins that could be of functional importance?*

WGCNA analysis utilises protein expression data to identify groups of proteins that are highly correlated in terms of their expression patterns. This will highlight groups of proteins that are potentially involved in synergistic functions, as well as refine the analysis by highlighting potential candidate biomarkers. The expression profile of each eigengene will be investigated to further group modules either as "Predominantly Upregulated (PU)" or "Predominantly Downregulated (PD)". This will allow us to focus on identifying biomarker candidates from groups that exhibit upregulation in response to decreased pH.

iii. *What are the predominating functional pathways/mechanisms that are induced in abalone haemocytes in response to acidic conditions?*

Further analysis of the PU group of eigengene modules will highlight key pathways of interest that could further elucidate the functions of candidate proteins, as well as aid in generating protein-protein interaction networks, which will ultimately explore the proteomic response of abalone haemocytes to the stressor.

iv. *Based on the statistical analyses conducted in Chapter 2, are there any statistically significant or biologically relevant proteins in the PU eigengene modules, and are they implicated in key pathways and thus, may prove suitable as potential biomarkers?*

This will allow us to refine the “suite” of protein candidates for biomarkers of ocean acidification stress and identify representative members of this stress response for validation using a label-free quantification proteomics approach. Focusing on statistically significant or biologically relevant proteins (i.e. consolidating quantitative and functional analyses) in the context of PU eigengene modules is expected to bolster the reliability of the potential candidate-biomarker group.

3.2 Methods and materials

To facilitate the functional analysis and ensure that a comprehensive approach was utilised, a workflow pipeline was established (Figure 1). This allowed for an "ease-of-flow" analysis that culminated in identifying key proteins of potential importance in the context of this study.

Figure 1. The workflow employed for comprehensive bioinformatics analysis of proteomics data obtained from LC-MS/MS analysis of iTRAQ-labelled peptides from haemolymph sampled from abalone exposed to reduced pH conditions for varying periods of time (12, 72 and 168 hours). Quantitative analysis was carried out to identify significantly differentially expressed proteins (Chapter 2), while qualitative analysis was conducted to functionally analyse protein groups identified during the iTRAQ experiment.

3.2.1 Functional analysis of 227 identified proteins in *H. midae* haemocytes

To facilitate functional analysis, full-length amino acid sequences of the identified proteins were obtained from UniProt (UniProt Consortium, 2018). The file of amino acid sequences was loaded into Blast2GO version 5 (Conesa *et al.*, 2005; Conesa and Götz, 2008; Götz *et al.*, 2008; BioBam Bioinformatics, 2018), which was used in the first step of the qualitative analysis pipe-line (Figure 1). This program serves as an all-in-one tool for the annotation of novel protein sequences, utilising a multi-database inferential approach. This annotation involves three steps: blasting to find homologous sequences, mapping to retrieve Gene Ontology (GO) terms associated with the blast results, and annotation that characterises the query sequences, while also identifying uncharacterised proteins (Conesa and Götz, 2008). Associated pathways and annotations from the KEGG (Kyoto Encyclopedia of Genes and Genomes) database were simultaneously assigned to members of the dataset via the Blast2GO interface, further annotating the proteins.

eggNOG-mapper and the eggNOG database were utilised for Clusters of Orthologous Groups (COG) analysis, which phylogenetically assigns proteins to different pre-determined functions (Tatusov *et al.*, 2000). This prediction of protein function relies on the inference of orthologous relationships between genes from different species, identifying similar domain motifs in query sequences to reference genomes, and assuming similar function. Functional data obtained from the COG analysis on eggNOG-mapper, in the form of a tab-delimited file, was used to generate a graphical representation of the entire dataset.

3.2.2 Weighted gene correlation network analysis (WGCNA)

Weighted Gene Correlation Network Analysis (WGCNA), included in the proteomics analysis package Perseus 1.6.5.0 (Tyanova *et al.*, 2016b; Tyanova and Cox, 2018), was utilised for facilitating co-expression network analysis. WGCNA in Perseus employs an R software package with R functions, in a user-friendly interface, for correlation analyses and network identification and construction (Langfelder and Horvath, 2008; Rudolph and Cox, 2018). A Spearman rank correlation statistical analysis was conducted on the expression data from the four-independent biological iTRAQ experiments. A signed hybrid, soft threshold network type was applied to the data, followed by a standard correlation statistical function. This generated groups of highly correlated proteins, namely eigengene modules. Protein expression profiles of each eigengene module were further assessed in Perseus, using the Profile plot viewer, to

group modules into either “Predominantly Upregulated” or “Predominantly Downregulated” categories. This served to further refine the analysis for biomarker discovery.

Each module and its corresponding proteomic data were exported as a .txt file for further functional and network analysis using Cytoscape and related analysis tools. This served as the foundation for the second step in the qualitative analysis workflow (Figure 1).

3.2.3 Network analysis, protein-protein interactions and GO assignment

The Uniprot database was used to obtain amino acid sequence .fasta files for each eigengene module generated in the WGCNA analysis. The STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database was subsequently utilised to observe protein-protein interactions (PPI) of the members of each eigengene module. STRING is a database of known and predicted proteins and their interactions, determined from a variety of sources such as experimental data, genomic predictions and automated text mining (Szklarczyk *et al.*, 2017).

The amino acid sequence of each of the proteins belonging to each eigengene module was searched against the STRING database. Due to underrepresentation of molluscan sequences in public databases, *Rattus norvegicus* was selected as the reference proteome (based on homologs/orthologs) as an initial search against all available proteomes identified *R. norvegicus* as the most suitable reference proteome to use based on the number of matched *H. midae* proteins. The eigengene module members matched proteins in the *R. norvegicus* database with varying degrees (%) of coverage. Matched proteins with the most query sequence coverage were selected for network construction on STRING. Network and annotation files of each eigengene module were downloaded for further analysis and editing in Cytoscape.

Cytoscape is an integrative, open-source program that generates networks using interaction evidence, as well as expression data and other molecular states (Shannon *et al.*, 2003). Protein-protein interaction networks of the PU eigengene modules were imported and merged in Cytoscape. The same protocol was carried out for the PD eigengene modules. Functional GO analysis of the PU and PD networks was graphically presented using ClueGO (Bindea *et al.*, 2009). ClueGO is a Cytoscape plug-in which visually groups functional terms that facilitate biological understanding of networks. Since protein IDs for the members of the PU and PD groups were reassigned with Ensemble Gene IDs for *R. norvegicus* in STRING, the reference organism for GO analysis on ClueGO was set to *R. norvegicus*. The GO term levels for analysis

were levels 3 to 10, while the minimum number of genes set per term was 3. The Kappa Score Threshold was set to 0.4 and the grouping of terms was 50%, such that 50% of genes were in common between terms.

3.2.4 Identifying potential candidates for biomarker discovery and further analysis

Since Chapter 2 focussed on identifying statistically significant ($p < 0.05$) and biologically relevant (fold-change > 1.2) proteins, it was essential to refine the process of biomarker discovery through reconciliation/consolidation of the quantitative and qualitative data. In Perseus, expression profile plots of the PU eigengene modules were screened for candidate biomarkers based on the aforementioned criteria and the data presented in Tables 2 to 4 of Chapter 2. In so doing, the aim was to select biomarker candidates that were 1. largely upregulated throughout the experimental period, and 2. could be considered outliers with regards to expression compared to the cohort of proteins for each eigengene module. Candidate proteins satisfying the criteria were exported to a separate matrix for further analysis.

STRING was employed (section 3.2.3) to generate a final PPI network of the candidate proteins that could be further edited and analysed in Cytoscape, using Reactome and Biorender. When the amino acid sequences of the candidates were processed through the STRING database in this instance, *Homo sapiens* was found to be the reference proteome with the most coverage of the query protein sequences. Thus, *H. sapiens* was used as the reference proteome for the PPI analysis of the candidate biomarker proteins. GO analysis data of the candidate proteins obtained using QuickGO and ClueGO were consolidated, while the Reactome Pathway browser was utilised for pathway analysis.

3.3 Results

3.3.1 Functional proteomic profiling of *H. midae* exposed to reduced pH conditions

To gain insight into the overall functional response of *H. midae* haemocytes to reduced pH conditions, eggNOG-mapper was utilised for conducting Cluster of Orthologous Groups (COG) analyses on the 227 proteins identified in the four replicate iTRAQ experiments. Sequence data for each protein was utilised in assigning COG categories, which is based on sequence similarity between query and database matches. The 227 identified proteins were assigned to 22 of the 25 COG categories, with "Cytoskeleton (Z)" being the most represented classification (52 sequences; Figure 2). "Translation, Ribosomal Structure and Biogenesis (J)" was the second most represented classification (43 sequences), while "Posttranslational Modification, Protein Turnover and Chaperones (O)" was the third most represented COG category (34 sequences). Due to the limited availability of molluscan sequence data, 14 protein sequences were categorized in the "Function Unknown (S)" class.

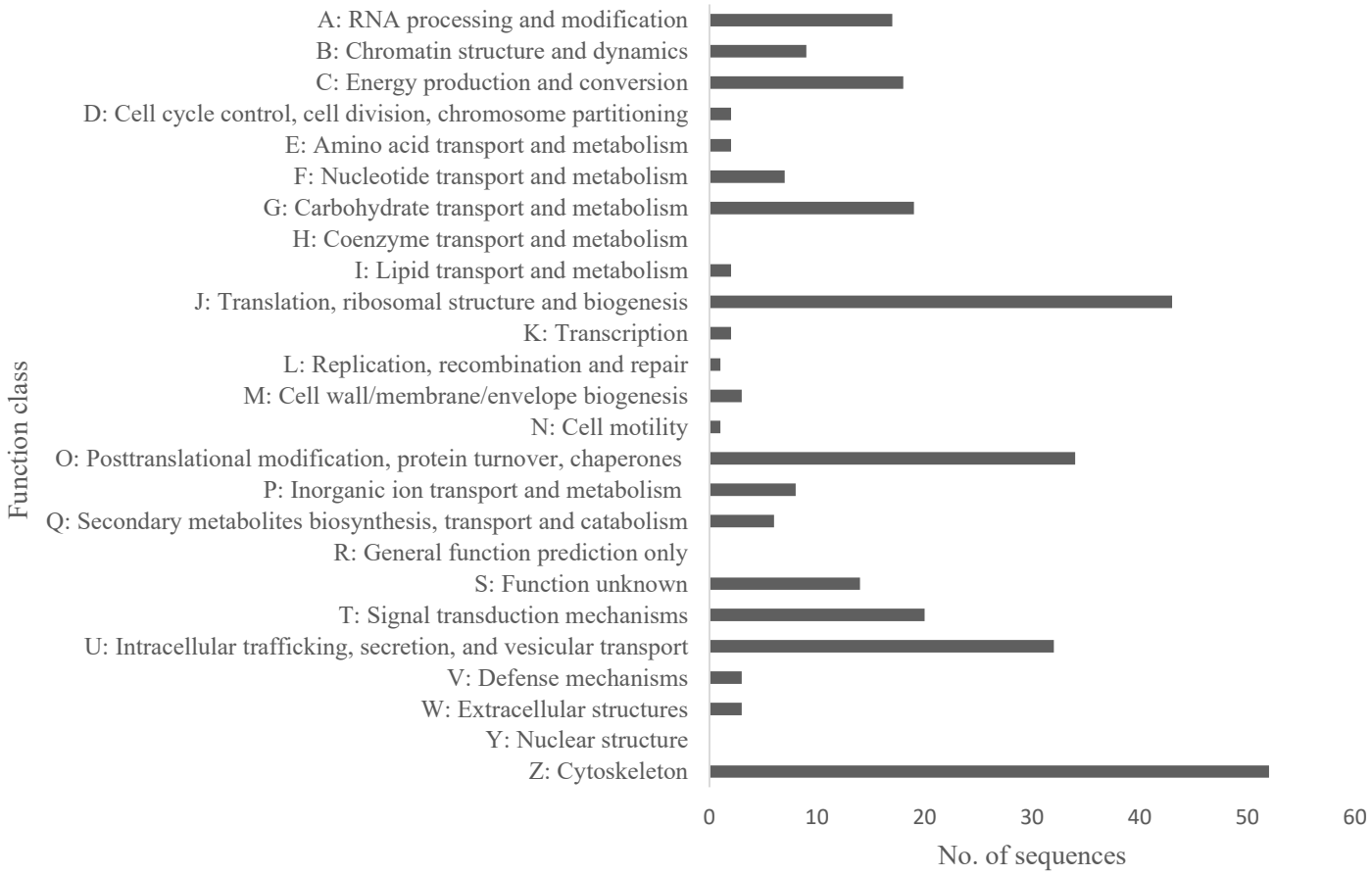

Figure 2. Cluster of orthologous groups (COG) classification of the 227 proteins identified in *H. midae* haemocytes in response to reduced pH conditions.

3.3.2 Weighted Gene Correlation Network Analysis (WGCNA)

To facilitate clustered network analysis, WGCNA was conducted using the expression data of all 227 proteins, to generate eigengene modules consisting of highly correlated proteins in relation to the respective sampling time points. WGCNA analysis facilitated the grouping of potentially co-expressed proteins through correlation statistics.

WGCNA identified eigengene modules of highly correlated proteins that could be investigated further with regard to protein-protein interactions, co-expression, and functional role(s) in specific biochemical pathways. To accomplish the analysis, eigengenes were grouped according to their overall expression profiles (i.e. Predominantly upregulated (PU) or Predominantly downregulated (PD)). The yellow, brown, blue, black and turquoise eigengene modules are PU, while those that are PD are the pink, magenta, red and green modules (Table 1).

Table 1. Eigengene modules generated following WGCNA analysis of the 227 differentially expressed proteins identified from *H. midae* haemocytes following OA stress.

Eigengene module	Number of protein members	Predominant expression profile
Yellow	37	Upregulated
Pink	23	Downregulated
Magenta	20	Downregulated
Brown	38	Upregulated
Black	24	Upregulated
Blue	49	Upregulated
Red	25	Downregulated
Green	25	Downregulated
Turquoise	51	Upregulated

3.3.3 Functional analysis of PU and PD eigengene modules

GO analysis of the PU and PD eigengene modules was conducted using ClueGO to obtain an overview of the functional response of the grouped eigengene modules, and to identify key processes and their associated proteins that could be investigated with regard to their candidacy as biomarkers of ocean acidification stress in *H. midae*.

Figures 3 and 4 are GO term (level 4 - 10) interaction networks of the PU and PD eigengene modules respectively, showing the GO biological processes (BP) and Reactome pathways associated with proteins in each group. Circular nodes (GO terms/Reactome pathways) are connected by edges, where edge thickness represents the degree of confidence that two nodes are associated. Node size represents the level of significance ($p \leq 0.05$) of terms associated with protein members, while colour assignment is arbitrarily grouping related terms. Emboldened terms represent the most significant functional term/pathway of grouped coloured terms, as well as (in most cases) being the "parent" term. Cellular protein localisation (GO: 0034613), translation (GO: 0006412), intracellular transport (GO: 0046907), actin cytoskeleton organisation (GO: 0030036), oxidoreductase activity (GO: 0016620) and response to epidermal growth factor (GO: 0070849) are some of the significant terms associated with proteins of the PU module (Figure 3). This suggests that most proteins in the PU modules are associated with or are involved in these processes or related functions, and as such, are "induced" under OA stress.

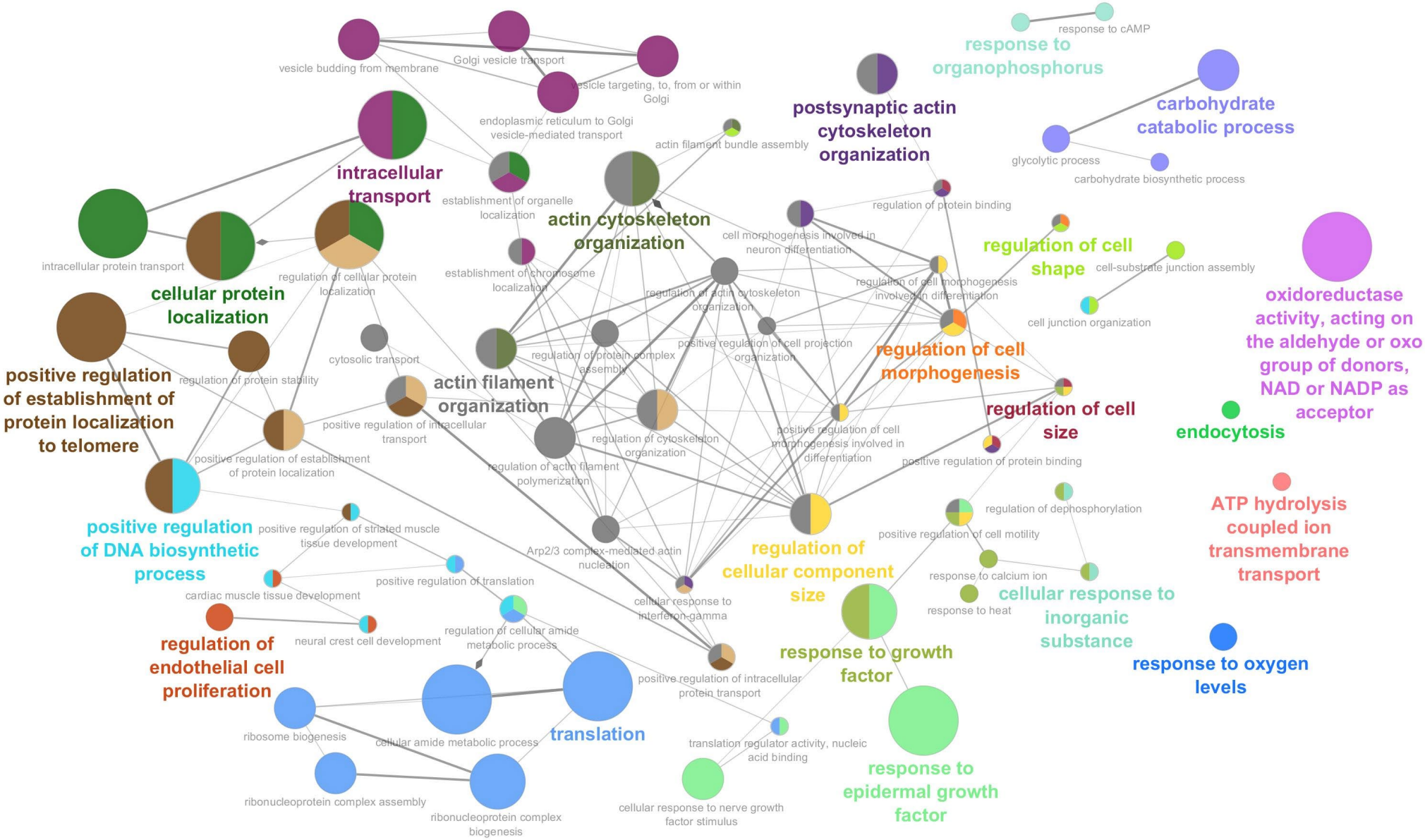
response to cAMP
response to organophosphorus
carbohydrate catabolic process
glycolytic process
carbohydrate biosynthetic process
vesicle budding from membrane
Golgi vesicle transport
vesicle targeting, to, from or within Golgi
endoplasmic reticulum to Golgi vesicle-mediated transport
intracellular transport
establishment of organelle localization
actin filament bundle assembly
postsynaptic actin cytoskeleton organization
regulation of protein binding
actin cytoskeleton organization
cell morphogenesis involved in neuron differentiation
regulation of cell shape
cell-substrate junction assembly
oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor
intracellular protein transport
cellular protein localization
regulation of cellular protein localization
establishment of chromosome localization
regulation of actin cytoskeleton organization
regulation of cell morphogenesis involved in differentiation
cell junction organization
positive regulation of establishment of protein localization to telomere
regulation of protein stability
cytosolic transport
actin filament organization
regulation of protein complex assembly
positive regulation of cell projection organization
regulation of cell morphogenesis
endocytosis
positive regulation of intracellular transport
regulation of cell size
positive regulation of establishment of protein localization
regulation of cytoskeleton organization
positive regulation of cell morphogenesis involved in differentiation
positive regulation of protein binding
regulation of actin filament polymerization
regulation of dephosphorylation
positive regulation of cell motility
ATP hydrolysis coupled ion transmembrane transport
positive regulation of DNA biosynthetic process
positive regulation of striated muscle tissue development
Arp2/3 complex-mediated actin nucleation
regulation of cellular component size
response to calcium ion
cellular response to inorganic substance
positive regulation of translation
cardiac muscle tissue development
cellular response to interferon-gamma
response to heat
regulation of endothelial cell proliferation
neural crest cell development
regulation of cellular amide metabolic process
response to growth factor
response to oxygen levels
positive regulation of intracellular protein transport
ribosome biogenesis
cellular amide metabolic process
translation
translation regulator activity, nucleic acid binding
response to epidermal growth factor
ribonucleoprotein complex assembly
ribonucleoprotein complex biogenesis
cellular response to nerve growth factor stimulus

Figure 3. ClueGO Gene Ontology analysis of the PU eigengene modules, showing the biological processes in which they function. Node size reflects term significance (term p-value corrected with Bonferroni step down), while edges are associated with the Kappa Score, which defines the degree of term-term interrelatedness. Terms in bold colour represent the most enriched/significant term, while groups of colours signify related parent and child GO terms.

Figure 4 shows the PD eigengene module and the assigned GO and Reactome terms associated with its members. Considering that this is a predominantly downregulated group of modules, it can be said that processes represented in Figure 3 are "uninduced". Carbohydrate catabolic process (GO: 0016052), positive regulation of translation (GO: 0045727), regulation of apoptotic processes (GO: 0010665; GO: 0010662), plasma membrane repair (GO: 0001778) and clathrin adaptor activity (GO: 0035615) are some of the most significant terms associated with protein members of the PD modules.

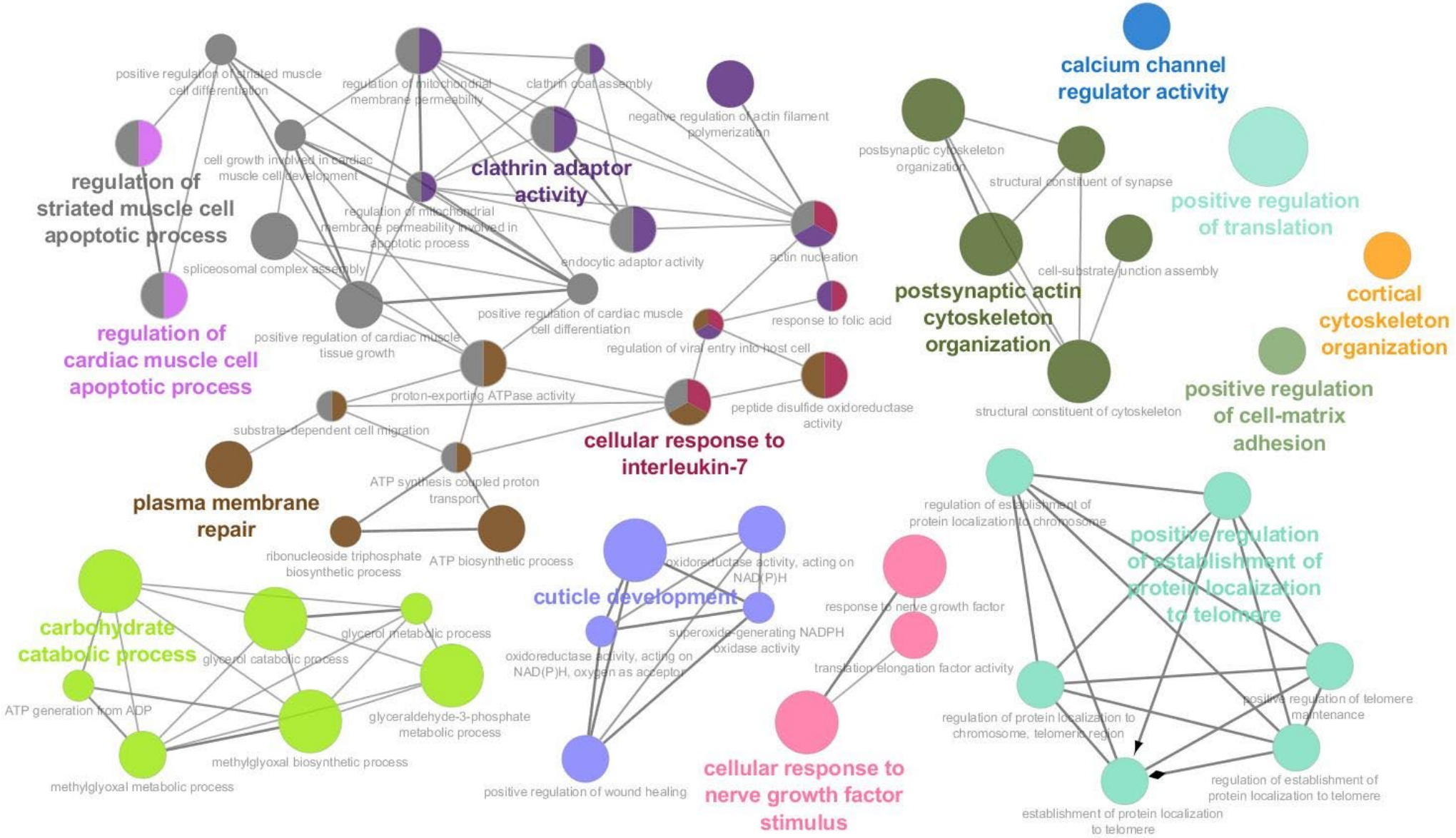
positive regulation of striated muscle cell differentiation
regulation of mitochondrial membrane permeability
clathrin coat assembly
negative regulation of actin filament polymerization
cell growth involved in cardiac muscle cell development
clathrin adaptor activity
regulation of striated muscle cell apoptotic process
regulation of mitochondrial membrane permeability involved in apoptotic process
spliceosomal complex assembly
endocytic adaptor activity
actin nucleation
positive regulation of cardiac muscle cell differentiation
response to folic acid
regulation of cardiac muscle cell apoptotic process
positive regulation of cardiac muscle tissue growth
regulation of viral entry into host cell
proton-exporting ATPase activity
peptide disulfide oxidoreductase activity
substrate-dependent cell migration
cellular response to interleukin-7
ATP synthesis coupled proton transport
plasma membrane repair
ribonucleoside triphosphate biosynthetic process
ATP biosynthetic process
calcium channel regulator activity
postsynaptic cytoskeleton organization
structural constituent of synapse
positive regulation of translation
postsynaptic actin cytoskeleton organization
cell-substrate junction assembly
cortical cytoskeleton organization
structural constituent of cytoskeleton
positive regulation of cell-matrix adhesion
carbohydrate catabolic process
glycerol metabolic process
glycerol catabolic process
ATP generation from ADP
glyceraldehyde-3-phosphate metabolic process
methylglyoxal biosynthetic process
methylglyoxal metabolic process
cuticle development
oxidoreductase activity, acting on NAD(P)H
superoxide-generating NADPH oxidase activity
oxidoreductase activity, acting on NAD(P)H, oxygen as acceptor
positive regulation of wound healing
response to nerve growth factor
translation elongation factor activity
cellular response to nerve growth factor stimulus
regulation of establishment of protein localization to chromosome
positive regulation of establishment of protein localization to telomere
positive regulation of telomere maintenance
regulation of protein localization to chromosome, telomeric region
regulation of establishment of protein localization to telomere
establishment of protein localization to telomere

Figure 4. ClueGO Gene Ontology analysis of the PD eigengene modules, showing the biological processes in which they function. Node size reflects term significance (term p-value corrected with Bonferroni step down), while edges are associated with the Kappa Score, which defines the degree of term-term interrelatedness. Terms in bold colour represent the most enriched/significant term, while groups of colours signify related parent and child GO terms.

3.3.4 Eigengene modules of interest: Toward candidate biomarker discovery

To incorporate the previously identified statistically significant ($p < 0.05$) and/or biologically relevant (fold-change > 1.2) proteins identified in Chapter 2 (Tables 2-4), comparative analyses were conducted between the quantitative and qualitative data.

Following screening the PU eigengene modules for statistically significant and/or biologically relevant proteins based on the criteria used in Chapter 2, a total of 33 potential candidate biomarkers of OA stress were identified for further analysis. The proteins are involved in signalling, cytoskeletal functions, translation, intracellular protein transport and metabolism, and are mostly grouped in the brown, black and turquoise eigengene modules. The fold-change of the proteins ranged from 1.2 (i.e. 78 kDa glucose-regulated protein, tubulin, and dual oxidase) to 1.6 (i.e. charged multivesicular body protein 1a, vacuolar protein sorting-associated protein 13A, and calmodulin-dependent protein kinase) (Table 2). Statistically significant proteins ($p < 0.05$) included 78 kDa glucose-regulated protein, enolase, calmodulin-dependent protein kinase, and mitogen-activated protein kinase.

Table 2. Statistically significant ($p < 0.05$) and/or biologically relevant (fold-change > 1.2) proteins identified in 5 of the 9 WGCNA eigengene modules (indicated by the coloured bars).

Uniprot Accession	Protein name	*P*-value*	Fold-change*
K1QEZ3_CRAGI	Fascin	0.30	1.3
A0A1L7H9J9_HALDI	Mitogen-activated protein kinase	0.05	1.4
K1QNQ9_CRAGI	Hypothetical protein CGI_10024082	0.12	1.3
A0A2T7PHL9_POMCA	60S ribosomal protein L23a	1.00	1.3
A0A2T7P5E9_POMCA	60S ribosomal protein L24-like	1.00	1.3

K1PCW5_CRAGI	Alpha-1 3-mannosyltransferase ALG2	1.00	1.2	
B3TK60_HALDV	Calmodulin-dependent protein kinase (Fragment)	0.02	1.5	
K1QX37_CRAGI	Enolase	0.04	1.3	
K1R5M2_CRAGI	Myosin-IXa	0.28	1.2	
O61472_APLCA	Reductase-related protein	0.02	1.2	
G8XRU5_9CAEN	Histone H3 (Fragment)	0.35	1.2	
A0A2T7PIY3_POMCA	60S ribosomal protein L27a	1.00	1.3	
A0A0B7BF17_9EUPU	Radixin isoform X1	0.33	1.2	
V4A8R6_LOTGI	Putative uncharacterized oxidoreductase YDR541C	0.01	1.3	
H9AWU2_HALDH	Chaperonin containing T-complex polypeptide subunit zeta	0.49	1.3	
V4ADA9_LOTGI	Phosphoglycerate kinase	0.05	1.4	
A0A0G2SSZ6_HALDI	Serpin B-like protein 2	0.27	1.3	
A0A0G2R2A9_9BIVA	Tubulin beta chain (Fragment)	0.17	1.2	
A0A210PT48_MIZYE	Vacuolar protein sorting-associated protein 13A	0.14	1.6	
A0A210QG69_MIZYE	Ras GTPase-activating-like protein IQGAP1	0.08	1.4	
A0A0B6ZMQ4_9EUPU	Solute carrier family 2, facilitated glucose transporter member 1	0.29	1.3	
A0A0L8FL54_OCTBM	Charged multivesicular body protein 1a-like	0.06	1.6	
A0A2T7PEM0_POMCA	Hemocyanin G-type, units Oda to Odg-like	0.55	1.4	
K1RM18_CRAGI	Dual oxidase	0.42	1.2	

K1QG65_CRAGI	rRNA 2'-O-methyltransferase fibrillarin	0.52	1.4	
A0A0L8H0E1_OCTBM	Sorting nexin	0.10	1.4	
A0A2T7NG09_POMCA	Ras GTPase-activating-like protein IQGAP1 isoform X1	0.14	1.3	
K1RH70_CRAGI	6-phosphogluconate dehydrogenase decarboxylating	0.03	1.2	
A0A2C9JK44_BIOGL	Obg-like ATPase 1	0.05	1.2	
A0A0B6Y9I8_9EUPU	Probable small nuclear ribonucleoprotein Sm D2	0.07	1.4	
A0A2C9JRW8_BIOGL	Uncharacterized oxidoreductase YajO-like	0.39	1.4	
V3ZW18_LOTGI	Probable methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	0.02	1.3	
V4BUJ1_LOTGI	78 kDa glucose-regulated protein	0.05	1.2	
A0A2T7PHH2_POMCA	Major vault protein-like	0.26	1.3	

*Note: p-value and fold-change of candidate proteins are not averaged across time points, but rather reflect significance and biological relevance at one of the sampling points over the course of the experiment.

To investigate the ontologies associated with the 33 candidate proteins, ClueGO was utilised for GO analysis in Cytoscape, thus generating a network of associated/interacting functional processes (Figure 5). Mapped proteins were found to be associated with more than one biological process and/or pathway (Figure 5). Sorting nexin (SNX6) was found to be associated with cytosolic transport and response to epidermal growth factor, while ras GTPase-activating-like protein IQGAP1 isoform X1 (IQGAP1) is involved in the Signalling by Ras mutants pathway and is strongly associated with response to epidermal growth factor. Further biological processes that IQGAP1 is associated with includes regulation of focal adhesion assembly and regulation of adherens junction organisation. Mitogen-activated protein kinase (MAPK1) is

another protein associated with several pathways and biological functions such as cytosolic transport, signalling, response to epidermal growth factor, cellular response to heat stress and oxidative stress induced by senescence.

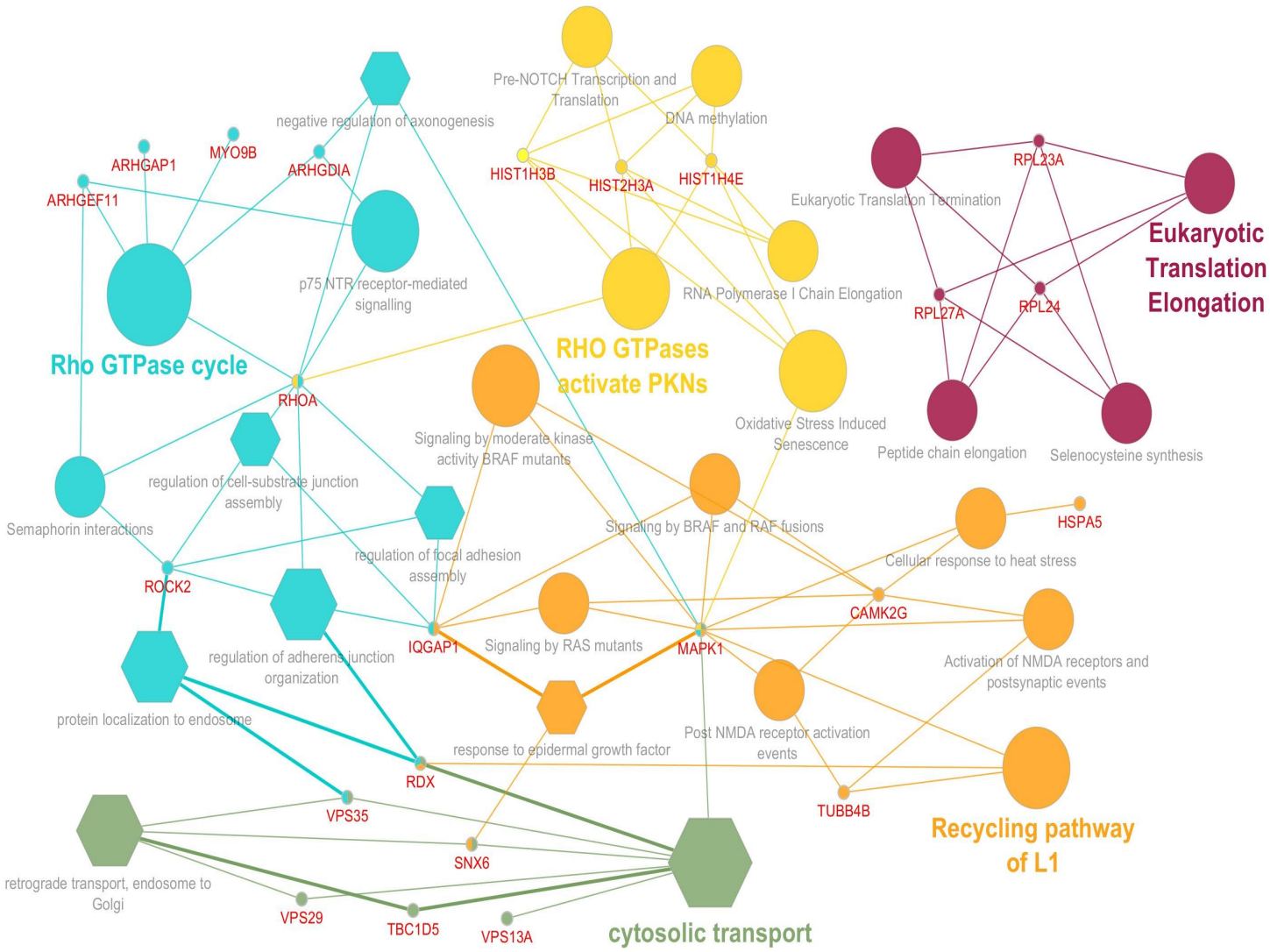

Figure 5. ClueGO network of the statistically significant (p-value corrected with Bonferroni step-down; $p < 0.05$) Gene Ontology biological processes (hexagonal nodes) and Reactome pathways (circular nodes). Nodes of protein members associated with the significant terms are labelled in red. Node size indicates the degree of significance of a term, while edge thickness represents the degree of confidence that two nodes are connected and related.

Since a significant amount of functional information is associated with members of this candidate biomarker group, I felt it necessary to summarise this information (Table 3). The 33 identified candidate biomarker proteins were grouped according to their overall functions. The assignment of colour is entirely arbitrary and serves the purpose of mere differentiation between the generalised functional groups of proteins. The orange group is comprised of proteins involved in metabolic and energy use functions; proteins in the blue group are involved in stress/immune-related functions; the yellow group includes proteins involved in cytoskeletal and intracellular protein sorting functions; proteins in the green group are involved in oxidation-reduction related processes; the purple group includes proteins involved in transcription and translation related functions.

Table 3. Functional analysis of the 33 candidate proteins identified as potential biomarkers of ocean acidification-induced stress in *Haliotis midae*, grouped according to general (as indicated by the colours), overarching functions.

