# Discovering Novel Lignocellulose Degrading Enzymes from the Marine Wood Borer, *Limnoria quadripunctata*

William Scott Eborall

PhD

University of York

Biology

September 2013

# Abstract

Transportation accounts for a large proportion of global $CO_2$ emissions, estimated at 22% of all anthropogenically produced $CO_2$ in 2010, with most of this coming from the use of liquid transportation fuels. Second generation biofuels offer a sustainable opportunity to decarbonise this sector of global energy use. To realise the potential of second generation biofuels more efficient methods of deconstructing their lignocellulosic starting material must be developed. Understanding how the marine wood borer, *Limnoria quadripunctata*, is able to do this biochemically without the assistance of microbial symbionts may inspire new techniques to achieve this in an industrial setting.

Whilst the anatomy of *Limnoria* has been well studied, and the transcriptome of its hepatopancreas (HP), a secretory organ of the digestive system, has been previously elucidated, little is known about how the animal is able to derive nutrients from its diet of wood without the aid of microbial symbionts. In this work a proteomic study of the HP and gut tissue of *Limnoria* was carried out and the data generated analysed in concert with that of the HP transcriptome to identify proteins which may be involved in lignocellulose digestion in the animal. In this way the glycosyl hydrolases, hemocyanins (HCs), ferritins and leucine rich repeat proteins were identified as having the potential to be involved in lignocellulose digestion.

Work was undertaken to characterise the function of HC proteins in *Limnoria*. Reverse transcriptase quantitative polymerase chain reaction analysis showed all five *Limnoria* HC genes to be solely expressed in the HP, whilst Western blot and proteomic examination showed HC protein to be present in the HP and gut, as well as the rest of the body. This indicates that HCs are transported post translation to the gut, the site where ingested wood particles are segregated and presumably digested. *In vivo* experiments suggested that a phenoloxidase enzyme activity was associated with the production of peroxides in the gut of *Limnoria*. Using a soluble extract of *Limnoria* tissue a di-phenoloxidase (DPO) enzyme activity was demonstrated which it was possible to inhibit using a phenoloxidase inhibitor. Mass spectrometry analysis implicated HC protein as the species responsible for this activity. Multiple sequence alignment showed that the *Limnoria* HC sequences possess the conserved features associated with the ability to carry out a DPO enzyme activity described in HC proteins from other arthropod species.

Attempts were made to heterologously express *Limnoria* HC proteins in a number of systems with limited success being achieved using the ArcticExpress strain of *Escherichia coli*. Partially pure heterologous protein produced in this way was able to show a DPO enzyme activity which was inhibited by a phenoloxidase inhibitor. Mass spectrometry analysis implicates HC as the protein species responsible for this activity.

Based on the findings of this work it seems possible that the hemocyanin proteins of *Limnoria* aid the animal in digesting wood by contributing to the formation of reactive peroxide compounds in its gut which attack lignocellulose. This process may also contribute towards the sterility of the gut by creating an inhospitable environment for microbial colonisation. Further work is now required to determine the mechanisms by which it occurs.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgements

Despite the purpose of a PhD being to train a young scientist to become an independent researcher, science is never a solitary process. Throughout the course of this work many people contributed help and assistance, either in a direct way by helping me to carry out the work, offering advice or indirectly by encouraging me to keep going and helping me to maintain my sanity.

I'd like to thank my supervisors Prof. Neil Bruce and Prof. Simon McQueen-Mason for putting their faith in me and putting me forwards for the project in the first place. I thank them for all of their advice and support during the four years and will remember our drunken discussions fondly. I'd also like to thank my sponsor, the BBSRC, for providing the funding which allowed me to carry out this work.

I would like to thank Dr. Katrin Beßer, Mrs. Luisa Elias and Dr. Marcelo Kern for their help and support in and out of the lab. It was a pleasure to work with them on the Gribble project and I hope I wasn't too much of a burden to them with my silly questions and terrible memory! They taught me a lot of things over the years and we shared the despair that is *Limnoria* protein expression on a regular basis. I would like to thank Dr. Jared Cartwright of the Protein Production laboratory at York for letting me work in his lab for several months and for always being willing to give advice. I must also acknowledge my gratitude to Dr. Graham Malyon and Dr. Simon Cragg. Graham provided me with a seemingly never ending supply of dissected Gribble tissue without hesitation and was a genuinely lovely person to meet and discuss ideas with. Simon Cragg was a great source of support and advice during this project and through many beer fuelled conversations acquainted me with the anatomy and physiology of *Limnoria* which was a great help.

I also extend my thanks to the many members of the Bruce and McQueen-Mason groups who I had the pleasure to work with over the course of the project and who provided much support and friendship: Drs. Joseph Bennett, Federico Sabbadin, Andy Taylor, Chun Shiong Chong, Astrid Lorenz, Leonardo Gomez, Shaza Mohamed, Susannah Bird, Anna Alessi, Hazel Housden, Clare Steele-King and Dana Sabir, Emily Johnston and Margaret Cafferky. Joe and Fede helped me with my chemistry problems in a lab full of biologists. Andy, Chong, Dana, Hazel, Sue, Anna and Emily were always there to make me laugh and to try and offer advice when things didn't go to plan. Astrid reviewed many of my interim reports and cheered me up when I was

feeling low. Shaza worked with me to try and express HC proteins and shared in my bafflement and despair when it didn't work. Leo provided guidance into the world of oxidative chemistry. Margaret and Clare kept me organised and on track over the whole process.

I really do believe that a lab full of scientists doesn't work properly without dedicated support technicians looking after us. There are days when scientists struggle to get dressed properly without assistance let alone keep a lab stocked and tidy and with equipment that works. Two hard-working people made sure that this happened in my lab and they were Dave Neale and Louise Haigh. I thank them for all of the little things that they did every day to make my job easier and making me laugh so often.

Part of this work wouldn't have been possible without my mother in law Celia and brother in law Jim obliging me with a very strange request. They provided me with several milk bottles full of sea water from the Bristol Channel that allowed me to keep a colony of *Limnoria* alive in the lab in York which formed the basis of many experiments.

I want to thank my father in law Steve for always showing interest in my work, even when I didn't, and for offering me honest and objective advice on issues in my life and career and for helping to put things into perspective.

None of this work would have been possible without the lifetime of help, support and love I was given by my mum, dad and sister. They have supported me through everything I've done and have provided me help in every way to put me where I am today. Words can't really express enough thanks to them so I won't try any further.

Finally I thank my beautiful wife Katherine. Completing a PhD is a very difficult thing to do, but sharing a life with someone completing a PhD is just as difficult. In a way Kat and I put our lives on hold for me to do this and there is no way I can thank her enough for that. She has been nothing but supportive and caring and without her I really would have given up half way through.

# Author’s Declaration

I declare that I am the sole author of the work in this dissertation and that it is original except where indicated by special reference in the text. No part of the dissertation has been submitted for any other degree to any other institution.

# Chapter 1: Introduction

As the world's supplies of oil are depleted, mankind must look ahead to find a source of fuel which will allow it to sustain its ever growing energy consumption (1). A large portion of the energy consumed by humans is used for transportation, and most of this energy is in the form of liquid transportation fuels (1). The use of these liquid fuels results in the release of greenhouse gases such as carbon dioxide ($CO_2$). In 2010, 22 % of the $CO_2$ released worldwide was from transportation (Figure 1.1) (2). Many novel ways of powering transport are being developed across the world including electric battery powered cars, hydrogen powered fuel cells and even wind powered-kite assisted super tankers (www.skysails.info). However, even if some of these new methods of providing energy became widely adopted, almost all of them will require the building of a completely new distribution infrastructure.

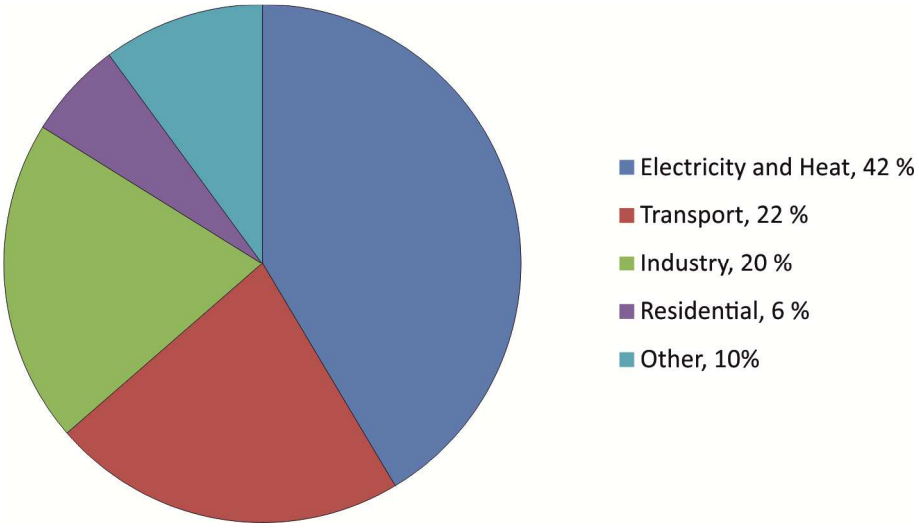



**Figure 1.1. Sources of global $CO_2$ emissions.**
Other includes commercial/public services, agriculture/forestry, fishing, energy industries other than electricity and heat generation and other sources which do not fall into the listed categories. Figure reproduced from (2).

Until one or more of these technologies reach maturation and become widely adopted we will still require liquid transportation fuels for our current fleet of vehicles. Biofuels offer us a way to displace some of the fossil-oil based fuels we use with "drop-in" replacement fuels that are compatible with our current transportation systems and fuel distribution networks.

Broadly, biofuels are defined as any source of fuel derived from recently living material and for which the $CO_2$ released during its use could theoretically be re-sequestered by living matter which could itself by processed into a fuel creating a closed loop system. The simplest example of a biofuel is wood. Wood is made by a plant from $CO_2$, water and light, and when it is burnt it releases $CO_2$, water and heat. The $CO_2$ and water released can then be used by another plant to create more wood, closing the cycle and demonstrating the term given to biofuels; renewable energy. Examples of more advanced biofuels are methane ($CH_4$), produced by anaerobically fermenting any organic matter; bio-ethanol, ethanol produced by the fermentation of plant derived sugars; bio-diesel, a mixture of long chain esters produced by the transesterification of fats from plant or animal sources. A recent comprehensive review of biofuels is given in (3).

## 1.1 First Generation Biofuel Production

First generation biofuels are produced using pre-existing technologies with sugar, starch or oil derived from the edible parts of plants or animals as a feedstock. In this section biogas produced via anaerobic digestion will be considered a first generation biofuel technology. Although anaerobic digestion is capable of utilising non-edible parts of food crops, and indeed certain lightly lignified non-food crops (such as grasses) it is considered a first generation technology as the microbes involved in the process are unable to deconstruct lignin. This prevents the technique from being able to use all of the potential nutrients present in the feedstock. In order to fully utilise all of the gas potential of the non-edible biomass to make their utilisation economical digester residence must be longer than for other feedstocks.

### 1.1.1 Bioethanol

Bioethanol is identical to ethanol; the only difference is that it has been produced from a biomass source rather than from a petrochemical source. Bioethanol is made by using microorganisms (bacteria, yeasts, fungi) to anaerobically ferment sugars into ethanol. In the case of first generation biofuels the sugars used in the fermentation are derived from the edible parts of plants. This could be sugar from sugar cane, sugar beet, fruits or palm juice. Starch from crops such as corn, wheat, rice, potatoes or barley could also be used as the breakdown of starch to glucose is simple to do on an industrial scale using a simple two enzyme cocktail containing amylase and glucoamylase.

Many issues surround the production of first generation bioethanol. First and foremost of these issues is the so called “food vs. fuel” debate (4, 5). If the edible part of a plant is used to make biofuel then it cannot also be used as a foodstuff. This creates extra competition for the commodity in the global market which will push up the price of that commodity potentially pricing it out of the reach of populations who need it as part of their diet. This issue could be mitigated by cultivating more of these useful crops; however, such food crops require good quality land for growth and high inputs of fertiliser. The amount of such land available not already cultivated is limited without having to change the use of existing land (i.e. by deforestation, damming). Also, the application of large amounts of fertiliser which is often of petrochemical origin can increase the amount of fossil fuel associated with the biofuel’s production and

consequently reduce the environmental benefit gained by its use over a fossil oil based transportation fuel. Finally, the edible part of food crops represents only a small fraction of the crops total biomass and so the use of only the edible part of the crop results in a large proportion of unused material from the crops cultivation.

### 1.1.2 Biodiesel

Biodiesel is a mixture of long chain alkyl esters produced by the transesterification of edible plant or animal derived fats (reviewed in (6)). In the production process fats (triacylglycerides) are transesterified with an alcohol (most commonly methanol) in an acid or base catalysed reaction resulting in glycerol and alkyl esters. If methanol is used as an alcohol the resulting esters are called fatty acid methyl esters (FAMEs). Once the glycerol has been removed from the reaction products the resulting alkyl esters can be used as a “drop-in” replacement for standard petro-diesel. The removed glycerol, once sufficiently purified, can be used in the food and cosmetic industry and used to be a valuable co-product. However, with the large increase in amounts of biodiesel produced, the value of glycerol has fallen (7, 8). Investigations are underway to use this supply of glycerol as a chemical precursor for more useful chemicals such as propanediols and epoxides (9). The fat used to produce first generation biodiesel can come from crops such as rapeseed (*Brassica napus*), soya bean (*Glycine max*), palm kernel (*Elaeis guinessnsis*) and used cooking oil.

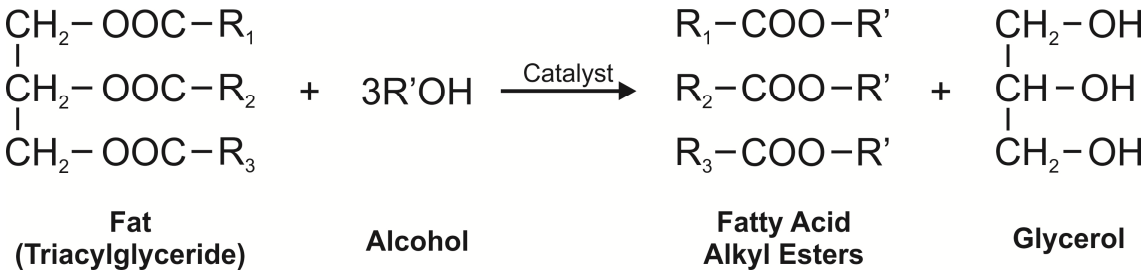



**Figure 1.2. General reaction of biodiesel production.**
The transesterification of triglycerides with an alcohol using a catalyst to form a mixture of fatty acid alkyl esters (biodiesel) and glycerol. $R_1$, $R_2$, and $R_3$, long alkyl chains; R’, short alkyl chain depending on which alcohol is used.

As has been mentioned with respect to bioethanol, biodiesel is also subject to the “food vs. fuel” debate and issues surrounding land use change and fertiliser inputs. It also wastes the non-edible parts of the crop used for its production.

### 1.1.3 Biogas

Biogas is produced by the anaerobic digestion of organic material and contains between 40 – 60 % methane. The process is reviewed in (10) and a schematic is shown in Figure 1.3. Producing methane in this way is complex and proceeds via four phases of biochemical processing namely; hydrolysis, acidogenesis, acetogenesis/dehydrogenation and methanation. Each of these phases is carried out by different communities of microorganisms. The resulting product of this digestion is biogas; a mixture of methane, carbon dioxide and trace amounts of other contaminant gases such as hydrogen sulphide and ammonia; and a nitrogen rich slurry called digestate. Biogas can be produced from very diverse types of organic matter such as sewage, animal slurry, food waste, agricultural waste (bedding, straw, etc.), food waste and non-food, bioenergy crops (such as grasses). Heavily lignified materials are avoided due to the long digester residence time required to effectively break them down (11). Anaerobic digestion can be carried out under mesophillic (35 – 42 ℃) or thermophillic (45 – 60 ℃) conditions with thermoph illic digestion resulting in larger biogas yields. Both systems require an input of energy to maintain digesters at their required temperatures; however, the amount of energy needed is small compared to that produced by the process and can often be supplied by the downstream use of produced gas (12).

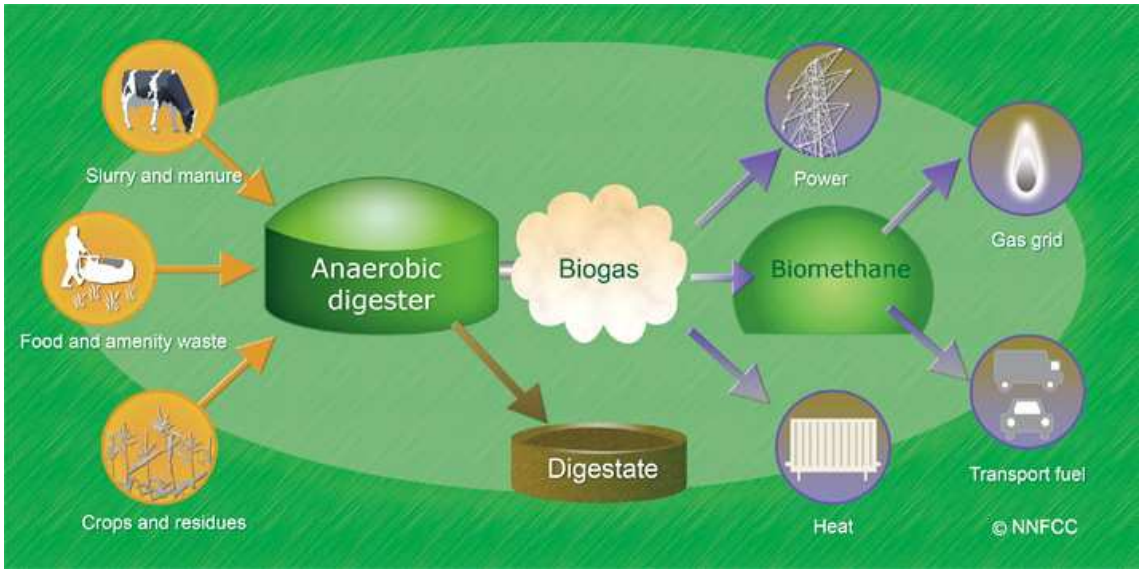



**Figure 1.3. Schematic representation of the anaerobic digestion process.** Organic material from a range of sources are mixed and placed into a digester. The digestion results in the production of biogas and digestate. Digestate can be used as a nitrogen rich fertiliser. Biogas can be used to produce heat and electricity, or it can be upgraded into biomethane to be used for heat, electricity, as a transport fuel or injected into the national gas grid. Figure is taken from www.nnfcc.co.uk.

The resulting biogas can be burnt directly as a fuel or it can be upgraded by the removal of carbon dioxide and other contaminants to produce biomethane. Biomethane is identical to petro-methane and as such can be used as a fuel, can be sold into the national grid or can be used in industrial chemistry where petro-methane would ordinarily be used.

The use of anaerobic digestion to produce biogas is not a new technology, however, it has gained popularity in recent years due to the increase in price of petro-methane (natural gas) and electricity, and the increase in cost of disposal of materials via landfill and the waste water network (10, 13). Installation of anaerobic digestion plant can save adopters money in three ways. In the case of a farm or food processing plant, anaerobic digestion can reduce the amount of waste going to landfill or being disposed of in the waste water network and therefore reduce the amount of landfill tax or waste water fees paid. The biogas (or biomethane) produced can be used as a fuel for a combined heat and power (CHP) system which generates onsite heat and electricity. The heat produced can be used to heat the digester tanks, an onsite greenhouse or a co-located industry reducing the sites heating bill. The electricity can similarly be used onsite to reduce the amount of grid electricity required. Alternatively excess biogas, biomethane or electricity can be exported to the national grid to generate an income

rather than a saving. The leftover digestate can be used as a nitrogen rich fertiliser and soil improver negating the need for traditional petrochemical synthesised fertiliser.

Unlike first generation bioethanol and biodiesel, biogas produced by anaerobic digestion need not be subject to the “food vs. fuel” debate. Although capable of using edible plant material, it can, and more often uses commercial, domestic, industrial and agricultural organic wastes as a feedstock. This allows it to use waste material that ordinarily would not be put to any other use. Indeed biogas production enables energy to be captured from waste that ordinarily would be landfilled or passed through a waste water treatment facility where it would release carbon dioxide and methane to the atmosphere (14). When an anaerobic digester and CHP facility is co-located with an industry which can use the resulting heat and energy great value can be achieved from the partnership. Scope to create these partnerships can be limited due to the creator and user of the biogas being geographically distant, or the capital cost that would be required to re-locate an industry to the site where biogas is created. When not co-located, or when the feedstock for a digester comes from an off-site source (such as municipal kerb-side food waste collection) anaerobic digestion suffers from the common requirement of all bioenergy sources which is for large freight to transport the bulky feedstock to the site of use (12).

## 1.2 Second Generation Biofuels

Second generation biofuels are produced from the inedible parts of plants. This could be the non-food parts of a food crop, agricultural/forestry residues or the whole of dedicated non-food crops such as grasses and tress. The USA believes that it alone could produce in the order of 1.3 billion dry tons of such biomass annually by the middle of this century (15). Dedicated bioenergy crops typically are fast growing, require little fertiliser input and can be grown on “marginal land,” that is land not suitable for the cultivation of more demanding food-crops (16-18). By using such feedstocks second generation biofuels bypass the “food vs. fuel”, land-use change and fertiliser issues discussed above in relation to first generation biofuels. If biofuels are to be produced on a scale that would enable them to substitute a sizeable proportion of our current fossil-oil based transportation fuels then second generation biofuels must be embraced.

To make second generation biofuels economical a major barrier must be overcome. This barrier is the difficulty in releasing the energy stored inside the lignocellulose structure of plant cell walls.

### 1.2.1 The Structure of Plant Cell Walls

Plants are able to photosynthesise, using light energy to chemically modify carbon dioxide and water into complex sugars which they can use for energy and as a building material. In-fact 30 – 50 % of a plants structural material is composed of glucose (19). This presents an abundant potential source of glucose as well as other sugars which could be used for the production of second generation bioethanol. However, plants have evolved to resist attempts to gain access to these sugars from microbes, pests and ultimately our biofuel industry. They have done this by locking up this source of sugar in a strong, waterproof, degradation recalcitrant, heterogeneous material called lignocellulose. Lignocellulose is composed of cellulose, hemicellulose and lignin. How this material is able to resist attempts to degrade it are described below. If second generation biofuels are to be realised on a global scale science must find a way to break down this barrier to sugar release.

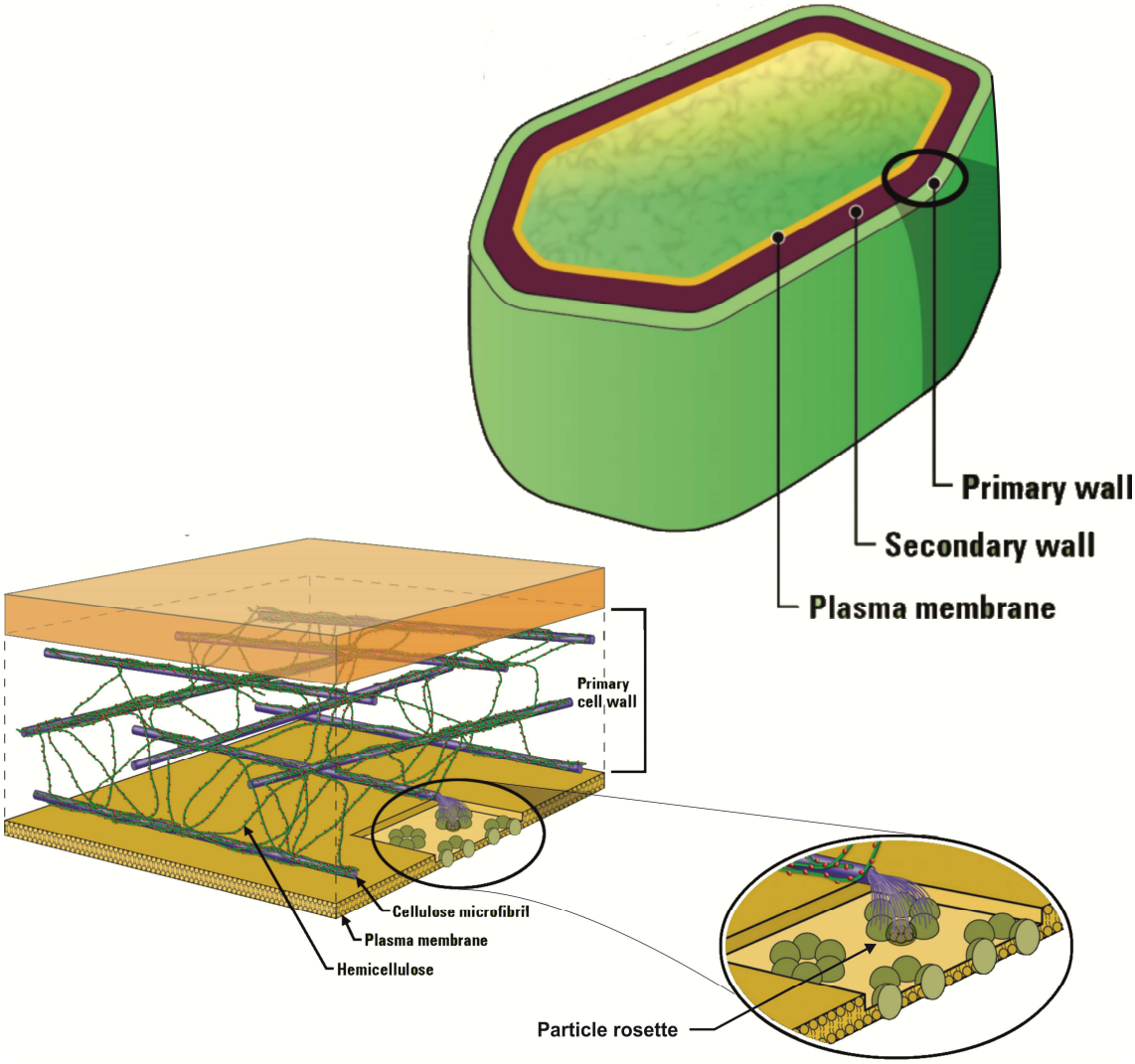



**Figure 1.4. Overview of the organisation of plant cell walls with a focus on the primary cell wall.**
Figure reproduced from (20), originally adapted from (21).

All plant cells possess a primary cell wall composed of cellulose, hemicelluloses, pectins and structural proteins (Figure 1.4) (22). Once cells have ceased to grow, some cell types which require greater mechanical strength lay down an internal secondary cell wall which is composed of cellulose, hemicelluloses and lignin; collectively called lignocellulose (Figure 1.5) (23). Secondary cell walls represent a large proportion of a plant's biomass and lignocellulose is the most abundant polymer on the planet. The recalcitrance of woody material to biological degradation is due to the structure and arrangement of the cellulose, hemicelluloses and lignin making it up (24).

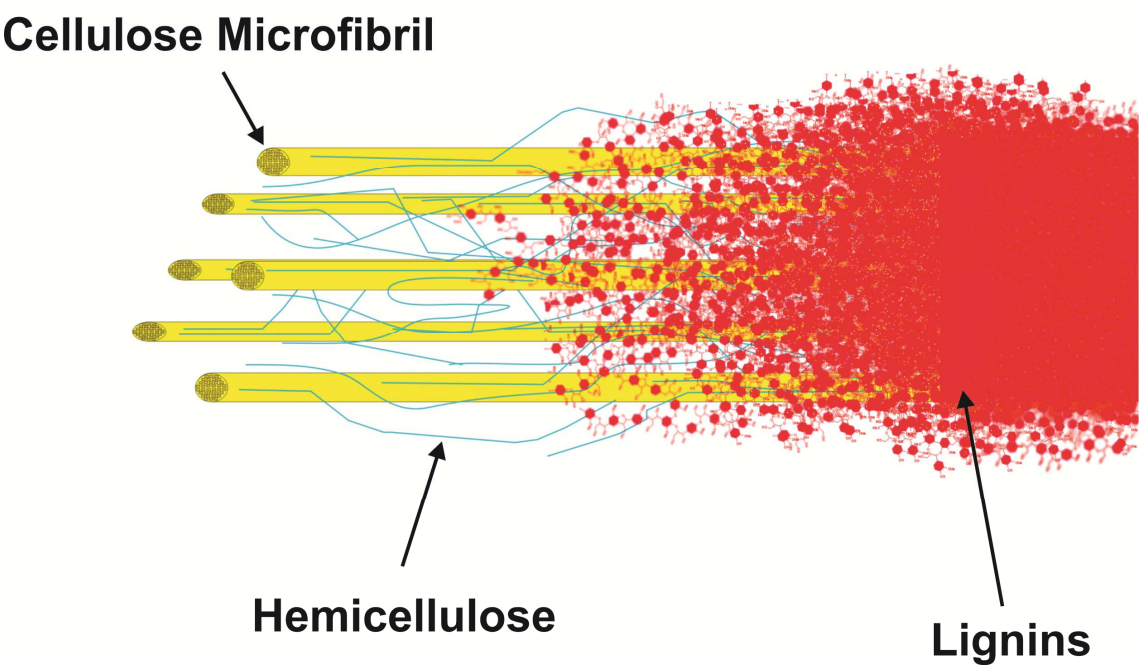



**Figure 1.5. A representation of the structure of lignocellulose from a secondary cell wall.**
Hemicellulose hydrogen bonds to the surface of cellulose microfibrils and spans between individual microfibrils, holding them together. Lignin fills the gaps between cellulose and hemicellulose rendering the polysaccharides very difficult to access. The inclusion of lignin has been thinned to the left side of the figure simply to make the polysaccharide components visible. Figure reproduced from (25).

Cellulose is made up of glucose units, covalently bonded into a rigid, linear, un-branched polymer connected by β-1,4 glycosidic bonds. Cellulose makes up 35 – 50 % of the dry weight of typical lignocellulosic biomass (26). In plants cellulose is made by cellulose synthase genes (CESA) embedded in the plasma membrane (22). Each CESA protein is capable of producing one β-1,4 glucan chain. Six CESA proteins associate together to form a complex called a rosette subunit which produces a crystalline ribbon of cellulose (Figure 1.6). Six rosette subunits associate together to form a large hexamer complex called a particle rosette which is capable of spooling a crystalline cellulose microfibril (36 β-1,4 glucan chains). The crystalline nature of the microfibril renders it insoluble. The cellulose chains spontaneously form such microfibril structures due to interchain hydrogen bonding holding them together. This renders most of the cell wall cellulose unreachable to enzymes as the insoluble crystalline substrate is too large to fit into an active site.

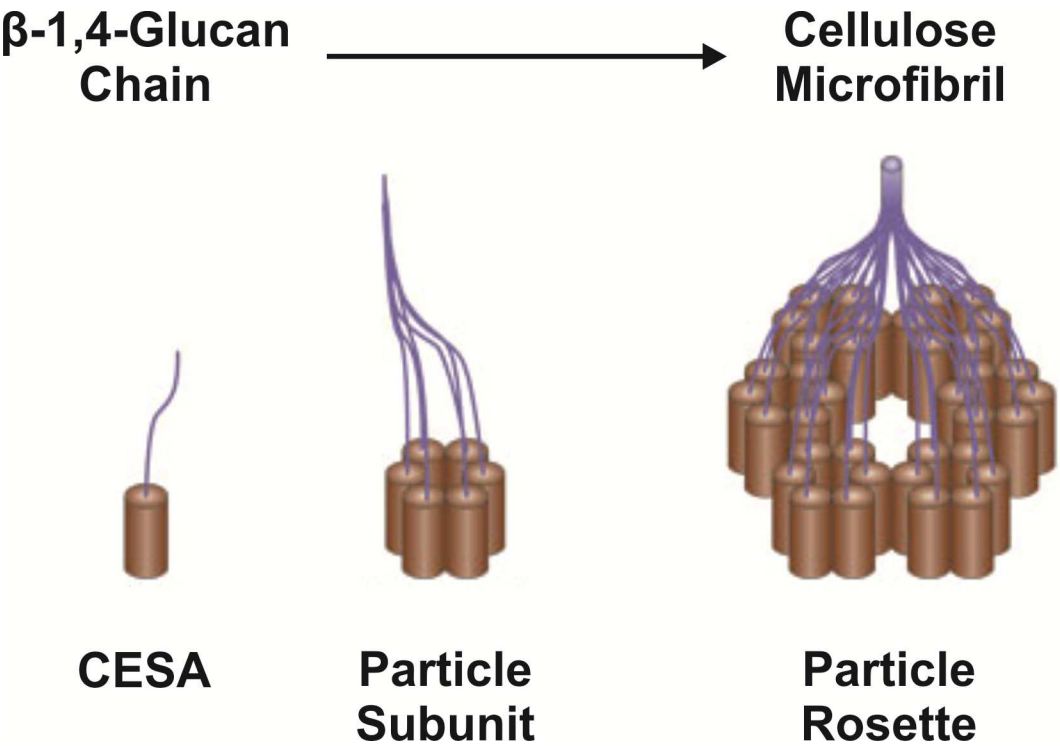



**Figure 1.6. Diagram of the structures responsible for the formation of cellulose.** A single cellulose synthase gene (CESA) is able to produce a single β-1,4-glucan cellulose chain. Six CESA proteins associate together to form a particle subunit able to create a crystalline ribbon of cellulose. A particle rosette is formed by the association of six particle subunits. Together the complex of 36 CESA proteins is able to spool a crystalline cellulose microfibril. Figure adapted from (22).

Hemicelluloses are a very diverse group of polysaccharides typically composed of β-1,4 linked backbones of glucose, xylose or mannose decorated with other sugars such as arabinose, fucose and galactose (27). Between 20 – 35 % of the dry weight of typical bioenergy crops can represent hemicelluloses (26). The predominant hemicellulose present in dicotyledon species of plants is xyloglucan, that is a glucose backbone decorated with xylose residues (28). Hemicelluloses of monocotyledons are primarily xylans and arabinoxylans (28). Hemicelluloses form hydrogen bonds to the surface of cellulose microfibrils, forming a hairy coat that can stretch between adjacent microfibrils to form a cohesive material (22).

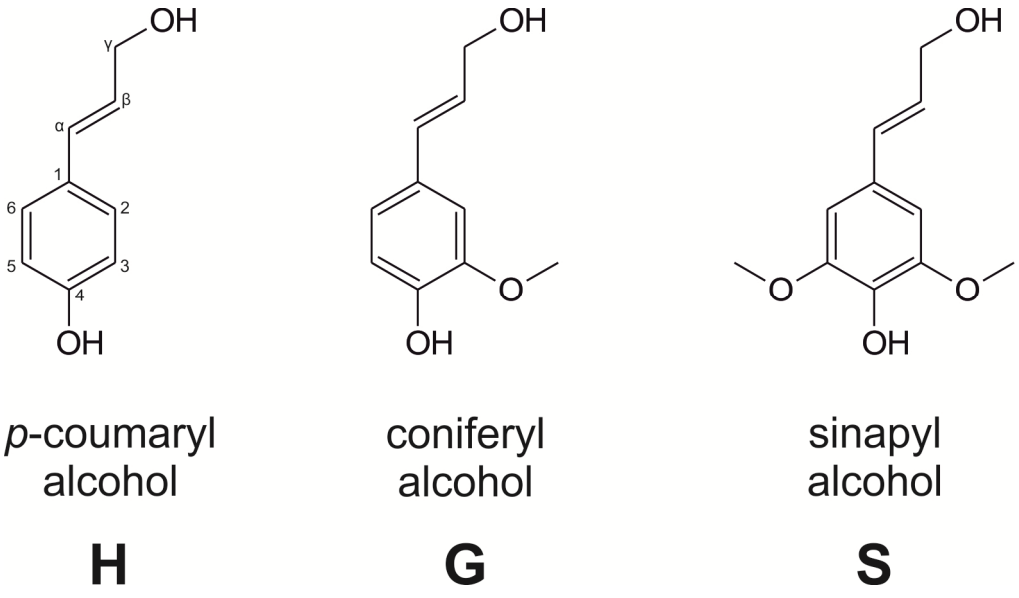



**Figure 1.7. The three major monomers from which lignin is constructed.** The monomers are referred to as both 4-hydroxyphenylpropanoids and hydroxycinnamyl alcohols.

Within the secondary cell wall the resistance to degradation is compounded further by the presence of lignin. Lignin is the name given to a huge, amorphous polymer of aromatic 4-hydroxyphenylpropanoid compounds which can constitute 10 – 30 % of the dry weight of biomass (26, 29). Lignin molecules are mostly built from three different hydroxycinnamyl alcohol monomers, with the amount of each varying depending on plant and tissue; those monomers are *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol (Figure 1.7) (30). Once incorporated into a lignin molecule these monomers constitute units, and these units are identified differently to their monomers; becoming *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) units in respect to the order of the previous sentence. Different groups of plants build their lignins using different proportions of G, S and H units. Dicotyledonous angiosperms (hardwoods) tend to be made of G and S units with trace amounts of H units. Gymnosperm species (softwoods) use predominantly G units with some H units and trace amounts of S units. Monocot grass species, like angiosperms, use mainly a mixture of G and S units, but use a higher proportion of H units (31).

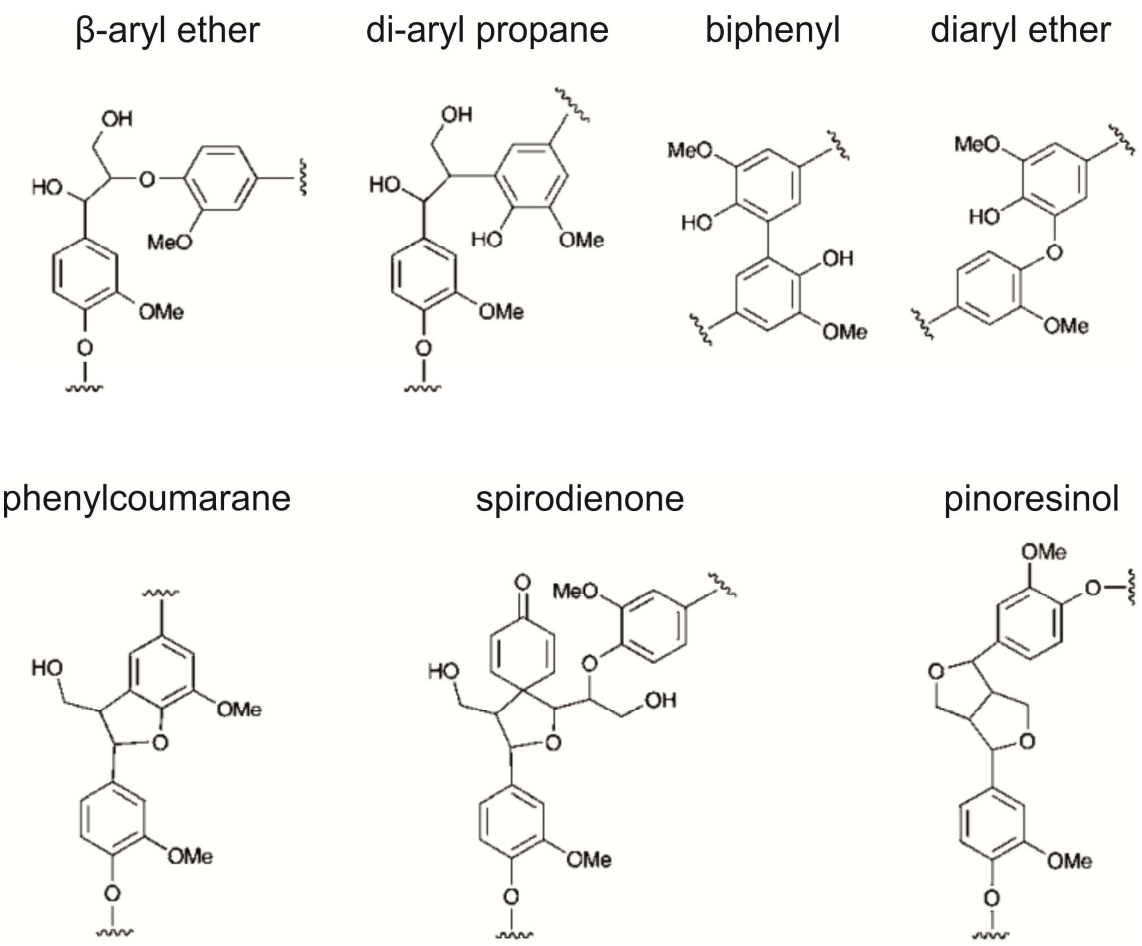



**Figure 1.8. Structures of chemical linkages between monolignol units found in lignin (shown as G units).**
Figure reproduced from (32).

In plants, the monomers are polymerised by a series of random/non-targeted oxidative free radical polymerisation reactions (30). The reactions are thought to be facilitated by peroxidases, laccases, polyphenol oxidases and coniferyl alcohol oxidases with these enzymes oxidatively dehydrogenating the monolignols resulting in monolignol radicals. These monolignol radicals are relatively stable as they are able to delocalise the unpaired radical electron around their conjugated system. Two monolignol radicals come together and couple via formation of a covalent bond using their radical electrons. This coupling reaction quenches the radicals. To elongate the nascent lignin chain further rounds of dehydrogenation must occur. Due to the ability of the radical electron to delocalise a large variety of couplings are observed depending on where the electron was when the coupling took place. The most common linkage to form is the β-aryl ether linkage, followed by the biphenyl linkage (Figure 1.8) (32). A selection of other important linkages which are possible are shown in Figure 1.8.

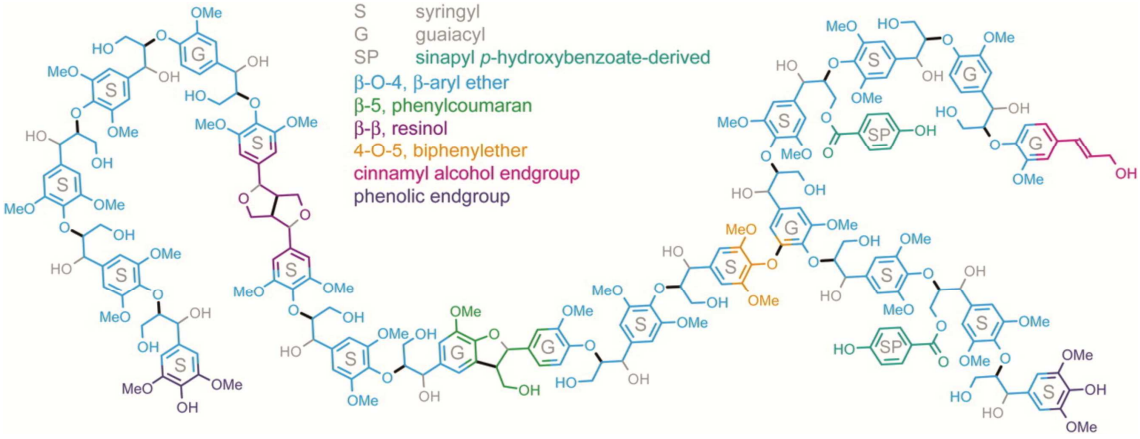



**Figure 1.9. Representative lignin molecule from *Populus sp.* as predicted by NMR analysis.**
Figure reproduced from (30), based on data from (33).

Polymerising lignin incorporates itself into the spaces between cellulose microfibrils and can even form covalent cross links with saccharide residues in the cell wall (34). This excludes water from the cell wall and encases the microfibrils in a hydrophobic barrier, protecting the cell from biodegradation by enzymes. Enzymes cannot evolve a specific active site to attack lignin due to its random structure, large size and insolubility – an example lignin is shown in Figure 1.9.

## 1.3 Second Generation Biofuel Production

In order to access the energy stored in plant cell walls, and by extension all biomass, one of two approaches must be taken. The first option is to thermochemically deconstruct the biomass in a non-specific way, degrading the complex lignocellulose polymer fully into simpler molecules such as carbon monoxide or methane which can be used as fuels. Pyrolysis and gasification are examples of such an approach. Alternatively a suite of enzymes can be used to biochemically and specifically break apart the components of lignocellulose resulting in the release of monomeric sugars which can be subsequently fermented into bioethanol. Both approaches have their advantages and disadvantages which will be discussed below.

### 1.3.1 Thermochemical Deconstruction

#### *1.3.1.1 Pyrolysis*

Pyrolysis is the thermal decomposition of organic matter in the absence of oxygen. Temperatures of between 300 – 500 ℃ are used to de compose biomass into oil, char and a gas containing methane and hydrogen (35). The oil produced contains a complex mixture of organic carbon molecules and can be used directly as a fuel oil, or can be upgraded to other more useful compounds (36). The gas produced can also be used directly as a fuel for CHP applications. The remaining solid component of the process, char, can be burnt as a solid fuel or used as a fertiliser and soil improver. The amounts of each component produced in the reaction can vary depending on the conditions under which the feedstock is heated, as well as the composition of the feedstock used (37).

The pyrolysis technique is able to completely convert lignocellulosic biomass into second generation biofuels, however the types of fuels produced are not “drop-in” substitutes for currently used liquid transportation fuels and therefore are of a lower utility for displacing fossil fuels. Relatively high temperatures are also required to carry out pyrolysis which lowers the net energy yield of the process, reducing the amount of fossil oil it is able to displace.

### *1.3.1.2 Gasification*

Gasification is a similar process to that of pyrolysis; it is the thermal decomposition of organic matter in the presence of a controlled and limiting amount of oxygen, and also sometimes steam. The decomposition is carried out at temperatures of between 750 – 1500 ℃ and in some systems at increased pressure (38). The products of the decomposition are a mixture of carbon monoxide, carbon dioxide, hydrogen and methane gases collectively known as syngas (synthesis or producer gas) and a glass like inert slag waste material (38). Syngas can be used directly as a fuel for CHP. It can also be upgraded into straight chain, liquid alkanes via the Fischer-Tropsch process (39). In this way biomass, via gasification, can be made into alkanes suitable for direct substitution with currently used liquid transportation fuels.

Like pyrolysis, gasification is able to completely convert lignocellulosic biomass into a second generation biofuel, syngas. Although syngas itself is not a “drop-in” replacement for our current liquid transportation fuels, via the Fischer-Tropsch process, it can be made so. However both the gasification process itself, as well as the Fischer-Tropsch process, require very high temperatures and in some cases pressures in order to work. This requires large amounts of energy to be used to provide this heat, reducing the benefit of using a renewable biomass feedstock for the process. The heat requirement, coupled with the many production steps involved in the conversion of syngas to liquid fuels, raises the production costs of the gasification route to levels that make it uncompetitive with current fossil fuels (40). As the cost of oil rises there may reach a point where this method of producing liquid transportation fuels becomes economical; or indeed if the production method is made cheaper and more efficient it may become more widely adopted as has been seen in the case of shale gas extraction (41). However, it is believed that in the near term the use of gasification will remain niche and will be more often used by the fossil fuel industry rather than the bioenergy sector.

### *1.3.1.3 Microbial Gas Fermentation*

Bacterial species have been discovered which are able to anaerobically ferment carbon monoxide and hydrogen into short chain fatty acids and alcohols (42); whilst other species are able to ferment carbon monoxide and water into carbon dioxide and hydrogen (43, 44). Both pathways could potentially offer a cheaper, less energy

intensive method for converting syngas into liquid transportation fuels. Indeed, a proprietary form of this technology has been commercially exploited by LanzaTech to capture flue gases from steel manufacturing, oil refining and chemical production industries (www.lanzatech.com).

### 1.3.2 Biochemical Deconstruction

Current industrial attempts to biochemically deconstruct lignocellulosic biomass use a two-step process.

First the biomass is pre-treated. The pre-treatment can constitute multiple stages, but usually employs a milling step and a physicochemical step (45, 46). The milling step aims to reduce the particle size of the material. This has little effect alone on the amount of extractable sugars present but makes the physicochemical step more effective by increasing the amount of available surface area (47). The physicochemical step aims to change the nature of the material on a molecular scale, attempting to make sugars more easily extractable. The most common physicochemical pre-treatment used is to soak the milled biomass in acidic or alkaline solutions; this is often done at elevated temperatures and pressures (70 – 200 ℃ and 45 – 220 PSI) (48). Another physio-chemical pre-treatment used is steam or ammonia fibre explosion. In this process the biomass is suspended in water or ammonia and heated to 60 – 100 ℃ under 250 – 300 PSI of pressure; once these conditions have been maintained for ~5 minutes the pressure is rapidly released causing liquid which has penetrated the lignocellulose structure to quickly expand, exploding, and opening up the molecular structure of the biomass (49). The use of ionic liquids as solvents to disrupt the lignocellulosic structure is being investigated and shows promise (50, 51); however, there remain several practical hurdles and issues of sustainability that must be solved before they prove commercially useful (52).

The nature of these pre-treatment methods makes them energy intensive (high temperature and pressure) and expensive (53, 54). The use of chemicals such as acids, alkalis and solvents also gives them a large water footprint. Considering the scale and final goal of such pre-treatments, to make liquid transportation fuel to replace petro-based fuels, the scale and source of chemical required must also be considered. For example, it would be foolish to use a petroleum derived solvent to make a biofuel to replace petrol unless it were possible to recycle that solvent. These factors make it desirable to negate or reduce to a minimum any pre-treatment necessary.

The second stage industrially used to release sugars from pre-treated biomass is the use of glycosyl hydrolase (GH) enzyme cocktails to hydrolyse polysaccharides into monomer sugars, a process called saccharification. Currently used commercially available cocktails contain enzymes able to break down cellulose to glucose (described below), and enzymes able to break down hemicellulose to some degree. Examples of such cocktails are marketed as Cellic® CTec3 (Novozymes) and Accellerase® Trio™ (Genencor) however their exact composition is not disclosed. For the effective hydrolysis of cellulose three groups of enzymes are needed which are collectively known as cellulases (55, 56): endo-β-1,4-glucanases (EC 3.2.1.4) which cleave cellulose chains internally at random points; exo-1,4-β-glucanases (also known as cellobiohydrolases EC 3.2.1.91) which cleave cellobiose units (a glucose disaccharide) from the end of a cellulose chain (enzymes exist which can facilitate this from the reducing and non-reducing ends of the cellulose chain); and 1,4-β-glucosidases (EC 3.2.1.21) which separate cellobiose into two glucose monomers. A representation of this process is shown in Figure 1.10. Due to the highly variable and branched nature of hemicellulose, a large array of enzymes is required to reduce hemicellulose to its constituent monomers including endo-1-4,-β-xylanase, β-xylosidase, α-glucuronidase, α-L-arabinofuranosidase, β-mannanase and β-mannosidase (55, 57). Enzymes are not commonly used to degrade or modify the lignin component of the biomass.

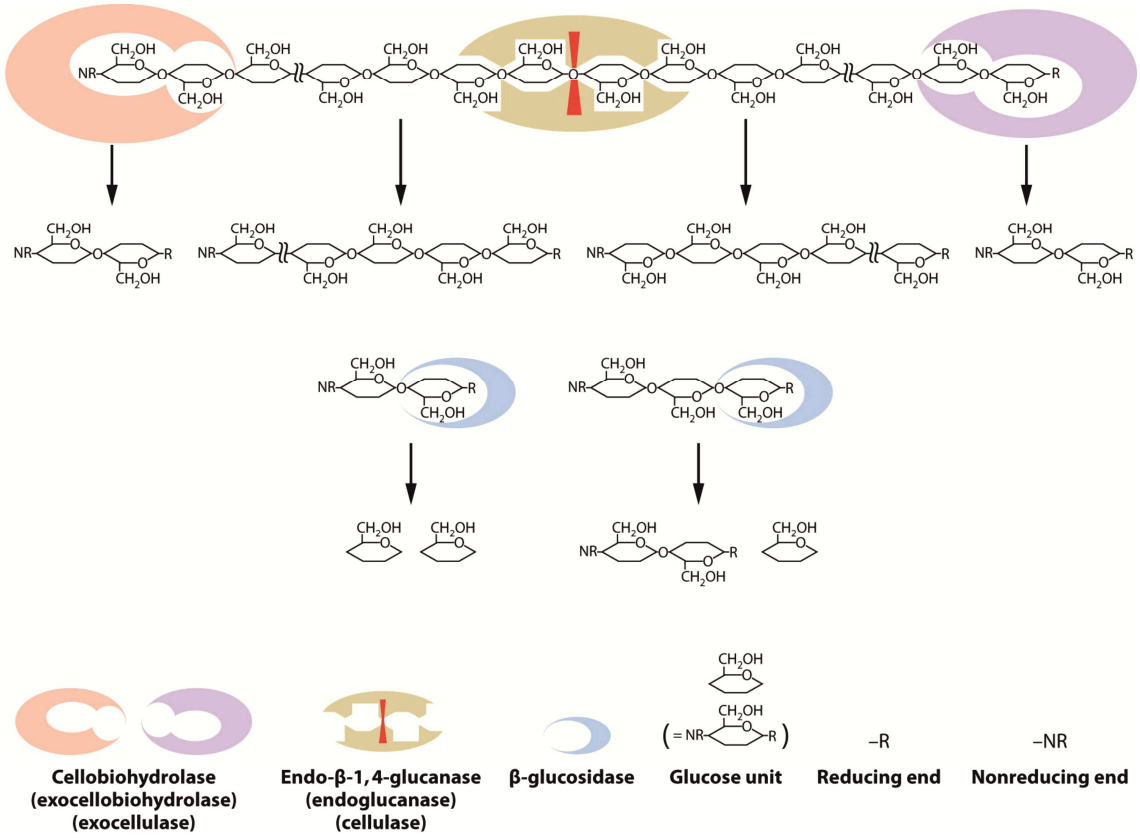



**Figure 1.10. Overview of the essential enzymatic activities required to completely depolymerise cellulose to glucose.**
Endo-β-1,4-glucanases are able to cleave cellulose chains internally, resulting in smaller cellulose chains. Cellobiohydrolases cleave cellobiose (glucose disaccharide) units from the ends of cellulose chains. Cellobiohydrolases exist which can act from the reducing or non-reducing end of cellulose chains. β-glucosidases hydrolyse glucose units from the ends of cellulose chains or break cellobiose down into glucose monomers. Figure reproduced from (58).

The standard biochemical reaction used by GH enzymes to hydrolase glycosidic bonds is one of general acid catalysis and can proceed via either a stereochemistry retaining, or inverting mechanism (Figure 1.11) (59). The inverting mechanism takes place via a single displacement reaction and results in a stereo inversion at the anomeric carbon (Figure 1.11 – A). The glycosidic oxygen is protonated by a general acid, whilst a general base deprotonates a water molecule situated beneath the substrate, activating it to behave as a nucleophile which attacks the anomeric carbon resulting in its hydroxylation. The retaining mechanism takes place via a double displacement reaction involving a glycosyl enzyme intermediate and results in the maintenance of the substrates stereochemistry (Figure 1.11 – B). In the first displacement reaction a general acid protonates the glycosidic oxygen, whilst at the same time a general base attacks the anomeric carbon. This results in a glycosyl enzyme intermediate. A water molecule is then deprotonated by the deprotonated general acid to form a nucleophile

which attacks the anomeric carbon, hydrolysing the glycosyl enzyme bond. This results in the regeneration of the enzyme and the product maintaining its original stereochemistry.