Uniprot Accession	Protein name (STRING protein ID)	Associated function	Predicted pathway (accession)	GO biological process
A0A0B6ZMQ4_9EUPU	Solute carrier family 2, facilitated glucose transporter member 1 (SLC2A4/GLUT1)	Glucose transmembrane transport	Integration of energy metabolism (R-HSA-163685.1)	Cellular response to glucose starvation (GO:0042149); Protein-containing complex assembly (GO:0065003)
K1QX37_CRAGI	Enolase (ENO1)	Glucose metabolism	Gluconeogenesis (R-HAS-70263.2), Glycolysis (R-HSA-70171.6)	Positive regulation of muscle contraction (GO:0045933); Glycolytic process (GO:0006096); Response to virus (GO:0009615); Gluconeogenesis (GO:0006094)
V4ADA9_LOTGI	Phosphoglycerate kinase (PGK1)	Glucose metabolism	Glycolysis (R-HSA-70171.6)	Cellular response to hypoxia (GO:0071456); Glycolytic process (GO:0006096); Plasminogen activation (GO:0031639); Gluconeogenesis (GO:0006094)
V3ZW18_LOTGI	Probable methylmalonate semialdehyde dehydrogenase (ALDH6A1)	Oxidoreductase	Metabolism of amino acids and derivatives (R-HSA-71291.6)	Oxidation-reduction process (GO: 0055114); Methylmalonate-semialdehyde dehydrogenase (acylating) activity (GO: 0004491)

K1RH70_CRAGI	6-phosphogluconate dehydrogenase decarboxylating (XP_005109585.1)	Oxidative carboxylase involved in the pentose phosphate pathway	Pentose phosphate pathway (R-HSA-71336.6)	Pentose-phosphate shunt (GO: 0006098); Oxidation-reduction process (GO: 0055114)
K1PCW5_CRAGI	Alpha-1 3-mannosyltransferase ALG2 (ALG2)	Associated with post-translational protein modification, specifically Asparagine N-linked glycosylation.	Metabolism of proteins (R-HSA-392499.7)	Protein glycosylation (GO: 0006486); Transferase activity (GO: 0016740)
K1RM18_CRAGI	Dual oxidase (DUOX1)	Part of family of NADPH oxidases that generate ROS	Metabolism of amino acids and derivatives (R-HSA-71291.5); MAPK signalling pathway (KO04013); Toll and IMD signalling pathway (KO04624)	Oxidation-reduction process (GO:0055114); Response to cAMP (GO:0051591); Cytokine-mediated signalling pathway (GO:0019221); Response to oxidative stress (GO:0006979)
A0A210QG69_MIZYE	Ras GTPase-activating-like protein IQGAP1 isoform X1 (IQGAP1)	Scaffold protein regulating various functions of the immune system such as signal transduction, cell adhesion and cytoskeletal organisation.	MAPK family signalling cascades (R-HSA-5683057.3); Neutrophil degranulation (R-HSA-6804803)	Signal transduction (GO:0007165); Neutrophil degranulation (GO:0043312)
A0A0G2SSZ6_HALDI	Serpin B-like protein 2 (SERPINB6B)	Protease inhibitor, implicated in various functions of the innate immune system	Complement and coagulation cascade (KO04610)	Negative regulation of apoptotic process (GO:0043066); Wound healing (GO:0042060); Interleukin-12-mediated signalling pathway (GO:0035722)

K1R5M2_CRAGI	Myosin-IXa (MYO9B)	Cell migration and cell-cell interactions; pseudopodia formation during phagocytosis	Signalling by Rho GTPases (R-HSA-194315.4); Fcgamma receptor (FCGR) dependent phagocytosis (R-HSA-2029480.1)	Intracellular signal transduction (GO:0035556); Cell junction assembly (GO:0034329)
K1QEZ3_CRAGI	Fascin (FSCN2)	Cell motility and phagocytosis	Interleukin-4 and 13 signalling (R-HSA-6789835)	Positive regulation of lamellipodium assembly (GO:0010592); Cell-cell junction assembly (GO:0007043); Regulation of actin cytoskeleton organisation (GO:0032956)
B3TK60_HALDV	(Calcium/) calmodulin-dependent protein kinase (CAMK)	Regulation of several signalling cascades; mediates learning and memory	MAPK1/MAPK3 signalling (R-HSA-5684996.4); Signalling by WNT (R-HSA-195721.5)	Intracellular signal transduction (GO:0035556); Activation of protein kinase activity (GO:0032147); Protein phosphorylation (GO:0006468)
A0A1L7H9J9_HALDI	Mitogen-activated protein kinase (MAPK1)	Links extracellular stimuli to intracellular functions for the regulation of cell proliferation, inflammation and apoptosis.	Advanced glycosylation end-product receptor signalling of innate immune system (R-HSA-879415.1); MAPK1/MAPK3 signalling (R-HSA-5684996.4); Cytokine signalling in immune system (R-HSA-1280215.3); Signalling by Receptor Tyrosine Kinases (R-HSA-9006934.4)	Cellular response to ROS (GO:0034614); Stress-activated MAPK cascade (GO:0051403); Signal transduction (GO:0007165); Apoptotic process (GO:0006915); Regulation of cytoskeleton organisation (GO:0051493); Regulation of cellular response to heat (GO:100034); Neutrophil degranulation (GO:0043312)

G8XRU5_9CAEN	Histone H3 (HIST1H3F)	Chromatin structure; potential antimicrobial activity	Oxidative stress induced senescence (R-HSA-2559580.3); Signalling by NOTCH (R-HSA-157118.4); Signalling by Nuclear Receptors (R-HSA-9006931.5); Chromatin organisation (R-HSA-4839726.2); Signalling by WNT (R-HSA-195721.5); DNA methylation (R-HSA-5334118.1)	Nucleosome assembly (GO:0006334); Regulation of gene silencing (GO:0060968); Blood coagulation (GO:0006325); Interleukin-7-mediated signalling pathway (GO:0038111);
A0A2T7PHH2_POMCA	Major vault protein (MVP)	Cellular transport, signalling and immune defence.	Neutrophil degranulation (R-HSA-6803325)	ERBB signalling pathway (GO:0038127); Negative regulation of signalling (GO:0023057); mRNA transport (GO:0051028);
V4BUJ1_LOTGI	78 kDa glucose-regulated protein (HSPA5)	Stress response; unfolded protein response	Cellular responses to stress (R-HSA-2262752.5); Unfolded protein response (UPR) (R-HSA-381119.2)	Chaperone cofactor-dependent protein refolding (GO:0051085); Posttranslational protein targeting to membrane (GO:0031204); Cellular response to unfolded protein (GO:0034620); Negative regulation of apoptotic process (GO:0043066); Cellular response to heat (GO:0034605)

A0A2T7PEM0_POMCA	Haemocyanin G-type (Hc)	Oxygen carrier in haemolymph; Potential antimicrobial activity		Oxygen transport (GO:0015671); Oxidation-reduction process (GO:0055114)
H9AWU2_HALDH	Chaperonin containing T-complex polypeptide subunit zeta (CCT6A/TCP-1)	Member of the TCP1 ring complex. Cytoskeletal organisation, intracellular assembly and cell cycle	Protein folding (R-HSA-391251.1)	Protein folding (GO:0006457)
A0A0B7BF17_9EUPU	Radixin isoform X1 (RDX)	Structural and regulatory functions as membrane-cytoskeletal crosslinkers	Regulation of actin cytoskeleton (KO04810); Tight junction (KO04530)	Establishment of protein localisation (GO:0045184); Cellular response to platelet-derived growth factor stimulus (GO:0036120); Positive regulation of early endosome to late endosome transport (GO:2000643)
A0A0L8H0E1_OCTBM	Sorting nexin (SNX6)	Endocytosis and endosomal signalling.	Signal transduction (R-HSA-162582)	Intracellular protein transport (GO:0006886); Regulation of macroautophagy (GO:0016241)
A0A0G2R2A9_9BIVA	Tubulin beta chain (Fragment) (TUBB4B)	Intracellular transport	Phagosome (PTR04145); Gap Junction (PTR04540)	Cytoskeleton-dependent intracellular transport (GO:0030705); Microtubule-based process (GO:0007017); Regulation of synapse organisation (GO:0050807)
A0A210PT48_MIZYE	Vacuolar protein sorting-associated protein 13A (VPS13C)	Intracellular sorting and transport of proteins	Membrane trafficking (R-HSA-199991.6)	Protein localisation (GO:0008104); Autophagy (GO:0006914); Golgi to

				endosome transport (GO:0006895)
A0A0L8FL54_OCTBM	Charged multivesicular body protein 1a-like (CHMP1a)	Multivesicular sorting of proteins in the interiors of lysosomes	Cellular response to stress (R-HSA-2262752.5); Signalling by MET (R-HSA-6806834.2)	Vesicle-mediated transport (GO:0007034); ESCRT-III complex disassembly (GO:1904903); Nucleus organisation (GO:0006997)
A0A0B6Y9I8_9EUPU	Probable small nuclear ribonucleoprotein Sm D2 (SNRPD2)	Involved in the formation of spliceosomes	Metabolism of non-coding RNA (R-HSA-194441.1); Processing of capped intron-containing pre-mRNA (R-HSA-72203.3)	mRNA processing (GO:0006397); Spliceosomal complex assembly (GO:0000245)
A0A2C9JK44_BIOGL	Obg-like ATPase 1 (OLA1)	ATP/GTP hydrolysation	Response to elevated platelet cytosolic Ca^{2+}(R-HSA-76005.2)	Platelet degranulation (GO:002576); ATP metabolic process (GO:0046034)
K1QG65_CRAGI	rRNA 2'-O-methyltransferase fibrillarin (FBL)	Component of snRNP particle involved in pre-ribosomal (r)RNA processing	rRNA modification in the nucleus and cytosol (R-HSA-6790901.4)	rRNA processing (GO:0006364); Histone glutamine methylation (GO:1990258)
A0A2T7PHL9_POMCA	60S ribosomal protein L23A (RPL23A)	Involved in translation and rRNA processing	Major pathway of rRNA processing in the nucleolus and cytosol (R-HSA-6791226.3); Eukaryotic translation initiation (R-HSA-72613.3)	Translation (GO:0006412); Cell population proliferation (GO:0008283)
A0A2T7P5E9_POMCA	60S ribosomal protein L24-like (ENSRNOG00000001611)	Involved in translation and rRNA processing	Major pathway of rRNA processing in the nucleolus and cytosol (R-HSA-6791226.3); Eukaryotic	Translation (GO:0006412); Cell population proliferation (GO:0008283)

			translation initiation (R-HSA-72613.3)	
A0A2T7PIY3_POMCA	60S ribosomal L27A (RPL27A)	Involved in translation and rRNA processing	Major pathway of rRNA processing in the nucleolus and cytosol (R-HSA-6791226.3); Eukaryotic translation initiation (R-HSA-72613.3)	Translation (GO:0006412); Cell population proliferation (GO:0008283)

To temporally and spatially visualise the 33 candidate proteins, Cytoscape and Biorender were utilised to generate a PPI network (Figure 6). A number of signalling proteins, such as mitogen-activated protein kinase, chaperonin containing T-complex polypeptide subunit zeta, calmodulin-dependent protein kinase and ras GTPase-activating-like protein IQGAP1, exhibited a fold-change of ≥1.2. Most of these proteins are localised to the cytoplasm and are predicted to interact with one another and cytoskeletal proteins (such as fascin and radixin), suggesting that they might be involved in more than one function. Other proteins, such as vacuolar protein sorting-associated protein 13A, sorting nexin and charged multivesicular body protein 1a-like, were found to be localised to the endosome, being implicated in protein sorting and trafficking. Although not directly interacting with one another, these endosomal proteins all belong to a group of proteins that interact with one another, functioning as either downstream or upstream effectors in vesicular trafficking and endosomal sorting.

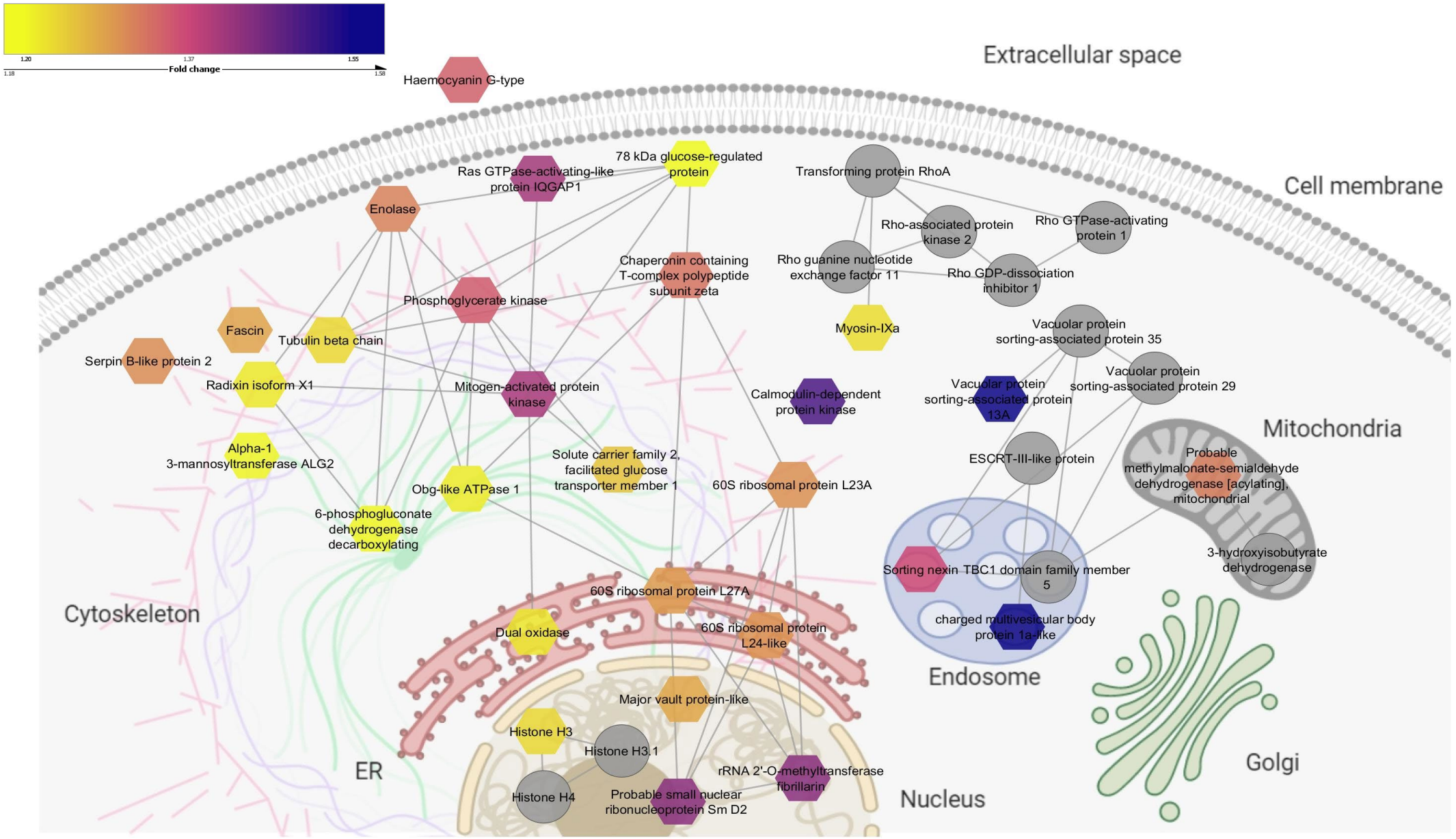
1.20
1.37
1.55
Fold change
1.18
1.58
Extracellular space
Haemocyanin G-type
Cell membrane
78 kDa glucose-regulated protein
Ras GTPase-activating-like protein IQGAP1
Transforming protein RhoA
Enolase
Rho-associated protein kinase 2
Rho GTPase-activating protein 1
Chaperonin containing T-complex polypeptide subunit zeta
Rho guanine nucleotide exchange factor 11
Rho GDP-dissociation inhibitor 1
Phosphoglycerate kinase
Fascin
Tubulin beta chain
Myosin-IXa
Vacuolar protein sorting-associated protein 35
Serpin B-like protein 2
Radixin isoform X1
Mitogen-activated protein kinase
Calmodulin-dependent protein kinase
Vacuolar protein sorting-associated protein 13A
Vacuolar protein sorting-associated protein 29
Mitochondria
Alpha-1 3-mannosyltransferase ALG2
Solute carrier family 2, facilitated glucose transporter member 1
ESCRT-III-like protein
Probable methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial
Obg-like ATPase 1
60S ribosomal protein L23A
6-phosphogluconate dehydrogenase decarboxylating
Sorting nexin
TBC1 domain family member 5
3-hydroxyisobutyrate dehydrogenase
60S ribosomal protein L27A
Cytoskeleton
Dual oxidase
60S ribosomal protein L24-like
charged multivesicular body protein 1a-like
Endosome
Major vault protein-like
Histone H3
Histone H3.1
rRNA 2'-O-methyltransferase fibrillarin
ER
Histone H4
Probable small nuclear ribonucleoprotein Sm D2
Nucleus
Golgi

Figure 6. Protein-protein interaction network of the statistically significant ($p \leq 0.05$) and/or biologically relevant (fold-change ≥ 1.2) proteins assigned to their putative GO cellular components. Grey nodes represent proteins associated with statistically significant and/or biologically relevant proteins which were not included in the previously identified group of proteins. These are predicted interactors obtained from the STRING database. The colours of the hexagonal nodes represent the relative fold change of the proteins of interest (fold-change ≥ 1.2) as indicated by the node legend. Diagram created with Biorender.com.

3.4 DISCUSSION

Qualitative analysis of proteomics data underscores a crucial step in a bottom-up, shotgun proteomics study. Fortunately, an array of analytical tools exist that facilitate functional analysis of large sets of -omics data, which facilitate investigations of the effects of various environmental conditions on model organisms. The conundrum arises when conducting discovery-based studies of non-model organisms. Since genetic data for non-model organisms can be scarce, unvalidated or insufficiently annotated in public databases, many studies require an inferential and homology-based analytical approach, which does come with its own limitations. This chapter sought to functionally characterise the stress response of *Haliotis midae* to reduced pH conditions, representative of future ocean acidification predictions, following iTRAQ and LC-MS/MS analysis. This was achieved through comparative amino acid sequence searches, assigning suitable reference proteomes and subsequently inferring protein network interactions and functions, using the highest degree of homology between the query proteins and reference proteomes. In so doing, key processes and associated proteins could be identified and investigated for their potential as biomarkers of ocean acidification stress.

3.4.1 Overall functional response to ocean acidification conditions

One aspect of functional and comparative -omics studies is the classification of proteins across taxa that are biochemically similar with regard to their molecular functions within a cell. Such a classification, although purposefully inferential, can be achieved through sequence homology to well-characterised proteins (Kaufmann, 2006). Satisfying this need led to the development of clusters of orthologous groups (COGs) by Tatusov *et al.* (2000). With the aid of COG and Gene Ontology (GO), Duan *et al.* (2013) were able to functionally classify the 593 unique sequences they identified in their immune gene discovery study of the ridgetail white prawn. In the current study, the 227 proteins identified in *H. midae* haemocytes that respond to reduced pH stress were functionally classified using COG imputed on eggNOG-mapper (Figure 2). Over 50 of the 227 sequences were classified as being associated with the “cytoskeleton (Z)”, while more than 40 sequences were ascribed to “translation, ribosomal structure and biogenesis (J)” functions” (Figure 2). Since cell viability is affected by changes in volume, maintenance of cell volume is crucial, particularly under conditions that expose cells to varying osmotic

environments (Di Ciano-Oliveira *et al.*, 2006). The sheer number of proteins found to be associated with cytoskeletal functions in this study might be attributed to the fundamental principle of cytoskeletal re-organisation and stabilisation under cellular stress, which seeks to regain cellular structural integrity. Epelboin *et al.* (2016) investigated expression of the *C. gigas* kinome in response to environmental conditions affecting an estuary, focusing on salinity changes. Under salinity constraints, a number of differentially expressed protein kinase genes involved in metabolism, cytoskeletal organisation and the immune system were identified, suggesting cytoskeletal re-organisation under hyposaline conditions (Epelboin *et al.*, 2016). Thus, under reduced pH conditions, decreases in the carbonate ion concentration of seawater might have adverse osmotic effects on *H. midae* haemocytes by affecting the structural integrity of cells.

Very few studies investigating the effects of environmental stress have employed COG analysis, particularly studies involving molluscs. This might be due to the limited availability of molluscan sequences in public databases, resulting in poor annotation efficiency (Zhang *et al.*, 2019). However, in a recent transcriptomic study looking at the effects of multi-stressors (hypoxia and thermal challenge) on the immune response of *H. diversicolor*, COG analysis revealed "General function prediction only (R)" and "Signal transduction mechanisms (T)" were the most represented classes (Zhang *et al.*, 2019). This is in contrast to the current study. This might be due to the use of a different database in this study which also focused primarily on the proteome and not the transcriptome. Of the 227 *H. midae* sequences analysed, 14 were classified with "Function unknown (S), which could be attributed to working with a non-model marine invertebrate, and as such, the appropriate annotation of proteins is hampered due to the limited molluscan sequence data available on public databases (Calvete, 2014). However, considering the limited genetic data available, it is impressive that only 6% of the protein sequences remained uncharacterised.

Due to the size and expression variability of the dataset, Weighted Gene Co-expression Network Analysis (WGCNA) was employed to facilitate the identification of groups of highly correlated proteins based on expression, which simplified functional analysis of the dataset. WGCNA can be employed for exploratory analysis of datasets, as well as for identifying modules for pathway analyses (Langfelder and Horvath, 2008). In this study, nine eigengene modules of co-expressed proteins were identified, which were further grouped into those that were Predominantly Upregulated (PU) and Predominantly Downregulated (PD) over the time course of the experiment (Table 1). This step further refined the dataset to focus primarily on

upregulated proteins for biomarker discovery, especially since reliable biomarkers should be sufficiently altered to discriminate between normal and stressed/treatment conditions (Moore *et al.*, 2017). A total of 5 of the 9 eigengenes were classified as PU, while the remaining 4 eigengene modules were PD.

Functional analysis of the PU and PD groups using ClueGO provided further insight into the predominating biological processes represented by the downregulated and upregulated proteins identified in the four-independent biological iTRAQ trials (Figures 3 and 4). While COG is concerned with gene homology based on speciation events from a common ancestor, Gene Ontology (GO) involves the functional annotation of genes and their products through protein-term matching, as well as providing an integrated source of knowledge and related tools. The GO database is considered the most comprehensive source of functional information applicable to computational analysis of genes and their products (GO Constortium, 2019). The GO database follows an ontological flow, where categories of gene functions (GO terms) have definitive relations with one another. This allows the user to apply inferential analyses using logical bases, while also evolving with the growing age of genetic studies (GO Constortium, 2019). For the purposes of this study, we will only focus on discussing the PU eigengene modules.

In this study, the PU eigengene modules were found to be associated with a number of GO biological processes, most notably "cellular protein localisation" (and related GO terms), "response to epidermal growth factor", "oxidoreductase activity", "translation" and "actin cytoskeleton organisation" (Figure 3). These processes could be considered activated since this group consists of proteins that are predominantly upregulated throughout the experimental period. As a reminder, these functional analyses utilized the *H. sapiens* and *R. norvegicus* gene product databases, and as such, functional annotation is based on inference through homology. Predictions in function provide an initial understanding of potential molecular functions of proteins, while also serving as the basis upon which further experimental studies can be designed to elucidate the true function of query candidates (Sjölander, 2004). Less significant GO terms associated with the PU eigengene modules include "carbohydrate catabolic process", "Golgi vesicle transport" (and related terms), "positive regulation of DNA biosynthetic process", "regulation of cellular component size", and "intracellular transport" (Figure 3). This might suggest that a possible shift from energetically costly aerobic metabolism to more efficient anaerobic metabolism is occurring to enable protein synthesis, intracellular transport, and related functions to proceed in *H. midae* haemocytes under the stressor. This could be

attributed to increased protein turnover following misfolded or damaged proteins under increased oxidative stress. To a lesser degree, GO terms such as endocytosis (GO: 0006897), positive regulation of cell motility (GO: 2000147) (and related terms such as response to heat and calcium ion), regulation of cell morphogenesis (GO: 0022604) and regulation of protein stability (GO: 0031647), were underrepresented in the PU dataset (Figure 3). However, in the context of this study, proteins associated with these terms are worth investigating further due to their potential as biomarkers of OA stress.

In a transcriptomic study investigating the effects of near-future pH conditions on the pteropod *Heliconoides inflatus*, ion transport (GO:0006811) was the only GO term that was significantly enriched, and was associated with genes that were found to be upregulated (Moya *et al.*, 2016). A microarray study testing the effects of elevated seawater pCO_2 on gill gene expression in the green crab *Carcinus maenas*, found structural molecule activity (GO:0005198) and cell (GO:0005623) as the only over-represented GO terms (Fehsenfeld *et al.*, 2011). Upregulated transcripts associated with these terms included 8 ribosomal proteins, alpha-tubulins and several membrane proteins (Fehsenfeld *et al.*, 2011). A transcriptomic study investigating the effects of climate-associated stressors (i.e. ocean acidification and elevated sea-surface temperatures) on *Haliotis tuberculata*, identified signalling, biological regulation and response to stimulus, as some of the over-represented GO biological processes identified in adult abalone compared to larvae (Harney *et al.*, 2016). Proteins associated with similar GO terms in the current study include cytoskeletal proteins (such as chaperonin containing T-complex, tubulin, radixin and fascin) and structural proteins (myosin, Cdc42, ras GTPase-activating-like protein IQGAP1 and sorting nexin). Differences observed between the current study and published research could be attributed to differences in tissue type (i.e. gills vs. haemocytes), as well as the use of updated GO databases in the current study which allows for improved annotation of the dataset. Furthermore, the type of -omics study can have an impact on data analysis, where, for example, proteomic datasets can be incomplete due to discrepancies in identification and the limited sensitivity experienced by some detection technologies (Pei *et al.*, 2017).

3.4.2 Functional analysis of candidate proteins for biomarker discovery

Following the identification of statistically significant ($p < 0.05$) and biologically relevant (fold-change > 1.2) proteins in Chapter 2, it was necessary to reconcile the two main components of the proteomics study (i.e. quantitative and qualitative analyses) to ensure

effective biomarker discovery. As such, a total of 33 statistically significant and/or biologically relevant proteins (Table 2) were identified from Tables 3-5 (Chapter 2) and the functional analysis of the PU eigengene modules (Table 1).

i. Metabolism & ATP hydrolysis

The physiological impact of ocean acidification is largely dependent on a marine organism's ability to tolerate the upper limits of environmental physico-chemical parameters, which can result in adverse effects on energy homeostasis and other downstream functions (Lannig *et al.*, 2010). Maintaining an equilibrium between energy generation and expenditure is critical in cellular homeostasis, where maintaining homeostasis while conducting "business as usual" can result in trade-offs in different organisms (Sokolova, 2013). Normal environmental conditions allow for optimal energy allocation in organisms, suggesting unencumbered aerobic scope for complete investment into Darwinian fitness (Sokolova, 2013). Interestingly, under moderate stress, a marine organism might be capable of compensating for the increased energy requirements. This is achieved through increasing energy generation and integration, as well as, in some cases, through metabolic regulation to meet elevated ATP demands (Lannig *et al.*, 2010). Conversely, extreme stress conditions are more likely to hamper this compensatory mechanism, leading to trade-offs such as metabolic depression to conserve energy until conditions return to normal, which can lead to reduced development and growth rates.

Following a 28-day exposure of *Mytilus galloprovincialis* to low pH, metabolic depression was observed that resulted in low energy expenditure, while the upregulation of carbonic anhydrase was attributed to acid-base regulation (Freitas *et al.*, 2017). Freitas *et al.* (2017) suggest this response is likely to lead to impaired reproductive activity, growth and immune system defences in mussels. Contrastingly, *C. gigas* exposed to ocean acidification conditions for 28 days experienced energy modulation, where aerobic metabolism was hampered and anaerobic metabolism was stimulated, with an observable shift *towards*, and an increase *in*, glycolysis (Cao *et al.*, 2018a). Following this, the authors challenged stressed oysters with a marine *Vibrio* spp. to determine the risk of pathogenic infection in the low pH environment. The study concluded that, despite the compensation of reduced aerobic metabolism by anaerobic metabolism and increased glycolytic enzyme activity, energy modulation and antioxidant regulation of *C. gigas* was notably hampered. This was reflected by increased oxidative damage, a shift from aerobic to anaerobic metabolism, and a decline in energy reserves, thus

suggesting an inability to survive for prolonged periods in acidic waters, coupled with an increased risk of pathogenic infection (Cao *et al.*, 2018).

Three of the 33 candidate proteins were found to be predominantly involved in metabolic related functions, which include glucose metabolism and ATP hydrolysis. Solute carrier family 2 (GLUT1/SLC2A4) facilitates the movement of glucose across membranes and is implicated in cellular response to glucose starvation (Table 3). Being a glucose carrier protein, GLUT1 is found in most cells that require carbohydrate metabolism and has been sparsely studied in a number of crustaceans (Wang *et al.*, 2016; Wang *et al.*, 2017). In oysters, GLUT1 was found to be upregulated under low salinity conditions, while under combined stress conditions (elevated temperature and low salinity), GLUT1 was downregulated (Ertl *et al.*, 2019). Little is known about the components of carbohydrate transport in marine invertebrates, but the predominating function of GLUT transporters is to regulate glucose homeostasis (Martínez-Quintana and Yepiz-Plascencia, 2012).

Enolase (ENO1) and phosphoglycerate kinase (PGK) are both implicated in glucose metabolism either through glycolysis, gluconeogenesis, or both (Table 3). Enolase is a highly conserved glycolytic enzyme that is activated by a number of divalent metal ions such as magnesium and manganese (Pancholi, 2001). It catalyses the dehydration of 2-phospho-D-glycerate (PGA) to phosphoenolpyruvate (PEP) (Pancholi, 2001). It has been reported that enolase is differentially expressed following stress exposure, such as elevated temperatures, carbohydrate deprivation and exposure to hypoxic conditions (Vincenzetti *et al.*, 2017). In a study investigating ENO1 expression in a micro-alga (*Dunaliella salina*) following thermal stress, ENO1 was found to be upregulated under elevated temperatures compared to its expression at lower temperatures, suggesting a possible contribution to *D. salina's* thermal tolerance (Ruan *et al.*, 2009). PGK is only involved in ATP synthesis in the glycolytic pathway, and has been implicated in the generation of ATP alongside pyruvate kinase under prolonged hypoxic conditions in *H. midae* (Venter *et al.*, 2018). PGK and ENO1 have not been extensively studied in marine invertebrates, particularly in response to changing environmental conditions, and as such, the response of these proteins in *H. midae* under reduced pH conditions is novel.

The PPI network showed a predicted interaction between ENO1 and PGK (Figure 6), suggesting they are likely co-expressed under reduced pH conditions, and are potentially implicated in the glycolysis pathway in *H. midae* (Table 3). Glycolysis, the anaerobic

breakdown of glucose, has been identified as one of the essential adaptations of metabolic activity in animals exposed to variable environmental conditions, particularly when the oxygen demand exceeds the supply (Drozdov and Drozdov, 2016). For example, the gills, foot muscle, adductor muscle and haemocytes of abalone have been found to respond to low-oxygen/hypoxic conditions by opting for anaerobic glycolytic metabolism to generate ATP (Venter *et al.*, 2018). Due to their variable environment, several intertidal marine invertebrate species have evolved effective anaerobic glycolytic pathways to yield more ATP and possess fewer metabolic proteins, relatively higher levels of glycolytic substrates and the ability to reduce excessive cellular acidosis through improved tissue buffering capabilities (Sokolova, 2018).

Although bioenergetics and pathogenic infection weren't closely monitored in this study, a plethora of marine studies focusing on the impacts a changing climate has on metabolism and susceptibility to infection have been published in the literature (Cheng *et al.*, 2004; Tomanek, 2011; Dang *et al.*, 2012; Cao *et al.*, 2018a; Cao *et al.*, 2018b). In molluscs, the overall effect appears to be a state of immunosuppression induced by environmental changes, leading to metabolic depression and shifts towards anaerobic metabolism to cope with oxidative stress and pathogenic infection. However, it is important to note that there are instances of species-specific responses, which should be bore in mind.

ii. Stress/immune response

The molluscan innate immune system is complex and efficient in protecting against pathogenic infection, enabling molluscs to successfully inhabit a range of aquatic and terrestrial environments (Guo *et al.*, 2015; Sokolova, 2009). Indeed, the stress response is closely linked to the immune response of molluscs, both conferring host defence/protection in a complex interplay of the two components (Guo *et al.*, 2015). Furthermore, functionally divergent immune genes could have overlapping or dual mechanisms of action under stress conditions, especially since numerous molecular pathways have diverse functions upon induction (Zhang *et al.*, 2015). The cellular defence component of the invertebrate innate immune system is represented by a group of immune cells, namely the haemocytes. Haemocytes of abalone and other molluscs, are the drivers of both the stress response and immune response, being critical in producing mediators of stress and immunological responses (Hooper *et al.*, 2007). Haemocytes respond to a variety of biotic and abiotic stimuli, such as through the recognition

of pathogen-associated molecular patterns (PAMPs) following pathogenic invasion (Adamo, 2012). And while all invertebrates lack acquired immunity, molluscs have evolved astute mechanisms for generating thousands of proteins from a limited number of genes, allowing for a well-equipped innate immune and stress response system. Though poorly characterised in molluscs, vertebrate studies implicate a complex relationship between the nervous, endocrine and immune systems for cellular communication. The stress response commences in the endocrine system with changes in norepinephrine, culminating in the release of biogenic amines that initiate the secondary stress response which is a miscellany of metabolic changes with the sole purpose of maintaining cellular homeostasis (Hooper *et al.*, 2007).

In the current study, 10 of the 33 statistically significant and/or biologically relevant proteins were associated with stress/immune-related functions (Table 3). Several of these proteins are also involved in signal transduction and cytoskeletal organisation. In a transcriptomic study of the pearl oyster mantle tissue following elevated CO_2 and temperature stresses, "cell stress responses" was one of the enriched GO terms associated with the differentially expressed unigenes (Li *et al.*, 2016b). Heat shock protein 70, associated with this GO category, was found to be upregulated in response to all four treatments (low pH and elevated water temperatures). In a comparative proteomics study of a wild-type and a selectively bred Sydney rock oyster population under ocean acidification conditions, "stress response" was one of the functional categories enriched in the wild type group following exposure to reduced pH conditions (Thompson *et al.*, 2015). This generalised stress response was represented by several heat shock proteins, antioxidant proteins and cytoskeletal proteins (Thompson *et al.*, 2015). Thus, in the current study, it is unsurprising that a generalised stress response induces a range of proteins with key roles in various processes within the cell.

Ras GTPase-activating-like protein IQGAP1, a highly conserved scaffold protein that is multifunctional in the immune system of most organisms, was found to have a fold-change of approximately 1.3 in *H. midae* haemocytes following reduced pH conditions (Table 2). IQGAP1, through its multidomain structure, facilitates the formation of various protein complexes necessary for carrying out innumerable functions, particularly in cell signalling pathways and the actin cytoskeleton (Pathmanathan *et al.*, 2011). Indeed, ClueGO analysis of the candidate proteins found IQGAP1 to be associated with several Ras signalling cascades and related biological processes, as well as being implicated in cell membrane functions (Figure 5). IQGAP1 and its interactors have been extensively studied in vertebrates, where, for example, IQGAP1 favourably binds to the active form of Cdc42. GTP-activated Cdc42 has

been found to be associated with the formation of filopodia and cell membrane extensions, thus implicating the overexpression of IQGAP1 in also regulating cell morphology through filopodia formation (Briggs and Sacks, 2003). Furthermore, the calponin homology domain of IQGAP1 interacts with F-actin to form cross-links, which is made possible through the coupling of the GTP-activated form of Cdc42 to F-actin (Hedman *et al.*, 2015). IQGAP1 has also been found to interact with calmodulin and calmodulin-like proteins through its four IQ motifs. F-actin cross-linking via IQGAP1 is inhibited in the presence of Ca^{2+}-activated calmodulin by reducing the association between Cdc42 and IQGAP1 (Briggs and Sacks, 2003; Pathmanathan *et al.*, 2011). This suggests that Ca^{2+} has a pivotal role in regulating the actin cytoskeleton, and by association IQGAP1. IQGAP1 also serves as a scaffold protein for several signalling cascades, including the ERK/MAPK pathway where it has been found to bind to B-Raf, MEK 1/2 and ERK 1/2, thus regulating MAPK activity (McNulty *et al.*, 2011).

Mitogen-activated protein kinase (MAPK1), a super family involved in the ERK/MAPK signalling cascade, transduces signals from external stimuli at the cell surface, while also regulating cell proliferation, inflammation and apoptotic-related cellular functions (Table 3). The MAPK superfamily participates in three main pathways: the extracellular signal-regulated protein kinase pathway (ERKs), the c-Jun N-terminal kinase pathway (JNK), and the p38 kinase family pathway (p38) (Cowan, 2003). The ERK pathway functions as an archetypal MAPK cascade, being largely involved in cell proliferation and development by responding to environmental stressors and other growth factors (Perera *et al.*, 2016). In vertebrates, the ERK pathway has been identified as a crucial regulator of innate and adaptive immunity. MAPK1 is the second member of the ERK subfamily, with homologs being found in humans, mice, frogs and zebrafish (Perera *et al.*, 2016). ERK induction in zebrafish could be linked to heat shock protein activation following thermal stress. However, the functional role of MAPK1 in molluscan immunity has largely remained unexplored, thus further investigation into its role in innate immunity is required (Perera *et al.*, 2016). Perera *et al.* (2016) identified a MAPK1 homolog in the disk abalone and were able to characterise its role in trochophore development, as well as its putative response to pathogenic infection, suggesting that MAPK1 is prominently involved in the immune response. Novel p38 MAPK and JNK were identified in *Crassostrea hongkongensis*, and found to be upregulated in response to pathogenic infection and stimulation of PAMPs (Wang *et al.*, 2018). Conversely, the MAPK genes, ERK and JNK, were significantly upregulated in response to pathogenic infection in the Yesso scallop compared to p38, suggesting that only ERK and JNK are involved in the innate immune response of the

scallop (Sun *et al.*, 2016). This could infer species-specific differences in the roles of the various MAPK proteins. In the present study, MAPK1 was found to have a fold-change of 1.4, which was statistically significant ($p \leq 0.05$), following exposure to reduced pH conditions. This upregulation suggests a link between the stress and immune responses in *H. midae*. The endoplasmic reticulum is a crucial site for protein folding and modification, as well as serving as a store for intracellular Ca^{2+}, which is essential in signal transduction (Darling and Cook, 2014). Changes in environmental Ca^{2+} or redox states can adversely affect protein folding, leading to the accumulation of unfolded proteins and subsequently, a series of signalling cascades known as the unfolded protein response (UPR). Ultimately, the UPR seeks to restore ER homeostasis through adaptation, but can also lead to cell death as a consequence of redox species accretion (Darling and Cook, 2014). MAPK signalling pathways have been shown to play a role in ER stress, each pathway having varying kinetics and being responsible for different functions. Furthermore, ERK/MAPK signalling cascades have been found to regulate a number of heat shock proteins (HSPs) responsible for correctly folding unfolded/misfolded proteins in response to various stressors (i.e. heat stress, Figure 5) (Wang *et al.*, 2013). 78 kDa glucose regulated protein (GRP78/BiP) is a member of the HSP70 family and chaperonin containing T-complex (TCP-1), a member of the HSP60 family, are both examples of proteins included in the candidate biomarker group that are regulated by ERK/MAPK signalling cascades (Table 3, Figure 6).