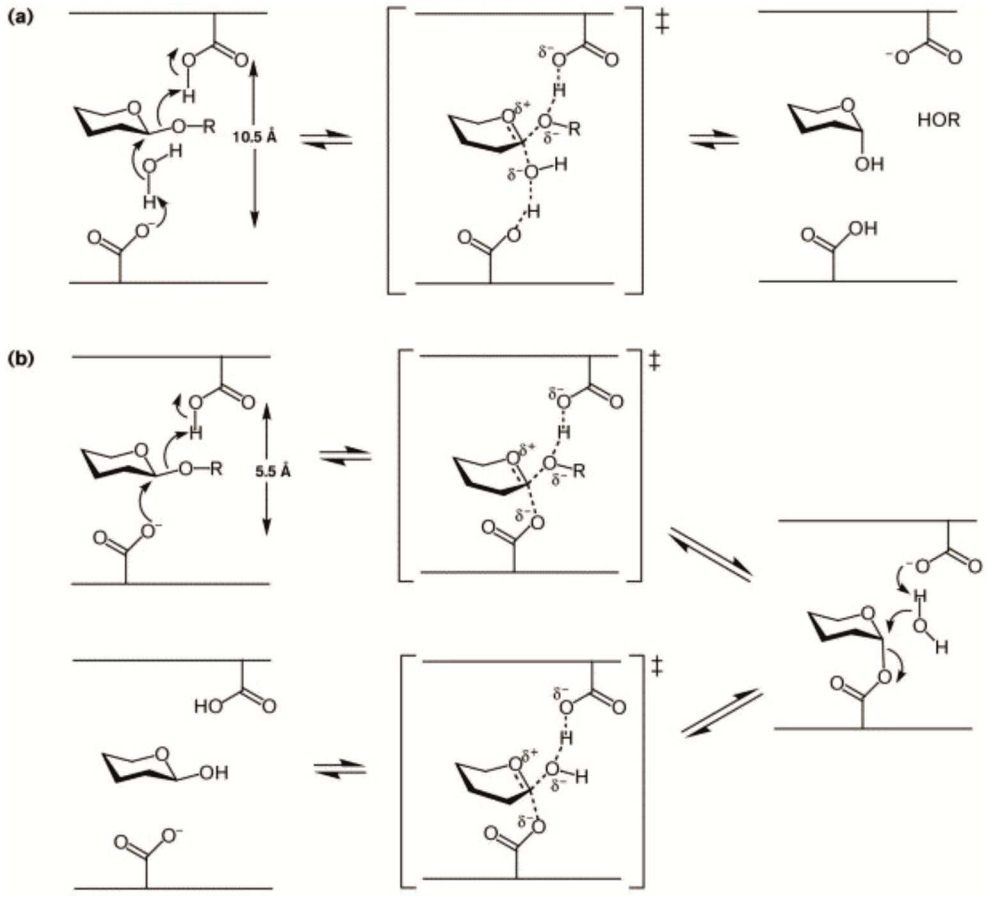



**Figure 1.11. General enzymatic mechanism of inverting (A) and retaining (B) glycoside hydrolases.**
Reproduced from (60).

Despite being able to catalyse the breakdown of a highly diverse range of substrates the topology of GH active sites is thought to be represented by just three general classes (59). These are the pocket or crater, cleft or groove and tunnel topologies with examples of each shown in Figure 1.12. The pocket or crater substrate binding topology is most often encountered in enzymes which target substrates having many accessible ends and which cleave monosaccharides from those ends such as glucoamylase which cleaves glucose units from the many available chain ends of starch granules. The cleft or groove active site is found in enzymes which attack large chain substrates like cellulose or xylan. The open cleft of the active site allows the enzyme to bind in the middle of a chain, making these enzymes endo acting in general. Tunnel topology enzymes are similar to cleft enzymes in that they possess a substrate

binding cleft, but it is covered by long flexible loops forming a tunnel. These enzymes are able to thread long chain substrates such as cellulose or xylan through their tunnel to an active site where the chain is hydrolysed into mono/disaccharides or smaller oligosaccharides, and the smaller hydrolysed fragment can then exit the other end of the tunnel and diffuse away. This allows these enzymes to act processively, moving along the long chain substrate carrying out repeat cycles of hydrolysis. Hybrid enzymes, behaving as cleft and tunnel type hydrolases have been documented which are able to open their long tunnel forming loops exposing an active site cleft (61). They are able to bind in an endo fashion on a long chain substrate, close their loops over the substrate, cleave the chain and then processively hydrolase that chain.

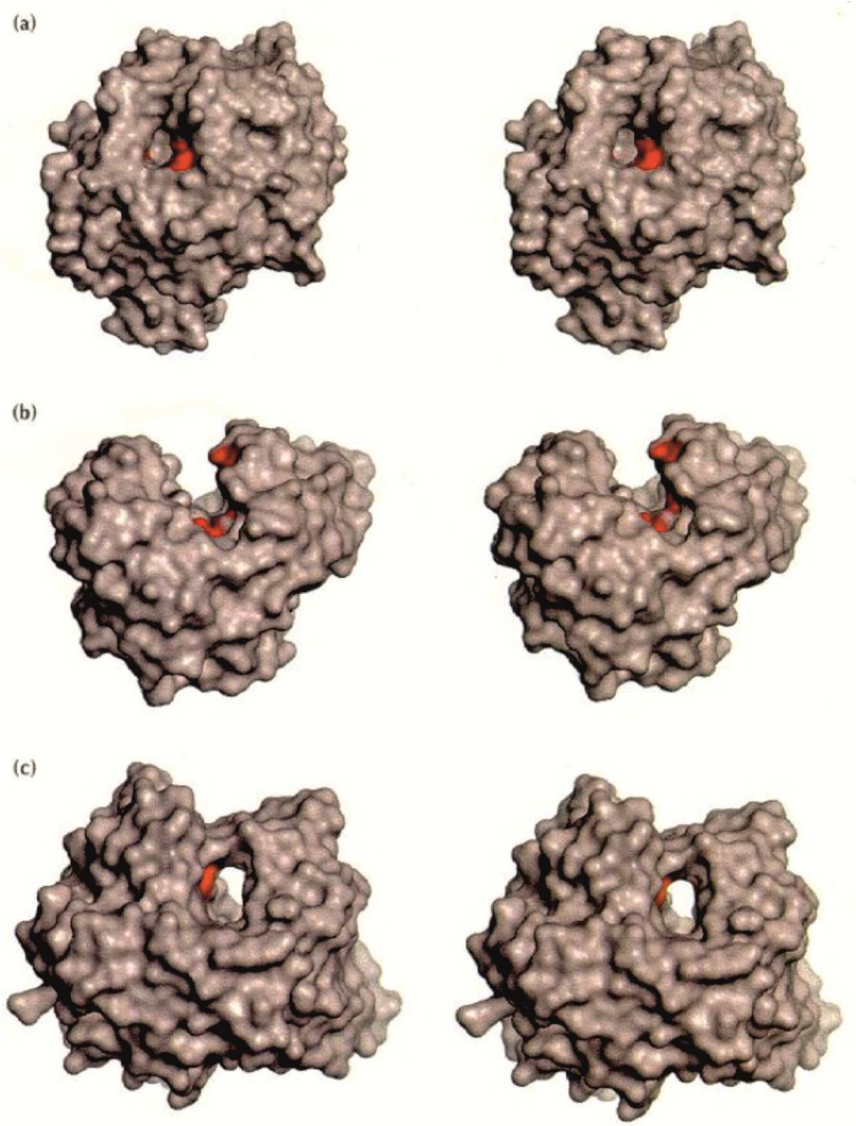



**Figure 1.12. The three surface topologies of glycosyl hydrolase enzymes.** The pocket or crater (a), the cleft or groove (b) and the tunnel (c). The proposed catalytic residues are shown in red. Reproduced from (59).

Glycosyl hydrolases often also contain a domain known as a carbohydrate binding module (CBM) which are reviewed in (62). These CBMs are often attached to the GH protein via a linker peptide and as their name suggests they bind carbohydrates. It is thought that CBMs help to increase the effective concentration of GHs on the surface of insoluble substrates increasing the rate at which the substrate is degraded. As discussed for GHs themselves, there are three topologies used by CBMs which influence the type of substrate they bind to. Type A CBMs are known as surface binding CBMs. They possess flat platform like binding sites which are thought to be complementary to the surface of crystalline substrates such as cellulose (63). Type B CBMs have been called glycan-chain-binding CBMs and have cleft or groove like binding sites able to bind several sugar units. This allows them to bind to a glycan chain enhancing the processivity of their attached GH enzyme. Type C CBMs are similar to type B CBMS in that they also possess cleft or groove binding sites. These binding sites are only large enough to bind mono, di or tri saccharides and are more prevalent in bacteria where they are used to bind cell surfaces or particular glycan structures in a lectin type manner (64).

In order to make the biochemical deconstruction of biomass an economical alternative to both thermal decomposition methods and fossil fuels the process must be made more efficient. Many pathways to achieve this are being actively researched. At the beginning of the process work is being undertaken to increase the yield of biomass obtainable from a given area of land, increasing the amount of feedstock available for no extra inputs; for example by modifying the morphology of the plants leaf canopy to maximise the amount of light useable for energy capture (65). Work is being carried out to determine which genes in plants can be up or down regulated to increase the saccharification potential of plants using current technologies without detrimentally impacting the desirable qualities of the plant, such as stem strength (66, 67). Much effort is being invested into improving the potential of the saccharification process itself by improving the enzyme cocktails used for lignocellulose breakdown (61, 68).

Currently, as described above, enzyme cocktails focus on breaking down cellulose and hemicellulose. If more active enzymes were available, or they were able to be made more efficiently, this would reduce the amount of enzyme necessary to treat a given amount of biomass reducing the cost and energy associated with that enzyme. Enzymes less affected by lignocellulose breakdown product inhibition would also help to make the process more economical by allowing more sugar to be released from biomass. The discovery and use of enzymes able to degrade lignin, or able to break

the covalent bonds between lignin and polysaccharides would also increase the amount of sugar able to be released during the process by breaking down the tough, waterproof lignin barrier protecting the cellulose and hemicellulose fibres. Degrading lignin would also weaken the superstructure of lignocellulose, potentially allowing milder, less energy intensive pre-treatments to be used. The compounds that lignin is made of are also of value to the flavour/fragrance and materials industry and so if they were able to be released from the lignin and recovered this would provide an additional revenue stream to make the process economical (69).

The sugars released by biochemical deconstruction of biomass cannot be used as a fuel directly; first they must be fermented into bioethanol which can be used as a liquid transportation fuel. Current first generation bioethanol is produced using adapted strains of yeast to ferment glucose to ethanol anaerobically. However, the current strains and methods available are not sufficient to make fermentation of biomass sugars economical (70, 71). Numerous aspects of the fermentation of deconstructed biomass mean that the conversion to ethanol is not as efficient as is seen for the fermentation of corn starch sugars. These issues include fermentation inhibition by lignocellulose break down products such as acetate, lignin and mono-lignols, inhibition of yeast growth and fermentation by ethanol, unnecessary yeast growth wasting resources and inability to utilise 5-carbon pentose sugars (C5) such as xylose (72-75).

Work is being carried out to address these issues with the following approaches being used. Biochemical pathways allowing the use of pentose sugars (and therefore the hemicellulose fraction of biomass) for ethanol production are being incorporated into yeast and bacteria (76, 77). The regulation of sugar use in these organisms is also being investigated and manipulated to allow the simultaneous use of pentose and hexose sugars for fermentation, reducing the amount of time required to ferment all available sugars (78). Metabolic engineering is being carried out upon ethanol producing yeast and bacteria to minimise the amount of by-products and pathways which intersect with sugar fermentation, maximising the amount of flux from sugar to ethanol (77, 79, 80). Such metabolic engineering has been successfully used in other industries to incrementally increase the maximum possible yield of a conversion process (81). Attempts are also being made to engineer organisms capable of simultaneously degrading lignocellulose and fermenting the released sugars. This would allow the overall process to involve less steps and expensive plant (82-84). In a similar vein, experiments are being carried out to create stable mixed microbial

communities to perform the same function, with specific strains being present to carry out certain biochemical conversions with maximum efficiency (85).

## 1.4 Lignocellulose Breakdown in Nature

Despite the resistance of lignocellulose to bio-degradation, there are examples of organisms spread across the phyla of life able to degrade it. Indeed, the ability to degrade lignocellulose is essential to the global carbon cycle, releasing carbon sequestered into plant biomass back to the atmosphere as carbon dioxide. Different organisms have evolved different ways to attack and break down plant biomass, some being able only to release a limited amount of sugar from it and others able to entirely mineralise biomass. The major classes of life able to do this, and a description of their evolved tactics are discussed below.

### 1.4.1 Bacteria and Fungi

Bacteria are a tremendously prolific and successful form of life having evolved to survive in practically every environment present on the planet, from saline ponds to near boiling water in the depths of the ocean (86, 87). Fungi are also a very successful form of life with 1.5 million species predicted to exist (88). Both groups of organisms have evolved to utilise many forms of chemical energy for their growth and reproduction and lignocellulose is one of those forms of energy. Aerobic bacteria and basidomycete fungi have been discovered which secrete soluble GH enzymes to break down the cellulose and hemicellulose fractions of lignocellulose whilst several species of anaerobic bacteria have been documented to use a different method to help them exploit this feedstock, the cellulosome.

#### *1.4.1.1 The Cellulosome*

The cellulosome is a name given to a very large protein complex composed of a modular assortment of GH enzymes bound to a protein framework anchored to the bacterial cell surface at one end and a polysaccharide at the other. The first cellulosome described was discovered in the thermophilic bacterium, *Clostridium thermocellum*. Many variations of the cellulosome have since been discovered but below is a description of the simplest from *C. thermocellum* (Figure 1.13). An account of the discovery of the first documented cellulosome as well as details of subsequent cellulosomal research is reviewed in (89).

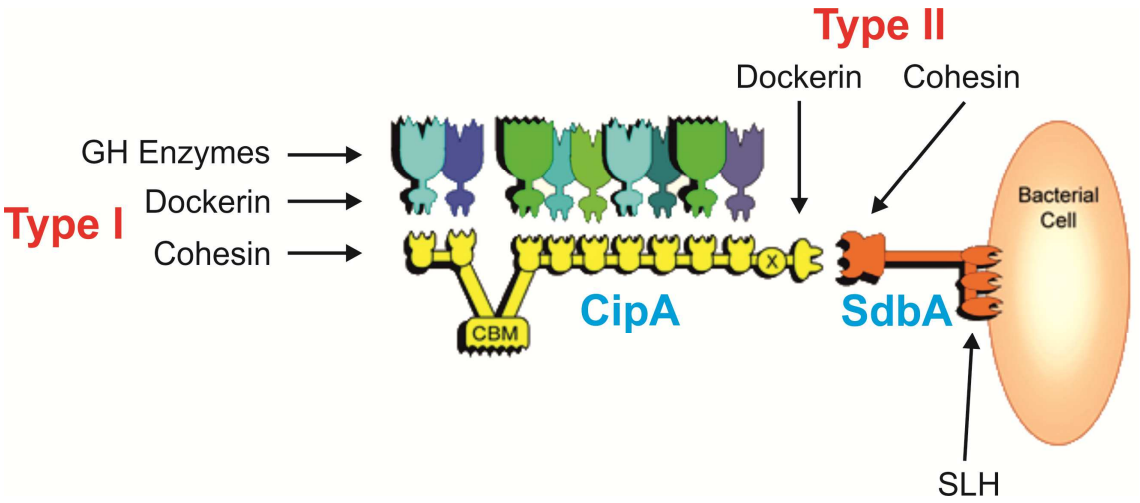



**Figure 1.13. Schematic representation of a cellulosome from *Clostridium thermocellum*.**
Glycosyl hydrolase enzymes bearing type I dockerin modules attach to the type I cohesin modules of the CipA scaffoldin protein. The carbohydrate binding module (CBM) of CipA is able to attach the scaffoldin to a cellulose chain. The whole CipA/GH complex is anchored to the bacterial cell surface by binding with a SdbA scaffoldin via a type II dockerin/cohesin interaction. The SdbA scaffoldin itself attaches to the cell surface via a S-layer homology (SLH) module. Figure adapted from http://www.cazypedia.org/index.php/Cellulosome.

The first part of the complex is an anchoring scaffoldin protein (SdbA); it binds to the surface of the bacterial cell via an S-layer homology module (SLH) and in doing so tethers the whole cellulosome complex to the bacterium. At the other end of the SdbA protein is a type II cohesin module. Bound to this cohesin module via an X-dockerin dyad is a large, 210 kDa scaffoldin protein called CipA. The CipA scaffoldin forms the protein framework onto which the GH proteins bind. The CipA protein contains nine type I cohesin modules and one CBM. Each type I cohesin module is able to bind one dockerin module containing protein. Dockerin modules were found to be possessed by a suite of GH enzymes and each of these enzymes could bind to any of the CipA cohesin modules allowing the scaffold to be randomly populated by any available GH enzymes. This could result in a CipA scaffold bearing nine GH enzymes. The CBM of the CipA scaffold is able to bind cellulose, and in doing so anchors the end of CipA distal to the bacterial cell to a cellulosic substrate. The virtue of this system is that it anchors a bacterium in close proximity to a source of sugar and in the direct vicinity of that sugar it concentrates a number of GH enzymes. These enzymes can then liberate sugars from the food supply that are in within their reach.

Other arrangements of cellulosomes exist, but they all share in common the use of modular cohesin and dockerin modules to allow various arrangements of scaffoldins

and enzymes to be connected with the goal of maintaining the bacterium in close vicinity to a polysaccharide food source and increasing the local concentration of enzymes able to degrade that food source for the bacterium (89).

### *1.4.1.2 Lignin Degradation*

Bacteria and fungi have been discovered which can degrade lignin and its breakdown products (90-92). Methods by which this occurs are reviewed in (32, 93). There are many species of fungi able to degrade lignocellulose, and these have been broadly classified into the white and brown rot fungi. White rot fungi are able to completely breakdown lignocellulose, eventually catalysing the degradation of lignin to $CO_2$ and $H_2O$ (94). Brown rot fungi however seem to modify the structure of lignin, leaving it as a polymeric residue rather than fully degrading it (95). A brief outline of some of the major lignin breakdown methods are detailed below.

The primary method of lignin degradation employed by white rot fungi is of oxidative attack via the use of extracellular oxidative enzymes such as peroxidases and laccases. Bacteria have also been discovered which secrete extracellular peroxidases which are able to catalyse the breakdown of β-aryl ether linkages (96). As enzymes are not able to enter the lignocellulosic structure directly it seems that the use of peroxidases has evolved as a method to break down lignin via small molecule mediators which are able to diffuse away from the peroxidase to act upon the surface of lignin.

White rot fungi have been found to possess three classes of peroxidases, namely manganese peroxidases (MnP), lignin peroxidases (LiP) and more recently versatile peroxidases (VP) (97-100). All three classes are heme containing enzymes. Bacteria have been found to possess LiPs and potentially VPs (96, 101, 102).

Manganese peroxidases use hydrogen peroxide to oxidise $Mn^{2+}$ to $Mn^{3+}$ (103). Highly reactive $Mn^{3+}$ forms stable complexes with oxalic acid and other similar organic acids. In this way $Mn^{3+}$(oxalate) complexes can act as powerful and stable diffusing oxidants, and are able to oxidise a variety of phenolic and non-phenolic substrates which are distal to the MnP active site.

The active site of LiP is only accessible to small non-phenolic substrates and so is unlikely to act directly upon lignin (104). Rather, it is likely that LiP oxidises small non-

phenolic compounds such as veratryl alcohol (3,4-dimethoxybenzyl alcohol), a product of ligninolysis, to its cation radical counterpart which could in-turn act as a diffusible oxidant in an analogous fashion to $Mn^{3+}$. The cation radical of veratryl alcohol has a very short half-life however, and so by itself is unlikely to fulfil this role. It has been shown that the LiP protein acts to stabilise the veratryl alcohol radical by binding it in an acidic environment which reduces the rate of decay of the radical by ~1000 fold as opposed to a chemically generated, non-enzyme bound veratryl alcohol radical (105). A redox active conserved amino acid has also been discovered in LiP proteins, W171. It is thought that this residue may act as part of an electron transfer pathway between veratryl alcohol and the enzymes heme active site (106).

Versatile peroxidases are hybrids of MnPs and LiPs being able to both oxidise $Mn^{2+}$ to $Mn^{3+}$ and oxidise non-phenolic aromatic substrates to their cation radical counterparts (97).

The peroxidases discussed are therefore able to generate diffusible powerful oxidants which allow them to perform single electron reductions of lignin. These reductions result in the generation of radicals which can delocalise through the lignin molecule resulting in cleavage. An example of the radicals formed when a model β-aryl ether compound (a model of the most common linkage found in lignin) is attacked by such diffusible oxidants is shown in Figure 1.14. Oxidation of the model compound can result in the generation of a phenoxy radical (I) or a cyclohexadienyl radical (II) which go on to cleave the compound by $C_\alpha$ – $C_\beta$ or alkyl – phenyl cleavage respectively (107, 108). This fashion of oxidative $C_\alpha$ – $C_\beta$ cleavage is found to be used by white rot fungi to cleave β-aryl ether, diarylpropane and phenylcoumarane linkages (103, 109-112).

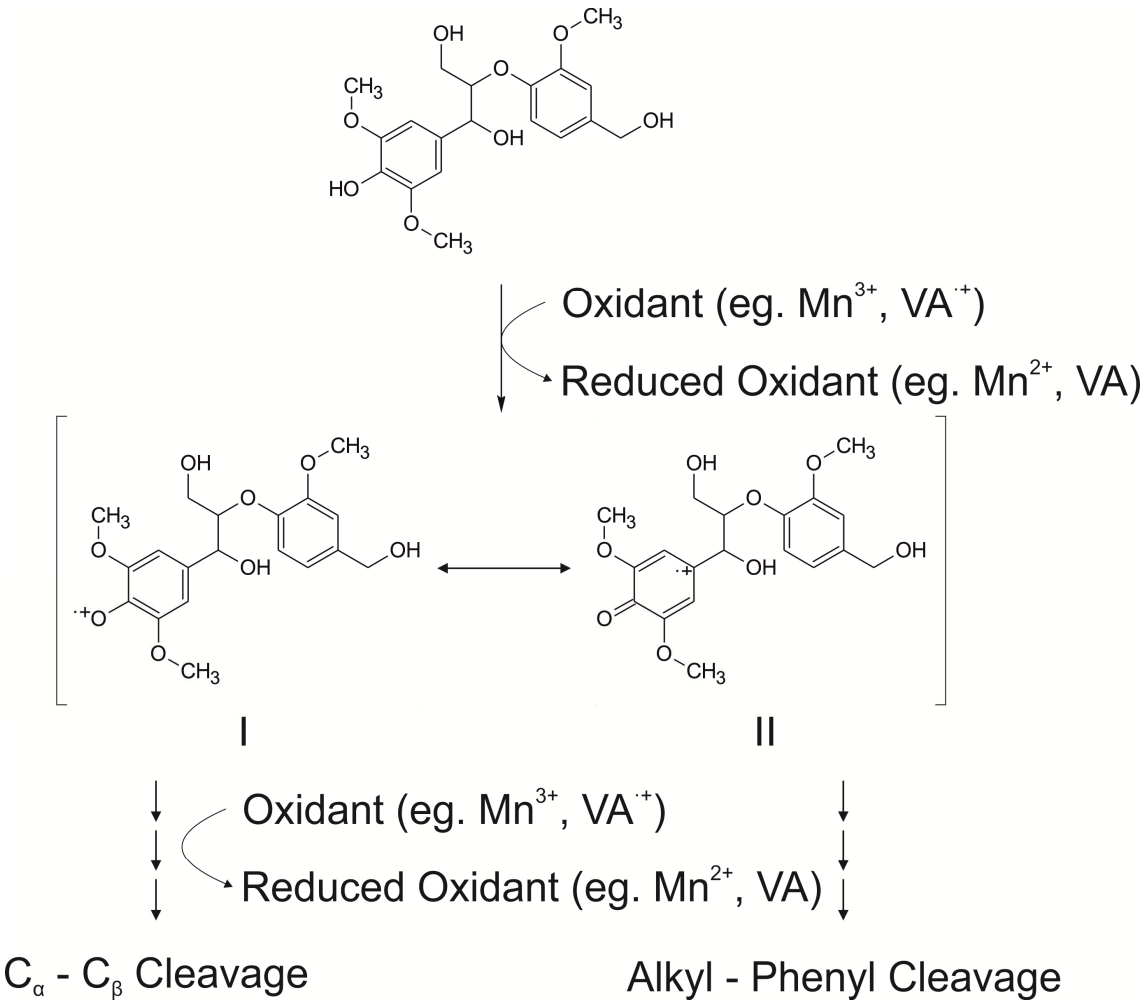



**Figure 1.14. Partial mechanism of the oxidative radical cleavage of a model β-aryl ether compound leading to either $C_\alpha$ – $C_\beta$ or alkyl – phenyl cleavage.**
Model compound, 1-(3,5-dimethoxy-4-hydroxyphenyl)-2-(4-(hydroxy-methyl)-2-methoxyphenoxy)-1,3-dihydroxypropane. Figure adapted from (107).

In this way the lignin structure is fragmented, allowing improved access for GHs to cellulose. Cellulose itself can also be attacked by these oxidative species, disrupting the crystalline nature of the cellulose microfibrils, increasing the number of sites at which GHs can attack.

Whilst bacteria have also been shown to produce extracellular peroxidases, and consequently are likely to also use the above described method of disrupting lignin, they have also been shown to use extensive arrays of more specific enzymes to cleave particular bonds in lignin breakdown compounds (32). A review of all such known mechanisms is beyond the scope of this report, however a brief description of such a pathway used to cleave the common β-aryl ether linkage is given below.

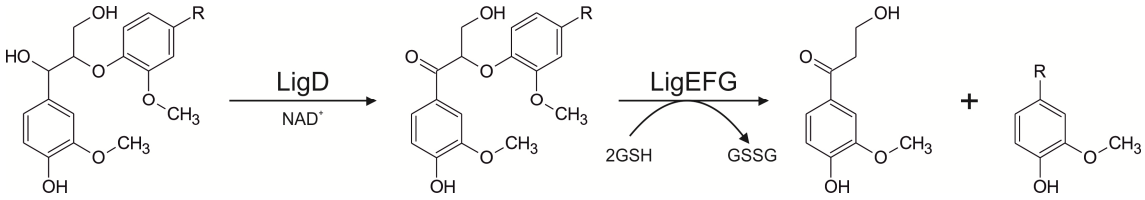



**Figure 1.15. Bacterial pathway for the cleavage of β-aryl ether linkages using the LigD, E, F and G enzymes.**
GSH, reduced glutathione; GSSG, glutathione disulphide. Figure adapted from (32).

The bacterium, *Pseudomonas paucimobilis* SYK-6, is able to take up various dimeric lignin compounds (113). Upon taking up a di-lignin compound containing a β-aryl ether linkage the NAD dependant dehydrogenase, LigD, catalyses the oxidation of the α-hydroxyl group of the compound (Figure 1.15). The β-aryl ether link is then cleaved by a collection of three enzymes; LigE, F and G. LigE and F catalyse the same reaction, but do so to different enantiomer substrates. They catalyse the cleavage of the β-aryl ether link using reduced glutathione resulting in a β-thioether intermediate. LigG then uses another molecule of reduced glutathione to cleave the β-thioether intermediate.

Another form of oxidative attack used to gain access to cellulose has been proposed for white and brown rot fungi. This method uses the oxidation of $Fe^{2+}$ by hydrogen peroxide to produce $Fe^{3+}$ and the hydroxyl radical ($^{\cdot}OH$), the most powerful oxidising agent found in living cells in a Fenton type reaction. It is thought that the hydroxyl radical can go on to oxidatively cleave cellulose fibres and modify lignin allowing GHs better access to cellulose and hemicellulose. Hydroxyl radicals are able to attack lignin, but are thought to result in the oxidation of the polymer, modifying and opening up its structure, rather than its total degradation (95). The $Fe^{2+}$ form of iron is rarely found in aerobic conditions where it is quickly oxidised to $Fe^{3+}$, and the half-life of the hydroxyl radical is measured in nanoseconds. These two factors mean that for this method to work iron must be reduced and the hydroxyl radical generated in very close proximity to the substrate they are to attack. Several pathways have been described to allow this to be so.

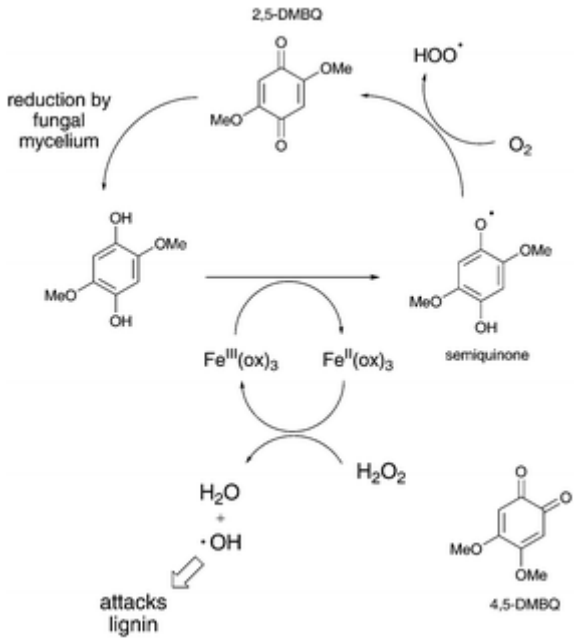



**Figure 1.16. Proposed scheme for generation of hydroxyl radicals via Fenton chemistry in white and brown rot fungi.**
$Fe_{III}$ (chelated to oxalate) is reduced to $Fe_{II}$ by the coupled oxidation of a hydroquinone to its semiquinone counterpart. Hydrogen peroxide is able to react with $Fe_{II}$ to form water and a hydroxyl radical which is able to attack lignin. The semiquinone is reduced back to a hydroquinone via its quinone form by the fungus. 2,5-DMBQ, 2,5-dimethoxy-1,4-benzoquinone; 4,5-DMBQ, 4,5-dimethoxy-1,2-benzoquinone; ox, oxalate. Figure reproduced from (32).

One such method to produce reduced iron involves the redox cycling of a (hydroxy)quinone with its semi-quinone radical counterpart to reduce $Fe_{III}$ to $Fe_{II}$ (114, 115) (Figure 1.16). Several other methods are shown in Figure 1.17, including a method which up-regulates biosynthetic pathways producing phenolate compounds capable of reducing $Fe^{3+}$, such as variegatic acid (116). An iron reductase enzyme has been discovered in the brown rot, *Serpula lacrymans*, and the white rot, *Phanerochaete chrysosporium* which possess a CBM which would anchor it to a crystalline cellulose fibre allowing it to produce $Fe^{2+}$ ions in very close vicinity to the target cellulose (116). *S. lacrymans* was also found to greatly up-regulate oxidoreductase enzymes when feeding on wood which would allow a steady supply of hydrogen peroxide to be produced to fuel the hydroxyl radical generating Fenton reaction (116). Coupled with a suitable suite of GH enzymes, it is thought that this is how brown rot species are able to efficiently saccharify lignocellulosic substrates without mineralising the lignin component.

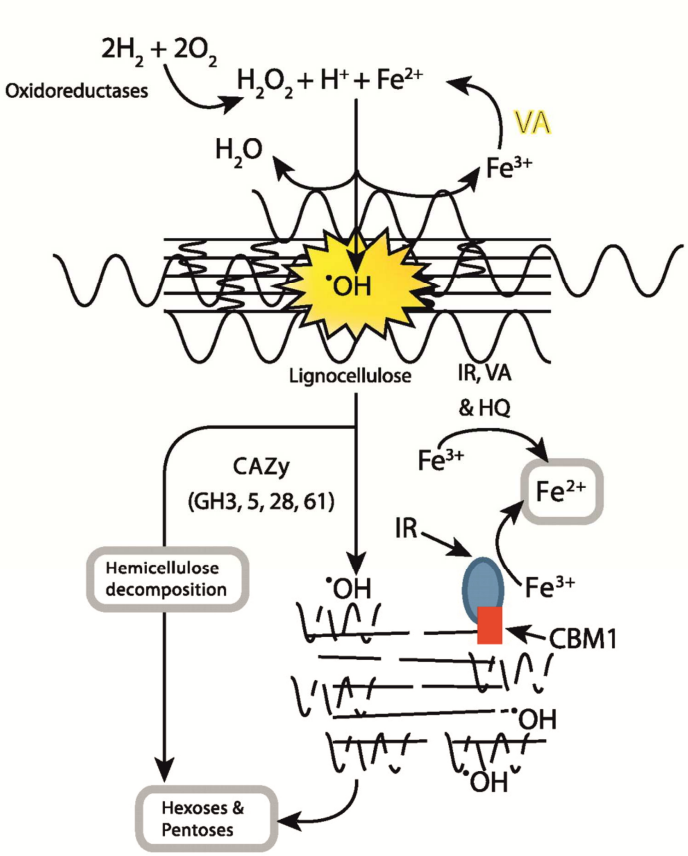



**Figure 1.17. Schematic overview of the deconstruction of lignocellulose by hydroxyl radicals.**
Several mechanisms facilitate the reduction of $Fe^{3+}$ to $Fe^{2+}$; reduction by phenolate compounds such as variegatic acid (VA), reduction by hydroxyquinones (HQ) and reduction by iron reductase enzymes (IR) which may possess carbohydrate binding modules (CBM1) which bind them to cellulose chains. Extracellular oxidoreductases produce hydrogen peroxide to facilitate the production of hydroxyl radicals via the Fenton reaction $H_2O_2 + H^+ + Fe^{2+} \rightarrow H_2O + {}^{\cdot}OH + Fe^{3+}$. Figure reproduced from (116).

### *1.4.1.3 Polysaccharide Monooxygenases*

Recently a new class of cellulose degrading enzymes were discovered in fungi which have been called polysaccharide monooxygenases (PMOs) and which were previously, incorrectly classified as members of a GH family (GH61) (117). Polysaccharide monooxygenases are extracellular, copper containing proteins which require either a soluble redox active chemical cofactor such as gallate or ascorbate, or an electron donating companion enzyme such as cellobiose dehydrogenase for activity (118, 119).

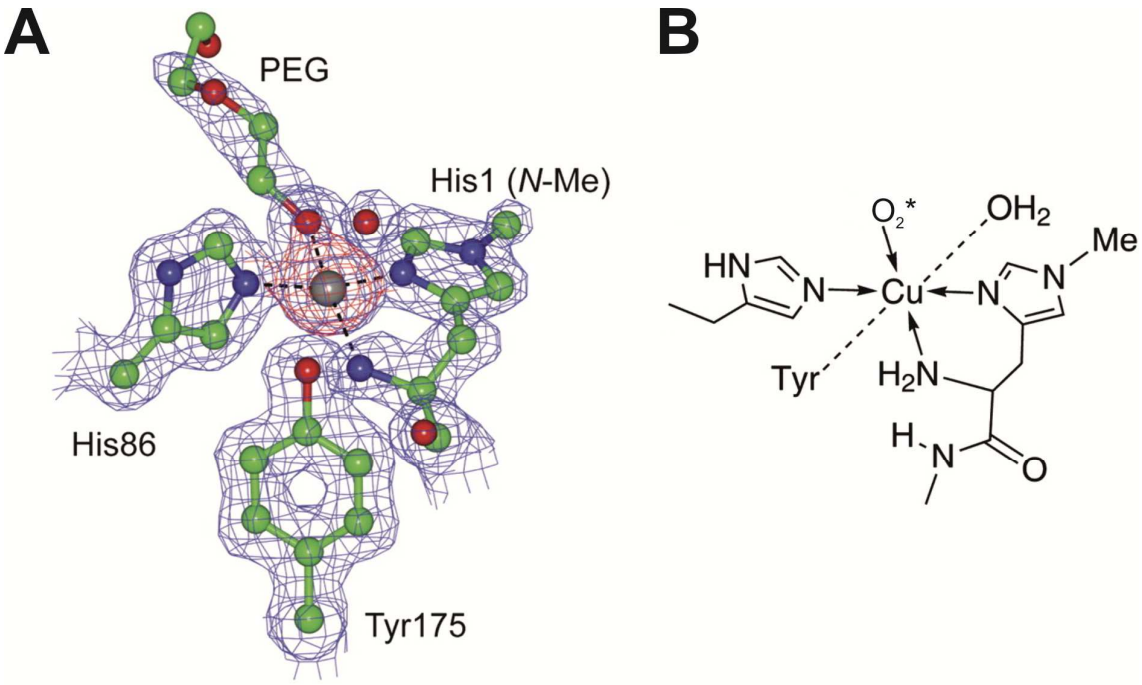



**Figure 1.18. The copper II binding site of polysaccharide monooxygenases.** (A) 3D structure of the copper binding site from *Thermoascus aurantiacus* (PDB entry 2YET). (B) Schematic diagram of the copper binding site of polysaccharide monooxygenases. Figure adapted from (118).

The three dimensional structures of several PMOs from fungi have been previously published and together with electroparamagnetic resonance data show the proteins possess type-2 copper binding sites populated with a $Cu^{2+}$ ion (117, 118, 120, 121). The copper ion is coordinated by two histidine residues (at the Nε of one residue and the amino terminus and Nδ of another), the oxygen atom of a tyrosine residue and a water molecule. The sixth coordination site is thought to be occupied by oxygen during the enzymes catalytic cycle. The structures also revealed that the copper binding site (active site) is present on a flat surface of the proteins (Figure 1.19).

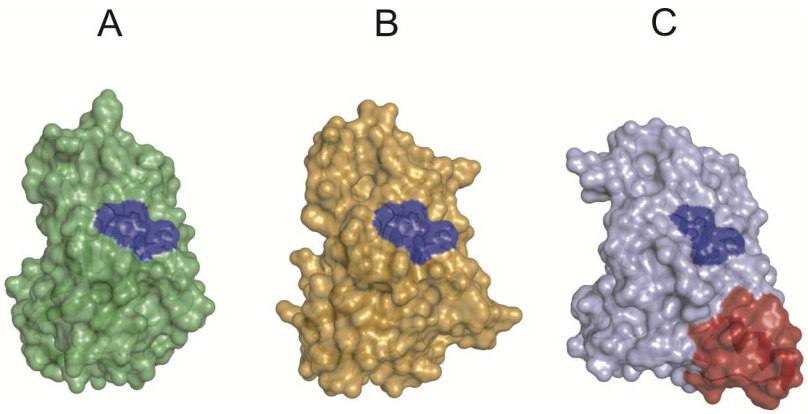



**Figure 1.19. Surface positioning of copper II binding sites in polysaccharide monooxygenases.**
Surface representations of polysaccharide monooxygenases from *Thielavia terrestris* (A), *Phanerochaete chrysosporium* (B – homology model based on A) and *Hypocrea jecorina* (C). PDB entries: A, 3EII; C, 2VTC. Figure adapted from (121).

A reaction mechanism has been proposed for PMOs in which oxygen is inserted at either the C1 or C4 position of glucose units joined by a glycosidic bond which is shown in Figure 1.20 (119, 122). In the first step of the reaction the $Cu^{2+}$ ion of PMO is reduced to $Cu^{+}$ by either an electron from cellobiose dehydrogenase, or a reducing cofactor such as gallate. Molecular oxygen then binds to the $Cu^{+}$ ion to form a copper superoxo intermediate. Depending on whether the PMO in question acts at the reducing or non-reducing end of sugars, the copper superoxo then abstracts a proton from either the C1 or C4 of glucose residues either side of a glycosidic bond generating a copper hydroperoxo intermediate and a substrate radical. A further electron is donated as described above to the copper hydroperoxo intermediate which facilitates O-O bond cleavage, releasing water and generating a copper oxo radical. The copper oxo radical couples with the substrate radical which results in hydroxylation of the substrate at either the C1 or C4 position. This destabilises the glycosidic bond resulting in an elimination reaction taking place, breaking the glycosidic bond. As a result one of the sugars originally involved in the glycosidic bond has been oxidised to a lactone if oxygen was inserted at the C1 position, or a ketoaldose if inserted at the C4 position.

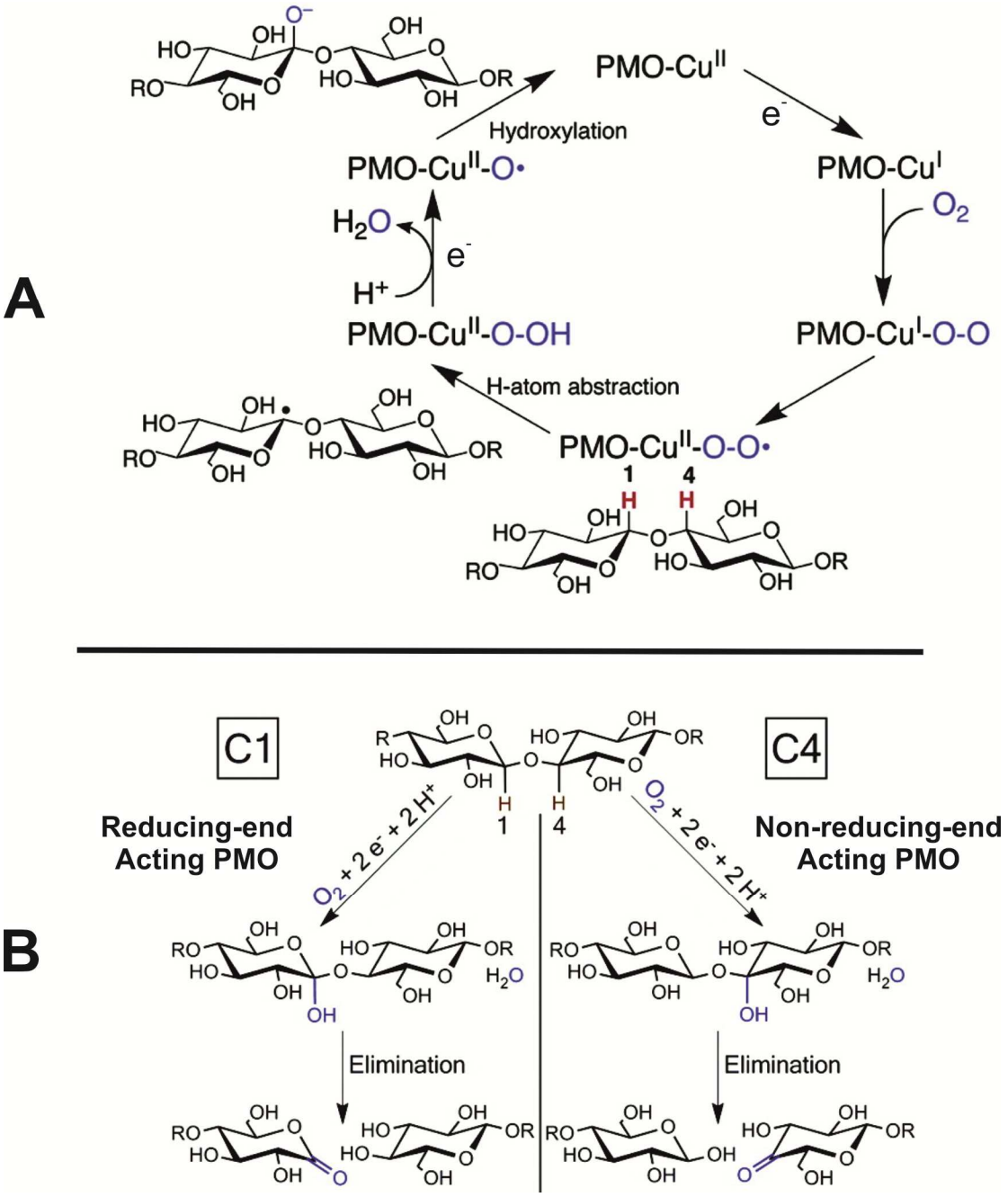



**Figure 1.20. The proposed mechanism and reaction pathway for the oxidative cleavage of cellulose by polysaccharide monooxygenases.**
(A) Proposed mechanism for a reducing end cleaving polysaccharide monooxygenase. (B) Overall reaction pathway. Electrons ($e^-$) are donated either by soluble reducing equivalents such as gallate or enzymatically by cellobiose dehydrogenase. Figure adapted from (119, 122).

The activity of PMOs has been shown to be synergistic with that of conventional cellulases, enhancing the amount of sugar they are able to release in a given time (117-119). It seems likely that as the copper active site is present on a flat surface of the proteins that they are interacting with the surface of crystalline cellulose microfibrils

and inducing breaks in cellulose chains. These enzymes offer a very attractive advantage to the industrial saccharification of lignocellulose; allowing the use of a lower quantity of GH enzymes (loading) to achieve the same level of saccharification, therefore reducing the cost of the enzymatic step. Indeed, the commercially available cellulase cocktail Cellic CTec3 (Novozymes) already includes a PMO.

### 1.4.2 Animal – Symbiont Systems

As well as bacteria and fungi, some animals have developed methods of degrading lignin and utilising cellulose as a source of energy (123). Whilst termites have been shown to possess their own genes to contribute to this process, most studied animals such as woodlice, beetles and cows rely on microbial and fungal gut symbionts to assist them (123-126). However it is thought at least 99% of the symbiont species are not readily culturable (127). This means that deciphering the contributions of each species, as well as those of the host, will prove extremely challenging using today's technologies. Progress is being made however with several recent studies obtaining the host and symbiont transcriptomes or proteomes from termites and cows (124, 125, 128).

## 1.5 *Limnoria quadripunctata* as an Unusual System

The marine wood-boring crustacean, *Limnoria quadripunctata*, commonly known as gribble, is also an example of an animal able to survive on a diet of lignocellulose. However, unlike the animals mentioned above, *Limnoria* has a sterile gut and does not rely on symbionts to aid its digestion of lignocellulosic material (68, 129). This suggests that *Limnoria* possesses all of the enzymes it requires to obtain sugar from wood. In order to gain enough nutrition from lignocellulose *Limnoria* must possess some system for breaking down, or at least modifying the structure of lignin to be able to access a sufficient amount of cellulose.

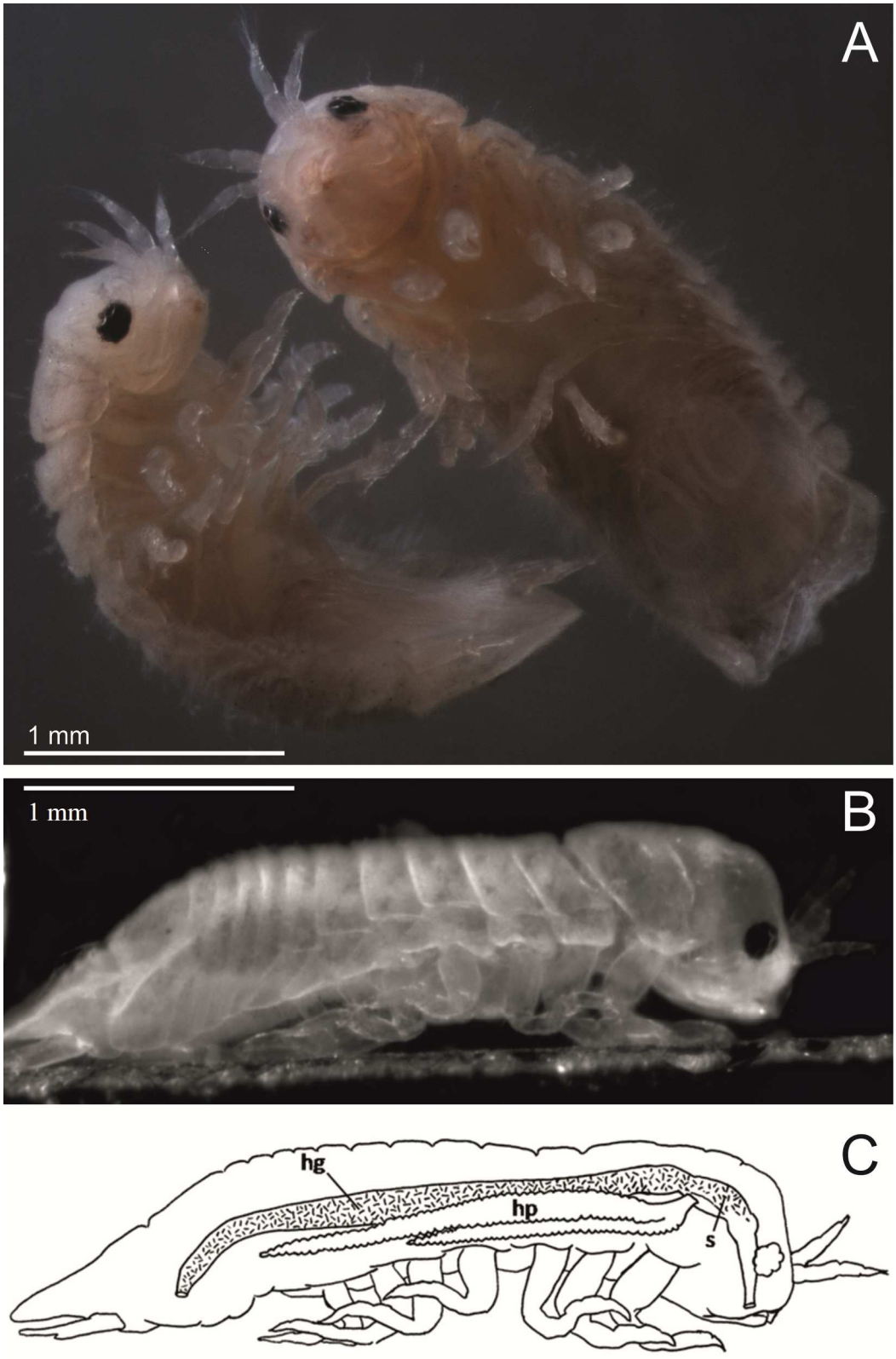



**Figure 1.21. Photomicrographs and a diagram of *L. quadripunctata*.**
Photomicrographs of two specimens of *L. quadripunctata* (A) and a single animal in an artificial burrow (B). Diagrammatic representation of the digestive system of *L. quadripunctata* (C). Abbreviations: hg, hindgut; hp, hepatopancreas; s, stomach. Images courtesy of Dr. Graham Malyon, University of Portsmouth.

A diagram of the digestive system of *Limnoria* is shown in Figure 1.21 – C. The gut of *Limnoria* is composed of a long tube running almost the full length of the animal, which is divided into different sections along its length (130). The gut is surrounded by longitudinal and circumferential muscles (131). Within the head of the animal the tube forms a stomach (s), which is followed by the foregut, then the hindgut (hg). The hindgut is packed full of ingested wood particles (Figure 1.22). A four lobed organ, called the hepatopancreas (HP), sits abreast of the gut, two lobes on either side of the animal, joined to the posterior end of the stomach by two filter channels (132, 133). It is thought that fluids are able to pass between the gut and the HP via these filter channels, whilst ingested wood particles are restricted to the gut. The lobes of the HP are surrounded by rings of muscle, and the interior of the organ is lined with secretory cells (130, 134, 135). The HP is believed to produce digestive enzymes which are secreted into the lumen of the HP and which are subsequently pumped out of the HP into the stomach where they would act upon ingested wood.

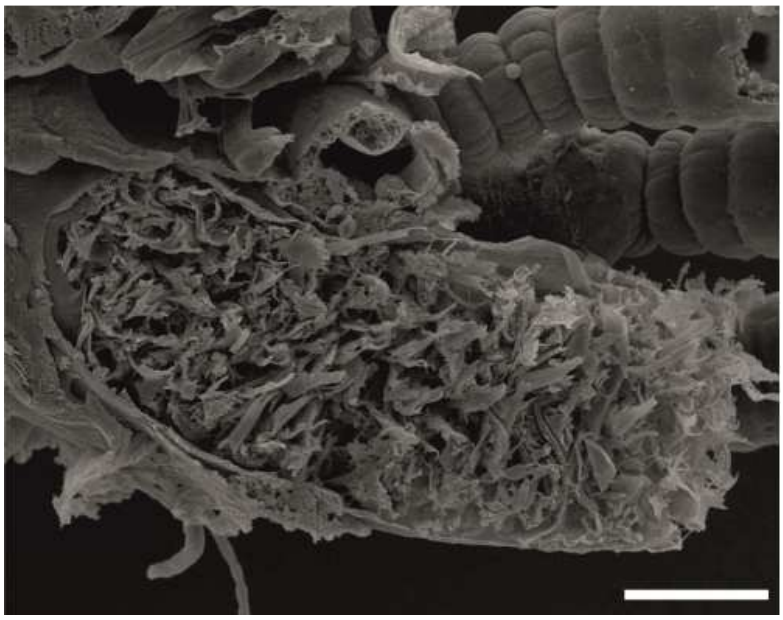


**Figure 1.22. Scanning electron micrograph showing an obliquely sectioned hindgut from *Limnoria*.**
The image shows how densely packed the hindgut is with wood particles. Scale bar, 50 µm. Image reproduced from (61).

The transcriptome of *Limnoria*'s HP has previously been elucidated and has provided some interesting insights as to how *Limnoria* may digest lignocellulose (68). A chart of the distribution of the most represented gene families found in (68) has been reproduced in Figure 1.23. By far the most abundant group of transcripts found in the HP were GH enzymes (27 % of all detected transcripts), of which the GH7 family (as

classified by the CAZy system – www.CAZy.org (136)) represented 53.3%. This is very interesting as previous to this work the GH7 enzymes had not previously been found in an animal system, only fungal and mutualist systems (68).

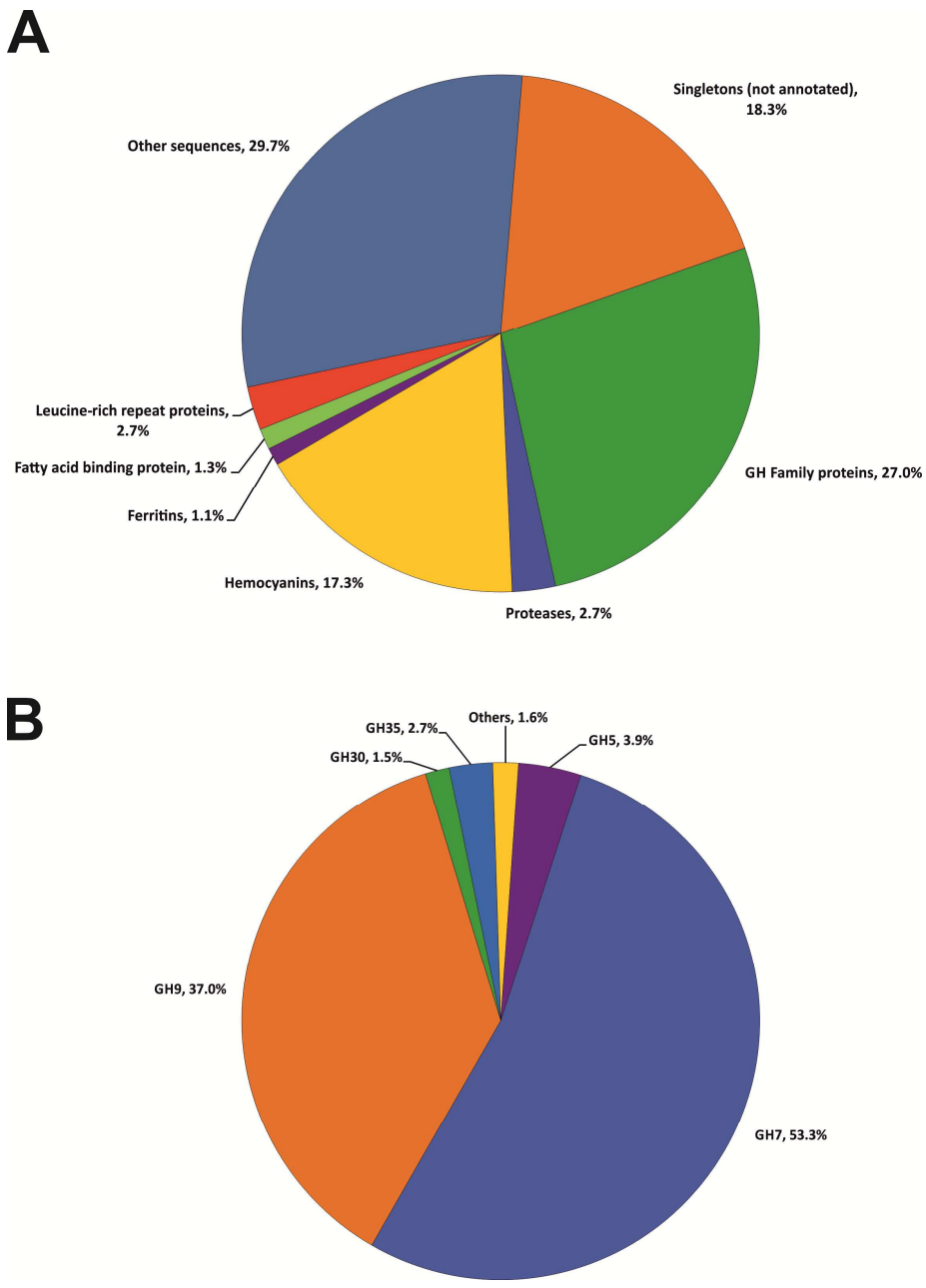



**Figure 1.23. Diagrammatic overview of the transcriptome of *Limnoria*'s HP.** Overview of the most abundantly represented sequence types (A) and a breakdown of the glycosyl hydrolase (GH) families detected (B). Singletons represent individual EST sequences which could not be assembled into larger contiguous sequences and which did not contain enough information to allow annotation.

A GH7 enzyme from *Limnoria*, designated LqCel7B, has since been heterologously expressed and extensively characterised (61). The enzyme was found to behave as a

cellobiohydrolase capable of endoinitiation. This means that LqCel7B processively removes cellobiose units from the end of a cellulose chain; but also that it is able to bind to the middle of a cellulose chain, cleave it open and then proceed to processively remove cellobiose units. Given the high transcript abundance for this family of enzymes, and its mode of activity, this may help explain how *Limnoria* is able to survive without the aid of symbionts on a lignocellulosic diet; being able to attack amorphous regions of cellulose and processively degrade them. It was also discovered by reverse transcription quantitative polymerase chain reaction (qPCR) that LqCel7B is solely expressed in the HP of *Limnoria*, whilst Western blot analysis showed that LqCel7B protein is present in both the HP and the animal's gut (61). This is the first piece of evidence to directly support the above view that enzymes involved in digestion in *Limnoria* are expressed in the HP, but are moved to and act in the gut.

As well as family 7, GH families 9, 5, 35 and 30 in order of decreasing abundance were found to represent more than 1 % of all GH transcripts. Families 2, 13, 16, 18, 20, 31 and 38 were found to represent less than 1 % of all GH transcripts. This shows that in the HP of *Limnoria* were found transcripts representing a broad range of glycoside hydrolase specificities which by CAZy assignment are predicted to cover the core activities responsible for breaking down cellulose and hemicellulose (136).

Another abundant group of transcripts present belonged to the hemocyanin (HC) genes (17.3% of all transcripts). Hemocyanins are type III copper containing proteins which exist freely soluble in the hemolymph of many arthropod and mollusc species and traditionally are thought of as oxygen carrying proteins (137). They are discussed in more detail in Chapter 4.

Other well represented groups of genes found to be present were, in order of decreasing abundance, proteases, leucine-rich repeat containing proteins, fatty acid binding proteins and ferritins. As it was found that LqCel7B was abundantly expressed only in the HP, but that its protein was present in the animals gut were it is involved in digestion, it appears possible that some of these other groups of genes found to be highly expressed in the HP of *Limnoria* may also play roles in lignocellulose digestion.

## 1.6 Aims of this Thesis

The broad aim of the work described in this thesis is to discover new enzymes which could be used to improve the industrial process of producing second generation bioethanol from lignocellulosic feedstocks. In order to do this the marine wood borer, *Limnoria quadripunctata*, is being used as a source from which to prospect for enzymes. *Limnoria* is able to utilise wood, along with any organisms with which it is infested, as a sole source of nutrition without the need for microbial or fungal gut symbionts. Therefore the animal is able to gain sufficient nutrition from the wood using its own, genome encoded, enzymes. Understanding how the animal is able to do this may provide information which could be applied in industrial fermenters for the improved saccharification of lignocellulosic material. A previous study has already investigated which genes are expressed in the HP of *Limnoria*, the organ believed to be the site of production of digestive enzymes. This study has suggested certain genes which may be involved in the digestive process.

The work described in this thesis will aim to determine which genes could be involved in the process of digesting wood. A proteomic approach will be carried out on HP and gut tissue from *Limnoria* to determine which proteins dominate these two organs. This data will be used in conjunction with that of the previous transcriptomic study to identify genes for further study. A promising candidate gene identified will be subject to *in vitro*, *in vivo* and *ex vivo* investigations to determine how it contributes to the process of lignocellulose breakdown.

# Chapter 2: Materials and Methods

## 2.1 Suppliers

Unless otherwise stated all chemicals and consumables used in this work were supplied by common scientific suppliers.

DNA modifying enzymes such as restriction endonucleases, ligases, nucleases and polymerases were sourced from Promega, Finnzymes and Fermentas.

DNA oligonucleotides were produced by Sigma Aldrich.

All water used to prepare buffers and solutions was purified by an Elga Purelab Ultra system to genetic or analytic grade.

## 2.2 Organisms and Plasmids

### 2.2.1 *E. coli* Strains

The *E. coli* strains used in this work are detailed in Table 2.1.

**Table 2.1. Details of the *E. coli* strains used in this work.**

| Strain | Properties | Manufacturer |
|---|---|---|
| StrataClone SoloPack | Routine cloning strain expressing Cre recombinase to facilitate the circularising of topoisomerase I mediated ligations | Agilent Technologies |
| DH5α | Routine cloning strain | Invitrogen – Life Technologies |
| XL1-Blue Sub-cloning | Routine cloning strain | Agilent Technologies |
| XL1-Blue Supercompetent | Highly competent routine cloning strain | Agilent Technologies |
| XL10-Gold Ultracompetent | Very highly competent routine cloning strain | Agilent Technologies |
| BL21 (DE3) pLysS | Protein expression strain expressing T7 lysozyme to suppress background recombinant expression | Invitrogen – Life Technologies |
| BL21 (DE3) Star™ | Protein expression strain possessing a mutated RNaseE gene to enhance mRNA stability | Invitrogen – Life Technologies |
| Rosetta-gami 2 (DE3) | Protein expression strain with mutations in thioredoxin reductase and glutathione reductase genes to enhance disulphide bond formation in in the cytoplasm.<br>Also carries the pRARE2 plasmid which encodes genes for seven tRNAs which recognise codons rarely found in bacteria. These codons are AUA (Ile), AGG (Arg), AGA (Arg), CUA (Leu), CCC (Pro), CGG (Arg) and GGA (Gly). | Novagen-Merck |
| ArcticExpress (DE3) RIL | Protein expression strain expressing the Cpn10 and Cpn60 chaperonin genes from the psychrophilic bacterium *Oleispira antartica* which facilitate protein refolding between 4 – 12 ℃.<br>Also carry genes encoding for tRNAs which recognise four codons rarely found in bacteria. These codons are AGA (Arg), AGG (Arg), AUA (Ile) and CUA (Leu). | Agilent Technologies |

### 2.2.2 Plasmids

The StrataClone blunt PCR cloning plasmid (pSC-B, Agilent Technologies) was routinely used for the sub-cloning of PCR products. Protein expression was carried out using a range of pET vectors (Novagen - Merck) and the pGEX-4T-3 vector (GE Healthcare Life Sciences). The features of each of these vectors are shown in Table 2.2.

**Table 2.2. Features of the plasmids used in this work for sub-cloning and expression.**

| Vector | Antibiotic Resistance | Features | Manufacturer |
|---|---|---|---|
| pSC-B | Carbenicillin or kanamycin | Allows for the efficient sub-cloning of blunt ended PCR products using a topoisomerase I/Cre recombinase mediated mechanism | Agilent Technologies |
| pET26b(+) | Kanamycin | Carries an N-terminal *pelB* signal sequence which has the potential to locate the expressed protein to the periplasm | Novagen - Merck |
| pET28a(+) | Kanamycin | Fuses a $His_6$ affinity tag to the N-terminus which is cleavable using thrombin | Novagen - Merck |
| pET52b(+) | Carbenicillin | Fuses a Strep·Tag® II affinity tag to the N-terminus which is cleavable using the HRV 3C protease | Novagen - Merck |
| pGEX-4T-3 | Carbenicillin | Fuses a glutathione-S-transferase gene tag to the N-terminus which is cleavable using thrombin | GE Healthcare Life Sciences |

## 2.3 Media

SOC media was used during the recovery phase of bacterial transformation procedures and was purchased as sterile aliquots from Invitrogen – Life Technologies.

### 2.3.1 Luria-Bertani Broth, Miller

Luria-Bertani broth, Miller (LB) was produced by dissolving 10 g of tryptone, 5 g of yeast extract powder and 10 g of sodium chloride in a final volume of 1 L of water. The solution was then autoclaved prior to use.

### 2.3.2 Luria-Bertani Broth Agar

Luria-Bertani broth agar (LBA) was used for the culturing of *E. coli* in petri dishes. To make LBA, agar was added to LB prior to autoclaving at a concentration of 1.5 % w/v.

### 2.3.3 2x YT broth

During protein expression 2x YT broth was often used rather than LB. The medium contained 16 g of tryptone, 10 g of yeast extract powder and 10 g of sodium chloride dissolved in a final volume of 1 L of water. The solution was autoclaved prior to use.

## 2.4 Molecular Biology Techniques

### 2.4.1 Polymerase Chain Reaction (PCR)

Oligonucleotide primers for PCR were designed to meet several criteria. Each primer should have a melting temperature ($T_m$) between 60 and 70 ℃, and the $T_m$s of each primer pair should not differ by more than 3 ℃. The GC content of each primer should not exceed 60 %. No predicted secondary structures should greatly exceed a ΔG value of -10.0 kcal/mol. $T_m$ values were calculated using the Thermo Scientific Tm Calculator (http://www.thermoscientificbio.com/webtools/tmc) as recommended for the Phusion DNA polymerase. Predicted secondary structures were assessed using NetPrimer (http://www.premierbiosoft.com/netprimer/ - Premier Biosoft).

Amplification of genes of interest was carried out using the Phusion® Hot Start High-Fidelity DNA Polymerase (Thermo Fisher Scientific) and was carried out using a PTC-200 Peltier Thermal Cycler (MJ Research). Reactions ranged from 20 – 50 µl in volume and each contained final concentrations of 1x HF buffer, 200 µM of each dNTP, 0.5 µM of each primer, 0.02 U/µl of Phusion DNA polymerase and template DNA was used within the range of 0.02 - 0.2 ng/µl of reaction. The following thermal cycle was used unless otherwise stated; initial denaturation at 98 ℃ for 5 minutes, 35 cycles of denaturation at 98 ℃ for 10 seconds, annealing for 20 seconds, extension at 72 ℃ for 15 seconds/kb and a final extension period of 5 minutes at 72 ℃. Annealing temperatures varied according to each primer pair but were calculated as 3 ℃ higher than the $T_m$ of the primer with the lower $T_m$.

### 2.4.2 Purification of DNA Fragments

PCR products from PCR reactions were purified to remove contaminants such as enzymes, buffer and dNTPs using a QIAquick PCR Purification kit (QIAGEN) following the manufacturer's instructions. DNA was eluted using 40 µl of Buffer EB (QIAGEN).

### 2.4.3 Nucleic Acid Quantification

The concentration of RNA or double stranded DNA in a sample was quantified by analysing 1.2 µl of the sample using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

### 2.4.4 Agarose Gel Electrophoresis

DNA fragments were visualised and separated by agarose gel electrophoresis. Gels were typically composed of 1 % w/v agarose in 0.5x Tris-borate-EDTA buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA - TBE) and 0.5 µg/ml ethidium bromide. Samples were mixed with GelPilot 5x DNA Loading Dye (QIAGEN) prior to running. Hyperladder™ 1kb (Bioline) was used as a molecular weight marker. Gels were typically ran for 30 – 60 minutes at between 90 and 130 V depending on the size of the gel. DNA was visualised with UV light using a UviTec transilluminator (Cambridge).

### 2.4.5 Extraction of DNA Fragments from Agarose Gel

DNA contained in agarose gel slices was extracted using a QIAquick® Gel Extraction kit (QIAGEN) following the manufacturer's instructions. DNA was eluted using 40 µl of Buffer EB (QIAGEN).

### 2.4.6 Plasmid DNA Preparation

Plasmid DNA was prepared by inoculating 5 ml of LB containing the appropriate antibiotic with freshly transformed bacteria carrying the plasmid of interest. The culture was grown for 16 hours at 37 ℃ shaking at 200 rpm. The cells were harvested by centrifugation and the plasmid DNA extracted using a QIAprep® Spin Miniprep kit (QIAGEN). DNA was eluted using 40 µl of Buffer EB (QIAGEN).

### 2.4.7 DNA Sequencing and Analysis

DNA was sequenced by the Genomics Laboratory, Technology Facility, University of York, using a 3130XL Genetic Analyser (Applied Biosystems). Primers used for sequencing are shown in Table 2.3.

**Table 2.3. A list of the primers used in this work for DNA sequencing.**

| Primer | Sequence (5' → 3') |
|---|---|
| M13 Forward | TGTAAAACGACGGCCAGT |
| M13 Reverse | AGGAAACAGCTATGACCAT |
| HC3 Internal F | CAAGCGTATTCAGCTAAGATGAC |
| HC3 Internal R | GATCCAGGTACCAATGTTCTCC |
| T7 Promoter | TAATACGACTCACTATAGGG |
| T7 Terminator | GCTAGTTATTGCTCAGCGG |
| pBACSEC Forward | ACGGATCTCTAGAGG |
| pBACSEC Reverse | CTCTAGTTTATTAGG |

Sequence chromatograms were analysed using ContigExpress (part of Vector NTI Advance® software package version 11.2 – Invitrogen – Life Technologies).

### 2.4.8 Colony Screening

After bacterial transformations colonies were routinely screened by PCR to identify those colonies which did possess the correctly transformed DNA (positive transformants). Screening was performed by picking a colony from a plate using a sterile pipette tip and stirring the tip in a prepared PCR reaction mixture. If blue/white screening was being used only white or very pale blue colonies were picked. The PCR reaction mixture contained 5 µl of DreamTaq™ Green PCR Master Mix 2x (Fermentas), 0.5 µl each of forward and reverse primer (0.5 µM final concentrations) and 4 µl of sterile, nuclease free water. DreamTaq™ Green PCR Master Mix 2x contains a proprietary Taq DNA polymerase, $MgCl_2$, dNTPs and a loading dye for later agarose gel electrophoresis analysis. The primer pair used varied depending on the vector transformed but typically was either the M13 forward and reverse primers or the T7 promoter and terminator primers detailed in Table 2.3.

The following thermal cycle was applied to the PCR reactions; initial denaturation at 95 ℃ for 2 minutes, 35 cycles of denaturation at 95 ° C for 30 seconds, annealing at 54 ℃ for 30 seconds, extension at 72 ℃ for 60 seconds/k b, a final extension period of 5 minutes at 72 ℃. Afterwards 7 µl of each reaction was run on an agarose gel to determine whether or not the picked colony was a positive transformant (see above).

### 2.4.9 Restriction Endonuclease Digestion of DNA

Digestion of DNA was carried out using restriction endonuclease enzymes and compatible buffers obtained from Promega. DNA was typically digested using 10U of enzyme per 0.5 µg of DNA, 1x suitable buffer and water to make the reaction up to 20 µl. Reactions were usually incubated at 37 ℃ for 2 hours prior to separation of cut and un-cut DNA by agarose gel electrophoresis.

### 2.4.10 StrataClone Mediated Sub Cloning

Genes of interest (GOI) that were to be ultimately cloned into expression vectors were routinely sub cloned into the StrataClone Blunt (pSC-B, Agilent Technologies) vector first. This sub cloning system was extremely efficient and made the subsequent cloning into an expression vector using the DNA ligase mediated cloning procedure much easier and more likely to succeed first time.

The gene to be cloned was amplified by PCR using a proof reading blunt end forming DNA polymerase such as Phusion as described in Section 2.4.1. The DNA produced in the PCR reaction was then purified as described in Section 2.4.2. A 2 µl aliquot of the purified PCR product was then visualised on an agarose gel. If the DNA produced a strong single band on the gel then the remaining DNA was diluted 10 times with nuclease free water.

The cloning reaction was set up in 0.2 ml thin walled PCR tubes and the following components added to the tube in the following order: 3 µl of StrataClone Blunt Cloning Buffer (Agilent Technologies), 2 µl of diluted PCR reaction, 1µl of StrataClone Blunt Vector Mix (Agilent Technologies). The components were mixed by gentle pipetting and allowed to sit at room temperature for 5 minutes before being placed on ice. A 1 µl aliquot of the mixture was then transformed into StrataClone SoloPack competent cells (Agilent Technologies) using the standard procedure described in Section 2.4.14 followed by blue/white screening.

### 2.4.11 DNA Ligase Mediated Cloning

To place most GOIs into their destination expression vector a standard cloning procedure was followed.

Double restriction endonuclease digestions were set up for both the sub cloned GOI (i.e. pSC-B-GOI construct) and the destination expression vector as described in Section 2.4.9. These digestions were typically performed in 20 µl volumes and contained as much DNA as was possible in this volume. The digest reactions were then separated by agarose gel electrophoresis and the bands containing cut DNA excised and the DNA within them extracted as previously described. The amount of DNA retrieved from each digestion was determined and ligation reactions set up.

A typical ligation reaction contained 1 µl of T4 DNA Ligase (Promega), 1 µl of 10x Ligase Buffer (Promega), vector and insert DNA double digested with the same restriction endonuclease pair, and water to make up the reaction volume to 10 µl. The reactions were incubated at 4℃ for 16 hours.

The amount of DNA used varied. Usually two reactions were set up per ligation, one using a 3:1 ratio of insert to vector DNA, and another using a 1:1 ratio. Reactions were

set up to include the maximum amount of DNA possible to yield the desired ratio in the available volume.

After ligation 5 µl of the ligated DNA were transformed into any suitable highly competent cloning strain of *E. coli* listed in Table 2.1 using a standard transformation procedure (see below).

## 2.4.12 In-Fusion® HD Cloning

The In-Fusion® HD cloning system (Clontech Laboratories) allows DNA inserts to be ligated into any vector without the need to go via a sub cloning step, whilst still retaining the much higher efficiency and simplicity of a topoisomerase based sub cloning system.

Suitable primers were designed for each GOI following the guidance given in the In-Fusion® HD manual. The major different feature of these primers compared to conventional primers is the inclusion of a 5' 15 base pair region of homology to the cut 3' end of the destination vector. These primers were used to amplify the desired GOIs as described in Section 2.4.1. A 1 µl aliquot of the PCR reactions was analysed by agarose gel electrophoresis to confirm the success of the amplification.

The destination expression vector was linearised using a single restriction endonuclease as described in Section 2.4.9 to yield cut ends compatible with the regions of homology present in the primers used.

Both the GOI PCR product and linearised vector were purified as described in Section 2.4.2 and the concentration of DNA recovered quantified. In-Fusion® cloning reactions were then prepared containing 2 µl of 5x In-Fusion® HD Enzyme Premix, 50 – 200 ng of linearised vector, 10 – 200 ng of purified PCR product and sufficient water to make the volume up to 10 µl. The cloning reactions were then incubated for 15 minutes at 50 ℃ before placing them on ice.

Finally 2.5 µl of the cloning reaction was transformed into any suitable highly competent cloning strain of *E. coli* listed in Table 2.1 using a standard transformation procedure.

### 2.4.13 Preparation of Chemically Competent *E. coli* Cells

Cells of the desired bacterial strain were grown on an LBA plate containing all appropriate selection antibiotics. A colony was picked from the plate and used to inoculate 100 ml of LB containing the same selection antibiotics. This culture was incubated whilst shaking at 200 rpm until an optical density at 600 nm ($OD_{600}$) of 0.6 was reached. The culture was then chilled on ice for 20 minutes after which the cells were harvested by centrifugation at 4 ℃. The cells were resuspended in 20 ml of ice cold SEM buffer (250 mM KCl, 55 mM $MnCl_2$, 15 mM $CaCl_2$, 10 mM PIPES). The cell suspension was kept on ice for 10 minutes before harvesting the cells by centrifugation at 4 ℃. The cells were resuspended in 6 ml of fresh ice cold SEM buffer and 460 µl of DMSO (final concentration 7 % v/v). The cell suspension was aliquoted and snap frozen in liquid nitrogen before storage at -80 ℃. Transformation using cells produced with this method was still efficient after storage of up to 18 months.

### 2.4.14 Transformation of Chemically Competent *E. coli* Cells

An aliquot of frozen chemically competent cells was allowed to thaw on ice. Meanwhile 1 µl of a ten-fold dilution of a plasmid preparation or 5 µl of a ligation reaction was placed into a sterile 1.5 ml tube and chilled on ice. Once thawed, 50 µl of cells were added to the tube containing DNA and the mixture gently mixed by stirring. The cells were incubated with DNA for 20 minutes on ice before exposing them to a heat shock at 42 ℃ for 45 seconds. After placing the cells on ice for 2 minutes 250 µl of SOC medium was added and the cells allowed to recover at 37 ℃ with shaking at 200 rpm for 1 hour. The cells were then plated onto LBA plates containing the appropriate selection antibiotic. If blue/white screening was to be carried out the LBA plates had previously been spread with 40 µl of 2 % w/v X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) in dimethylformamide. The plates were then incubated at 37 ℃ for 16 hours.

# 2.5 Protein Techniques

## 2.5.1 Protein Quantification

Protein concentration was determined using a Bradford assay. In a 96 well optical microtitre plate 5 µl of samples and standards were added to 250 µl of Coomassie Plus™ Protein Assay Reagent (Thermo Fisher Scientific). The plate was incubated in the dark at room temperature for 20 minutes before reading the $OD_{595}$ of each sample using a Sunrise™ plate reader (Tecan).

In this way a linear standard curve was produced using bovine serum albumin (BSA) solutions over a concentration range of 0.031 – 1 mg/ml and this curve used to back-calculate the concentration of samples.

## 2.5.2 Concentrating Protein Samples

Fractions obtained from protein chromatography were routinely concentrated by centrifugation. Samples were loaded into either Vivaspin 20 10,000 MWCO (Sartorius) or U-Tube™ 6 – 10 10,000 MWCO (Novagen) concentrating devices and span according to the manufacturer's recommendations in swinging bucket centrifuge until the samples had reduced to the desired volume.

## 2.5.3 Buffer Exchange (Desalting)

Zeba™ Spin Desalting Columns (7,000 MWCO, Thermo Fisher Scientific) were used to exchange the buffer of a protein sample. This was done according to the manufacturer's instructions.

## 2.5.4 Bacterial Cell Lysis for Protein Extraction

### *2.5.4.1 Small Scale*

Small scale relates to the cells from the equivalent of 1 – 100 ml of culture.

Cells were resuspended in an appropriate lysis buffer with any desired additives, specific compositions will be discussed where relevant later. Lysis buffer was added in the ratio of 0.5 ml of lysis buffer per 10 ml of culture. The vessel containing the cell suspension was secured in crushed ice and the Microtip™ probe of an S-4000 sonicator (Misonix Sonicators) positioned 3 – 5 mm into the suspension. The cells were then lysed by sonication using 10 cycles of 5 seconds of 60 % maximum amplitude, 5 seconds rest.

The cell lysate was then fractionated into a soluble fraction and insoluble debris fraction by centrifugation at 14,000 rpm for 25 minutes at 4 ℃ using a benchtop centrifuge.

#### *2.5.4.2 Large Scale*

Large scale relates to the cells from the equivalent of 100 ml of culture or more.

Cells were resuspended in an appropriate lysis buffer with any desired additives, specific compositions will be discussed where relevant later. Lysis buffer was added in the ratio of 2 ml of lysis buffer per gram of wet cell paste. The cell suspension was then carefully poured into the inlet of a TS2 Benchtop Cell Disruptor (Constant Systems Ltd.). The cell disruptor was precooled and maintained at 4 ℃ by a circulation of chilled water. The cell disruption was performed automatically by the disruptor using a pressure of 25 kPSI.

The cell lysate was then fractionated into a soluble fraction and insoluble debris fraction by centrifugation at 15,000 rpm for 25 minutes at 4 ℃ using a RC-5B centrifuge (Sorvall) equipped with an SS-34 rotor. The soluble fraction was often then subjected to protein purification.

### 2.5.5 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Standard discontinuous SDS-PAGE systems provided by Bio-Rad and Expedeon were used as described by Laemmli (138). A general procedure is given with the variables for each system shown in Table 2.4. Gels were loaded into gel tanks and both electrode chambers filled with running buffer. Prior to loading, samples were mixed with an appropriate amount of 5x SDS sample loading buffer and boiled for 5 minutes. The

samples were then briefly centrifuged to bring all liquid to the bottom of the tube and to loosely pellet any non-dissolved material. At least one lane of each gel contained 5 µl of PageRuler Plus Prestained Protein Ladder (Thermo Fisher Scientific) as a molecular weight marker. Gels were run for approximately 1 hour until the dye front ran off of the end of the gel.

**Table 2.4. SDS-PAGE system variables.**
Details of the different equipment, reagents or conditions used for running SDS-PAGE gels with both Bio-Rad and Expedeon systems.

| | **Bio-Rad System** | **Expedeon System** |
|---|---|---|
| Gels | 4 % v/v acrylamide stacking region, 10 % v/v acrylamide resolving region cast using a Mini-PROTEAN® II system (Bio-Rad) | 10 % or 4 – 20 % v/v acrylamide RunBlue SDS-PAGE Pre-cast gels (Expedeon) |
| Gel Tank | Mini-PROTEAN® II (Bio-Rad) | Mini-PROTEAN® II (Bio-Rad) or RunBlue (Expedeon) |
| Running Buffer | 25 mM Tris, 192 mM glycine, 0.1 % w/v SDS | RunBlue RAPID buffer (for reducing conditions) |
| | pH 8.2 – 8.3 | 30 mM MOPS, 60 mM Tris, 0.1 % w/v SDS, 6.5 mM sodium bisulphite |
| | | pH 8.2 – 8.3 |
| Loading Buffer | 5x SDS sample loading buffer | |
| | 66 mM Tris-HCl pH 6.8, 21 % v/v glycerol, 2.1 % w/v SDS, 5.3 % v/v 2-mercaptoethanol, 39 % v/v water, 0.025 % w/v bromophenol blue | |
| Run Conditions | 200 V | 180 V, 90 mA/gel |

After electrophoresis the stacking gel was cut away from the resolving gel and discarded. Gels to be visualised with Coomassie were placed into a 9 cm square petri dish containing ~20 ml of InstantBlue™ stain (Expedeon). After a minimum of 1 hour in the Coomassie stain gels were washed three times in water for 5 minutes. Coomassie stained gels were then visualised using a UviTec transilluminator (Cambridge) with a white light convertor screen.

## 2.5.6 Western Blot Analysis

All incubation and wash steps were carried out with gentle agitation on a gel rocker unless otherwise stated.

### *2.5.6.1 Transfer*

SDS-PAGE gels were blotted using a semi-dry method. SDS-PAGE resolving gels were first briefly rinsed in water before equilibrating them in Towbin buffer for 20 minutes (25 mM Tris, 192 mM glycine, 20 % v/v methanol, pH 8.2 – 8.3).

To blot the gels the following stack was assembled in a Trans-Blot® SD Semi-Dry Transfer Cell (Bio-Rad): Bottom to top; Extra Thick Blot Paper (Bio-Rad), Protran BA 85 Nitrocellulose membrane (Whatman™ - GE Healthcare Life Sciences), SDS-PAGE gel, Extra Thick Blot Paper. Each layer was first immersed in Towbin buffer, and after adding each layer a smooth glass tube was rolled over the stack to remove air bubbles. A voltage of 25 V was applied across the stack for 20 minutes.

Once the transfer was complete the nitrocellulose membrane (blot) was removed from the unit and the rest of the stack discarded. The blot was incubated with Ponceau S solution (0.1 % w/v Ponceau S, 5 % v/v acetic acid) for 5 minutes and then partially de-stained with water to verify the transfer. If the transfer was successful the blot was completely de-stained by washing with tris-buffered saline tween solution (TBST; 100 mM Tris-HCl pH 7.5, 20 mM NaCl, 0.1 % w/w Tween-20).

### *2.5.6.2 Blocking*

Blots were incubated in blocking buffer (5 % w/v skimmed milk powder in TBST) for 16 hours at 4 ℃ with gentle agitation.

### *2.5.6.3 Primary Antibody*

Blocking solution was discarded and replaced with 10ml of a dilution of primary antibody in blocking buffer. The amount of dilution required depended on the antibody used; for commercially obtained antibodies the manufacturer's guidelines were

followed, for custom antibodies made from recombinant *Limnoria* protein the dilution was determined empirically during the purification of the antibody from crude serum.

The primary antibody was incubated with blots for 2 hours at room temperature. Blots were then washed for 5 minutes three times with TBST. If the primary antibody used was conjugated to horse radish peroxidase (HRP) the next step was to visualise the blot, else a secondary antibody was applied.

### *2.5.6.4 Secondary Antibody*

The secondary antibody (always a commercially obtained antibody) was diluted as recommended by the manufacturer in blocking solution and 10 ml of this dilution applied to blots for 2 hours at room temperature. Blots were then washed for 5 minutes three times with TBST before proceeding to visualise the blot. A 1 in 1250 fold dilution of a horseradish peroxidase conjugated anti-rabbit antibody (product number #32460, Thermo Fisher Scientific) was used as a secondary antibody when using the anti-hemocyanin primary antibody. A 1 in 10000 fold dilution of a horseradish peroxidase conjugated anti-sheep antibody (product number #A3415, Sigma Aldrich) was used as a secondary antibody when using the anti-glycosyl hydrolase 7 primary antibody.

### *2.5.6.5 Visualisation*

The two substrate components of a SuperSignal® West Pico Chemiluminescent Substrate kit (Thermo Fisher Scientific) were mixed in a 1:1 ratio to provide 2 ml of solution per blot to be visualised. This solution was poured on top of blots making sure to cover the whole surface of the membrane and the membranes undisturbed for 5 minutes. After 5 minutes the solution was discarded and all excess liquid carefully dripped off blots, which were then placed between two layers of plastic wrap. In a dark room the blot was exposed to Amersham Hyperfilm™ ECL x-ray film (GE Healthcare Life Sciences) for various amounts of time. The films were then automatically processed using a Xograph Compact 4 developer (Xograph Healthcare).

## 2.5.7 Native Polyacrylamide Gel Electrophoresis (Native PAGE)

When native PAGE gels were run with a molecular weight marker the HMW Native Marker Kit (GE Healthcare Life Sciences) was used. The marker is supplied as lyophilised protein. Each vial of lyophilised protein was reconstituted in 100 µl of 1x Native-PAGE Sample Buffer (Expedeon) and 5 µl of this solution was used per lane on a native gel. Reconstituted marker was aliquoted for single use and stored at -20 ℃.

### *2.5.7.1 Expedeon System*

Expedeon 3 – 20 % Native/DNA-PAGE RunBlue precast gels were used. Gels were loaded into either a RunBlue gel tank (Expedeon) or Mini-PROTEAN® II gel tank (Bio-Rad). Both chambers of the gel tank were filled with Native-PAGE Running Buffer (40 mM tricine, 60 mM Tris, the solution should naturally have a pH of ~8.3 at 25 ℃). Protein samples were mixed with an appropriate amount of 4x Native-PAGE Sample Buffer (40 % v/v glycerol, 4 % v/v Ficoll-400, 0.8 M triethanolamine-Cl pH 7.7, 0.025 % w/v phenol red, 0.025 % w/v Coomassie Brilliant Blue G250, 2 mM di-sodium EDTA - Expedeon).

After loading the protein samples the gels were ran at 140 V with a limit of 110 mA of current per gel until the loading dye ran off the gel.

After electrophoresis gels were either Coomassie stained or Western blotted as described above.

### *2.5.7.2 Amersham ECL Gel System*

Amersham ECL 4 – 12 % precast gels (GE Healthcare Life Sciences) were used. Approximately 90 ml of native running buffer (192 mM glycine, 25 mM Tris-HCl pH 8.3) was added to each buffer reservoir of the horizontal gel tank. A gel was placed onto the gel tank so that each of the legs of the gel cassette was placed into a buffer reservoir. Gels were pre-run for 12 minutes at 160 V.

After pre-running the gel's comb was removed and ~6 ml of native running buffer added to the well section. Protein samples were mixed with an appropriate amount of

4x Native-PAGE Sample Buffer (Expedeon) before being loaded into the wells. Gels were run at 160 V until the dye front had reached the level of the anode buffer reservoir.

After electrophoresis gels were treated as already described for standard SDS-PAGE gels.

## 2.6 Analytical Techniques

### 2.6.1 In-Gel Di-Phenoloxidase (DPO) Enzyme Activity Assay

This DPO assay is based on the method described by Mason (139) utilising the improvements to that assay later described by Winder and Harris (140). In the DPO assay a phenoloxidase enzyme catalyses the oxidation of dopamine to its corresponding quinone, dopaminequinone. The compound 3-methyl-2-benzothiazolinone hydrazine hydrochloride hydrate (MBTH) present in the assay then undergoes a spontaneous Michael addition reaction with dopaminequinone to form a highly visible stable pink/purple precipitant product. When the enzyme carrying out the phenoloxidation reaction is restrained in a polyacrylamide gel matrix a band of pink/purple staining appears in the gel at the location of the enzyme. These areas of staining are referred to as activity bands.

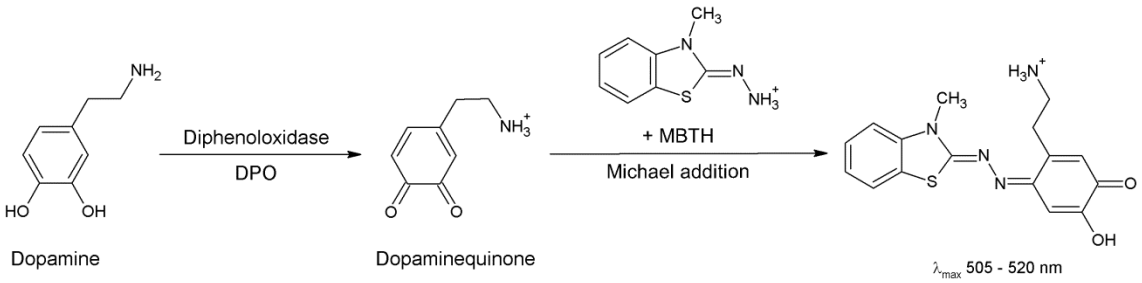



**Figure 2.1. Diphenoloxidase (DPO) assay reaction scheme.**

Protein samples to be tested for DPO activity were ran on a native PAGE gel as described in Section 2.5.7. Aliquots of 0.5 µg of a tyrosinase from the mushroom species *Agaricus bisporus* (product number #T3824, Sigma Aldrich) were used as a positive control. After electrophoresis gels were then placed into 9 cm square petri dishes with assay buffer (100 mM Tris-HCl pH 7.0, 500 mM NaCl). Activity was

visualised by adjusting the assay buffer to contain 2.75 mM SDS, 2 mM dopamine and 2 mM MBTH and incubating the gels on a gel rocker in the dark at room temperature. Gels were incubated in the activity stain for up to 16 hours to allow activity bands to develop. If a DPO enzyme inhibitor were to be tested during the assay, gels were pre-incubated in assay buffer containing inhibitor for 10 minutes prior to the addition of SDS, dopamine and MBTH. The final concentrations of inhibitors used were 2 mM for both phenylthiourea (PTU) and kojic acid. To image gels they were placed between two sheets of plastic and scanned using an Epson Perfection 1670 flat-bed scanner.

## 2.6.2 Polyacrylamide Gel Band Protein Identification

To identify which proteins were present in a slice of polyacrylamide gel the proteins within that gel were digested with trypsin and the resulting peptides analysed by matrix-assisted laser desorption/ionisation tandem time of flight mass spectrometry (MALDI-TOF/TOF). This procedure is described below.

### *2.6.2.1 Gel Band Excision*

Protein bands to be identified from either native or SDS-PAGE were excised from the gel using a razor blade taking care to collect the minimum amount of gel necessary. The gel band was then cut into 3 – 4 pieces and placed in a 0.5 ml LoBind tube (Eppendorf). The razor blade was cleaned between each band excised using ethanol. Bands were stored at -20 ℃ until further processin g.

### *2.6.2.2 Protein Trypsinisation*

Coomassie stain and other contaminants were removed from the gel pieces by rehydrating them in 100 µl of ABC/MeCN solution (25 mM ammonium bicarbonate, 50 % v/v acetonitrile) for 20 minutes. The solution was discarded and this de-staining step repeated once.

The gel pieces were resuspended in 100 µl of acetonitrile and allowed to stand for 5 minutes before removing and discarding the acetonitrile. Next the gel pieces were dried in a SC210A SpeedVac® Plus (Thermo Savant) for 20 minutes on the machine's medium setting.

The cysteine residues of the proteins contained in the gel pieces were reduced by rehydrating the gel with 100 µl of DTE solution (100 mM ammonium bicarbonate, 10 mM dithioerythrietol) and incubating them at 56 ℃ for 1 hour. The DTE solution was discarded. The reduced cysteines were then alkylated by adding 100 µl of IAA solution (100 mM ammonium bicarbonate, 50 mM iodoacetamide) to the gel pieces and incubating them at room temperature in the dark for 30 minutes.

The IAA solution was then discarded and the gel pieces washed with 100 µl of 100 mM ammonium bicarbonate solution for 15 minutes, 100 µl of ABC/MeCN solution for 15 minutes and 100 µl of acetonitrile for 5 minutes, discarding each wash before applying the next. The gel pieces were dried in a SpeedVac for 20 minutes on the machine's medium setting.

A working solution of trypsin was then prepared. A vial of Sequencing Grade Modified Trypsin (Promega) containing 20 µg of lyophilised protein was resuspended in 200 µl of Trypsin Resuspension Buffer (Promega). A 20 µl aliquot of this solution was further diluted with 80 µl of 100 mM ammonium bicarbonate solution.

A 10 µl aliquot of trypsin working solution was added to the gel pieces and the gel allowed to re-swell for 5 – 10 minutes. If the gel pieces were not covered in liquid 25 mM ammonium bicarbonate was added in 5 µl aliquots up to a maximum total volume of 40 µl until the gel pieces were covered. The gel pieces were incubated at 37 ℃ for 16 hours in the dark for trypsinisation to occur. After this incubation the liquid that the gel pieces were immersed in contained trypsinised peptides ready for spotting onto a MALDI plate.

### *2.6.2.3 Preparing Samples for Mass Spectrometry*

The MALDI matrix was prepared by mixing 10 mg of α-cyano-4-hydroxycinnamic acid with 1 ml of 0.1% v/v formic acid, 50 % v/v acetonitrile. This suspension was vortexed at maximum speed for 1 minute to create a saturated solution. The suspension was centrifuged at 13,000 x g to sediment any non-dissolved solids before carefully aspirating the top 500 µl into a new tube. This supernatant was then diluted with a further 500 µl of 0.1% v/v formic acid, 50 % v/v acetonitrile.

Samples were then spotted onto a MALDI plate by pipetting 1 µl of trypsinised peptides into the centre of an etched circle on the MALDI plate. Before the sample dried 1 µl of

MALDI matrix was pipetted into the drop of sample and the two mixed together taking care not to move the solution outside of the etched circle or to touch the plate with the pipette tip. The spots were then allowed to air dry under cover.

The same spotting technique was carried out for calibrant solution. Calibrant solution was prepared by the Proteomics Laboratory, Technology Facility, University of York and contains a mixture of 6 peptides described in Table 2.5. This calibrant would be referred to by the mass spectrometer between samples to maintain inter-run calibration.

Spots were arranged on the MALDI plate as a 3 x 3 grid with a spot of calibrant in the centre.

**Table 2.5. Details of the six peptides used to produce calibrant solution for MALDI-TOF/TOF analysis.**
All peptides were purchased from Sigma Aldrich.

| Peptide | Mass/Charge | Product Number |
|---|---|---|
| des-Arg$^1$ Bradykinin | 904.468 | B1901 |
| Angiotensin I | 1296.685 | A9650 |
| Glu$^1$-Fibrinopeptide B | 1570.677 | F3261 |
| Adrenocorticotropic Hormone fragment 1-17 | 2093.086 | A0637 |
| Adrenocorticotropic Hormone fragment 18-39 | 2465.198 | A2407 |
| Adrenocorticotropic Hormone fragment 7-38 | 3657.929 | A1527 |

### *2.6.2.4 Peptide Mass Spectrometry*

Positive-ion MALDI mass spectra were obtained using a Bruker Ultraflex III in reflectron mode, equipped with a Nd:YAG smart beam laser. Mass spectra (MS) were acquired over a mass range of m/z 800-4000. Final mass spectra were externally calibrated against an adjacent spot containing a calibrant of 6 peptides (see Table 2.5). Monoisotopic masses were obtained using a sophisticated numerical annotation procedure (SNAP) averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) and a signal:noise ratio (S:N) threshold of 2.

For each spot the ten strongest peaks of interest, with a S:N greater than 30, were selected for tandem mass spectrometry (MS/MS) fragmentation. Fragmentation was performed in LIFT™ mode without the introduction of a collision gas. The default

calibration was used for MS/MS spectra, which were baseline-subtracted and smoothed (Savitsky-Golay (141), width 0.15 m/z, cycles 4); monoisotopic peak detection used a SNAP averaging algorithm (C 4.9384, N 1.3577, O 1.4773, S 0.0417, H 7.7583) with a minimum S:N of 6. Bruker flexAnalysis software (version 3.3) was used to perform the spectral processing and peak list generation for both the MS and MS/MS spectra.

Tandem mass spectral data were submitted to database searching using a locally-running copy of the Mascot program (Matrix Science Ltd., version 2.2.0), through the ProteinScape interface (Bruker, version 2.1). Search criteria included: Enzyme, Trypsin; Fixed modifications, Carbamidomethyl (C); Variable modifications, Oxidation (M); Peptide tolerance, 250 ppm; MS/MS tolerance, 0.5 Da; Instrument, MALDI-TOF-TOF; Database, *Limnoria* transcriptomic database and a nightly updated version of the NCBI non-redundant database.

### 2.6.3 Purification and Analysis of Anti – Hemocyanin Antibody

An anti-hemocyanin antibody was raised in a rabbit by Covalab using denatured recombinant HC1 protein produced in mammalian cell culture by Dr. Katrin Beßer (University of York). This antibody was used throughout the work discussed in this thesis. The antibody was received as a crude serum from the terminal bleed of the animal. This crude serum produced a high non-specific background when used for Western blots so was affinity purified by myself using denatured recombinant HC3 produced in *E. coli*.

Denatured HC1 protein from mammalian cells was used to produce the antibody as at the time it was the only system available which could produce enough protein for this purpose. When the time came to purify the antibody it was possible to produce enough denatured HC3 protein to do so using *E. coli*. Using these two different proteins and expression systems was actually beneficial for the purification for the following reasons:

- HC1 and HC3 have a sequence identity of ~84 % calculated using blastp (http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastHome). This means that the proteins are highly likely to be similar enough to both be detected by the same polyclonal antibody. Indeed all four other *Limnoria* HC

proteins possess a sequence identity of 62 – 87 % with HC1 when calculated in a similar way.

- This means that by purifying the crude antibody using a different, but similar protein to that which it was raised against, should favour the enrichment of the purified antibody with those antibodies able to detect the non-gene-specific regions of the *Limnoria* HC proteins resulting in an antibody fraction able to detect the other *Limnoria* HCs.
- Purifying the antibody serum against protein raised in a different expression system should reduce the amount of non-GOI background detection. This should result as the contaminating proteins present from the expression host should be different in the protein the antibody was raised against and that used to purify it from the crude serum.

### *2.6.3.1 Preparation of Antibody Affinity Column*

Briefly, a $His_6$-tagged form of HC3 was recombinantly expressed in Rosetta-gami 2 *E. coli* cells. The cells were grown in 2x YT medium at 37 ℃ until they had reached an $OD_{600}$ of 1.0 after which they were induced to express HC3 by adjustment of the medium to 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and moved to 15 °C for 3 hours before the cells were harvested by centrifugation.

The cells were resuspended in lysis buffer (0.1 M $KH_2PO_4/K_2HPO_4$ pH 8.0, 0.15 M NaCl, 1x HALT EDTA Free Protease Inhibitor Cocktail – Thermo Fisher Scientific), lysed by cell disruption and separated into soluble and insoluble fractions by centrifugation. The insoluble fraction containing inclusion bodies of insoluble HC3 was washed three times in lysis buffer and then dissolved in denaturing buffer (0.1 M $NaH_2PO_4$, 10 mM Tris-NaOH pH 8.0, 8 M urea). Any of the pellet which could not be dissolved was sedimented by centrifugation. The urea dissolved protein was batch bound to Ni-NTA resin for 16 hours at 4 ℃ before w ashing the resin (denaturing buffer, pH 6.3). Protein bound to the resin was eluted using a decrease in pH (denaturing buffer, pH 5.9 and 4.5 respectively – shown in Figure 2.2). Elution fractions were pooled and dialysed twice against 4 L of coupling buffer (0.1 M $NaHCO_3$, 0.5 M NaCl, pH 8.3) at 4 ℃. The dialysed protein was concentra ted to 1 ml and was found to contain 0.67 mg/ml of protein. SDS-PAGE gels were used to monitor the process throughout and showed the final concentrated, dialysed protein sample to be partially

pure with a single dominant species at approximately the correct molecular mass to be HC3 (Figure 2.2).

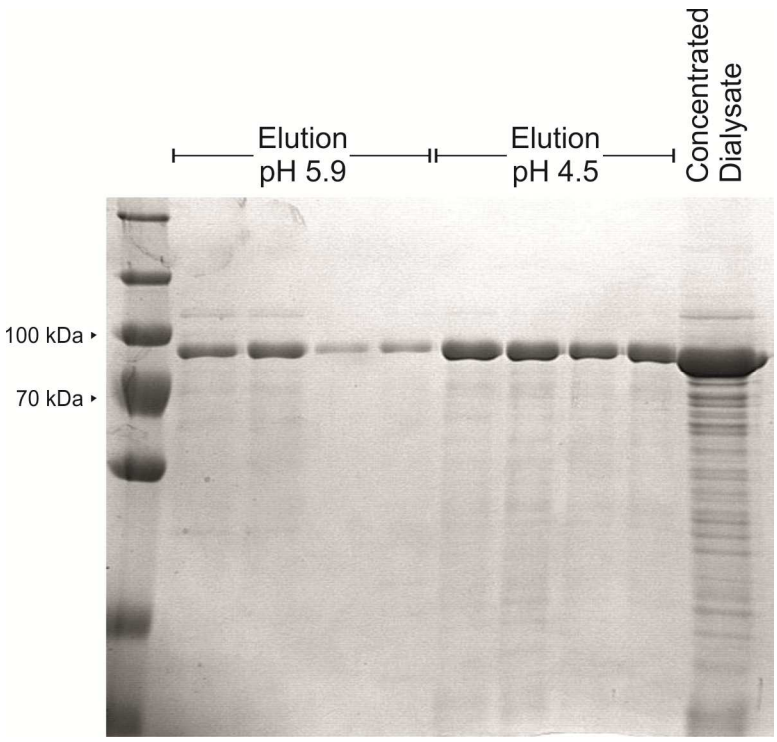



**Figure 2.2. Coomassie stained SDS-PAGE gel of denatured HC3 purification fractions.**

To make the antibody affinity column CNBr-activated Sepharose™ 4 Fast Flow resin (GE Healthcare Life Sciences) was used. This resin will spontaneously react with primary amine groups and so will react with protein molecules anchoring them to the resin.

The concentrated, dialysed protein sample was incubated with CNBr resin for 16 hours at 4 ℃. The solution was then removed from the res in and any un-filled binding sites in the resin blocked by incubation with 0.1 M Tris-HCl pH 8.0. The resin was then washed with high and low pH solutions (0.1 M Tris-HCl, 0.5 M NaCl at pH 8.0 and 0.1M $CH_3CO_2H/CH_3CO_2Na$, 0.5 M NaCl pH 4.0 respectively) before equilibrating it in phosphate buffered saline (PBS).

### 2.6.3.2 Antibody Purification

An aliquot of crude anti-HC antibody serum was buffered to ~0.1 M Tris-HCl pH 8.0 and incubated with the prepared HC3-CNBr resin for 2 hours at room temperature. The serum was then removed from the resin before washing it with PBS. Finally the purified antibodies were eluted using 0.1 M glycine-HCl pH 2.0 directly into tubes containing 1 M Tris-HCl pH 8.0.

### 2.6.3.3 Analysis of Antibody

To assess the effectiveness of the antibody purification on reducing non-specific background detection during Western blotting a series of dot blots of recombinant HC3 protein were used to test the pre-immune serum, crude serum and antibody purification elution fractions. A comparison was also made to a commercial anti-$His_6$ tag antibody.

Purified denatured recombinant HC3 was produced in a similar way to that discussed in Section 2.6.3.1 yielding protein of a similar degree of purity. This protein was spotted onto nitrocellulose membrane to produce dot blots of the following amounts of protein: 100, 50, 10, 1, 0.5 and 0.1 ng. These dot blots were tested by Western analysis using the standard procedure discussed in Section 2.5.6. This was done using 1 in 50 and 1 in 1000 fold dilutions of anti-HC antibody pre-immune serum, crude serum and antibody purification elutions 1 – 6 as a primary antibody followed by a 1 in 1250 fold dilution of anti-rabbit antibody (product number #32460, Thermo Fisher Scientific) as a secondary antibody. A comparison was also produced using a 1 in 2000 dilution of a monoclonal anti-polyHistidine-Peroxidase antibody produced in mouse (product number #A7058, Sigma Aldrich). All commercial antibodies were used at dilutions recommended by their respective suppliers. The resulting dot blots are shown in Figure 2.3.

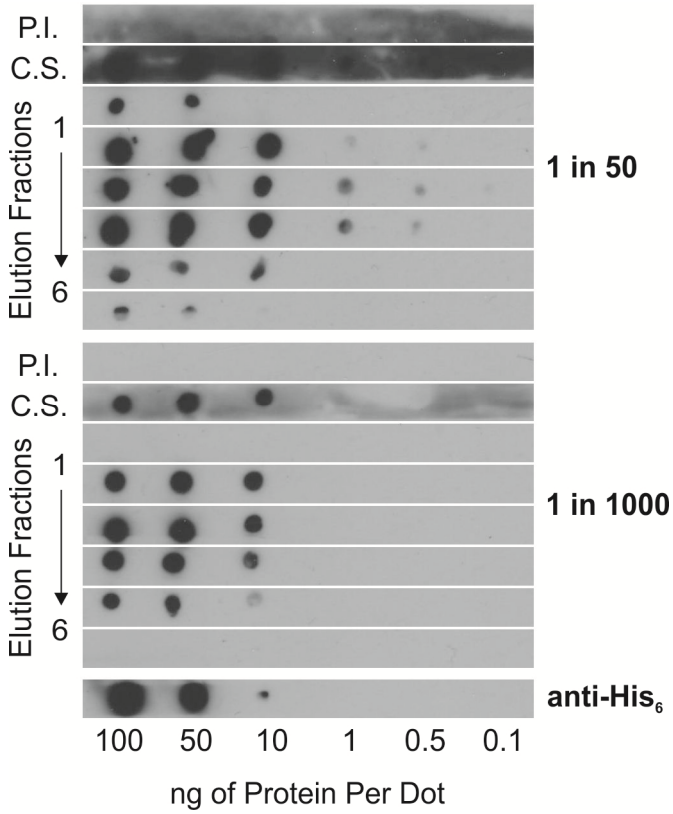



**Figure 2.3. Evaluation of anti-HC antibodies purification from crude serum.** Dot blots were produced by spotting various amounts of purified denatured HC3 protein from *E. coli* onto nitrocellulose membranes. The membranes were then Western blotted using as a primary antibody either a 1 in 50 or 1 in 1000 dilution of pre-immune serum (P.I.), post-immune final bleed crude serum (C.S.) or purified anti-HC antibody fractions. A dot blot was also probed with a commercial anti-$His_6$ antibody as a comparison.

Using a 1 in 50 dilution, which is a quite small dilution for an effective antibody, as little as 0.1 ng of HC protein could be detected using the dot blot method and an extremely strong non-specific background signal was seen with both pre-immune and crude serums.