78 kDa glucose-regulated protein (GRP78/BiP) is a molecular chaperone that is highly conserved across taxa, being an important promoter of protein folding, particularly following ER stress (Darling and Cook, 2014). Members of the HSP70 subfamily are also involved in a number of other functions, such as protein assembly, intracellular localisation, secretion, regulation and degradation of proteins (Hamer *et al.*, 2004). GRP78 is considered a key regulator of the UPR, being largely localised to the ER lumen. The ER has come to represent the site of pro- and anti-apoptotic processes, where tolerating ER stress requires the UPR to inhibit general translational processes while prioritizing the expression of chaperone molecules such as GRP78 (Lee, 2005). Since GRP78 has been linked to anti-apoptotic activities via the disruption of caspase activation, upregulation of GRP78 in response to pathogenic infection suggests it could serve as part of a "protective mechanism" for surviving ER stress (Lee, 2005). In the current study, GRP78/BiP was significantly upregulated ($p \leq 0.05$) following exposure to the stressor (Table 2). This might suggest that anti-apoptotic "protective mechanisms" are being activated in *H. midae* in response to reduced pH conditions, since GRP78 induction could

be considered a suitable indicator of ER stress and likely the effector of the UPR pathway. This further supports the supposition that the stress and immune responses are closely linked.

Ca^{2+} is a crucial secondary messenger for numerous processes such as muscle contraction, cell proliferation, gene expression and neurotransmitter release (Picciotto *et al.*, 1995). Calmodulin and calmodulin-like proteins are effectors of Ca^{2+} regulation via the CAMK and MAPK signalling cascades. Calmodulin-dependent protein kinase (CAMK) is a member of the CAMK signalling cascade, acting as a multi-functional protein kinase upon calmodulin activation via Ca^{2+}-binding. In vertebrates, Ca^{2+}-facilitated signalling is essential during fertilisation and embryonic development, signal transduction via the MAPK cascade, intracellular protein transport, protein secretion and memory (Epelboin *et al.*, 2016). Since Ca^{2+} is essential in mediating several cellular processes, oscillations in intracellular Ca^{2+} concentrations can initiate the translocation of transcription factors in the cytoplasm to the nucleus, such as NF-κB (Pinto *et al.*, 2015). This occurs through the initial phosphorylation cascade that results in dissociation of the inhibitor IκB from NF-κB in the cytoplasm, exposing the nuclear localisation signals of NF-κB proteins. IκB is activated for dissociation by Ca^{2+}/CAMKs, after which NF-κB translocation induces the transcription of genes involved in inflammatory or immune response pathways, such as cell proliferation (Pinto *et al.*, 2015). Thus, Ca^{2+} and CAMK are key components of cell signalling related to NF-κB and subsequent immune system responses. Several CAMKs have been implicated in host defence in the horseshoe crab following pathogenic infection (Belinda *et al.*, 2008). Another study identified an homologous CAMK-related protein in the pearl oyster, which was found to play a crucial role in Ca^{2+} signalling and subsequent downstream cellular processes, as well as in shell and pearl formation (Dai *et al.*, 2005). All this highlights the diverse roles Ca^{2+} and CAMK (and CAMK cascade members) play in vertebrates and invertebrates alike. In the present study, the significant ($p \leq 0.05$) upregulation of CAMK may indicate induction of crucial signalling cascades (i.e. MAPK and NF-κB pathways) as part of a stress response in *H. midae* exposed to reduced pH conditions. Pointedly, CAMK (as indicated by homology with CAMK2G) was found to be associated with "cellular response to heat stress" and Ras-related signalling events in the ClueGO analysis (Figure 5), further suggesting an overlap/link between the stress and immune responses in *H. midae*.

Serpin B-like protein 2 (SerpinB2) is part of a family of serine protease inhibitors involved in numerous innate immune functions such as phagocytosis, antimicrobial peptide synthesis and coagulation (Bathige *et al.*, 2015). They are essential in that superfluous protease activity can

be detrimental. Serpins are part of a group of trapping inhibitors that form irreversible complexes and are well-documented in humans. SerpinB2 is a member of Clade B serpins. Although not yet fully characterised in other organisms, and especially in molluscs, Clade B serpins are of particular interest since they specifically prevent unnecessary cytotoxic apoptotic protease-activity (Law *et al.*, 2006). Thus, it is possible that the SerpinB2 identified in haemocytes sampled from *H. midae* exposed to reduced pH conditions has an anti-apoptotic function. Indeed, serpin B2 had a fold-change of 1.3 which we deemed to be biologically relevant (Table 2). The reactive-centre loop (RCL) of molluscan serpins, which exhibits inhibitory activity, interacts with target proteases secreted by pathogens and inhibits their activity (Bathige *et al.*, 2015). Thus, SerpinB2 may also serve as an important constituent of the immune response, particularly in response to pathogenic infection. Although the serpin family isn't well characterised in molluscs, three proteins similar to clade B serpins have been identified in the disk abalone (Bathige *et al.*, 2015). Following pathogenic challenge, expression analysis revealed an upregulation of the serpin B proteins, suggesting a possible role in defence against pathogenic infection and associated tissue injury in the disk abalone (Bathige *et al.*, 2015).

Haemocyanin (Hc), a member of the phenoloxidase (PO) family, relies on copper binding to function as a transport protein. Being abundant in the haemolymph of invertebrates, Hc is essentially an oxygen transporter, but has also been implicated as a homeostatic protein involved in processes such as osmoregulation, hormone transport and protein storage (Coates and Nairn, 2014). Since Hc also exhibits PO activity, it is considered a constituent of the immune system. Following the detection of pathogenic infection in invertebrates, through processes such as PAMP binding and cuticle damage, induced POs initiate the proPO activation cascade that results in the formation of melanin which is an antimicrobial pigment (Coates and Nairn, 2014). As a result of PO activity, reactive oxygen species (ROS) are generated as a by-product, which can prove detrimental to the host's cells if not removed efficiently. It is believed that the mechanisms involved in proPO activation are involved in the conversion of Hc to a PO-like protein which functions in the immune response of numerous invertebrates. Hc can be converted into a PO-type enzyme either through *in vivo* proteolytic digestion or by altering protein conformation (Coates and Nairn, 2014). Hc has been implicated as an immune response gene in several marine invertebrates exposed to environmental change, which has been shown to cause immunosuppression and increased susceptibility to infection (Castellanos-Martínez *et al.*, 2014; Duan *et al.*, 2013; Zhuang *et al.*, 2015). Although there is a paucity of published

information regarding the effect of ocean acidification on molluscan Hc, abalone haemolymph has been found to have an inefficient pH buffering ability, and thus haemolymph is susceptible to tissue acidosis, particularly under hypoxic conditions (Venter *et al.*, 2018). Furthermore, Hc reportedly exhibits a reverse Bohr Effect, meaning oxygen binds more effectively to the molecule at lower pH conditions (Venter *et al.*, 2018). This is especially beneficial under hypoxic conditions since oxygen saturation can be maintained. In a transcriptomic study investigating the effect high CO_2 has on the hypoxic response of white leg shrimp Hc isoforms, it was found that all isoforms were upregulated under hypoxic conditions alone, while their expression was unaffected by high CO_2 coupled with hypoxia stress (Johnson *et al.*, 2015). Conversely, elevated sea surface temperatures have a suppressive effect on the expression of Hc in *Octopus maya*, despite eliciting enhanced PO activity (Pascual *et al.*, 2019). This reflected the metabolic stress induced by higher temperatures, as well as indicating an immune compensatory mechanism with higher phenoloxidase activity (Pascual *et al.*, 2019). *Haliotis fulgens* facing short term environmental extremes has exhibited higher levels of Hc expression under hypoxic conditions than under elevated temperatures or combined stressors (Samuel *et al.*, 2019). The authors suggest that this could be due to an immediate increase in metabolic activity, although this did trigger an increase in respiration despite a decline in Hc abundance. Notwithstanding this contradiction, the authors note that abalone Hc has low oxygen carrying capacity, with most of the oxygen being transported freely in the haemolymph. Consequently, it is suggested that Hc best functions in serving as an energy source under elevated temperatures (Samuel *et al.*, 2019). As such, the upregulation of Hc in the present study could be in response to adverse effects on the abalone metabolism, leading to an increased demand for more effective oxygen transport to fuel protein turnover. Furthermore, Hc could also potentially be involved in phenoloxidase activity to mitigate ER stress and remove damaged tissues/haemocytes. This could suggest that Hc is not only an immune response protein, but also serves as a stress protein induced by ocean acidification stress. However, Hc and PO activity will need to be monitored in future studies to confirm this hypothesis.

iii. Cytoskeleton, protein sorting & signal transduction

The cytoskeleton is a complex and integral component in regulating innumerable processes such as cell motility and division, cellular organisation, intracellular transport, and the formation of cell-cell junctions (Chang *et al.*, 2016). It has been postulated that ocean acidification induces a generalised stress response in bivalves through changes in the electron

transport chain and the generation of ROS that induce oxidative stress. In turn, this might lead to cellular damage and the disruption of the cytoskeleton (Goncalves *et al.*, 2016; Cao *et al.*, 2018b). Specifically, the actin and tubulin complex is considered the cytoskeletal component most vulnerable to oxidative stress (Cao *et al.*, 2018b). Upregulation of cytoskeletal proteins might be a compensatory mechanism following increased protein turnover as a result of oxidative stress. Furthermore, signalling peptides with an overall pK_a around the pH of seawater are likely to be protonated under ocean acidification conditions, thus altering their charge and impacting efficient cell-cell signalling and intracellular signal transduction (Melzner *et al.*, 2020). Cao *et al.* (2018b) conducted an integrated proteomics study on the Pacific oyster, which investigated the effects of ocean acidification on the immune response following pathogenic infection. The actin cytoskeleton KEGG pathway was found to be over-represented in both single-stress treatments, as well as in the combined stress treatment (Cao *et al.*, 2018b). Similar results were observed in the Sydney rock oyster following exposure to elevated pCO_2 (Ertl *et al.*, 2016b). Interestingly, short-term exposure of Pacific oyster larvae to ocean acidification conditions (pH 7.5) resulted in substantial suppression of cytoskeletal proteins (Dineshram *et al.*, 2012). This suppression was also observed in geoduck clam larvae under similar conditions (Timmins-Schiffman *et al.*, 2019). This suggests that molluscs respond differently to an environmental stress at different stages of their life cycle. In our study, six cytoskeletal-related/sorting proteins were included in the 33-member list of candidate biomarkers. ClueGO analysis of the candidate biomarker group showed an over-representation of "cytosolic transport" and "endosome to Golgi transport" functions (Figure 5).

TCP-1/CCT is considered an HSP60-like protein that is a member of the TRiC (TCP-1 Ring Complex) family, which functions as a chaperone protein during protein folding in the cytoskeleton, while also playing an essential role in tubulin biogenesis in the cytoplasm (Truebano *et al.*, 2013). This is observed in the PPI network where TCP-1 and tubulin beta chain are predicted interactors (Figure 6). Furthermore, members of this family are responsible for targeting aggregations of damaged proteins for degradation, thus mitigating cellular damage, particularly under adverse environmental conditions (Zhang and Storey, 2018). TCP-1 is unique in its structure as a chaperone protein in that it is comprised of 8 subunits arranged to form a ring (TRiC) that serves as a substrate for the binding and folding of proteins (Brackley and Grantham, 2009). In the current study, TCP-1 had a fold-change of 1.3 in response to the stressor (Table 2). Following exposure of brine shrimp to ocean acidification conditions (pH 7.6 and 7.8), TCP-1 (among other cytoskeletal proteins) was upregulated after 7 days of OA

stress (Chang *et al.*, 2016). TCP-1 has been found to respond differently to other abiotic stressors. For example, TCP-1 was found to be downregulated in a transcriptomic study investigating the response of invasive (*Mytilus galloprovincialis*) and native (*Mytilus trossulus*) blue mussels to salinity stress (Lockwood and Somero, 2011). Conversely, Truebano *et al.* (2013) observed an upregulation of TCP-1 and other cytoskeletal proteins in the Antarctic clam (*Laternula elliptica*) following exposure to elevated temperatures. Upregulation of TCP-1 observed in our study might be in response to an increased need for cytoskeletal proteins such as tubulin. Furthermore, TCP-1 upregulation could be a consequence of increased protein turnover, as well as oxidative stress that ultimately disrupts the cytoskeleton, due to reduced pH conditions.

Sorting nexin (SNX6) is part of a group of nexins that function in endosome formation, protein transport, as well as the formation of protein complexes (Worby and Dixon, 2002). Being well-characterised in vertebrates, sorting nexins function closely with a range of vacuolar protein sorting proteins (Carlton *et al.*, 2005). This can be seen in the PPI network (Figure 6). BLAST analysis confirmed homology between sorting nexin identified in this study and sorting nexin 6 from *Mizuhopecten yessoensis.* GO analysis of sorting nexin 6 found "intracellular protein transport", "response to epidermal growth factor" and "regulation of macroautophagy" as biological processes associated with the protein (Figure 5 and Table 3). SNX6 is grouped with other sorting nexins and VPS5 in the cytoplasm. The latter is a vacuolar sorting protein involved in vacuolar to trans Golgi network trafficking (Worby and Dixon, 2002). Sorting nexins contain characteristic phox homology (PX) domains, an important region for recruitment of phosphoinositides for the formation of endosomes of the early endocytic network (Cullen, 2008). PX-containing proteins have been implicated in several human diseases (Teasdale and Collins, 2012; Worby and Dixon, 2002). Some sorting nexins have complementary functions in haemopoietic cells during inflammation, where, in mammalian platelets and endothelial cells, they are involved in mediating the adhesion and extravasation of leucocytes following inflammation (Teasdale and Collins, 2012).

Among the other proteins involved in vesicular trafficking identified in the present study, charged multivesicular body protein 1a (CHMP1a) was found to be upregulated, with a fold-change of 1.5 (Table 2). Vesicular transport allows membrane-bound organelles within cells to communicate, particularly those that are essential in protein synthesis, degradation and surface receptor signalling (Stauffer *et al.*, 2001). CHMP1a is one of the two isoforms of CHMP1 in humans and the exact function of each remains unclear (Li and Blissard, 2015). CHMP1 is a

highly conserved protein that has been implicated in multivesicular body (MVB) formation in mammals. MVBs are cargo-filled late endosomes that transport proteins, such as receptors and hydrolases, to the interior of lysosomes (Stauffer *et al.*, 2001). Essentially, the formation of MVBs, and their subsequent trafficking, serves as one of two major systems for mediating protein degradation within eukaryotic cells (Urbé, 2005). Furthermore, MVBs have been found to fuse with autophagosomes in autophagy, a lysosomal degradation pathway, to form an amphisome (Fader and Colombo, 2009). Amphisomes associate with lysosomes to degrade the cargo being transported in them. This highlights the importance of CHMPs and their involvement in forming MVBs, especially since changing environmental conditions can lead to cellular oxidative stress, and subsequent induction of compensatory cellular mechanisms for protein damage. In *Saccharomyces cerevisiae*, gene silencing of several CHMP genes resulted in the disruption of the endosomal membrane and incorrect sorting of two key hydrolases, thus validating their importance in endosomal formation and sorting (Stauffer *et al.*, 2001). CHMP1, in mammals, is considered an accessory protein of the ESCRT-III complex, which functions in membrane scission and vesicle release (Li and Blissard, 2015). The ESCRT (endosomal sorting complex for transport) machinery has 3 highly conserved and distinct functions that target ubiquitylated membrane proteins to lysosomes (Raiborg and Stenmark, 2009). CHMP proteins have not been investigated in molluscan studies, particularly those pertaining to environmental stress. CHMP1 (and other CHMPs) has been found to interact with vacuolar protein sorting 4 (VPS4) necessary for MVB sorting in yeast (Stauffer *et al.*, 2001; Urbé, 2005; Raiborg and Stenmark, 2009).

Vacuolar protein sorting 13a (VPS13a), identified in the present study (fold-change ≥ 1.2, Table 2), is a member of a group of conserved class A VPS13 proteins that are involved in an array of functions in humans, most notably vesicular trafficking and protein transport (Muñoz-Braceras *et al.*, 2019). VPS13a is diverse in functionality, where, in humans, it has been reported to be involved in cytoskeletal organisation, phosphoinositide regulation, calcium homeostasis and autophagy (Kumar, 2018; Muñoz-Braceras *et al.*, 2019). VPS13 proteins are predominantly localised at the interface between the mitochondria and vacuole, and are hypothesised to facilitate lipid transfer from vacuoles and other membranes, potentially serving as an alternative route for the movement of specific lipids between the ER and mitochondria (Kumar *et al.*, 2018). In a study investigating the phagosome proteome of the ciliate *Tetrahymena thermophila*, GFP-tagged VPS13 was found to associate closely with the phagosome membrane throughout phagocytosis, while VPS13-knockout cells exhibited

inefficient phagocytosis (Samaranayake *et al.*, 2011). This suggests a crucial role for VPS13 in phagocytosis. VPS13a gene silencing in human and *Dictyostelium discoideum* cell lines led to defects in autophagy, suggesting a likely outcome of general inhibition of the endolysosomal pathway (Muñoz-Braceras *et al.*, 2019). Degradation by autophagy is essential for maintaining cellular homeostasis, particularly in neuronal survival, and defects in this cellular pathway can lead to neurodegenerative diseases in humans (Muñoz-Braceras *et al.*, 2019). VPSs (especially VSP13a) have not been widely investigated in molluscs, but a microarray study of thermal adaptation in an Antarctic fish *Trematomus bernacchii* identified an upregulated VPS26 after 4-hours of exposure to heat stress (Buckley and Somero, 2009). In another study investigating haemocyte gene expression in Manila clams (*Ruditapes philippinarum*) 10-hours post-pathogenic infection, two VPS-like proteins were among the upregulated proteins, while "ion transmembrane transport processes" was one of several enriched GO-terms (Hasanuzzaman *et al.*, 2018). Thus, their exact role requires further elucidation in *H. midae* exposed to OA stress conditions.

Sorting nexins, CHMPs and VPSs clearly have an interactive role in maintaining homeostasis through protein sorting and trafficking, as well as through lysosomal degradation of damaged cellular components through autophagy. As such, it is unsurprising that these proteins were substantially upregulated in the present study, while PPI network analysis identified interacting complexes between these proteins and their family members, as well as other proteins involved in vesicular trafficking and protein sorting (Figure 6).

3.5 Summary & Conclusion

Discovery proteomics is a thrilling avenue for investigating the effects of climate change on non-model organisms. With the aid of innumerable bioinformatics tools and solutions, homology-based functional analysis is possible. This study utilised several popular tools to functionally analyse LC-MS/MS data using homology with other molluscan species, and the well-sequenced genomes of *Homo sapiens* and *Rattus norvegicus*. The functional analysis pipeline used in this study aided in the identification of 33 candidate biomarkers that were further investigated for their potential as biomarkers of OA stress in *H. midae* through data mining. Functional analysis and data mining of these candidates proposed a dynamic and complex interplay of the stress and immune responses in *H. midae*, particularly through the upregulation of cytoskeletal proteins in an attempt to restore cellular homeostasis through cytoskeletal re-organisation and stabilisation, under stress conditions. Additionally, the induction of several signalling cascades by virtue of the upregulation of their drivers (i.e. MAPK, CAMK, IQGAP1) highlights the importance these pathways have in eliciting a stress response, such as the UPR and inflammation. Data mining further emphasized the importance of Ca^{2+} in regulating and driving innumerable biological processes, thus being of notable importance in the context of ocean acidification and related studies. This chapter consequently provides supporting evidence of the potential several of these candidate proteins have in serving as biomarkers of OA stress.

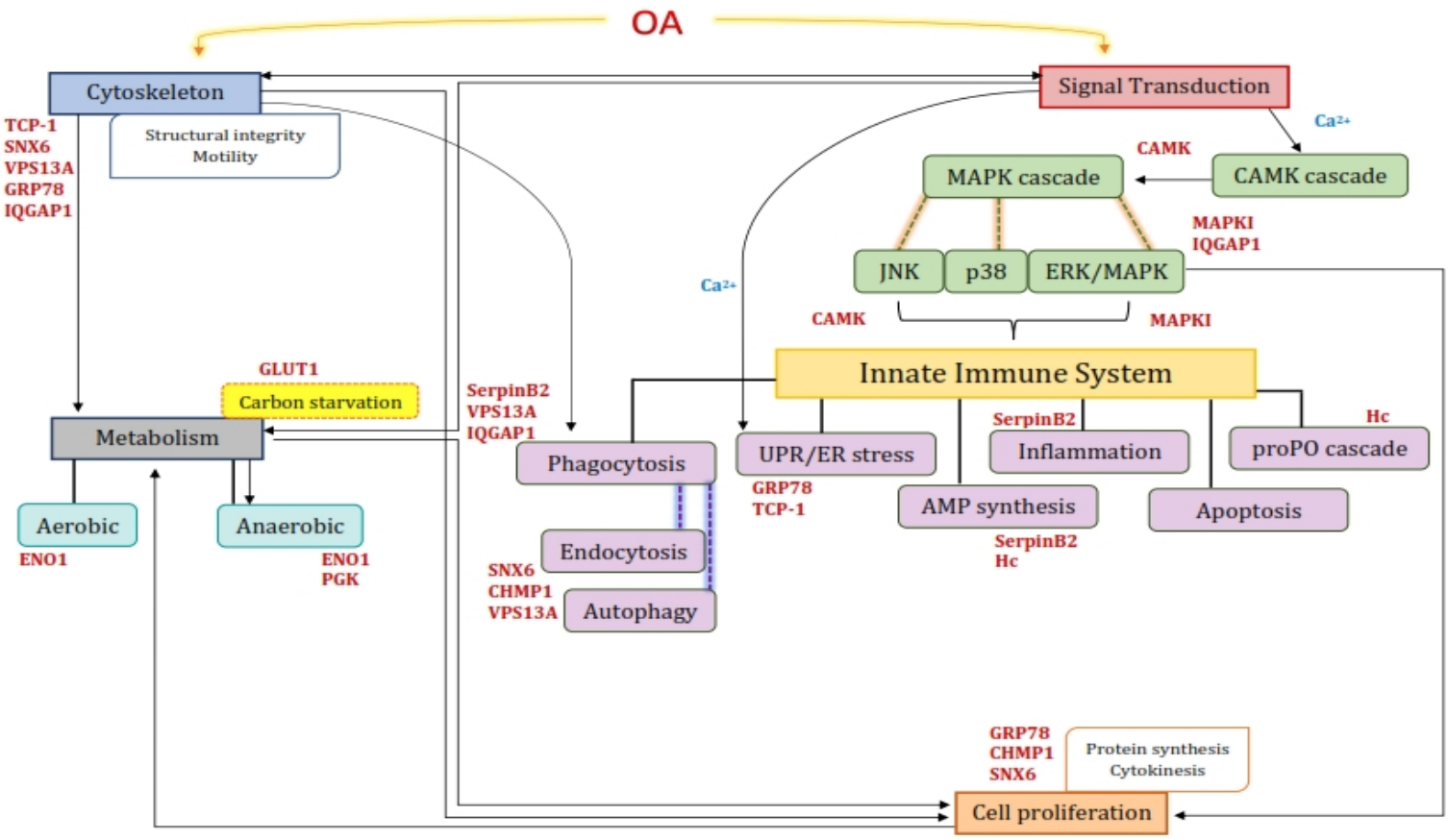

Figure 7. Schematic representation of potential key proteins of OA stress in relation to their putative functional roles and linked cascades in *H. midae* exposed to reduced pH conditions.

Ocean acidification (OA) induces a series of responses within the cell, specifically related to signal transduction and the cytoskeleton. Consequently, intracellular Ca^{2+} concentrations fluctuate, leading to initiation of the CAMK and MAPK signalling cascades (Figure 7), which are ultimately directed by calmodulin and CAMKs to initiate MAPK signalling. MAPK1 and IQGAP1, separately, mediate various sub-pathways of the MAPK signalling cascade (i.e. JNK, p38 and ERK/MAPK), all of which have a role to play in an array of immune-related responses (Figure 7). CAMK also facilitates various aspects of the innate immune system. Phagocytosis, an active immune defence against pathogen infections, is usually associated with related functions such as endocytosis and autophagy. Key proteins involved in phagocytosis, endocytosis and autophagy include SerpinB2, VPS13A (vacuolar protein sorting 13A), IQGAP1 (ras GTPase-activating protein IQGAP1), SNX6 (sorting nexin), and CHMP1 (charged multivesicular body protein 1) (Figure 7). SerpinB2, also involved in antimicrobial peptide synthesis and the inflammatory response in vertebrates, functions as a serine protease inhibitor as excessive protease activity can be damaging. VPS13a is a member of a group of

conserved VPS13 class A proteins, largely involved in vesicular trafficking and protein transport. It has been implicated in autophagy, and also regulates calcium homeostasis and cytoskeletal organisation. Although not well-characterised in molluscs, VPS13-like proteins have been detected in thermal stress and pathogenic challenge studies in fish and clams, respectively. SNX6 and CHMP1 have been found to interact in related pathways with VPS13 and other VPS proteins, where they are involved in endosome formation and intracellular protein transport. Furthermore, both proteins are also involved in cell proliferation in that they transport components necessary for protein synthesis, particularly between the Golgi and early endosome. IQGAP1, a highly conserved scaffold protein, is involved in a number of processes, particularly in the cytoskeleton and signal transduction pathways such as the MAPK cascade. During phagocytosis in vertebrates, IQGAP1 binds preferentially to GTP-activated Cdc42, forming a complex that functions in filopodia and cell membrane extensions. In the cytoskeleton, IQGAP1's role in F-actin cross-linking is inhibited in the presence of Ca^{2+}-bound calmodulin, thus limiting its ability to confer structural integrity within the actin cytoskeleton. Other immune-related functions, such as the UPR and apoptosis, are regulated or mediated by other members of the candidate biomarker group. TCP-1 and GRP78 are both involved in the UPR as a result of ER stress (Figure 7). GRP78 serves as a member of the HSP70 family, promoting protein folding following ER stress. It also plays a role in other functions such as protein assembly and localisation, as well as the secretion, regulation and degradation of proteins. TCP-1 is a key cytoskeletal protein comprising the TRiC complex, which is essential in tubulin synthesis. It is a member of the HSP60 family and as a chaperone protein of the TRiC complex, TCP-1 is critical in protein folding in the cytoskeleton. Furthermore, under cellular stress, TCP-1 targets aggregations of damaged proteins for degradation to prevent cellular damage. Haemocyanin (Hc), an invertebrate oxygen transporter, is also involved in homeostatic functions such as osmoregulation, hormone transport and protein storage. As an immune-related protein, Hc has been implicated in phenoloxidase-like activity following pathogenic infection in various molluscan species. Mechanisms involved in activating the proPO cascade aid in Hc conversion to a PO-like enzyme, resulting in the formation of the antimicrobial pigment melanin. In some invertebrates, Hc has been found with an antimicrobial peptide region, implicated it in AMP synthesis (Figure 7). OA stress was also found to impact several metabolic-related proteins. GLUT1 is directly involved in the cellular response to carbon starvation and the regulation of glucose homeostasis through facilitating glucose transport across membranes. Under abiotic stress, ENO1 and PGK evoke anaerobic metabolism to compensate for reduced aerobic metabolism. ENO1 is a glycolytic enzyme involved in both

anaerobic and aerobic ATP synthesis, while PGK functions alongside pyruvate kinase in the glycolytic pathway to generate ATP anaerobically. Although ENO1 and PGK have not been extensively investigated in molluscs, both proteins are highly conserved and are known to play crucial roles in ATP synthesis in eukaryotes.

In biomarker discovery it is important to identify candidates that are not only demonstrably differentially changing under treatment conditions but are also implicated in a stress response following comprehensive functional analysis. As such, the candidates specifically discussed here (section 3.4) have been identified as potential biomarkers of OA stress, and consequently the expression of these proteins in response to OA stress will be validated using an LFQ proteomics approach (Chapter 4).

Chapter 4

Towards the validation of candidate biomarkers of abiotic stress in *Haliotis midae.*

Contents

4.1 Introduction

Whole proteome analysis of an organism through shotgun proteomics (i.e. iTRAQ) typically generates large quantities of data. This is advantageous for biomarker discovery, which entails the detection and identification of biological markers of a specific physiological or diseased state (Depledge and Fossi, 1994; Feng *et al.*, 2004; Ilyin *et al.*, 2004; Sarkar *et al.*, 2006; Martyniuk *et al.*, 2012; Moore *et al.*, 2017). Experimental validation of MS-obtained data is becoming increasingly important, particularly in comparative proteomics studies. Validation serves to confirm the quantitative differences observed in protein levels in cells or tissues sampled from an organism subjected to various treatment conditions are in fact experimentally reproducible. This has led to several failed clinical trials in the medical field, especially since there are often trade-offs in validation experiments, such as resources, time and the possible failure of said experiments. However, experimental validation of quantitative proteomics analyses, particularly in biomarker discovery, buttresses the significance of the data (Handler *et al.*, 2018).

While western blotting has been the most common approach for validation of MS data to date, this method ultimately has pitfalls of its own, especially where antibody performance is concerned (Aebersold *et al.*, 2013; Handler *et al.*, 2018). This is notably evident in trying to obtain species-specific antibodies, specifically in studies that involve non-model organisms; while antibody development itself is an expensive and time-consuming process that dissuades many researchers (Handler *et al.*, 2018). On the other hand, commercially-obtained antibodies oftentimes bind non-specifically in complex protein samples, compromising the validation outcome. Furthermore, sometimes antibodies don't recognise similar homologues or sequence variants occurring from mutations or polymorphisms, and can be hindered in reliably distinguishing between different modification sites on proteins (Liebler and Zimmerman, 2013). This is especially true when working on non-model systems.

Isotopic labelling, such as iTRAQ and SILAC, have proven to be popular, high-throughput technologies capable of reducing variation frequently introduced through sample preparation (Sandin *et al.*, 2015). To this end, isotopic labelling augments the multiplexing capability of LC/MS-MS analysis. However, the drawbacks of this technology include reagent costs and the need for pairwise comparisons between samples, which in turn limits the scope of large studies (Old *et al.*, 2005). Label-free protein quantification (LFQ) methods are proving suitable alternatives, where, for example, LFQ surpasses SILAC's performance in proteome coverage

and dynamic range (Sandin *et al.*, 2015). LFQ allows for both relative and absolute quantitation. Relative quantitation occurs through quantifying peptides using peak area or spectral counting, while absolute quantitation is achievable through sample spiking with synthetic peptide standards (Stevenson *et al.*, 2009). Spectral counting is essentially a measure of the number of matched MS/MS spectra of a protein in a sample (Sandin *et al.*, 2015). Stevenson *et al.* (2009) compared spectral counting and peak integration in an effort to validate applications for seed allergen profiling. Their study found that although both approaches were reproducible in terms of quantitation, bioinformatics analysis of spectral count data proved simpler, highlighting its adaptability to various high-throughput sample preparations (Stevenson *et al.*, 2009).

Due to LFQ's cost-effectiveness and reliability, it is unsurprising that multi-factor studies investigating the effects of environmental change on marine invertebrates are increasingly opting for this proteomics approach. For example, a label-free shotgun proteomics protocol was utilised for investigating the response of *Mytilus edulis* to low salinity and anthropogenic contaminants found in European coastal waters (Campos *et al.*, 2016). Following this approach, the authors were able to identify 2071 proteins, while proteome profiling of 587 highly reproductive proteins were found to implicate salinity as a driving force in the response of *M. edulis* to the environmental contaminant propranolol (Campos *et al.*, 2016). In another study, a shotgun label-free proteomics analysis was conducted to identify the main functional protein groups occurring in haemolymph from *M. edulis*, and at the same time, elucidate the underlying mechanisms of the innate immune system of the mussel (Campos *et al.*, 2015). Through integration of transcriptomic data, their proteomics strategy enabled the identification of 595 haemolymph proteins, of which some were shown to play a role in cellular metabolism, oxidative stress defence, immune response and xenobiotic detoxification (Campos *et al.*, 2015). Timmins-Schiffman *et al.* (2014) employed an integrative approach that included label-free proteomics to characterise the stress response of *Crassostrea gigas* to ocean acidification. The incorporation of a proteomics approach proved useful, highlighting its applicability in studying the environmental stress response of marine invertebrates, which could enable modelling of their stress response to future climatic conditions (Timmins-Schiffman *et al.*, 2014b). It is clear that LFQ is a suitable, cost-effective tool for investigating responses of important marine animals to environmental stress, particularly with regard to multiple stressors.

However, few studies have investigated the interactive effects multiple environmental stressors have on marine invertebrates. Focusing on single-stress studies is unrealistic since

environmental stressors rarely occur in isolation. Furthermore, different environmental stressors elicit different physiological and molecular responses, making it impossible to predict the effects of combined stress exposure from the effects of a single stress (Sokolova, 2013). In combination, various environmental stressors may act additively, synergistically or antagonistically, which could differ from the effects of the stressors in isolation (Griffen *et al.*, 2016). In studies investigating the physicochemical conditions of tidal pools, it was found that during the summer, water temperatures rose due to increased solar radiation and higher ambient air temperatures (Gunderson *et al.*, 2016). Furthermore, increased photosynthetic activity by algae has been observed, which was shown to increase the pH and dissolved O_2 levels of the seawater. The converse occurs at night in summer, highlighting the dynamic interplay between changing physicochemical properties of ocean water (Gunderson *et al.*, 2016). In Puget Sound, a study focused on analysing the relationships between ocean carbonate chemistry (pCO_2), temperature, oxygen and salinity (Reum *et al.*, 2014). The analysis found that ocean water acidity had a strong correlation with temperature and oxygen, but the magnitude and direction of the correlation was dependent on the season and geographical location (Reum *et al.*, 2014). It is clear that physicochemical properties of the ocean are naturally fluctuating and exhibit various relationships with one another. Climate change is likely to unequivocally exacerbate this interplay of abiotic factors, thus understanding the combined effects of stressors on the stress response of invaluable marine organisms, such as abalone and oysters, is imperative to understanding their potential to adapt and survive.

4.1.1 Aims of this chapter

The experimental validation of previously obtained and analysed MS data is important in the fields of comparative proteomics and biomarker discovery. Experimental reproducibility underpins the objective of validation studies; to confirm previous quantitative data. Traditional experimental approaches, such as western blots, have their drawbacks, especially where research on non-model organisms is concerned. Furthermore, with the advancement of proteomics technology, researchers are opting for more sensitive techniques for validation, such as label-free protein quantification using LC-MS/MS. LFQ offers advantages that make it a suitable avenue for not only validating previously obtained MS data, but it offers a platform for more in-depth studies, such as investigating the combined effects of multiple environmental stressors on abalone. This is of particular relevance to biomarker discovery, since abiotic stressors usually occur concomitantly.

Thus, the aims of this chapter are three-fold:

1. To validate the quantitative MS data (Chapter 2) of the potential biomarkers selected in response to reduced seawater pH (Chapter 3, Table 2).
   This was achieved through the use of an identical aquarium-based experimental setup to that described in Chapter 2, but utilising LFQ for validating the selected candidate biomarkers. The data-dependent acquisition (DDA) aspect of the iTRAQ experiments in Chapter 2 eliminates the need for utilising synthetic peptide standards in the LFQ validation, as relative quantification of potential target biomarkers is suitable for the aims of this chapter. Thus, peptide matching and peak intensity calculations were employed for relatively quantifying the abundance of candidate protein biomarkers.

2. To validate candidate biomarkers of acute temperature stress (Calder and Coyne, unpublished).
   Candidate biomarkers of acute temperature stress (Calder and Coyne, unpublished) were validated employing aquarium-based experiments with the inclusion of "elevated temperature only" and "combined stressors" treatments.

3. To determine the specificity of the candidate OA and acute temperature biomarkers in relation to elevated temperature and reduced pH, respectively, as well as to a combination of the two stressors.

   This will allow us to select biomarker candidates that respond specifically to decreased pH, or elevated temperature, or a combination of the stressors for future selected reaction monitoring (SRM) studies of haemolymph samples obtained from local abalone farms. Statistical analyses were conducted on the expression data from all three treatments to identify candidate proteins specific to each stressor.

4.2 Methods and Materials

4.2.1 Label-free protein quantification workflow

The LFQ approach utilised in this study employed the same workflow as the iTRAQ workflow described in Chapter 2 (Sections 2.2.1-2.2.3; 2.2.6-2.2.7.3). Adjustments that were made to the protocol were as follows:

i. The sampling times employed in this experiment included 12, 24 and 72 hours only (Figure 1a). Since the severity of prolonged exposure to the combined stressors on the abalone was unknown, a shorter time course was followed to avoid abalone mortality.

ii. The experimental design utilised in this experiment differed from that in Chapter 2. Three treatments were set-up alongside a control treatment (ambient seawater conditions): pH only (pH 7.15), temperature only (increased from 17 to 25°C), combined stressors (pH 7.15 + elevated temperature adjusted in the same manner as the temperature only treatment). Incremental increases of the water temperature was accomplished by adjusting the water heaters by 5°C per day (2.5°C every 12 hours). Four abalone were sampled per time point and were treated as four-independent biological replicates, resulting in a total of 40 samples (Figure 1a).

iii. The trypsin digested peptides were not labelled with iTRAQ® reagents, nor were the samples pooled, but instead each sample was treated as an individual biological replicate (Figure 1a & b).

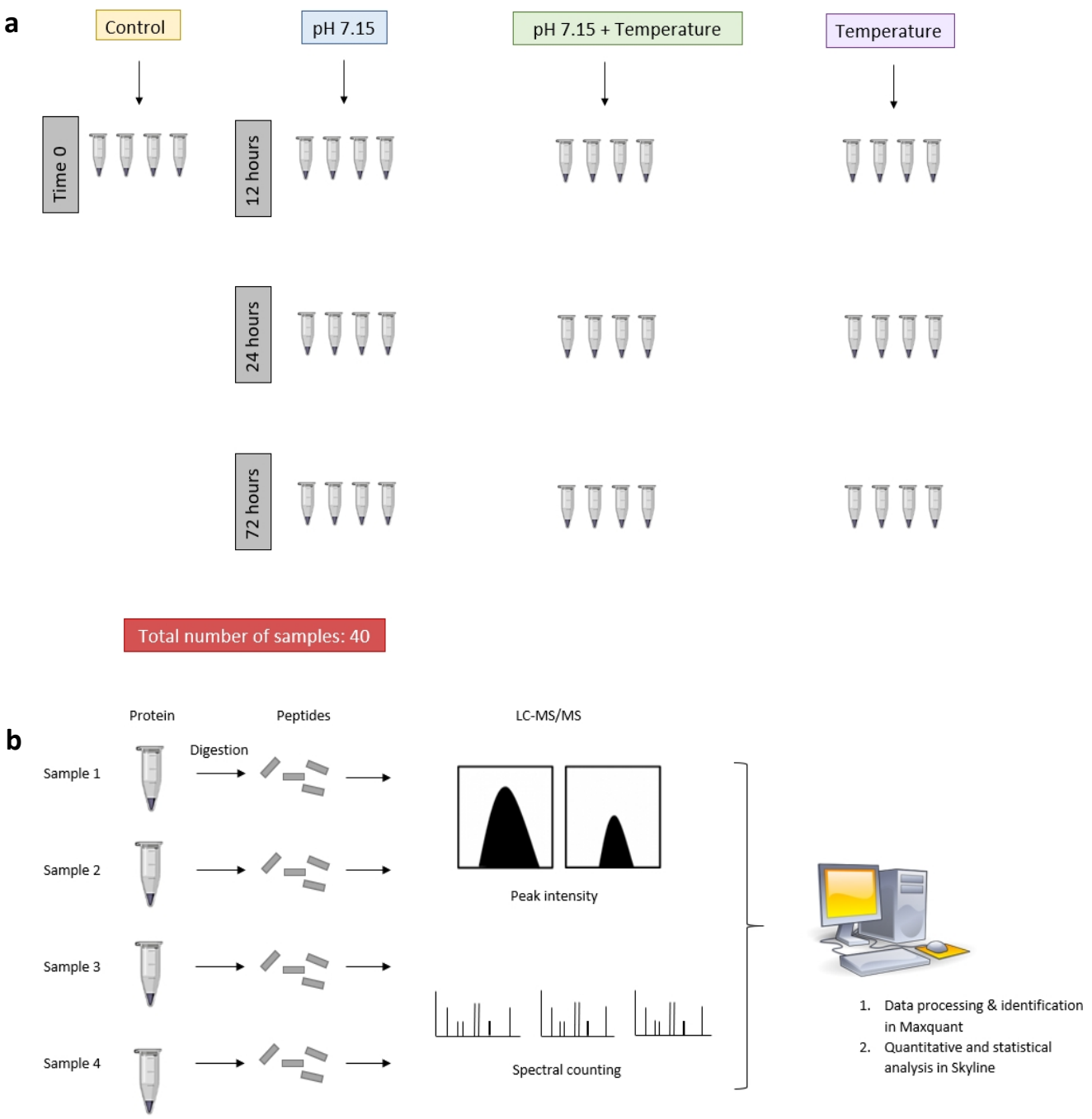

Figure 1. a: The experimental set-up of the aquarium-based stress experiments conducted on *H. midae* in order to validate candidate biomarker proteins. b: The post-sample acquisition workflow for LFQ.

4.2.2 LC-MS/MS analysis

The peptide samples were analysed using a Q-Exactive quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, U.S.A) coupled with a Dionex Ultimate 3000 nano-UPLC system. Data was acquired using Xcalibur version 4.1.31.9, Chromelean version 6.8 (SR13), Orbitrap MS version 2.9 (build 2926) and Thermo Foundations 3.1 (SP4). Peptide samples were resuspended in 0.1% formic acid (Sigma Aldrich) and 2% acetonitrile, and subsequently loaded onto a C18 trap column (PepMap 100, 300 µm × 5 mm × 5 µm). Prior to elution on the analytical column, samples were washed for 3 minutes. Chromatographic separation was performed with a Waters nanoEase (Zenfit) M/Z Peptide CSH C18 column (75 µm × 25 cm × 1.7 µm) as described below. The solvent system employed was solvent A: LC water (Burdick and Jackson BJLC365), 0.1% formic acid and solvent B: acetonitrile, 0.1% formic acid. The multi-step gradient (Table 1) for peptide separation was generated at 300 nL/min as follows: time change 5 min, gradient change: 2 – 5% Solvent B, time change 40 min, gradient change 5 – 18% Solvent B, time change 10 min, gradient change 18 – 30% Solvent B, time change 2 min, gradient change 30 – 80% Solvent B. The gradient was then held at 80% Solvent B for 10 minutes before returning it to 2% Solvent B for 15 minutes. All data acquisition was obtained using Proxeon stainless steel emitters (Thermo Fisher).

Table 1. The multi-step gradient for peptide separation of peptide samples processed from *H. midae* haemocytes following exposure to abiotic stress conditions.

Time (min)	Flow rate (µl/min)	Solvent A (%)	Solvent B (%)
0.00	0.3	98	2
0.00	0.3	98	2
3.00	0.3	98	2
8.00	0.3	95	5
48.00	0.3	82	18
58.00	0.3	70	30
60.00	0.3	20	80
70.00	0.3	20	80
70.10	0.3	98	2
85.00	0.3	98	2

The mass spectrometer was utilised in positive ion mode, with a capillary temperature of 320°C, while the electrospray voltage was 1.95 kV. Data acquisition is detailed below:

Full scan:

Resolution	70.000 (at m/z 200)
AGC target value	3e6
Scan range	320 – 2000 m/z
Maximal injection time (ms)	100

Data-dependent MS/MS:

Inclusion	Off
Resolution	17.500 (at m/z 200)
AGC target value	1e5
Maximal injection time (ms)	50
Loop count	10
Isolation window width (Da)	3
NCE (%)	27

Data-dependent settings:

Underfill ratio (%)	1
Charge exclusion	Unassigned, 1, 7, 8, >8
Peptide match	Preferred
Exclusion isotopes	On
Dynamic exclusion (s)	60

4.2.3 Data analysis and peptide identification using Maxquant

Raw data files (.raw) obtained from the LC-MS/MS of each unlabelled digested sample were processed and peptides identified using Maxquant 1.6.2.10 (Tyanova *et al.*, 2016a). Database search parameters were as follows:

Mass error tolerance:

First search peptide tolerance: 20.0 ppm Isotope mass tolerance: 2.0 ppm
Maximum peptide mass: 4600 Da
Enzyme: Trypsin
Non-specific cleavage: 1
Maximum missed cleavages: 2

PTM:

Fixed: Carbamidomethyl (C) Variable: Oxidation (M); Acetyl (Protein N-term)
Fragmentation mode: HCD
Mass analyzer: FT-ICR/Orbitrap
FDR threshold (%): 1.0

Database: The Uniprot database was accessed on the 8th of August 2018 to retrieve all updated molluscan protein sequences.

4.2.4 LFQ analysis in Skyline

Skyline (Pino *et al.*, 2017) was utilised initially to identify and select suitable peptides (with at least one unique peptide) of the candidate biomarkers, for peptide matching with the LFQ dataset. This was achieved through independent importation of the amino acid sequences of the candidate OA biomarkers (Chapter 3, section 3.3.4, Table 2) and the candidate acute temperature biomarkers (Calder and Coyne, unpublished) from the Uniprot database as .fasta files. The msms.txt file of the LFQ data generated by Maxquant was imported to Skyline for mass spectral analysis and subsequent relative peptide quantification through peak intensity analysis. Default settings were employed except for the following:

Resolving power: 70000 at *m/z*: 200
Precursor mass analyser: Orbitrap
Peaks: 3
Isotope peaks included: Count
Added modifications: Carboxymethyl (N-term)
Oxidation (M)

Candidate proteins matched to the LFQ dataset through peptide matching were further analysed for their response to the three treatment conditions. This was achieved through MS1 peak intensity analysis of target peptides and their peak areas. Statistical analysis (Multiple-hypothesis testing using a Two-way Students T-test with a Benjamini-Hochberg correction) was conducted to identify significantly upregulated proteins ($p \leq 0.05$), as well as those that are of biological importance (fold-change ≥ 1.2), following exposure to each stressor. Venn diagram analysis was conducted using a web-based tool (Bioinformatics Evolutionary Genomics, n.d.). to identify the biomarker candidates specific to each stressor.

4.3 Results

Following the iTRAQ experiments conducted in this study (Chapter 2) and by Calder and Coyne (unpublished), an aquarium-based experiment was conducted over a shorter period with sampling times at 12, 24 and 72 hours.

4.3.1 Haemocyte proteomic profiling of LFQ data

Mass spectrometry analysis and processing was conducted on the LFQ MS spectral data using Maxquant (Table 2). A total of 430,568 MS/MS spectra were submitted for analysis, while only 852 tandem MS spectra were identified and annotated, with a total of 2,276 unique peptide sequences being identified (Table 2).

Table 2. LC-MS/MS data processing summary conducted in Maxquant, following an LFQ-based proteomics approach.

	Total [a]
MS/MS Submitted (PEAK) [b]	430568
MS/MS identified (PEAK) [c]	852
Peptide sequences identified [d]	2276

[a] Sum of all samples; [b] Number of tandem MS spectra submitted for analysis where the precursor ion was detected as a single peak; [c] Total number of identified tandem MS spectra, where the precursor ion was detected as a single peak; [d] Total number of unique peptide amino acid sequences identified from the recorded tandem mass spectra.

4.3.2 Validation of the response of ocean acidification protein biomarkers to decreased pH, elevated temperature and a combination of the two stressors.

Following data processing and statistical analysis in Skyline, we were able to identify 23 of the 33 potential biomarkers included in the LFQ dataset. Of the 23 proteins, only 7 were validated as upregulated candidates in response to OA stress (Figure 2). Serpin B2 (SerpinB2) was not only significantly upregulated ($p \leq 0.05$) but also had an overall fold-change (FC) of 6.3 following exposure to the stressor (Figure 2). Histone H3 (H3) and 78 kDa glucose-regulated protein (GRP78/BiP) both had a fold-change of approximately 2, while phosphoglycerate kinase (PGK) was upregulated by approximately 50% following exposure to reduced pH conditions. TCP-1/CCT6a, IQGAP1 and RPL23a were all moderately upregulated and validated as biomarker candidates of OA stress (Figure 2).

Following exposure of *H. midae* to elevated temperature, 10 of the 23 candidate proteins for OA stress were found to be upregulated (Figure 3). SerpinB2 and major vault protein (MVP) were found to be significantly upregulated compared to expression levels under control conditions. However, the fold-change of SerpinB2 was 3.6 under elevated temperature, which is comparatively lower than the fold-change of this protein in response to reduced pH (Figures 2 & 3). Tubulin beta chain (TUBB4B), 6-phosphogluconate dehydrogenase decarboxylating (PGD), ENO1 and MVP were found to be uniquely upregulated in response to elevated temperature compared to the reduced pH treatment (Figures 2 & 3).

Although 7 of the 23 identified biomarker candidates were upregulated in the combined stress treatment, none were significantly differentially expressed proteins (Figure 4). Interestingly, SerpinB2 is downregulated under combined stress conditions, while radixin (RDX) and mitogen-activated protein kinase (MAPK) were found to be upregulated in the combined stress treatment only, compared to their downregulation observed under pH stress (Figure 2 & 4, Table 3).

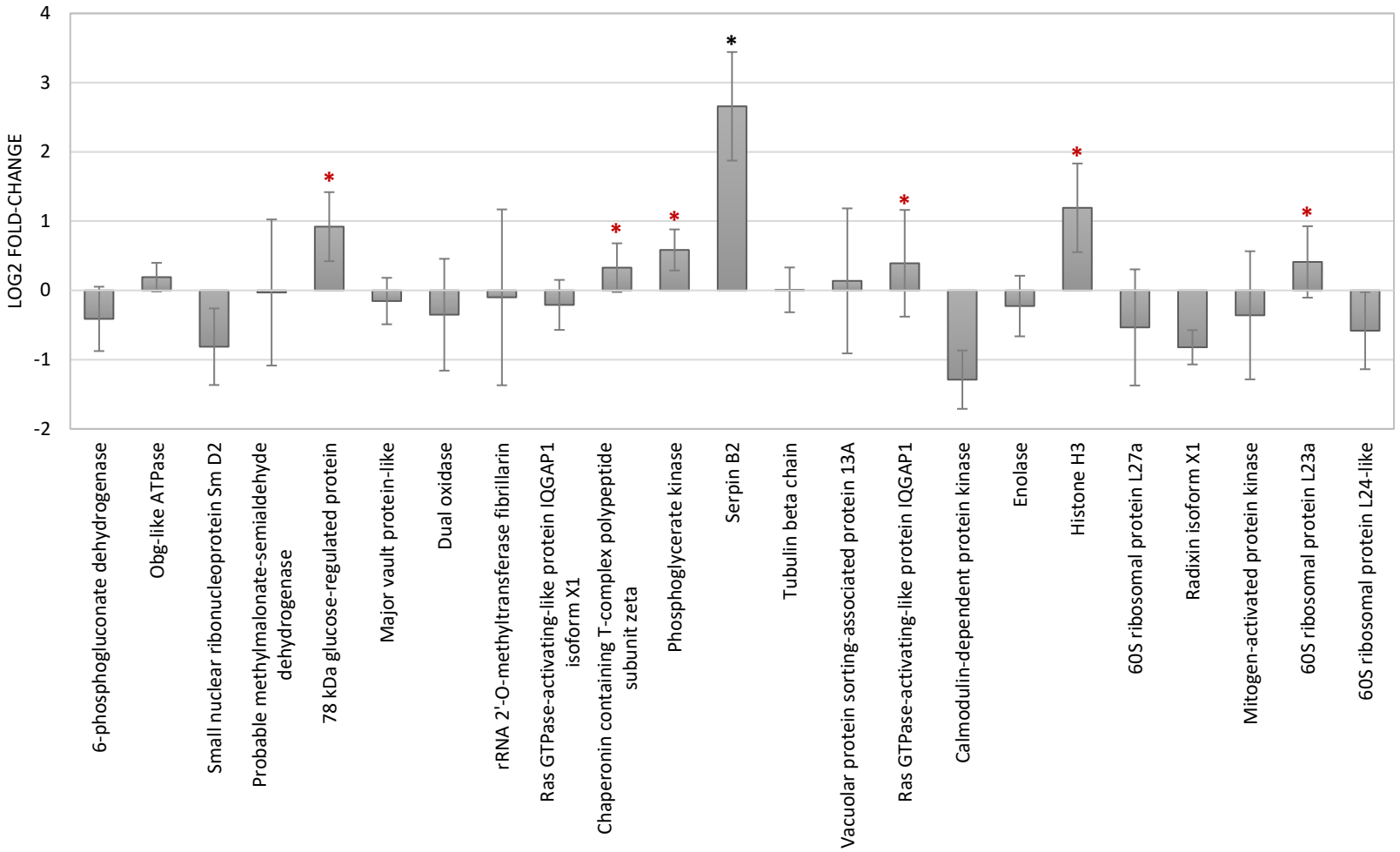

Figure 2. Average relative quantification (Log2 fold-change) of matched candidate proteins in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* exposed to reduced pH conditions. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while the red * indicates proteins with a log2 FC cut-off of 0.33 (fold-change ≥ 1.2). Vertical lines depict the standard error of the mean.

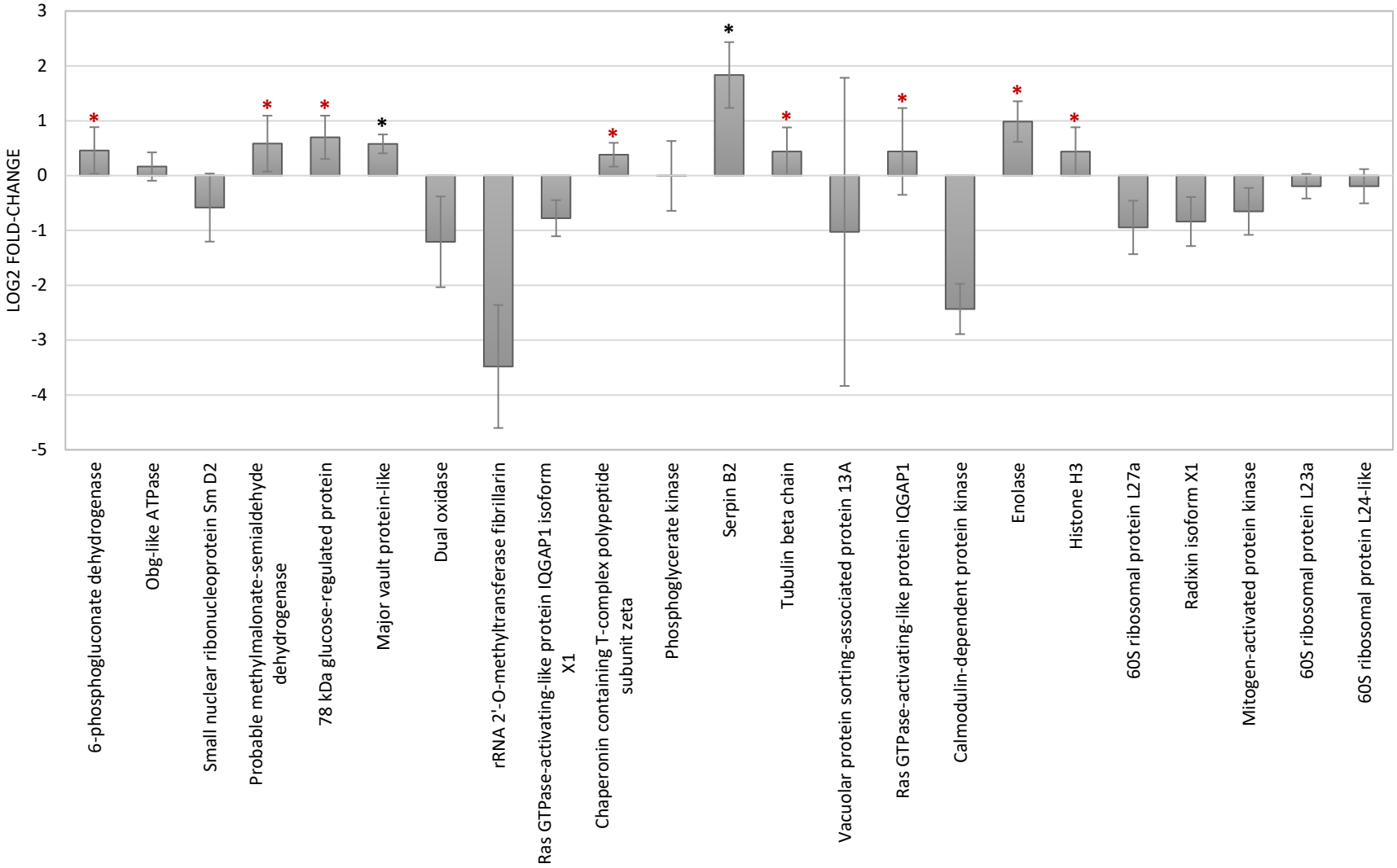

Figure 3. Average relative quantification (Log2 fold-change) of matched candidate proteins in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* exposed to elevated temperature. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while red * indicates proteins with a log2 FC cut-off of 0.33 (fold-change ≥ 1.2). Vertical lines depict the standard error of the mean.

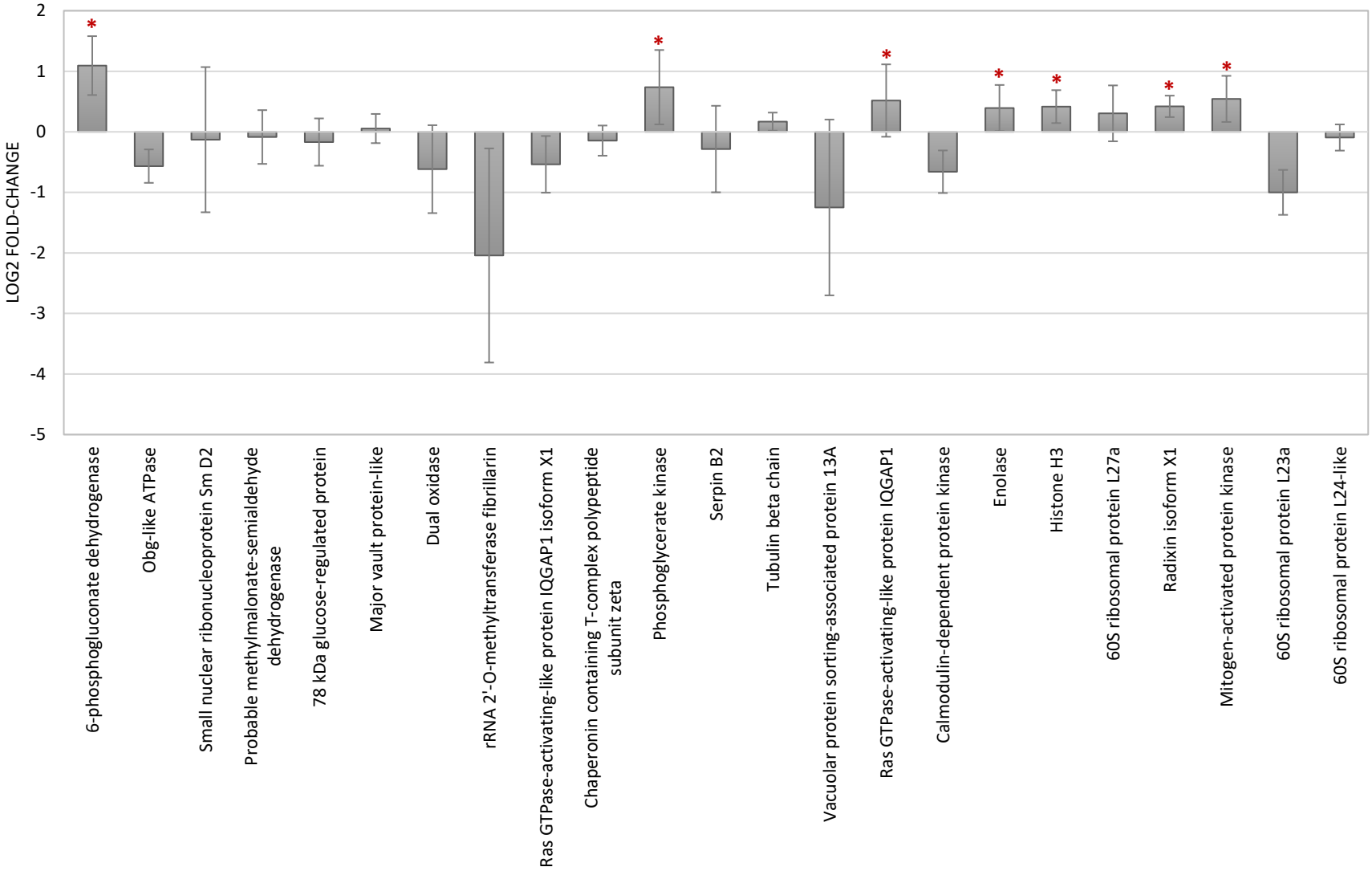

Figure 4. Average relative quantification (Log2 fold-change) of matched candidate proteins in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* simultaneously exposed to elevated temperature and reduced pH. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while red * indicates proteins with a log2 FC cut-off of 0.33 (fold-change ≥ 1.2). Vertical lines depict the standard error of the mean.

4.3.3 Validation of the response of acute temperature protein biomarkers to decreased pH, elevated temperature and a combination of the two stressors.

To validate the 31 proteins identified previously as candidate biomarkers of acute thermal stress in *H. midae* in a study conducted by Calder and Coyne (unpublished), the LFQ dataset generated in the current study was processed and analysed to observe the response of these potential biomarkers to elevated temperature, reduced pH and a combination of the two stressors. There were no significantly upregulated proteins across all treatments (Figures 5-7). Of the 31 candidates, only 10 were found to have a fold-change ≥ 1.2 under reduced pH conditions. Glucose-6-phosphate 1-dehydrogenase (G6PD1) and tropomyosin (TPM1) had a fold-change of approximately 3 and 2.5 respectively, while small G protein, 60 kDa heat shock protein (HSP60) and peptidyl-prolyl cis-trans isomerase were found to have increased 2-fold under the stressor (Figure 5).

Under elevated temperature conditions, 12 of the 31 candidates were validated (Figure 6). Interestingly, although G6PD1 doubled in expression, TPM1 was downregulated under increased water temperature. In contrast to the pH treatment results, the elevated temperature treatment elicited a response from two HSPs: 78 and 84 kDa heat shock proteins, with fold-changes of 1.5 and 2.3 respectively (Figures 5 & 6). Calreticulin (CALR) was validated as a potential biomarker with a fold-change of 1.6, while small G-protein was also upregulated in response to elevated temperature.

Of the 31 candidates identified as potential biomarkers of acute thermal stress by Calder and Coyne (unpublished), 15 proteins had a fold-change ≥ 1.2 in response to simultaneous exposure to the two stressors (Figure 7). Three metabolic proteins, namely 6-phosphogluconate dehydrogenase (PGD), G6PD1 and fructose-bisphosphate aldolase (FBPA), had a fold-change of ≥ 2. A single HSP (84 kDa heat shock protein) was upregulated by approximately 60%, which is comparatively lower than expression of the protein in response to elevated temperature alone. The combined stressors led to upregulation of MAPK, with a fold-change of 1.5, which had not been observed in the other treatments (Figures 5-7).

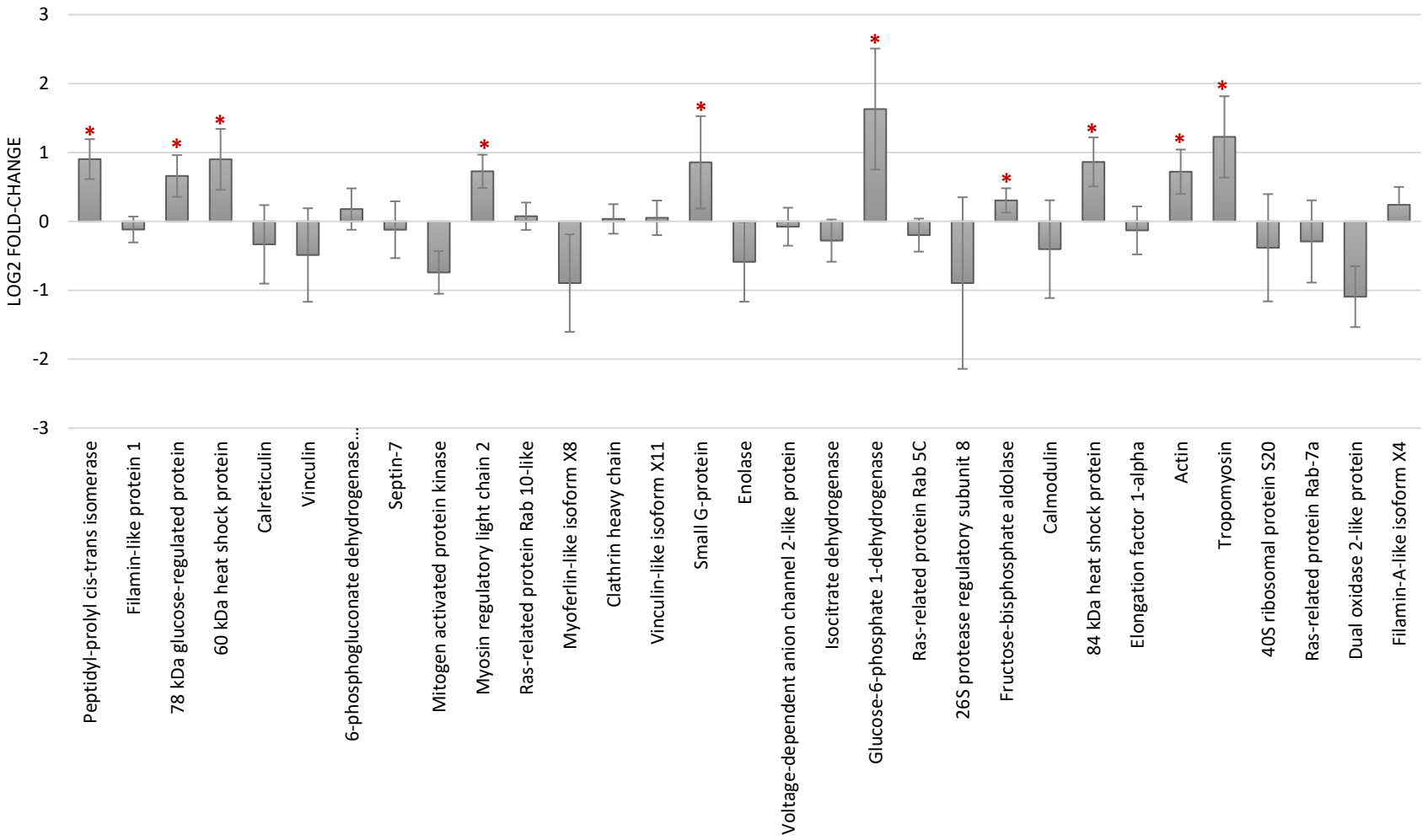

Figure 5. Average relative quantification (Log2 fold-change) of matched candidate proteins (Calder and Coyne, unpublished) in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* exposed to reduced pH conditions. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while red * indicates proteins with a fold-change ≥ 1.2. Vertical lines depict the standard error of the mean.

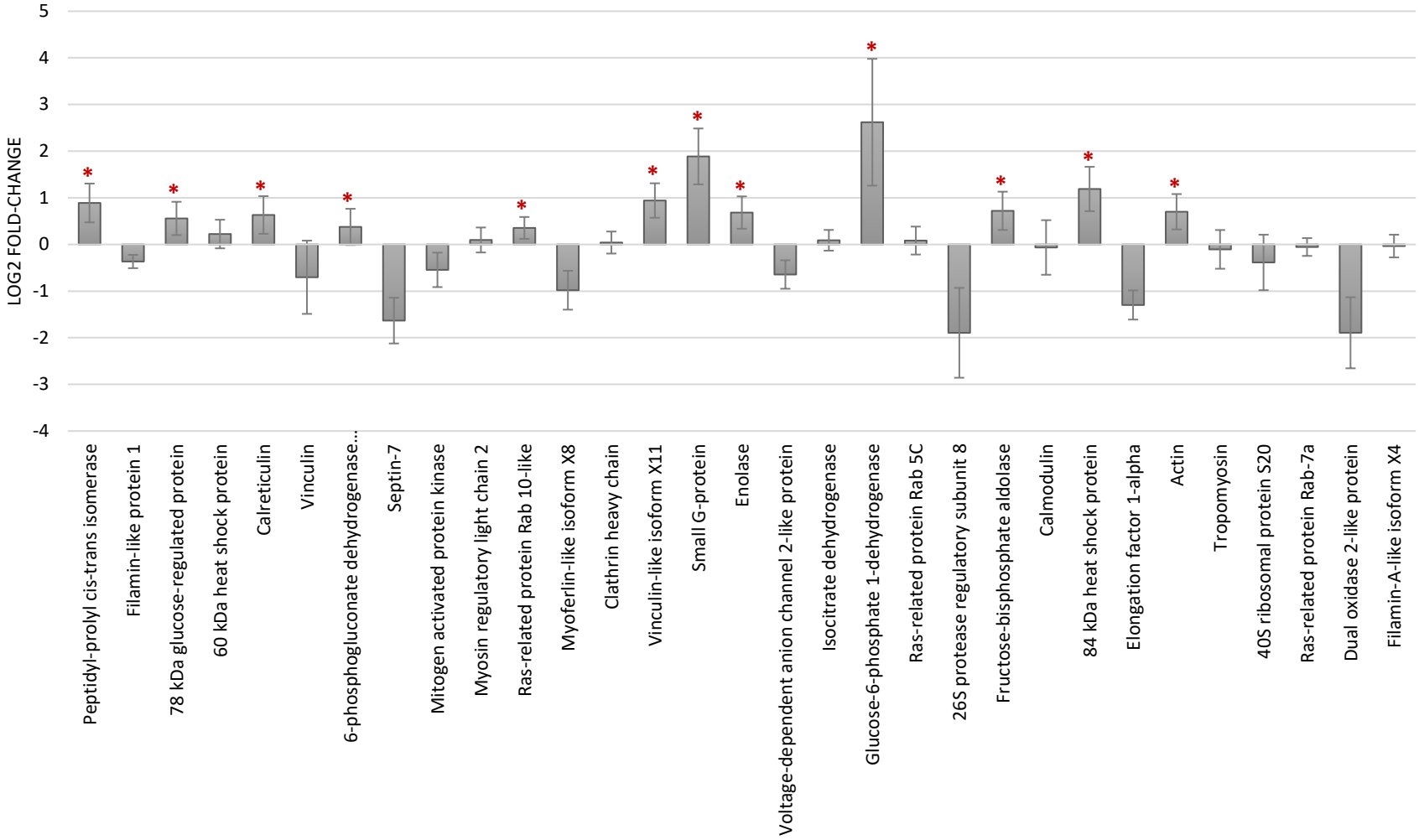

Figure 6. Average relative quantification (Log2 fold-change) of matched candidate proteins (Calder and Coyne, unpublished) in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* exposed to elevated temperature. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while red * indicates proteins with a fold-change ≥ 1.2. Vertical lines depict the standard error of the mean.

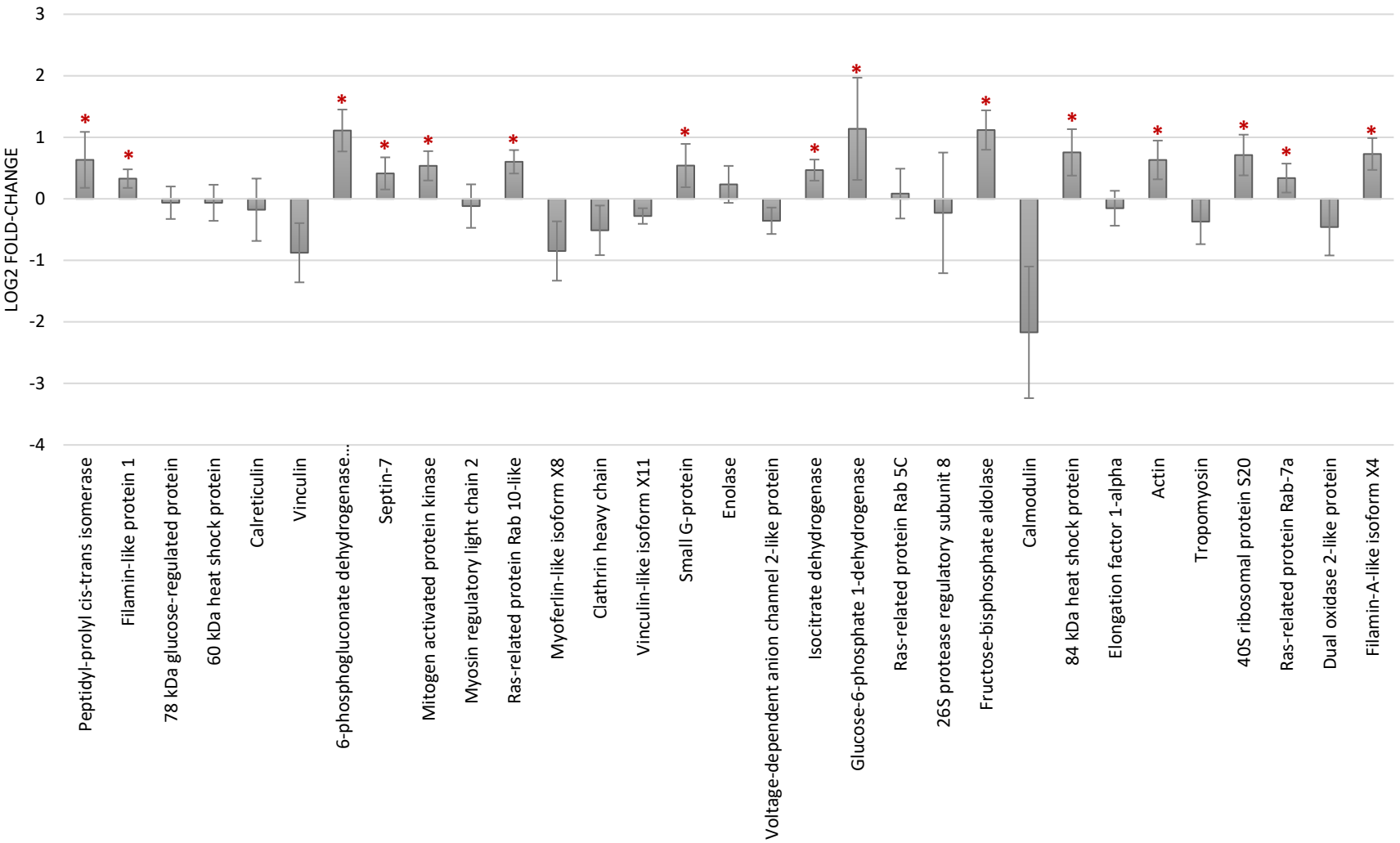

Figure 7. Average relative quantification (Log2 fold-change) of matched candidate proteins (Calder and Coyne, unpublished) in the LFQ dataset following LC-MS/MS analysis of haemocyte proteins obtained from *H. midae* exposed to both stressors. Black * indicates proteins that are significantly upregulated ($p < 0.05$), while red * indicates proteins with a fold-change ≥ 1.2. Vertical lines depict the standard error of the mean.

4.3.4 Candidate biomarkers unique to climate change stressors

To identify upregulated candidate proteins that respond to either pH, temperature or a combination of both stressors, Venn diagram analysis was conducted on proteins that were statistically significant ($p \leq 0.05$) or biologically relevant (fold-change ≥ 1.2) in Figures 2-7. The results are summarised in Table 3. In reconciling the LFQ data and the two candidate biomarker datasets, we were able to refine the biomarker "suite" to a total of 22 upregulated proteins in response to abiotic stress representative of future climate conditions (Table 3). Of the 22 candidates, 7 proteins were found to be unique to OA stress, 8 proteins were unique to temperature stress and 7 proteins were uniquely upregulated under a combination of the stressors. H3 and SerpinB2 (*) were found to be substantially upregulated under OA stress compared to elevated temperature, thus are included as candidates of OA conditions (Table 3).

Table 3. Candidate protein biomarkers identified as unique to either OA stress, elevated temperature or a combination of the stressors, through Venn diagram analysis of upregulated candidates across treatments.

Stressors	Candidate proteins	General function
Ocean acidification	PGK	Metabolism
	RPL23a	Translation
	HSP60	Protein folding
	MYL9	Cytoskeletal structure & protein transport
	TPM1	Cytoskeletal structure & function
	*H3	Chromatin packaging & DNA damage repair
	*SERPINB2	Protease inhibitor & innate immune functions
Elevated temperature	PGD	Metabolism
	ALDOA	Metabolism
	TUBB4B	Cytoskeletal structure & function
	MVP	Signal transduction & innate immune responses
	ENO1	Metabolism
	CALR	Signal transduction & oxidative stress response
	VCL	Cytoskeletal structure & function
	RAB10	Endocytic pathway
Combined	RPS20	Translation
	FLNA	Cytoskeletal organisation & signal transduction
	IDH2	Metabolism
	MAPK	Signal transduction & stress response
	RDX	Cytoskeletal structure & function
	RAB7A	Cytoskeleton-associated involved in ER morphology
	Sep7	Cytoskeletal structure & innate immune responses

4.4 Discussion

Biomarker discovery strongly relies on reproducibility, emphasising the significant role validation experiments have in proteomics-based studies. Several techniques have been employed in validating discovery proteomics datasets, though western blotting has been the most popular approach. Albeit widely used, western blotting has its caveats. As such, more sensitive, affordable and reliable approaches are increasingly sought after. In the present study, LFQ was employed with the aim of validating the candidate protein group members identified in Chapter 3, as well as validating previously identified candidate proteins that responded to acute temperature stress (Calder and Coyne, unpublished). The suitability of the LFQ approach can be found in its affordability, multiplexing capability, as well as its simplicity. Utilising this approach for validation also provided us with the opportunity to conduct a multi-factor study, investigating the response of the candidate biomarker proteins to a combination of reduced pH and elevated temperature conditions.