Using a dilution of anti-HC antibodies of 1 in 1000, which is a reasonable dilution for an antibody, the pre-immune serum produces no signal. The crude serum however produces a strong signal where there were dots of HC protein, as well as a strong non-specific background signal. Antibody elution fractions 2 – 5 also show strong signals where there were dots of HC protein, but no non-specific background. The crude serum and elution fractions 2 – 4 equally well detected as little as 10 ng of HC protein, with fraction 5 having a slightly weaker signal with 10 ng of HC protein. This detection range is comparable with the commercial anti-$His_6$ antibody used.

As elution fractions 2 – 4 performed equally well these fractions were pooled and adjusted to contain 0.05 % w/v sodium azide. This batch of antibody solution was used throughout this work wherever “anti-HC antibody” is referred to.

To determine the dilution of antibody to use routinely to detect HC protein in SDS-PAGE gels a dilution series of antibody was probed against recombinant denatured HC3 protein.

An SDS-PAGE gel with a single lane the width of the gel was ran with 1.5 µg of the recombinant denatured HC3 protein used for the previously mentioned dot blots. This gel was transferred to nitrocellulose membrane as described in Section 2.5.6.1. The membrane was Ponceau S stained, partially de-stained, dried and then cut into 3 mm wide vertical strips. Each strip showed a single band at the correct apparent molecular mass to be HC3 and which would correspond to ~66 ng of protein. Several of these strips were then Western blotted as described in Section 2.5.6 using the following dilutions of anti-HC antibody as a primary antibody: 1 in 500, 1 in 1000, 1 in 1500, 1 in 2500, 1 in 5000 and 1 in 10000. The secondary antibody was a 1 in 1250 dilution of the previously mentioned commercial HRP-anti-rabbit antibody. The resulting blots are shown in Figure 2.4.

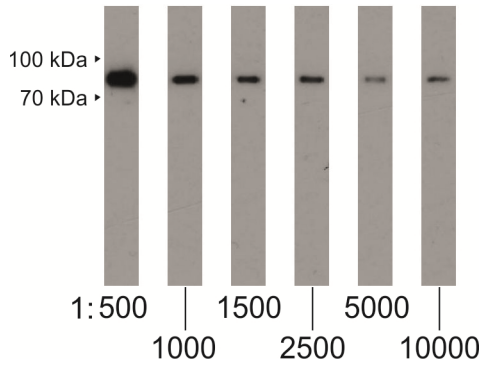



**Figure 2.4. Determining the dilution of anti-HC antibody to routinely use for Western analysis.**
A dilution series of pooled purified anti-HC antibody (fractions 2-4, see above) was probed against strips of nitrocellulose, each containing ~66 ng of purified denatured HC3 protein from *E. coli* separated by SDS-PAGE.

Figure 2.4 shows that even as large as a 1 in 10000 fold dilution of the anti-HC antibody was able to detect a band of ~66 ng of HC3 protein. All dilutions only detected a single band and not any of the remaining *E. coli* contaminants known to be present.

### *2.6.3.4 Discussion of Anti-HC Antibody Analysis*

From this analysis it was seen that the crude anti-HC antibody serum had been sufficiently purified to be used as a specific detection agent for HC protein in Western blotting procedures. As the antibody was raised against HC1 but has been shown to be capable of detecting HC3 it is apparent that the specificity of the antibody is broad enough to cross react with other *Limnoria* HC proteins. The evaluation of the antibody against strips of SDS-PAGE separated purified HC3 protein suggest that a 1 in 1000 fold dilution of the antibody would be a suitable dilution to use in future work. This dilution is able to produce a strong but not over-exposed signal with ~66 ng of HC protein, which is enough protein to produce an easily visible band on an SDS-PAGE gel with Coomassie staining, so it should also be sufficient to produce a signal with much less protein also.

# Chapter 3: Prospecting for Lignocellulose Degrading Enzymes

## 3.1 Introduction

The hepatopancreas (HP) of *Limnoria* is lined with secretory cells which are believed to secrete enzymes and proteins involved in digestion into the lumen of the HP (130, 134, 135). The gut of *Limnoria* is lined with a cuticle and is not thought to actively contribute proteins to the digestive process (132). The HP of *Limnoria* is joined to the gut at the posterior end of the stomach by two filter channels which are thought to allow the passage of fluids and proteins between the two organs, but to segregate the ingested wood particles to the gut only (132, 133).

This anatomical arrangement leads to the hypothesis that for a protein to be involved in digestion and directly act upon ingested wood this protein would need to be produced in cells in the HP, secreted into the HP lumen and then moved into the gut to be in proximity to the wood particles. It also implies that proteins involved in digestion will likely be found abundantly in the HP and gut of the animal as they move from one organ to the other in sufficient quantities to facilitate the degradation of the lignocellulose food source.

The transcriptome of the HP has previously been sequenced and analysed which should provide part of the data towards testing this hypothesis (68); namely which genes are transcribed in the HP of *Limnoria*. The other data that will be needed are details of which proteins are present in each of the two organs, their proteomes, as well as some idea of each protein's relative abundance. Proteins present solely in one organ, or in both organs but much more abundantly in one tissue may be of interest as this partitioning may imply a role in the digestion process.

In order to acquire the proteomic data necessary to address this hypothesis several techniques were considered. Proteomic techniques have evolved over time to be more sophisticated; however each technique has its own advantages and disadvantages and many are still used routinely. The following described techniques have all been successfully used to determine the differences in gene expression patterns and protein profiles between different tissues (142-144).

The most established of the techniques is two-dimensional polyacrylamide gel electrophoresis (2D-PAGE). This technique separates the proteins in a sample in two physical, perpendicular dimensions in a polyacrylamide gel allowing the proteins in the sample to be visualised as an array of dots. By doing this for two or more different tissues or treatments one can visualise the presence, absence or change in size of a dot at a particular point in each of the gels. Spots can be excised and the proteins contained within the gel identified. This technique was successfully employed for example to identify 17 proteins with modified expression from a total of 675 visible proteins from tumorous and non-tumorous epithelium tissue (145). However the technique is subject to human bias when selecting spots for identification if an automated machine is not available to do this function. The technique also requires a relatively large amount of protein to work efficiently.

A more recently developed technique utilises liquid chromatography systems to separate trypsin digested proteins before using a rapidly scanning mass-spectrometer to identify the peptides in that mixture. This technique is seen as being superior to 2D-PAGE as it can be quantitative or semi-quantitative depending on whether a chemical label is added to the peptides or not, and in doing so makes the results more amenable to statistical analysis. It also requires less protein than a 2D gel. A small disadvantage of this technique is that compared to 2D-PAGE there is no visual data produced. A great deal of data processing is required to interpret the results of these experiments. Such techniques have been used successfully to identify biomarkers of tumour metastasis and non-small-cell lung cancer (146, 147).

Hybrid approaches have also been developed utilising one dimensional gel separation followed by liquid chromatography – mass spectrometry analysis of peptides made by trypsinising the proteins within bands of the first dimension gel, so called GeLC-MS (148). This technique suffers most of the disadvantages of both the 2D-PAGE and LC-MS approaches, but can result in simpler to interpret data. If only a certain mass range of proteins is of interest one dimensional gel separation can allow all other mass ranges to be discarded before LC-MS analysis. This would result in a smaller dataset enriched with information about the proteins one is interested in.

### 3.1.1 Aims of this Chapter

The work undertaken in this chapter aimed to identify interesting proteins from the gut and HP of *Limnoria* which may be involved in the animal's digestive processes. To achieve this 2D-PAGE and label-free semi-quantitative proteomic approaches were taken to analyse the proteins present in gut and HP tissue from *Limnoria*. The data generated, taken in context with the existing HP transcriptome data, were used to identify proteins which may play a role in lignocellulose digestion for further study.

## 3.2 Methods

### 3.2.1 Two Dimensional - Polyacrylamide Gel Electrophoresis (2D-PAGE)

Conventional SDS-PAGE separates protein molecules based on their electrophoretic mobility when possessing a uniform charge, which is a function of the protein molecules size and molecular weight. However a typical SDS-PAGE mini-gel (8 – 10 cm in length) is not capable of resolving the thousands of individual proteins which make up a proteome into individually separated bands. When trying to analyse the differences in proteomes of tissues, a higher level of resolving power is required to detect the differences between samples. To do this a second form of separation is needed.

The 2D-PAGE technique increases the resolving power of SDS-PAGE by hyphenating it to a second separation technique. This second dimension of separation must be orthogonal to the first; meaning it must separate proteins based on an entirely un-related property to that used in the first dimension, in order to achieve the maximum gain in resolving power. Isoelectric focussing (IEF) is a common separation technique used in 2D-PAGE as a first dimension, followed by SDS-PAGE. It is based on the separation of proteins by their isoelectric point (pI), the point at which a protein carries no net charge, and is a function of the pH of the environment the protein is in. In order to focus proteins based on their pI the protein samples are applied to an immobilised pH gradient gel strip (IPG strip) and an electric current passed across the strip for a defined number of volt hours (V-hr). When applied to the strip proteins will take on a charge appropriate to their pI and the pH of the part of the gel which they are in. When the electric field is applied the proteins migrate towards the electrode with the opposite charge to their net charge until they reach a point in the pH gradient which causes their charge to reverse polarity at which point they migrate in the reverse direction. This continues until the protein has focussed at a point in which it has no net charge and is not attracted to either electrode.

The IPG strip is then attached to a conventional SDS-PAGE gel such that the direction of separation from IEF is perpendicular to the direction of separation by SDS-PAGE. After running the SDS-PAGE gel the many proteins present at each pH on the IPG strip

have been separated based on their molecular size and form well defined spots in the SDS-PAGE gel.

Although not capable of separating every protein present in the complex mixtures of tissue lysates, 2D-PAGE is able to provide a much greater degree of resolution allowing the presence or absence of different proteins in different samples to be observed. The gel spots containing protein can then be excised and the proteins contained within the spot identified using standard mass spectrometry identification techniques.

#### *3.2.1.1 Method Development*

Given the considerable effort involved in collecting dissected *Limnoria* material several practice 2D-PAGE gels were prepared using an *E. coli* protein mixture as a sample.

Various sample homogenisation methods were evaluated such as sonication and boiling the sample in an SDS containing buffer. Problems encountered with the IEF step were trouble-shooted and were found to be due to the urea in pre-prepared rehydration buffer having decayed. This resulted in a high current across the IPG strip during focussing which prevented efficient focussing. Preparing the rehydration buffer fresh solved this issue.

Once a working protocol was established it was used to try and identify differences in protein expression between *Limnoria* HP and gut tissue. However further difficulties were encountered with incomplete sample homogenisation and again with the IEF step resulting in a failure of the samples to focus.

Further method development was undertaken with whole *Limnoria* as a starting material. Exposing the sample to liquid nitrogen and then grinding it with a Teflon pestle was found to be a more effective homogenisation technique for this sort of material resulting in a more complete homogenisation. The failure to focus of *Limnoria* samples was thought to be due to the presence of charged impurities in the homogenate coming either from the material itself, or from residual sea water collected with the material. To overcome this, a protein clean-up kit designed for use with 2D-PAGE was used. This kit allowed collected protein to be precipitated and washed before re-dissolving it in rehydration buffer ready for IEF. These modifications to the protocol resulted in a more robust protocol. Once a successful practice gel had been

prepared using this protocol with whole *Limnoria* as a starting material it was used to carry out the experiment described below.

### *3.2.1.2 E. coli Protein Standard Preparation*

During the sample preparation of *Limnoria* tissues for 2D-PAGE, the tissues are resuspended in a solution containing 0.2 % w/v SDS. The Coomassie Plus™ Protein Assay Reagent used to carry out Bradford assays is only compatible with final SDS concentrations up to 0.016 % w/v; using *Limnoria* tissue samples resuspended in this way results in SDS concentrations of ~0.04 % w/v. It would not be possible to quantify the amount of protein in the *Limnoria* tissue samples using such a Bradford assay. To estimate the protein concentration of *Limnoria* tissue homogenates a different method was used. This involved running an aliquot of *Limnoria* protein on an SDS-PAGE gel alongside known quantities of *E. coli* lysates and estimating by eye which *E. coli* standard the *Limnoria* sample most resembled in terms of amount of overall stain.

A colony of *E. coli* carrying the empty pGEX-4T-3 vector was picked and used to inoculate 10 ml of LB medium. The culture was grown for 16 hours at 37 ℃ whilst shaking at 200 rpm. The cells were harvested and resuspended in 1 ml of lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % v/v TWEEN 20). The cells were sonicated and the lysate cleared of insoluble debris as described in Section 2.5.4.1. The concentration of protein present in the soluble lysate was determined as described in Section 2.5.1 and found to be 3.21 mg/ml. The lysate was diluted to 2.00 mg/ml with lysis buffer, aliquoted and stored at -20 ℃ until needed.

### *3.2.1.3 Sample Processing and Protein Quantification*

*Limnoria* material was kindly collected by Dr. Graham Malyon (University of Portsmouth). The guts and HPs of 70 *Limnoria* living on Balau hardwood were dissected out of the animals and plunged immediately into a micro-centrifuge tube containing and floating on liquid nitrogen. Once all tissue had been collected the nitrogen inside the tube was allowed to evaporate before sealing the tubes and storing them at -80 ℃. The material was transported to York on dry ice.

Before processing the samples, a tube rack was prepared by packing it into a shallow polystyrene container with wet ice and repeatedly pouring liquid nitrogen into the centre

wells of the rack. Once the nitrogen ceased to boil away very rapidly the rack was ready to use.

Processing one tube at a time, a tube containing *Limnoria* material was floated on liquid nitrogen in the tube rack. The tube was carefully filled with more liquid nitrogen before grinding the contents with a pre-cooled steel pestle until it had a powdery consistency. To the tube was added 110 µl of solubilisation buffer (50 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.2 % w/v SDS, 5 mM DTE, 2 mM EDTA, 200 µM 4-(2-Aminoethyl)benzenesulphonyl fluoride (AEBSF)) which instantly froze. The tube contents were ground again to a powdery consistency. The tube was allowed to gradually warm up to room temperature. Once the liquid in the tube had melted the contents were ground to a smooth consistency. The tube was kept on wet ice and the other *Limnoria* tissue sample prepared in the same way.

To determine the protein concentration of the *Limnoria* HP and gut homogenates 10 µl of each were ran on an SDS-PAGE gel alongside 20, 10, 5, 2.5 and 1 µg of *E. coli* protein prepared as described above.

The remaining tissue homogenates were prepared for IEF using a ReadyPrep™ 2-D Cleanup Kit (Bio-Rad) following the kit's instructions. This resulted in nearly dry protein pellets. These pellets were resuspended in 125 µl of freshly prepared rehydration buffer (6 M urea, 2 M thiourea, 4 % w/v CHAPS, 65 mM DTE, 0.8 % w/v Bio-Lyte 3/10 carrier ampholytes/resolytes (product number #163-1112, Bio-Rad)). To aid dissolution the samples were vortexed for 30 seconds and floated in a sonic bath containing ice water at full power for 2 minutes.

### *3.2.1.4 Isoelectric Focussing (IEF) – First Dimension*

Re-dissolved protein samples were then used to rehydrate IPG strips. To wells of an IEF focussing tray 125 µl of each sample were applied. A ReadyStrip IPG, pH 3 – 10, 7 cm, linear gradient strip (Bio-Rad) was then laid on top of each sample and overlaid with 1 ml of mineral oil. The strips were allowed to passively rehydrate for 15 hours.

Once rehydration was complete filter paper wicks soaked in water were inserted between the IPG strip and the electrodes of the tray before placing the tray in a PROTEAN IEF Cell (Bio-Rad) and applying the following voltage scheme: 250 V for 30

minutes (rapid), 1000 V for 30 minutes (rapid), 8000 V for 60 minutes (linear), 8000 V for 10000 V-hr (rapid), 500 V for 24 hours (holding step).

### *3.2.1.5 SDS-PAGE – Second Dimension*

Once focussed the IPG strips were placed gel side up in an equilibration tray and each overlaid with 2 ml of equilibration buffer (10 % w/v Tris-HCl pH 8.8, 36 % w/v urea, 30 % v/v glycerol, 2 % w/v SDS) plus 67 mM DTE for 12 minutes at room temperature to reduce the proteins in the strip. The proteins were then alkylated by moving the strips to fresh wells and overlaid with equilibration buffer plus 42 mM iodoacetamide for 12 minutes at room temperature. The strips were finally rinsed in running buffer (NuPAGE® MES SDS Running Buffer – Invitrogen Life Technologies).

The strips were inserted into the wells of precast NuPAGE® 4 – 12 % Bis-Tris ZOOM™ 1.0 mm x IPG well polyacrylamide gels (Invitrogen Life Technologies). The strips were overlaid with molten agarose (0.5 % w/v agarose, 0.025% bromophenol blue in running buffer) to join the strip and gel. The gels were run at 200 V until the dye front ran off the gels.

### *3.2.1.6 Fixing and Staining*

The gels were removed from their cassettes and rinsed in water before covering them with fix solution (40 % v/v methanol, 10 % v/v acetic acid) and placing them on a gel rocker at room temperature for 30 minutes. After fixing the gels they were again rinsed in water before covering them in InstantBlue™ stain and placing them on a gel rocker for an hour.

### *3.2.1.7 Gel Spot Protein Identification*

The gel spot containing proteins to be identified was excised using a spot picker pen into a 0.5 ml LoBind tube (Eppendorf). The gel was placed into 100 µl of 25 mM ammonium bicarbonate in 50 % v/v acetonitrile for 20 minutes to de-stain the gel. This was repeated once. The gel was washed with 100 µl of acetonitrile before drying it in a SC210A SpeedVac® Plus (Thermo Savant) on the machines medium setting for 20 minutes.

The proteins within the gel were trypsinised as described in Section 2.6.2.2. The peptides were then subjected to mass spectrometry analysis as described in Section 2.6.2.3 to identify the proteins that the peptides came from.

## 3.2.2 Label-Free Semi-Quantitative Proteomic Analysis

### *3.2.2.1 Collection of Material*

*Limnoria* material was kindly collected by Dr. Graham Malyon (University of Portsmouth). The guts and HPs of 100 *Limnoria* living on Balau hardwood were dissected out of the animals and plunged immediately into a micro-centrifuge tube containing and floating on liquid nitrogen. Once all tissue had been collected the nitrogen inside the tube was allowed to evaporate before sealing the tubes and storing them at -80 ℃. The material was transported to Yor k on dry ice.

### *3.2.2.2 Filter-Aided Sample Preparation (FASP)*

The *Limnoria* tissue samples were lysed in a buffer containing SDS to aid the dissolution of the proteins within the tissue; however SDS is not compatible with down-stream mass spectrometry analysis. To circumvent this issue the FASP technique was followed essentially as described by Wiśniewski et al. (149). This technique works by capturing the dissolved proteins in a centrifugal filter device and washing away the SDS in the sample with 8 M urea, before exchanging this urea for ammonium bicarbonate. The proteins in the sample are alkylated and trypsinised whilst in the filter unit and once trypsinisation is complete the peptides ready for analysis can simply be recovered from the unit by centrifugation. The full protocol is detailed below.

The samples of *Limnoria* tissue described above were resuspended in 500 µl of SDS lysis buffer (100 mM Tris-HCl pH 7.6, 4 % w/v SDS, 100 mM DL-dithiothreitol (DTT)) and ground using a Teflon pestle. Once the samples were homogeneous they were sonicated on ice in a sonic bath for 5 minutes before being placed in a heating block at 100 ℃ for 5 minutes. Samples were then centrifuged at 10,000 x g for 5 minutes and the resulting aqueous portions of the supernatant removed into tubes, pre-washed with ethanol, and diluted 10 fold with UA solution (100 mM Tris-HCl pH 8.5, 8 M urea, freshly prepared). Aliquots of each diluted sample were applied to Amicon Ultra-0.5, Ultracel-30 Membrane, 30 kDa MWCO filter devices (product number #UFC503096,

Millipore) and centrifuged at 14,000 x g for 10 minutes at 20 ℃, discarding the flow-through. This was repeated until all of the samples had been loaded into the filter devices. The samples were washed twice with 250 µl of UA solution, spinning away the wash at 14,000 x g for 10 minutes at 20 ℃ each time.

The samples were alkylated by adding 100 µl of IAA solution (40 mM iodoacetamide, 100 mM Tris-HCl pH 8.5, 8 M urea, freshly prepared) to each filter device and mixing the samples at 500 rpm for 1 minute using a micro-centrifuge tube shaker at room temperature before allowing the samples to stand without mixing for 15 minutes. The IAA solution was removed by centrifugation at 14,000 x g for 10 minutes at 20 ℃.

Urea was washed away from the samples with five batches of 250 µl of ABC solution (50 mM ammonium bicarbonate), spinning away the wash at 14,000 x g for 10 minutes at 20 ℃ each time. A vial of Sequencing Grade Modified Trypsin (Promega) containing 20 µg of lyophilised protein was resuspended in 200 µl of Trypsin Resuspension Buffer (Promega). A 20 µl aliquot of this solution was further diluted with 80 µl of ABC solution and all of this applied to the filter unit. The filter units were sealed with Parafilm, placed into new collection tubes which had been pre-washed with ethanol and mixed at 600 rpm for 1 minute using a micro-centrifuge tube shaker before incubating the samples for 16 hours in a heating block at 37 ℃.

The trypsinised peptides were recovered by centrifugation at 14,000 x g for 10 minutes at 20 ℃ after which 100 µl of water was added to the filter units and they were again centrifuged.

### *3.2.2.3 Mass Spectrometry Analysis*

Peptide samples were kindly analysed by Dr. David Ashford in the Proteomics Lab, Technology Facility, University of York. A suitable amount of sample was loaded onto a nanoAcquity ultra-performance liquid chromatography (UPLC) system (Waters) equipped with a nanoAcquity Symmetry $C_{18}$, 5 µm trap (180 µm x 20 mm – Waters) and a nanoAcquity BEH130 1.7 µm $C_{18}$ capillary column (75 µm x 250 mm – Waters). The trap wash solvent was 0.1 % v/v aqueous formic acid and the trapping flow rate was 10 µL/min. The trap was washed for five minutes before switching flow to the capillary column. The separation used a gradient elution of two solvents (solvent A: 0.1 % v/v formic acid; solvent B: acetonitrile containing 0.1 % v/v formic acid). The flow rate for the capillary column was 300 nL/min Column temperature was 60℃ and the

gradient profile was as follows: initial conditions 5 % solvent B, followed by a linear gradient to 30 % solvent B over 125 minutes, then a linear gradient to 50 % solvent B over five minutes, followed by a wash with 95 % solvent B for 10 min. The column was returned to initial conditions and re-equilibrated for 30 minutes before subsequent injections.

The nanoLC system was interfaced to a maXis LC-MS/MS System (Bruker Daltonics) with a nano-electrospray source fitted with a steel emitter needle (180 µm O.D. x 30 µm I.D. – Proxeon). Positive ESI – MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.3 SR3 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were: ion spray voltage, 1,500 V; dry gas, 6 L/min; dry gas temperature, 160 ℃; ion acquisition range, *m/z* 50-2200;. AutoMSMS settings were: MS, 0.5 seconds (acquisition of survey spectrum); MS/MS, (CID with $N_2$ as collision gas); ion acquisition range, *m/z* 300-1500; 0.1 seconds acquisition for precursor intensities above 100000 counts; acquisition time increase linearly to 1.0 for precursor intensities down to 1000 counts. The collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table: five precursor ions, absolute threshold 1,000 counts, preferred charge states 2 – 4, singly charged ions excluded. One MS/MS spectrum was acquired for each precursor and former target ions were excluded for 30 seconds. MS & MS/MS results were submitted to a locally running copy of the Mascot program (Matrix Science Ltd., version 2.2.0) through the ProteinScape interface (Bruker, version 2.1) for searching against the *Limnoria* HP transcriptome generated by King et al. (68) supplemented with the full-length sequences of genes subsequently cloned from *Limnoria*.

### *3.2.2.4 Data Processing*

Data for each sample was returned as a list of peptides which had been detected, grouped by the contiguous sequence (contig) from the *Limnoria* HP transcriptome that those peptides most likely represented. For each contig the Mascot program had calculated an exponentially modified protein abundance index (emPAI) value. The emPAI "offers approximate, label-free, relative quantitation of the proteins in a mixture based on protein coverage by the peptide matches in a database search result" – Matrix Science. The emPAI was developed by Ishihama et al. (150) and is calculated

as shown below where $N_{observable}$ is the theoretical number of peptides a protein could be digested into by trypsin, and $N_{observed}$ is the number of those peptides detected.

$$Protein\ Abundance\ Index\ (PAI) = \frac{N_{observed}}{N_{observable}}$$

$$exponentially\ modified\ PAI\ (emPAI) = 10^{PAI} - 1$$

In brief, the emPAI value reflects the fact that the amount of a protein present in a mixture relative to all other proteins in a mixture is best represented by the ratio of how much of the sequence of each protein is detected in an analysis, rather than by the number of peptides found for each protein (i.e. spectral count). The way emPAI values are calculated by the Mascot program is based on an estimated, rather than definite, value for $N_{observable}$. This is done for reasons of speed and the unavailability of all of the necessary data required in producing a definite value. Matrix Science claim that their method for estimating $N_{observable}$ results in a negligible difference to the final emPAI values calculated. More information about how Mascot calculates emPAI values can be found at http://www.matrixscience.com/help/quant_empai_help.html.

To make the data more useful, the emPAI value for each contig was converted into a molar fraction percentage (MFP) value using the formula described by Shinoda et al. (151) as shown below where Σ(emPAI) is the sum of all individual emPAI values from a single LC-MS run dataset.

$$Molar\ Fraction\ Percentage\ (mol\ \%) = \frac{emPAI}{\sum(emPAI)} \times 100$$

This value represents as a percentage how much of the total amount of protein in the sample a particular protein represented.

The nucleotide sequence of each contig detected was retrieved and submitted to a BLASTx search (http://blast.ncbi.nlm.nih.gov). The highest ranking sensible BLAST hit was recorded for each contig. Here the word sensible is used to mean that if the highest ranking hit was an unknown protein, and the second highest hit was a known/named protein with near identical match scores, the known/named protein was recorded to make the downstream data more meaningful. The UniProt identifier for each of the recorded BLAST assignments was obtained and recorded

(http://www.uniprot.org). The UniProt identifier for each contig was used to select a suitable gene ontology term to allow contigs to be grouped by function (http://www.ebi.ac.uk/QuickGO). If a contig was found to represent a glycosyl hydrolase (GH), the UniProt identifier for that contig was used to determine which GH family the contig belonged to by searching the Carbohydrate-Active enZYme database (CAZy http://www.cazy.org).

#### *3.2.2.5 Notes on Molar Fraction Percentage (MFP) Values*

As MFP values are calculated by dividing each individual emPAI value by the sum of all emPAI values in a particular sample, MFP values can only be used to make direct comparisons within that sample. For example, it would be incorrect to claim that there was twice as much of protein x in the gut sample as compared to the HP sample, one could only claim that protein x was twice as abundant in the gut sample as compared to the HP sample. The actual amount of protein one per cent represents in each sample could be vastly different, therefore, differences must be discussed in terms of relative abundance.

# 3.3 Results

## 3.3.1 Two Dimensional - Polyacrylamide Gel Electrophoresis (2D-PAGE)

The HP of *Limnoria* is lined with secretory cells which are thought to secrete proteins and enzymes into the HP lumen whilst the ingested wood particles are segregated in the inertly lined gut. This arrangement suggests that digestive proteins are likely to be expressed in the HP and moved to the gut to act upon the ingested wood. Therefore examining the proteome of these two organs may indicate proteins which play a major role in the digestive process; it may also help determine how the digestive process is partitioned between these two organs by highlighting the differences in protein localisation between them. To this end 70 HPs and 70 guts from *Limnoria* living on Balau hardwood were subjected to 2D-PAGE analysis.

The first step in this process was to determine how much protein was in each sample of HPs and guts. This was done by running a 10 µl aliquot of homogenised tissue extract next to known amounts of *E. coli* protein on an SDS-PAGE gel. The resulting gel is shown in Figure 3.1. The amount of overall staining produced by the HP sample seems most comparable to that produced by 10 µg of *E. coli* protein, whilst that produced by the gut sample most similar to 5 µg. This rough estimate would mean that the remaining 100 µl of HP and gut tissue extracts contained maximums of approximately 100 and 50 µg of protein respectively.

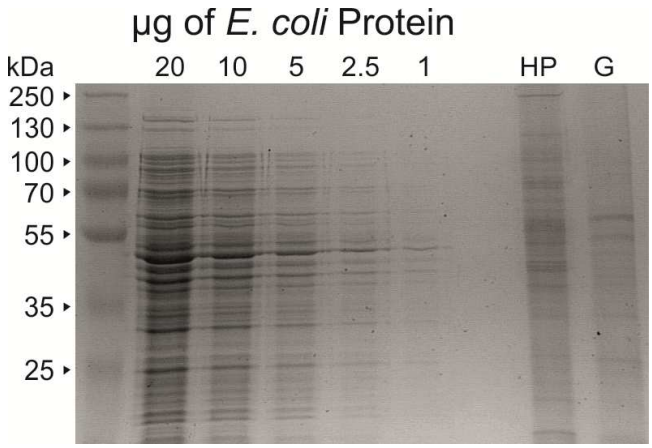



**Figure 3.1. Determination of protein content of *Limnoria* tissue homogenates.** A Coomassie stained SDS-PAGE gel was ran with various quantities of *E. coli* protein to produce a series of standards against which aliquots of *Limnoria* HP and gut (G) homogenates could be compared to determine their approximate protein content.

Once the amount of protein in each extract was estimated it was cleaned up using a ReadyPrep™ 2-D Cleanup Kit (Bio-Rad) which resulted in clean, precipitated semi-dry protein pellets. These protein pellets were re-dissolved before soaking the solutions into IPG strips. The IPG strips used recommended applying a minimum 100 µg of protein to them; as the maximum estimated amount of protein contained in a *Limnoria* tissue extract was 100 µg, and assuming the clean-up and re-dissolving steps would have resulted in some protein loss it was decided to apply all of the available extracts to the IPG strips rather than attempt to load an equal amount of protein on each. This would maximise the chance of obtaining usable data at the end of the process. After performing IEF on the IPG strips they were attached to SDS-PAGE gels and the proteins on the strips separated based on their molecular mass. This produced the two 2D-PAGE gels shown in Figure 3.2.

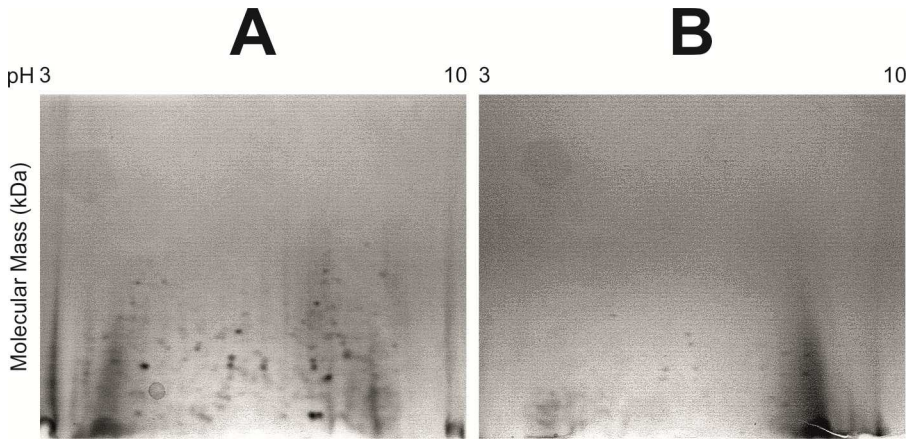



**Figure 3.2. 2D-PAGE gels of *Limnoria* HP and gut tissue.**
Cleaned up protein from homogenised *Limnoria* HPs (A) and guts (B) were used to produce 2D-PAGE gels which were stained with Coomassie.

The two 2D-PAGE gels produced succeeded in separating proteins in both dimensions resulting in well resolved spots (Figure 3.2). However the gels are not as informative as was hoped with relatively few spots visible. This suggests that either there was a considerable over-estimate of the amount of protein loaded, or that the clean-up and re-dissolution steps suffered large amounts of protein loss. The gut 2D-PAGE gel especially (Figure 3.2 – B) is disappointing, with hardly any spots clearly visible. Also most of the spots that are visible in either gel are present in the bottom half of the gels meaning they are of low molecular weight. This suggests that there was degradation of the proteins in the sample during processing.

Twenty of the strongest spots visible on the HP gel (Figure 3.2 – A) and six from the gut gel (Figure 3.2 – B) were excised and subjected to protein identification.

Of the twenty spots analysed from the HP gel, eleven failed to produce data of sufficient quality to allow a match to be made. The spots that did yield data which resulted in a protein being identified did not reveal any proteins of interest; half of them containing actin or myosin and the others metabolic or DNA binding proteins. From the gut gel only one spot resulted in identification of a protein, which was actin.

### 3.3.2 Label-Free Semi-Quantitative Proteomic Analysis

As well as 2D-PAGE, another approach was taken to study the proteomes of the digestive organs of *Limnoria*. This approach involved trypsinising the proteins from 100 HPs and guts of *Limnoria* and separating the resulting peptides by liquid chromatography before finally detecting the peptides by mass spectrometry. The calculation of emPAI values from the resulting data would allow a semi-quantitative analysis to be carried out for each tissue.

Analysis of the proteins from the HPs and guts of 100 *Limnoria* resulted in the acquisition of 2737 and 1587 spectra able to be assigned to peptides respectively. From the HP these spectra represented 484 unique peptides which allowed 210 contigs from *Limnoria* to be identified. For the gut there were 357 unique peptides identifying 148 contigs from *Limnoria*. These contigs could be organised into 25 different classes of genes. The emPAI values for each of these contigs were converted into MFP values and a summary of these data are shown in Figure 3.3. A table showing each detected contig and its associated analysis is shown in Appendix A. For clarity and for comparison only the gene classes highlighted in the transcriptomic analysis of HP tissue (68) are shown in the pie chart of Figure 3.3. In the figure the “unknown” class of proteins refers to those which matched contigs which when subjected to a BLASTx search produced either no hits or a hit labelled as uncharacterised or predicted and for which no further information could be found to identify its function.

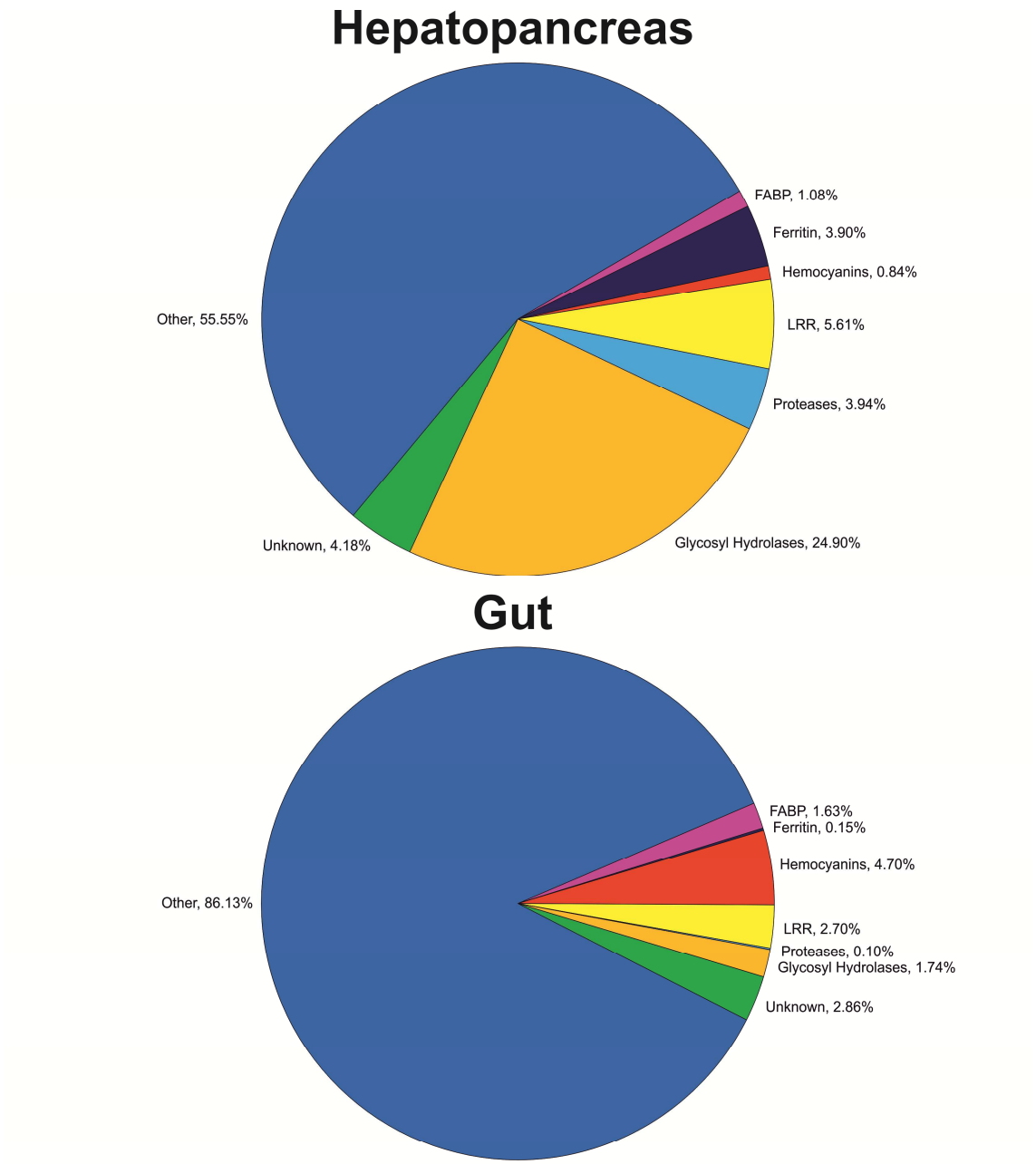



**Figure 3.3. Relative abundances of protein classes from *Limnoria* HP and gut proteomes.**
The molar fraction percentages (MFPs) of each class of proteins found in the HP and gut proteomes were plotted as pie charts to illustrate the relative abundance of each protein class in each tissue. Protein classes featured in the transcriptomic analysis of *Limnoria* HP (68) are shown individually with all extra classes collectively grouped as "other". The percentage values shown in other represent the MFP values for the HP and gut respectively. Categories such as carbohydrate metabolism, hydrolase and metabolic process refer to intracellular proteins involved in homeostatic functions rather than those involved in the digestion of ingested material – see Appendix A for further

details. Abbreviations: FABP, fatty acid binding protein; LRR, leucine rich repeat protein.

All of the classes of proteins highlighted in the transcriptomic analysis of *Limnoria*'s HP were represented in the proteomes of both the HP and the gut (68), although not necessarily to the same degree as their transcripts. Most of the classes were much less abundant in the gut tissue compared to the HP, namely; proteases 39.4 fold, ferritin 26 fold and GHs 14.3 fold less. The LRR proteins were moderately less abundant in the gut representing 2.1 fold less of the total protein than was found in the HP. The unknown proteins, about which no information could be obtained, represented slightly less of the total protein of the gut at 1.46 fold lower than the HP. However, two of the protein classes accounted for larger proportions of the total gut protein compared to the HP; hemocyanins (HCs) 5.6 fold more and the FABPs a moderate 1.5 fold more.

The most abundant single class of proteins present in the HP were the GHs, representing a little under a quarter of all of the protein of the HP (24.90 %). However these proteins represented only 1.74 % of the gut proteome. Not only was a larger proportion of the HP protein made up of GHs, but more GH families were detected in that tissue with the HP having members of 8 GH families represented compared to only 4 in the gut Figure 3.4.

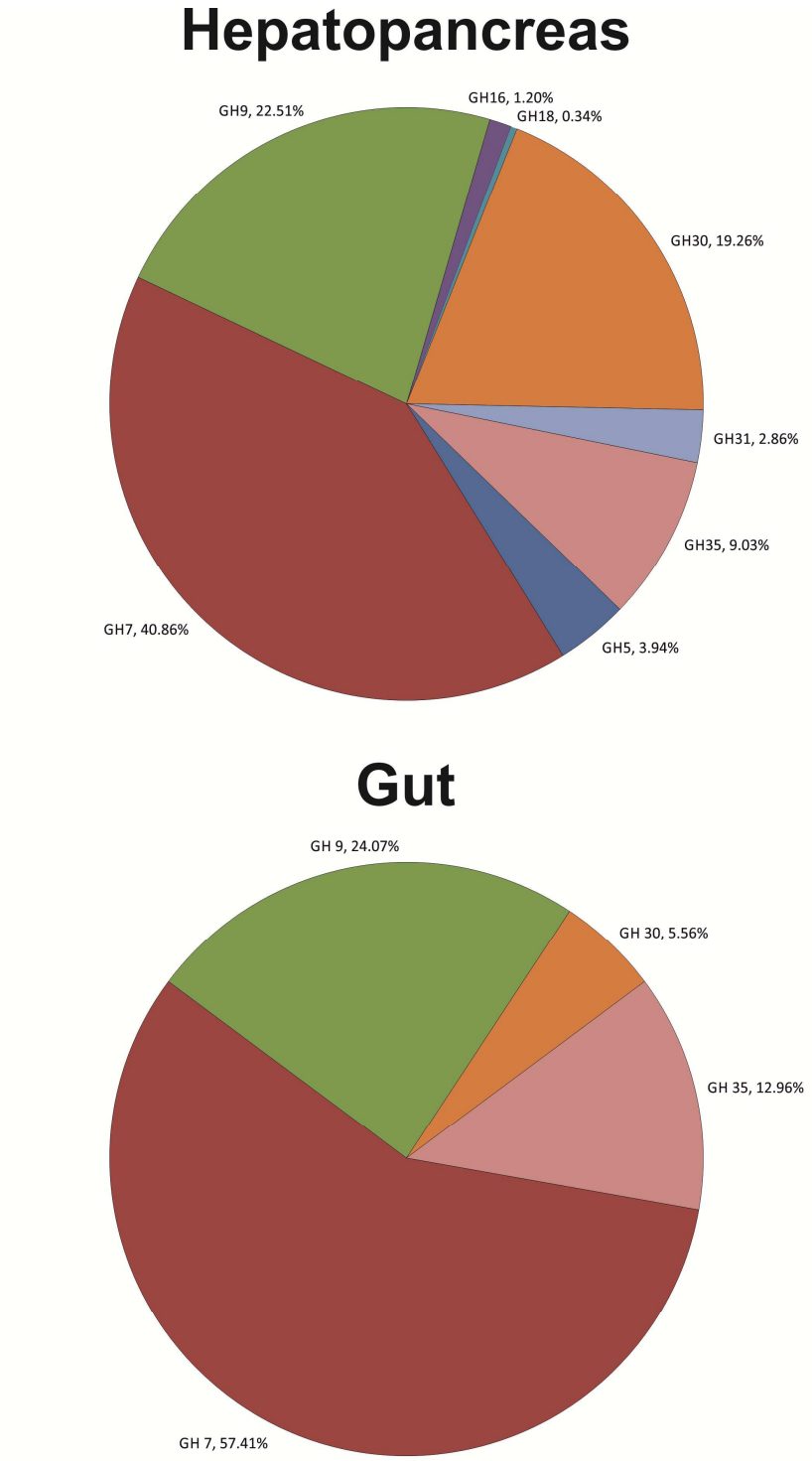



**Figure 3.4. Relative abundances of GH families from *Limnoria* HP and gut proteomes.**
The molar fraction percentage (MFP) of GHs from each tissue were divided into the proportions of those values represented by each GH family and these proportions were plotted as pie charts. The total amount of protein represented by each pie chart is different, with the HP chart representing 24.90 % of the total protein content of that organ and the gut chart representing 1.74 %.

A major difference between the proteomes of the two different organs is the proportion of proteins designated as “other,” with 86.13 % of the gut falling into this category as opposed to only 55.55 % of the HP. In the gut some classes of house-keeping proteins are relatively much more abundant than they are in the HP such as ATP metabolism (~2.4 fold more), cytochrome-c oxidase (~21 fold more), cytoskeleton (~2 fold more), metabolic processes (~3 fold more) and transferases (~2.4 fold more). The increased abundance of the classes of protein listed in this paragraph account for 29.1 % of the gut proteome and goes some way to explaining why the “other” category is a considerably larger proportion of gut protein as opposed to HP protein.

# 3.4 Discussion

## 3.4.1 2D-PAGE versus Label Free Semi Quantitative

Ultimately it was possible to successfully obtain 2D-PAGE gels for both HP and gut tissues of *Limnoria*. However, despite undertaking multiple rounds of method development to improve the procedure for producing 2D-PAGE gels very little data was obtained using this method. The resulting gels (Figure 3.2) showed very few well resolved spots and most of the spots were present at the low molecular weight end of the gels indicating possible sample degradation. This was despite taking every reasonable precaution to avoid protein degradation during processing. In comparison to the label-free semi-quantitative approach the 2D-PAGE method was quite labour intensive and yielded very little data whilst requiring quite a large number of animals to be sacrificed. The label-free semi-quantitative approach was much more effective resulting in the generation of more useful data for a modest increase in the number of animals used.

## 3.4.2 Gut and Hepatopancreas Proteome Comparison

The label-free semi-quantitative approach was a much more effective way to visualise the proteomes of the digestive organs of *Limnoria*. The technique was able to identify a total of 694 peptides across both tissues compared to the 45 peptides found using 2D-PAGE. The technique also removed human sampling bias as all peptides in the samples would have had a chance to be detected during the LC-MS run rather than only those in gel spots chosen for analysis.

Glycosyl hydrolases are enzymes which catalyse the hydrolysis or re-arrangement of glycosidic bonds, the bonds between sugar molecules. The wood food source which *Limnoria* feeds on is composed broadly of cellulose, hemicellulose and lignin. Both cellulose and hemicellulose are polysaccharides composed of sugar molecules joined together by an extensive array of different glycosidic bonds. It would make sense therefore for an animal which utilises such a substrate as its primary source of nutrition to possess glycosyl hydrolase genes and express them relatively abundantly whilst feeding. For an enzyme to be involved in digestion it would likely be made in and secreted into the lumen of the HP, and then moved into the gut to act upon ingested wood particles. It comes as little surprise therefore that nearly a quarter of all protein

present in the HP constitutes GHs (24.90 %). This also makes sense in the light of the transcriptomic data previously published (68) which showed that GHs were very highly transcribed in the HP (27.0 % of all transcripts).

It does seem surprising however that glycosyl hydrolase protein makes up such a small proportion of the gut proteome (1.74 %). Although it has been shown that this GH protein found in the gut must have been transported there from the HP as representatives of the most abundantly transcribed glycosyl hydrolase gene families expressed in the HP (GH families 5, 7 and 9) have been shown to be exclusively transcribed in the HP ((61) and Dr. Katrin Beßer – University of York, unpublished data).

As wood particles in the gut have no way of entering the HP of the animal, for the wood particles to be digested enzymes must be brought to them. Therefore if GHs are involved in the breakdown of the wood substrate one would expect to find them in the gut and as the polysaccharides present in the wood would constitute the animals primary carbon source one would expect them to be abundantly present in order to maximise the amount of sugar that could be solubilised from the wood particles. As glycosyl hydrolase proteins are not abundantly present in the gut of *Limnoria* this suggests two possibilities.

The first possibility is that the glycosyl hydrolase proteins are bound tightly to the wood particles and either during dissection these wood particles are lost along with the protein or the GH proteins are not removed from the insoluble wood particles during protein extraction. It would seem unlikely that the proteins are lost during dissection as when dissecting a gut out of an animal; some gut contents are occasionally lost, but not always, and not so completely as to reflect the stark difference in abundance between the two tissue types. It is possible that the GHs are bound so tightly to the wood particles that they are not removed during sample processing. Indeed faecal pellets produced by the animal have been shown by Western blot to contain GH family 7 protein (Dr. Katrin Beßer – University of York; unpublished data), the same protein family which was the most abundantly represented GH in *Limnoria*'s gut (Figure 3.4). However, no GH family 9 protein was found present in the faecal pellets despite being the second most abundantly represented GH family in *Limnoria*'s gut. This favours the second possibility.

The second possibility is that *Limnoria* does not use specific glycosyl hydrolase breakdown of wood as a primary digestive tool; but rather primarily utilises a non-specific mechanism to attack the substrate releasing smaller soluble oligosaccharides. Glycosyl hydrolases are present in the gut so it would seem that they may then act as a secondary form for breakdown, reducing the size of the oligosaccharides further reducing them into smaller oligosaccharides or disaccharides, which can be withdrawn back to the HP where other GHs can break them down to their constituent sugars for absorption. This proposal seems to be supported by several pieces of evidence. Firstly the much lower relative abundance of GH proteins in the gut suggests that most of the GHs produced by the HP are retained there by an unknown mechanism. Of the GH proteins that are found in the gut 57.41 % of them are members of family 7, known as cellulases (Figure 3.4). These enzymes are able to facilitate cleavage within a cellulose chain liberating smaller oligosaccharides which would agree with the general hypothesis detailed above. Secondly data from our lab which shows that the relative sugar composition of the wood that *Limnoria* feeds on and the faecal pellets that the animal egests are little changed (Dr. Graham Malyon – University of Portsmouth, Dr. Leonardo Gomez – University of York; unpublished data). If glycosyl hydrolase attack of the wood particles were taking place one would expect this to favour the removable of a particular moiety from the substrate reducing the amount of a particular sugar relative to others. This is not seen.

A major difference between the tissues is the almost lack of proteases in the gut proteome, 0.1 %, 39.4 fold lower than the proportion of protein dedicated to them in the HP proteome. Primarily this could be explained by the need of a protein secretory organ to be able to post-translationally process proteins using proteases. As this type of activity would be much lower in the gut this could explain the difference in protease content seen. However proteases are also part of the digestive system, being required to break down nutritional protein into peptides or amino acids for absorption and re-use as well as the recycling of self-proteins. The wood that *Limnoria* feeds on has a very low protein content (152) and it is likely that most of the protein that *Limnoria* consumes comes from microbes or algae living on the wood that it ingests. It is known that none of these organisms survive the digestive process to be able to colonise the gut of *Limnoria* (68), it is also known that the faecal pellets of *Limnoria* are devoid of bacterial life so there must be some process occurring in the gut to kill any creatures living on the ingested wood (153). This process would likely result in the lysis of the microbes which would liberate proteins from the killed organisms which could be passed back to the HP. In this way it would seem unnecessary for there to be proteases in the gut

where they would risk loss in the digestive transit or destruction by the hypothesised non-specific lignocellulose attack mechanism; rather partition the proteases into the HP where they could degrade any proteins retrieved from wood infesting microbes.

The ferritin proteins were also present as a very minor component of the gut proteome representing only 0.15 % it; 26 fold less than their proportion of the HP protein. Ferritin proteins store dangerous $Fe_{(II)}$ ions from their environment safely inside themselves as a core of hydrous ferric oxide ($[Fe_{(III)}O \cdot OH]$) composed of up to 4500 iron atoms which if left free in solution could go on to cause oxidative damage (154, 155). This difference in protein abundance suggests that it is much more important for potentially dangerous $Fe_{(II)}$ to be scavenged and sequestered in the HP than it is in the gut. This may be because the participation of $Fe_{(II)}$ in oxidative damage in the gut is actually beneficial. Furthermore, the gut is a sclerotized tube designed to survive the potential for rupture from the passage of sharp wood particles and therefore is perhaps not as susceptible to oxidative damage as the highly active HP.

Another interesting feature of the proteomic data is the difference in the abundance of the HC proteins. They were found to be 5.6 fold more abundant in the gut tissue than in the HP. This class of proteins were the second most abundantly expressed group of genes in the HP transcriptome, representing 17.3 % of all ESTs produced (68), and yet they only constituted 0.84 % of the HP proteome. Hemocyanins are classically regarded as the oxygen carrying component in arthropods and molluscs, analogous to haemoglobin (156). However, the literature suggests that they may be capable of other functions, possibly even an enzymatic function. This literature is discussed in detail in Chapter 4. The possible enzyme function and the potential roles that this implies for the proteins, together with the relatively high abundance of HCs in the gut tissue of *Limnoria* marks them as an interesting group of proteins for further study.

The transcriptome of *Limnoria*'s HP showed two further families which were highly represented. The leucine-rich repeat proteins (LRRs) and fatty acid binding proteins (FABPs) each represented 2.4 and 1.1 % of all of the transcripts sequenced from *Limnoria*'s HP respectively (68). Seven full length LRR genes were cloned by other group members from *Limnoria* HP cDNA along with two FABP genes.

Neither of the FABP genes possessed secretory signal sequences (as assessed by SignalP 3.0 – http://www.cbs.dtu.dk/services/SignalP-3.0/) and so are likely to be intracellular proteins. The FABP genes were also found to be expressed in both HP

and gut tissues (Dr. Katrin Beßer – University of York; unpublished data). Taken together, these two pieces of information make it unlikely that the FABP genes are involved directly in digestion as they are likely to be present inside the cells of both tissues, rather than secreted where they can act upon ingested wood particles.

The LRR genes were found to only be transcribed in HP tissue (Dr. Katrin Beßer – University of York; unpublished data) and all seven LRR genes are predicted to be secreted (as assessed by SignalP 3.0), and indeed LRR protein was found to constitute 5.61 and 2.70 % of the HP and gut proteomes respectively. This suggests that the LRRs may play a role in digestion in *Limnoria* as they are produced in the HP but are moved to the gut subsequently where they could function in the digestive process. Performing a BLASTx search on the LRR protein sequences reveals several proteins with portions of similar sequence; however these portions seem to only relate to the LRR motif and therefore shed no light on the possible function of these proteins in *Limnoria*. Structural modelling reveals that the *Limnoria* LRR proteins thread well into the known structure of a polygalacturonase-inhibiting protein (PGIP) from *Phaseolus vulgaris* (John McGeehan, University of Portsmouth, unpublished data). Polygalacturonase-inhibiting proteins are able to bind pectin and fungal polygalacturonases, and in doing so inhibit the fungal enzymes from breaking down the pectin found in plant cell walls preventing fungal invasion of the plant tissue (157). Although the *Limnoria* LRRs show conservation with and model well with PGIP from *P. vulgaris* the pectin binding residues are not conserved between the two. This indicates that the LRRs are not acting as pectin binding molecules, but it does open up the possibility that these proteins may act to complex both other proteins and oligosaccharides. If this is so it would likely be a very useful feature of *Limnoria*'s digestive system.