4.5.1 Haemocyte proteomic profiling using LFQ

There are three main types of LFQ: 1. Spectral counting, where protein abundance is proportional to the number of mapped spectra to a given protein; 2. Data-independent acquisition methods that employs product ion intensity; and 3. MS1 peak intensity, where the peak area of target peptides are used for calculating protein abundance (Shalit *et al.*, 2015). Following identification of 2276 peptide sequences in the LFQ dataset using Maxquant (Table 2), the peak intensities of the candidate proteins were determined using Skyline to quantify their response. Similarly, Xu *et al.* (2019) employed an LFQ-based proteomics approach to investigate the proteome of yellow shell and black shell *Pinctada fucata martensii*. Their comparative transcriptomic and proteomic approach yielded 1684 quantifiable proteins, allowing them to functionally analyse 85 proteins, some of which were found to dictate the yellow shell phenotype (Xu *et al.*, 2019). In a study on selectively bred Sydney rock oysters, LFQ was utilised for identifying candidate protein biomarkers of QX disease resistance (Vaibhav *et al.*, 2018). Through their approach they were able to identify a total of 43 differentially expressed proteins, some of which were found to be involved in cytoskeletal reorganisation, immune response and energy metabolism (Vaibhav *et al.*, 2018). These studies

highlight the usefulness, higher coverage capabilities, and dynamic range of LFQ-based proteomics for biomarker discovery and related studies. Indeed, Sandin *et al.* (2015) highlight the utility of LFQ workflows in their comparative review of label-free LC-MS/MS. Most notably, the growing number of studies supporting the viability of LFQ for large-scale biomarker studies (Sandin *et al.*, 2015) buttresses the usefulness of the technique in the current study.

4.5.2 Response of candidate OA biomarkers to reduced pH

Analysis of the MS data obtained by LFQ MS, validated 23 of the 33 candidate biomarkers previously identified by iTRAQ MS/MS (Chapter 2). Quantitative and statistical analysis validated only 7 of the candidates under OA stress (Figure 2). Among these proteins, SerpinB2 was significantly upregulated ($p \leq 0.05$), with a fold-change of 6.3. Being six times more abundant in the LFQ study, SerpinB2, identified in Chapter 2 (Table 4, section 2.3.3) as upregulated in response to OA stress, could have potential as a biomarker. Clade B serpins, although understudied in molluscs, have been implicated in anti-apoptotic roles and host defence (Bathige *et al.*, 2015).

Validated candidates H3 and GRP78 exhibited a 2-fold increase in protein expression in response to OA, but were not significantly upregulated (Figure 2), suggesting a potential biological significance under stress conditions. H3, a core histone protein, is part of an important octameric complex involved in chromatin packaging and structure maintenance (Nikapitiya and Dorrington, 2013). H3, together with other histone proteins, has been found at the cell surface of immune cells in response to stress, while in humans, increased levels of circulating histones have been observed following infection and inflammation (Chen *et al.*, 2014). García-Giménez *et al.* (2019) report on oxidative post-translational modifications in histones and their roles, where H3 S-glutathionylation is proposed to protect cells against oxidative stress. Furthermore, H3 and its variants have been implicated in DNA damage repair induced by environmental stress (Chen and Jin, 2017). As such, H3 (and GRP78) could serve as a suitable biomarker of OA stress since it has been reported to induce oxidative stress in cells.

PGK and RPL23a were found to be the only candidate proteins uniquely upregulated in response to OA stress (Table 3). Since the function of PGK has been discussed in Chapter 3 (Section 3.4.2 i.), it will not be discussed here. RPL23a is a ribosomal protein which has not

been functionally characterised in molluscs. In humans, RPL23a is localised at the interface of ribosomes, chaperones, a translocon and signal recognition particle (Hu *et al.*, 2014). Ribosomal proteins are involved in an array of functions, including gene transcription regulation, translation and apoptosis (Hu *et al.*, 2014). In a recent study investigating the effects of climate change stressors and two harmful algal species on the bay scallop, *Argopecten irradians*, RPL23a was one of 10 transcripts with the greatest increase in expression in larvae exposed to reduced pH conditions relative to control conditions, with a fold-change of 23.2 (Griffith *et al.*, 2019). Although poorly understood in terms of its function in response to stress, RPL23a appears to exhibit differential expression under pH stress only and so, it is worth examining its role further as a biomarker of OA stress in *H. midae*.

4.5.3 Response of candidate OA biomarkers to elevated temperature

Ten of the 23 candidate protein biomarkers of OA stress were found to be upregulated in response to elevated temperature (Figure 3). Of these ENO1, TUBB4B, PGD and MVP were unique to the elevated temperature treatment. Despite these candidate proteins being initially identified in the iTRAQ experiment (Chapter 2) investigating reduced pH effects, they were not validated as biomarkers of pH stress by LFQ. As such, their role as biomarkers of temperature stress should be investigated further.

Increased abundance of TUBB4B in response to elevated temperature is postulated to be due to thermal stress altering cytoskeletal structure and subsequent tubulin functionality. Additionally, increases in tubulin protein expression could be a result of increased ciliary activity to facilitate changes in metabolism in warmer conditions (Vasquez *et al.*, 2019). In a study investigating the proteomic response of *Mytilus* species to acute heat shock, several tubulin isoforms were upregulated in the 24 and 28°C treatments in *M. galloprovincialis* (Tomanek and Zuzow, 2010). Tomanek and Zuzow (2010) suggested that these changes are likely triggered through thermal stress effects on the stability of cytoskeletal proteins. Furthermore, the observed heat-induced ROS production could be co-effectors of cytoskeleton instability, thus leading to an upregulation of tubulin and actin to overcome this (Tomanek and Zuzow, 2010).

PGD is also induced following oxidative stress induced by excess ROS. Thermal stress has been shown to induce oxidative stress in marine organisms such as *Saccostrea glomerate* (Ertl *et al.*, 2019). Thus, this suggests an indirect effect of thermal stress on PGD. Being involved in

carbohydrate metabolism, PGD is essential for ER structural maintenance and protein secretion (Li *et al.*, 2018). Furthermore, PGD is the second enzyme in the pentose phosphate pathway and, alongside G6PD1, is an important source of NADPH, reducing complexes essential for biosynthesis and antioxidant protection (Storey and Wu, 2013). Thus, with increasing temperatures, oxidative stress is likely induced in *H. midae*, leading to the induction of antioxidant defences and an upregulation of PGD.

MVP is ubiquitous and highly conserved in the cytoplasm of organisms belonging to a number of taxa, forming barrel-like structures that regulate processes such as signal transduction and immune responses (Berger *et al.*, 2009; Das *et al.*, 2016). MVPs bind numerous effectors of stress-induced signalling pathways, such as the EGFR-induced MAPK cascade (Tomanek and Zuzow, 2010). In a study by Tomanek and Zuzow (2010), MVP was upregulated at elevated temperatures in one of the mussel congeners, *M. trossulus*, again highlighting intraspecies variability. Das *et al.* (2016) investigated the role of MVP in modulating growth/survival signalling in human airway smooth muscle cells, induced by oxidative modifications and related to bronchial thermoplasty. Their analyses found that MVP can be glutathionylated following receptor-mediated oxidation, which enables its binding to myosin-9, a cell death co-factor (Das *et al.*, 2016). Binding to its co-factor, MVP inhibited heat-induced cellular death of smooth muscle cells in human airways following bronchial thermoplasty treatment. This suggests a protective role of MVP in mitigating cellular death, which may be a possibility in *H. midae*.

4.5.4 Response of candidate OA biomarkers to a combination of the two stressors

Under combined stress conditions, MAPK and RDX were two of the 23 matched candidate proteins responding to OA stress (Table 3). Both these proteins were downregulated under reduced pH conditions, thus negating their potential as OA stress biomarkers in *H. midae*. Radixin is a cytoskeletal protein member of the ezrin-radixin-moesin (ERM) family, which has been implicated in structural and regulatory functions by serving as membrane-cytoskeletal cross-linkers in actin-rich cellular protrusions such as filopodia, as well as between cell surface receptors and the actin cytoskeleton (Hoeflich and Ikura, 2004; Valderrama *et al.*, 2012). Few studies have investigated the ERM family, specifically radixin, in molluscs, and as such merits further investigation. Following exposure to prolonged environmental stress, radixin was identified as one of three highly upregulated proteins in *S. glomerata* (Melwani *et al.*, 2016).

Environmental conditions and the effects thereof were compared between "high impact" and "low impact" populations of *S. glomerata*, which consisted of sites near sources of chemical input (i.e. agricultural run-off, storm drains) and sites 5 km away from these points. Factors in high-impact sites led to low water pH, reduced dissolved oxygen, and an accretion of harmful chemicals (Melwani *et al.*, 2016). Thus, energetics and the cytoskeleton were adversely impacted in these areas. Interestingly, the effects of a 28-day exposure period to OA conditions on *Acropora millepora*, a reef-building coral, elicited downregulation of radixin (Kaniewska *et al.*, 2012). However, an initial early upregulation of radixin (and related cytoskeletal proteins) was observed in *A. millepora* under OA stress, suggesting changes in interactions between the cell membrane and cytoskeleton over time. The effect of combined stressors on radixin in *H. midae* could be such that, together, they exacerbate oxidative stress likely experienced under single-stress exposure, thus leading to an upregulation in signal transduction and cytoskeletal remodeling to restore cellular homeostasis.

4.5.5 Response of candidate acute temperature biomarkers to reduced pH

Under OA stress conditions, 10 of the 31 previously identified biomarker candidates of acute temperature stress (Calder and Coyne, unpublished) were found to have a fold-change ≥ 1.2 following analysis of the LFQ data, but none were significantly upregulated (Figure 5). Candidates that were unique to OA stress from this group of proteins included a HSP60, myosin regulatory light chain 2 (MYL9) and TPM1 (Table 3).

HSP60 belongs to a large conserved family of molecular chaperones that assist in binding non-functional proteins to functionally alter their conformational states (Zhou *et al.*, 2010). Heat shock proteins (HSPs) are induced under different physiological conditions, such as thermal stress, oxygen deprivation and pathogenic infection (Zhou *et al.*, 2010). HSPs have been implicated in an array of functions from metabolism and growth, to immune responses and programmed cell death. HSP60 has been found to play a role in autoimmune diseases and the innate immune response, playing a critical role in mediating ATP-dependent protein folding (Zhou *et al.*, 2010). A higher abundance of HSP60 under OA stress was observed compared to exposure to elevated seawater temperatures (Figure 5 & 6). In a study investigating the transcriptomic response of *C. gigas* to thermal stress (32°C), several HSPs were substantially upregulated, with HSP60 being one of two HSPs that remained upregulated throughout the study (Lim *et al.*, 2016). Similarly, HSP60 was highly upregulated in *Apostichopus japnoicus*

following thermal stress (26°C) (Xu *et al.*, 2014). These studies support the possibility that HSP60 also plays a prominent role in the heat stress response in *H. midae*. Following pH stress (pH 7.0) for 96 hours, HSP60 and HSP10 mRNA were significantly upregulated in *Penaeus monodon* compared to control conditions, suggesting an involvement in the pH stress response (Shi *et al.*, 2016). Deviations from optimum pH conditions can induce ROS generation, as well as upregulate manganese superoxide dismutase (MnSOD) and other antioxidants (Shi *et al.*, 2016). Since HSP60 is essential for the correct folding and functioning of MnSOD, an increase in MnSOD in *H. midae* exposed to decreased pH is likely to incur an upregulation of HSPD1/HSP60 under reduced pH conditions.

MYL9 is a cytoskeletal protein involved in a number of crucial functions such as facilitating cell division and migration, intracellular protein transport, and maintaining cell shape (Chang *et al.*, 2016). TPM1 is the main allergen in shellfish allergies in humans, while MYL9 has been identified as another, less prominent, allergenic protein in shellfish (Faber *et al.*, 2017). Being considered a "prototype" of myosin, TPM1's dynamic role involves regulating actin filament machinery and function (Parry and Squire, 2017). Wei *et al.* (2015a) conducted a combined metabolomic and proteomic study on mantle tissue of *C. gigas* under OA conditions (pH 7.6). Their analysis found an upregulation of several cytoskeletal proteins, such as tropomyosin and myosin light chain, in response to the stressor. TPM1 and MYL9 have been reported to be concomitantly upregulated in other OA studies in molluscs and other marine organisms exposed to reduced pH conditions, as well as other stressors (Rodríguez-Ortega *et al.*, 2003; Chang *et al.*, 2016; Wäge *et al.*, 2016; Hasanuzzaman *et al.*, 2018). The co-upregulation of TPM1 and MYL9 in these studies and the present study in response to OA stress supports the premise that molluscan cytoskeletal proteins are affected by environmental stress. It is unsurprising that these proteins are co-regulated since they both function as co-regulators of actin-myosin interactions (Wei *et al.*, 2015a, b). Thus, their dynamic interplay could be of importance in mitigating the effects of OA stress on *H. midae* and consequently, their potential role as biomarkers of OA stress should be investigated.

4.5.6 Response of candidate acute temperature biomarkers to elevated temperature

A total of 5 candidate biomarkers were validated following exposure to elevated temperature (Table 3). PGD, CALR, ENO1, vinculin (VCL) and ras-related protein Rab-10-like (RAB10) were validated candidates for heat stress in *H. midae* (Figure 6, Table 3). CALR is a calcium-binding chaperone protein that is localized to the ER, being involved in functions such as maintaining Ca^{2+} homeostasis, protein chaperoning, as well as the temperature and oxidative stress responses (Udayantha *et al.*, 2016; Visudtiphole *et al.*, 2010). CALR has been implicated in the unfolded protein response (UPR) signaling pathway in molluscs, playing a crucial role as an ER chaperone during ER stress (Huang *et al.*, 2018). A thermal stress study on *C. gigas* by Lim *et al.* (2016) reported upregulation of CALR following prolonged exposure to elevated temperatures, suggesting a prominent role in stabilizing the cellular landscape under the stressor. Thus, it is unsurprising that CALR appears to play a prominent role in heat-stressed *H. midae*. VCL, an actin regulator and cell adhesion protein (also implicated in apoptosis-related functions), has previously been found to be modulated by CALR (Opas *et al.*, 1996). Cell adhesion is closely linked to and regulated by Ca^{2+}, where low extracellular Ca^{2+} concentrations impact cell viability. Furthermore, since CALR regulates intracellular Ca^{2+} stores, it can be inferred that CALR, via vinculin, regulates cellular adhesion (Opas *et al.*, 1996). This was confirmed in work carried out on mouse L fibroblasts, where overexpression of CALR led to increased cell-substrate and cell-cell adhesion, as well as vinculin-rich cell-cell junctions (Opas *et al.*, 1996). As such, it is possible that elevated temperatures result in an upregulation of CALR that potentially induces an increase in VCL expression, as observed in *H. midae*.

RAB10 is an ER-localised Rab GTPase that facilitates ER structure maintenance (English and Voeltz, 2013). As far as we are aware, there are no studies investigating RAB10 in molluscs. However, RAB10 has been extensively studied in *Caenorhabditis elegans*, being implicated in the endocytic pathway as a recycling regulator (Chen *et al.*, 2006). Vega *et al.* (2010) found that HSP70 played a major role in regulating the endocytic pathway in the ridgetail white prawn, *Exopalaemon carinicauda*. Their study highlighted an aspect of the stress response that led to an increase in the endocytosis of extracellular macromolecules (Vega *et al.*, 2010). With the exposure of *H. midae* to elevated seawater temperatures in the present study, a stress response was induced via the upregulation of several HSPs, including GRP78 which belongs to the HSP70 family. Consequently, it is possible that this interplay of HSP70s with endocytosis can result in upregulation of RAB10, a key recycler in the pathway. As such, this

might explain the response observed with regard to RAB10 and GRP78 expression in *H. midae* under heat stress in the present study.

4.5.7 Response of candidate acute temperature biomarkers to a combination of the two stressors

The following proteins were found to be solely upregulated in *H. midae* exposed to a combination of elevated temperature and reduced pH: MAPK, RDX, Septin-7 (Sep7), ras-related protein Rab-7a (RAB7A), filamin (FLNA), isocitrate dehydrogenase (IDH2) and 40S ribosomal protein S20 (RPS20) (Table 3).

Filamins are a group of actin-binding proteins, being implicated in cross-linking F-actin filaments, and thus functioning as central cytoskeleton organisers (Méndez-López *et al.*, 2012). They have also been associated with membrane receptor proteins, signal transduction proteins and transcription factors (Méndez-López *et al.*, 2012). In a study investigating the effect of pathogenic infection on *Octopus vulgaris* using a proteomic approach, filamin was found to be upregulated in the infected cohort (Castellanos-Martínez *et al.*, 2014). Filamin is reported to be an active member of signaling cascades that induce NF-κB as part of an immune response to pathogenic infection (Castellanos-Martínez *et al.*, 2014). Similar responses in filamin expression were observed in *Ostrea edulis* following pathogenic challenge (Morga *et al.*, 2011). This suggested an increase in cytoskeletal polymerization leading to an increase in pathogenic engulfment (Morga *et al.*, 2011). Filamin has also been shown to be upregulated in response to oxidative stress induced by pathogenic infection of the surf clam *Mesodesma donacium* (Maldonado-Aguayo *et al.*, 2014). Under high stress conditions, filamin dissociates from Rac activation, leading to actin polymerization (Diz-Muñoz *et al.*, 2013). The upregulation of filamin in *H. midae* might be in response to oxidative stress induced by the combined stress, which could be considered “high stress conditions”. Thus, upregulation of cytoskeletal proteins such as FLNA under combined stress conditions potentially aids in maintaining cell integrity.

RAB7A is a member of the small G proteins and has been identified as a key player in regulating intracellular vesicular trafficking to the late endosome (Boutet *et al.*, 2009). Specifically, RAB7A in mammals regulates transport from early endosomes to lysosomes (Guerra and Bucci, 2016). Mateus *et al.* (2018) provide evidence that supports the theory that RAB7A is essential in regulating ER morphology through the maintenance of cellular

homeostasis during ER stress. However, RAB7A has yet to be studied in molluscs. It can be postulated that the upregulation of RAB7A in *H. midae* under combined stress conditions is borne out of the need to maintain ER morphology following ER stress.

Sep7 is part of a family of septins that are ubiquitous GTP-binding proteins and are considered scaffold proteins for structural integrity of the cytoskeleton (Kinoshita, 2006). Septin was identified as an immune effector in the transcriptomic analysis of *S. glomerata* immunity (Ertl *et al.*, 2016a), thus implicating its potential role in the immune response upon pathogenic infection. Similarly, septin was among the upregulated immune-related genes in *Mya arenaria* in response to *Vibrio splendidus* challenge (Araya *et al.*, 2010). Under OA conditions, septin was downregulated in the Sydney rock oyster (Goncalves *et al.*, 2017), but coupled to elevated temperature might induce cytoskeletal reorganization in *H. midae* in order to maintain cellular homeostasis under oxidative stress conditions.

4.6 Conclusion

LFQ served as a suitable avenue for pursuing the validation of candidate biomarkers of OA stress in *H. midae*, as well as validating candidate biomarker proteins of acute temperature stress previously identified by iTRAQ MS/MS by Calder and Coyne (unpublished). A total of 5 potential protein biomarkers of OA stress and 10 candidate protein biomarkers of temperature stress were validated as specific stress biomarkers in *H. midae*. A total of 7 proteins were found to be upregulated in response to a combination of the stressors. Energetics and the cytoskeleton appear to be the most affected components of *H. midae* haemocytes when the animal is exposed to high stress conditions, likely suggesting that proteins associated with these cellular processes play an essential role in the stress response of *H. midae*.

Chapter 5

Final discussion

Haliotis midae farming is an economically important sector of the South African aquaculture industry. However, with the increasing global population, intensive anthropogenic activities have contributed significantly to the current climate crisis through the accretion of greenhouse gases such as CO_2. Climate change is not only leading to increased temperature differentials across the globe, but also changes in other environmental physico-chemical properties. For example, ocean acidification (OA) is of global concern, with changes in pH having far-reaching consequences for marine life and community structures. As such, physico-chemical parameters on abalone farms in South Africa might be altered under future climatic conditions, resulting in increased stress and consequently, the health of an important commodity. Since *H. midae* does not show distinct physiological responses to pH stress, mitigating abalone losses could prove difficult as the detection of animal stress might be too late. Furthermore, abiotic stress has been reported to cause immunosuppression in molluscs, increasing susceptibility to pathogenic infection. However, little is known about the abalone stress response, particularly following exposure to pH conditions representative of predicted climate conditions for the year 2100. Consequently, in order to find solutions that mitigate stress on abalone farms, it is crucial to understand the underlying mechanisms that characterize the stress response of *H. midae* under reduced pH conditions. In doing so, biomarkers of stress can be identified and potentially utilized in a health monitoring program that could be implemented on commercial farms in the country. Thus, this study sought to accomplish the following aims (discussed in more detail in Chapter 1):

1. To quantitatively characterize the haemocyte proteome of *H. midae* exposed to reduced pH conditions, utilising a shotgun proteomics approach with isobaric tagging for relative and absolute quantification (iTRAQ), coupled to LC-MS/MS analysis.
2. To determine the functional response of the *H. midae* haemocyte proteome to OA conditions, employing an array of bioinformatics tools and workflows that, in combination, would aid in the analysis of proteomics data derived from a non-model organism. Additionally, by incorporating statistical analyses of the data, we sought to

identify candidate biomarker proteins of OA stress following a functional analysis of the upregulated proteins.

3. To further investigate and validate candidate biomarker proteins of OA stress, while also validating previously identified candidate protein biomarkers of acute temperature stress (Calder and Coyne, unpublished). Furthermore, we sought to determine how the candidate biomarker proteins respond when *H. midae* is exposed to a combination of elevated temperature and reduced pH.

As far as we know, this is the only comprehensive study investigating the *H. midae* haemocyte proteome response to predicted future climatic conditions, with a focus on ocean acidification (OA). The main empirical and functional findings of this study are presented in detail in Chapters 2-4. The outcomes of this study provide a holistic proteomics analysis of the haemocyte proteome of *H. midae* exposed to reduced pH conditions. The aim of this final chapter is to not only provide a summation of the data generated and to highlight key findings, but to also acknowledge the caveats and limitations of aspects of this study, while also presenting what I believe is worth focusing on in future studies and investigations. As such, I will progress accordingly through each of the three main aims of this study.

5.1 Quantitative characterisation of the haemocyte proteome using iTRAQ and LC-MS/MS

The use of -omics technologies to comprehensively answer complex questions about organisms and their environment has grown in popularity since the 1970s, especially with the enhanced development and availability of an array of techniques and reagents. Unlike gel-based approaches, shotgun proteomics has superior data efficiency, as well as allowing for the identification of a greater number of proteins from a single sample. The shotgun isobaric labelling approach employed in this study was suitable for the milieu and desired outcomes of this chapter. Protein labelling strategies have become increasingly popular as MS technologies and machinery have evolved. A simple Google Scholar search for iTRAQ generated nearly 2500 results for 2019 alone, underscoring the value and popularity of iTRAQ in a plethora of studies ranging from whole proteome analysis of Chinese Wild Rice, to identifying key proteins for diagnosing coronary heart disease in humans. The use of iTRAQ in Chapter 2 was

demonstrably invaluable to our comparative stress study, since it couples peptide identification with relative or absolute quantitation, as well as allowing the option of multiplexing up to 8 independent samples.

Phagocytosis activity utilising heat-killed *V. anguillarum* was measured to assess the physiological response of *H. midae* haemocytes following exposure to OA stress. Interestingly, contrary to most studies investigating the effect of environmental change on haemocytes, phagocytic activity was initially boosted under the stressor, followed by a decline in activity after prolonged exposure to OA conditions. This could be attributed to the other roles phagocytosis reportedly plays, which is principally the removal and recycling of circulating dead haemocytes following cellular stress induced by OA conditions. As such, this might have extensive consequences under changing climatic conditions, especially since abiotic stress has been linked to immunosuppression in other molluscs. Immunosuppression can lead to infection and, compounded with OA stress, could put a strain on the energetics of abalone due to the necessity of having to not only mitigate the effects of OA stress by clearing dead haemocytes, but also through energetically costly activities such as haematopoiesis for increasing the number of circulating haemocytes for the phagocytosis of invading pathogens. This could affect the sustainability of South African abalone farms as climate change worsens. Intrinsically, this study highlights the importance of elucidating the physiological, as well as the molecular, response of abalone under environmental change, since some responses are not only species-specific, but may also not only be demonstrably evident. Thus, future studies related to the effect of ocean acidification on *H. midae* could build on the current study by examining other parameters such as the number of circulating dead haemocytes, determining the phagocytic index and measuring oxidative stress indices.

Compared to the 100 differentially expressed *H. midae* haemocyte proteins identified in response to pathogenic challenge utilising similar technologies (Beltran, 2015), this study highlights the growing dynamic range of iTRAQ LC-MS/MS with identification of a total of 227 differentially expressed proteins, particularly as more sequence data is incorporated in public databases. One can only expect this trajectory to continue, especially since more studies are opting for -omics technologies to answer complex scientific questions, and MS-related technologies and workflows are analogously advancing. This is particularly salient for biomarker discovery studies. Dalman *et al.* (2012) point towards the dangers of solely relying on statistical significance in the biological interpretation of quantitative data in microarray and -omics studies, suggesting that utilising a combination of fold-change and statistical

significance proves more reliable. Thus, proteins of statistical ($p \leq 0.05$) and biological significance (fold-change ≥ 1.2) were identified throughout the time course of this study. In doing so, we were able to satisfy the first aim of this study, while laying the foundation for biomarker discovery of OA stress in Chapter 3. These proteins were found to be associated with an array of cellular processes; most notably energetics and metabolism, cytoskeletal-related functions, signal transduction and stress/immune-related responses. Several similar studies conducted on other marine invertebrates appear to support the findings of the present study and are noted in Chapter 2 (Section 2.3). This bolsters the current knowledge surrounding the effects of climate change on marine molluscs, but also sheds light on the previously unknown response of *H. midae* to OA stress, and so provides a solid foundation for directing further studies on specific cellular processes *in vitro*. In its entirety, the data generated in this chapter also contributes to the growing body of *H. midae* MS spectral data, as well as stressing the value of utilising an iTRAQ-labelling shotgun proteomics approach.

However, some limitations were experienced in satiating the first aim:

i. Limited resource availability and financial constraints allowed for only four independent biological replicate studies. The experimental set-up and subsequent analysis could benefit from the inclusion of more biological replicates, which would improve statistical analyses, while generating more genetic data and enhancing the study's scope.

ii. Due to financial constraints, single MS injections were conducted in this study. Gan *et al.* (2007) assessed the effect of multiple injections on the reproducibility of iTRAQ and found that applying three LC-MS/MS injections reduced MS variation significantly. Thus, by including multiple MS injections in future studies, peptide identification and quantification can be enhanced, while also improving the depth of iTRAQ applications.

iii. Stringent filtering employed during the iTRAQ analysis led to a large number of proteins, which could be of biological importance, not being retained and identified. Since iTRAQ analyses conducted on non-model organisms should be approached with caution, we opted for a conservative filtering approach to ensure greater confidence in the identification process. Additionally, due to the limited abalone genetic data available, a number of peptide spectra could not be identified. As more sequence data is generated and made publicly available, the data in the present study should be re-analysed to characterise the unidentified peptide spectra.

Subsequently, together with the addition of the omitted proteins, this additional data could be investigated for the functional roles played by these haemocyte proteins in response to reduced pH conditions, which could bolster or shed more light on the current results. This underpins the importance of carrying out -omics studies: to not only investigate hypotheses of interest, but to contribute to the growing body of genetic data available to the scientific community. Furthermore, due to the predominantly qualitative nature of shotgun proteomics, supplementary technologies, which have been extensively explored and reviewed, need to be considered for quantitation in the future (Martyniuk *et al.*, 2012; Aggarwal *et al.*, 2019; Gao and Yates 2019; Pappireddi *et al.*, 2019; Wang and Kuruc 2019; Corrales and Odriozola 2020).

In essence, iTRAQ and LC-MS/MS served as reliable and reproducible tools for relatively quantifying and identifying proteins comprising the haemocyte proteome of *H. midae* following OA stress. Although there are caveats to label and label-free shotgun proteomics approaches, for the purposes of this study and the aims set out in Chapter 2, the use of an isobaric labelling approach was found suitable for whole haemocyte proteome analysis of a non-model organism.

5.2 Functional characterisation of the haemocyte proteome utilising bioinformatics

Albeit an important aspect of proteomics studies, quantitative analysis of proteins does not provide information on the molecular mechanisms at play in an organism under a particular environmental condition. To satisfy the second aim of this study, a range of bioinformatics tools were employed to cluster and functionally analyse the identified proteins, elucidating the molecular stress response of *H. midae* to OA conditions. Furthermore, in reconciling the quantitative and functional results we were able to refine our analysis by focusing on specific proteins (33), described in detail in Chapter 3, that could potentially serve as biomarkers of OA stress in *H. midae*. Functional annotation and network analysis implicated several key proteins in cellular processes such as response to carbohydrate starvation, cytoskeleton stabilization and re-organisation, signal transduction and the stress/immune response. In combination with other molluscan studies, we were able to shed further light on, as well as substantiate, the proposed

complex interplay of the immune and stress response of *H. midae* under OA conditions. As such, Chapter 3 highlights the importance of employing a comprehensive range of bioinformatics tools to holistically analyse the functional response of *H. midae* haemocyte proteins to OA stress, further emboldening the crux of functional proteomics analyses.

The employment of WGCNA analysis facilitated the analysis of groups of proteins that had statistically similar expression patterns, serving as a foundation upon which further inspection of induced protein groups can be conducted. Functional analysis of the Predominantly Downregulated eigengene modules suggest a predominant aerobic metabolic depression and downregulation of proteins related to translation, growth and development under OA stress. Conversely, functional analysis of the Predominantly Upregulated modules implicates the induction of oxidoreductase activity, intracellular transport and signaling, translation and cytoskeletal organisation as the predominating roles of proteins comprising these modules. This suggests a potential depression of energetically costly activities and a metabolic shift under OA stress, with a diversion of anaerobically-generated energy to important transport and signaling functions, as well as the maintenance of cellular homeostasis through oxidative stress responses and cytoskeletal stabilization. In other molluscan studies, environmental stress has been found to disrupt energy metabolism in a way that correlates with immunosuppression. As such, the potential metabolic depression and subsequent metabolic shift observed in *H. midae* are likely to have long-term consequences for adaptation and survivability under future climatic conditions. Cao *et al.* (2018) evaluated antioxidative responses, lipid peroxidation, metabolic gene expression, glycolytic enzyme activity and the composition of energy content of *C. gigas* subjected to OA conditions to determine the risk of pathogenic infection. The results of their study provided strong evidence for energy modulation and immunosuppression under OA stress, and as such, are crucial parameters that I believe require further exploration in *H. midae* to confirm this hypothesis.

Notable reference should be made to the proposed link between the stress and immune responses observed in *H. midae*. Indeed, the induction of functionally divergent immune-related genes during abiotic stress or biotic defense could have dual functions in overlapping pathways (Zhang *et al.*, 2015), and are of particular interest in the present study. Malham *et al.* (2003) provide evidence for this dynamic interplay between the stress and immune response in *H. tuberculata* exposed to mechanical stress. In the current study, this is also evidenced by the upregulation of proteins such as mitogen-activated protein kinase (MAPK), calmodulin-dependent protein kinase, serpin B2 and haemocyanin, which have been found to perform

crucial roles in the stress response *and* innate immune system of a number of molluscs. Although the functional roles of these proteins have been putatively elucidated through network modelling and comparative species analysis of biological processes, their exact roles in *H. midae* haemocytes are worth further elucidation and experimentation in the laboratory. Furthermore, observing changes in norepinephrine (NE) might further support the intertwined nature of the stress and immune response of *H. midae*, since tyrosine derivatives are associated with acute stress responses across taxa (Adamo, 2008). NE mediates a number of stress-related responses, including the modulation of immune function, which ultimately serves as an adaptive function under stress conditions, preparing the animal for potential immunosuppression (Adamo 2008). In conjunction with our findings, this might provide a more informed indication of the role the stress *and* immune responses of *H. midae* will have under potentially immunosuppressive future OA conditions.

Consolidation of the quantitative (Chapter 2) and qualitative (Chapter 3) aspects of this study facilitated the identification and protein-protein network analysis of several candidate biomarker proteins of OA stress, providing several avenues for future studies, especially in the context of stress-induced immunity in *H. midae*. It was unsurprising that a number of metabolic-related proteins were found to comprise this group of candidate proteins. Being mostly observed in vertebrate studies, the upregulation of enolase (ENO1) in *H. midae* exposed to OA stress is novel. ENO1 has previously been implicated as a thermal stress protein in *D. salina* (Ruan *et al.*, 2009), suggesting a possibly important role in the metabolic shift observed in *H. midae* to maintain homeostasis under the stressor. Little information exists regarding the molecular mechanism of action of ENO1 in marine invertebrates, and as such, it presents a potential avenue for functional characterisation *in vitro*. Ras GTPase activating protein IQGAP1, another protein extensively studied in vertebrate immune systems, is a key scaffold protein in cell signalling and the actin cytoskeleton. Since IQGAP1 is affected by changes in intracellular Ca^{2+}, its precise role in *H. midae* haemocytes under OA stress is of particular importance and should be elucidated *in vitro*. This would not only contribute to understanding its role in OA stress, but will expand our current knowledge regarding the *H. midae* stress and immune responses since IQGAP1 has not been investigated in marine invertebrates to date. In vertebrates, IQGAP1 has been implicated in mediating MAPK activity. MAPK1 upregulation in *H. midae* haemocytes in response to OA stress confirms the potential link between IQGAP1 and MAPK1 in the stress response under OA conditions. As a vital signal transduction protein, MAPK1 and its relatives mediate innumerable functions such as

inflammation, ER stress and apoptosis, via the ERK/MAPK cascade. This cascade is also affected by changes in environmental Ca^{2+}, which is of importance with regard to the framework of this study. Fluctuations in Ca^{2+} can adversely affect protein folding, which usually culminates in ER stress and the subsequent mediation of the Unfolded Protein Response (UPR) to restore ER homeostasis. In light of OA stress in *H. midae*, it is likely that MAPK and related proteins mediate functions that serve to restore cellular homeostasis. This putative function requires further validation from *in vitro* studies since it has not been extensively studied in molluscs. Furthermore, the role of MAPK1 in OA-stress-induced immunosuppression of *H. midae* exposed to pathogens should be further investigated, as it has been previously implicated in the immune response of infected *H. discus discus*. The identification of a clade B serine protease inhibitor (SerpinB2) in *H. midae* is particularly interesting as the protein was previously identified in the disk abalone in response to infection and associated tissue injury. This highlights its putative role in the *H. midae* immune and stress responses, especially since it exhibits anti-apoptotic activity. Thus, under OA stress, it is likely that SerpinB2 functions in preventing the accumulation of cytotoxic proteases that would otherwise result in cell death. In essence, SerpinB2 could serve as a cellular component responsible for maintaining homeostasis, particularly under stressful conditions. This putative role should be further explored in *in vitro* experiments to confirm this hypothesis.

The functional analysis component of this proteomics study, although largely based on sequence homology and inferences, underpins the importance of utilising a range of bioinformatics tools when working on non-model organisms. Furthermore, it satisfies the two-fold second aim of this study by not only functionally characterizing the stress response of *H. midae* to OA stress, but also by providing a foundation for identifying candidate biomarker proteins of OA stress for further validation (Chapter 4).

Challenges faced in satisfying the aims of this chapter included:

i. Data handling and analysis proved challenging since abalone are non-model organisms and as such, there was limited suitable genetic data available to functionally characterize the stress response of *H. midae* with a high level of confidence. Thus, numerous tools were employed to facilitate an in-depth analysis of the proteomics data to the best of our ability. However, this could be overcome in future studies through the integration of some -omics platforms and data to provide an exhaustive analysis of molecular mechanisms and pathways in *H. midae*

under environmental conditions. Indeed, successful integration of the 3 main layers of -omics datasets (transcriptomics, proteomics and metabolomics) can reveal the complex interplay of these layers, as well as the vital dynamic biological processes and networks involved in the stress-immune response relationship in abalone (Buescher and Driggers, 2016; Misra *et al.*, 2019).

ii. The protein-protein interaction network (Chapter 3, Figure 5) generated for the 33 candidate protein biomarkers using the STRING database, included predicted interactors (grey nodes) which could be investigated in the future for their potential role as biomarkers of OA. This would provide further insight into the response of these interactors in relation to the 33 candidates under OA stress, enhancing our knowledge of the *H. midae* stress response.

5.3 Validation of candidate biomarker proteins of OA stress and acute temperature stress utilising LFQ.

To validate candidate biomarker proteins identified in Chapter 3, as well as those previously identified in response to acute temperature stress (Calder and Coyne, unpublished), another aquarium-based experimental set-up was employed with adult abalone. Three treatment conditions were tested and digested peptide samples from *H. midae* haemocytes were subsequently analysed using a LFQ LC-MS/MS approach. Furthermore, we sought to observe the relative quantitative response of the candidate proteins following exposure of *H. midae* to a combination of stress conditions (reduced pH + elevated temperature). Chapter 4 details the outcomes of this experiment.

As with any biomarker discovery approach, validation is a crucial step in the proteomics pipeline. For all intents and purposes, validation of proteomics data serves to confirm that the quantitative differences observed in cell types under various treatment conditions are in fact experimentally reproducible. This is particularly important following MS-based data acquisition, as MS technologies can have their own pitfalls. In the context of this study, validation was especially necessary since only single iTRAQ MS injections were carried out in Chapter 2. The decision to utilize an LFQ proteomics approach for validation purposes satisfied the aims of this chapter. This decision is bolstered by several biomarker studies that have employed LFQ for validating candidate immune proteins in marine invertebrates, highlighting its growing popularity and utility as an alternative and cost-effective technique. A

total of 2276 peptide sequences were identified, allowing us to validate our candidate biomarker proteins.