The proportion of proteins from each tissue designated as "other" differs greatly. Certain classes of house-keeping proteins such as those involved in ATP metabolism, cytochrome-c oxidase, cytoskeleton, metabolic processes and transferases are relatively much more abundant in the gut than in the HP (each representing at least twice as much of the total protein compared to the HP). Exactly why this should be is not readily apparent but perhaps indicates that the musculature surrounding the gut is more substantial than that around the HP. This may be due to the extra effort involved in peristaltically moving wood particles down the gut compared to merely contracting the lumen of the fluid filled HP. The more general finding that the "other" category of proteins was much more abundant in the gut when compared to the HP may be due to

a magnification of the difference caused by the relative nature of the way the MFPs are calculated. For example, if the HP is an organ designed to secrete large amounts of a few classes of proteins then this tissue may produce much more protein per volume of tissue as compared to another tissue which is not secretory. This would lead to those secreted proteins representing a large fraction of that tissue, whilst at the same time making the MFP of ordinary house-keeping genes disproportionately small (most of which are represented under “other”).

From the proteomic data discussed in this chapter the GH, HC, Ferritin and LRR proteins all fit the criteria set out in my initial hypothesis to suggest they may have a role in digestion in *Limnoria*. They are all transcribed in the HP of *Limnoria* (61, 68) but their protein is found to a greater or lesser extent in both the HP and the gut where they could be involved in digestion. Reinforcing this suggestion are the findings that all of the genes are exclusively expressed in the HP and possess secretion signal peptides (HCs, see Chapter 4; GH7, see (61); all other genes, Dr. Katrin Beßer – University of York, unpublished data). This means that for these proteins to be present in the gut of *Limnoria* they will have had to be transported to that organ after translation and that this would occur as those proteins are needed for a specific role in that organ. As the main role of the gut is as a place for digestion to occur this implicates these proteins in that process.

# Chapter 4: The Involvement of Hemocyanins in Digestion

## 4.1 Introduction

Hemocyanins (HC) are members of the type III copper protein family and exist freely soluble in the hemolymph of many arthropod and mollusc species (137). Other members of this family of proteins include phenoloxidases (PO) and tyrosinases (137).

Arthropod HCs are made up of a single polypeptide chain arranged into three distinct domains (158, 159). Each polypeptide chain has a molecular mass of approximately 75 kDa. The N-terminal domain (NTD) or all-alpha domain is composed of 6 alpha helices, 5 of which surround a central helix. The middle domain or HC copper containing domain has a multi-helical arrangement containing the di-copper centre, and coordinates two copper atoms, each via three histidine ligands. The C-terminal or Ig-like domain has been classified as containing an immunoglobulin-like beta-sandwich fold, and as a member of the E set domain superfamily (Structural Classification of Proteins – http://http://scop.mrc-lmb.cam.ac.uk/scop/). *In vivo* arthropod HCs aggregate into hetero hexamers, or multiples of hexamers up to 48mers depending on the species of origin (reviewed in (160)).

Molluscan HCs are structurally completely different to those of arthropods with the exception of the fashion in which the type III copper centre binds copper and oxygen (161). Molluscan HC molecules consist of 10 polypeptide chains arranged as large cylindrically shaped macromolecules which can oligomerise themselves into dimers or further. Each polypeptide chain has a molecular mass of 0.34 – 0.43 MDa and consists of 7-8 functional units, meaning a HC molecule consisting of a single decamer could weigh up to 4.3 MDa (160, 161). A functional unit is divided into two domains, domains II and III. Domain II, also known as the core domain is analogous to the HC middle domain of arthropod HC, and contains the two copper atoms and is where oxygen is bound (162). Domain III, or the β sandwich domain is analogous to the NTD of arthropod HC and folds over the copper centre excluding it from the solvent (162).

As *Limnoria* is a member of the arthropod phylum, from this point on any reference to HCs, unless otherwise stated, refers to those of arthropods.

Each HC molecule is able to reversibly bind one molecule of di-oxygen. In doing so the copper atoms are oxidised to the $Cu_{(II)}$ state, and the oxygen is bound as a peroxide anion, the complex having $\mu$-$\eta^2$: $\eta^2$ coordination geometry (163).

The long established physiological role of HCs is as the oxygen carrying component of the hemolymph (137). The affinity of HC molecules for oxygen is highly variable across different species, reflected in the wide range of habitats occupied by species utilising them (160). Within a particular species the affinity is also highly tuneable and is dependent on such variables as hexamer subunit composition (164), pH (165), concentration of various allosteric effectors (165, 166) and cooperativity effects (167-169).

Recently, various arthropod HCs have been shown to catalyse an enzymatic reaction, namely the oxidation of mono/di-phenols to *o*-quinones, when placed under certain conditions (137, 170). This PO activity of HCs has been elicited by various salts, mild denaturation by detergents such as SDS and phospholipids, limited trypsin proteolysis and complex formation with other proteins (i.e. clotting factors) (171-177).

Work has been done to understand the mechanism of this activation, and it is now believed that all of these activators somehow bring about a change in the spatial location of the NTD of HC. The NTD contains a conserved aromatic residue (often phenylalanine) which several crystal structures have shown to point towards the di-copper centre of HC, blocking access to the metal centre for bulky phenolic substrates (178) (Figure 4.5). It is thought that the activators in some way dislodge this conserved residue, either by limited conformational change in the case of detergents, salts and effector proteins, or by cleaving the N -terminal domain from the rest of the HC protein in the case of mild trypsinolysis.

In arthropods POs are required for several essential processes, namely melanin production, sclerotization (hardening) of the cuticle after moulting, wound healing, and are involved in the immune response (179-181). Puzzlingly the EST library created for *Limnoria* showed no transcripts for a (pro)phenoloxidase enzyme (68). It did however show a large abundance of transcripts for HC genes, and given the potential for activation of HCs to perform a PO function, it is tempting to believe that *Limnoria* contains no *sensu stricto* PO enzymes, but rather has evolved to give its HC genes a dual role.

Phenoloxidase enzymes catalyse the conversion of endogenous mono/di-phenols to their more reactive *o*-quinones. This represents the first committed step in the pathways of the important processes mentioned earlier. Many recent reviews are available which discuss in detail how PO activity is involved in these processes, so they shall not be covered here.

From the work discussed in Chapter 3 it is known that HCs are transcribed in the HP of *Limnoria* and that HC protein is found to represent a large proportion of the total protein content of the gut. It is also known that the gut of *Limnoria* is an anoxic environment (Figure 4.1) (153). If *Limnoria* HC does indeed also function as a PO enzyme *in vivo*, then there are four potential processes in the gut which they may be involved in. These processes are bringing oxygen to the gut for an oxygen consuming process, sclerotization of the gut, contributing to the aseptic environment of the gut and lignin modification. How PO activity could be involved in these processes will be discussed in more detail below.

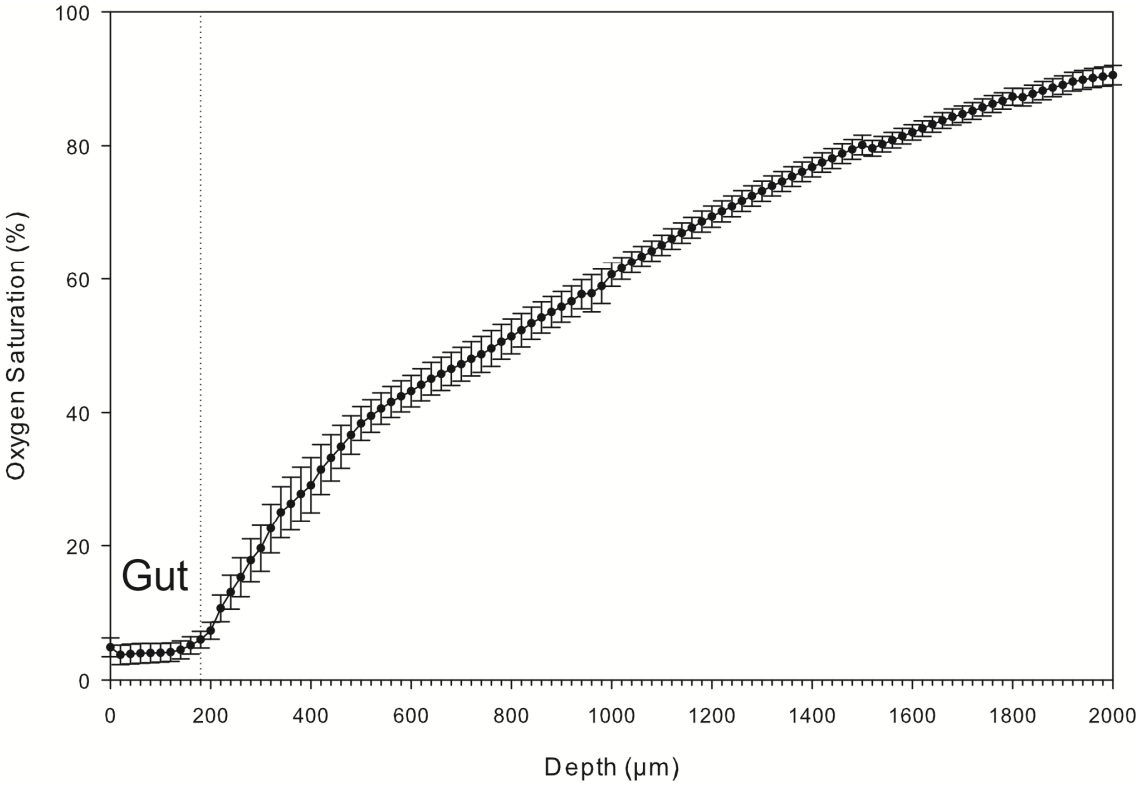



**Figure 4.1. Oxygen concentration of the gut of *Limnoria*.**
An oxygen microsensor was inserted ventrally into the gut of *Limnoria* (depth 0 µm) and withdrawn from the animal whilst recording. At 180 µm the probe left the gut of the animals. This was performed on 10 animals and the plot shows the mean and standard error of these 10 measurements. Data reproduced from (153) with the permission of Dr. Graham Malyon (University of Portsmouth).

Given that the long established role of HCs in arthropods is to transport oxygen around the body to where it is needed, then the simplest explanation for why HC protein is found in *Limnoria*'s gut is that it is there to transport oxygen. As it has been shown that in fact the gut of *Limnoria* is a very anoxic environment then this raises two possibilities. First, that this hypothesised role is incorrect as the evidence shows there is not oxygen in the gut. The second possibility is that the oxygen brought into the gut is rapidly consumed by a chemical process. If *Limnoria* HCs are able to carry out a PO enzyme activity then this could be the oxygen consuming process.

The second potential role for PO activity in the gut is sclerotization of the gut lining. In arthropods, all surfaces exposed to the environment are covered with a cuticle, which is sclerotized to a greater or lesser extent, i.e. the exoskeleton, gut lining, and tracheae (182). This sclerotization provides structural strength, a barrier to pathogens and prevents dehydration (179, 183). The gut of *Limnoria* is packed full of masticated wood particles which could potentially rupture the gut lining as they pass through during digestion (68). Active PO enzymes in the gut could serve to reinforce the lining of the gut by increasing its degree of sclerotization. This theory would support the view that the gut does not actively contribute to the digestive process, but rather functions mostly as an inert tube containing the ingested wood and HP secreted enzymes. In this way it is behaving essentially as an enzyme reactor, as the sclerotization would render the gut useless for nutrient uptake and protein secretion, but would improve its ability to withstand hostile digestion chemistries.

Phenoloxidase activity has also been shown to have an anti-microbial effect in arthropods (180, 181). This is thought to be mediated by encapsulating the microorganisms in melanin once they have penetrated the arthropod's exoskeleton, essentially isolating them from being able to perform any further harm to the animal (184, 185). However there is another possibility which may be more relevant to PO activity in the gut of *Limnoria*.

A recent study has shown that purified HC from a horseshoe crab (*Carcinoscorpius rotundicauda*), once activated by bacterial proteases, is able to catalyse the oxidation of 4-methylcatechol (a diphenol) to its corresponding quinone (186). The study went on to show that the production of this quinone was sufficient to kill bacteria *in-vitro*. The cytotoxicity of quinones has been known for some time, and is believed to arise through their generation of reactive oxygen species (ROS), which are well known to be able to oxidise organic molecules and kill bacteria (187-189).

The ROS are believed to be created by the non-enzymatic reduction of PO created quinones to their semi-quinone radical counterparts in the presence of trace amounts of reduced transition metal ions (eg. $Fe_{(II)}$). These semi-quinone radicals are in turn able to reduce molecular oxygen to the superoxide anion radical ($O_2^{-\cdot}$) which sets up the possibility, via Fenton type reactions (190), for the creation of ROS in the gut of *Limnoria* (Figure 4.2).

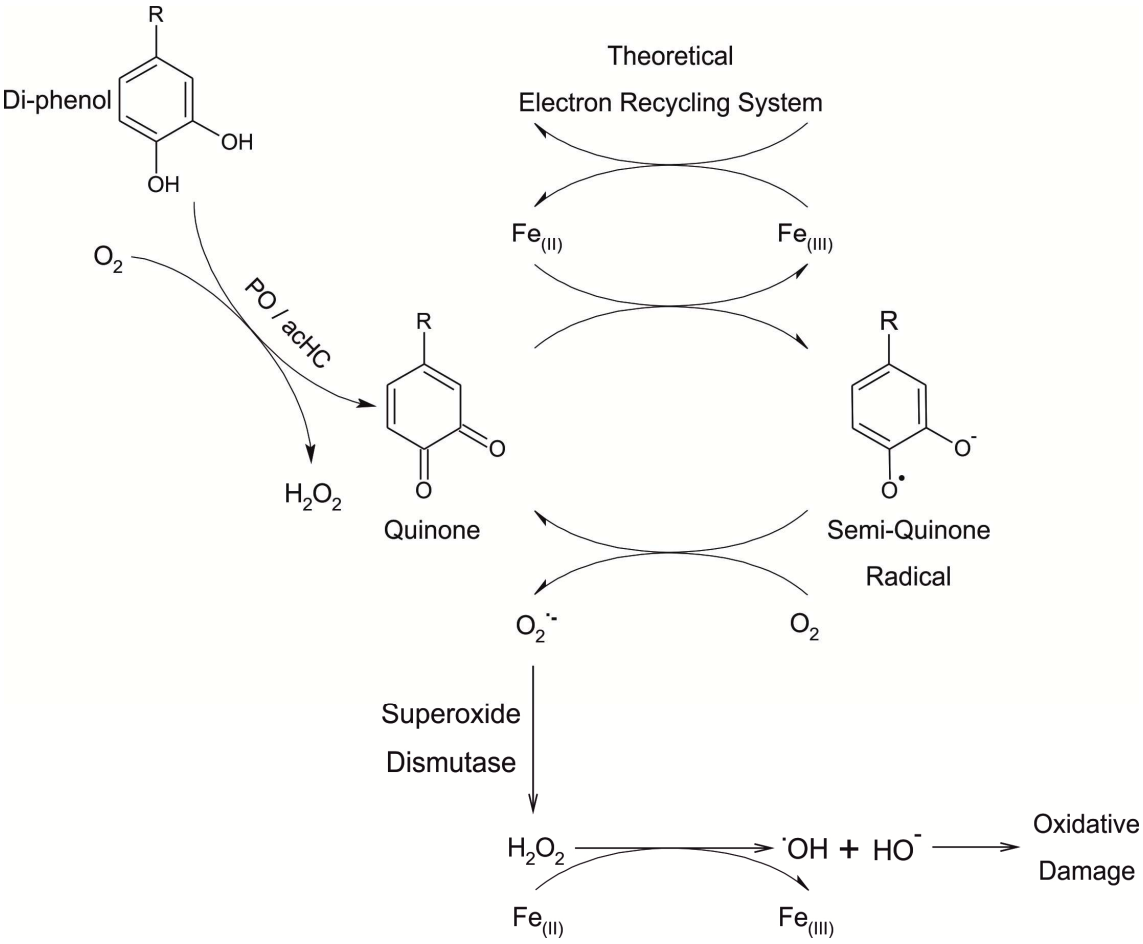



**Figure 4.2. Possible reaction scheme for the generation of reactive oxygen species.**
A reaction scheme demonstrating how the enzymatic oxidation of di-phenol compounds could result in the generation of reactive oxygen species causing oxidative damage. PO, phenoloxidase enzyme; acHC, activated hemocyanin protein. Adapted from (188, 190-192).

The final role proposed for the possible PO activity of HC is in the modification of lignin. Previous studies of white and brown rot fungal species have shown that ROS play an important role in the modification and breakdown of lignin and probably of cellulose (193-197). It seems that the ROS achieve this by non-specifically attacking the super

structure of both components, resulting in bond cleavage. If the above process for creating ROS via quinone redox cycling exists in the gut of *Limnoria*, this may provide a mechanism which would allow *Limnoria* to disrupt the structure of the lignin, and open up the crystalline cellulose fibres giving its GH enzymes access to more exposed cellulose strands (Figure 4.2).

### 4.1.2 Aims of this Chapter

The work described in this chapter aimed to show how the HC proteins of *Limnoria* could be involved in some aspect of the animal's digestive system. This revolved around the potential for *Limnoria*'s HCs to carry out a PO activity as suggested by the literature. To see if the HC proteins of *Limnoria* possessed the potential to carry out a PO activity a multiple sequence alignment (MSA) was carried out to compare them with other HC/PO proteins which could carry out PO reactions and for which structural data was available.

Building upon the proteomic data shown in Chapter 3, work was carried out to find which tissues HCs were transcribed in, and in which tissues HC protein was found. It was hoped that this data would support the hypothesis that proteins thought to be involved in digestion were expressed in the HP and transported into the gut to take part in digestion.

Finally work was undertaken to produce evidence that a HC protein in *Limnoria* tissue was able to be activated to carry out a PO activity, and to show that this PO activity was linked with the production of oxidative species such as peroxides.

## 4.2 Methods

### 4.2.1 Comparison of Hemocyanins from *Limnoria* and other Arthropods

The three dimensional structures of several arthropod HCs and a prophenoloxidase (PPO) have been previously elucidated and annotated. These sequences, structures and annotations were used to look for conserved features which would allow PO activity to be possible in the *Limnoria* HCs such as the copper atom binding sites.

#### *4.2.1.1 Sequence Analysis*

The ClustalX v.2.1 software program (http://www.clustal.org) was used to produce a MSA of the five full-length *Limnoria* HC protein sequences and the sequences of three other HC or PPO proteins from arthropods. The protein sequences were also compared with the *Limnoria* sequences using the BLASTp tool in order to calculate their degree of sequence identity (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). Finally the *Limnoria* HC sequences were analysed using the Pfam sequence analysis tool (http://pfam.sanger.ac.uk) to predict the presence of conserved protein domains in these sequences.

#### *4.2.1.2 Three-Dimensional Structures of HCs*

The three dimensional structures of the proteins used to generate the MSA have been previously elucidated. These molecular structures have revealed which amino acid residues are responsible for various features and behaviours of these HC molecules, which have been annotated alongside their sequences in the Protein Data Bank (http://www.pdb.org) as well as written about in several peer reviewed papers. These annotations were used in concert with the MSA described above to determine which features are conserved and which are not between HCs known to be able to carry out PO enzyme activity and the *Limnoria* HCs.

## 4.2.2 Transcript and Protein Distributions of Hemocyanins in *Limnoria*

### *4.2.2.1 Reverse Transcription Quantitative Polymerase Chain Reaction (qPRC) Analysis*

All qPCR reactions were set up in MicroAmp® Fast Optical 96-well reaction plates, 0.1 ml (Applied Biosystems) and sealed with Optical Adhesive Films (Applied Biosystems). The qPCR reactions were carried out in a StepOnePlus™ instrument with the following cycle conditions; initial denaturation 95 ℃, 20 seconds; cycling stage, 40 cycles, denaturation 95 ℃, 3 seconds, annealing and extension 60 ℃, 30 seconds; melt curve stage, denaturation 95 ℃, 15 seconds, annealing 60 ℃, 60 seconds, rising to 95 ℃ in 0.3 ℃ increments.

### *4.2.2.2 Primer Design*

Oligonucleotide primers for qPCR were designed to meet several criteria. Each primer should have a melting temperature ($T_m$) between 58 and 60 ℃ as evaluated by Primer Express version 2.0 (Invitrogen Life Technologies). The five terminal (3') bases must not consist of a total of more than two cytosine or guanine bases. A primer pair should produce an amplicon of between 50 and 150 nucleotides. No primer should contain a repeat of more than three bases (i.e. > GGG). As the five *Limnoria* HC genes possess a high degree of homology, where possible the terminal 3' nucleotide of the primer was chosen so as to be different to the nucleotide present in the equivalent place in the other four HC genes. The gene sequences of the *Limnoria* HC genes are shown in Appendix B.

Using these criteria primers were designed for *Limnoria* HC 2 and 3 genes using a combination of "by eye" evaluation and $T_m$ checking with Primer Express.

Primers for qPCR of the *Limnoria* HC 1, 4 and 7 genes were previously designed and evaluated by Dr. Katrin Beßer (University of York); as well as primers for the *Limnoria* ubiquitin gene and a middle and 5' fragment of the glyceraldehyde 3-phosphate dehydrogenase (GAPDH) gene.

The primers designed and to be used in qPCR experiments are shown in Table 4.1.

**Table 4.1. Details of the oligonucleotide primers used for qPCR analysis.**

| Primer Pair | F/R | Sequence (5'→3') | Amplicon Length (bp) |
|---|---|---|---|
| HC 1.3 | F | TCTGCTATTGTTTCACGACTTAATCAT | 85 |
| | R | TAGCGAAGACATCTGCTGCATT | |
| HC 2.1 | F | TCTAACCACATTACCAGGAAATCAAA | 92 |
| | R | GCAGCGTCAGCTTGTTCAAG | |
| HC 2.2 | F | CACATTACCAGGAAATCAAAGGATACT | 89 |
| | R | ACGGCAGCGTCAGCTTGT | |
| HC 2.3 | F | GCTGCCGTTTCATCTGGATCT | 112 |
| | R | CTGCAAACTCCATGCCTTCA | |
| HC 3.1 | F | TTCTCGAGGGTGTGGTATTCCT | 83 |
| | R | CAGCTACAATAACGGCGAACTC | |
| HC 3.2 | F | GATTTGGATTTGTCCGCTTATTCT | 77 |
| | R | CCTTGGGCATTGCCCTTA | |
| HC 3.3 | F | GATCTAACCACATAACACGAAAATCAT | 133 |
| | R | GAATAAGCGGACAAATCCAAATCT | |
| HC 4.3 | F | AACCACGTAACCCGAAAATCAT | 129 |
| | R | TGAATAAGCAGACAAGTCCAAATCC | |
| HC 7.1 | F | GACAAATTCTGGGTGATTGTAAATGA | 85 |
| | R | TCCTGTCGGGAATGGTCACT | |
| Ubiquitin | F | GGTTGATCTTTGCCGGAAAG | 92 |
| | R | TCTCAAAACGAGGTGAAGTGTTG | |
| GAPDH (mid) | F | CTCTACCTCCGCGCCAATC | 65 |
| | R | CGCTGTAACGGCTACTCAGAAGA | |
| GAPDH (5') | F | TGTAATTTTCCTTCCATCGACAAC | 76 |
| | R | CTCCACACACGGTCGCTACA | |

**Table 4.2. Distance of amplicon start from the 3' end of a target gene.**
Listed are the number of base pairs that a particular primer pair begins amplifying their amplicon from the 3' end of their target gene.

| Primer Pair | Number of Base Pairs Amplicon Start from Target Gene 3' End |
|---|---|
| HC 1.3 | 621 |
| HC 2.2 | 410 |
| HC 3.3 | 398 |
| HC 4.3 | 417 |
| HC 7.1 | 447 |
| Ubiquitin | 341 |
| GAPDH (5') | 865 |
| GAPDH (mid) | 478 |

### *4.2.2.3 Primer Testing*

The qPCR primers for the HC 2 and 3 genes were tested for their ability to effectively amplify their single intended product. To assess if they were able to amplify efficiently and only produce one amplification product they were used to carry out qPCR amplifications using *Limnoria* HP cDNA donated by Dr. Katrin Beßer (University of York) as a source of template DNA. Each microliter of this cDNA represented 6.3 ng of RNA starting material.

For each primer pair triplicate reactions were set up containing 1 µl of HP cDNA (equivalent to 6.3 ng of RNA), 2 µl of primers (final concentration of 400 nM per primer), 9.5 µl of RNAse free water and 12.5 µl of Fast SYBR® Green Master Mix (Applied Biosystems). Reactions were also set up in triplicate for each primer pair substituting template cDNA for an extra microliter of RNAse free water. The reaction plate's contents were mixed and briefly centrifuged to bring down any liquid before carrying out the amplification.

Once the amplification was complete 1 µl of each reaction was ran on a 3 % w/v agarose gel (Section 2.4.4) to confirm the production of a single correctly sized amplicon.

To verify that the amplifications had resulted in the intended specific sequences being amplified, the HC 2 and 3 primer pairs chosen to take forward, along with the already

designed primer pairs for HCs 1, 4 and 7 were amplified by standard PCR, cloned and sequenced using standard procedures described in Chapter 2.

### *4.2.2.4 Determining Primer Efficiency*

The PCR and doubling efficiency of primer pairs was determined to be used as a correction factor when quantifying transcript levels in *Limnoria* tissues. Reactions were set up using a doubling dilution of HP cDNA to produce standard curves. The cDNA was donated by Dr. Katrin Beßer (University of York) and each microliter of the top standard of this cDNA represented 6.3 ng of RNA starting material. The following dilutions were used as standards: 1 in 1, 2, 4, 8, 16, 32, 64. Each reaction was set up in triplicate and contained 3 µl of HP cDNA (equivalent to ~19 ng of RNA in the neat standard), 1.6 µl of primers (final concentration of 400 nM per primer), 5.4 µl of RNAse free water and 10 µl of Fast SYBR® Green Master Mix (Applied Biosystems). Negative controls containing water instead of cDNA were also ran for each primer pair in triplicate.

Once the amplifications were run the StepOnePlus Software program automatically produced standard curves for each primer pair from the data, plotting dilution factor against cycle threshold values ($C_t$). The slopes of these standard curves were used to calculate the doubling efficiencies for each primer pair using the formula shown below.

$$Doubling\ Efficiency = 10^{-1/slope}$$

### *4.2.2.5 Isolation of Limnoria RNA and Protein*

The RNA and protein were extracted from *Limnoria* HPs and guts using a TRIzol® reagent (Invitrogen Life Technologies) based method. Throughout the isolation procedure certified RNAse free water (Ambion – Invitrogen Life Technologies) was used rather than ordinary double distilled lab water to avoid the risk of RNA degradation.

Three sets of five *Limnoria* were dissected to separate their guts, HPs and the rest of the body (RoB). Immediately upon dissection the tissue was placed into a tube containing 100 µl of ice cold TRIzol and stored on ice.

Once all of the required tissue had been collected each sample of tissue was homogenised using a Teflon pestle. Material bound to the pestle was rinsed off using 400 µl of TRIzol.

### *4.2.2.6 RNA Isolation*

The homogenised samples were incubated at room temperature (RT) for 5 minutes before adding 100 µl of chloroform to each sample. The tubes were shaken vigorously for 15 seconds, allowed to stand at RT for 3 minutes before centrifuging them at 12000 x g for 15 minutes at 4 ℃. The upper, colourless, aqueous, RNA containing layers were removed to fresh tubes. The remaining red, organic layers were retained for later DNA and protein extraction. A further 250 µl of chloroform was added to the aqueous fractions and they were shaken and centrifuged as before. The upper, aqueous, RNA containing layers were removed to fresh tubes and the chloroform containing tubes discarded.

The RNA in the aqueous samples was precipitated by the addition of 10 µg of RNAse free glycogen and 250 µl of isopropanol to each sample. The samples were incubated at RT for 10 minutes before pelleting the RNA by centrifugation at 12000 x g for 10 minutes at 4 ℃. The aqueous supernatants were removed and the RNA pellets washed by adding 500 µl of 75 % v/v ethanol to them and briefly vortexing them. The RNA was re-pelleted by centrifugation at 7500 x g for 5 minutes at 4 ℃. The wash was discarded and the pellets air-dried for 5 to 10 minutes ensuring they did not completely dry out. The semi-dry pellets were resuspended in 20 µl of RNAse free water by gentle repeat pipetting. Aliquots of the resuspended RNA samples were used to produce 1 in 10 dilutions and the RNA concentrations of these dilutions were determined. All RNA samples were stored at -80 ℃ until later analysis.

### *4.2.2.7 Removal of DNA from the Organic Phase*

The DNA contained in the organic phase samples was precipitated by adding 150 µl of ethanol to each sample. The tubes were inverted several times, incubated at RT for 3 minutes before pelleting the DNA by centrifugation at 2000 x g for 5 minutes at 4 ℃. The protein containing supernatants were removed to fresh tubes and the DNA pellets discarded.

### 4.2.2.8 Protein Isolation

The protein contained in the organic phase samples was precipitated by adding 750 µl of isopropanol to the samples and allowing them to stand at RT for 10 minutes. The precipitated protein was pelleted by centrifugation at 12000 x g for 10 minutes at 4 ℃. The supernatant was discarded. The protein pellets were washed three times by incubating them with 1 ml of 0.3 M guanidine hydrochloride in 95 % v/v ethanol before centrifuging them at 7500 x g for 5 minutes at 4 ℃ and discarding the wash solution. The pellets were then washed once in the same way using only ethanol as the wash solution. The washed protein pellets were allowed to air-dry partially for 5 – 10 minutes. Finally the protein pellets were resuspended by pipetting them up and down in 50 µl of 8 M urea. The samples were sonicated in a sonic bath for 1 minute before centrifuging them at 10000 x g for 10 minutes to remove any insoluble material. The protein containing supernatants were removed to fresh tubes and stored at -20 ℃ until analysis.

### 4.2.2.9 Production of cDNA

The quality and integrity of the *Limnoria* RNA samples described above was first determined. To do this 5 ng of RNA from each sample were submitted to the Genomics Lab, Technology Facility, University of York to be analysed on a Pico chip using a 2100 Bioanalyzer (Agilent Technologies). All samples were found to contain good quality RNA (Section 4.3.4).

### 4.2.2.10 DNAse Treatment

Aliquots of each RNA sample were DNAse treated to remove any contaminating DNA present. Reactions were set up for each RNA sample containing 8 µl of RNA sample (containing 528 ng of RNA, diluted if necessary), 1 µl RQ1 RNAse-free DNase (Promega), 1 µl RQ1 RNAse-free DNase 10x buffer (Promega), 0.5 µl RiboLock RNAse inhibitor (Fermentas). The reactions were incubated at 37 ℃ for 30 minutes before adding 1 µl of Stop Solution (Promega) to each reaction and heating them to 65 ℃ for 10 minutes.

### 4.2.2.11 First-Strand cDNA Synthesis

For each sample the following was prepared in a 0.2 ml thin walled PCR tube; 1 µl Oligo(dT)12-18 (0.5 µg/µl) primer (Invitrogen Life Technologies), 1 µl dNTP mix (10 mM of each dNTP), 10 µl of DNase treated RNA. The tubes were heated to 65 ℃ for 5 minutes before quickly chilling them on ice and centrifuging them to bring all liquid to the bottom of the tube.

To each sample was then added 4 µl of First-Strand 5x buffer (Invitrogen Life Technologies), 2 µl of 0.1 M DL-dithiothreitol (DTT) and 1 µl RiboLock RNAse inhibitor. The solutions were stirred with a pipette tip then incubated at 42 ℃ for 2 minutes. To initiate first strand synthesis 1 µl of SuperScript® II Reverse Transcriptase (Invitrogen Life Technologies) was added to each sample. The samples were incubated at 42 ℃ for 50 minutes before inactivating the reactions by heating them to 70 ℃ for 15 minutes. All samples were diluted with 80 µl of RNAse free water.

One microliter of the resulting 100 µl cDNA solutions represented 5 ng of RNA starting material.

This process was carried out in duplicate, however with the addition of RNAse free water instead of SuperScript® II Reverse Transcriptase. These duplicates are later referred to as no-RT controls and would show if any contaminating DNA was either still present in the DNase treated RNA starting material, or had been introduced during sample handling.

### 4.2.2.12 Determining Transcript Levels in Limnoria Tissues

The cDNAs produced from *Limnoria* tissue were used to determine the transcript levels of all five *Limnoria* HC genes using the HC 1, 2, 3, 4 and 7 pairs of primers listed in Table 4.1. Amplifications were also carried out with these cDNAs using ubiquitin, GAPDH (mid) and GAPDH (5') primers. Negative controls for each pair of primers were carried out using water instead of cDNA. The GAPDH (mid) primer pair was also used to carry out a negative control by using them in amplifications with no-RT cDNA from each tissue. All amplifications were carried out in triplicate (i.e. three technical replicates) and contained 2 µl of cDNA (equivalent to 10 ng of RNA), 3.2 µl of primers (final concentration of 400 nM per primer), 4.8 µl of RNAse free water and 10 µl of Fast

SYBR® Green Master Mix (Applied Biosystems). The amplifications were carried out as described above.

For each cDNA the $C_t$ values of the technical replicates were averaged to produce a mean value ($mean_{Ct}$). The $mean_{Ct}$ values were adjusted to take account of the efficiency of their respective primer pairs using the formula shown below.

$$(Eff^{Ct})_{target} = Doubling\ Efficiency^{meanCt}$$

Ubiquitin was previously found to be the most stably expressed reference gene out of eight assessed genes using the geNorm, BestKeeper and NormFinder software packages (61). The primer efficiency corrected Ct values, $(Eff^{Ct})_{target}$, were normalised to the ubiquitin primer efficiency corrected Ct value, $(Eff^{Ct})_{reference}$, for the respective cDNA using the formula shown below (198).

$$Ubiquitin\ normalised\ relative\ expression\ level = \frac{(Eff^{Ct})_{reference}}{(Eff^{Ct})_{target}}$$

The mean and standard deviation of the ubiquitin normalised relative expression levels were calculated for each body partition (gut, HP or rest of body).

### *4.2.2.13 Western Blot Analysis of Limnoria Tissues*

The protein isolated from samples of gut, HP and rest of body tissue described above were thawed on ice before vortexing them to thoroughly resuspend the protein. Any protein which could not be resuspended was removed by centrifugation. The amount of protein contained in the samples was quantified as described in Section 2.5.1. Aliquots of each sample were ran on duplicate 10 % acrylamide RunBlue SDS-PAGE pre-cast gels as described in Section 2.5.5. One gel was run with 2.5 µg of protein loaded per lane and was Coomassie stained after running. The other gel was run with 5.0 µg of protein loaded per lane and was Western blotted after running as described in Section 2.5.6 using a 1 in 2500 dilution of anti-HC antibody as a primary antibody.

### 4.2.3 *In-vivo* Peroxide Detection

The presence of peroxides in the gut of *Limnoria* were tested for using horseradish peroxidase (HRP) and the Amplex® Red reagent (Invitrogen Life Technologies). Horseradish peroxidase is able to oxidise Amplex Red using hydrogen peroxide as a co-substrate in a 1:1 stoichiometric ratio to produce the highly fluorescent compound resorufin which has an excitation/emission maxima of 570/585 nm (Figure 4.3). This method is able to detect as low as a 50 nM concentration of hydrogen peroxide and the fluorescent signal produced is stable enough to permit injection and visualisation of the sample in a comfortable time frame (199). Amplex Red is also able to detect other kinds of hydroperoxides such as lipid hydroperoxides (200, 201).

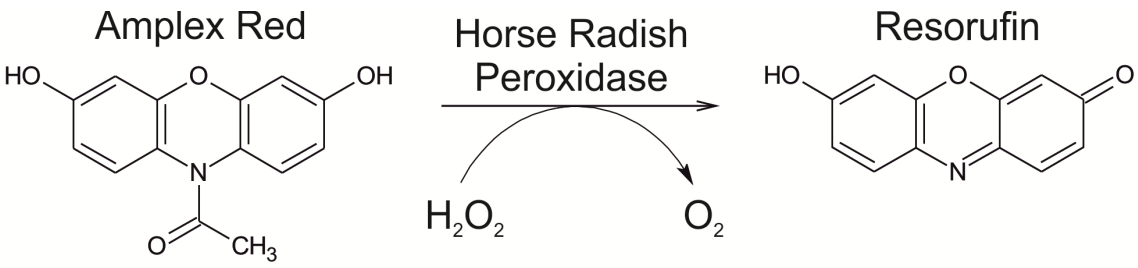



**Figure 4.3. Oxidation of Amplex Red to the highly fluorescent compound resorufin.**

Previous preliminary work carried out by Dr. Leonardo Gomez and Dr. Laura Fass (University of York) and Dr. Simon Cragg (University of Portsmouth) suggests that co-injecting Amplex Red and HRP into the gut of *Limnoria* produced a strong red fluorescence indicating the presence of peroxides (data not shown). Co-injection of a saturated bromophenol blue solution did not affect the production of this signal but did behave as a tracer dye allowing the site of injection to be visualised (data not shown). Inclusion of catalase in the substrate mixture, which should decompose any hydrogen peroxide present to water and oxygen, resulted in the inhibition of the generation of the red fluorescent signal to a large extent (data not shown). This work suggested that hydrogen peroxide or possibly other peroxides were present in the gut of *Limnoria* and established a method for its detection *in-vivo*.

Further work was carried out to determine if the presence of peroxides in the gut of *Limnoria* was linked to a PO enzyme activity. This further work was carried out with the assistance of Dr. Katrin Beßer, Dr. Leonardo Gomez and Dr. Laura Fass (University of

York). To determine if PO activity is involved in the presence of hydrogen peroxide the well-known PO inhibitor *N*-phenylthiourea (PTU) was used (202-204).

The experiment involved placing live *Limnoria* into sea water containing PTU prior to injection with Amplex Red to detect the presence or absence of peroxides. Before this could be done it was therefore necessary to determine a PTU concentration that *Limnoria* could survive in.

Into two wells of a six well culture dish were each placed five live *Limnoria*. One well was filled with seawater and the other with seawater containing 5 mM PTU. Each well contained a match stick of short rotation coppice willow wood which had been pre-soaked in the solution added to that well. After 24 hours all five animals in the seawater containing well were alive and feeding. After four hours in the PTU containing well one animal had died and the remaining four animals were lethargic and did not exhibit normal behaviour; after 24 hours in the solution all five animals were dead. This procedure was repeated using instead 50 µM PTU. After 72 hours one animal from each well had died whilst the other four animals were feeding normally. This concentration of PTU was therefore used for the rest of the experiment.

### *4.2.3.2 Preparation of Solutions*

A 9.7 mM stock solution of Amplex Red (Fisher Scientific) was prepared in DMSO and stored at -20 ℃ until needed. When required a work ing solution was prepared by diluting 20 µl of the Amplex Red stock solution into 1 ml of 100 mM potassium phosphate buffer pH 8.0 resulting in a final concentration of 0.2 mM Amplex Red. To this solution was added 1 µl of HRP (1 U/µl in 50 mM potassium phosphate buffer pH 7.4).

A saturated bromophenol blue solution was prepared by dissolving 10 mg of bromophenol blue sodium salt in 100 µl of 100 mM potassium phosphate buffer pH 8.0. After vigorous vortexing the solution was centrifuged at 13000 x g for 1 minute and the saturated supernatant removed to a new tube.

#### *4.2.3.3 Preparation of Animals*

Approximately 20 live *Limnoria* living on Balau hardwood were collected and 10 animals were placed into each of two wells of a six well culture dish. One well was filled with seawater and the other with 50 µM PTU dissolved in seawater. Into each well was placed a match stick of short rotation coppice willow wood which had been pre-soaked in the solution added to that well. Animals were taken at random for analysis after living under these conditions for ~21 hours.

#### *4.2.3.4 Analysis of Animals*

The following was carried out on randomly selected animals from each of the above treatments.

Animals were placed in a pool of pre-chilled seawater kept cold on ice to make them lethargic and easier to handle. Bromophenol blue and Amplex Red working solution were co-injected ventrally into the animal's gut using a glass capillary needle. Keeping the animals on ice they were then visualised using a MZ FLIII microscope (Leica) under bright field illumination to verify that the tracer dye was present in the correct part of the animal. Ultraviolet illumination and a DsRed filter were used to visualise the resorufin fluorescence. Some animals were dissected post injection to visualise the signal produced specifically by the gut tissue. As a positive control a drop of Amplex Red working solution was placed on a microscope slide. To this drop was added a drop of sea water containing 50 µM PTU and a drop of hydrogen peroxide to determine if either sea water or PTU inhibited the oxidation of Amplex Red by hydrogen peroxide.

### 4.2.4 *Ex-vivo* Phenoloxidase Activity

To determine if a protein present in *Limnoria* tissue was able to carry out a PO activity an *ex-vivo* PO activity assay was carried out.

Forty five *Limnoria* living on Balau hardwood were removed from the wood into a petri dish containing seawater. Once all animals were collected they were transferred to a 1.5 ml microcentrifuge tube and tapped gently to the bottom of the tube. Excess sea water was removed from the animals using a pipette. The tube was then sealed and immersed in liquid nitrogen to snap freeze the animals.

Whilst frozen 50 µl of homogenisation buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 % v/v Tween 20, 1x HALT EDTA Free Protease Inhibitor Cocktail – Thermo Fisher Scientific) were added to the tube. The animals and buffer were ground to a fine powder using a pre-cooled Teflon pestle and the powder allowed to come to room temperature. Once the tubes contents had melted the slurry was ground further to a homogenous suspension. The slurry was incubated on ice for 30 minutes and vortexed occasionally before sedimenting any non-soluble debris by centrifugation at 13000 x g for 5 minutes at 1 ℃. The supernatant was removed to a fresh pre-cooled tube and the volume of supernatant measured.

The amount of protein in the homogenate was determined as described in Section 2.5.1 using 1 µl of the homogenate in triplicate. The homogenate was estimated to contain a minimum of 2.7 µg/µl of protein. A suitable volume of sample loading buffer was added to the remaining homogenate before loading an equal volume into three lanes of an Amersham ECL 4 – 12 % precast native gel. Into a further three lanes was loaded 0.2 µg of mushroom tyrosinase to act as positive control. The gel was run as described in Section 2.5.7.2.

The gel was cut up to yield the three *Limnoria* homogenate and three mushroom tyrosinase lanes as separate strips of gel. A *Limnoria* homogenate and MT lane together constituted a pair. All three pairs of gel strips were subjected to an in-gel di-phenoloxidase (DPO) assay as described in Section 2.6.1. Before doing this, however, two of the pairs were subjected to different pre-treatments.

One pair was pre-incubated for 10 minutes in assay buffer containing 2 mM PTU before carrying out the DPO assay on them. During the assay 2 mM PTU was also included in the assay buffer.

The other pair was subjected to a reduction and an alkylation step to reduce the side chains of any cysteine residues (including those in di-sulphide bonds) and to derivatise them with an alkyl group. This would denature any proteins contained in the gel in a gentle fashion rather than heating the gel. The pair was first reduced by incubating them with 10 mM DTE in 100 mM ammonium bicarbonate solution at 56 ℃ for one hour. The gel strips were then moved to a dish containing 50 mM IAA in 100 mM ammonium bicarbonate solution and incubated at room temperature in the dark for 30 minutes. Finally the gel strips were rinsed in assay buffer for 5 minutes before subjecting them to the assay.

Once the assays were complete the gel strips were scanned to record the results before subjecting them to a further Coomassie staining to visualise where proteins were in the gel in relation to where activity was seen.

# 4.3 Results

## 4.3.1 Conservation of Features Necessary for Phenoloxidase Activity

Hemocyanin proteins are type III copper centre containing proteins, a classification shared also by POs and tyrosinases. Many HC proteins have been shown to be able to carry out a PO enzyme activity after undergoing some form of activation and have been extensively characterised. The three dimensional structures of two of these proteins from arthropod species, along with a PPO also from an arthropod have been elucidated. These structures show that there are several features that are conserved amongst these proteins which are necessary for their ability to perform as POs. If as hypothesised *Limnoria* HCs are also able to behave as POs then given that they are also arthropodan HC proteins it would seem likely that they must also possess these conserved features.

The protein sequences of the five *Limnoria* HC genes were compared against three HC/PPO genes from other arthropods using a MSA (Figure 4.4) and a BLASTp search. The BLASTp search revealed that there was between 94.92 – 60.94% sequence identities between the *Limnoria* HC genes making them highly similar to each other. The *Limnoria* genes showed a minimum sequence identity of 56.21 %, 34.37 % and 28.59 % with the HC/PPO genes of *Panulirus interruptus*, *Limulus polyphemus* and *Manduca sexta* respectively. This indicates a high degree of identity with the HC from *P. interruptus* and moderate identity with the *L. polyphemus* and *M. sexta* genes.

The five *Limnoria* HC sequences were also analysed using the Pfam sequence analysis tool. This analysis predicts that all of the *Limnoria* HC genes will organise into three domains; a HC all-alpha domain, a HC copper containing domain and a HC Ig like domain.

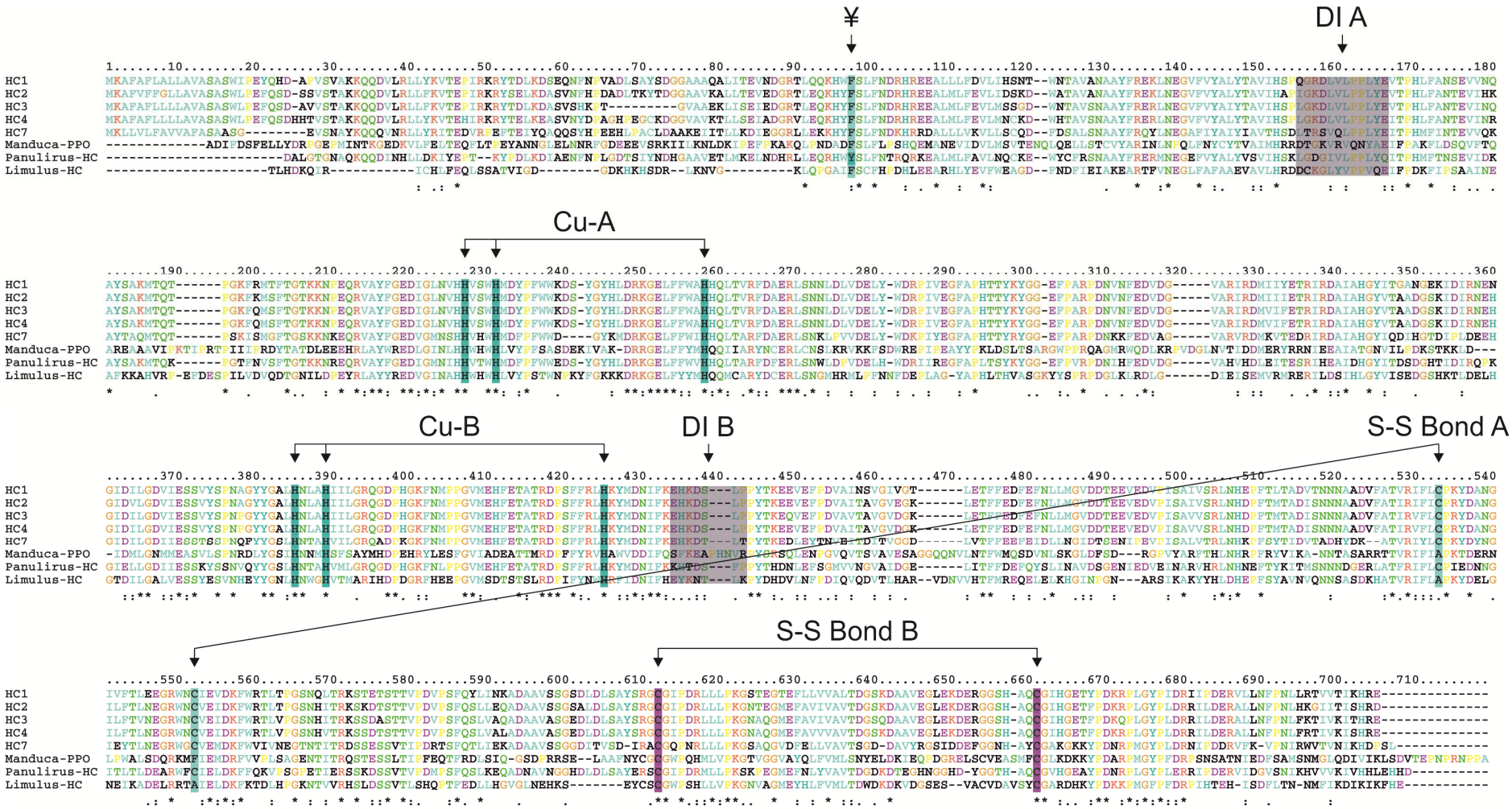

¥
DI A
Cu-A
Cu-B
DI B
S-S Bond A
S-S Bond B
HC1
HC2
HC3
HC4
HC7
Manduca-PPO
Panulirus-HC
Limulus-HC

**Figure 4.4. Multiple sequence alignment of *Limnoria* HC sequences and reference protein sequences.**
A multiple sequence alignment of the *Limnoria* HC 1, 2, 3, 4 and 7 protein sequences with those of a prophenoloxidase from *Manduca sexta* (Manduca-PPO), a HC from *Panulirus interruptus* (Panulirus-HC) and a HC from *Limulus polyphemus* (Limulus-HC). Features conserved between the sequences are highlighted: ¥, active site blocking aromatic residue; DI-A and DI-B, domain interface A and B; Cu-A and Cu-B, copper binding site A and B; S-S Bond A and B, disulphide bonds A and B.

All three comparison genes are composed of a single amino acid chain annotated as consisting of three well defined domains. At the amino terminus they all possess a HC N-terminal or all-alpha domain, followed by a HC copper containing domain and finally a HC Ig-like domain. The multiple sequence alignment shows that the five *Limnoria* HC genes are approximately the same length as the three reference genes and align well with them over most of their length. The *Limnoria* genes possess extra sequence at their N-terminus not seen in the other genes; the first 15-18 amino acids of this sequence are accounted for by their signal peptides whilst approximately a further 15 amino acids show poor sequence identity with all of the reference genes. Poor sequence matching is seen up to approximately position 100 in the MSA sequence after which much more sequence conservation is seen.

The MSA shows that the two transition points between the three domains (domain interfaces) are in the same regions of each protein sequence and indeed the five *Limnoria* genes show good sequence conservation with at least one of the reference sequences in these regions suggesting that they would also have a domain interface in these regions (DI-A and DI-B in Figure 4.4). This is also in agreement with the Pfam analysis.

### *4.3.1.2 Copper Binding Sites*

Proteins such as HCs contain a type III copper centre which is composed of two copper binding sites. Within each binding site a single copper atom is coordinated by three histidine residues. The 3D structures of the three reference genes show that they do indeed contain such type III copper centres. The MSA shows that the copper binding histidine residues and the spacing between these residues is conserved in all three reference genes and also in all five *Limnoria* genes (Figure 4.4).

### 4.3.1.3 Substrate Occluding Residue

Between the copper atoms of a type III copper centre is bound a single molecule of dioxygen as a peroxide ligand. In their inactive forms HCs and PPOs prevent access to this oxygen molecule by occluding it from the solvent by blocking the entrance to the copper centre using an aromatic amino acid (phenylalanine in the case of *L. polyphemus*) which is part of the NTD (Figure 4.5 A and B). After activation it is thought that the NTD is removed or displaced either by proteolysis removing the domain or mild denaturation resulting in movement of the NTD out of its blocking position. This removes the blocking aromatic residue from the entrance to the copper centre and reveals a solvent accessible channel directly to the bound oxygen molecule which would allow substrate access (Figure 4.5 – C). This aromatic residue is conserved in all reference genes and also in all five *Limnoria* genes (Figure 4.4).

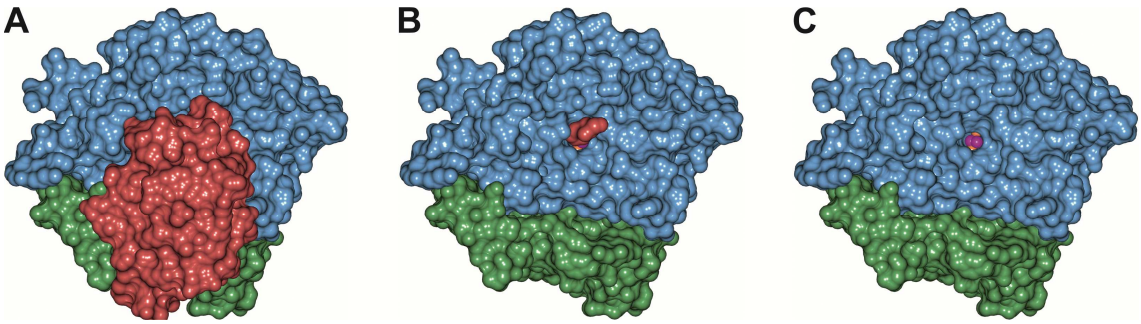



**Figure 4.5. Three-dimensional structure of *L. polyphemus* HC.**
A) The structure of one monomer of *L. polyphemus* HC is shown: red, N-terminal domain; blue, copper containing domain; green, Ig-like domain. B) The N-terminal domain has been removed from the structure with the exception of residue F49. C) The N-terminal domain is completely removed from the structure showing the solvent accessible channel to the copper and oxygen containing active site. Copper atoms are shown in orange, oxygen in purple. Figure generated using CCP4MG version 2.7.3 (University of York).

### 4.3.1.4 Disulphide Bonds

Disulphide bonds usually contribute to the stability and tertiary fold of proteins. The 3D structure of the *P. interruptus* HC shows that it contains two disulphide bonds in its C-terminal Ig-like domain whilst the *M.sexta* and *L. polyphemus* structures only share one of these bonds. The MSA shows that the cysteine residues that contribute to both of these disulphide bonds are conserved in all *Limnoria* HC genes (Figure 4.4).

### 4.3.2 qPCR Primer Testing

To allow the transcript abundance of the *Limnoria* HC 2 and HC 3 genes to be quantified three pairs of qPCR primers were designed for each gene. The primer pairs were used in amplifications to test that they produced a single amplicon of the correct length effectively. All six primer pairs resulted in strong amplification from *Limnoria* cDNA of amplicons of the expected length (Figure 4.6).

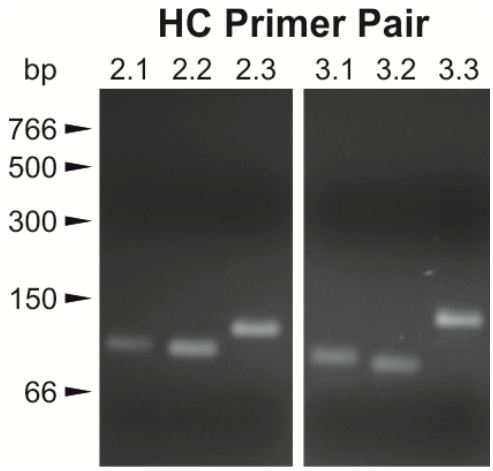



**Figure 4.6. Analysis of qPCR primer test reaction products.**

Primer pairs HC 2.2 and HC 3.3 were chosen to take forward for the quantification of the transcript levels of HC 2 and 3. As all primer pairs were capable of producing strong amplification of a single product these pairs were chosen for the following reasons: both pairs target a region of their target genes that are a similar distance from the 3' end of those genes and both pairs produced very similar $C_t$ values (19.9 and 20.86 respectively, n = 3). The DNA sequences of the amplicons produced using these chosen primer pairs, as well as those previously designed for HC 1, 4 and 7, were obtained and it was verified that the amplicons produced were those expected.