A total of 5 potential candidate biomarkers of OA stress and 10 candidate protein biomarkers of acute temperature stress were detected and validated in the LFQ dataset. For the combined stress condition, a total of 7 upregulated candidate biomarkers were detected. Although detected under acute temperature stress, SerpinB2 and H3 are candidate proteins worth investigating as biomarkers of OA stress as their fold-changes were substantially higher in response to OA compared to the temperature treatment. Additionally, the functions of these proteins should be further investigated *in vitro* to elucidate their mechanisms of action under stress conditions. Interestingly, myosin regulatory light chain 2 (MYL9), tropomyosin 1 (TPM1) and 60 kDa heat shock protein (HSP60) (Calder and Coyne, unpublished) were unique to the OA stress treatment, which could be explained by the fact that temperature stress in this study was chronic, while Calder and Coyne (unpublished) tested the effects of acute temperature stress on *H. midae*. TPM1 and MYL9 have been reported to be concomitantly upregulated in other marine invertebrates under OA stress. Once again this supports the notion that the cytoskeleton undergoes structural changes under environmental stress likely due to changes in intracellular Ca^{2+}, all in an effort to maintain cellular homeostasis through cytoskeletal re-organisation and stabilization.

The experimental limitations of this chapter relate to the low number of candidate proteins that were validated for their role as OA stress biomarkers, which is most likely due to the experimental set-up: the small sample size employed in the experiment, as well as the use of only a single MS injection of each sample. This was predominantly due to financial constraints, as well as the limited availability of acclimated abalone for conducting aquarium-based experiments. Future studies should include more biological replicates, as well as technical replicates (i.e. ≥ 3 MS injections). This would not only enhance the depth and range of the LFQ proteomics approach, but would also improve the statistical analyses, thus resulting in a list of candidate biomarkers supported by statistical confidence.

LFQ served as a suitable avenue for the relative quantification of candidate protein biomarkers of abiotic stress. The data presented in this chapter contributes to the growing body of MS spectral data for molluscs, especially *H. midae*, which in turn fortifies publicly available species-specific proteomics data. Furthermore, this data also provides a foundation upon which further studies can be pursued. Particularly, investigation of the suitability and reliability of the

candidate biomarkers for detecting stress affecting *H. midae* sampled on abalone farms located around the South African coast will prove invaluable. In doing so, we hope to utilize this "suite" of biomarkers as a component of a health monitoring program for farmed abalone, which would ultimately aid in improving farming techniques and mitigating major losses as climate change intensifies over the next few decades.

5.4 Conclusion

The outcomes of this research project have fulfilled the aims set out in Chapter 1: to quantitatively and functionally characterize the haemocyte stress response of *H. midae* to OA conditions, as well as to validate candidate protein biomarkers of OA and acute temperature stress using an LFQ proteomics approach. The data presented in this dissertation provide a solid foundation for understanding the functional stress response of an economically important marine mollusc, while contributing towards the fields of proteomics and biomarker discovery in aquaculture, especially in light of climate change. The survivability and adaptation of *H. midae* to future climatic conditions is contingent on a number of factors, most notably the complex interplay between the stress and immune responses. Haemocyte proteins upregulated in response to abiotic stress were investigated for their potential as biomarkers, providing the groundwork for future on-farm experimentation to put to test the viability of a "suite" of biomarkers that can constitute a health monitoring program for farmed *H. midae*.

The validated candidate proteins together with the functional analysis from this research provide further insight into the stress response of *H. midae* under a changing climate, particularly in response to ocean acidification. Figure 1 presents a synthesis of the candidate biomarker proteins in relation to their respective functional roles following OA stress (blue), acute temperature stress (red), or a combination of the two stressors (green). Candidate proteins of OA stress are suggested to mainly be involved in functions of the innate immune system, which, to reiterate, has been reported to overlap with the generalised stress response (Figure 1). Acute temperature stress appears to elicit responses from candidate proteins predominantly associated with energetics and oxidative ER stress, and the effects thereof on the cytoskeleton (i.e. stabilization and re-organisation) (Figure 1). Finally, biomarker candidates associated with signal transduction and intracellular trafficking via the Endocytic pathway, as well as the cytoskeleton, were upregulated in response to a combination of the two stresses (which has never before been investigated in *H. midae*) (Figure 1). Collectively, this study provides a

comprehensive insight into the putative responses (as represented by the associated candidate biomarker proteins) of the *H. midae* haemocyte proteome following exposure to stressors associated with climate change.

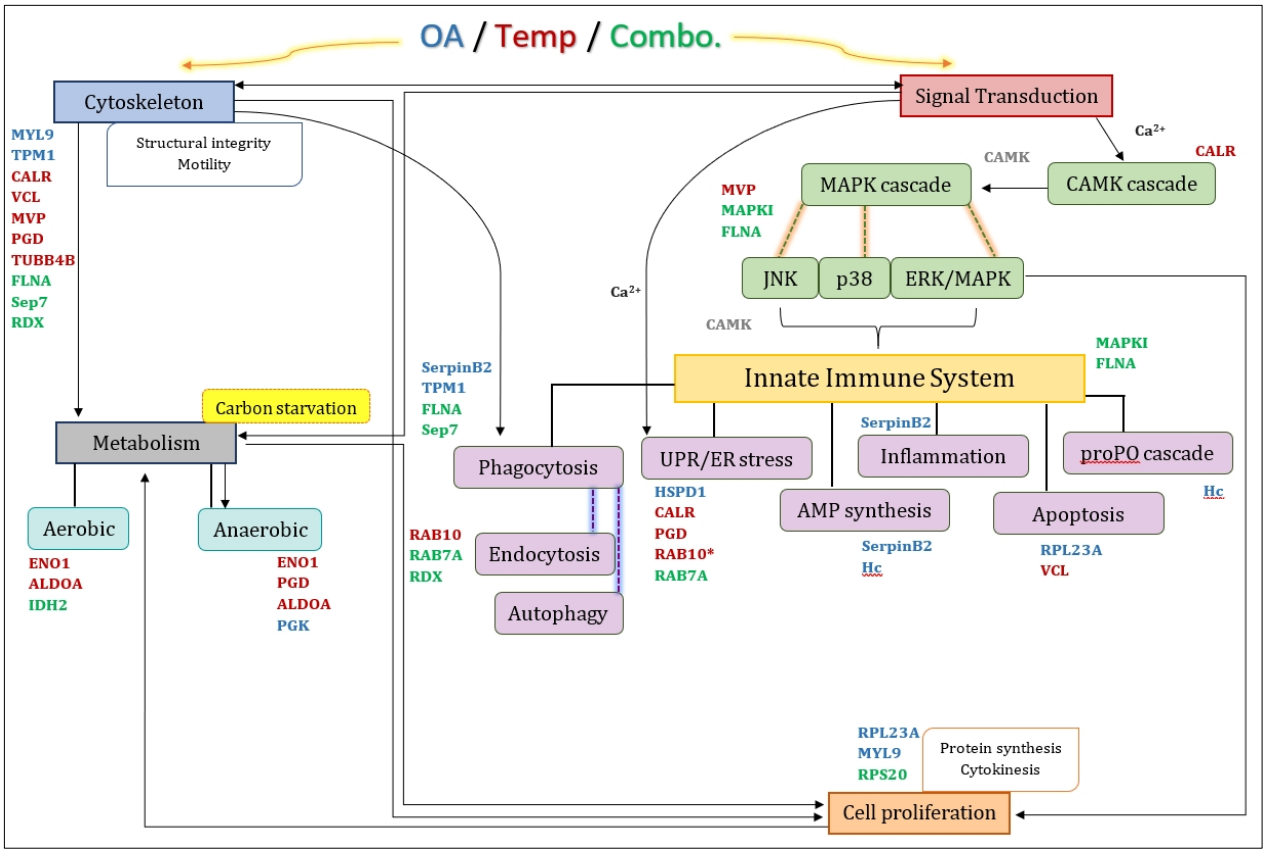

Figure 1. Candidate proteins of OA stress (blue), acute temperature stress (red) and a combination of the stressors (green), showing their putative functions (boxes) in *H. midae* haemocytes.

In conclusion, mass spectrometry and the -omics fields have transformed the way we approach scientific research related to climate change and the effects thereof on the environment. The global climate is currently changing at an unprecedented rate and the fate of marine ecosystems and key species is a crucial avenue for research. In carrying out this research, my goal was to contribute to this field and to provide further insight into an enigmatic species of the Haliotidae family. As such, the outcomes of this research provide a putative model of the effects of abiotic stress on the *H. midae* haemocyte proteome, while providing a foundation upon which further confirmatory studies in the wild or on abalone farms can be conducted. In its entirety, this research highlights the avenues for developing health monitoring programs on abalone farms, which could ultimately improve the sustainability of the South African abalone aquaculture industry in a changing climate.

Appendix A

Supplementary information

Table S1. The 227 differentially expressed proteins identified in *H. midae* following exposure to reduced pH conditions for 12, 72 and 168 hours. Protein fold change across the 4 replicates has been expressed as average ratios, following Log2 transformation and normalisation to the control group (T0).

Uniprot Accession [a]	Protein [b]		Average ratios			
		-10lgP [c]	T0	T12	T72	T168
tr V4C4V3 V4C4V3_LOTGI	14-3-3 protein beta/alpha	69.07	0	0.00	-0.09	0.12
tr A0A2C9JTK9 A0A2C9JTK9_BIOGL	14-3-3 protein epsilon	89.29	0	0.03	0.05	0.03
tr A0A2C9JJZ0 A0A2C9JJZ0_BIOGL	26S protease regulatory subunit 8	61.27	0	0.14	-0.10	-0.12
tr A0A0L8GT57 A0A0L8GT57_OCTBM	26S proteasome non-ATPase regulatory subunit 1	37.69	0	0.07	0.02	-0.14
tr A0A2C9JUD8 A0A2C9JUD8_BIOGL	40S ribosomal protein S16	64.93	0	-0.01	-0.08	0.11
tr B6RB81 B6RB81_HALDI	40S ribosomal protein S18	69.92	0	-0.37	-0.17	-0.53
tr A0A210Q0W5 A0A210Q0W5_MIZYE	40S ribosomal protein S25	36.94	0	0.00	0.16	0.04
tr K1QWC3 K1QWC3_CRAGI	40S ribosomal protein S3	81.14	0	-0.22	-0.17	-0.20
tr A0A210PKG8 A0A210PKG8_MIZYE	40S ribosomal protein S5	48.28	0	0.04	0.03	0.11
tr B3TK66 B3TK66_HALDV	40S ribosomal protein S8	101.77	0	-0.19	-0.09	-0.06
tr A0A210PM31 A0A210PM31_MIZYE	40S ribosomal protein S9	62.71	0	-0.29	-0.05	-0.17
tr K1QN55 K1QN55_CRAGI	60S acidic ribosomal protein P1	39.88	0	0.13	-0.18	0.01
tr A0A0B6Z7B1 A0A0B6Z7B1_9EUPU	60S ribosomal protein L13	52.41	0	0.01	-0.12	0.00
tr K1Q273 K1Q273_CRAGI	60S ribosomal protein L14	67.44	0	-0.16	-0.10	-0.09
tr A0A2T7PHL9 A0A2T7PHL9_POMCA	60S ribosomal protein L23a	44.94	0	0.38	-0.01	-0.08
tr A0A2T7P5E9 A0A2T7P5E9_POMCA	60S ribosomal protein L24-like	34.20	0	0.34	0.38	-0.41

tr A0A2T7PIY3 A0A2T7PIY3_POMCA	60S ribosomal protein L27a	59.52	0	0.37	0.03	0.09
tr K1RH70 K1RH70_CRAGI	6-phosphogluconate dehydrogenase, decarboxylating	70.98	0	-0.11	0.06	0.27
tr V4BUJ1 V4BUJ1_LOTGI	78 kDa glucose-regulated protein	103.57	0	-0.04	-0.05	0.23
tr B3SNT2 B3SNT2_HALDV	Actin	231.94	0	0.02	-0.05	-0.05
tr B3SND4 B3SND4_HALDV	Actin depolymerisation factor/cofilin	75.83	0	-0.11	-0.28	0.11
tr A0A2C9KD73 A0A2C9KD73_BIOGL	actin, plasmodial isoform-like	159.70	0	0.52	0.42	-0.13
tr A0A210QJS7 A0A210QJS7_MIZYE	Actin-2	222.52	0	0.01	-0.24	-0.18
tr A0A2C9JUN2 A0A2C9JUN2_BIOGL	actin-5C	158.74	0	0.19	0.26	-0.23
tr A0A0L8GB97 A0A0L8GB97_OCTBM	actin-related protein 2 isoform X2	56.90	0	-0.05	0.07	0.01
tr A0A0L8FI18 A0A0L8FI18_OCTBM	Actin-related protein 2/3 complex subunit 4	78.49	0	-0.09	0.03	-0.14
tr A0A210Q1K8 A0A210Q1K8_MIZYE	Actin-related protein 3	96.32	0	0.03	0.01	-0.12
tr A0A0B7BQ86 A0A0B7BQ86_9EUPU	ADP, ATP carrier protein 3, mitochondrial-like	87.37	0	-0.55	-0.15	-0.28
tr A0A2C9JJL0 A0A2C9JJL0_BIOGL	ADP-ribosylation factor 1	87.07	0	-0.29	-0.08	-0.23
tr B6RB72 B6RB72_HALDI	ADP-ribosylation factor 2	81.33	0	-0.42	-0.25	-0.23
tr A0A0B7BKX0 A0A0B7BKX0_9EUPU	ADP-ribosylation factor 4	82.89	0	-0.08	-0.03	-0.05
tr A0A2T7NDZ2 A0A2T7NDZ2_POMCA	aldehyde dehydrogenase, mitochondrial-like	38.27	0	-0.14	0.12	-0.54
tr A8DU73 A8DU73_9MOLL	Alpha enolase (Fragment)	83.27	0	0.22	0.31	0.39
tr K1PCW5 K1PCW5_CRAGI	Alpha-1 3-mannosyltransferase ALG2	47.83	0	0.27	-0.05	0.16
tr A0A2C9KFK2 A0A2C9KFK2_BIOGL	Alpha-1 4 glucan phosphorylase	69.83	0	-0.32	-0.12	0.18
tr K1RH58 K1RH58_CRAGI	Alpha-actinin, sarcomeric	152.64	0	-0.15	-0.07	-0.29
tr K1Q811 K1Q811_CRAGI	Alpha-centractin	37.97	0	0.08	-0.14	-0.01
tr A0A2C9JQJ6 A0A2C9JQJ6_BIOGL	alpha-L-fucosidase-like	40.38	0	0.16	0.14	0.24
tr A0A1I9W307 A0A1I9W307_HALDH	Aminopeptidase	41.68	0	0.20	-0.10	-0.11
tr K1R2I9 K1R2I9_CRAGI	Ankyrin-3	63.27	0	-0.04	0.09	0.07
tr K1PYW8 K1PYW8_CRAGI	Annexin	45.37	0	-0.13	-0.18	0.12
tr A0A2C9JZ65 A0A2C9JZ65_BIOGL	AP complex subunit beta	108.30	0	-0.05	-0.20	-0.44
tr K1RG10 K1RG10_CRAGI	AP complex subunit sigma	74.29	0	0.13	-0.04	0.14
tr V4A112 V4A112_LOTGI	AP-2 complex subunit alpha	72.08	0	0.01	-0.06	-0.17
tr A0A2T7NW66_A0A2T7NW66_POMCA	AP-2 complex subunit sigma	74.29	0	0.13	-0.04	0.14
tr A0A210QPW6 A0A210QPW6_MIZYE	Ariadne-like RING finger protein	56.46	0	0.07	0.03	0.04
tr V4A6V0 V4A6V0_LOTGI	ATP synthase subunit alpha	157.33	0	-0.20	-0.12	0.16

tr Q45Y90 Q45Y90_HALRU	ATP synthase subunit beta (Fragment)	136.44	0	-0.24	-0.07	-0.22
tr A0A1S5WH81 A0A1S5WH81_CONMI	ATPase subunit B (Fragment)	42.97	0	-0.15	-0.08	-0.11
tr B3TK83 B3TK83_HALDV	Beta-actin 2	231.94	0	0.02	-0.05	-0.05
tr A0A2T7P4L5 A0A2T7P4L5_POMCA	beta-parvin-like isoform X2	122.69	0	0.07	0.21	0.04
tr A0A2T7P4Z1 A0A2T7P4Z1_POMCA	bromodomain and WD repeat-containing protein 1	62.08	0	-0.03	0.04	-0.13
tr V4ABU5 V4ABU5_LOTGI	CAD protein-like isoform X4	38.66	0	0.01	0.05	0.05
tr Q6EEV2 Q6EEV2_PINFU	Calmodulin	93.77	0	-0.24	-0.61	0.03
tr B3TK26 B3TK26_HALDV	Calmodulin 2	44.79	0	0.16	-0.03	0.26
tr B3TK60 B3TK60_HALDV	Calmodulin-dependent protein kinase (Fragment)	66.54	0	0.52	0.22	0.58
tr V3ZFZ8 V3ZFZ8_LOTGI	calpain-B-like isoform X19	68.59	0	-0.06	-0.21	-0.05
tr A0A140C6Z8 A0A140C6Z8_HALDI	Calreticulin	88.03	0	0.02	-0.14	0.08
tr E6Y9P6 E6Y9P6_BIOGL	Cdc24-like protein	70.69	0	0.10	0.04	0.19
tr Q86DH9 Q86DH9_APLCA	Cdc42	70.69	0	0.10	0.04	0.19
tr A0A0B7AYN3 A0A0B7AYN3_9EUPU	cdc42 homolog	70.69	0	-0.15	0.01	0.19
tr A0A0L8G456 A0A0L8G456_OCTBM	cdc42 homolog	70.69	0	-0.05	0.06	0.17
tr H9AWU2 H9AWU2_HALDH	Chaperonin containing T-complex polypeptide subunit zeta	118.97	0	0.08	0.41	0.11
tr B6RB18 B6RB18_HALDI	Chaperonin containing tcp1	123.5	0	0.21	0.13	0.13
tr A0A0L8FL54 A0A0L8FL54_OCTBM	charged multivesicular body protein 1a-like	37.99	0	-0.31	0.00	0.65
tr A0A2C9JRV1 A0A2C9JRV1_BIOGL	Clathrin heavy chain	91.11	0	-0.17	-0.04	0.06
tr A0A2T7PDX8 A0A2T7PDX8_POMCA	clathrin heavy chain 1	101.79	0	0.07	0.14	0.09
tr V4A0B8 V4A0B8_LOTGI	Coatomer subunit alpha	95.40	0	-0.32	-0.11	0.06
tr A0A2T7NIY3 A0A2T7NIY3_POMCA	copine-3-like	49.34	0	-0.06	0.01	0.01
tr A0A210PIS5 A0A210PIS5_MIZYE	Cytoplasmic aconitate hydratase	51.59	0	-0.06	-0.16	0.08
tr A0A2C9JEV1 A0A2C9JEV1_BIOGL	cytoplasmic dynein 1 heavy chain 1-like isoform X1	105.96	0	-0.10	-0.20	0.03
tr V4BGR4 V4BGR4_LOTGI	cytoplasmic dynein 1 heavy chain 1-like isoform X2	126.05	0	-0.10	-0.06	0.04
tr A0A2T7NSR7 A0A2T7NSR7_POMCA	cytosolic 10-formyltetrahydrofolate dehydrogenase-like	73.47	0	-0.22	-0.11	-0.11
tr V3YZ86 V3YZ86_LOTGI	deoxyribose-phosphate aldolase-like	46.04	0	0.14	0.16	0.15
tr K1RM18 K1RM18_CRAGI	Dual oxidase	41.76	0	-0.69	0.28	-0.17
tr A0A210QZA7 A0A210QZA7_MIZYE	Dual oxidase 2	34.48	0	-0.32	-0.19	-0.10
tr V4A9P6 V4A9P6_LOTGI	dual oxidase 2-like	47.87	0	-0.19	-0.19	0.03

tr V4A2S8 V4A2S8_LOTGI	dual oxidase 2-like isoform X1	97.32	0	-0.08	-0.14	0.00
tr K1QHK9 K1QHK9_CRAGI	Dynein heavy chain, cytoplasmic	122.11	0	-0.25	-0.31	-0.19
tr D1H0L8 D1H0L8_HALTU	Elongation factor 1-alpha	89.03	0	0.21	0.26	0.03
tr K1QFW9 K1QFW9_CRAGI	elongation factor 2	35.44	0	-0.03	-0.01	-0.11
tr K1QX26 K1QX26_CRAGI	Endoplasmin	64.23	0	-0.03	-0.06	-0.02
tr K1QX37 K1QX37_CRAGI	Enolase	83.27	0	0.22	0.31	0.39
tr A0A2T7NB66 A0A2T7NB66_POMCA	eukaryotic initiation factor 4A-I	72.99	0	-0.04	0.07	-0.12
tr G9K379 G9K379_HALDV	Eukaryotic translation initiation factor 5A	35.22	0	-0.10	-0.11	-0.07
tr K1QEZ3 K1QEZ3_CRAGI	Fascin	42.45	0	0.18	0.36	0.20
tr A0A0K0YAX9 A0A0K0YAX9_MYTCO	Filamin-like protein-1	62.35	0	0.07	-0.14	-0.09
tr A0A2T7NM92 A0A2T7NM92_POMCA	flotillin-1-like isoform X2	37.57	0	0.12	0.15	0.14
tr A0A2T7PI06 A0A2T7PI06_POMCA	fructose-bisphosphate aldolase	60.92	0	-0.12	-0.04	0.17
tr A0A1R7T095 A0A1R7T095_HALMK	Galectin	142.12	0	-0.22	-0.23	-0.32
tr B6RB97 B6RB97_HALDI	Gelsolin	46.48	0	-0.16	-0.12	0.06
tr A0A0B6ZFV9 A0A0B6ZFV9_9EUPU	gelsolin-like protein 2	61.66	0	-0.03	-0.03	0.07
tr V3ZMG1 V3ZMG1_LOTGI	glucose-regulated protein 94	87.37	0	0.10	0.12	-0.02
tr A0A1R7T098 A0A1R7T098_HALMK	Glutathione reductase	83.97	0	0.02	-0.14	-0.02
tr B3TK24 B3TK24_HALDV	Glutathione-S-transferase isoform	56.29	0	0.03	0.11	0.07
tr B6RB30 B6RB30_HALDI	Glyceraldehyde-3-phosphate dehydrogenase	128.59	0	0.10	0.01	0.30
tr A0A2C9JK49 A0A2C9JK49_BIOGL	glycerol-3-phosphate dehydrogenase, mitochondrial-like isoform X1	64.72	0	-0.16	-0.77	-0.09
tr A0A2T7NWS0 A0A2T7NWS0_POMCA	Glycogen phosphorylase	36.98	0	0.17	-0.01	-0.22
tr A0A210QVP2 A0A210QVP2_MIZYE	GTP-binding protein SAR1b	49.93	0	-0.12	-0.11	0.07
tr K1R2S6 K1R2S6_CRAGI	Guanine nucleotide-binding protein G(O) subunit alpha	32.37	0	-0.31	-0.06	-0.19
tr Q8I0U4 Q8I0U4_HALTU	H2 protein (Fragment)	199.09	0	-0.25	-0.13	0.04
tr Q564J1 Q564J1_APLCA	Haemocyanin	51.63	0	-0.22	0.26	-0.18
tr M1JNK7 M1JNK7_HALDV	Heat shock protein 90	110.73	0	0.18	-0.02	0.13
tr Q1MVA1 Q1MVA1_MEGCR	Hemocyanin 2	132.42	0	-0.26	-0.11	-0.09
tr A0A2T7PEM0 A0A2T7PEM0_POMCA	hemocyanin G-type, units Oda to Odg-like	50.55	0	0.07	0.35	0.43
tr A0A2C9JTB5 A0A2C9JTB5_BIOGL	heterogeneous nuclear ribonucleoprotein R-like isoform X1	40.11	0	-0.08	0.11	0.03

tr A0A2C9JTB9 A0A2C9JTB9_BIOGL	heterogeneous nuclear ribonucleoprotein R-like isoform X2	61.80	0	-0.10	-0.19	-0.08
tr A0A2C9JU94 A0A2C9JU94_BIOGL	hillarin isoform X1	78.88	0	0.15	-0.29	0.10
tr K0J2U1 K0J2U1_9BIVA	Histone 3 (Fragment)	32.65	0	0.05	-0.19	0.30
tr K1RXS6 K1RXS6_CRAGI	Histone H2A	57.04	0	-0.29	0.10	0.07
tr A0A2C9KJV3 A0A2C9KJV3_BIOGL	Histone H2B	36.47	0	-0.38	-0.30	-0.03
tr G8XRU5 G8XRU5_9CAEN	Histone H3 (Fragment)	32.65	0	0.05	-0.19	0.30
tr V4A0D9 V4A0D9_LOTGI	Histone H4	159.57	0	-0.49	-0.19	-0.19
tr A0A2C9LW88 A0A2C9LW88_BIOGL	huntingtin-interacting protein 1-like isoform X4	40.15	0	0.05	-0.31	0.25
tr K1QNQ9 K1QNQ9_CRAGI	hypothetical protein CGI_10024082	57.27	0	0.26	0.34	-0.18
tr A0A210Q018 A0A210Q018_MIZYE	Innexin	40.55	0	-0.68	-0.35	-0.36
tr V4AWB5 V4AWB5_LOTGI	inositol polyphosphate 1-phosphatase-like	53.95	0	-0.09	-0.02	-0.05
tr K1QR98 K1QR98_CRAGI	Integrin-linked protein kinase	74.22	0	-0.15	-0.06	-0.12
tr A0A2T7PNT8 A0A2T7PNT8_POMCA	integrin-linked protein kinase-like isoform X2	76.26	0	-0.13	-0.17	0.10
tr A0A2T7P1P2 A0A2T7P1P2_POMCA	kelch-like protein 24 isoform X1	38.76	0	0.05	0.14	0.20
tr Q6KC55 Q6KC55_MEGCR	Keyhole limpet hemocyanin2	132.42	0	-0.26	-0.11	-0.09
tr A0A2T7PHH2 A0A2T7PHH2_POMCA	major vault protein-like	87.26	0	0.16	0.36	0.36
tr B6RB90 B6RB90_HALDI	Malate dehydrogenase (Fragment)	75.47	0	-0.01	-0.01	-0.04
tr A0A2T7NTX8 A0A2T7NTX8_POMCA	mitochondrial-processing peptidase subunit beta	53.50	0	0.10	0.13	-0.05
tr A0A1L7H9J9 A0A1L7H9J9_HALDI	Mitogen-activated protein kinase	68.82	0	0.49	0.16	0.07
tr A6MD73 A6MD73_HALDI	Mx	75.29	0	-0.33	0.16	-0.34
tr A0A2C9KA20 A0A2C9KA20_BIOGL	myosin heavy chain, non-muscle-like isoform X1	83.4	0	-0.03	-0.20	-0.14
tr A0A2C9K9W0 A0A2C9K9W0_BIOGL	myosin heavy chain, non-muscle-like isoform X3	84.4	0	0.06	-0.33	-0.14
tr A0A2C9K9W7 A0A2C9K9W7_BIOGL	myosin heavy chain, non-muscle-like isoform X4	83.4	0	-0.05	-0.25	-0.13
tr K1R5M2 K1R5M2_CRAGI	Myosin-IXa	38.93	0	0.30	0.00	0.16
tr A0A2T7PRR5 A0A2T7PRR5_POMCA	N-chimaerin isoform X2	54.82	0	-0.11	0.08	0.14
tr A0A210QXB7 A0A210QXB7_MIZYE	Nesprin-1	75.88	0	0.02	-0.13	-0.09
tr A0A159ZKT0 A0A159ZKT0_HALDV	Non-specific serine/threonine protein kinase	61.34	0	0.08	0.20	0.07
tr A0A2C9JK44 A0A2C9JK44_BIOGL	Obg-like ATPase 1	71.28	0	0.04	0.00	0.28
tr A0A1R7T0A2 A0A1R7T0A2_HALMK	Omega class glutathione-s-transferase 1	74.20	0	-0.01	0.10	0.00
tr A0A286QYA2 A0A286QYA2_HALDH	Paramyosin	39.41	0	-0.07	0.27	0.33

tr K1RDX5 K1RDX5_CRAGI	Peptidyl-prolyl cis-trans isomerase	30.52	0	-0.08	-0.21	-0.25
tr V4B404 V4B404_LOTGI	phosphoglucomutase-1-like	45.33	0	-0.10	-0.25	-0.10
tr V4ADA9 V4ADA9_LOTGI	Phosphoglycerate kinase	77.87	0	0.12	0.43	0.23
tr C7EAA2 C7EAA2_HALAI	PL10-like protein	120.62	0	0.16	0.07	-0.01
tr V3Z1G4 V3Z1G4_LOTGI	plexin-A2-like isoform X1	62.58	0	-0.30	-0.10	-0.05
tr A0A2C9KJB3 A0A2C9KJB3_BIOGL	polyubiquitin-C isoform X4	31.33	0	-0.12	-0.28	-0.19
tr A0A2T7NTI6 A0A2T7NTI6_POMCA	PRA1 family protein 2-like	58.37	0	0.05	0.11	0.19
tr A0A2C9LDE5 A0A2C9LDE5_BIOGL	PREDICTED: uncharacterized protein LOC106070598	42.01	0	0.03	-0.02	0.20
tr A0A0L8HWW8 A0A0L8HWW8_OCTBM	PREDICTED: uncharacterized protein LOC106868120	53.74	0	-0.12	-0.06	-0.06
tr A0A2T7P0I9 A0A2T7P0I9_POMCA	prefoldin subunit 2	57.28	0	-0.01	0.14	0.05
tr V3ZW18 V3ZW18_LOTGI	probable methylmalonate-semialdehyde dehydrogenase [acylating], mitochondrial	76.54	0	0.14	0.30	0.40
tr A0A0B6Y9I8 A0A0B6Y9I8_9EUPU	probable small nuclear ribonucleoprotein Sm D2	55.55	0	-0.72	0.44	0.52
tr A0A2T7PYV8 A0A2T7PYV8_POMCA	prohibitin isoform X1	86.50	0	0.13	-0.16	0.01
tr A0A1X9T5X9 A0A1X9T5X9_HALDH	Prolyl endopeptidase	75.33	0	-0.12	-0.13	0.02
tr A0A0B6Z7I4 A0A0B6Z7I4_9EUPU	Proteasome subunit alpha type	45.98	0	0.08	0.04	0.01
tr A0A0B6Z599 A0A0B6Z599_9EUPU	protein ABHD17B-like	67.57	0	-0.06	0.16	0.06
tr B6RB63 B6RB63_HALDI	Protein disulfide-isomerase	108.22	0	-0.13	-0.20	-0.04
tr A0A2T7NEV5 A0A2T7NEV5_POMCA	protein disulfide-isomerase A3-like	52.83	0	0.21	-0.01	0.10
tr A0A0L8GV67 A0A0L8GV67_OCTBM	protein FAM151B	45.11	0	-0.32	0.16	-0.06
tr A0A210PWU9 A0A210PWU9_MIZYE	Protein N-terminal glutamine amidohydrolase	30.17	0	0.14	0.19	0.05
tr K1R866 K1R866_CRAGI	Puromycin-sensitive aminopeptidase	69.01	0	-0.02	0.17	0.11
tr G9K381 G9K381_HALDV	Putative 40S ribosomal protein S18	82.47	0	-0.31	0.04	-0.29
tr A0A194AKV2 A0A194AKV2_PINFU	Putative 40S ribosomal protein S25-like protein	36.94	0	0.00	0.16	0.04
tr A0A2C9M1J7 A0A2C9M1J7_BIOGL	putative ankyrin repeat protein RF_0381	44.66	0	-0.43	-0.05	-0.11
tr Q70MT3 Q70MT3_CRAGI	Putative ribosomal protein S25 (Fragment)	36.94	0	-0.03	0.01	-0.17
tr B3TK70 B3TK70_HALDV	Putative RNA-binding protein	44.21	0	-0.09	-0.07	0.07
tr V4A8R6 V4A8R6_LOTGI	putative uncharacterized oxidoreductase YDR541C	34.91	0	0.37	0.27	0.17
tr K1PJ46 K1PJ46_CRAGI	Pyruvate kinase	34.13	0	-0.17	-0.11	-0.15
tr B8XW76 B8XW76_HALDV	QM-like protein	78.58	0	0.01	-0.04	-0.19

tr A0A210PLN1 A0A210PLN1_MIZYE	Quinone oxidoreductase	53.32	0	-0.12	0.00	-0.22
tr B6RB76 B6RB76_HALDI	RAB protein	58.72	0	-0.11	-0.01	-0.08
tr A0A0B7BF17 A0A0B7BF17_9EUPU	radixin isoform X1	64.11	0	0.16	0.28	0.30
tr A0A2T7NG09 A0A2T7NG09_POMCA	ras GTPase-activating-like protein IQGAP1 isoform X1	81.63	0	-0.12	0.42	-0.06
tr B6RB23 B6RB23_HALDI	Ras-related protein Rab-1A	103.41	0	0.01	0.03	0.05
tr A0A210PFE4 A0A210PFE4_MIZYE	Ras-related protein Rab-2	60.51	0	-0.26	-0.18	0.01
tr K1PZ08 K1PZ08_CRAGI	Ras-related protein Rab-7a	86.18	0	-0.16	-0.14	-0.07
tr O61472 O61472_APLCA	Reductase-related protein	39.22	0	0.30	0.06	0.28
tr A0A210Q7D8 A0A210Q7D8_MIZYE	Retinal dehydrogenase 2	55.21	0	0.09	0.17	0.31
tr A0A0N9HJE9 A0A0N9HJE9_9BIVA	Rho GTPase	70.69	0	-0.19	0.03	0.13
tr Q70MN8 Q70MN8_CRAGI	Ribosomal protein L24 (Fragment)	34.20	0	0.34	0.38	-0.41
tr B6RB96 B6RB96_HALDI	Ribosomal protein l5 (Fragment)	79.07	0	0.18	0.00	0.08
tr V4BY94 V4BY94_LOTGI	ribosomal protein LP1	39.88	0	0.13	-0.18	0.01
tr J9Q550 J9Q550_OSTED	Ribosomal protein P1	39.88	0	0.06	-0.23	-0.08
tr A0A077H3J9 A0A077H3J9_MYTTR	Ribosomal protein S16 (Fragment)	75.36	0	-0.08	0.01	-0.06
tr E3P7H5 E3P7H5_LEPCI	Ribosomal protein S25	36.94	0	-0.07	-0.02	-0.11
tr B6RB83 B6RB83_HALDI	Ribosomal protein S4 (Fragment)	71.04	0	-0.16	-0.11	-0.13
tr A0A2C9JCF0 A0A2C9JCF0_BIOGL	Ribosomal protein S6 kinase	70.27	0	-0.10	-0.04	0.12
tr A9LMJ6 A9LMJ6_HALDI	Ribosomal protein S9	81.53	0	-0.12	-0.09	-0.04
tr K1QG65 K1QG65_CRAGI	rRNA 2'-O-methyltransferase fibrillarin	43.63	0	-0.21	0.52	-0.07
tr V3ZJR3 V3ZJR3_LOTGI	serine/arginine-rich splicing factor 10-like isoform X1	37.11	0	-0.03	-0.15	0.19
tr B6RB22 B6RB22_HALDI	Serine/threonine-protein phosphatase	79.08	0	-0.07	0.06	-0.03
tr A0A2T7P0L9 A0A2T7P0L9_POMCA	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	97.98	0	0.19	0.19	0.20
tr A0A0G2SSZ6 A0A0G2SSZ6_HALDI	Serpin B-like protein 2	52.44	0	0.11	0.38	0.21
tr Q8MUE3 Q8MUE3_9BIVA	Small nuclear ribonucleoprotein D2-like protein	55.55	0	-0.83	0.31	0.56
tr A0A0L8FUD9 A0A0L8FUD9_OCTBM	small nuclear ribonucleoprotein Sm D2	42.46	0	-0.10	-0.12	0.19
tr A0A2T7PQU0 A0A2T7PQU0_POMCA	sodium/potassium-transporting ATPase subunit alpha-like	43.25	0	-0.26	-0.37	-0.15
tr A0A0B6ZMQ4 A0A0B6ZMQ4_9EUPU	Solute carrier family 2, facilitated glucose transporter member 1	39.32	0	0.11	0.28	0.32

tr A0A0L8H0E1 A0A0L8H0E1_OCTBM	Sorting nexin	43.62	0	-0.12	0.45	0.09
tr A0A2T7NLR4 A0A2T7NLR4_POMCA	spectrin alpha chain-like isoform X5	49.85	0	0.05	0.10	-0.03
tr A0A2C9JSV3 A0A2C9JSV3_BIOGL	spliceosome RNA helicase DDX39B	59.54	0	0.23	-0.01	0.03
tr A0A210QMS8 A0A210QMS8_MIZYE	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	81.04	0	-0.17	-0.15	0.08
tr A0A0L8G346 A0A0L8G346_OCTBM	SWI/SNF complex subunit SMARCC2-like	29.11	0	-0.18	0.17	0.14
tr A0A2T7PSF5 A0A2T7PSF5_POMCA	talin-1-like isoform X1	164.86	0	-0.12	-0.23	-0.16
tr V4AQH1 V4AQH1_LOTGI	TAR DNA-binding protein 43-like	50.01	0	-0.03	-0.12	0.04
tr A0A2C9K119 A0A2C9K119_BIOGL	T-complex protein 1 subunit beta	57.20	0	0.24	0.11	0.06
tr A0A2C9JP35 A0A2C9JP35_BIOGL	T-complex protein 1 subunit delta	67.76	0	-0.04	-0.01	0.08
tr A0A0B6ZUU6 A0A0B6ZUU6_9EUPU	T-complex protein 1 subunit epsilon-like	103.72	0	-0.04	0.05	0.03
tr A0A210QB77 A0A210QB77_MIZYE	T-complex protein 1 subunit gamma	68.32	0	-0.20	-0.14	-0.25
tr K1QH85 K1QH85_CRAGI	Trafficking protein particle complex subunit	43.97	0	-0.12	0.08	-0.16
tr A0A210PM24 A0A210PM24_MIZYE	Transaldolase	62.39	0	0.07	-0.11	0.09
tr A0A2C9KGC1 A0A2C9KGC1_BIOGL	transcriptional activator protein Pur-beta-like isoform X1	84.68	0	0.01	0.11	0.20
tr A0A0K0WS32 A0A0K0WS32_HALDI	Transferrin-like protein	54.53	0	-0.37	0.12	-0.25
tr K1RBC9 K1RBC9_CRAGI	Transketolase-like protein 2	69.54	0	-0.09	0.04	-0.04
tr A0A2T7NW92 A0A2T7NW92_POMCA	transmembrane emp24 domain-containing protein 10-like	50.13	0	-0.13	-0.05	0.13
tr Q45Y86 Q45Y86_HALRU	Triosephosphate isomerase (Fragment)	77.77	0	-0.08	-0.51	-0.12
tr B7XC62 B7XC62_HALDI	Tropomyosin	79.43	0	0.03	0.05	0.03
tr Q7YZR1 Q7YZR1_HALAI	Tropomyosin 4 (Fragment)	46.91	0	-0.14	-0.32	0.01
tr B6RB46 B6RB46_HALDI	Troponin	42.32	0	0.12	-0.06	-0.06
tr V4B157 V4B157_LOTGI	Tubulin alpha chain	93.70	0	-0.03	0.20	-0.08
tr A0A0G2R2A9 A0A0G2R2A9_9BIVA	Tubulin beta chain (Fragment)	101.07	0	-0.16	-0.02	0.30
tr V4B045 V4B045_LOTGI	tubulin beta-4B chain isoform X2	109.12	0	-0.55	-0.12	-0.45
tr A0A2T7PAU9 A0A2T7PAU9_POMCA	U5 small nuclear ribonucleoprotein 200 kDa helicase	67.97	0	-0.34	0.02	-0.33
tr A0A077H3L5 A0A077H3L5_MYTTR	Ubiquitin-ribosomal protein S27a (Fragment)	54.23	0	-0.05	-0.30	-0.05
tr A0A2C9JRW8 A0A2C9JRW8_BIOGL	uncharacterized oxidoreductase YajO-like	37.49	0	-0.33	0.44	0.38
tr A0A0L8IEF8 A0A0L8IEF8_OCTBM	unconventional myosin-Ie-like isoform X1	52.77	0	0.01	-0.11	-0.30
tr A0A210PT48 A0A210PT48_MIZYE	Vacuolar protein sorting-associated protein 13A	64.69	0	0.03	0.05	0.66

tr A0A210QAT4 A0A210QAT4_MIZYE	Vacuolar protein sorting-associated protein VTA1-like	56.48	0	-0.19	0.00	0.00
tr A0A2S0WB76 A0A2S0WB76_TRISQ	Vacuolar-type proton ATPase subunit A	70.86	0	-0.22	0.11	-0.14
tr A0A2C9LFJ1 A0A2C9LFJ1_BIOGL	vegetative incompatibility protein HET-E-1-like	40.37	0	-0.12	-0.18	0.04
tr A0A2T7PIF5 A0A2T7PIF5_POMCA	vinculin-like isoform X11	176.89	0	-0.06	-0.16	0.05
tr K1QI28 K1QI28_CRAGI	V-type proton ATPase subunit B	104.78	0	-0.05	0.04	0.03

[a] Uniprot SwissProt/Trembl accession number; [b] Protein names assigned by Uniprot, [c] PEAKS peptide identification score

Appendix B

References Cited

Abdallah, C., Sergeant, K., Guillier, C., Dumas-Gaudot, E., Leclercq, C.C., Renaut, J., 2012. Optimization of iTRAQ labelling coupled to OFFGEL fractionation as a proteomic workflow to the analysis of microsomal proteins of Medicago truncatula roots. Proteome Sci. 10, 37. https://doi.org/10.1186/1477-5956-10-37

Adamo, S.A., 2012. The effects of the stress response on immune function in invertebrates: An evolutionary perspective on an ancient connection. Horm. Behav. 62, 324–330. https://doi.org/10.1016/J.YHBEH.2012.02.012

Adamo, S.A., 2008. Norepinephrine and octopamine: Linking stress and immune function across phyla. Invertebr. Surviv. J. 5, 12–19.

Aebersold, R., Burlingame, A.L., Bradshaw, R.A., 2013. Western blots versus selected reaction monitoring assays: time to turn the tables? Mol. Cell. Proteomics 12, 2381–2. https://doi.org/10.1074/mcp.E113.031658

Aggarwal, K., Choe, L.H., Lee, K.H., 2006. Shotgun proteomics using the iTRAQ isobaric tags. Briefings Funct. Genomics Proteomics 5, 112–120. https://doi.org/10.1093/bfgp/ell018

Aggarwal, S., Talukdar, N.C., Yadav, A.K., 2019. Advances in Higher Order Multiplexing Techniques in Proteomics. J. Proteome Res. 18, 2360–2369. https://doi.org/10.1021/acs.jproteome.9b00228

Araújo, J.E., Madeira, D., Vitorino, R., Repolho, T., Rosa, R., Diniz, M., 2018. Negative synergistic impacts of ocean warming and acidification on the survival and proteome of the commercial sea bream, *Sparus aurata*. J. Sea Res. 139, 50–61. https://doi.org/10.1016/J.SEARES.2018.06.011

Araya, M.T., Markham, F., Mateo, D.R., McKenna, P., Johnson, G.R., Berthe, F.C.J., Siah, A., 2010. Identification and expression of immune-related genes in hemocytes of soft-shell clams, *Mya arenaria*, challenged with *Vibrio splendidus*. Fish Shellfish Immunol. 29, 557–564. https://doi.org/10.1016/j.fsi.2010.05.017

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M., 2014. Non-

model organisms, a species endangered by proteogenomics. J. Proteomics 105, 5–18. https://doi.org/10.1016/J.JPROT.2014.01.007

Barcia, R., Ramos-Martínez, J.I., 2011. Stress-based modulation of the immune response in molluscan hemocytes: a two-receptor model. Invertebr. Surviv. J. 8, 56–58.

Barkai, R., Griffiths, C.L., 1986. Diet of the South African abalone *Haliotis midae*. South African J. Mar. Sci. 4, 37–44. https://doi.org/10.2989/025776186784461891

Barkai, R., Griffiths, C.L., 1987. Consumption, absorption efficiency, respiration and excretion in the South African abalone *Haliotis midae*. South African J. Mar. Sci. 5, 523–529. https://doi.org/10.2989/025776187784522441

Bathige, S.D.N.K., Umasuthan, N., Godahewa, G.I., Whang, I., Kim, C., Park, H.-C., Lee, J., 2015. Three novel clade B serine protease inhibitors from disk abalone, *Haliotis discus discus*: Molecular perspectives and responses to immune challenges and tissue injury. Fish Shellfish Immunol. 45, 334–341. https://doi.org/10.1016/J.FSI.2015.04.020

Bednaršek, N., Tarling, G.A., Bakker, D.C., Fielding, S., Cohen, A., Kuzirian, A., McCorkle, D., Lézé, B., Montagna, R., 2012. Description and quantification of pteropod shell dissolution: a sensitive bioindicator of ocean acidification. Glob. Chang. Biol. 18, 2378–2388. https://doi.org/10.1111/j.1365-2486.2012.02668.x

Belinda, L.W.C., Wei, W.X., Hanh, B.T.H., Lei, L.X., Bow, H., Ling, D.J., 2008. SARM: a novel Toll-like receptor adaptor, is functionally conserved from arthropod to human. Mol. Immunol. 45, 1732–1742. https://doi.org/10.1016/j.molimm.2007.09.030

Beltran, C.G.G., 2015. A proteomic investigation of the immune response of the South African abalone, *Haliotis midae*. University of Cape Town.

Berger, W., Steiner, E., Grusch, M., Elbling, L., Micksche, M., 2009. Review Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity. Cell. Mol. Life Sci 66, 43–61. https://doi.org/10.1007/s00018-008-8364-z

Bibby, R., Widdicombe, S., Parry, H., Spicer, J., Pipe, R., 2008. Effects of ocean acidification on the immune response of the blue mussel Mytilus edulis. Aquat. Biol. 2, 67–74. https://doi.org/10.3354/ab00037

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: A Cytoscape plug-in to

decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–1093. https://doi.org/10.1093/bioinformatics/btp101

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., Apweiler, R., 2009. QuickGO: A web-based tool for Gene Ontology searching. Bioinformatics 25, 3045–3046. https://doi.org/10.1093/bioinformatics/btp536

Boehm, A.M., Pütz, S., Altenhöfer, D., Sickmann, A., Falk, M., 2007. Precise protein quantification based on peptide quantification using iTRAQ^{TM}. BMC Bioinformatics 8, 214. https://doi.org/10.1186/1471-2105-8-214

Boutet, I., Jollivet, D., Shillito, B., Moraga, D., Tanguy, A., 2009. Molecular identification of differentially regulated genes in the hydrothermal-vent species *Bathymodiolus thermophilus* and *Paralvinella pandorae* in response to temperature. BMC Genomics 10, 222. https://doi.org/10.1186/1471-2164-10-222

Brackley, K.I., Grantham, J., 2009. Activities of the chaperonin containing TCP-1 (CCT): Implications for cell cycle progression and cytoskeletal organisation. Cell Stress Chaperones. https://doi.org/10.1007/s12192-008-0057-x

Briggs, M.W., Sacks, D.B., 2003. IQGAP1 as signal integrator: Ca^{2+}, calmodulin, Cdc42 and the cytoskeleton. FEBS Lett. 542, 7–11. https://doi.org/10.1016/S0014-5793(03)00333-8

Britz, P., Venter, S., 2016. Spotlight on Africa at World Aquaculture 2017.

Buckley, B.A., Somero, G.N., 2009. cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish *Trematomus bernacchii* to alter gene expression in response to heat stress. Polar Biol. 32, 403–415. https://doi.org/10.1007/s00300-008-0533-x

Buescher, J.M., Driggers, E.M., 2016. Integration of omics: More than the sum of its parts. Cancer Metab. https://doi.org/10.1186/s40170-016-0143-y

Calvete, J.J., 2014. Challenges and prospects of proteomics of non-model organisms. J. Proteomics 105, 1–4. https://doi.org/10.1016/J.JPROT.2014.04.034

Campos, A., Apraiz, I., Da Fonseca, R.R., Cristobal, S., 2015. Shotgun analysis of the marine mussel *Mytilus edulis* hemolymph proteome and mapping the innate immunity elements. Proteomics 15, 4021–4029. https://doi.org/10.1002/pmic.201500118

Campos, A., Danielsson, G., Farinha, A.P., Kuruvilla, J., Warholm, P., Cristobal, S., 2016a. Shotgun proteomics to unravel marine mussel (*Mytilus edulis*) response to long-term exposure to low salinity and propranolol in a Baltic Sea microcosm. J. Proteomics 137, 97–106. https://doi.org/10.1016/j.jprot.2016.01.010

Campos, A., de Almeida, A. 2016b. Top-Down Proteomics and Farm Animal and Aquatic Sciences. Proteomes 4, 38. https://doi.org/10.3390/proteomes4040038

Cao, R., Liu, Y., Wang, Q., Yang, D., Liu, H., Ran, W., Qu, Y., Zhao, J., 2018a. Seawater Acidification Reduced the Resistance of *Crassostrea gigas* to *Vibrio splendidus* Challenge: An Energy Metabolism Perspective. Front. Physiol. 9, 880. https://doi.org/10.3389/fphys.2018.00880

Cao, R., Wang, Q., Yang, D., Liu, Y., Ran, W., Qu, Y., Wu, H., Cong, M., Li, F., Ji, C., Zhao, J., 2018b. CO_2-induced ocean acidification impairs the immune function of the Pacific oyster against *Vibrio splendidus* challenge: An integrated study from a cellular and proteomic perspective. Sci. Total Environ. 625, 1574–1583. https://doi.org/10.1016/J.SCITOTENV.2018.01.056

Carlton, J., Bujny, M., Rutherford, A., Cullen, P., 2005. Sorting Nexins - Unifying Trends and New Perspectives. Traffic 6, 75–82. https://doi.org/10.1111/j.1600-0854.2005.00260.x

Castellanos-Martínez, S., Diz, A.P., Álvarez-Chaver, P., Gestal, C., 2014. Proteomic characterization of the hemolymph of *Octopus vulgaris* infected by the protozoan parasite *Aggregata octopiana*. J. Proteomics 105, 151–163. https://doi.org/10.1016/J.JPROT.2013.12.008

Chang, X., Zheng, C., Wang, Y., Meng, C., Xie, X., Liu, H., 2016. Differential protein expression using proteomics from a crustacean brine shrimp (*Artemia sinica*) under CO_2-driven seawater acidification. Fish Shellfish Immunol. 58, 669–677. https://doi.org/10.1016/J.FSI.2016.10.008

Chen, C.C.-H., Schweinsberg, P.J., Vashist, S., Mareiniss, D.P., Lambie, E.J., Grant, B.D., 2006. RAB-10 Is Required for Endocytic Recycling in the *Caenorhabditis elegans* Intestine. Mol. Biol. Cell 16, 1286–1297. https://doi.org/10.1091/mbc.E05

Chen, R., Kang, R., Fan, X.-G., Tang, D., 2014. Release and activity of histone in diseases.

Cell Death Dis. 5, e1370. https://doi.org/10.1038/cddis.2014.337

Chen, D., Jin, C., 2017. Histone variants in environmental-stress-induced DNA damage repair. Mutat. Res. Mutat. Res. https://doi.org/10.1016/J.MRREV.2017.11.002

Chenau, J., Michelland, S., Sidibe, J., Seve, M., 2008. Peptides OFFGEL electrophoresis: a suitable pre-analytical step for complex eukaryotic samples fractionation compatible with quantitative iTRAQ labeling. Proteome Sci. 6, 9. https://doi.org/10.1186/1477-5956-6-9

Cheng, W., I-Shan, H., Chen, J.-C., 2004. Change in water temperature on the immune response of Taiwan abalone (*Haliotis diversicolor*) and its susceptibility to *Vibrio parahaemolyticus*. Fish Shellfish Immunol. 17, 235–243. https://doi.org/10.1016/j.fsi.2004.03.007

Coates, C.J., Nairn, J., 2014. Diverse immune functions of hemocyanins. Dev. Comp. Immunol. 45, 43–55. https://doi.org/10.1016/J.DCI.2014.01.021

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Conesa, A., Götz, S., 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics 2008. https://doi.org/10.1155/2008/619832

Constortium, G., 2019. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338. https://doi.org/10.1093/nar/gky1055

Cooper, T.F., De'ath, G., Fabricius, K.E., Lough, J.M., 2008. Declining coral calcification in massive Porites in two nearshore regions of the northern Great Barrier Reef. Glob. Chang. Biol. 14, 529–538. https://doi.org/10.1111/j.1365-2486.2007.01520.x

Corrales, F.J., Odriozola, L., 2020. Proteomic Analyses. Princ. Nutr. Nutr. 69–74. https://doi.org/10.1016/b978-0-12-804572-5.00009-4

Cottrell, J.S., 2011. Protein identification using MS/MS data. J. Proteomics 74, 1842–1851. https://doi.org/10.1016/J.JPROT.2011.05.014

Cowan, K.J., 2003. Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. J. Exp. Biol. 206, 1107–1115.

https://doi.org/10.1242/jeb.00220

Cox, J., Mann, M., 2007. Is Proteomics the New Genomics? Cell 130, 395–398. https://doi.org/10.1016/j.cell.2007.07.032

Cullen, P., 2008. Endosomal sorting and signalling : an emerging role for sorting nexins. Nature: Perspectives. 9, 574-582.

Cunningham, S.C., Smith, A.M., Lamare, M.D., 2016. The effects of elevated pCO_2 on growth, shell production and metabolism of cultured juvenile abalone, *Haliotis iris*. Aquac. Res. 47, 2375–2392. https://doi.org/10.1111/are.12684

Dai, Y., Xie, L., Xiong, X., Chen, L., Fan, W., Zhang, R., 2005. Cloning and characterization of a homologous Ca 2+ /calmodulin-dependent protein kinase PSKH1 from pearl oyster *Pinctada fucata*. Tsinghua Sci. Technol. 10, 504–511. https://doi.org/10.1016/S1007-0214(05)70108-5

Dalman, M.R., Deeter, A., Nimishakavi, G., Duan, Z.-H., 2012. Fold change and p-value cutoffs significantly alter microarray interpretations. BMC Bioinformatics 13, S11. https://doi.org/10.1186/1471-2105-13-S2-S11

Dang, V.T., Speck, P., Benkendorff, K., 2012. Influence of elevated temperatures on the immune response of abalone, *Haliotis rubra*. Fish Shellfish Immunol. 32, 732–740. https://doi.org/10.1016/j.fsi.2012.01.022

Darling, N.J., Cook, S.J., 2014. The role of MAPK signalling pathways in the response to endoplasmic reticulum stress. Biochim. Biophys. Acta - Mol. Cell Res. 1843, 2150–2163. https://doi.org/10.1016/J.BBAMCR.2014.01.009

Das, D., Wang, Y.-H., Hsieh, C.-Y., Suzuki, Y.J., 2016. Major vault protein regulates cell growth/survival signaling through oxidative modifications. Cell. Signal. 28, 12–18. https://doi.org/10.1016/J.CELLSIG.2015.10.007

Depledge, M.H., Fossi, M.C., 1994. The role of biomarkers in environmental assessment (2). Invertebrates. Ecotoxicology 3, 161–172. https://doi.org/10.1007/BF00117081

DeSouza, L. V., Siu, K.W.M., 2013. Mass spectrometry-based quantification. Clin. Biochem. 46, 421–431. https://doi.org/10.1016/J.CLINBIOCHEM.2012.10.025

Di Ciano-Oliveira, C., Thirone, A.C.P., Szaszi, K., Kapus, A., 2006. Osmotic stress and the

cytoskeleton: the R(h)ole of Rho GTPases. Acta Physiol. 187, 257–272. https://doi.org/10.1111/j.1748-1716.2006.01535.x

Dineshram, R., Wong, K.K.W., Xiao, S., Yu, Z., Qian, P.Y., Thiyagarajan, V., 2012. Analysis of Pacific oyster larval proteome and its response to high-CO_2. Mar. Pollut. Bull. 64, 2160–2167. https://doi.org/10.1016/j.marpolbul.2012.07.043

Dineshram, R., Thiyagarajan, V., Lane, A., Ziniu, Y., Xiao, S., Leung, P.T.Y., 2013. Elevated CO_2 alters larval proteome and its phosphorylation status in the commercial oyster, *Crassostrea hongkongensis*. Mar. Biol. 160, 2189–2205. https://doi.org/10.1007/s00227-013-2176-x

Dineshram, R., Chandramouli, K., Ko, G.W.K., Zhang, H., Qian, P.-Y., Ravasi, T., Thiyagarajan, V., 2016. Quantitative analysis of oyster larval proteome provides new insights into the effects of multiple climate change stressors. Glob. Chang. Biol. 22, 2054–2068. https://doi.org/10.1111/gcb.13249

Diz-Muñoz, A., Fletcher, D.A., Weiner, O.D., 2013. Use the force: Membrane tension as an organizer of cell shape and motility. Trends Cell Biol. https://doi.org/10.1016/j.tcb.2012.09.006

Donaghy, L., Hong, H.K., Lambert, C., Park, H.S., Shim, W.J., Choi, K.S., 2010. First characterisation of the populations and immune-related activities of hemocytes from two edible gastropod species, the disk abalone, *Haliotis discus discus* and the spiny top shell, *Turbo cornutus*. Fish Shellfish Immunol. 28, 87–97. https://doi.org/10.1016/j.fsi.2009.10.006

Donaghy, L., Kraffe, E., Le Goïc, N., Lambert, C., Volety, A.K., Soudant, P., 2012. Reactive Oxygen Species in Unstimulated Hemocytes of the Pacific Oyster *Crassostrea gigas*: A Mitochondrial Involvement. PLoS One 7, 1–10. https://doi.org/10.1371/journal.pone.0046594

Doney, S.C., Fabry, V.J., Feely, R.A., Kleypas, J.A., 2009. Ocean Acidification: The Other CO_2 Problem. Ann. Rev. Mar. Sci. 1, 169–192. https://doi.org/10.1146/annurev.marine.010908.163834

Dowd, W.W., 2012. Challenges for Biological Interpretation of Environmental Proteomics Data in Non-model Organisms. Integr. Comp. Biol. 52, 705–720.

https://doi.org/10.1093/icb/ics093

Drozdov, K.A., Drozdov, A.L., 2016. Anaerobic glycolysis in sea urchins as adaptation to life in a habitat lacking oxygen. Biol. Bull. 43, 517–520. https://doi.org/10.1134/S1062359016060066

Duan, Y., Liu, P., Li, Jitao, Li, Jian, Chen, P., 2013. Immune gene discovery by expressed sequence tag (EST) analysis of hemocytes in the ridgetail white prawn *Exopalaemon carinicauda.* Fish Shellfish Immunol. 34, 173–182. https://doi.org/10.1016/J.FSI.2012.10.026

English, A.R., Voeltz, G.K., 2013. Rab10 GTPase regulates ER dynamics and morphology. Nat. Cell Biol. 15, 169–178. https://doi.org/10.1038/ncb2647

Enricuso, O.B., Conaco, C., Sayco, S.L.G., Neo, M.L., Cabaitan, P.C., 2019. Elevated seawater temperatures affect embryonic and larval development in the giant clam *Tridacna gigas* (Cardiidae: Tridacninae). J. Molluscan Stud. 85, 66–72. https://doi.org/10.1093/mollus/eyy051

Epelboin, Y., Quintric, L., Guévélou, E., Boudry, P., Pichereau, V., Corporeau, C., 2016. The Kinome of Pacific Oyster *Crassostrea gigas*, Its Expression during Development and in Response to Environmental Factors. PLoS One 11, e0155435. https://doi.org/10.1371/journal.pone.0155435

Eravci, M., Sommer, C., Selbach, M., 2014. IPG Strip-Based Peptide Fractionation for Shotgun Proteomics. Humana Press, New York, NY, pp. 67–77. https://doi.org/10.1007/978-1-4939-0685-7_5

Ernoult, E., Gamelin, E., Guette, C., 2008. Improved proteome coverage by using iTRAQ labelling and peptide OFFGEL fractionation. Proteome Sci. 6, 27. https://doi.org/10.1186/1477-5956-6-27

Ernoult, E., Guette, C., 2011. Neuroproteomics 57, 145–158. https://doi.org/10.1007/978-1-61779-111-6

Ertl, N.G., O'Connor, W.A., Papanicolaou, A., Wiegand, A.N., Elizur, A., 2016a. Transcriptome analysis of the Sydney rock oyster, *Saccostrea glomerata*: Insights into molluscan immunity. PLoS One 11, 1–33. https://doi.org/10.1371/journal.pone.0156649

Ertl, N.G., O'Connor, W.A., Wiegand, A.N., Elizur, A., 2016b. Molecular analysis of the

Sydney rock oyster (*Saccostrea glomerata*) CO_2 stress response. Clim. Chang. Responses 3, 6. https://doi.org/10.1186/s40665-016-0019-y

Ertl, N.G., O'Connor, W.A., Elizur, A., 2019. Molecular effects of a variable environment on Sydney rock oysters, *Saccostrea glomerata*: Thermal and low salinity stress, and their synergistic effect. Mar. Genomics 43, 19–32. https://doi.org/10.1016/J.MARGEN.2018.10.003

Faber, M.A., Pascal, M., El Kharbouchi, O., Sabato, V., Hagendorens, M.M., Decuyper, I.I., Bridts, C.H., Ebo, D.G., 2017. Shellfish allergens: tropomyosin and beyond. Allergy Eur. J. Allergy Clin. Immunol. 72, 842–848. https://doi.org/10.1111/all.13115

Fabregat, A., Sidiropoulos, K., Viteri, G., Forner, O., Marin-Garcia, P., Arnau, V., D'Eustachio, P., Stein, L., Hermjakob, H., 2017. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 18, 142. https://doi.org/10.1186/s12859-017-1559-2

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2018. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 46, D649–D655. https://doi.org/10.1093/nar/gkx1132

Fader, C.M., Colombo, M.I., 2009. Autophagy and multivesicular bodies: two closely related partners. Cell Death Differ. 16, 70–78. https://doi.org/10.1038/cdd.2008.168

Fehsenfeld, S., Kiko, R., Appelhans, Y., Towle, D.W., Zimmer, M., Melzner, F., 2011. Effects of elevated seawater pCO_2 on gene expression patterns in the gills of the green crab, *Carcinus maenas*. BMC Genomics 12, 488. https://doi.org/10.1186/1471-2164-12-488

Feng, Z., Prentice, R., Srivastava, S., 2004. Research issues and strategies for genomic and proteomic biomarker discovery and validation: a statistical perspective. Pharmacogenomics 5, 709–719. https://doi.org/10.1517/14622416.5.6.709

Figueiredo, D.A.L., Branco, P.C., dos Santos, D.A., Emerenciano, A.K., Iunes, R.S., Shimada Borges, J.C., Machado Cunha da Silva, J.R., 2016. Ocean acidification affects parameters of immune response and extracellular pH in tropical sea urchins *Lytechinus*

variegatus and *Echinometra luccunter*. Aquat. Toxicol. 180, 84–94. https://doi.org/10.1016/J.AQUATOX.2016.09.010

Forsgren, E., Dupont, S., Jutfelt, F., Amundsen, T., 2013. Elevated CO_2 affects embryonic development and larval phototaxis in a temperate marine fish. Ecol. Evol. 3, 3637–3646. https://doi.org/10.1002/ece3.709

Fortunato, H., 2015. Mollusks: Tools in Environmental and Climate Research. Am. Malacol. Bull. 33, 310–324. https://doi.org/10.4003/006.033.0208

Franchini, P., van der Merwe, M., Roodt-Wilding, R., 2011. Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. BMC Res. Notes 4, 59. https://doi.org/10.1186/1756-0500-4-59

Freitas, R., De Marchi, L., Bastos, M., Moreira, A., Velez, C., Chiesa, S., Wrona, F.J., Figueira, E., Soares, A.M.V.M., 2017. Effects of seawater acidification and salinity alterations on metabolic, osmoregulation and oxidative stress markers in *Mytilus galloprovincialis*. Ecol. Indic. 79, 54–62. https://doi.org/10.1016/j.ecolind.2017.04.003

Fuller, H.R., Morris, G.E., 2012. Quantitative Proteomics Using iTRAQ Labeling and Mass Spectrometry. Integr. Proteomics 442. https://doi.org/10.5772/31469

Gan, C.S., Chong, P.K., Pham, T.K., Wright, P.C., 2007. Technical , Experimental , and Biological Variations in Isobaric Tags for Relative and Absolute Quantitation (iTRAQ) research articles. J. Proteome Res. 6, 821–827.

Gao, Y., Yates, J.R., 2019. Protein Analysis by Shotgun Proteomics. Mass Spectrom. Chem. Proteomics 1–38. https://doi.org/10.1002/9781118970195.ch1

García-Campos, M.A., Espinal-Enríquez, J., Hernández-Lemus, E., 2015. Pathway Analysis: State of the Art. Front. Physiol. 6, 383. https://doi.org/10.3389/fphys.2015.00383

García-Giménez, J.L., Romá-Mateo, C., Pallardó, F. V., 2019. Oxidative post-translational modifications in histones. BioFactors biof.1532. https://doi.org/10.1002/biof.1532

Gazeau, F., Parker, L.M., Comeau, S., Gattuso, J.P., O'Connor, W.A., Martin, S., Pörtner, H.O., Ross, P.M., 2013. Impacts of ocean acidification on marine shelled molluscs. Mar. Biol. 160, 2207–2245. https://doi.org/10.1007/s00227-013-2219-3

Gene Ontology Consortium: going forward, 2015. . Nucleic Acids Res. 43, D1049–D1056.

https://doi.org/10.1093/nar/gku1179

Gomiero, A., Bellerby, R.G.J., Manca Zeichen, M., Babbini, L., Viarengo, A., 2018. Biological responses of two marine organisms of ecological relevance to on-going ocean acidification and global warming. Environ. Pollut. 236, 60–70. https://doi.org/10.1016/J.ENVPOL.2018.01.063

Goncalves, P., Anderson, K., Thompson, E.L., Melwani, A., Parker, L.M., Ross, P.M., Raftos, D.A., 2016. Rapid transcriptional acclimation following transgenerational exposure of oysters to ocean acidification. Mol. Ecol. 25, 4836–4849. https://doi.org/10.1111/mec.13808

Goncalves, P., Jones, D.B., Thompson, E.L., Parker, L.M., Ross, P.M., Raftos, D.A., 2017. Transcriptomic profiling of adaptive responses to ocean acidification. Mol. Ecol. 26, 5974–5988. https://doi.org/10.1111/mec.14333

Götz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435. https://doi.org/10.1093/nar/gkn176

Griffen, B., Belgrad, B., Cannizzo, Z., Knotts, E., Hancock, E., 2016. Rethinking our approach to multiple stressor studies in marine environments. Mar. Ecol. Prog. Ser. 543, 273–281. https://doi.org/10.3354/meps11595

Griffith, A.W., Harke, M.J., DePasquale, E., Berry, D.L., Gobler, C.J., 2019. The harmful algae, Cochlodinium polykrikoides and Aureococcus anophagefferens, elicit stronger transcriptomic and mortality response in larval bivalves (Argopecten irradians) than climate change stressors. Ecol. Evol. 9, 4931–4948. https://doi.org/10.1002/ece3.5100

Guerra, F., Bucci, C., 2016. Multiple Roles of the Small GTPase Rab7. Cells 5, 34. https://doi.org/10.3390/cells5030034

Gunderson, A.R., Armstrong, E.J., Stillman, J.H., 2016. Multiple Stressors in a Changing World: The Need for an Improved Perspective on Physiological Responses to the Dynamic Marine Environment. Ann. Rev. Mar. Sci. 8, 357–378. https://doi.org/10.1146/annurev-marine-122414-033953

Guo, X., He, Y., Zhang, L., Lelong, C., Jouaux, A., 2015. Immune and stress responses in

oysters with insights on adaptation. Fish Shellfish Immunol. 46, 107–119. https://doi.org/10.1016/J.FSI.2015.05.018

Hamer, B., Hamer, D.P., Müller, W.E.., Batel, R., 2004. Stress-70 proteins in marine mussel *Mytilus galloprovincialis* as biomarkers of environmental pollution: a field study. Environ. Int. 30, 873–882. https://doi.org/10.1016/J.ENVINT.2004.02.008

Hamp, T., Rost, B., 2015. Evolutionary profiles improve protein–protein interaction prediction from sequence. Bioinformatics 31, 1945–1950. https://doi.org/10.1093/bioinformatics/btv077

Han, X., Aslanian, A., Yates, J.R., 2008. Mass spectrometry for proteomics. Curr. Opin. Chem. Biol. 12, 483–490. https://doi.org/10.1016/J.CBPA.2008.07.024

Handler, D.C., Pascovici, D., Mirzaei, M., Gupta, V., Salekdeh, G.H., Haynes, P.A., 2018. The Art of Validating Quantitative Proteomics Data. Proteomics 18, 1800222. https://doi.org/10.1002/pmic.201800222

Harney, E., Dubief, B., Boudry, P., Basuyaux, O., Schilhabel, M.B., Huchette, S., Paillard, C., Nunes, F.L.D., 2016. De novo assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome. Mar. Genomics 28, 11–16. https://doi.org/10.1016/J.MARGEN.2016.03.002

Harvey, B.P., Gwynn-Jones, D., Moore, P.J., 2013. Meta-analysis reveals complex marine biological responses to the interactive effects of ocean acidification and warming. Ecol. Evol. 3, 1016–1030. https://doi.org/10.1002/ece3.516

Hasanuzzaman, A.F.M., Rubiolo, J.A., Robledo, D., Gómez-Tato, A., Álvarez-Dios, J.A., Fernández-Boo, S., Cao, A., Villalba, A., Pardo, B.G., Martínez, P., 2018. Gene expression analysis of *Ruditapes philippinarum* haemocytes after experimental *Perkinsus olseni* zoospore challenge and infection in the wild. Fish Shellfish Immunol. 72, 611–621. https://doi.org/10.1016/J.FSI.2017.11.033

Hedman, A.C., Smith, J.M., Sacks, D.B., 2015. The biology of IQGAP proteins: beyond the cytoskeleton. EMBO Rep. 16, 427–446. https://doi.org/10.15252/embr.201439834

Henson, S.A., Beaulieu, C., Ilyina, T., John, J.G., Long, M., Séférian, R., Tjiputra, J., Sarmiento, J.L., 2017. Rapid emergence of climate change in environmental drivers of marine ecosystems. Nat. Commun. 8, 14682. https://doi.org/10.1038/ncomms14682

Hoeflich, K.P., Ikura, M., 2004. Radixin: cytoskeletal adopter and signaling protein. Int. J. Biochem. Cell Biol. 36, 2131–2136. https://doi.org/10.1016/J.BIOCEL.2003.11.018

Hoegh-Guldberg, O., Mumby, P.J., Hooten, A.J., Steneck, R.S., Greenfield, P., Gomez, E., Harvell, C.D., Sale, P.F., Edwards, A.J., Caldeira, K., Knowlton, N., Eakin, C.M., Iglesias-Prieto, R., Muthiga, N., Bradbury, R.H., Dubi, A., Hatziolos, M.E., 2007. Coral Reefs Under Rapid Climate Change and Ocean Acidification. Science (80-.). 318, 1737–1742. https://doi.org/10.1126/science.1152509

Hooper, C., Day, R., Slocombe, R., Handlinger, J., Benkendorff, K., 2007. Stress and immune responses in abalone: Limitations in current knowledge and investigative methods based on other models. Fish Shellfish Immunol. 22, 363–379. https://doi.org/10.1016/j.fsi.2006.06.009

Hörth, P., Miller, C.A., Preckel, T., Wenz, C., 2006. Efficient Fractionation and Improved Protein Identification by Peptide OFFGEL Electrophoresis. Mol. Cell. Proteomics 5, 1968–1974. https://doi.org/10.1074/mcp.T600037-MCP200

Hu, P., He, X., Zhu, C., Guan, W., Ma, Y., 2014. Cloning and characterization of a ribosomal protein L23a gene from Small Tail Han sheep by screening of a cDNA expression library. Meta Gene 2, 479–488. https://doi.org/10.1016/J.MGENE.2014.06.005

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13. https://doi.org/10.1093/nar/gkn923

Huang, Y., Sun, J., Wang, L., Song, L., 2018. The unfolded protein response signaling pathways in molluscs 183–196.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Hüning, A.K., Melzner, F., Thomsen, J., Gutowska, M.A., Krämer, L., Frickenhaus, S., Rosenstiel, P., Pörtner, H.-O., Philipp, E.E.R., Lucassen, M., 2013. Impacts of seawater acidification on mantle gene expression patterns of the Baltic Sea blue mussel:

implications for shell formation and energy metabolism. Mar. Biol. 160, 1845–1861. https://doi.org/10.1007/s00227-012-1930-9

Ilyin, S.E., Belkowski, S.M., Plata-Salamán, C.R., 2004. Biomarker discovery and validation: technologies and integrative approaches. Trends Biotechnol. 22, 411–416. https://doi.org/10.1016/J.TIBTECH.2004.06.005

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

Johnson, J.G., Paul, M.R., Kniffin, C.D., Anderson, P.E., Burnett, L.E., Burnett, K.G., 2015. High CO2 alters the hypoxia response of the pacific whiteleg shrimp (*Litopenaeus vannamei*) transcriptome including known and novel hemocyanin isoforms. Physiol. Genomics 47, 548–558. https://doi.org/10.1152/physiolgenomics.00031.2015

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462. https://doi.org/10.1093/nar/gkv1070

Kaniewska, P., Campbell, P.R., Kline, D.I., Rodriguez-Lanetty, M., Miller, D.J., Dove, S., Hoegh-Guldberg, O., 2012. Major Cellular and Physiological Impacts of Ocean Acidification on a Reef Building Coral. PLoS One 7, e34659. https://doi.org/10.1371/journal.pone.0034659

Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., Hester, S. V., Lilley, K.S., 2010. Addressing Accuracy and Precision Issues in iTRAQ Quantitation. Mol. Cell. Proteomics 9, 1885–1897. https://doi.org/10.1074/mcp.M900628-MCP200

Kaufmann, M., 2006. The Role of the COG Database in Comparative and Functional Genomics 10.