### 4.3.3 qPCR Doubling Efficiency

The qPCR doubling efficiencies were determined for each set of primers to be used for quantifying transcript levels in *Limnoria* tissues. These efficiency values are shown in Table 4.3.

**Table 4.3. Calculated doubling efficiencies of qPCR primer pairs.**

| Primer Pair | Doubling Efficiency |
|---|---|
| HC 1.3 | 1.93 |
| HC 2.2 | 1.99 |
| HC 3.3 | 1.96 |
| HC 4.3 | 1.98 |
| HC 7.1 | 1.98 |
| Ubiquitin | 1.98 |
| GAPDH 5’ | 2.00 |
| GAPDH (mid) | 2.00 |

### 4.3.4 Transcript Levels in *Limnoria* Tissues

The RNA from the HP, gut and rest of the body (RoB) was extracted from three groups of five animals and reverse transcribed to produce cDNA. The quality and integrity of the RNA was determined prior to cDNA production using a Bioanalyzer instrument (Figure 4.7). This analysis showed the RNA to be of good quality and not degraded as well resolved bands are visible in all lanes of the simulated agarose gel shown in Figure 4.7 rather than a smear of signal which would indicate degradation.

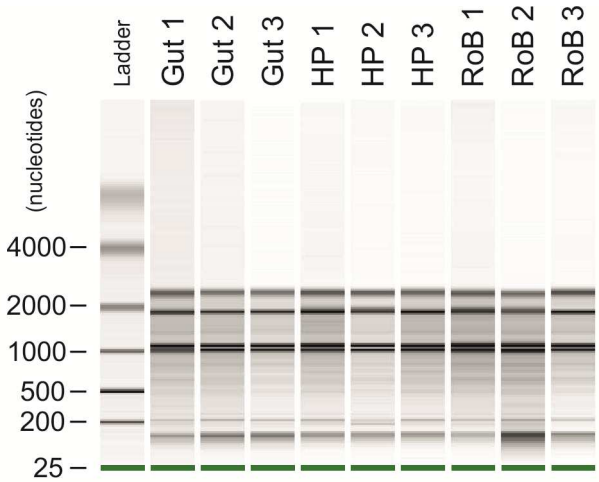



**Figure 4.7. Integrity of RNA used for cDNA production.**
An analysis of the integrity of the extracted *Limnoria* RNA to be used for cDNA production by a Bioanalyzer 2100 instrument.

The transcript levels of HC 1, 2, 3, 4 and 7 in each tissue type from each group of animals were determined. Transcript levels of ubiquitin were also determined as well as a mid-gene and 5’ fragment of the GAPDH gene. All values were corrected using their

pre-obtained primer efficiency values and normalised using the expression level of ubiquitin.

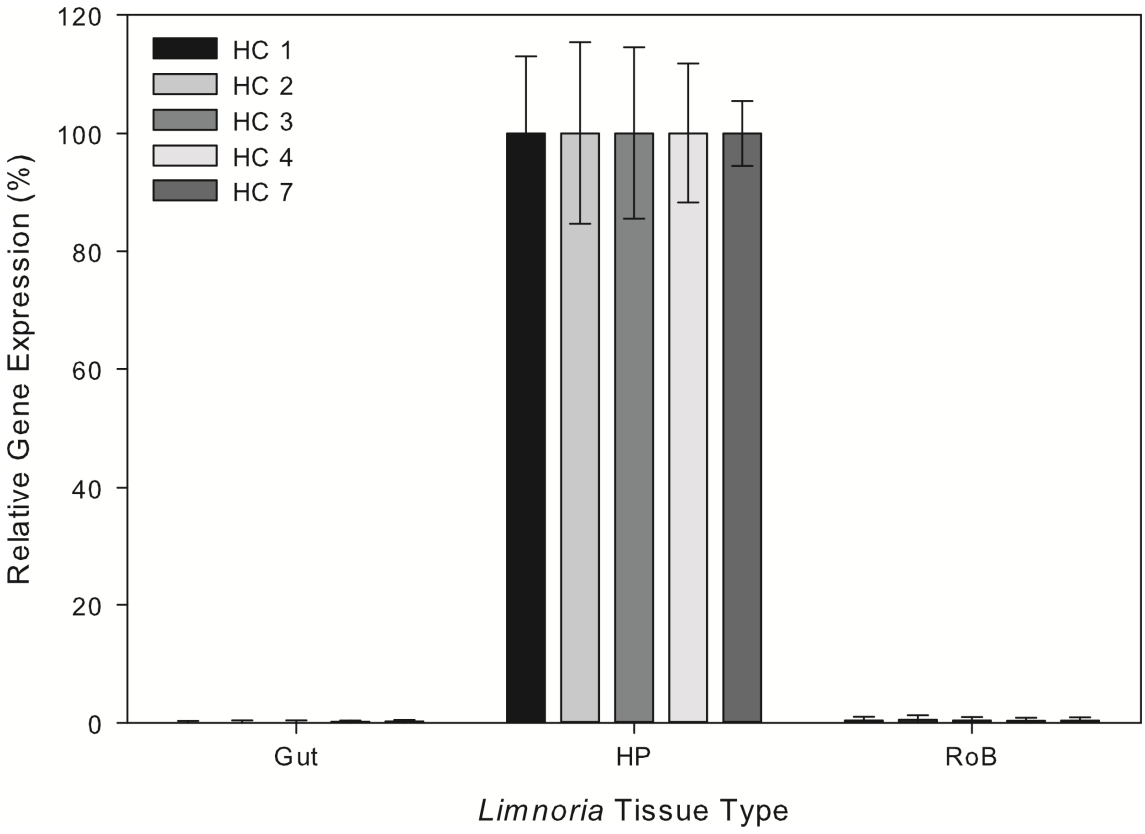



**Figure 4.8. Relative HC gene expression levels in various HC tissue types.**
The corrected and ubiquitin normalised expression levels of the *Limnoria* HC genes 1, 2, 3, 4 and 7 in Gut, HP and rest of body (RoB) tissues. Values for each gene were made relative to each other by allowing the highest expression value for each gene represent 100 %. Error bars represent the standard deviation.

The HC genes were all found to be exclusively expressed in HP tissue (Figure 4.8).

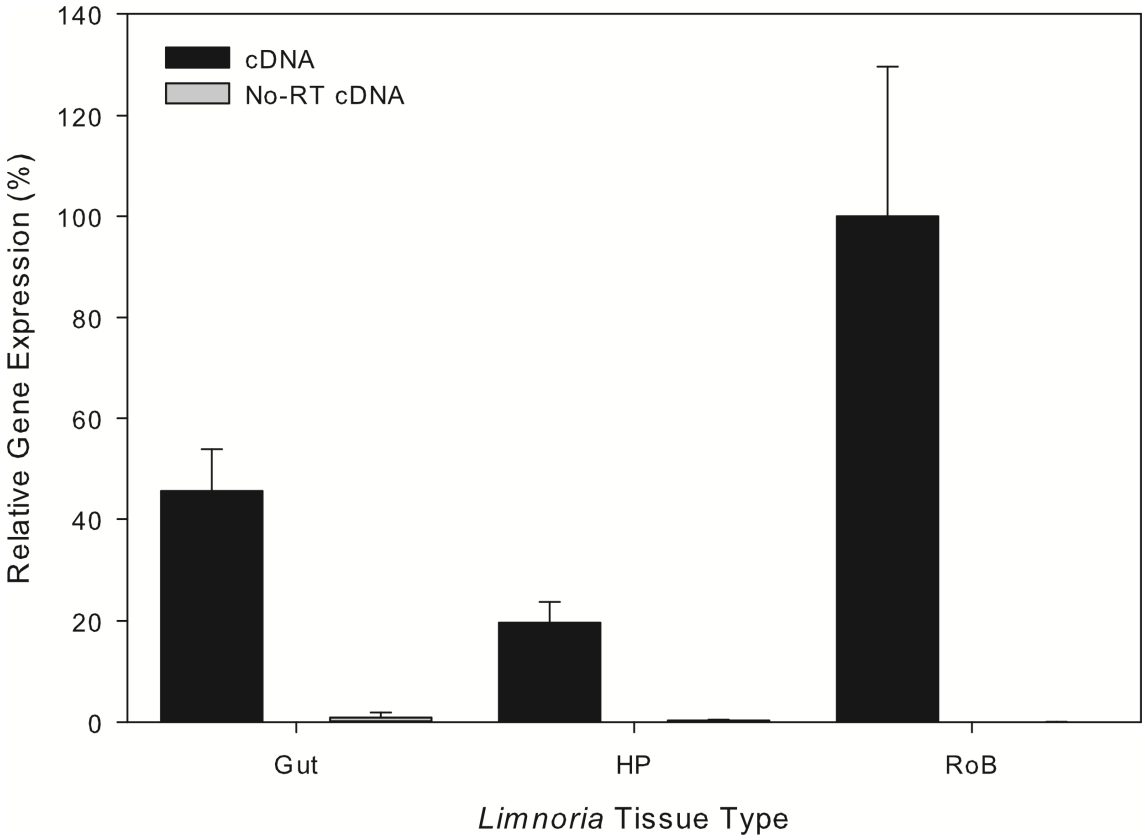



**Figure 4.9. No reverse transcriptase (RT) control.**
The corrected and ubiquitin normalised expression levels of GAPDH were measured in Gut, HP and rest of body (RoB) tissues using the GAPDH (mid) primer pair. Expression levels were determined using cDNA and reverse transcription reactions produced without reverse transcriptase enzyme. Values for each tissue were made relative to each other by allowing the highest expression value represent 100 %. Error bars represent the standard deviation.

Negative controls were carried out for each primer pair using water in the reaction instead of cDNA. No amplification was seen in any of these controls. Another form of negative control was carried out in which reverse transcription reactions were performed without the inclusion of reverse transcriptase enzyme; this should mean that the samples contain no cDNA. These controls, called no-RT controls, were amplified with GAPDH (mid) primers. Once the data from these controls was corrected and normalised it is apparent that in comparison to similar amplifications using cDNA there was effectively no amplification from the no-RT controls confirming that there was no contaminating genomic or other source of DNA in the samples (Figure 4.9).

An assessment of the quality of cDNA produced during the reverse transcription step and also of the integrity (completeness) of the starting mRNAs was also carried out. Following the recommendations of the Minimum Information for Publication of

Quantitative Real-Time PCR Experiments (MIQE) guidelines the 3':5' ratio of expression levels for the GAPDH gene were calculated for each tissue and group of animals (Table 4.4) (205). This was done using the GAPDH mid primers to represent the 3' region of the gene, and GAPDH 5' primers to represent the 5' end.

**Table 4.4. Ratio of 3':5' expression of the GAPDH gene.**
Corrected and ubiquitin normalised expression values of the GAPDH gene were measured using the GAPDH (mid) and GAPDH (5') primer pairs and the ratio between them calculated.

| **Tissue Type** | **3':5' Ratio** | | |
|---|---|---|---|
| | **Group of Animals** | | |
| | **1** | **2** | **3** |
| Gut | 322.3 | 235.8 | 515.9 |
| HP | 264.1 | 177.6 | 233.7 |
| Rest of Body | 275.3 | 178.2 | 474.8 |

Once the expression values for these two primers pairs had been corrected and normalised 3':5' ratios were calculated (Table 4.4). Despite the ratio values being far from optimal the cDNAs were judged to be of sufficient quality to support the findings of this experiment. The full implications of these ratio values are discussed further in Section 4.4.

### 4.3.5 Spatial Distribution of Hemocyanin Proteins

Protein was also extracted from the same samples of HP, gut and rest of the body (RoB) described above resulting in three samples of protein from each tissue, each sample from a group of five animals.

Aliquots of the protein were ran in duplicate on SDS-PAGE gels with one gel being Coomassie stained and the other Western blotted using the anti-HC antibody described in Section 2.6.3. Despite loading the same amount of protein per lane the Coomassie stained gel shows more staining in the RoB lanes (Figure 4.10 – A). The Western blot shows a strong signal indicating the presence of HC protein in all lanes with the signal being particularly strong in the RoB lanes (Figure 4.10 – B). This shows that HC protein is present in all of the tissue types of *Limnoria* analysed.

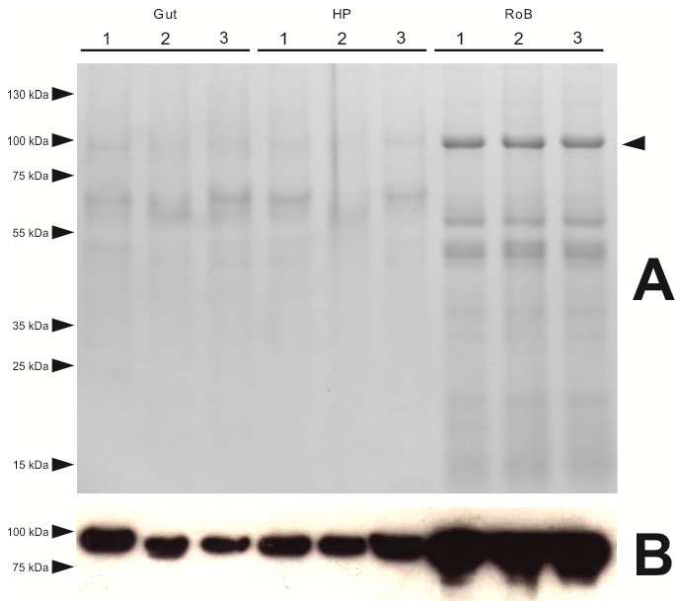



**Figure 4.10. Spatial distribution of hemocyanin proteins.**
Triplicate protein extracts were produced from dissected gut, HP and rest of body (RoB) tissue samples of *Limnoria*. Aliquots of each extract were ran on duplicate SDS-PAGE gels before being A) Coomassie stained or B) Western blotted using an anti-HC primary antibody. Expected position of HC band is indicated in A with an arrow.

### 4.3.6 Peroxide Production in the Hindgut of *Limnoria*

The results detailed below for this section of work should only be considered preliminary and more study is required to confirm and understand the findings.

Live *Limnoria* were assayed for the presence of peroxide in their hindguts, using Amplex Red as a detection agent, after spending ~21 hours living in sea water containing the PO inhibitor PTU. As a positive control the same assay was carried out on animals which had not been exposed to PTU. Non-injected animals which had not been exposed to PTU were visualised in the same way as injected animals acting as a negative control to confirm that neither sea water nor the animals themselves auto-fluoresced.

Positive control animals not exposed to PTU showed a very strong fluorescence in their hindgut after injection (example shown in Figure 4.11). The bromophenol blue tracer dye confirmed that the injections were into the hindgut of the animals but also that the injected liquid rapidly diffused throughout the animal possibly as a result of damage caused to the animal by the injection. A gut was dissected from an injected animal

which showed strong fluorescence (Figure 4.11). Other un-identifiable tissues still attached to the gut also fluoresced.

Animals exposed to PTU prior to injection showed greatly reduced fluorescence in their hindguts, either when intact inside an animal of when dissected out (example shown in Figure 4.11).

Non-injected negative control animals which had not been exposed to PTU showed no auto-fluorescence. An example of such an animal is shown in Figure 4.11.

Mixing Amplex Red working solution with sea water and hydrogen peroxide in a drop on a microscope slide resulted in the drop turning a red/pink colour and fluoresced highly when placed under UV light. This confirmed that sea water did not have an inhibitory effect on the oxidation of Amplex Red by hydrogen peroxide (data not shown).

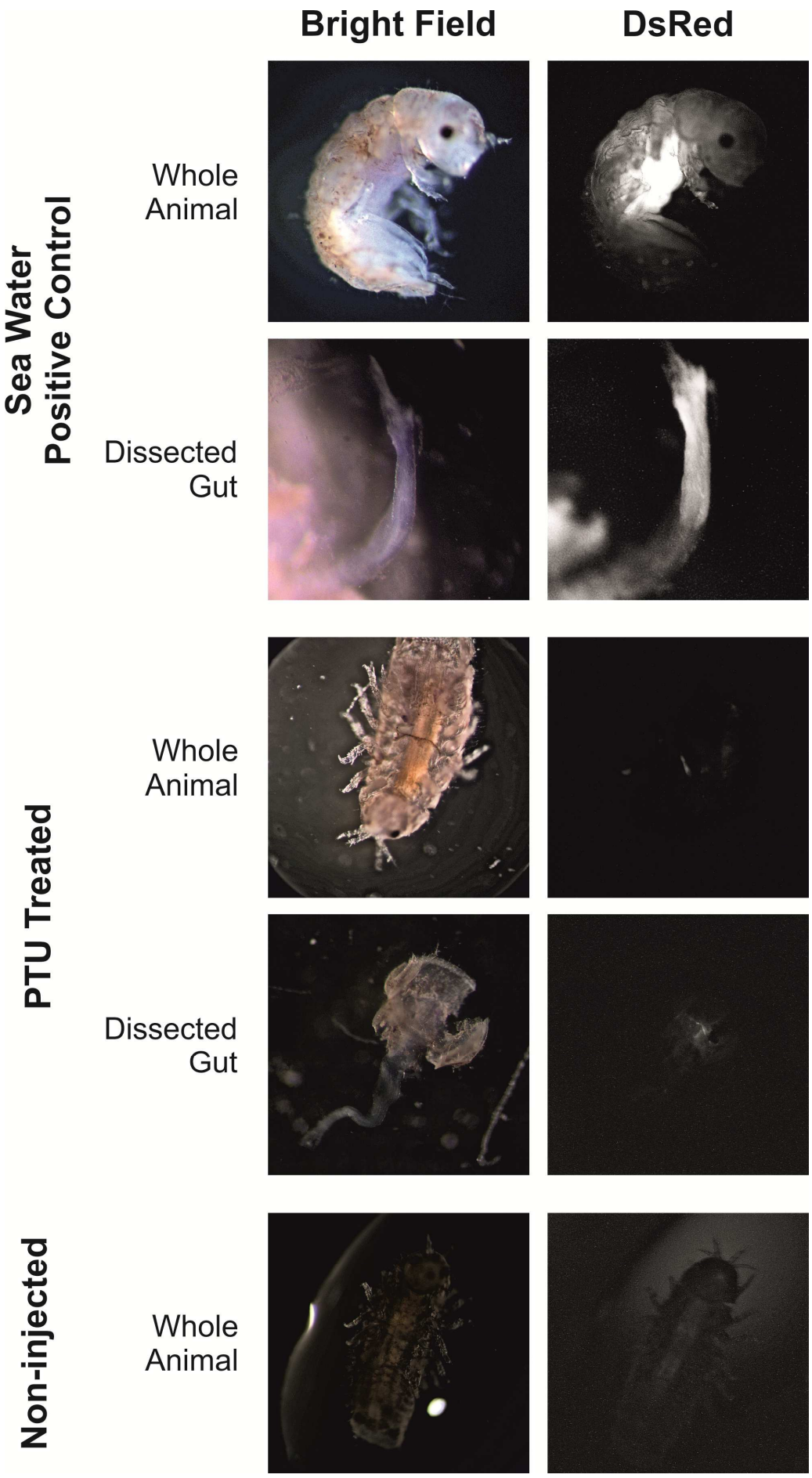

Bright Field
DsRed
Sea Water Positive Control
Whole Animal
Dissected Gut
PTU Treated
Whole Animal
Dissected Gut
Non-injected
Whole Animal

**Figure 4.11. Detection of peroxide in the gut of *Limnoria*.**
Example micrographs are shown demonstrating the presence or absence of peroxide in the region of the gut of intact *Limnoria* or in dissected gut tissue. Micrographs were either obtained under bright field illumination, or under UV illumination and any resorufin fluorescence visualised using a DsRed filter. White colouration in the DsRed obtained micrographs indicates the presence of resorufin and hence peroxide. The brightness and contrast of the PTU treated and non-injected DsRed micrographs were highly modified to allow faithful reproduction during printing.

### 4.3.7 Phenoloxidase Activity in Whole *Limnoria* Extract

A soluble protein extract was produced from whole *Limnoria* animals in order to determine if a protein in that extract was capable of carrying out a DPO enzyme activity. The extract was run on a native gel and the gel stained for DPO activity (Figure 4.12 A – C). A positive control enzyme, mushroom tyrosinase, was also treated in the same way (Figure 4.12 G – I). To confirm that any activity seen was caused by a protein capable of DPO activity several treatments were tested on the gel strips including a DPO activity inhibitor, PTU, as well as a chemical protein denaturation process. The gel strips where then subsequently counter-stained with Coomassie (Figure 4.12 D – F for *Limnoria* extract, J – L for mushroom tyrosinase). This counter-staining procedure was intended merely to show the pattern of proteins present in the gel strips, not the protein loading. The amount of protein loaded per strip is equal for each protein source.

When subjected to the standard DPO assay *Limnoria* extract does produce staining indicating DPO activity (Figure 4.12, A), as does the positive control enzyme (Figure 4.12, G). *Limnoria* extract produces five distinct bands of staining with smearing of staining between them. Subsequent Coomassie staining of the gel strips does not reveal any extra distinct bands (Figure 4.12, D). The major bands, labelled 1 – 5 in Figure 4.12, were excised and the proteins within them determined. The proteins identified are detailed in Table 4.5. According to the protein identification data HC protein was only found in bands 2 and 5.

**Table 4.5. Proteins identified in gel bands displaying DPO activity.**
Soluble *Limnoria* extract was used in a PAGE based DPO assay. Protein bands displaying DPO activity were subjected to protein identification, searching data against the NCBI database and the results are shown. LRR; leucine rich repeat motif containing protein of unknown function. A * indicates a single peptide match.

| Band No. | Proteins Identified | Species |
|---|---|---|
| 1 | Actin | *Penaeus monodon* |
| 2 | Hemocyanin 1 | *Limnoria quadripunctata* |
| | Hemocyanin 3 | *Limnoria quadripunctata* |
| | Hemocyanin 4 | *Limnoria quadripunctata* |
| 3 | Actin | *Penaeus monodon* |
| | LRR 2 | *Limnoria quadripunctata* |
| | LRR 3 | *Limnoria quadripunctata* |
| | LRR 10 * | *Limnoria quadripunctata* |
| | Arginine Kinase * | *Callinectes sapidus* |
| 4 | LRR 1 | *Limnoria quadripunctata* |
| | LRR 2 | *Limnoria quadripunctata* |
| | LRR 10 * | *Limnoria quadripunctata* |
| | Beta Actin * | *Pacifastacus leniusculus* |
| | Actin * | *Penaeus monodon* |
| 5 | LRR 1 | *Limnoria quadripunctata* |
| | LRR 4 | *Limnoria quadripunctata* |
| | GA16177-PA | *Drosophila pseudoobscura* |
| | Hemocyanin * | *Limnoria quadripunctata* |

When the gel strips were pre-incubated with the DPO inhibitor PTU the amount of DPO staining produced by *Limnoria* extract is greatly reduced, although not entirely removed (Figure 4.12, B). Coomassie staining confirms that the same distribution of proteins is present as is seen when only DPO staining is performed (Figure 4.12, E). This means that the reduction of signal seen is due to the inhibition of a DPO activity, not because of the loss or diffusion of protein. The signal produced at the interface between the stacker and resolving gels (above band 1 in Figure 4.12, B) seems to be unaffected by the inhibitor, still producing a dark brown band. The signal produced by the positive control enzyme is completely abolished by the presence of PTU (Figure 4.12, H).

Chemically denaturing the proteins within the gel reduced the amount of staining further (Figure 4.12, C). Coomassie staining the gel strip demonstrates that the denaturation treatment has not removed or greatly affected the distribution of proteins

within the gel (Figure 4.12, F). This means that the signal seen with DPO staining alone is likely caused by a folded protein. The staining at the stacker/resolving gel interface again persists and seems to be unaffected by the treatment. The positive control enzyme was effectively deactivated using this denaturation treatment (Figure 4.12, L).

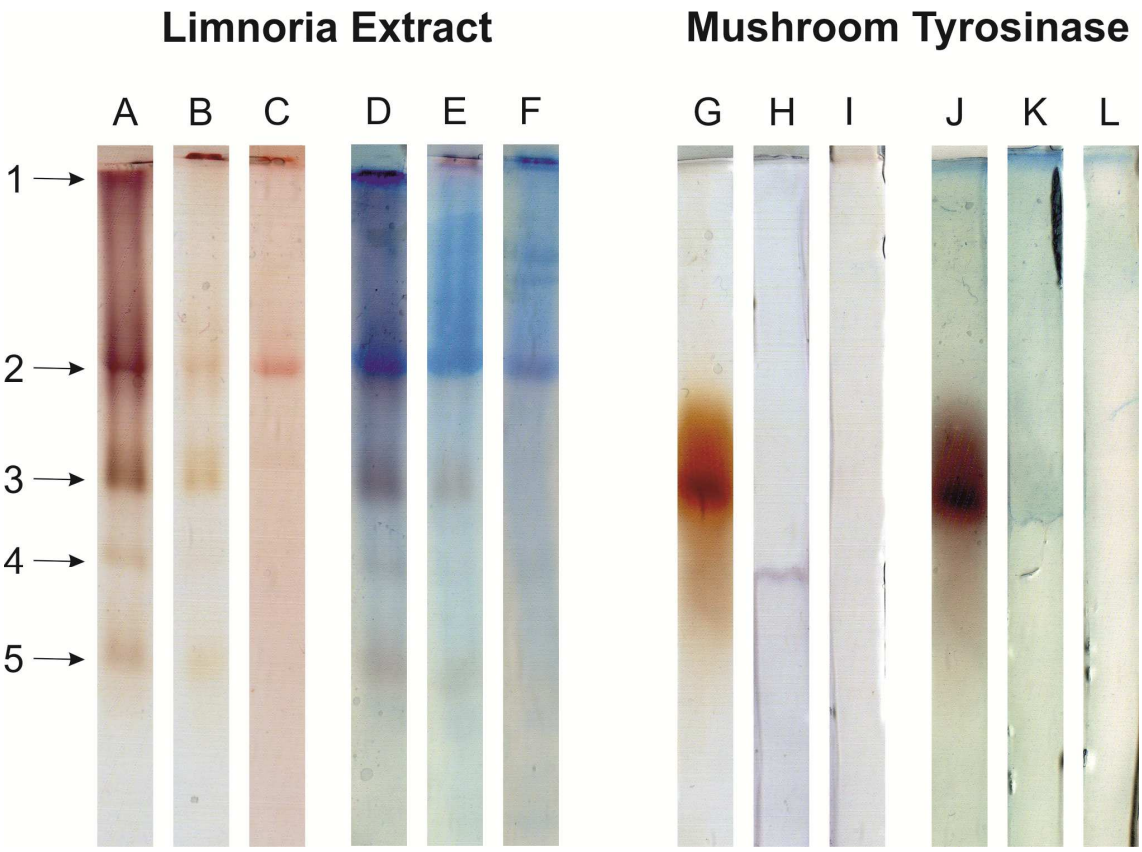



**Figure 4.12. Di-phenoloxidase activity stained native PAGE strips.**
Aliquots of soluble *Limnoria* extract and positive control mushroom tyrosinase enzyme were separated by native PAGE before subjecting them to different treatments and testing them for DPO activity (A – C and G – I respectively). The gel strips were subsequently Coomassie counter-stained to visualise the distribution of proteins within them (D – F and J – L respectively). Regions of gel positive for DPO activity are stained brown/red. A, D, G and J: No treatment. B, E, H and K: pre-incubated with 2 mM PTU. C, F, I and L: chemically reduced and cysteine side chains alkylated. Bands 1 – 5 indicated were excised from gel strip A and subjected to protein identification.

## 4.4 Discussion

Many arthropodan HCs possess an ability to carry out a PO enzyme activity after undergoing some sort of activating treatment (137, 170). A multiple sequence alignment of the five *Limnoria* HC protein sequences along with those of three arthropodan proteins for which 3D structures are available was produced. The *Limnoria* HC sequences were also analysed using the Pfam protein family prediction software. This revealed that all of the features believed to be necessary for a HC to demonstrate activatable PO activity were conserved. The *Limnoria* HCs are all predicted to organise into the classic arthropodan three domain structure (PFAM) and these domains align well with the domains of the reference proteins. All five *Limnoria* HCs show conservation of the two copper binding sites needed for a type III copper centre, as well as the substrate occluding residue believed to prevent activity when not exposed to an activator. Finally the *Limnoria* HCs show potential conservation of two disulphide bonds in their Ig-like domains suggesting a preservation of tertiary structure within this domain in comparison to the previously studied proteins.

This evidence shows that the main features of these protein families are well conserved across various phylogenetic classes and suggests that like their more thoroughly characterised counterparts they possess the potential to be activated to become POs.

Reverse transcription quantitative PCR (qPCR) and Western blotting techniques were used to determine in which tissues in *Limnoria* the HC genes were transcribed and in which tissues HC protein was present. Three groups of five animals were dissected into the three tissue groups; gut, HP and rest of body (RoB). Protein and RNA were extracted from these tissue samples for analysis. The quality of the extracted RNA was assessed and found to be of good quality before it was used to produce cDNA in a reverse transcription reaction. A set of no-RT controls were also produced by subjecting each sample of RNA to a reverse transcription reaction containing no reverse transcriptase.

Primers were successfully designed and evaluated to specifically quantify the transcript levels of HC 2 and 3 from *Limnoria* cDNA. The amplification products produced by the primers to be used for qPCR of each of the five *Limnoria* HC genes were sequenced and confirmed that each set of primers correctly amplified their correct target sequence

and only from their intended HC gene. The qPCR doubling efficiencies were determined for primers targeting HC 1, 2, 3, 4, 7, ubiquitin, a 5' fragment of GAPDH and a mid-gene fragment of GAPDH.

During the experiment to determine the relative transcript levels of the HCs, the data of which is described below, several controls and further quality assessments were carried out. Negative control amplifications were carried out for all primer pairs assessed using water in the amplification reactions in place of cDNA. All of these reactions showed no amplification demonstrating that no contaminating DNA was introduced during the experiment. A further negative control was employed; GAPDH mid primers were used in amplifications of the no-RT cDNAs from each tissue type and replicate. Once corrected and normalised the relative expression of these reactions were compared to their positive cDNA counterparts. This comparison showed that without reverse transcriptase included in the reverse transcription reaction no cDNA was produced. It also shows that no detectable contaminating genomic DNA was present in any of the reverse transcription reactions.

An assessment of the quality of cDNA produced was carried out as recommended by the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines. This was done by carrying out amplifications on the cDNA from each tissue type and replicate using primers which target a 5' (GAPDH 5') and 3' (GAPDH mid) region of the GAPDH gene, and creating a ratio of the resulting corrected normalised relative expression levels.

A ratio value of 1 indicates that the starting material mRNA was of good integrity and that the majority of the mRNAs were more than half complete (206). It is also an indication that the cDNA synthesis was efficient and resulted in the near complete reverse transcription of the available mRNAs. A ratio of greater than 5 suggests that either the mRNAs had degraded or that the cDNA production process had been inefficient/incomplete as the 5' portions of mRNAs had not been reverse transcribed into cDNA (206). The values obtained in this work are substantially higher than 5 (Table 4.4). This could indicate that the RNA starting material in this experiment was degraded however this seems unlikely as the Bioanalyzer analysis indicated that the RNA was not degraded and was of good quality. Another interpretation of this finding is that the cDNA synthesis was not complete or was inefficient. This could be due to the reverse transcription reagents being too old and not working optimally. Whatever the reason, the poorness of these ratios is unlikely to have affected the data described

above for several reasons. Firstly, the primers used for qPCR analysis were all designed to sections of their target genes which were closer to the 3' ends of those genes than the GAPDH 5' primers were (Table 4.2). Secondly, the ubiquitin primers showed good and comparable amplification in all tissue types and replicates, and these primers targeted a region of the ubiquitin gene which was a similar distance from the 3' end as the other primer pairs. Finally, the 3':5' ratios obtained are approximately similar to each other (within a 2.9 fold range). This suggests that any deficiency in the cDNA production process affected all tissue types and animal groups in approximately the same way but is unlikely to have influenced the data described below.

The relative transcript levels for HC 1, 2, 3, 4, 7 were determined with the data corrected using the obtained qPCR doubling efficiencies and normalised using the expression level of ubiquitin in each respective tissue sample. This demonstrated that in *Limnoria* all five HC genes are exclusively expressed in the HP organ. Western blot analysis using anti HC antibodies carried out on protein extracted from the same tissue samples used for qPCR reveals that HC protein is present in gut, HP and RoB tissues.

An interpretation of this data is that as HCs are known to function as oxygen carriers HC protein may be found in all parts of the body as all parts of the body would require oxygen. This work is unable to distinguish if the HC protein present in the gut is present intracellularly or in the lumen of the gut where it may be taking part in digestion.

However, two pieces of information suggest that the HC protein detected in the gut of *Limnoria* is present in the organ's lumen. Firstly, immunofluorescent confocal microscopy work carried out by Dr. Clare Steele-King (University of York) indicated that although HC protein was present in cells of the HP, it was not present in the cells of the gut (unpublished data). Secondly, further Western blot analysis by Dr. Katrin Beßer demonstrated that HC was present in the luminal fluid of the gut (unpublished data). The luminal fluid was obtained by gently centrifuging dissected guts to squeeze out their fluid content.

Therefore summarising this data, given that HC genes are solely transcribed in the HP and HC protein is found in the cells of the HP, and in the gut lumen this data supports the hypothesis that as a protein involved in digestion, HCs would be transcribed and translated in the HP before being transported through the luminal spaces to the gut to take part in digestion.

To investigate if there was a protein present in *Limnoria* capable of carrying out a PO reaction, and to determine if this protein was HC the following experiment was carried out. A soluble extract of whole *Limnoria* as well as a positive control tyrosinase enzyme were subjected to a PAGE based DPO activity assay in triplicate. One replicate was subjected to the assay without further treatment. Another replicate was incubated with PTU, a PO inhibitor, prior to being assayed. The final replicate was subjected to a chemical protein denaturation procedure which would break disulphide bonds and alkylate cysteine residues prior to the DPO assay.

The un-treated replicate produced five activity bands (visible as brown staining on the gel). Pre-treatment with PTU greatly reduced the intensity of all five bands showing partial inhibition of a DPO activity. The reason that this inhibition was incomplete may be due to either of two reasons. Firstly, an insufficient excess of inhibitor may have been used for the quantity of protein present in the assay. Alternatively, it may be that PTU is not an ideal inhibitor for this particular protein. Given that the activity assay measures DPO activity and the method of inhibition is believed to be competitive, PTU may not be the best inhibitor to use. Kojic acid, another known PO inhibitor appears to be a closer molecular match to a diphenol and may prove a more effective inhibitor. The replicate that underwent chemical protein denaturation showed an even greater reduction in DPO activity. Without treatment the positive control enzyme showed very strong activity staining, whilst both treatments resulted in complete abolition of enzyme activity.

All replicates were subsequently Coomassie stained and it was found that there were no other distinct major protein bands visible that had not also shown DPO activity.

These results show that a folded protein, probably containing disulphide bonds and possessing a DPO activity is responsible for the production of the activity staining seen in the *Limnoria* extract. In order to implicate HC as the protein responsible the five activity bands seen in the untreated *Limnoria* extract were excised and the proteins within them identified. From these five bands two were found in which at least one peptide was identified as belonging to the HCs. However, although HC protein was not detected in any of the other three bands this does not mean it was not present. It is likely that there are many proteins present in each of the bands analysed and the method of protein identification used in this experiment is not ideally suited to identifying all proteins present, only the most abundant. Therefore HC may have been present in the other three bands but in small quantities that did not permit identification.

Equally, due to the limitations of this technique it is possible to conclude that HC may not be the protein responsible for the DPO activity seen in these bands. Possibly a better method to be used would be to Western blot the DPO activity gels using anti-HC antibodies.

It is worth noting that the LRR proteins discussed briefly in Chapter 3 were identified in three of the activity bands analysed. There is no direct evidence to suggest that these proteins would be responsible for the DPO activity seen, however given their potential to bind to other proteins it is interesting to find them possibly associating with the protein responsible for the DPO activity seen. It is of course possible that the detection of the LRR proteins in these three bands is merely also due to the limitation of the identification technique discussed above.

The following case to justify the possible implication of HC as the protein responsible for the DPO activity seen in this experiment. Firstly, although HC is not identified as being present in every band analysed, neither was any other protein. This means that there is not an obvious identified candidate protein being ignored. Secondly, HC is the only protein detected in band 2 (Figure 4.12 – A), the most intensely stained activity band. This suggests that where the most DPO activity is seen HC is the most abundant protein. Finally, the literature discussed in this chapter shows that other highly related HCs are capable of DPO activity when assayed as in this experiment, and that PTU is able to inhibit this activity. As a protein in this experiment fits this data seen by other researchers studying HCs and as HC is identified in two of the bands it is reasonable to propose that HC may be responsible for carrying out a DPO activity in *Limnoria* extract.

Previous work has suggested that peroxide species may be present in the gut of *Limnoria* (Dr. Leonardo Gomez, Dr. Laura Fass (University of York) and Dr. Simon Cragg (University of Portsmouth), unpublished data). Building upon this work an experiment was carried out to investigate whether exposing the animals to a PO inhibitor would remove this peroxide from the gut.

Non-treated control animals recreated the finding mentioned above. They showed a strong fluorescence after they were injected with the peroxide detecting reagent, Amplex Red, indicating the presence of a peroxide. It was found that exposing the animals to the PO inhibitor, PTU, before injecting them with Amplex Red reagent resulted in a greatly reduced amount of fluorescence indicating very little to no peroxide being present. Non-injected animals that had not been exposed to PTU showed no

fluorescence confirming that it was oxidised Amplex Red which caused the fluorescence rather than the animal auto-fluorescing. Mixing Amplex Red working solution with sea water and hydrogen peroxide in a drop caused the drop to turn red/pink and fluoresce strongly confirming that the presence of sea water did not inhibit the oxidation reaction of Amplex Red.

These findings suggest that there is a PO enzyme reaction occurring in the gut of *Limnoria* which is linked to the production of a peroxide in that organ. Inhibiting the PO enzyme activity with a PO inhibitor resulted in little to no detection of peroxide in the gut. As has already been discussed, HC proteins may be able to carry out a PO activity.

Taken together, these last two findings suggest that in the gut of *Limnoria* HCs may carry out a DPO enzyme activity which may be linked with the production of a peroxide species in that organ. Heterologous expression and purification of *Limnoria* HC proteins followed by in-vitro assaying would allow this statement to be made with more confidence and is the focus of Chapter 5 of this work.

# Chapter 5: Expression and Activity of Heterologous *Limnoria* Hemocyanins

## 5.1 Introduction

When studying an individual gene or protein to determine its function or biochemistry *in vivo* or *ex vivo* approaches can be useful, but there is almost always uncertainty of the effects of other cellular processes. They are also limited in the types of information which can be produced. This makes it necessary to study the gene or protein in isolation, *in vitro*. When the target of study is a protein or enzyme it can be very difficult, impractical or unethical to obtain sufficient quantities of the protein from its native host. Extracting and purifying a single protein directly from *Limnoria* for the purpose of physical or biochemical characterisation would require thousands of animals. As *Limnoria* lives deeply burrowed into wood, every animal needed must be found and broken out of infested wood which is both difficult and time consuming. Such a difficult exercise would be further exacerbated if the target protein was exclusively found in a single organ and required dissection, with each dissection taking around three minutes. This would render the approach impractical. When faced with such scenarios the preferred method is to overexpress the target protein in a heterologous host.

Heterologous expression allows the target protein to be produced by a host system which is either unicellular or cell free. This avoids the ethical component involved in sourcing a protein from a host animal. The approaches can also be scaled up to industrial quantities if necessary, negating any impracticality associated with obtaining the protein from its native source.

Once overexpressed, in order to study the target protein in isolation it must be purified from either the host cells, or if a secreted protein, the growth media. An excellent reference describing the history and general principles of protein separation can be found in (207). Below is a brief summary of these principles.

There are many commercial systems available for separating proteins from complex mixtures. Most work on the basis of capturing the target protein from a mixture of proteins using chromatography media. Undesired proteins are then washed away before unbinding (eluting) the target protein in isolation. It is not always possible to do

this in a single step and systems based on different separation criteria are often coupled to achieve homogeneity of the target protein. Separation media separate the target protein from others in the mixture using a physical/chemical characteristic of the target protein. Size exclusion or gel filtration chromatography separates proteins based on their size. Ion exchange separates proteins based on their charge at a particular pH (i.e. their pI). Hydrophobic interaction chromatography separates proteins based on their hydrophobicity. It is also possible to specifically bind a target protein based on its affinity for a ligand, so called affinity chromatography. These ligands can be a natural interaction partner of the protein, for example a lectin protein ligand can bind a specific sugar decoration of a glycosylated target protein. They can also be part of an engineered affinity system which uses a polypeptide “tag” which is included at a terminus of the target protein which can bind a specific ligand. Examples of such systems include the poly-histidine ($His_6$) tag, Strep·Tag II and glutathione S-transferase (GST) tag which have strong binding affinities for divalent metal ions (most often $Ni^{2+}$), StrepTactin and glutathione respectively. Of these affinity chromatography techniques the $His_6$ tag is most commonly used in part because the tag is very small, only six histidine residues long, and therefore often does not need to be removed from target proteins to allow activity or protein crystallisation. The $His_6$ tag can strongly bind a number of divalent metal ions including $Ni^{2+}$, $Co^{2+}$, $Cu^{2+}$ and $Fe^{2+}$ and the technique is referred to as immobilised metal ion affinity chromatography (IMAC).

There are many commercial or widely used heterologous expression systems available the hosts of which span many of the kingdoms of life. Each have their own advantages and disadvantages and when choosing which system to use these are considered along with a hierarchy of cost and complexity associated with each system. Table 5.1 outlines the major systems available and mentions each systems major advantages and disadvantages.

**Table 5.1. Advantages and disadvantages of many commonly used heterologous protein expression systems.**

| | Advantages | Disadvantages |
|---|---|---|
| *Escherichia coli* | • Many different vectors and strains available<br>• Well established genetic tools<br>• Quick and cheap system | • No post-translational modification<br>• Eukaryotic genes often difficult to express and form insoluble aggregates |
| *Pichia pastoris* | • Potential to secrete proteins<br>• Post translational modification | • Time consuming clone screening often required |
| *Aspergillus sp.* | • High protein secretion capacity<br>• Post translational modification<br>• Long history of safe use in the food industry | • Possess many endogenous proteases<br>• Strongly acidifies the growth medium leading to secreted protein degradation<br>• Fungal morphology influences protein yield |
| Insect cell culture (baculovirus mediated) | • Post translational modification<br>• Potential to secrete proteins | • Expensive growth media requirements<br>• Complex system<br>• Tendency to hyper-glycosylate proteins |
| Stable transgenic *Arabidopsis thaliana* | • Once established provides a sustainable supply of heterologous protein<br>• Post translational modification | • Requires a very long time frame to establish<br>• Many transformants need to be screened to find a good expresser<br>• Often low expression levels<br>• Often difficult to purify proteins from plant tissue |
| Transient *Agrobacterium* mediated plant expression | • Quick and simple system<br>• Post translational modification | • Efficiency of heterologous expression unpredictable from batch to batch<br>• Often difficult to purify proteins from plant tissue |
| *Xenopus laevis* oocytes | • Post translational modification<br>• Facilitating protein expression is quick and simple | • Not suitable for scale up<br>• Intensive use of skilled labour required for microinjection of mRNA<br>• Maintenance of animals is complex and expensive |
| Human cell culture | • Extensive eukaryotic post translational modifications possible<br>• Good potential for protein secretion | • Expensive growth media requirements<br>• Complex system<br>• Difficult to scale up |
| *In vitro* translation | • Very quick simple system | • Expensive<br>• Not very amenable to scale up |

Within this chapter are described efforts to utilise the baculovirus mediated insect cell culture and *E. coli* expression systems. Below is a brief outline of how these systems work, and what can be done to optimise soluble protein expression within them.

### 5.1.2 Baculovirus Mediated Insect Cell Culture Expression System

The baculovirus mediated insect cell culture expression system is a eukaryotic cell based expression system capable of post-translational modification with a high secretion capacity. An overview of the system is given in (208, 209). The baculovirus expression system consists of two parts; a virus and a host cell line. The Sf9 and Sf21 cells lines used in commercial baculovirus systems are derived from the ovaries of *Spodoptera frugiperda*, the fall armyworm. The baculovirus used in these systems is the *Autographa californica* nucleopolyhedrovirus (AcMNPV). The virus possess a circular, double stranded DNA genome of ~134 kbp which is packaged inside a rod shaped nucleocapsid. The virus has a two phase life cycle; a budded virus (BV) stage which consists of the DNA containing nucleocapsid inside an envelope of insect cell derived membrane, and an occlusion derived virus (ODV) stage in which multiple nucleocapsids are surrounded by a shell of viral polyhedrin protein forming occlusion bodies (OB). The BV stage of the life cycle is very vulnerable to damage from the environment and is responsible for cell to cell transmission of the virus. The ODV stage of the life cycle is resistant to UV and drying conditions in the environment and is responsible for animal to animal transmission.

In the natural environment a *S. frugiperda* caterpillar may ingest OBs of the virus present on a leaf. Once inside the gut of the animal the polyhedrin coat dissociates and releases the ODV which is taken up by the cells of the intestine by endocytosis. Inside the intestinal cells the viral genome rapidly replicates and forms BV particles which bud away from the infected cells to propagate the infection. In the very late phase of infection (24 – 72 hours post infection) ODV is formed which are imbedded in polyhedrin and the host cell lyses releasing OBs to infect more cells. Eventually this cycle of virus induced cell lysis causes the animal to essentially liquefy depositing OBs of virus onto the leaf that the host died on ready for the next caterpillar to ingest and begin the process again.


**Figure 5.1. Schematic representation of the pBACSEC-HC3 insertion cassette.** Important features of the insertion cassette are shown. Sizes are not to scale. ORF 1629, an essential viral gene; eGFP, enhanced green fluorescent protein; ETL, early-to-late promoter; polh, polyhedrin promoter and 5' un-translated region; gp64 s.s., viral gp64 gene signal peptide sequence; $His_6$, poly-histidine tag; lef2, viral lef2 gene.

In the laboratory various aspects of the viral life cycle have been modified and exploited for the effective production of heterologous protein. A transfer vector called pBACSEC based on the pBAC4x-1 vector (Novagen – Merck) has been modified by the Protein Production Laboratory of the Technology Facility, University of York. The transfer vector is used to transfer the gene to be overexpressed into the viral genome, and is a plasmid based vector able to replicate inside *E. coli* and therefore allows standard molecular biology techniques to be used for its manipulation. The gene to be overexpressed along with the features described below in this paragraph together form the "insertion cassette" and a schematic of their arrangement is shown in Figure 5.1. The pBACSEC vector contains a ligation independent cloning (LIC) site which is used to insert the gene to be overexpressed coupled to a C-terminal poly-histidine tag. Immediately upstream of this site, so as to be in-frame and attached to the inserted gene, is a secretion signal peptide sequence derived from the viral *gp64* gene to direct the overexpressed protein for secretion (this signal peptide was found to be extremely efficient at directing proteins for secretion in this system, Dr. Jared Cartwright – Protein Production Laboratory, Technology Facility, University of York, unpublished data). Further upstream are the viral polyhedrin promoter and a 5' un-translated region (UTR) from the lobster, *Homarus americanus*. The lobster derived UTR has been shown to further increase expression levels from this promoter (210). In the laboratory the ODV stage of the viral life cycle is unnecessary, therefore the polyhedrin gene has been knocked out of the viral genome and it's very strong promoter can be used instead to drive the overexpression of the inserted gene. Further upstream still, but orientated in the opposite direction is the enhanced green fluorescent protein (eGFP) gene under the control of the viral early-to-late (ETL) promoter. Under the control of this promoter eGFP protein will be produced by virus infected cells 6 – 24 hours post infection. This will allow the success/failure and progress of a viral infection to be quickly and easily monitored using a suitably equipped microscope. All of these features are flanked by the viral *lef2* gene on one side and the essential viral ORF1629 gene on the other.

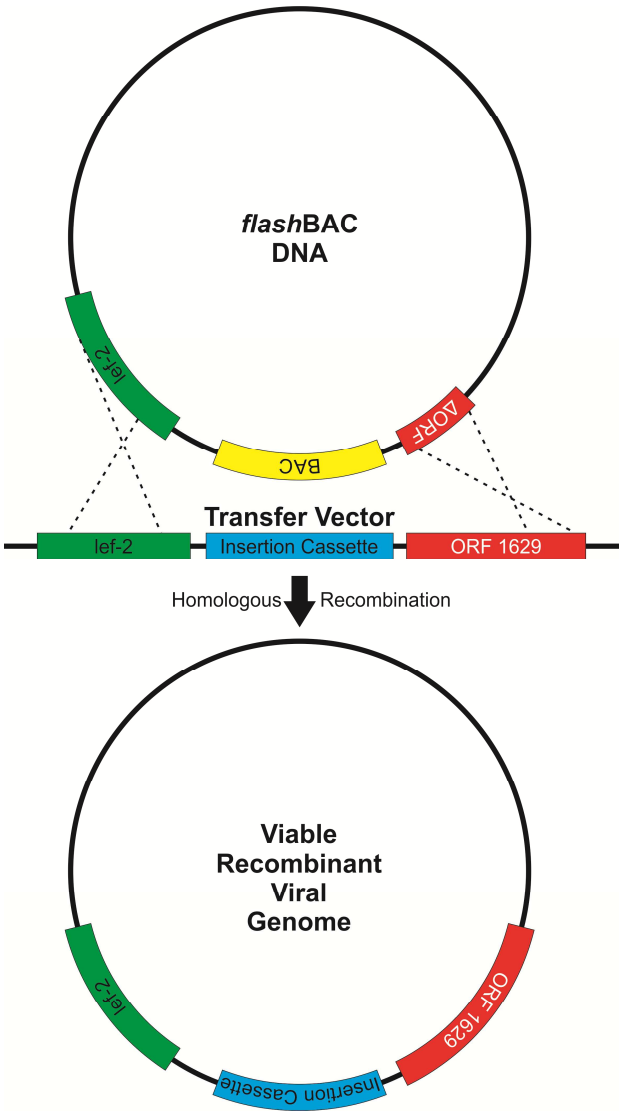



**Figure 5.2. Schematic representation of homologous recombination between a modified viral genome and transfer vector.**
Once co-transfected into insect cells the non-viable modified viral genome (in this case flashBAC DNA) and pBACSEC-HC3 transfer vector undergo homologous recombination resulting in a viable recombinant viral genome containing the gene to be overexpressed. Sizes are not to scale. ORF 1629, an essential viral gene; BAC, bacterial artificial chromosome; lef2, viral lef2 gene; insertion cassette, see Figure 5.1.

Once the gene to be overexpressed has been inserted into the transfer vector it is co-transfected along with a modified version of the viral genome, in this case the *flash*BAC DNA (Oxford Expression Technologies). This version of the viral genome has had its copy of the ORF1629 gene truncated making it non-functional, and therefore the genome none viable (211). It has also had its viral polyhedrin gene replaced with a bacterial artificial chromosome (BAC), making the genome a bacmid, allowing the genome to be replicated inside bacteria. Inside the insect cells homologous recombination occurs between the transfer vector and the modified viral genome DNA

resulting in a viral genome containing the gene of interest, eGFP and a full copy of the ORF1629 gene (Figure 5.2). This genome can go on to replicate and during the very late infection stage (24 – 72 hours post infection) will begin to overexpress the gene of interest. By using a non-viable modified viral genome in this way it is not necessary to verify the presence of the inserted gene in the virus by traditional plaque assay as no viable virus can be created without recombination with a transfer vector (211).

Due to the exacting growth conditions required by eukaryotic cultured cells there are few variables which can be adjusted to improve the soluble expression of a heterologous protein in the baculovirus system. One factor which can be varied is the way in which the cells are infected with the virus; either via the amount of cells infected or the amount of virus used as an initial inoculum. If the heterologous protein is harmful or toxic to the expressing insect cells then it would be beneficial for a large number of cells to be infected at the same time. In this way all of the cells would express the protein simultaneously with no need for further waves of infection. The amount of time an infection proceeds before harvesting can also be varied to allow the amount of target protein produced to build up over time, or to harvest it before degradation occurs. More substantial variables which can be investigated are the cell line used as a host (for example High Five cells – Invitrogen – Life Technologies) and the promoter used to drive overexpression (212).

### 5.1.3 *Escherichia coli* Expression System

The *E. coli* host expression system is one of the most widely used systems for the production of heterologous proteins. This is largely due to its low cost and relative simplicity, but also as one of the most used systems it has undergone much study and development resulting in the availability of many different strains and vectors tailored to aid in the expression of different kinds of proteins, and the knowledge of how to optimise the system for improved soluble expression. Several reviews provide a good overview of how the system works, has been developed and how it is still developing to meet the needs of today's biotechnology industry (213-216).

The work described within this chapter using *E. coli* as an expression host has all been carried out using the pET expression system. The pET expression system is based upon the biology of the *lac* operon and T7 bacteriophage and was originally developed by Studier and colleagues in the late 1980s (217-219). A brief outline of the pET expression system is given below and is reviewed well in (215).

The pET expression system is made up of two components, a heterologous gene vector (plasmid) and a compatible host strain of *E. coli*. The gene to be overexpressed is inserted into the multiple cloning site of a vector plasmid using standard molecular biology techniques. In the vectors used in the work described blow, upstream of the inserted gene in the plasmid sequence is located a copy of the T7 promoter, lac operator and a ribosome binding site. Downstream is a copy of the T7 terminator sequence. The vector also contains a copy of the lac repressor (*lacI*) gene and usually an antibiotic selection marker gene. To be able to allow inducible expression of the inserted gene in a pET vector, the vector must be transformed into a strain of *E. coli* which is a lysogen of the DE3 bacteriophage. The DE3 bacteriophage carries a fragment of DNA encoding the *lacI* gene, and the T7 phage RNA polymerase under the control of the *lac*UV5 promoter. This means that DE3 lysogen strains of *E. coli* possess a lactose/IPTG inducible copy of the T7 RNA polymerase gene in their chromosome.

When not induced the lac repressor complex binds to the lac operator sequence upstream of the chromosomal copy of the T7 RNA polymerase gene preventing its transcription. It also binds at the lac operator sequence upstream of the inserted gene in the pET vector. Native *E. coli* RNA polymerases are unable to initiate transcription from the T7 promoter sequence on the pET plasmid, and so without any T7 RNA polymerase able to be produced the inserted gene of interest is not expressed. If any un-induced basal expression of the T7 RNA polymerase were to occur it should still be unable to cause expression of the inserted gene due to the lac repressor complex present on the pET vector.

When induced by the addition of the gratuitous inducer IPTG the molecule binds to lac repressor complexes causing them to lose affinity for and dissociate from the lac operator sequences upstream of the T7 RNA polymerase gene and the inserted gene. This would then allow T7 RNA polymerase to be transcribed by the endogenous *E. coli* RNA polymerases, followed by T7 RNA polymerase transcription of the inserted gene resulting in overexpression.

Once induced the strength of interaction between the T7 RNA polymerase and its promoter, coupled with the high processivity of the enzyme result in large amounts of mRNA for the inserted gene being produced. However, despite this many factors can prevent that mRNA resulting in the production of large quantities of soluble, correctly folded and active heterologous protein. Much work has been done to find ways to work around these issues with many of the resulting discoveries having been

commercialised and becoming widely used. It is also possible to mitigate some of these barriers to successful overexpression by carefully controlling the conditions under which *E. coli* grows and overexpresses proteins. A general overview of these issues and ways to try and overcome them are discussed in (220-222). Those approaches used in this work are discussed in more detail in the results section of this chapter.

Most of the work discussed in this chapter was carried out with the *E. coli* expression system using the *Limnoria* HC1 and HC3 genes as targets for expression. However other lab members and an industrial partner, Novozymes, have made attempts to heterologously express other HC genes in alternative expression systems. These included trying to express the HC2 and HC4 genes in *E. coli*, and the HC1 gene in *Aspergillus sp.*, human cell culture and *in vitro* translation systems. The results obtained when trying to express HC2 and HC4 in *E. coli* largely mirror the results discussed below and will be discussed further in this context later. Attempts made within our lab and by our industrial partner to express HC1 in *Aspergillus sp.* resulted in the successful transcription of the gene, but an inability to generate any heterologous protein. Work carried out to express HC1 in cultured human cells resulted in protein expressed only in an insoluble form. It was possible to purify this insoluble protein from the human cells however, and it was used to raise an antibody against HC protein in a rabbit. Of the eukaryotic systems only the *in vitro* translation system was able to produce HC1 protein in a soluble form; however due to the scale and expense of the system it was not possible to purify the protein from the expression reactions and nor was it feasible to scale up the expression. It was also not possible to show enzymatic activity using this source of HC1 protein.

### 5.1.4 Aims of this Chapter

The work described in this chapter aimed to find a suitable procedure to heterologously produce soluble, active HC protein to allow its biochemical and physical characterisation.

# 5.2 Methods

## 5.2.1 Eukaryotic Expression – Baculovirus Mediated Insect Cell Expression

### *5.2.1.1 Cell Culture*

The Sf21 cell line was used throughout this work and was obtained as a line already adapted to growth in suspension culture in a serum free media from Gibco. This cell line is derived from the ovarian cells of the *Spodoptera frugiperda* (Fall Armyworm) insect species. All work undertaken with insect cells was carried out using standard aseptic techniques and a class II safety cabinet where appropriate. Cells were grown in Sf-900™ II serum free media (SFM) and were maintained in suspension culture at 27 ℃, shaking at 80 rpm. Cells were routinely passaged to a density of 0.4 – 1 x $10^6$ cells/ml. Cell line viability was checked whilst passaging by trypan blue exclusion assay (223).

### *5.2.1.2 Linearisation of Plasmid*

Twenty micrograms of pBACSEC vector were digested using the BseRI restriction endonuclease as described in Section 2.4.9. A negative control digestion was also set up containing 1 µg of pBACSEC vector DNA and no restriction endonuclease.

Both digestions were run on a 0.6 % w/v agarose gel to separate cut and un-cut vector. The bands visible on the gel representing cut (i.e. linear) vector DNA were excised and the DNA within them extracted as detailed in Section 2.4.5.

### *5.2.1.3 PCR Amplification of Gene Insert*

The HC3 gene to be inserted into the transfer vector was amplified by PCR as described in Section 2.4.1. The HC3 gene sequence is shown in Appendix B. The primers used are detailed in Table 5.2. An aliquot of the completed PCR reaction was ran on an agarose gel to verify correct amplification before purifying the remaining reaction volume as described in Section 2.4.5.

The template DNA used was present in the pJET1.2 vector and was kindly provided by Dr. Andrew King (Centre for Novel Agricultural Products, Department of Biology, University of York). The HC template vectors were produced by amplification of the full length sequences from Limnoria cDNA and subsequent cloning of these amplified sequences into the pJET1.2 vector, part of the CloneJET high efficiency blunt end cloning system (Thermo Fisher Scientific). Cloned sequences were verified by DNA sequencing prior to use.

**Table 5.2. Sequence of primers used for the amplification of HC3 for cloning into the pBACSEC vector.**

| Primer | Sequence (5' → 3') |
|---|---|
| INSECT_HC3_F | CACCACCACCACATGTCTTGGTTACCGGAATTCCAATCG |
| INSECT_HC3_R | GAGGAGAAGGCGCGTTATTCTCTGTGGGTAATCTTCACAATAG |

### *5.2.1.4 Ligation Independent Cloning*

Linearised pBACSEC vector and PCR amplified HC3 DNA were treated with T4 DNA polymerase as shown in Table 5.3 at 22 ℃ for 30 minutes before inactivating the reactions at 75 ℃ for 20 minutes. This process would produce compatible sticky ends on both pieces of DNA facilitating ligation independent cloning (LIC cloning).

**Table 5.3. Volumes of components required to set up T4 DNA polymerase reactions.**

| Component | Linearised pBACHIS | HC3 |
|---|---|---|
| LIC qualified T4 DNA polymerase | 1.2 µl | 0.4 µl |
| T4 DNA polymerase buffer (10x) | 5.0 µl | 2.0 µl |
| 25 mM dTTP (vector) or dATP (gene) | 5.0 µl | 2.0 µl |
| 100 mM DTT | 2.5 µl | 1.0 µl |
| DNA | 25.0 µl | 2.0 µl |
| Water | 11.3 µl | 12.6 µl |

The two sources of DNA were annealed together by mixing 1 µl of T4 DNA polymerase treated linear pBACSEC with 2 µl of T4 DNA polymerase treated PCR amplified HC3 for 10 minutes at 20 ℃. After incubation 1 µl of 25 mM EDTA was added to the mixture and incubated for 10 minutes at 20 ℃. All 4 µl of the annealing reaction was transformed into XL1-Blue Supercompetent cells as described in Section 2.4.14 using

ampicillin (100 µg/ml) as a selection pressure. A negative control was also carried out replacing HC3 DNA with water.

The transformation resulted in the production of a large number of colonies (>150) whilst the negative control reaction resulted in very few colonies (<15). Plasmid DNA was prepared from 8 randomly chosen colonies from the transformation as described in Section 2.4.6. The primers, pBACSEC Forward and Reverse, detailed in Table 2.3 (Chapter 2) were then used to sequence across the insertion site of the plasmids prepared. This resulted in one plasmid which was found to contain the correct HC3 gene inserted in frame and with no errors in the gene sequence. This plasmid carrying the *Limnoria* HC3 gene was designated the transfer vector (pBACSEC-HC3).

### *5.2.1.5 Sterilisation of Transfer Vector*

The transfer vector was sterilised by size exclusion filtration. The transfer vector was placed into a sterile Spin-X 0.22 µm cellulose acetate spin filter (Costar), and the spin filter placed into a sterile micro-centrifuge tube. The filter was spun at 13000 x g for 2 minutes. The spin filter was discarded and the sterilised transfer vector stored at -20 ℃ until further use.

### *5.2.1.6 Co-Transfection*

An aliquot of Sf21 insect cells were adjusted to a cell density of 0.75 x $10^6$ cells/ml and 420 µl of this preparation transferred into a well of a 24 well Nunclon® coated cell culture dish. The cells were allowed 30 minutes to drop out of suspension and adhere to the bottom of the well.

Meanwhile, the co-transfection mixture was prepared by mixing 1 µl of sterile transfer vector (~600 ng of DNA), 2.5 µl of *flash*BAC DNA (Oxford Expression Technologies), 22 µl of Sf-900 II SFM and 2 µl of Lipofectin® (Invitrogen – Life Technologies). The mixture was allowed to stand for 30 minutes to allow liposomes to form, encapsulating the DNA.

Once both incubation periods had elapsed 25 µl of the co-transfection mixture were carefully dripped into the media of the well containing cells so as not to disturb the

cells. The cell culture dish was sealed with Parafilm and carefully placed into a humidified sandwich box before incubating at 27 ℃.

After 24 hours, 25 µl of foetal calf serum (FCS) were added to the well to act as a protectant for any produced virus. The cells were incubated in the same manner for a further 4 days.

### *5.2.1.7 Harvesting Seed Virus*

Once the co-transfection incubation period had elapsed the cells were very briefly visualised under UV light to confirm that virus particles had been produced. If they had, enhanced green fluorescent protein would be made in the cytoplasm of infected cells causing them to fluoresce green under UV light. Fluorescing cells were observed indicating viable virus has been produced so the viral seed stock was harvested. A pipette was used to gently wash any remaining adherent cells off the bottom of the well before aspirating all of the mixture into a sterile cryovial. The seed virus was stored in the dark at 4 ℃ until a P2 virus had been successf ully created, after which the seed virus was snap frozen in liquid nitrogen and stored at -80 ℃.

### *5.2.1.8 P1 Amplification*

The amplification of the seed stock virus to produce a P1 virus was carried out in a µ-24 MicroReactor Central Vent Cell Culture Cassette (Pall). To one well of a cassette was added 5 ml of Sf21 cells at a density of 0.25 x $10^6$ cells/ml and 100 µl of seed stock virus. The well was sealed with a sterile breathable plate cover and the cassette loaded into a µ-24 Bioreactor (Applikon Biotechnology). The bioreactor was programmed to maintain the well at 27 ℃, shake the cassette at 500 rpm and ensure a minimum dissolved oxygen level within the well of 60 %. The amplification was carried out over 5 days, supplementing the infection to 5 % v/v FCS after 24 hours. The cell suspension was centrifuged at 1000 x g for 10 minutes and the virus containing supernatant removed and stored at 4 ℃ in the dark.

### *5.2.1.9 P2 Amplification*

A P2 virus was produced by inoculating 50 ml of Sf21 cells at a density of 0.5 x $10^6$ cells/ml with 250 µl of P1 virus in a 250 ml vented top disposable culture flask. The

infected cells were incubated at 27 ℃ for 1 week, supplementing the flask to 5 % v/v FCS after 24 hours. The virus containing supernatant was harvested by centrifugation at 1000 x g for 10 minutes and stored at 4 ℃ in the dark.

### *5.2.1.10 Titration of P2 Virus*

Aliquots of 50 µl of Sf21 cells at a density of 0.5 x $10^6$ cells/ml were added to wells of a Nuclon coated 96 well culture plate. The cells were allowed 30 minutes in the dark at room temperature to adhere to the culture plate.

A 10 fold dilution series of P2 virus was produced using Sf-900 II SFM media as a diluent, ranging from neat to a 1 x $10^{-7}$ fold dilution. In duplicate, a 50 µl aliquot of each virus dilution was dripped into a well of cells. This was also done with an aliquot of Sf-900 II SFM media as a negative control. The plate was covered with its lid, placed into a humidified sandwich box and incubated at 27 ℃ for 36 hours.

After incubation the plate was examined using a CK40 stereo microscope (Olympus) under UV light. At each dilution of virus the number of cells seen to fluoresce was counted with an upper limit of 50 cells.

The viral titre was calculated using the below formula. The viral dilution which resulted in ~10 cells being seen to fluoresce was used in the calculation.

$$\text{Number of fluorescing cells} \times \text{Viral dilution factor} \times \text{Volume correction} = \text{Viral titre}$$

The result of the calculation is measured in plaque forming units (pfu) per millilitre and each pfu represents a budded virion particle capable of causing the infection of one cell.

### *5.2.1.11 Protein Expression Trial*

A protein expression trial was carried out in a µ-24 MicroReactor Central Vent Cell Culture Cassette (Pall). To two wells of the cassette 5 ml of Sf21 cells at a cell density of 2 x $10^6$ cells/ml were added. One well was infected with the HC3 virus with a multiplicity of infection (see below) value of 0.1 and the other of 1.0. The cassette was

loaded into a µ-24 Bioreactor. The bioreactor was programmed to maintain the wells at 27 ℃, shake the cassette at 500 rpm and ensure a m inimum dissolved oxygen level within the wells of 60 %.

After 20 minutes to allow the wells to equilibrate to their set conditions, t=0 hours samples were taken. This was done by removing 250 µl of suspension from each well into micro centrifuge tubes. Cells were pelleted from the suspension by centrifugation at 5000 rpm. The supernatant (media) was removed to a fresh tube containing a suitable amount of 5x SDS sample loading buffer. The cell pellet was resuspended in 250 µl of 1x SDS sample loading buffer. Samples were frozen at -80 ℃ until processing. Further samples were taken in the same way at 24, 48, 72, 96 and 120 hours post inoculation.

A multiplicity of infection value (MOI) indicates the number of virions added to an infection per cell present in the culture and is calculated as shown below. Therefore a MOI value of 1 indicates that for every cell present in a culture a virion has been added, a value of 0.1 indicates that 1 virion has been added for every 10 cells. Varying the MOI of an infection can allow all cells to be infected at once resulting in synchronicity of infection and protein production, or it can allow waves of infection to occur allowing the number of cells in the culture to increase prior to the infection of the whole culture.

$$\frac{\text{Multiplicity of infection} \times \text{Number of cells}}{\text{Viral titre}} = \text{Volume of virus to achieve desired MOI}$$

Once all samples had been collected they were thawed at room temperature before being boiled for 5 minutes. Samples were then analysed in duplicate using 10 % v/v acrylamide RunBlue SDS-PAGE pre-cast gels (Expedeon) as described in Section 2.5.5. One duplicate gel was Coomassie stained, the other was blotted as described in Section 2.5.6 and probed with a 1 in 2000 dilution of a monoclonal anti-polyHistidine-Peroxidase antibody produced in mouse (product number #A7058, Sigma Aldrich).

### *5.2.2* Prokaryotic Expression – *Escherichia coli*

Throughout this work numerous methods of expressing HC proteins in *Escherichia coli* were evaluated in an attempt to obtain pure, soluble hemocyanin protein. These various methods broadly followed the general procedure described below. Unless

otherwise stated, suitable antibiotics were included at all stages to preserve uniformity of the cultures, and maintain plasmid selection pressure.

### *5.2.2.1 Cloning*

In order to express HC genes in *E. coli* they were amplified by PCR and cloned into various expression vectors. The target gene was amplified by PCR as described in Section 2.4.1. The HC gene sequences are shown in Appendix B. The template DNAs used were kindly provided by Dr. Andrew King (Centre for Novel Agricultural Products, Department of Biology, University of York) present in pJET1.2 vectors. The primers used are detailed in Table 5.4. Aliquots of completed PCR reactions were analysed by agarose gel electrophoresis to verify correct amplification before purifying the remaining reaction volume as described in Section 2.4. The purified PCR products were then cloned into their respective expression vectors using the cloning techniques discussed in Section 2.4.10-12.

Following transformation of plasmids containing the cloned PCR products into *E. coli* bacterial colonies were formed. To verify that these colonies carried the correct DNA construct they were screened by colony PCR as described in Section 2.4.8. In each cloning procedure several colonies identified as carrying the correct DNA construct were used to produce plasmid DNA (Section 2.4.6). The plasmid DNA from these colonies was sequenced as described in Section 2.4.7 to verify that the intended gene had been inserted into the vector correctly in-frame and that the inserted gene was free from any coding errors incurred during amplification. A construct found to be wholly correct was taken forward for protein expression testing.

**Table 5.4. Details of the primers used to amplify different HC genes for cloning into various pET based vectors using various cloning methods.**
HC1*, an *E. coli* codon optimised version of the HC1 gene; pET26b(+)§, a Strep·Tag II was added to the C-terminal of HC3 by PCR and the vector's native $His_6$ tag not used; pET28a(+)‡, primers designed to exclude the vectors N-terminal $His_6$ tag but include the C-terminal $His_6$ tag. See Figure 5.6 for a schematic of the constructs.

| Gene | Vector | Primers | Cloning Method |
|---|---|---|---|
| HC1* | pET28a(+)‡ | **F:**AGGAGATATACCATGGGTAGCTGGATTCCTGAATATCA | In-Fusion® HD Cloning |
| | | **R:**CCGCAAGCTTGTCGACTTCACGATGTTTAATGGTAAC | |
| HC3 | pET26b(+)§ | **F:**CCGGCGATGGCCATGTCTTGGTTACCGGAATTCC | In-Fusion® HD Cloning |
| | | **R:**GGTGGTGGTGCTCGATTACTTTTCGAACTGCGGGTGGCTCCATTCTCTGTGGGTAATCTTCACA | |
| HC3 | pET28a(+) | **F:**GTACGGATCCTCTTGGTTACCGGAATTCCAATCG | Strataclone sub cloning, DNA ligase cloning |
| | | **R:**TCTTGTCGACGCTTATTCTCTGTGGGTAATCTTCACAAT | |
| HC3 | pET52b(+) | **F:**ACTAGTCGACTCTTGGTTACCGGAATTCCAAT | Strataclone sub cloning, DNA ligase cloning |
| | | **R:**CCTTGCGGCCGCTGTTATTCTCTGTGGGTAATCTTCACAATAGTC | |

### *5.2.2.2 Expression*

A construct to be used for expression was transformed into a protein expression strain of *E. coli* as detailed in Section 2.4.14. Various strains of *E. coli* were trialled throughout this work; details of those strains along with their features and advantages are listed in Table 2.1 (Chapter 2). A fresh transformation was carried out for each protein expression experiment to minimise the potential for plasmid loss.

Freshly transformed single colonies were used to inoculate starter cultures. These cultures were usually 10 ml in size, contained in 50 ml Falcon tubes and used 2 x YT broth as a growth medium. The cultures were grown overnight at 37 ℃ shaking at 180 rpm. Occasionally expression experiments were carried out at this scale (10 ml of media). In these cases the cultures were grown at 37 ℃ shaking at 180 rpm until their optical density at 600 nm ($OD_{600}$) had reached a desired value. After this point they were treated as larger cultures were, skipping the procedures of the next paragraph.

The cells from starter cultures were harvested by centrifugation at 4000 x g and resuspended in fresh growth medium. This was done to discard antibiotic degrading enzymes which may have built up in the medium of the starter culture. The resuspended cells were used as an inoculum for larger amounts of growth media (25 – 500 ml).

Larger cultures were grown at 37 ℃, shaking at 180 rpm until their $OD_{600}$ had reached a desired value. At this point cultures were induced to express their plasmid encoded gene of interest by supplementing the growth media with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of either 0.1 or 1.0 mM (see Appendix C). Once induced, cultures were often then incubated at a different temperature to the growth phase, still shaking at 180 rpm, for various amounts of time as discussed later. After a defined period of expression time cultures were chilled on ice, and their cells harvested by centrifugation at 4000 x g. The spent growth media was discarded and cell pellets weighed before being snap frozen in liquid nitrogen. Cells were stored at -80 ℃ until processing.

### *5.2.2.3 Processing*

Cell pellets to be processed were thawed on ice before being resuspended in lysis buffer. The amount of liquid used for resuspension varied according to the lysis method to be used (see Section 2.5.4). Lysis buffer was typically composed of 100 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 % v/v Tween 20, 50 µg/ml hen egg white lysozyme, 20 U of DNase, 1 x HALT EDTA Free Protease Inhibitor Cocktail (Thermo Fisher Scientific). Resuspended cells were incubated at 4 ℃ on a rotary shaker at 12 rpm for 30 minutes. In certain circumstances an alternative buffer was used, and other components were added or removed from the lysis buffer depending on the downstream processing to take place.

The cells were lysed either using sonication or cell disruption as described in Section 2.5.4. Cell lysates were fractioned into soluble and insoluble fractions by centrifugation at 15,000 rpm for 25 minutes at 4 ℃ using a RC-5B centrifuge (Sorvall) equipped with an SS-34 rotor.

If no purification was to be carried out these soluble and insoluble fractions would be analysed by SDS-PAGE as described in Section 2.5.5. Often samples were also Western blotted to identify the presence or absence of the heterologously expressed

protein (Section 2.5.6). In certain circumstances bands were excised from SDS-PAGE gels and subjected to gel band protein identification as described in Section 2.6.2.

### *5.2.2.4 Purification*

All constructs produced in this work expressed their inserted genes fused to an affinity purification tag, either a poly-histidine tag or a Strep·Tag® II tag. Several different purification systems and types of purification media were tested in an attempt to separate the overexpressed target proteins from the mixture of host proteins. Such systems included batch purification using gel bead based affinity resin, pre-packed affinity columns driven by a peristaltic pump and semi-automated purification using pre-packed affinity columns and a an ÄKTA fast protein liquid chromatography (FPLC) system (GE Healthcare Life Sciences).

The methods used for each purification system followed the manufacturer's guidelines and for brevity are not included here; however all such methods follow a standard set of procedures which is described below.

The first step in a purification procedure is to equilibrate the purification media (gel beads or column) with the same buffer that the protein to be purified is in, typically lysis buffer. This step is typically carried out using 5-10 column volumes (CV) of buffer. In the case of gel bead media, a column volume is the volume of beads used. This washes away any compounds used for storage of the media which may cause proteins to precipitate out of solution, such as ethanol. It also adjusts the pH of the media to that of the solution the desired protein is present in; if there is a large pH difference between the lysate and the purification media this may also cause proteins to precipitate.

The next step in purification is to bind the target protein to the media. In the case of batch purification using gel beads this is achieved by incubating the soluble cell lysate with the beads with gentle agitation, typically for an hour at room temperature or 16 hours at 4 ℃. When using pre-packed columns this binding is achieved by pumping the soluble cell lysate through the column, typically at a flow rate 2-5 times slower than optimal column flow rate. This allows the lysate more time in contact with the media inside the column, increasing the chance for the target protein to bind.

After the target protein has been bound to the separation media, it is necessary to flush remaining lysate away from it, and to dissociate weakly or non-specifically bound proteins. This is done using a wash buffer. For affinity purifications this is typically the same as the lysis buffer but containing more NaCl. This is done to provide a greater ionic strength to dissociate weakly bound proteins from the separation media. For nickel affinity purification, as used by poly-histidine tagged proteins, a change in pH of the buffer or inclusion of a low amount of imidazole (10-40 mM) can also be used to remove non-specifically bound proteins.

Once the separation media have been washed the target protein is ready to be eluted. Poly-histidine tagged proteins bound to nickel affinity media are most often eluted using a high concentration of imidazole, the functional group side chain of histidine. Imidazole causes the elution of proteins by exchanging with histidine residues bound to nickel ions which are attached to the separation media. The use of the FPLC system allows a steady, increasing amount (gradient) of imidazole to be applied to the separation media. This is beneficial as any remaining non-specifically bound proteins may be eluted at lower imidazole concentrations before the target protein, which is likely to have the highest affinity for the separation media. It is also possible to conveniently collect smaller fractions using the FPLC system allowing for finer physical separation of the eluted proteins into separate aliquots. Poly-histidine tagged proteins can also be eluted using a lowering of the pH to 4.5. At this pH histidine residues are protonated and therefore cannot bind to the $Ni^{2+}$ of the separation media. Strep·Tag® II tagged proteins are eluted using a solution containing a compound which specifically competes with the Strep·Tag® II for its binding site in the separation media, *d*-desthiobiotin.

The final stage in a protein purification procedure is to regenerate the separation media for re-use. Nickel affinity media can be regenerated by simply flushing the column with 5-10 CVs of lysis buffer. If a more stringent regeneration is required the media can be stripped of its nickel ions, and any remaining protein, using a solution of EDTA, before re-charging the media with 0.1 M $NiSO_4$. Strep·Tag® II affinity media is regenerated by washing it with 0.1 M NaOH which removes remaining *d*-desthiobiotin. The media is then re-equilibrated with lysis buffer.

Aliquots of material from each stage of a purification were retained and analysed by SDS-PAGE as described in Section 2.5.5. Often samples were also Western blotted to identify the presence or absence of the heterologously expressed protein (Section

2.5.6). In certain circumstances bands were excised from SDS-PAGE gels and subjected to gel band protein identification as described in Section 2.6.2.

If the elution fractions were found to contain sufficient amounts of soluble target protein they may have been used in native PAGE based DPO enzymatic activity assays as described in Section 2.6.1. Depending on the concentration of the protein present in the elution fractions, concentrating or de-salting steps as described in Sections 2.5.2-3 may have been required prior to activity assay.

### 5.2.3 Prokaryotic Expression – ArcticExpress *Escherichia coli*

ArcticExpress (DE3) RIL *E. coli* cells were transformed with the pET28a(+) HC1 (optimised) construct as described in Section 2.4.14. Starter cultures were established as described above including 20 µg/ml gentamycin and 50 µg/ml kanamycin as selection pressures. After incubating the cultures at 37 ℃ for 16 hours they were used to inoculate larger amounts of media (10 ml of starter culture per 500 ml of media). Larger cultures were grown without antibiotic selection pressures at 30 ℃ whilst shaking at 180 rpm until they had reached an $OD_{600}$ of ~0.5. The temperature of the incubator was then reduced to 14 ℃ and the $OD_{600}$ of the cultures continuously monitored. Once the incubator had reached its set temperature, and the cultures had reached an $OD_{600}$ of 0.8 – 0.9 the cultures were induced to overexpress HC1 by adjusting the media to contain 0.5 mM IPTG. Copper sulphate was also added to the media to a final concentration of 100 µM in case a source of copper ions, which HCs are suspected to contain, are required for proper folding. Cultures were grown under these conditions for 22 – 24 hours before harvesting their cells by centrifugation as described above.

Harvested cells were processed as described in Section 5.2.2.3 with the following modifications. The soluble lysate fraction was vacuum filtered through a 0.22 µm membrane (Sartorius). To the soluble lysate fraction was added a suitable amount of 10x dissociation buffer (50 mM Tris-HCl pH 7.4, 10 mM $MgSO_4$, 2 mM ATP (disodium salt)) to aid in the reduction of co-purifying DnaK and GroEL chaperone proteins (224, 225). The lysate was incubated at 37 ℃ for 15 minutes to allow the dissociation to occur before proceeding.

Filtered soluble lysate fractions were purified using an ÄKTApurifier 10 FPLC system (GE Healthcare Life Science). The system was fitted with a 1 ml HisTrap HP column.

Solution A was 100 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.1 % v/v Tween 20, 5 mM imidazole and solution B was as A but containing 500 mM imidazole. The following purification program was used: equilibration, 5 CVs of 0 % B; sample loaded from a 50 ml superloop at a flow rate of 0.5 ml/ml; wash, 10 CVs of 4 % B; elution, a rising imidazole gradient over 20 CVs from 4 % to 100 % B collecting 1 ml fractions. The absorbance at 280 nm, used as an indicator of the presence of protein, was continually recorded. Samples from each purification step as well as elution fractions corresponding to peaks in the 280 nm trace were analysed as described above. Again, if elution fractions were found to contain sufficient amounts of soluble target protein they may have been used in native PAGE based DPO enzymatic activity assays as described in Section 2.6.1. Where a DPO enzyme inhibitor was tested, kojic acid at a final concentration of 2 mM was used. Depending on the concentration of the protein present in the elution fractions, concentrating or de-salting steps as described in Sections 2.5.2-3 may have been required prior to activity assay.

# 5.3 Results

## 5.3.1 Eukaryotic Expression System

In order to express the *Limnoria* HC3 gene in insect cells first a transfer vector had to be constructed. This would allow the HC3 gene to be incorporated into a viral genome ready for the infection of insect cells which would then express the gene to create heterologous HC3 protein.

To produce the transfer vector first the destination plasmid, pBACSEC, was linearised using a restriction endonuclease. When subjected to gel electrophoresis along with un-cut plasmid it can be seen that the restriction digestion has successfully linearised the plasmid (Figure 5.3 – A). Un-cut circular DNA forms supercoiled structures which migrate with lower apparent molecular weights due to the structures reducing the molecular radius of the DNA; in this case un-cut plasmid migrates with an apparent molecular weight of 3 kbps. Linearised DNA however does not form such structures and so migrate at their expected molecular mass. Figure 5.3 – A 1-3 show linearised pBACSEC migrating with an apparent molecular mass of 6 kbps, much closer to the expected mass of ~6.7 kbps.

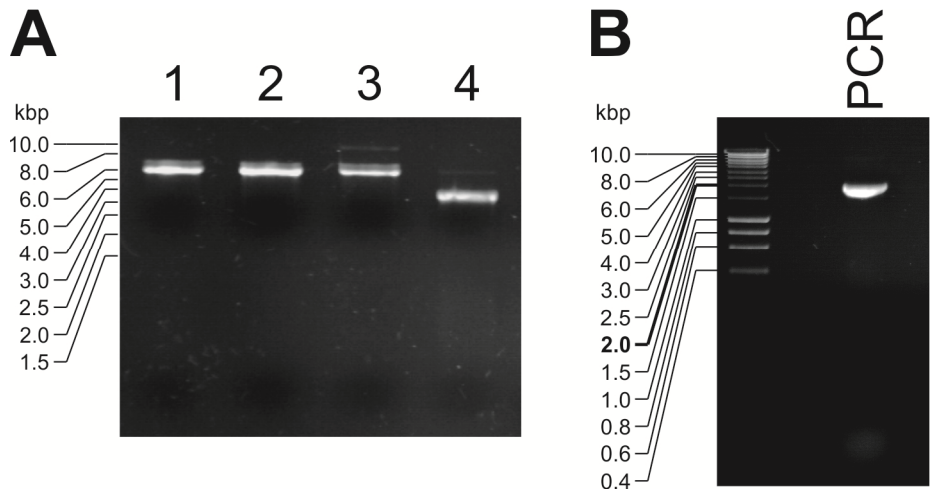



**Figure 5.3. Agarose electrophoresis gels of linearised pBACSEC and amplified HC3 DNA.**
A) 1 – 3, linearised pBACSEC DNA; 4, non-linearised pBACSEC DNA. B) PCR amplification of HC3.

The target gene, HC3, was PCR amplified using suitable primers to be cloned into the pBACSEC vector. An aliquot of this PCR reaction was analysed by gel electrophoresis

which showed that the reaction had successfully amplified a single product of approximately the correct size (~2 kbps) to be HC3 (Figure 5.3 – B).

A construct was produced from the PCR amplified HC3 target gene and the linearised pBACHIS vector using the LIC cloning method. Eight colonies produced by the subsequent transformation of the cloning reaction were used to produce plasmid DNA and these DNA samples were sequenced. One plasmid was found to contain the target gene inserted correctly in-frame and with no errors in its sequence. This was then referred to as the transfer vector (pBACSEC-HC3).

The transfer vector was co-transfected into insect cells with *flash*BAC DNA, a modified version of the baculovirus genome. After five days the cells transfected were visualised under UV light and cells were seen which fluoresced green indicating that they were expressing eGFP and therefore that a viable virus genome had successfully been produced (Figure 5.4). These cells and their media were harvested as a seed virus.

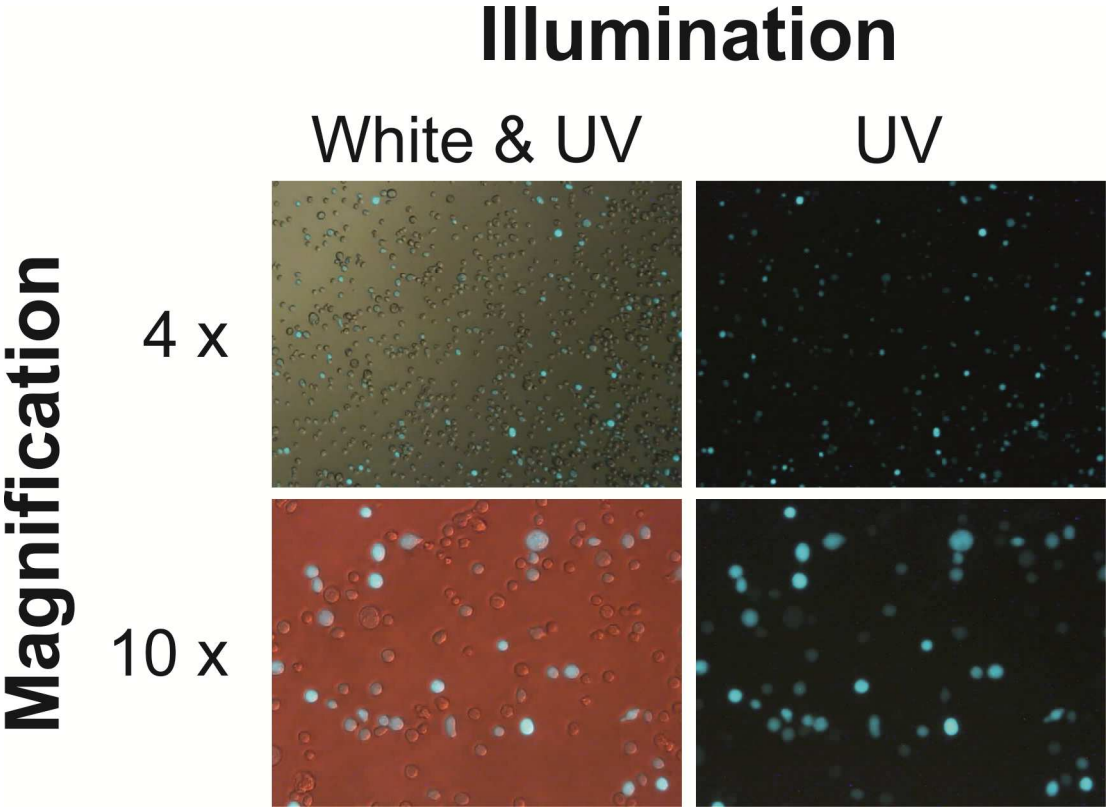



**Figure 5.4. Micrographs of seed virus producing insect cells.**
Micrographs were obtained using white light and UV illumination to visualise infected (green) and non-infected/very early infected cells (transparent); and UV illumination alone to better visualise cells in the early-to-late stage of infection or later (~6 hours post infection or greater).

The seed virus was amplified twice to produce a P2 virus. This P2 virus was titred to assess how many viable virus particles it contained. It was found that the virus contained 2 x $10^6$ pfu/ml of virus when used to infect an adherent culture of cells. Infections of cell suspensions are ~10 x more efficient than infections of adherent cells (Dr. Jared Cartwright, Protein Production Laboratory, Technology Facility, University of York), therefore the titre figure used in later calculations was taken to be ~2 x $10^7$ pfu/ml.

A protein expression trial was conducted using the P2 virus to infect insect cells in suspension with MOI values of 0.1 and 1.0. Each well in the expression trial contained 5 ml of cells at a density of 2 x $10^6$ cells/ml. This meant that for MOI values of 0.1 and 1.0, volumes of 0.05 and 0.5 ml of virus were added respectively.

Samples of cell pellets and culture medium were analysed by SDS-PAGE and Western blot from each time point of each infection of the protein expression trial. An aliquot of non-infected insect cells was also analysed. The samples taken from the MOI 0.1 infection showed no signal in the corresponding Western blot indicating that no poly-histidine tagged protein had been produced (data not shown). Samples taken from the MOI 1.0 infection showed production of a poly-histidine tagged protein in the insect cell pellets between 72 and 120 hours post inoculation (Figure 5.5 – B). No such signal was observed in the culture medium where one was expected due to the inclusion of a secretion signal in the expression construct. The signal observed in the cell pellet samples was at an apparent molecular mass slightly less than 70 kDa (i.e. potentially partially degraded, being less than the expected 75.1 kDa). It is not possible to observe the appearance of corresponding bands in the Coomassie stained counterpart gel (Figure 5.5 – A). An intense band is seen at all of the time points in the culture medium samples at an apparent molecular mass of just under 70 kDa. This band is likely to be albumin (~69.3 kDa) present in the FCS which was added to the P2 virus as a protectant. No signal from a poly-histidine tagged protein was observed in the sample of non-infected cells.

As no secreted poly-histidine tagged protein was observed to be produced and therefore it is unlikely that HC3 was produced in a soluble form no further work was carried out using this system (see the relevant section of the discussion for further details).

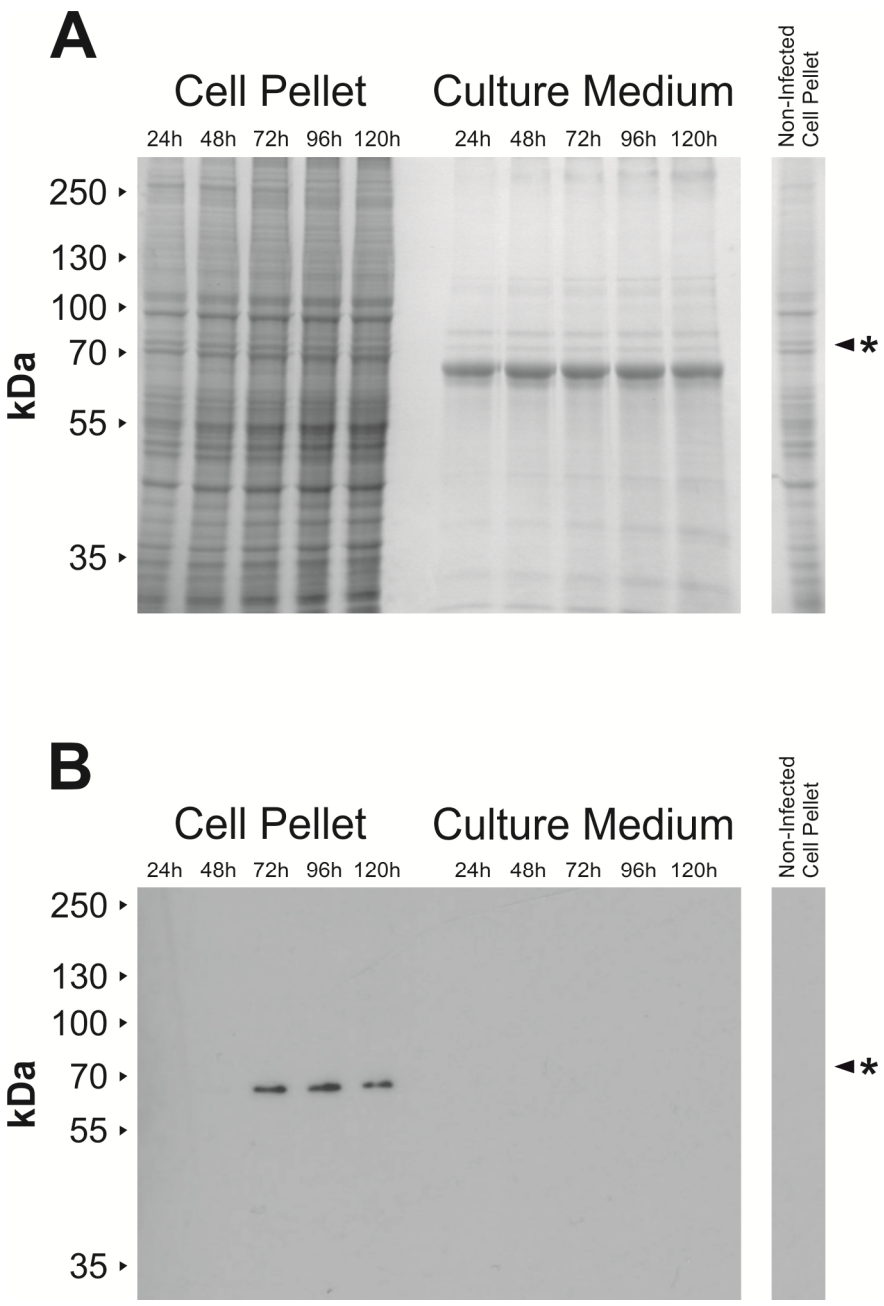



**Figure 5.5. SDS-PAGE gel and Western blot of an insect cell HC3 expression trial.** Cell pellet and culture medium samples from various time points of a HC3 protein expression trial in insect cells were analysed by SDS-PAGE (A) and Western blot (B). Also a non-infected sample of insect cells were analysed. Western blot carried out using a monoclonal anti-polyHistidine-Peroxidase antibody. Times indicated are hours post infection. Expected molecular mass of pBACSEC-HC3, 75.1 kDa, denoted by *.

## 5.3.2 Prokaryotic Expression

### *5.3.2.1 Escherichia coli*

During attempts to express *Limnoria* HC proteins in *E. coli* many different variables were assessed in an effort to improve the yield or purity of soluble heterologously expressed protein. Details of each of these variables along with the findings associated are given below and are summarised in Appendix C.

Several strains of *E. coli* were assessed for their ability to express HC3 in a soluble form. Each strain has its own properties which give it an advantage for heterologous expression and these are detailed in Table 2.1, Chapter 2. Briefly BL21 (DE3) pLysS co-expresses the T7 lysozyme protein which should effectively suppress background un-induced expression avoiding the possible selection pressure described below when discussing the inclusion of glucose in the culture medium and by Studier et al. (226). The BL21 (DE3) Star™ strain possesses a mutation which enhances the stability of mRNA molecules produced. This should result in HC3 mRNAs persisting in the cell for longer giving them a greater chance to be translated (227, 228). Finally the Rosetta-gami 2 (DE3) strain possess mutations in several genes which result in a more oxidising environment in the bacteria's cytoplasm, facilitating the easier formation of di-sulphide bonds (229, 230). It also possesses plasmids encoding seven tRNAs which are rarely used by *E. coli* (231); this should help alleviate the problem of codon bias when expressing genes with a different codon usage pattern. The BL21 (DE3) pLysS strain resulted in very little HC3 being expressed and only in an insoluble form. The BL21 (DE3) Star™ strain resulted in very high levels of HC3 expression, but again solely in an insoluble form. The Rosetta-gami 2 (DE3) strain was able to produce some soluble expression of HC3 and so has been used for all further development work.

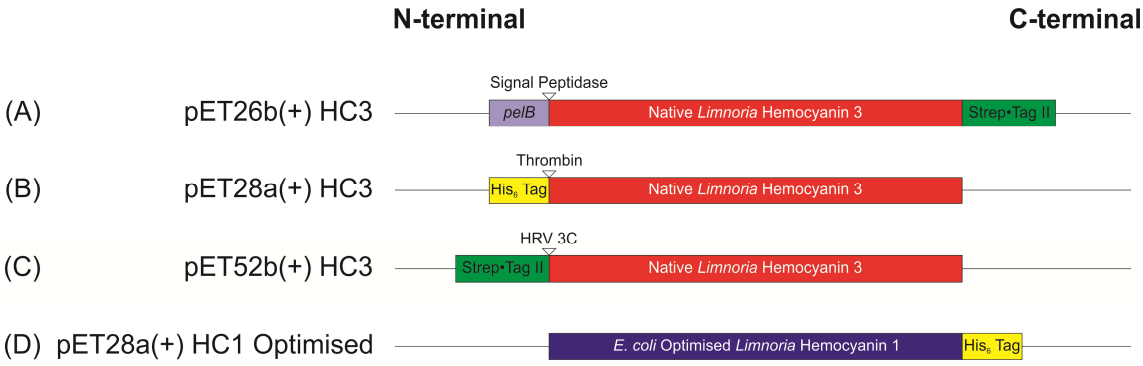



**Figure 5.6. Representation of the protein encoding regions of *E. coli* based HC expression vectors.**
(A) Vector included pelB leader sequence with natural signal peptidase cleavage site. PCR introduced Strep·Tag II tag. (B) Vector included $His_6$ tag with vector encoded thrombin cleavage site. (C) Vector included Strep.Tag II tag with vector encoded HRV 3C cleavage site. (D) *E. coli* codon optimised HC1 sequence. N-terminal vector encoded $His_6$ tag omitted by PCR. C-terminal non-cleavable vector encoded $His_6$ tag included.

Three different vectors were trialled each using the pET expression system. Those vectors were pET26b(+), pET28a(+) and pET52b(+) (see Figure 5.6). The pET28a(+) and pET52b(+) vectors are very similar to each other, and do not confer any advantage with regard to yielding more soluble expression of their controlled genes, they simply afford the use of different affinity purification technologies. The pET26b(+) vector, however, fuses a *pelB* signal sequence to the controlled gene which has the potential to locate the nascent protein to the periplasm of *E. coli* (232). The advantage of doing this is that *E. coli*'s periplasm is a naturally more oxidising environment than its cytoplasm and therefore should promote the formation of di-sulphide bonds which would enable proteins containing them to fold correctly. As the *Limnoria* HCs are predicted to contain di-sulphide bonds this may aid in the soluble expression of these proteins. However, when trialled with the HC3 gene the use of this vector did not result in an improved yield of soluble HC3 protein (Figure 5.7). It was not possible to obtain any pure or partially pure HC3 protein using this vector and so it was not possible to test HC3 produced in this way for enzyme activity (Figure 5.7).

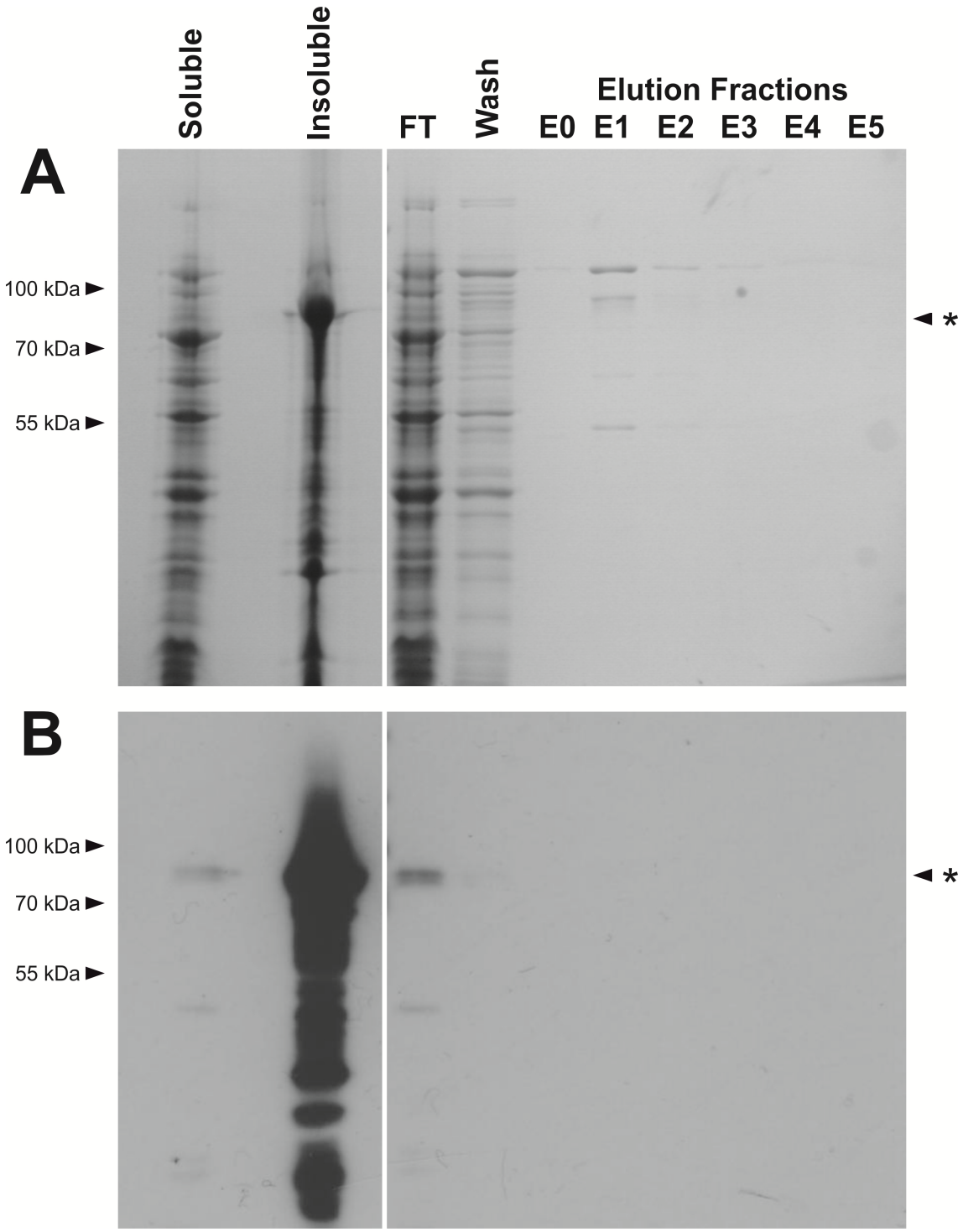



**Figure 5.7. Expression and purification of HC3 from a pET26 vector.**
Aliquots of soluble and insoluble cell lysate along with various fractions from a Strep·Tag II purification were analysed by Coomassie staining (A) and Western blot analysis (B). Rosetta-gami 2 cells were used to express HC3 from the pET26b(+)-HC3 vector. Cells were grown to an $OD_{600}$ of 1.0 before inducing expression with 1 mM IPTG and adjusting the growth temperature to 15 ℃ for 24 hours. Western blot produced using a monoclonal anti-polyHistidine-Peroxidase antibody. Expected molecular mass of pET26b(+)-HC3, 77.2 kDa, denoted by *. FT, column flow-through.

Expression and purification trials of HC3 using the pET28a(+) and pET52b(+) vectors showed no distinct and repeatable ability of one vector to cause more soluble expression of HC3. When purifying HC3 protein overexpressed using the pET52b(+) vector (Strep·Tag II purification), cleaner elution fractions were obtained with fewer co-purifying proteins present than when using IMAC purification and the pET28a(+) vector. However, Strep·Tag II purification produced lower yields of purified protein and still resulted in co-purifying proteins. The IMAC purification method resulted in slightly larger yields of purified protein. Elution using the Strep·Tag II purification system is binary in nature. A single concentration of a specific competitor for the binding site of the chromatography medium is applied which displaces bound tagged proteins. This system does not allow for much optimisation to try and reduce the levels of co-purifying contaminant proteins. Elution when using the IMAC purification procedure, however, is brought about by the gradual competitive displacement of several binding ligands per bound protein by imidazole. This can also be facilitated by a change in pH. This allows gradients of imidazole concentration or pH to be used to gradually remove contaminant proteins which have lower binding affinities for the chromatography medium than the target protein. For this reason it was decided to use the pET28a(+) vector for expressing *Limnoria* HC proteins.

Cultures were grown to a specific $OD_{600}$ before being induced to express their target protein. The $OD_{600}$ at which induction was performed was varied. If an overexpressed protein is toxic to or has a detrimental effect on the growth of *E. coli* then inducing the expression of such a protein will kill the bacteria in the culture and protein production will cease. In such cases it can be beneficial to induce expression when there are more cells present in the culture (i.e. at a higher $OD_{600}$) so that more protein is made before the toxic effects are seen. Cultures were induced to express HC3 at $OD_{600}$ of 0.5, 0.6 and 1.0. It was found that much more soluble protein was produced by the same volume of culture when induced at an $OD_{600}$ of 1.0 and this value was used in subsequent experiments (Figure 5.8).

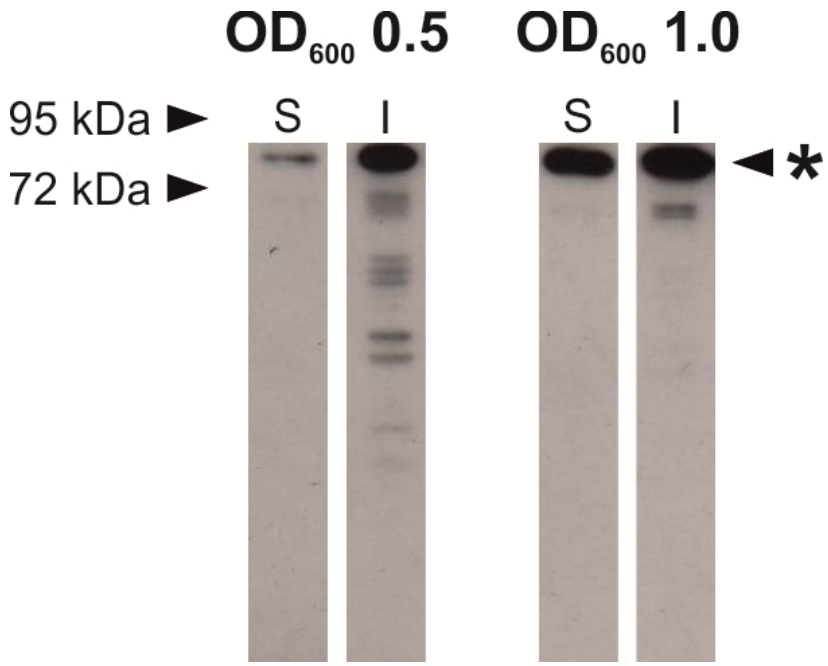



**Figure 5.8. The effect of $OD_{600}$ of a culture at the time of induction on the amount of heterologous HC3 produced.**
Western blots were produced from equal loadings (according to $OD_{600}$ at cell harvest) of insoluble and soluble fractions of pET28a(+)-HC3 expressing Rosetta-gami 2 cell lysates using a monoclonal anti-polyHistidine-Peroxidase antibody. Cells were induced with 1 mM IPTG and grown at 15 ℃ post induction fo r 3 hours. Expected molecular mass of pET28a(+)-HC3, 77.4 kDa, denoted by *. S, soluble fraction; I insoluble fraction.

Once cultures are induced to overexpress a protein fast growth is no longer important and so the temperature of the culture during expression can be changed. At lower temperatures cellular processes including protein expression are slowed down. When overexpressing a protein that has a tendency to fold incorrectly and form inclusion bodies, this quality can be beneficial, allowing more time for a protein to fold correctly and to form the correct disulphide bonds (233). Cultures were incubated at 37, 30, 20 and 15 ℃ whilst expressing HC3 post-induction with 1.0 mM IPTG. At 37 ℃ considerable insoluble expression but no soluble expression was seen at and below the expected molecular mass of HC3 when expressed in this system of 77.4kDa. As the temperature was lowered the amount of insoluble expression decreased, but no soluble expressed protein was observed until 20 ℃. At 20 ℃ a small amount of soluble expression was detected by Western blot, but considerably more was seen at 15 ℃. Therefore future expression studies incubated cells at 15 ℃ post induction.