Kemp, I.K., Coyne, V.E., 2011. Identification and characterisation of the Mpeg1 homologue in the South African abalone, *Haliotis midae*. Fish Shellfish Immunol. 31, 754–764. https://doi.org/10.1016/j.fsi.2011.07.010

Keskin, O., Tuncbag, N., Gursoy, A., 2016. Predicting Protein-Protein Interactions from the Molecular to the Proteome Level. Chem. Rev. 116, 4884–4909. https://doi.org/10.1021/acs.chemrev.5b00683

Khatri, P., Sirota, M., Butte, A.J., 2012. Ten Years of Pathway Analysis: Current Approaches

and Outstanding Challenges. PLoS Comput. Biol. 8, e1002375. https://doi.org/10.1371/journal.pcbi.1002375

Kinoshita, M., 2006. Diversity of septin scaffolds. Curr. Opin. Cell Biol. 18, 54–60. https://doi.org/10.1016/j.ceb.2005.12.005

Koh, H.Y., Lee, J.H., Han, S.J., Park, H., Shin, S.C., Lee, S.G., 2015. A transcriptomic analysis of the response of the arctic pteropod *Limacina helicina* to carbon dioxide-driven seawater acidification. Polar Biol. 38, 1727–1740. https://doi.org/10.1007/s00300-015-1738-4

Kultz, D., Fiol, D., Valkova, N., Gomez-Jimenez, S., Chan, S.Y., Lee, J., 2007. Functional genomics and proteomics of the cellular osmotic stress response in `non-model' organisms. J. Exp. Biol. 210, 1593–1601. https://doi.org/10.1242/jeb.000141

Kumar, C., Mann, M., 2009. Bioinformatics analysis of mass spectrometry-based proteomics data sets. FEBS Lett. 583, 1703–1712. https://doi.org/10.1016/j.febslet.2009.03.035

Kumar, N., Leonzino, M., Hancock-Cerutti, W., Horenkamp, F.A., Li, P., Lees, J.A., Wheeler, H., Reinisch, K.M., De Camilli, P., 2018. VPS13A and VPS13C are lipid transport proteins differentially localized at ER contact sites. J. Cell Biol. 217, 3625–3639. https://doi.org/10.1083/jcb.201807019

Lagos, N.A., Benítez, S., Duarte, C., Lardies, M.A., Broitman, B.R., Tapia, C., Tapia, P., Widdicombe, S., Vargas, C.A., 2016. Effects of temperature and ocean acidification on shell characteristics of *Argopecten purpuratus*: Implications for scallop aquaculture in an upwelling-influenced area. Aquac. Environ. Interact. 8, 357–370. https://doi.org/10.3354/AEI00183

Lamond, A.I., Uhlen, M., Horning, S., Makarov, A., Robinson, C. V., Serrano, L., Hartl, F.U., Baumeister, W., Werenskiold, A.K., Andersen, J.S., Vorm, O., Linial, M., Aebersold, R., Mann, M., 2012. Advancing Cell Biology Through Proteomics in Space and Time (PROSPECTS). Mol. Cell. Proteomics 11, O112.017731. https://doi.org/10.1074/mcp.O112.017731

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. https://doi.org/10.1186/1471-2105-9-559

Lannig, G., Eilers, S., Pörtner, H.O., Sokolova, I.M., Bock, C., Lannig, G., Eilers, S., Pörtner,

H.O., Sokolova, I.M., Bock, C., 2010. Impact of Ocean Acidification on Energy Metabolism of Oyster, *Crassostrea gigas*—Changes in Metabolic Pathways and Thermal Response. Mar. Drugs 8, 2318–2339. https://doi.org/10.3390/md8082318

Latosinska, A., Vougas, K., Makridakis, M., Klein, J., Mullen, W., Abbas, M., Stravodimos, K., Katafigiotis, I., Merseburger, A.S., Zoidakis, J., Mischak, H., Vlahou, A., Jankowski, V., 2015. Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis. PLoS One 10, e0137048. https://doi.org/10.1371/journal.pone.0137048

Law, R.H.P., Zhang, Q., McGowan, S., Buckle, A.M., Silverman, G.A., Wong, W., Rosado, C.J., Langendorf, C.G., Pike, R.N., Bird, P.I., Whisstock, J.C., 2006. An overview of the serpin superfamily. Genome Biol. https://doi.org/10.1186/gb-2006-7-5-216

Lee, A.S., 2005. The ER chaperone and signaling regulator GRP78/BiP as a monitor of endoplasmic reticulum stress. Methods 35, 373–381. https://doi.org/10.1016/J.YMETH.2004.10.010

Li, H., Ericsson, M., Rabasha, B., Budnik, B., Wagner, B., Garraway, L.A., Schreiber, S.L., 2018. 6-Phosphogluconate Dehydrogenase Links Cytosolic Carbohydrate Metabolism to Protein Secretion. bioRxiv 480129. https://doi.org/10.1101/480129

Li, S., Huang, J., Liu, C., Liu, Y., Zheng, G., Xie, L., Zhang, R., 2016a. Interactive Effects of Seawater Acidification and Elevated Temperature on the Transcriptome and Biomineralization in the Pearl Oyster *Pinctada fucata*. Environ. Sci. Technol. 50, 1157–1165. https://doi.org/10.1021/acs.est.5b05107

Li, S., Liu, C., Huang, J., Liu, Y., Zhang, S., Zheng, G., Xie, L., Zhang, R., 2016b. Transcriptome and biomineralization responses of the pearl oyster *Pinctada fucata* to elevated CO_2 and temperature. Sci. Rep. 6, 18943. https://doi.org/10.1038/srep18943

Li, Z., Blissard, G., 2015. The vacuolar protein sorting genes in insects: A comparative genome view. Insect Biochem. Mol. Biol. 62, 211–225. https://doi.org/10.1016/J.IBMB.2014.11.007

Liebler, D.C., Zimmerman, L.J., 2013. Targeted Quantitation of Proteins by Mass Spectrometry. https://doi.org/10.1021/bi400110b

Lim, H.J., Kim, B.M., Hwang, I.J., Lee, J.S., Choi, I.Y., Kim, Y.J., Rhee, J.S., 2016. Thermal

stress induces a distinct transcriptome profile in the Pacific oyster *Crassostrea gigas.* Comp. Biochem. Physiol. - Part D Genomics Proteomics 19, 62–70. https://doi.org/10.1016/j.cbd.2016.06.006

Liu, W., Huang, X., Lin, J., He, M., 2012. Seawater acidification and elevated temperature affect gene expression patterns of the pearl oyster *Pinctada fucata.* PLoS One 7. https://doi.org/10.1371/journal.pone.0033679

Lockwood, B.L., Somero, G.N., 2011. Transcriptomic responses to salinity stress in invasive and native blue mussels (genus *Mytilus*). Mol. Ecol. 20, 517–529. https://doi.org/10.1111/j.1365-294X.2010.04973.x

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G., 2003. PEAKS: powerful software for peptidede novo sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17, 2337–2342. https://doi.org/10.1002/rcm.1196

Malagoli, D., Casarini, L., Sacchi, S., Ottaviani, E., 2007. Stress and immune response in the mussel *Mytilus galloprovincialis*. Fish Shellfish Immunol. 23, 171–177. https://doi.org/10.1016/J.FSI.2006.10.004

Maldonado-Aguayo, W., Lafarga-De la Cruz, F., Gallardo-Escárate, C., 2014. Identification and expression of antioxidant and immune defense genes in the surf clam *Mesodesma donacium* challenged with *Vibrio anguillarum*. Mar. Genomics 19, 65–73. https://doi.org/10.1016/j.margen.2014.11.006

Malham, S.K., Lacoste, A., Gélébart, F., Cueff, A., Poulet, S.A., 2003. Evidence for a direct link between stress and immunity in the mollusc *Haliotis tuberculata*. J. Exp. Zool. Part A Comp. Exp. Biol. 295A, 136–144. https://doi.org/10.1002/jez.a.10222

Martínez-Quintana, J.A., Yepiz-Plascencia, G., 2012. Glucose and other hexoses transporters in marine invertebrates : A mini review 15. https://doi.org/10.2225/vol15-issue5-fulltext-12

Martyniuk, C.J., Alvarez, S., Denslow, N.D., 2012. DIGE and iTRAQ as biomarker discovery tools in aquatic toxicology. Ecotoxicol. Environ. Saf. 76, 3–10. https://doi.org/10.1016/j.ecoenv.2011.09.020

Mateus, D., Marini, E.S., Progida, C., Bakke, O., 2018. Rab7a modulates ER stress and ER

morphology. Biochim. Biophys. Acta - Mol. Cell Res. 1865, 781–793. https://doi.org/10.1016/j.bbamcr.2018.02.011

Matozzo, V., Mari, M.G., 2011. Bivalve immune responses and climate changes: is there a relationship? Invertebr. Surviv. J. 8, 70–77.

Mayne, J., Ning, Z., Zhang, X., Starr, A.E., Chen, R., Deeke, S., Chiang, C.-K., Xu, B., Wen, M., Cheng, K., Seebun, D., Star, A., Moore, J.I., Figeys, D., 2015. Bottom-Up Proteomics (2013−2015): Keeping up in the Era of Systems Biology. https://doi.org/10.1021/acs.analchem.5b04230

McNulty, D.E., Li, Z., White, C.D., Sacks, D.B., Annan, R.S., 2011. MAPK scaffold IQGAP1 binds the EGF receptor and modulates its activation. J. Biol. Chem. 286, 15010–15021. https://doi.org/10.1074/jbc.M111.227694

Melatunan, S., Calosi, P., Rundle, S.D., Moody, A.J., Widdicombe, S., 2011. Exposure to Elevated Temperature and Pco 2 Reduces Respiration Rate and Energy Status in the Periwinkle Littorina littorea. Physiol. Biochem. Zool. 84, 583–594. https://doi.org/10.1086/662680

Melwani, A.R., Thompson, E.L., Raftos, D.A., 2016. Differential proteomic response of Sydney rock oysters (*Saccostrea glomerata*) to prolonged environmental stress. Aquat. Toxicol. 173, 53–62. https://doi.org/10.1016/J.AQUATOX.2016.01.003

Melzner, F., Mark, F.C., Seibel, B.A., Tomanek, L., 2020. Ocean Acidification and Coastal Marine Invertebrates: Tracking CO_2 Effects from Seawater to the Cell . Ann. Rev. Mar. Sci. 12, 1–25. https://doi.org/10.1146/annurev-marine-010419-010658

Méndez-López, L., Hellman, U., Ibarguren, I., Villamarín, J.A., 2012. Filamin isoforms in molluscan smooth muscle. Biochim. Biophys. Acta - Proteins Proteomics 1824, 1334–1341. https://doi.org/10.1016/j.bbapap.2012.07.011

Meseck, S.L., Alix, J.H., Swiney, K.M., Long, W.C., Wikfors, G.H., Foy, R.J., 2016. Ocean Acidification Affects Hemocyte Physiology in the Tanner Crab (*Chionoecetes bairdi*). PLoS One 11, e0148477. https://doi.org/10.1371/journal.pone.0148477

Misra, B.B., Langefeld, C., Olivier, M., Cox, L.A., 2019. Integrated omics: tools, advances and future approaches. J. Mol. Endocrinol. R21–R45. https://doi.org/10.1530/jme-18-0055

Moore, R.E., Kirwan, J., Doherty, M.K., Whitfield, P.D., 2017. Biomarker Discovery in Animal Health and Disease: The Application of Post-Genomic Technologies. Biomark. Insights 2, 117727190700200. https://doi.org/10.1177/117727190700200040

Morash, A.J., Alter, K., 2016. Effects of environmental and farm stress on abalone physiology: perspectives for abalone aquaculture in the face of global climate change. Rev. Aquac. 8, 342–368. https://doi.org/10.1111/raq.12097

Morga, B., Arzul, I., Faury, N., Segarra, A., Chollet, B., Renault, T., 2011. Molecular responses of *Ostrea edulis* haemocytes to an in vitro infection with *Bonamia ostreae.* Dev. Comp. Immunol. 35, 323–333. https://doi.org/10.1016/j.dci.2010.10.005

Mortensen, S.H., Glette, J., 1996. Phagocytic activity of scallop (*Pecten maximus*) haemocytes maintained in vitro. Fish Shellfish Immunol. 6, 111–121. https://doi.org/10.1006/fsim.1996.0012

Moya, A., Howes, E.L., Lacoue-Labarthe, T., Forêt, S., Hanna, B., Medina, M., Munday, P.L., Ong, J.-S., Teyssié, J.-L., Torda, G., Watson, S.-A., Miller, D.J., Bijma, J., Gattuso, J.-P., 2016. Near-future pH conditions severely impact calcification, metabolism and the nervous system in the pteropod *Heliconoides inflatus*. Glob. Chang. Biol. 22, 3888–3900. https://doi.org/10.1111/gcb.13350

Mukherjee, S., Bhunia, A.S., Bhunia, N.S., Ray, M., Ray, S., 2016. Immunomodulatory effects of temperature and pH of water in an Indian freshwater sponge. J. Therm. Biol. 59, 1–12. https://doi.org/10.1016/J.JTHERBIO.2016.04.005

Muñoz-Braceras, S., Tornero-Écija, A.R., Vincent, O., Escalante, R., 2019. VPS13A is closely associated with mitochondria and is required for efficient lysosomal degradation. Dis. Model. Mech. 12, dmm036681. https://doi.org/10.1242/dmm.036681

Navarro, J.M., Duarte, C., Manríquez, P.H., Lardies, M.A., Torres, R., Acuña, K., Vargas, C.A., Lagos, N.A., 2016. Ocean warming and elevated carbon dioxide: multiple stressor impacts on juvenile mussels from southern Chile. ICES J. Mar. Sci. J. du Cons. 73, 764–771. https://doi.org/10.1093/icesjms/fsv249

Nikapitiya, C., Dorrington, T., 2013. The role of histones in the immune responses of aquatic invertebrates. ISJ 10, 94–101.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R.,

Resing, K.A., Ahn, N.G., 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol. Cell. Proteomics 4, 1487–1502. https://doi.org/10.1074/mcp.M500084-MCP200

Opas, M., Szewczenko-Pawlikowski, M., Jass, G.K., Mesaeli, N., Michalak, M., 1996. Calreticulin modulates cell adhesiveness via regulation of vinculin expression. J. Cell Biol. 135, 1913–1923. https://doi.org/10.1083/jcb.135.6.1913

Orr, J.C., Fabry, V.J., Aumont, O., Bopp, L., Doney, S.C., Feely, R.A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R.M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R.G., Plattner, G.-K., Rodgers, K.B., Sabine, C.L., Sarmiento, J.L., Schlitzer, R., Slater, R.D., Totterdell, I.J., Weirig, M.-F., Yamanaka, Y., Yool, A., 2005. Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. Nature 437, 681–686. https://doi.org/10.1038/nature04095

Pancholi, V., 2001. Multifunctional α-enolase: its role in diseases. Cell. Mol. Life Sci. 58, 902-920.

Pandey, a, Mann, M., 2000. Proteomics to study genes and genomes. Nature 405, 837–846. https://doi.org/10.1038/35015709

Pappireddi, N., Martin, L., Wühr, M., 2019. A Review on Quantitative Multiplexed Proteomics. ChemBioChem 20, 1210–1224. https://doi.org/10.1002/cbic.201800650

Parry, D.A.D., Squire, J.M., 2017. Fibrous Proteins: Structures and Mechanisms, Subcellular Biochemistry. https://doi.org/10.1007/978-3-319-49674-0

Pascual, C., Mascaro, M., Rodríguez-Canul, R., Gallardo, P., Sánchez, A.A., Rosas, C., Cruz-López, H., 2019. Sea Surface Temperature Modulates Physiological and Immunological Condition of *Octopus maya*. Front. Physiol. 10. https://doi.org/10.3389/fphys.2019.00739

Pathmanathan, S., Hamilton, E., Atcheson, E., Timson, D.J., 2011. The interaction of IQGAPs with calmodulin-like proteins. Biochem. Soc. Trans. 39, 694–699. https://doi.org/10.1042/BST0390694

Pei, G., Chen, L., Zhang, W., 2017. WGCNA Application to Proteomic and Metabolomic Data Analysis. Methods Enzymol. 585, 135–158.

https://doi.org/10.1016/BS.MIE.2016.09.016

Perera, N.C.N., Godahewa, G.I., Lee, J., 2016. Mitogen-activated protein kinase 1 from disk abalone (*Haliotis discus discus*): Roles in early development and immunity-related transcriptional responses. https://doi.org/10.1016/j.fsi.2016.10.031

Picciotto, M.R., Zoli, M., Bertuzzi, G., Nairn, A.C., 1995. Immunochemical localization of calcium/calmodulin-dependent protein kinase I. Synapse 20, 75–84. https://doi.org/10.1002/syn.890200111

Pila, E.A., Sullivan, J.T., Wu, X.Z., Fang, J., Rudko, S.P., Gordy, M.A., Hanington, P.C., 2016. Haematopoiesis in molluscs: A review of haemocyte development and function in gastropods, cephalopods and bivalves. https://doi.org/10.1016/j.dci.2015.11.010

Piñeiro, C., Cañas, B., Carrera, M., 2010. The role of proteomics in the study of the influence of climate change on seafood products. Food Res. Int. 43, 1791–1802. https://doi.org/10.1016/j.foodres.2009.11.012

Pino, L.K., Searle, B.C., Bollinger, J.G., Nunn, B., Maclean, B., Maccoss, M.J., 2017. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. Mass Spectrom. Rev. https://doi.org/10.1002/mas.21540

Pinto, M.C.X., Kihara, A.H., Goulart, V.A.M., Tonelli, F.M.P., Gomes, K.N., Ulrich, H., Resende, R.R., 2015. Calcium signaling and cell proliferation. Cell. Signal. 27, 2139–2149. https://doi.org/10.1016/J.CELLSIG.2015.08.006

Raiborg, C., Stenmark, H., 2009. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. Nature 458, 445–452. https://doi.org/10.1038/nature07961

Reum, J.C.P., Alin, S.R., Feely, R.A., Newton, J., Warner, M., McElhany, P., 2014. Seasonal Carbonate Chemistry Covariation with Temperature, Oxygen, and Salinity in a Fjord Estuary: Implications for the Design of Ocean Acidification Experiments. PLoS One 9, e89619. https://doi.org/10.1371/journal.pone.0089619

Rodrigues, P.M., Campos, A., Kuruvilla, J., Schrama, D., Cristobal, S., 2017. Proteomics in Aquaculture : Quality and Safety, Proteomics in Food Science. Elsevier Inc. https://doi.org/10.1016/B978-0-12-804007-2.00017-5

Rodríguez-Ortega, M.J., Grøsvik, B.E., Rodríguez-Ariza, A., Goksøyr, A., López-Barea, J.,

2003. Changes in protein expression profiles in bivalve molluscs (*Chamaelea gallina*) exposed to four model environmental pollutants. Proteomics 3, 1535–1543. https://doi.org/10.1002/pmic.200300491

Roggatz, C.C., Lorch, M., Hardege, J.D., Benoit, D.M., 2016. Ocean acidification affects marine chemical communication by changing structure and function of peptide signalling molecules. Glob. Chang. Biol. 22, 3914–3926. https://doi.org/10.1111/gcb.13354

Ruan, K., Duan, J., Bai, F., Lemaire, M., Ma, X., Bai, L., 2009. Function of *Dunaliella salina* (Dunaliellaceae) enolase and its expression during stress. Eur. J. Phycol. 44, 207–214. https://doi.org/10.1080/09670260802573105

Rudolph, J., Cox, J., 2018. A network module for the Perseus software for computational proteomics facilitates proteome interaction graph analysis. https://doi.org/10.1101/447268

Russeth, K.P., Higgins, L., Andrews, M.T., 2006. Identification of proteins from non-model organisms using mass spectrometry: Application to a hibernating mammal. J. Proteome Res. 5, 829–839. https://doi.org/10.1021/pr050306a

Salameh, A.I., Ruffin, V.A., Boron, W.F., 2014. Effects of metabolic acidosis on intracellular pH responses in multiple cell types. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 307, R1413–R1427. https://doi.org/10.1152/ajpregu.00154.2014

Sales, J., Britz, P.J., 2001. Research on abalone (*Haliotis midae* L.) cultivation in South Africa. Aquac. Res. 32, 863–874. https://doi.org/10.1046/j.1365-2109.2001.00629.x

Samaranayake, H.S., Cowan, A.E., Klobutcher, L.A., 2011. Vacuolar protein sorting protein 13A, TtVPS13A, localizes to the *Tetrahymena thermophila* phagosome membrane and is required for efficient phagocytosis. Eukaryot. Cell 10, 1207–1218. https://doi.org/10.1128/EC.05089-11

Samuel, C.L., Norma Yolanda, H.S., Salvador Emilio, L.C., Pedro, C.H., Felipe De Jesús, A.V., María Teresa, S., 2019. Survival and respiration of green abalone (*Haliotis fulgens*) facing very short-term marine environmental extremes. Mar. Freshw. Behav. Physiol. 52, 1–15. https://doi.org/10.1080/10236244.2019.1607734

Sandin, M., Chawade, A., Levander, F., 2015. Is label-free LC-MS/MS ready for biomarker

discovery? Proteomics - Clin. Appl. 9, 289–294. https://doi.org/10.1002/prca.201400202

Sarkar, A., Ray, D., Shrivastava, A.N., Sarker, S., 2006. Molecular Biomarkers: Their significance and application in marine pollution monitoring. Ecotoxicology 15, 333–340. https://doi.org/10.1007/s10646-006-0069-1

Shalit, T., Elinger, D., Savidor, A., Gabashvili, A., Levin, Y., 2015. MS1-based label-free proteomics using a quadrupole orbitrap mass spectrometer. J. Proteome Res. 14, 1979–1986. https://doi.org/10.1021/pr501045t

Shannon, P., Markiel, A., Owen Ozier, 2, Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2498–2504. https://doi.org/10.1101/gr.1239303.metabolite

Shi, J., Fu, M., Zhao, C., Zhou, F., Yang, Q., Qiu, L., 2016. Characterization and function analysis of Hsp60 and Hsp10 under different acute stresses in black tiger shrimp, *Penaeus monodon*. Cell Stress Chaperones 21, 295–312. https://doi.org/10.1007/s12192-015-0660-6

Sjölander, K., 2004. Phylogenomic inference of protein molecular function: Advances and challenges. Bioinformatics 20, 170–179. https://doi.org/10.1093/bioinformatics/bth021

Slattery, M., Ankisetty, S., Corrales, J., Marsh-Hunkin, K.E., Gochfeld, D.J., Willett, K.L., Rimoldi, J.M., 2012. Marine proteomics: A critical assessment of an emerging technology. J. Nat. Prod. 75, 1833–1837. https://doi.org/10.1021/np300366a

Sokolova, I.M., 2009. Apoptosis in molluscan immune defense, ISJ. 6, 49-58. ISSN1824-307X

Sokolova, I.M., 2013. Energy-Limited Tolerance to Stress as a Conceptual Framework to Integrate the Effects of Multiple Stressors. Integr. Comp. Biol. 53, 597–608. https://doi.org/10.1093/icb/ict028

Sokolova, I., 2018. Mitochondrial Adaptations to Variable Environments and Their Role in Animals' Stress Tolerance. Integr. Comp. Biol. 58, 519–531. https://doi.org/10.1093/icb/icy017

Stauffer, D.R., Howard, T.L., Nyun, T., Hollenberg, S.M., 2001. CHMP1 is a novel nuclear matrix protein affecting chromatin structure and cell-cycle progression. J. Cell Sci. 114,

2383–93.

Stevens, A.M., Gobler, C.J., 2018. Interactive effects of acidification, hypoxia, and thermal stress on growth, respiration, and survival of four North Atlantic bivalves. Mar. Ecol. Prog. Ser. 604, 143–161. https://doi.org/10.3354/meps12725

Stevenson, S.E., Chu, Y., Ozias-Akins, P., Thelen, J.J., 2009. Validation of gel-free, label-free quantitative proteomics approaches: Applications for seed allergen profiling. J. Proteomics 72, 555–566. https://doi.org/10.1016/j.jprot.2008.11.005

Stone, D.A.J., Harris, J.O., Wang, H., Mercer, G.J., Schaefer, E.N., Bansemer, M.S., 2013. Dietary Protein Level and Water Temperature Interactions for Greenlip Abalone *Haliotis laevigata*. J. Shellfish Res. 32, 119–130. https://doi.org/10.2983/035.032.0118

Stone, D.A.J., Bansemer, M.S., Lange, B., Schaefer, E.N., Howarth, G.S., Harris, J.O., 2014. Dietary intervention improves the survival of cultured greenlip abalone (*Haliotis laevigata* Donovan) at high water temperature. Aquaculture 430, 230–240. https://doi.org/10.1016/J.AQUACULTURE.2014.03.047

Storey, K.B., Wu, C.W., 2013. Stress response and adaptation: A new molecular toolkit for the 21st century. Comp. Biochem. Physiol. - A Mol. Integr. Physiol. https://doi.org/10.1016/j.cbpa.2013.01.019

Stumpp, M., Dupont, S., Thorndyke, M.C., Melzner, F., 2011. CO_2 induced seawater acidification impacts sea urchin larval development II: Gene expression patterns in pluteus larvae. Comp. Biochem. Physiol. - A Mol. Integr. Physiol. 160, 320–330. https://doi.org/10.1016/j.cbpa.2011.06.023

Sun, Y., Zhang, L., Zhang, M., Li, R., Li, Y., Hu, X., Wang, S., Bao, Z., 2016. Characterization of three mitogen-activated protein kinases (MAPK) genes reveals involvement of ERK and JNK, not p38 in defense against bacterial infection in Yesso scallop *Patinopecten yessoensis.* Fish Shellfish Immunol. 54, 507–515. https://doi.org/10.1016/J.FSI.2016.04.139

Surtida, A. 2000. Abalone. *SEAFDEC* Asian Aquaculture. 22, 14-15 & 28.

Szklarczyk, D., Jensen, L.J., 2015. Protein-Protein Interaction Databases. Humana Press, New York, NY, pp. 39–56. https://doi.org/10.1007/978-1-4939-2425-7_3

Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A.,

Doncheva, N.T., Roth, A., Bork, P., Jensen, L.J., von Mering, C., 2017. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 45, D362–D368. https://doi.org/10.1093/nar/gkw937

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., Mering, C. von, 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607–D613. https://doi.org/10.1093/nar/gky1131

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E. V, 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–6.

Teasdale, R.D., Collins, B.M., 2012. Insights into the PX (phox-homology) domain and SNX (sorting nexin) protein families: structures, functions and roles in disease. Biochem. J. 441, 39–59. https://doi.org/10.1042/bj20111226

Thompson, E.L., O'Connor, W., Parker, L., Ross, P., Raftos, D.A., 2015. Differential proteomic responses of selectively bred and wild-type Sydney rock oyster populations exposed to elevated CO_2. Mol. Ecol. 24, 1248–1262. https://doi.org/10.1111/mec.13111

Timmins-Schiffman, E., Nunn, B.L., Goodlett, D.R., Roberts, S.B., 2013. Shotgun proteomics as a viable approach for biological discovery in the Pacific oyster. Conserv. Physiol 1. https://doi.org/10.1093/conphys/cot009

Timmins-Schiffman, E, Coffey, W.D., Hua, W., Nunn, B.L., Dickinson, G.H., Roberts, S.B., 2014. Shotgun proteomics reveals physiological response to ocean acidification in *Crassostrea gigas*. Conserv. Physiol. 1, 1–59. https://doi.org/10.7287/peerj.preprints.388v1

Timmins-Schiffman, E., Guzmán, J.M., Elliott, R., Vadopalas, B., Roberts, S.B., 2019. Dynamic response in the larval geoduck clam proteome to elevated pCO_2. bioRxiv 613018. https://doi.org/10.1101/613018

Tomanek, L., Zuzow, M.J., 2010. The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. J. Exp. Biol.

https://doi.org/10.1242/jeb.041228

Tomanek, L., 2011. Environmental Proteomics: Changes in the Proteome of Marine Organisms in Response to Environmental Stress, Pollutants, Infection, Symbiosis, and Development. Ann. Rev. Mar. Sci. 3, 373–399. https://doi.org/10.1146/annurev-marine-120709-142729

Tomanek, L., Zuzow, M.J., Ivanina, A. V, Beniash, E., Sokolova, I.M., 2011. Proteomic response to elevated pCO_2 level in eastern oysters, *Crassostrea virginica*: evidence for oxidative stress. J. Exp. Biol. 214, 1836–44. https://doi.org/10.1242/jeb.055475

Tomanek, L., 2014. Proteomics to study adaptations in marine organisms to environmental stress. J. Proteomics 105, 92–106. https://doi.org/10.1016/J.JPROT.2014.04.009

Troell, M., Robertson-Andersson, D., Anderson, R.J., Bolton, J.J., Maneveldt, G., Halling, C., Probyn, T., 2006. Abalone farming in South Africa: An overview with perspectives on kelp resources, abalone feed, potential for on-farm seaweed production and socio-economic importance. Aquaculture 257, 266–281. https://doi.org/10.1016/j.aquaculture.2006.02.066

Truebano, M., Diz, A.P., Thorne, M.A.S., Clark, M.S., Skibinski, D.O.F., 2013. Proteome response to heat stress in the Antarctic clam *Laternula elliptica*. J. Integr. OMICS 3, 34–43. https://doi.org/10.5584/jiomics.v3i1.125

Tyanova, S., Temu, T., Cox, J., 2016a. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat. Protoc. 11, 2301–2319. https://doi.org/10.1038/nprot.2016.136

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J., 2016b. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat. Methods 13, 731–740. https://doi.org/10.1038/nmeth.3901

Tyanova, S., Cox, J., 2018. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. Humana Press, New York, NY, pp. 133–148. https://doi.org/10.1007/978-1-4939-7493-1_7

Udayantha, H.M.V., Godahewa, G.I., Bathige, S.D.N.K., Wickramaarachchi, W.D.N., Umasuthan, N., De Zoysa, M., Jeong, H.-B., Lim, B.-S., Lee, J., 2016. A molluscan calreticulin ortholog from *Haliotis discus discus*: Molecular characterization and

transcriptional evidence for its role in host immunity. Biochem. Biophys. Res. Commun. 474, 43–50. https://doi.org/10.1016/J.BBRC.2016.04.056

UniProt Consortium, The, 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 46, 2699–2699. https://doi.org/10.1093/nar/gky092

Unwin, R.D., 2010. Quantification of Proteins by iTRAQ. Humana Press, Totowa, NJ, pp. 205–215. https://doi.org/10.1007/978-1-60761-780-8_12

Urbé, S., 2005. Ubiquitin and endocytic protein sorting. Essays Biochem. 41, 81. https://doi.org/10.1042/eb0410081

Vaibhav, V., Thompson, E.L., Raftos, D.A., Haynes, P.A., 2018. Potential protein biomarkers of QX disease resistance in selectively bred Sydney Rock Oysters. Aquaculture 495, 144–152. https://doi.org/10.1016/J.AQUACULTURE.2018.05.035

Valderrama, F., Thevapala, S., Ridley, A.J., 2012. Radixin regulates cell migration and cell-cell adhesion through Rac1. J. Cell Sci. 125, 3310–3319. https://doi.org/10.1242/jcs.094383

Vasquez, M.C., Lippert, M.R., White, C., Walter, R.K., Tomanek, L., 2019. Proteomic changes across a natural temperature gradient in a marine gastropod. Mar. Environ. Res. 149, 137–147. https://doi.org/10.1016/J.MARENVRES.2019.06.002

Vega, V.L., Charles, W., De Maio, A., 2010. A new feature of the stress response: Increase in endocytosis mediated by Hsp70. Cell Stress Chaperones 15, 517–527. https://doi.org/10.1007/s12192-009-0165-2

Venter, L., Du, ·, Loots, T., Lodewyk, ·, Mienie, J., Peet, ·, Jansen Van Rensburg, J., Mason, S., Vosloo, A., Lindeque, J.Z., 123AD. Uncovering the metabolic response of abalone (*Haliotis midae*) to environmental hypoxia through metabolomics. Metabolomics 1, 49. https://doi.org/10.1007/s11306-018-1346-8

Vertommen, A., Panis, B., Swennen, R., Carpentier, S.C., 2011. Challenges and solutions for the identification of membrane proteins in non-model plants. J. Proteomics 74, 1165–1181. https://doi.org/10.1016/J.JPROT.2011.02.016

Vincenzetti, S., Felici, A., Ciarrocchi, G., Pucciarelli, S., Ricciutelli, M., Ariani, A., Polzonetti, V., Polidori, P., 2017. Comparative proteomic analysis of two clam species: *Chamelea gallina* and *Tapes philippinarum*. Food Chem. 219, 223–229.

https://doi.org/10.1016/J.FOODCHEM.2016.09.150

Visudtiphole, V., Watthanasurorot, A., Klinbunga, S., Menasveta, P., Kirtikara, K., 2010. Molecular characterization of Calreticulin: A biomarker for temperature stress responses of the giant tiger shrimp *Penaeus monodon*. Aquaculture 308. https://doi.org/10.1016/j.aquaculture.2010.06.040

Wäge, J., Lerebours, A., Hardege, J.D., Rotchell, J.M., 2016. Exposure to low pH induces molecular level changes in the marine worm, Platynereis dumerilii. Ecotoxicol. Environ. Saf. 124, 105–110. https://doi.org/10.1016/j.ecoenv.2015.10.008

Wang, H., Zhang, H., Wong, Y.H., Voolstra, C., Ravasi, T., B. Bajic, V., Qian, P.-Y., 2010. Rapid transcriptome and proteome profiling of a non-model marine invertebrate, *Bugula neritina.* Proteomics 10, 2972–2981. https://doi.org/10.1002/pmic.201000056

Wang, L., Song, X., Song, L., 2018. The oyster immunity. Dev. Comp. Immunol. 80, 99–118. https://doi.org/10.1016/J.DCI.2017.05.025

Wang, L.L., Yang, C., Song, L.S., 2013. The molluscan HSP70s and their expression in hemocytes. Isj 77–83.

Wang, Q., Cao, R., Ning, X., You, L., Mu, C., Wang, C., Wei, L., Cong, M., Wu, H., Zhao, J., 2016. Effects of ocean acidification on immune responses of the Pacific oyster *Crassostrea gigas.* Fish Shellfish Immunol. 49, 24–33. https://doi.org/10.1016/J.FSI.2015.12.025

Wang, X., Li, E., Chen, L., Wang, X., Li, E., Chen, L., 2016. A Review of Carbohydrate Nutrition and Metabolism in Crustaceans A Review of Carbohydrate Nutrition and Metabolism in Crustaceans 2055. https://doi.org/10.1080/15222055.2016.1141129

Wang, X., Li, E., Xu, Z., Li, T., Xu, C., Chen, L., 2017. Molecular response of carbohydrate metabolism to dietary carbohydrate and acute low salinity stress in pacific white shrimp *Litopenaeus vannamei*. Turkish J. Fish. Aquat. Sci. 17, 153–169. https://doi.org/10.4194/1303-2712-v17_1_18

Wang, X., Kuruc, M., 2019. Functional Proteomics Methods and Protocols Methods in Molecular Biology 1871. Springer Science + Business Media, LLC, New York.

Wei, L., Wang, Q., Ning, X., Mu, C., Wang, C., Cao, R., Wu, H., Cong, M., Li, F., Ji, C., Zhao, J., 2015a. Combined metabolome and proteome analysis of the mantle tissue from

Pacific oyster *Crassostrea gigas* exposed to elevated pCO_2. https://doi.org/10.1016/j.cbd.2014.12.001

Wei, L., Wang, Q., Wu, H., Ji, C., Zhao, J., 2015b. Proteomic and metabolomic responses of Pacific oyster Crassostrea gigas to elevated pCO_2 exposure. J. Proteomics 112, 83–94. https://doi.org/10.1016/j.jprot.2014.08.010

Wernberg, T., Russell, B.D., Moore, P.J., Ling, S.D., Smale, D.A., Campbell, A., Coleman, M.A., Steinberg, P.D., Kendrick, G.A., Connell, S.D., 2011. Impacts of climate change in a global hotspot for temperate marine biodiversity and ocean warming. J. Exp. Mar. Bio. Ecol. 400, 7–16. https://doi.org/10.1016/J.JEMBE.2011.02.021

Westermeier, R., Naven, T., 2002. Proteomics in Practice Related title from Wiley-VCH Electrophoresis in Practice Third Edition, Analysis.

Wiese, S., Reidegeld, K.A., Meyer, H.E., Warscheid, B., 2007. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. Proteomics 7, 340–350. https://doi.org/10.1002/pmic.200600422

Worby, C.A., Dixon, J.E., 2002. Sorting out the cellular functions of sorting nexins. Nat. Rev. Mol. Cell Biol. 3, 919–931. https://doi.org/10.1038/nrm974

Wu, F., Xie, Z., Lan, Y., Dupont, S., Sun, M., Cui, S., Huang, X., Huang, W., Liu, L., Hu, M., Lu, W., Wang, Y., 2018. Short-Term Exposure of *Mytilus coruscus* to Decreased pH and Salinity Change Impacts Immune Parameters of Their Haemocytes. Front. Physiol. 9, 166. https://doi.org/10.3389/fphys.2018.00166

Wu, W.W., Wang, G., Baek, S.J., Shen, R., 2006. Comparative Study of Three Proteomic Quantitative Methods , DIGE , cICAT , and iTRAQ , Using 2D Gel- or LC - MALDI TOF / TOF research articles. J. Proteome Res. 5, 651–658. https://doi.org/10.1021/pr050405o

Wu, X., Hasan, M. Al, Chen, J.Y., 2014. Pathway and network analysis in proteomics. J. Theor. Biol. 362, 44–52. https://doi.org/10.1016/J.JTBI.2014.05.031

Xu, D., Sun, L., Liu, S., Zhang, L., Ru, X., Zhao, Y., Yang, H., 2014. Molecular cloning of heat shock protein 10 (Hsp10) and 60 (Hsp60) cDNAs and their expression analysis under thermal stress in the sea cucumber *Apostichopus japonicus.* Comp. Biochem. Physiol. Part - B Biochem. Mol. Biol. 171, 49–57.

https://doi.org/10.1016/j.cbpb.2014.03.009

Xu, M., Huang, J., Shi, Y., Zhang, H., He, M., 2019. Comparative transcriptomic and proteomic analysis of yellow shell and black shell pearl oysters, *Pinctada fucata martensii.* BMC Genomics 20. https://doi.org/10.1186/s12864-019-5807-x

Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics 11, M111.010587. https://doi.org/10.1074/mcp.M111.010587

Zhang, L., Li, L., Guo, X., Litman, G.W., Dishaw, L.J., Zhang, G., 2015. Massive expansion and functional divergence of innate immune genes in a protostome. Sci. Rep. 5. https://doi.org/10.1038/srep08693

Zhang, P., Li, C., Zhang, Peng, Jin, C., Pan, D., Bao, Y., 2014. ITRAQ-based proteomics reveals novel members involved in pathogen challenge in sea cucumber Apostichopus japonicus. PLoS One 9. https://doi.org/10.1371/journal.pone.0100492

Zhang, X., Shi, J., Sun, Y., Habib, Y.J., Yang, H., Zhang, Z., Wang, Y., 2019. Integrative transcriptome analysis and discovery of genes involving in immune response of hypoxia/thermal challenges in the small abalone Haliotis diversicolor. Fish Shellfish Immunol. 84, 609–626. https://doi.org/10.1016/J.FSI.2018.10.044

Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., Yates, J.R., 2013. Protein Analysis by Shotgun/Bottom-up Proteomics. https://doi.org/10.1021/cr3003533

Zhang, Y., Sun, J., Mu, H., Li, J., Zhang, Yuehuan, Xu, F., Xiang, Z., Qian, P., Qiu, J., Yu, Z., 2015. Proteomic Basis of Stress Responses in the Gills of the Pacific Oyster *Crassostrea gigas*.

Zhang, Y., Storey, K.B., 2018. Life in Suspended Animation: Role of Chaperone Proteins in Vertebrate and Invertebrate Stress Adaptation. pp. 95–137. https://doi.org/10.1007/978-3-319-74715-6_5

Zhao, X., Yu, H., Kong, L., Li, Q., 2012. Transcriptomic Responses to Salinity Stress in the Pacific Oyster *Crassostrea gigas.* PLoS One 7, e46244. https://doi.org/10.1371/journal.pone.0046244

Zhao, X., Shi, W., Han, Y., Liu, S., Guo, C., Fu, W., Chai, X., Liu, G., 2017. Ocean acidification adversely influences metabolism, extracellular pH and calcification of an economically important marine bivalve, *Tegillarca granosa.* Mar. Environ. Res. 125, 82–89. https://doi.org/10.1016/J.MARENVRES.2017.01.007

Zhou, J., Wang, W.N., He, W.Y., Zheng, Y., Wang, L., Xin, Y., Liu, Y., Wang, A.L., 2010. Expression of HSP60 and HSP70 in white shrimp, *Litopenaeus vannamei* in response to bacterial challenge. J. Invertebr. Pathol. 103, 170–178. https://doi.org/10.1016/j.jip.2009.12.006

Zhuang, J., Coates, C.J., Zhu, H., Zhu, P., Wu, Z., Xie, L., 2015. Identification of candidate antimicrobial peptides derived from abalone hemocyanin. Dev. Comp. Immunol. 49, 96–102. https://doi.org/10.1016/j.dci.2014.11.008

Web based sources

Earth System Research Laboratory, n.d. *Trends in Atmospheric Carbon Dioxide*. National Oceanic and Atmospheric Administration, United States of America, accessed 7 May 2018. https://www.esrl.noaa.gov/gmd/ccgg/trends/

Venn diagram generator: http://bioinformatics.psb.ugent.be/webtools/Venn/ Date accessed: 7/12/2019

World Population Trends, UNFPA. Last updated 29 August 2017. Date accessed: 21 December 2019. <https://www.unfpa.org/world-population-trends>

Somewhere, something incredible is waiting to be known – Carl Sagan