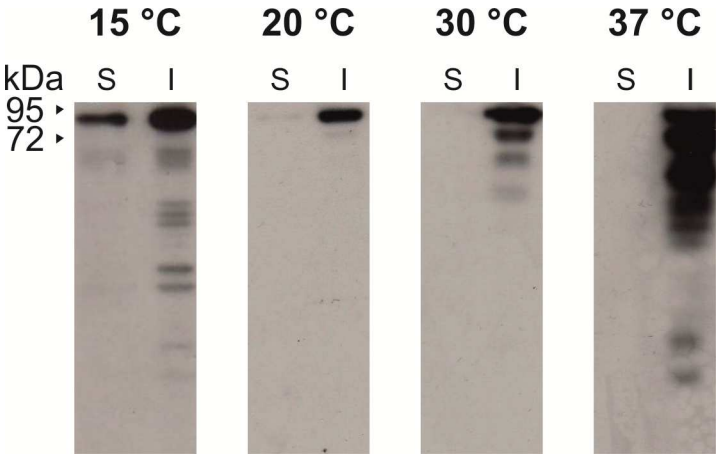



**Figure 5.9. The effect of temperature during protein expression on the solubility of heterologous HC3.**
Western blots were produced from equal loadings (according to $OD_{600}$ at cell harvest) of insoluble and soluble fractions of pET28a(+)-HC3 expressing Rosetta-gami 2 cell lysates using a monoclonal anti-polyHistidine-Peroxidase antibody. Expected molecular mass of pET28a(+)-HC3, 77.4 kDa. S, soluble fraction; I insoluble fraction.

After cultures have been induced to overexpress a protein the amount of time they are given to do so can affect the recovered yield of protein. Stable proteins that do not interfere with the cell's growth tend to accumulate with time, so a long expression period is beneficial. Unstable proteins, or proteins which form inclusion bodies can be degraded by proteases over time and proteins toxic to *E. coli* can result in cell death and so will degrade. In these cases a shorter amount of time for expression is preferred. Rosetta-gami 2 cells expressing HC3 were sampled at 3, 4, 5, 6, 8 and 21 hours post induction. It was found that 21 hours after induction no HC3 protein could be detected, soluble or insoluble (Figure 5.10). Between 3 and 8 hours post induction some soluble expression was observed with the amount detected by Western blot over these samples not varying, whilst the amount of insoluble expression seen decreased as the time post induction increased (Figure 5.10). Due to the timing practicalities of expression experiments, and as no perceptible benefit in using another time period could be seen, incubating cultures for 3 hours post induction was chosen as the best condition.

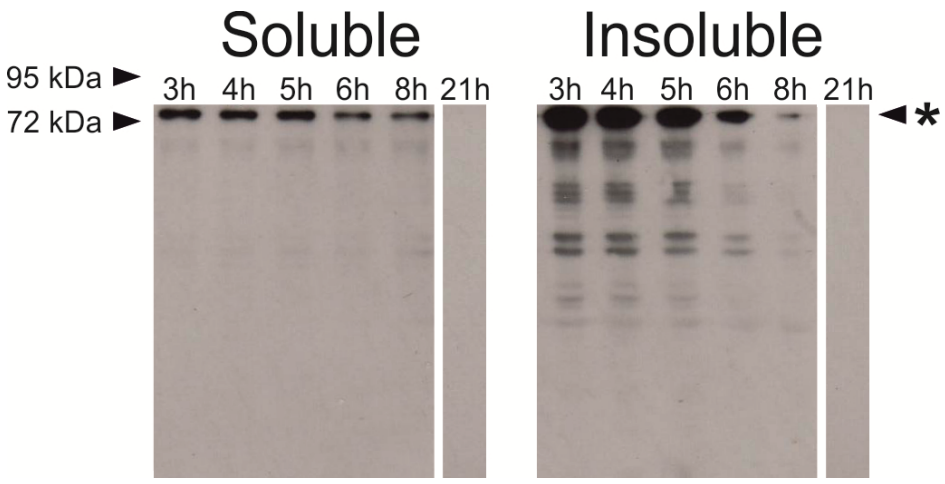



**Figure 5.10. The effect of the amount of growth time post-induction on the amount of heterologous HC3 produced.**
Western blots were produced from equal loadings (according to $OD_{600}$ at cell harvest) of insoluble and soluble fractions of pET28a(+)-HC3 expressing Rosetta-gami 2 cell lysates using a monoclonal anti-polyHistidine-Peroxidase antibody. Cells were induced with 1 mM IPTG and grown at 15 ℃ post induction fo r various amounts of time. Expected molecular mass of pET28a(+)-HC3, 77.4 kDa, denoted by *. S, soluble fraction; I insoluble fraction.

Whether or not the type of culture medium used whilst overexpressing proteins had an effect on the yield of soluble HC3 protein was investigated. Yields of HC3 protein obtained from cultures growing on Luria-Bertani medium or 2x YT broth were determined and were found to be the same by Western blot. Auto-induction media (Overnight Express Instant TB – Novagen) was also assessed for its ability to produce soluble HC3 expression; however, it was found to only result in degraded HC3 protein present in the insoluble cellular fraction. As a result of these findings 2x YT, the initial media used in other expression experiments, continued to be used.

As HCs are copper containing proteins it was thought that if the growth media of expression cultures did not contain sufficient quantities of copper then this might prevent overexpressed HC protein from folding correctly, or from being active if it was able to fold correctly. Copper chloride and copper sulphate at a final concentration of 1 mM were separately added to overexpressing cells to assay their effect on soluble HC3 levels. It was observed that the presence of either form of copper, or its absence, had no detectable effect on the amount of soluble HC3 produced. The presence of copper did not slow the growth rate of the bacteria. It was therefore decided that copper chloride should be included in the growth media during expression experiments as although it did not increase the yield of HC3 protein it may result in the HC3 produced being correctly folded and containing copper.

When using the pET expression system it is possible for a small amount of non-induced background expression to occur prior to induction. If the gene under the control of the pET system is toxic to the cell then this background expression could lead to a selection pressure being exerted on cells all the way from the first single transformed cell, through all subsequent growth stages. This selection pressure may be to mutate or deactivate the toxic gene to remove the selection pressure. As HCs in other species have been demonstrated to be potentially toxic to bacteria this may occur during attempts to express HC3 (186). To reduce the possibility of this negative selection all growth media from the initial transformation of the HC3 construct through to the expression culture were supplemented to a final concentration of 1 % w/v glucose, and cells were not allowed to reach stationary growth phase at any point to avoid the supply of glucose being depleted. *E. coli* will preferentially use glucose as a carbon source avoiding the production of cyclic adenosine monophosphate (cAMP) which is necessary for induction of expression in the pET system (234). This should therefore reduce the amount of non-induced expression via catabolite repression. When tested the inclusion of glucose in all growth media did not result in a greater yield of HC3 being expressed in a soluble form, but it did reduce some of the experiment to experiment variability seen during method development. Glucose was therefore included in all subsequent experiments.

Most of the heterologous expression work conducted in *E. coli* used the HC3 gene as a target gene. It was used in its native sequence form with one exception; the signal peptide portion of the gene was removed by PCR. Towards the end of the method development process a version of the HC1 gene was synthesized by GeneArt® (now part of Invitrogen – Life Technologies) with its sequence optimised for *E. coli*'s codon usage pattern. The optimised HC1 gene was cloned into the pET28 vector such as to have a C-terminal poly-histidine tag rather than an N-terminal tag. Expression of this construct using the parameters determined to be best by the work with HC3 resulted in a slight increase in the amount of soluble HC obtained.

As a result of the *E. coli* heterologous expression method development described above, the following expression procedure was settled upon as the best available method. The HC3 gene present in the pET28a(+) vector was transformed into Rosetta-gami 2 cells. The transformed cells were always grown in the presence of 1 % w/v glucose. A transformed colony was used to seed a starter culture which was grown to an $OD_{600}$ of 0.6 before harvesting and using the cells to inoculate a larger amount of 2x YT media supplemented to 1 mM $CuCl_2$. Once the culture had reached an $OD_{600}$ of 1.0

the cells were induced to express HC3 by adjusting the culture to contain 1 mM IPTG and placed in an incubator at 15 ℃ shaking at 180 rpm. After incubating for 3 hours under these conditions the cells were harvested, processed and HC3 purified from the cell lysate using IMAC. An example elution profile, SDS-PAGE and Western blot analysis of such an expression and purification is shown in Figure 5.11.

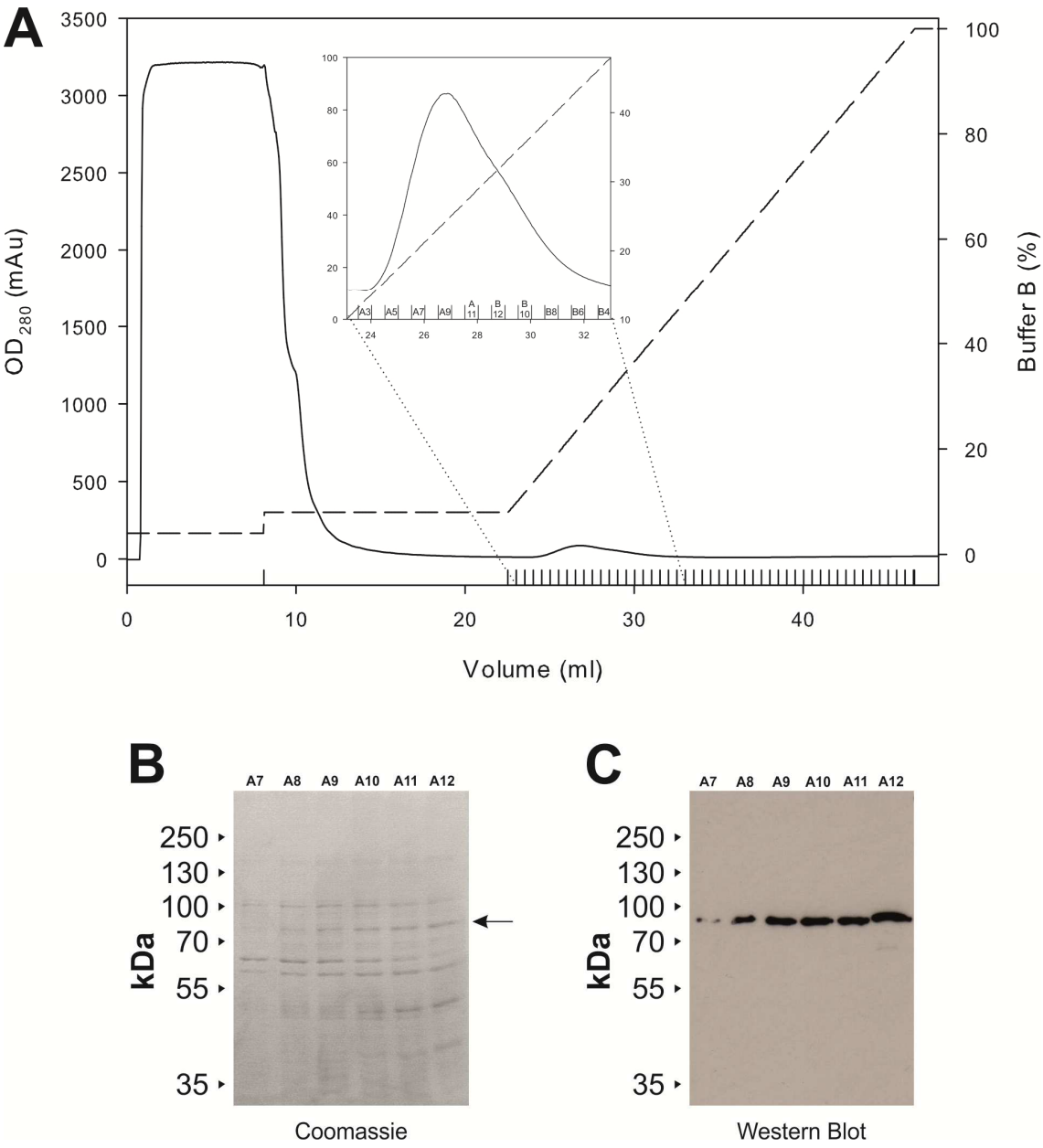



**Figure 5.11. Purification profile and analysis of IMAC purified heterologous HC3 protein.**
(A) Example IMAC purification profile of HC3. Solid line, $OD_{280}$ trace; dashed line, buffer B percentage. (A-Inset) Magnification of elution region of the purification profile. (B) Coomassie stained SDS-PAGE analysis of elution fractions A7 – A12. Band indicated by arrow in A12 was subjected to gel band protein identification. (C) Western blot of elution fractions A7 – A12 using anti-hemocyanin primary antibody and anti-

rabbit-HRP secondary antibody. Expected molecular mass of heterologous HC3, 77.4 kDa.

Using this method a small elution peak can be observed in the 280 nm absorbance trace of the purification profile (Figure 5.11 – A-INSET). Analysing by SDS-PAGE the fractions which make up such a peak shows that the peak does not represent a single protein species, but rather many proteins of various sizes (Figure 5.11 – B). A band is visible between 70 – 100 kDa, which is consistent with the expected molecular mass of HC3 (77.4 kDa), was found to be HC3 by protein identification (indicated with an arrow). This band is the only band present which generates a signal in a corresponding Western blot created using an anti-hemocyanin antibody (Figure 5.11 – C). This indicates that the other bands observed when using Coomassie staining are not HC3 or its degradation products, but are co-purifying contaminant proteins. Therefore, the expression and purification procedure developed here results in partially purified fractions of HC3 protein.

Partially purified fractions of HC3 obtained during protein expression method development were subjected to DPO enzyme activity assays as described in Section 2.6.1. None of these assays resulted in activity being observed.

#### *5.3.2.2 ArcticExpress Escherichia coli*

It has already been mentioned that one tactic to favour the production of soluble heterologous protein in the *E. coli* expression system is to lower the temperature at which the expression stage is carried out. However, at temperatures of 15 ℃ or lower *E. coli*'s chaperone proteins such as GroEL and DnaK are much less efficient at aiding the folding of partially folded proteins (235). To overcome this problem the ArcticExpress strain of *E. coli* has been engineered to co-express two chaperone proteins, Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira antarctica*, which have been shown to possess high protein refolding activities at such low temperatures (236).

Working in conjunction with Mrs Luisa Elias (University of York) an *E. coli* codon optimised version of the *Limnoria* HC1 gene was cloned into the pET28a(+) vector and heterologously overexpressed in ArcticExpress (DE3) RIL cells. Cells expressing the protein were harvested and processed as described earlier before purifying HC1 from the cellular lysate. A representative purification profile along with SDS-PAGE and

Western blot analysis of fractions from various stages of the purification from 1 L of cell culture are shown in Figure 5.12.

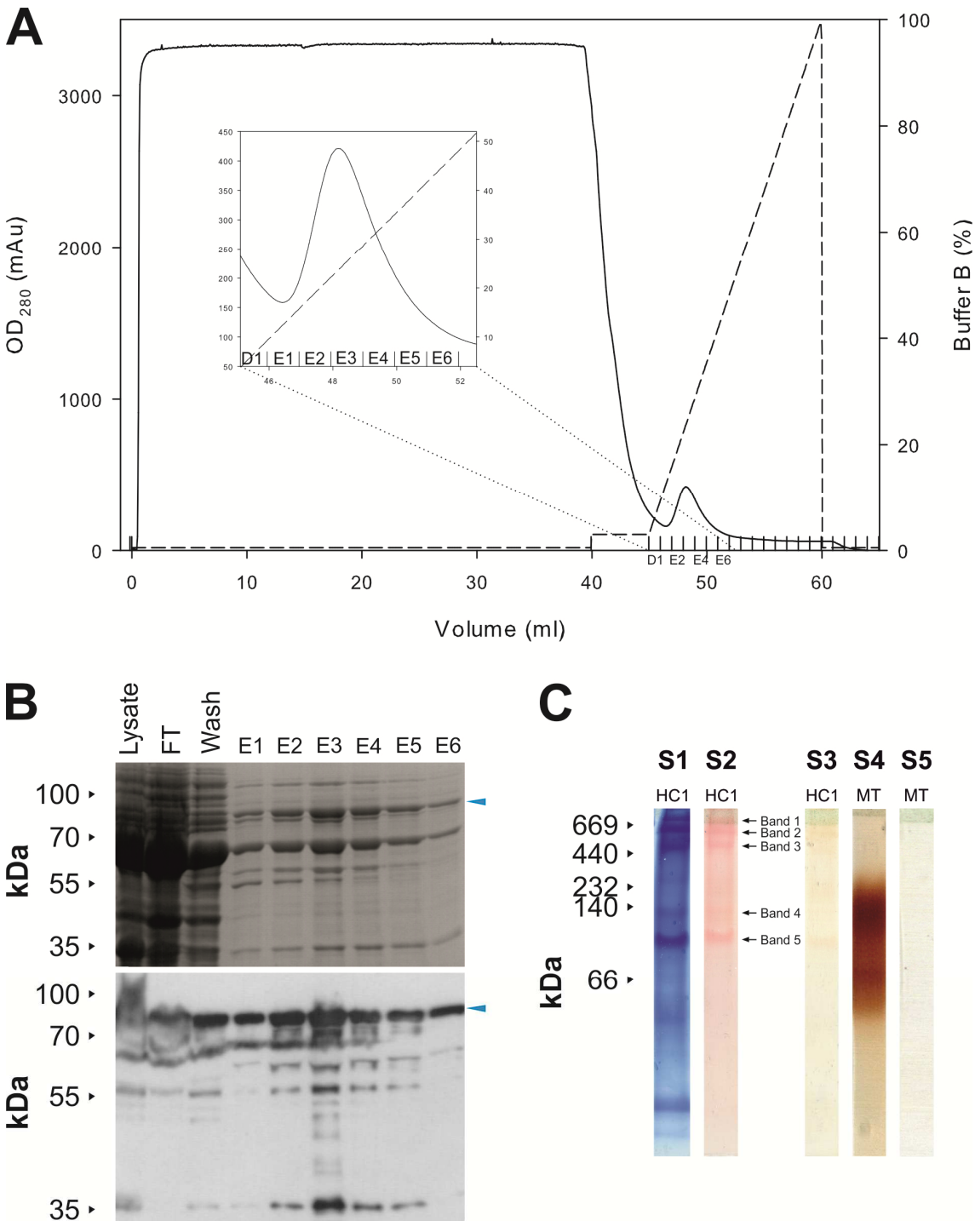



**Figure 5.12. Purification profile, PAGE analysis and DPO enzyme activity assay of IMAC purified heterologous HC1 protein produced in ArcticExpress *E. coli* cells.**
(A) Example IMAC purification profile of HC1. Solid line, $OD_{280}$ trace; dashed line, buffer B percentage. (A-Inset) Magnification of elution region of the purification profile. (B-upper) Coomassie stained SDS-PAGE analysis of various purification fractions and

elution fractions E1 – E6. (B-lower) Western blot of of various purification fractions and elution fractions E1 – E6 using anti-hemocyanin primary antibody and anti-rabbit-HRP secondary antibody. Expected molecular mass of heterologous HC1, 78.6 kDa, blue arrows. (C) PAGE gel based DPO enzyme activity assay. (S1) Coomassie stained. (S2) DPO assay. Bands 1 – 5 were excised and subjected to gel band protein identification. (S3) Pre-treated with the DPO inhibitor kojic acid prior to DPO assay. (S4) MT, mushroom tyrosinase positive control enzyme, DPO assay. (S5) MT pre-treated with the DPO inhibitor kojic acid prior to DPO assay.

The absorbance trace at 280 nm in the purification profile shows a peak in absorbance across the elution fractions E1 – E6 (Figure 5.12 A-INSET). These fractions were sampled for SDS-PAGE and Western blot analysis. The Coomassie stained SDS-PAGE gel in Figure 5.12 – B-UPPER shows several dominant bands in the peak fractions E1 – E6. From previous identification, one of the bands seen in a double band at an apparent molecular mass between 70 – 100 kDa, indicated with a blue arrow, is likely to represent HC1 (expected mass 78.6 kDa). The corresponding Western blot, carried out with an anti-hemocyanin antibody, seems to support this showing a strong signal at a mass between 70 – 100 kDa (blue arrow). The Western blot also indicates that several of the lower molecular weight bands seen in the Coomassie stained gel are HC1 degradation products, although exactly which bands these are is not possible to pin point. As there are more bands visible in the Coomassie stained gel than there are signals detected by the Western blot, this shows that there are still other co-purifying proteins. This indicates that the purification of HC1 from a lysate of HC1 overexpressing cells has been successful in removing a majority, but not all other proteins present in the lysate.

Elution fractions E1 – E6 were pooled and concentrated from 6 ml to 0.5 ml as described in Section 2.5.2. This pooled sample was buffer exchanged into assay buffer (Section 2.5.3) before mixing with a suitable amount of 4x Native-PAGE Sample Buffer and running 30 µl aliquots on a Native PAGE gel. HC1 containing PAGE lanes were subjected to in-gel DPO activity staining and Coomassie staining and the results are shown in Figure 5.12 – C. The positive control mushroom tyrosinase (MT) lane shown in strip S4 of that figure indicates that the DPO assay has worked correctly as a heavily brown/red pigmented band is seen. Also, strip S5 of the same figure shows that when gel lanes are pre-incubated in the PO inhibitor kojic acid before DPO activity staining the MT positive control is unable to produce staining demonstrating that these conditions are sufficient to allow inhibition to occur.

The Coomassie stained lane of concentrated partially pure HC1 shown in Figure 5.12 – C S1 shows several dominant bands spread across the whole mass range, as well as much smearing of signal between these bands. This confirms what has already been shown in Figure 5.12 – B, that the sample contains co-purifying contaminants. Strip S1 also shows that several of these bands may represent high molecular weight complexes (between 440 – 669 kDa). It is possible that these bands could represent hexameric HC1 which would be expected to have a mass of 471.6 kDa. However, due to the nature of native PAGE these bands may simply be proteins which possess little overall charge at the pH that the gel was run at and consequently have not migrated far. Strip S2 from the same figure shows that when stained for DPO activity the concentrated HC1 fraction produces five reddish pigmented bands indicating activity. When pre-incubated with kojic acid (Strip S3) these activity bands show greatly reduced intensity. The five bands showing DPO enzyme activity in strip S2 were excised and the proteins within those bands identified as described in Section 2.6.2. The results are shown in Table 5.5.

No single protein is identified as being present in all five bands, however HC1 is found in four of the five bands.

**Table 5.5. Proteins identified in gel bands displaying DPO activity.**
A buffer exchanged, concentrate of purification fractions from heterologously expressing HC1 ArcticExpress cells was used in a PAGE based DPO assay. Protein bands displaying DPO activity were subjected to protein identification and the results are shown. A * indicates a single peptide match.

| Band No. | Proteins Identified | Species |
|---|---|---|
| 1 | UDP-L-Ara4N formyltransferase/UDP-Glc C-4'-decarboxylase | *Escherichia coli* |
| | Hemocyanin 1 * | *Limnoria quadripunctata* |
| 2 | 60 kDa chaperonin | *Oleispira antartica* |
| | UDP-L-Ara4N formyltransferase/UDP-Glc C-4'-decarboxylase | *Escherichia coli* |
| | Chaperonin GroEL * | Many possible species |
| | Hemocyanin 1* | *Limnoria quadripunctata* |
| 3 | UDP-L-Ara4N formyltransferase/UDP-Glc C-4'-decarboxylase | *Escherichia coli* |
| | Hemocyanin 1 * | *Limnoria quadripunctata* |
| 4 | Alkyl hydroperoxide reductase, F52a subunit | *Escherichia coli* |
| | Molecular chaperone DnaK * | Many possible species |
| | Formyltetrehydrofolate deformylase * | *Escherichia coli* |
| | Hemocyanin 1 * | *Limnoria quadripunctata* |
| 5 | Glucoseamine-fructose-6-phosphate aminotransferase | *Escherichia coli* |

## 5.4 Discussion

The work described in this thesis has formed part of a larger project of enzyme discovery from *Limnoria*. Throughout the course of the project attempts have been made to heterologously express many different *Limnoria* genes, all of which were mentioned as being of interest in Chapter 2 of this work. Specifically this has included a single member of each of the glycosyl hydrolase (GH) families 5, 9, 30 and 35, three members of family 7, one leucine rich repeat (LRR) gene and four of the five HC genes.

With only one exception all of these genes have proven very difficult to express. A gene belonging to the GH 30 family was successfully expressed in *E. coli* resulting in soluble protein which purified and demonstrated xylosidase activity *in vitro* (unpublished data). Success in expressing several other genes has been achieved, however, requiring much more time and effort. Three members of the GH 7 family were eventually produced in a soluble, active form in *Aspergillus sp.* after attempts using six other expression systems (61).

Work to heterologously express members of the GH 5, 9 and 35 families in *E. coli* has so far only resulted in partially soluble protein and no activity has currently been unequivocally demonstrated. As well as the GH genes, work has been carried out to express an LRR gene using the *E. coli*, *P. pastoris*, *Aspergillus sp.*, baculovirus mediated insect cell and human cell expression systems. To date either no protein expression has been observed, or protein is only expressed in an insoluble, inactive form.

Outside of the work described in this chapter, attempts to heterologously express HC protein in eukaryotic expression systems have resulted in much the same outcome as has been seen with LRR expression attempts. Work was carried out by another group member and by our project's industrial partner, Novozymes, to express a codon optimised version of the HC1 gene in *Asergillus niger* and *Aspergillus oryzae*. In all attempts it was possible to detect transcription of the HC1 gene, but no protein production.

Use of a human cell expression system to try and heterologously express a non-codon optimised version of HC1 resulted in the production of insoluble protein only. This

finding was largely mirrored in the work described in this chapter using the baculovirus mediated insect cell expression system to express HC3. Despite being able to successfully produce recombinant, viable virus, it was only possible to achieve intracellular, insoluble expression of a slightly truncated form of HC3 in insect cells when using a relatively high MOI of virus. The protein was not successfully secreted into the culture medium in measureable amounts as was expected. Little work was done to optimise expression in this system for several reasons. Firstly, as already mentioned the HC3 protein produced appeared not to be full length by SDS-PAGE analysis. Compared to an expression system like *E. coli* or *Aspergillus sp.*, there are relatively few parts of the insect cell system which can be varied to try and improve the likelihood of successfully achieving soluble, secreted expression. Of these variables, only the changing of the expressing insect cell line is relatively easy to carry out. Other variables such as the promoter used to drive expression, the secretion signal peptide used, or codon optimising the gene for this system may be able to influence the production of more soluble, secreted protein. However to test these variables would require the re-cloning of the HC3 gene into different transfer vectors and reproducing and amplifying virus. This process could take up to two months to complete and would be a time consuming and costly endeavour. As some soluble expression had already been seen using the *E. coli* expression system it was decided not to invest further time into the insect cell system.

After some initial optimisation of the *E. coli* strain used, the $OD_{600}$ at which expression was induced and the temperature at which expression was carried out, it was possible to obtain some soluble expression of HC3 using *E. coli*. However, any further efforts to increase the yield of soluble expression resulted in marginal or un-certain improvements. There were also, despite meticulous care, frequently issues regarding the reproducibility of the findings. Conditions found to produce good yields of soluble HC3 protein in one experiment could almost completely fail to yield any soluble expression in a repeat of those conditions. Also, attempts to recover more soluble protein by pooling the bacteria from several growth flasks failed, commonly resulting in less protein recovered than would have been expected from a single flask. These same issues of reproducibility and yield were also witnessed by a colleague working separately to express the HC2 and HC4 genes in *E. coli* (unpublished data). It seems, therefore, that the successful expression and/or recovery of HC protein using the *E. coli* expression system may be dependent on the control of exacting growth conditions, such as the dissolved oxygen content or pH of the growth media during expression, which are beyond our ability to control in our laboratory. Whether or not this is so could

be investigated by carrying out the expression in a small scale (5 L) fermenter vessel in which such variables can be monitored and actively controlled. It is possible though that there is another unknown condition or cofactor which could be the cause of this un-reproducibility. It is worth noting that there is currently no literature evidence to suggest successful heterologous expression of an arthropodan hemocyanin.

When assayed, soluble protein produced using the *E. coli* expression system was unable to demonstrate DPO activity. This could be because of several reasons. Firstly, the soluble protein produced may still be improperly folded, having perhaps the wrong network of disulphide bonds or lacking the copper atom co-factors suggested to be present in Chapter 4. It could also simply be that HC3 is not capable of carrying out a DPO reaction, or at least not under the conditions tested in the DPO assay used.

Use of the ArcticExpress strain of *E. coli* along with its set of manufacturer recommended growth conditions did result in slightly higher yields of soluble HC1 protein being obtained compared to the other *E. coli* systems tested. It also reduced some of the inter-experiment variability seen, although occasions did still occur when little or no HC1 protein was recoverable from the bacterial cells. The most significant advantage of this system over the other *E. coli* approaches documented is that the soluble protein obtained was able to demonstrate DPO activity, albeit relatively weakly in comparison to the mushroom tyrosinase positive control enzyme. The signal generated by the positive control enzyme was visible after ~5 minutes of incubation whilst the HC1 generated signal appeared after an overnight incubation. It would not be correct under the circumstances to make a direct comparison between the amount of or rate of activity seen between these two proteins; however, it is possible that the speed of development of the signal caused by HC1 is as slow as it is due to the same reasons mentioned above as to why soluble HC3 protein was inactive. It may be that only a small proportion of the soluble HC1 protein obtained is correctly folded and therefore capable of carrying out a DPO activity. It may also simply be that HC1 only possesses a weak capability to carry out a DPO reaction. Another possibility that must be taken into account is that the conditions used in the DPO assay may not be optimal for allowing HC1 to carry out a DPO activity. The protein may require a different buffer or pH to operate at; it may require a different substrate or method of activation. Without a sufficient quantity of heterologously expressed pure protein to subject to other analyses it is difficult to say. It also appears that the known DPO inhibitors PTU and kojic acid are unable to completely inhibit the DPO activity of HC1. This may be due to

the active site of HC1 not being able to optimally bind these inhibitor molecules resulting in incomplete inhibition.

Heterologous, soluble HC protein produced either using the standard *E. coli* expression system, or using ArcticExpress cells has also proved difficult to effectively purify away from their background of host proteins. Whilst it was possible to achieve a marginally higher level of purity with HC protein produced bearing a Strep·Tag II tag this resulted in lower HC protein yields and co-purifying proteins were still present. Therefore, IMAC purification of poly-histidine tagged HC was preferred as it resulted in higher yields of HC protein. During the method development phase it was repeatedly found that the *E. coli* chaperone proteins DnaK and GroEL co-purified with HC. Whilst this is a useful indicator that the *E. coli* cells are actively trying to aid the folding of HC protein it does nevertheless constitute a contaminant of the purified HC protein. In an attempt to reduce the amount of these chaperones that co-purify with HC, a dissociation step was built into the method. This involved incubating the soluble cell lysate with $MgSO_4$ and ATP prior to purification to facilitate the dissociation of the chaperone proteins (224, 225). Inclusion of this step was able to reduce, but not eliminate the amount of chaperone protein co-purifying with HC and so was used in further experiments.

In summary, a method has been developed to successfully heterologously express soluble HC protein for which it has been possible to show evidence of DPO activity. Unfortunately, the most successful method developed is not capable of producing enough HC protein of sufficient purity to permit detailed biochemical and physical characterisation of the HC proteins at this point. More work is required to improve the efficiency of the expression system used to produce larger quantities of protein. Approaches to do this could include the use of *E. coli* vectors incorporating solubility enhancing fusion proteins. Alternatively the use of an actively controlled fermenter may allow the growth conditions of *E. coli* during the expression phase to be precisely monitored and controlled.

# Chapter 6: Final Discussion and Future Perspectives

## 6.1 Introduction

Second generation biofuels offer our modern, global society a valuable potential source of “drop-in” replacement transportation fuels that could help to reduce our dependence on fossil oil. In order to realise the potential offered by second generation biofuels their method of production must be made more efficient. There are several barriers to overcome in order to do this including improving the digestibility of the lignocellulose feedstock, improving the amount of soluble monomeric sugars which can be released from the feedstock, and increasing the amount of ethanol that can be produced during fermentation of those sugars. The marine wood borer, *Limnoria quadripunctata*, has found a way to overcome the second of these barriers and is able to breakdown lignocellulose to release enough sugar to allow it to thrive and reproduce. Most notably it is able to do this without the aid of symbiotic or mutualistic microorganisms. The work described in this study sought to shed light on how the animal achieves this, and to begin the exploitation of that knowledge to improve the process of industrially breaking down lignocellulosic material for biofuel production. The approach taken to do this involved first identifying potential digestion related proteins via a proteomic analysis of the animal’s digestive system, followed by the *in vivo* and *in vitro* characterisation of one group of identified target proteins.

The digestive system of *Limnoria* is arranged as a two compartment system. Ingested wood is held and broken down in the sclerotized gut, whilst enzyme secretion and nutrient absorption is thought to occur in the hepatopancreas (HP) organ which is lined with secretory cells and is attached to the gut, posterior to the stomach. Wood particles are prevented from entering the HP by two filter channels. The transcriptome of the HP has been previously determined and it was found to be dominated by only a few classes of genes; namely the glycosyl hydrolases (GHs), hemocyanins (HCs), leucine rich repeat proteins (LRRs), fatty acid binding proteins (FABPs) and ferritins. This arrangement leads to the hypothesis that the gut of *Limnoria* is akin to an enzyme bioreactor with an enzyme cocktail added to it from the HP to breakdown the wood particles before withdrawing the digested material and enzymes back to the HP to allow enzyme recycling and nutrient uptake. This is unlike the digestive systems of

other lignocellulose feeding animals such as termites, beetles and cows which are thought of more as microbial bioreactors or fermenters (58, 125, 237, 238).

## 6.2 Discussion

To determine if the classes of genes abundantly transcribed in the HP are found in the gut, and how abundantly the corresponding proteins are expressed, a label free semi quantitative proteomic analysis was successfully performed on both gut and hepatopancreas (HP) tissue from *Limnoria*. The analysis revealed that the protein products of all major classes of genes found in the HP transcriptome were found in both the gut and HP proteome. Quantitative PCR (qPCR) analysis revealed that the HCs, GHs, LRRs and ferritin were only transcribed in the HP of *Limnoria*, not any other part of the body. This means that those proteins have been transcribed and translated in the HP and then moved into the gut. This finding supports the hypothesis that proteins required for digestion of lignocellulose are produced in the secretory HP organ, and are passed into the gut to carry out their function. Determination of the full length gene sequences of the FABP genes revealed that they did not possess secretion signal sequences. It was also found by qPCR that they were expressed in both the gut and HP. This data means that the FABP genes do not fit the hypothesis described above, and as they are produced intracellularly in both the gut and HP are unlikely to contribute directly to lignocellulose breakdown in *Limnoria*.

Glycosyl hydrolase protein was found much more abundantly in the HP than in the gut. As the function of a GH is to depolymerise sugars it is highly likely they are involved in lignocellulose digestion. However, faecal pellets from *Limnoria* have been found to have approximately the same composition as the comparable non-digested wood. Glycosyl hydrolase attack would release specific saccharides from ingested wood, leaving the resulting faecal pellets enriched in certain types of sugars which were not targeted by the available suite of GH enzymes used. This therefore suggests that GH attack is not the primary method of digestion happening to ingested wood particles and may explain why they are not found abundantly in the gut. It may instead be that the wood is attacked by a non-specific mechanism first, liberating small soluble oligosaccharides that are able to be withdrawn back into the HP where they could be broken down into monomers by GHs.

The non-specific method of attack may be oxidative in nature as has been documented amongst the white rot fungi (93). It could be possible that the HCs are responsible for mediating this oxidative attack (196). Hemocyanins from other arthropods have been shown to have a dual role; oxygen carrier and phenoloxidase (PO). Quinones, the reaction products of POs, are known to be toxic to bacteria and this may be a reason for the sterile conditions seen in *Limnoria*'s gut. Quinones themselves, via a system of redox cycling with divalent transition metal ions may also be able to generate reactive oxygen species (ROS). If this redox cycling is happening in the gut of *Limnoria* it suggests another reason as to why the gut is devoid of microorganisms. It might also offer a method of non-specific oxidative attack of the lignocellulose in the gut. To determine if HCs may be able to carry out these roles they were investigated *in vivo* and *in-vitro*.

As has already been stated the HCs fit the hypothesis described above to be involved in digestion in *Limnoria* as they are solely transcribed in the HP, but their protein is found in the gut. As well as this, their amino acid sequences show good identity with those of HCs from other arthropods which have been shown to be able to carry out PO enzyme reactions; including conservation of the main features likely to be necessary for function.

Indeed a component present in soluble *Limnoria* extract (*ex vivo*) is able to carry out a di-phenoloxidase (DPO) activity. It is possible to greatly inhibit this activity by reducing and alkylating cysteine residues present in the soluble extract, suggesting that the component resulting in activity is a correctly folded, di-sulphide bond containing protein. Furthermore it is possible to partially inhibit this activity using the PO inhibitor phenylthiourea (PTU). Protein identification has implicated HC as the agent responsible for this activity.

Attempts to heterologously express sufficient quantities of HC protein for characterisation have proven very difficult. Some soluble protein has been successfully produced and partially purified from the ArcticExpress strain of *E. coli* cells. The partially purified fractions were able to demonstrate DPO activity which was to a great extent inhibited by the use of the PO inhibitor, kojic acid. Once again protein identification has implicated that HC is associated with this activity although further investigation with suitable controls is necessary to be certain.

Taken together, the *ex vivo* and *in vitro* protein DPO activity assays imply that *Limnoria* HC protein has DPO-like enzyme activity. Preliminary work carried out *in vivo* suggests that a PO enzyme activity may be associated with the production of peroxides in the gut of *Limnoria*. Injection of the peroxide detection reagent AmplexRed into healthy animals showed the presence of peroxides in the gut of *Limnoria*. Animals pre-incubated with PTU prior to assaying however showed greatly reduced levels of peroxide detection. This early data suggests that a PO enzyme activity, which could be caused by HC acting as a DPO enzyme, is responsible for the creation of peroxides in the digestive system of *Limnoria*. This could potentially provide a source of ROS which could non-specifically attack the structure of lignocellulose, disrupting the lignin structure which protects the hemicellulose and cellulose microfibrils, breaking from them small soluble oligosaccharides.

## 6.3 Future Perspectives

Whilst it has been possible in this work to show a protein species in *Limnoria* capable of carrying out a DPO activity, the highly complex mixture of proteins present gave only limited confidence that the activity is indeed solely caused by HC. This is also true, although to a lesser extent, for semi-pure heterologously expressed HC. Ideally, to be certain that it is indeed HC, and only HC, which is carrying out the DPO activity the assay would be performed using a homogeneous, pure sample of heterologous HC.

In order to do this the expression and purification procedures for HC will need to be improved. Most purification media work optimally when the applied source of protein contains enough of the target protein to be purified to saturate the available binding sites. In order to do this with HC the level of heterologous expression will need to be increased. An increased level of expression and effectiveness of the purification procedure will also be required to generate large enough quantities of sufficiently pure protein for physical and biochemical characterisation.

The HC proteins have proven very difficult to express. Within this work have been discussed many different expression systems and approaches tested to try and express high yields of HC protein in a soluble form. So far the *E. coli* expression system has been the most promising used for HC proteins. Further avenues to explore using the *E. coli* system in an aid to improve the soluble expression of HC proteins are the use of solubility enhancing polypeptide tags such as the glutathione-S-transferase

(239) and maltose-binding protein (240) tags or the use of tags designed to aid with the correct folding of their attached polypeptides such as DsbA, DsbC and thioredoxin (241-243). It is possible also that the co-expression of more than one *Limnoria* HC protein in a system may favour correct folding of HC protein by favouring correct heterohexamer formation over protein aggregation.

Larger available quantities of pure, heterologous protein would allow a liquid cuvette/microplate based DPO activity assay to be developed. Such an assay was developed during the course of this project using mushroom tyrosinase and hemocyanin purified from *Limulus polyphemus* (Sigma Aldrich) as positive control enzymes. However, the resulting assay proved incompatible with *Limnoria* extract and crude bacterial cell lysates. Such an assay would provide a much quicker, more convenient method to allow many more variables to be tested and kinetic data to be collected. Access to larger amounts of pure HC protein would also allow detailed physical characterisation to be performed. Size exclusion chromatography – multi-angle laser light scattering (SEC-MALLS) could be used to determine the accurate molecular mass of heterologously expressed HC, revealing its oligomerisation state in solution. Electron paramagnetic resonance (EPR) could reveal whether or not copper is bound inside HC as theorised. The 3D structure of *Limnoria* proteins could be determined with the use of X-ray crystallography. These could be compared to those of previously characterised HCs from other species to scrutinise what differs between them, and what they have in common. It could also aid in the explanation of biochemical data through a deeper understanding of how substrates or inhibitors interact with the proteins.

In order to determine if HCs behaving as POs are participating in the breakdown of lignocellulose in the digestive process of *Limnoria* attempts should be made to determine if such behaviour is able to influence the levels of saccharification possible. Pure, heterologously expressed HC could be included in a standard saccharification assay using powdered wood as a lignocellulose source, along with cofactors such as $Fe_{(II)}$ or a di-phenol substrate that might be required for activity, to determine if it is able to increase the amount of sugar solubilised from wood. After such a treatment, the amount of lignin remaining in the sample could be compared with that of a non-HC treated control to determine if HC was able to cause the loss of lignin (244). Oxidative damage caused by HC treatment may not result in a reduction in the amount of lignin in a sample, but rather simply modify the lignin revealing more access to the cellulose and hemicellulose fractions. Spectrophotometric assays and mass spectrometry

analysis could be used to determine what, if any, changes had been made to the lignin (245).

Preliminary findings detailed in this work showed the production of peroxides in the gut of *Limnoria*, and the inhibition of this production in the presence of PTU. A more detailed investigation of this phenomenon should be carried out to confirm whether this finding is sound by examining more animals. Animals should also be examined at different stages of their lifecycle as immediately before, and during the moulting process animals stop feeding. This may have a bearing on whether or not peroxides are detected as during and immediately after moulting the animals should be actively sclerotizing their new cuticle to protect themselves; this process is likely to involve PO activity.

## 6.4 Final Summary

Second generation biofuels offer the potential to greatly reduce our reliance on petroleum based transportation fuels; but to realise this potential we must find ways to make their production more efficient and cost effective. Barriers faced in science have often already been solved in nature, and so in this case it makes sense to look to nature to uncover how it has found ways to complete the carbon cycle – recycling the nutrients from wood back into carbon dioxide, water and minerals for the production of the next generation of plants. The marine wood borer, *Limnoria quadripunctata*, has evolved a way to use the oligosaccharide material locked-up inside the structure of lignocellulose as a primary source of nutrition. In doing so it has developed ways to overcome the problems of breaking past the water proof lignin layer of lignocellulose, and to liberating cellulose from crystalline cellulose fibres. Understanding how this is done may inspire new approaches to be exploited in industrial processes to improve the yield of soluble sugars from the saccharification step of second generation biofuel production. The work described in this study has contributed towards the beginning of this understanding.

# Appendix A – Label-Free Semi-Quantitative Proteomic Analysis Data

Summary of the label-free semi-quantitative proteomic data obtained from analysis of ***Limnoria* HP tissue**. Protein accession values relate to the designations of contigs produced from the Limnoria HP transcriptome (68) or gene names of fully cloned genes. EMPAI, exponentially modified protein abundance index; % Molar Fraction, molar fraction percentage.

| Protein Accession | EMPAI | % Molar Fraction | Category | Uniprot ID | Top *Sensible* BLAST Hit (Search Performed on 06/01/2012) |
|---|---|---|---|---|---|
| 454limoContig11301 | 0.47 | 0.67 | ATP metabolic process | E0W6V9 | F1F0-ATP synthase beta subunit [Marsupenaeus japonicus] |
| 454limoContig3909 | 0.83 | 1.18 | ATP metabolic process | D3W7E4 | mitochondrial ATP synthase subunit alpha precursor [Litopenaeus va |
| 454limoContig463 | 0.32 | 0.46 | ATP metabolic process | Q2F5T3 | ATP synthase [Bombyx mori] >gb\|ABD36284.1\| ATP synthase [Bomb |
| 454limoContig9119 | 0.65 | 0.92 | ATP metabolic process | D3W7E4 | mitochondrial ATP synthase subunit alpha precursor [Litopenaeus vannamei] |
| 454limoContig10536 | 0.22 | 0.31 | ATP metabolic process | D6WNL2 | PREDICTED: similar to ATP synthase delta chain, mitochondrial [Trib |
| 454limoContig5235 | 0.69 | 0.98 | ATP metabolic process | A0T2T1 | mitochondrial ATP synthase subunit [Cyclorana alboguttata] |
| 454limoContig4594 | 0.13 | 0.18 | ATP metabolic process | B4NBL2 | GK11166 [Drosophila willistoni] >gb\|EDW81176.1\| GK11166 [Drosoph |
| 454limoContig110 | 0.43 | 0.61 | Carbohydrate binding | E2CV66 | lectin A isoform 4 [Marsupenaeus japonicus] |
| 454limoContig9667 | 0.56 | 0.80 | Carbohydrate metabolism | D0NA42 | xylose isomerase 1 [Phytophthora infestans T30-4] >gb\|EEY54296.1 |
| 454limoContig7445 | 0.33 | 0.47 | Carbohydrate metabolism | Q1HQU8 | transaldolase [Aedes aegypti] |
| 454limoContig3080 | 0.12 | 0.17 | Carbohydrate metabolism | B0WIN7 | peritrophic membrane chitin binding protein [Culex quinquefasciatus] |
| 454limoContig5018 | 0.38 | 0.54 | Carbohydrate metabolism | E9HGX0 | mitochondrial malate dehydrogenase [Daphnia pulex] |
| 454limoContig7630 | 0.13 | 0.18 | Carbohydrate metabolism | F4WT75 | Pyruvate carboxylase, mitochondrial [Acromyrmex echinatior] |
| 454limoContig8815 | 0.12 | 0.17 | Carbohydrate metabolism | Q9U0S1 | Phosphoenolpyruvate carboxykinase [Nephrops norvegicus] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig795 | 0.2 | 0.28 | Carbohydrate metabolism | E2S087 | fructose 1,6-bisphosphatase [Marsupenaeus japonicus] |
| 454limoContig4903 | 0.12 | 0.17 | Carbohydrate metabolism | Q1HQU8 | transaldolase [Aedes aegypti] |
| 454limoContig8276 | 0.08 | 0.11 | Cytochrome-C oxidase | Q68EI6 | Cyc1 protein [Danio rerio] |
| 454limoContig5402 | 1.97 | 2.80 | Cytoskeleton | O96657 | actin 1 [Penaeus monodon] |
| 454limoContig11020 | 1.5 | 2.13 | Cytoskeleton | D6NTM4 | actin [Marsupenaeus japonicus] >gb\|ADH93988.1\| actin [Marsupena |
| 454limoContig2943 | 1.86 | 2.65 | Cytoskeleton | G4VXL0 | beta-actin [Onthophagus binodis] >gb\|AEI91061.1\| beta-actin [Onth |
| 454limoContig7621 | 0.2 | 0.28 | Cytoskeleton | C7TZK1 | Actin-2 [Schistosoma japonicum] |
| 454limoContig10913 | 0.32 | 0.46 | Cytoskeleton | A7KCY8 | alpha-tubulin [Heliconius melpomene] |
| 454limoContig10753 | 0.18 | 0.26 | Cytoskeleton | O96657 | actin 1 [Penaeus monodon] |
| 454limoContig5067 | 0.55 | 0.78 | Cytoskeleton | Q25009 | RecName: Full=Tubulin beta-1 chain; AltName: Full=Beta-I tubulin |
| 454limoContig1269 | 0.6 | 0.85 | Cytoskeleton | Q9DFT2 | alpha tubulin [Notothenia coriiceps] |
| 454limoContig4628 | 0.24 | 0.34 | Cytoskeleton | D4P8F7 | myosin light chain [Marsupenaeus japonicus] |
| 454limoContig11083 | 0.3 | 0.43 | Cytoskeleton | O61379 | RecName: Full=Tropomyosin; AltName: Full=Allergen Pan s I; |
| 454limoContig5641 | 0.21 | 0.30 | Cytoskeleton | Q964E1 | RecName: Full=Actin, cytoplasmic; Flags: Precursor >gb\|AAK68713.1 |
| 454limoContig3472 | 0.29 | 0.41 | Cytoskeleton | Q8I6V4 | beta-I tubulin [Callinectes sapidus] |
| 454limoContig7805 | 0.33 | 0.47 | Cytoskeleton | B8YDT8 | alpha-tubulin [Tetranychus cinnabarinus] |
| 454limoContig3259 | 0.28 | 0.40 | Cytoskeleton | Q7YZU2 | myosin heavy chain [Alpheus lottini] |
| 454limoContig3664 | 0.26 | 0.37 | Cytoskeleton | A9CSJ9 | tropomyosin [Oratosquilla oratoria] |
| 454limoContig311 | 0.29 | 0.41 | DNA binding | B9ENT8 | Histone H3.3 [Salmo salar] |
| 454limoContig9788 | 0.32 | 0.46 | DNA binding | Q3S349 | heat shock protein 70 [Pachygrapsus marmoratus] |
| 454limoContig2617 | 0.29 | 0.41 | DNA binding | A8IDE5 | heat shock protein 70 [Diplodon chilensis] |
| 454limoContig10216 | 0.1 | 0.14 | DNA binding | Q3T007 | HSPCA protein [Bos taurus] |
| 454limoContig4185 | 0.34 | 0.48 | DNA binding | B4E485 | heat shock cognate 70kDa protein [Paraceradocus gibber] |
| 454limoContig2734 | 0.31 | 0.44 | DNA binding | C5I773 | small ubiquitin-like modifier-1 [Litopenaeus vannamei] |
| 454limoContig9947 | 0.39 | 0.55 | DNA binding | C3KHE1 | ADP-ribosylation factor 1 [Anoplopoma fimbria] |
| FABP1(SC1)_seq | 0.76 | 1.08 | FABP | - | - |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig11839 | 0.82 | 1.17 | Ferritin | G5DFE0 | ferritin peptide [Fenneropenaeus indicus] |
| 454limoContig5306 | 0.97 | 1.38 | Ferritin | G5DFE0 | ferritin peptide [Fenneropenaeus indicus] |
| 454limoContig1071 | 0.95 | 1.35 | Ferritin | D3YHP5 | ferritin 3 [Eriocheir sinensis] |
| 454limoContig76 | 0.69 | 0.98 | Glycolysis | D7RF64 | glyceraldehyde 3-phosphate dehydrogenase [Eriocheir sinensis] >gb\| |
| 454limoContig10709 | 1.65 | 2.35 | Glycolysis | D7RF64 | glyceraldehyde 3-phosphate dehydrogenase [Eriocheir sinensis] >gb |
| 454limoContig10199 | 0.2 | 0.28 | Glycolysis | A8CWB3 | putative fructose-bisphosphate aldolase [Lutzomyia longipalpis] |
| 454limoContig485 | 0.49 | 0.70 | Glycolysis | Q6QWP6 | 2-phospho-D-glycerate hydrolase [Callinectes sapidus] |
| 454limoContig2206 | 0.14 | 0.20 | Glycolysis | O96656 | phosphopyruvate hydratase [Penaeus monodon] |
| GH5A | 0.37 | 0.53 | Glycosyl hydrolase (5) | D5IEF5 | - |
| 454limoContig7212 | 0.07 | 0.10 | Glycosyl hydrolase (5) | D5IEF7 | GH5 family protein GH5E [Limnoria quadripunctata] |
| GH5E | 0.17 | 0.24 | Glycosyl hydrolase (5) | D5IEF7 | - |
| GH5C | 0.08 | 0.11 | Glycosyl hydrolase (5) | D5IEF6 | - |
| GH7B | 2.43 | 3.46 | Glycosyl hydrolase (7) | D4HRL0 | - |
| 454limoContig4910 | 1.09 | 1.55 | Glycosyl hydrolase (7) | D4HRL0 | GH7 family protein [Limnoria quadripunctata] |
| 454limoContig2963 | 0.69 | 0.98 | Glycosyl hydrolase (7) | D4HRL0 | GH7 family protein [Limnoria quadripunctata] |
| GH7A | 1.18 | 1.68 | Glycosyl hydrolase (7) | D4HRK9 | - |
| 454limoContig11642 | 0.89 | 1.27 | Glycosyl hydrolase (7) | D4HRK9 | GH7 family protein [Limnoria quadripunctata] |
| 454limoContig11385 | 0.66 | 0.94 | Glycosyl hydrolase (7) | D4HRK9 | GH7 family protein [Limnoria quadripunctata] |
| GH7C | 0.21 | 0.30 | Glycosyl hydrolase (7) | D4HRL1 | - |
| GH9D | 0.98 | 1.39 | Glycosyl hydrolase (9) | - | - |
| GH9A | 0.57 | 0.81 | Glycosyl hydrolase (9) | D4HRL2 | - |
| 454limoContig81 | 0.36 | 0.51 | Glycosyl hydrolase (9) | D4HRL2 | GH9 family protein [Limnoria quadripunctata] |
| GH9B | 0.75 | 1.07 | Glycosyl hydrolase (9) | D4HRL3 | - |
| GH9H | 0.2 | 0.28 | Glycosyl hydrolase (9) | - | - |
| GH9I | 0.21 | 0.30 | Glycosyl hydrolase (9) | - | - |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig6507 | 0.19 | 0.27 | Glycosyl hydrolase (9) | D4HRL4 | GH9 family protein [Limnoria quadripunctata] |
| GH9F | 0.21 | 0.30 | Glycosyl hydrolase (9) | D4HRL4 | - |
| 454limoContig731 | 0.28 | 0.40 | Glycosyl hydrolase (9) | D4HRL2 | GH9 family protein [Limnoria quadripunctata] |
| 454limoContig4137 | 0.07 | 0.10 | Glycosyl hydrolase (9) | D4HRL2 | GH9 family protein [Limnoria quadripunctata] |
| 454limoContig4926 | 0.12 | 0.17 | Glycosyl hydrolase (9) | D4HRL4 | GH9 family protein [Limnoria quadripunctata] |
| 454limoContig11268 | 0.21 | 0.30 | Glycosyl hydrolase (16) | Q70EW4 | beta-1,3-glucan binding protein [Homarus gammarus] |
| 454limoContig10898 | 0.06 | 0.09 | Glycosyl hydrolase (18) | C4P6W4 | chitinase 4 precursor [Litopenaeus vannamei] |
| 454limoContig9008 | 0.54 | 0.77 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig4242 | 0.3 | 0.43 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] >gb\|EEC06943 |
| 454limoContig2926 | 0.41 | 0.58 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig11319 | 0.23 | 0.33 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig11358 | 0.49 | 0.70 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig6835 | 0.26 | 0.37 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig2536 | 0.21 | 0.30 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig8614 | 0.52 | 0.74 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig8807 | 0.08 | 0.11 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig10032 | 0.26 | 0.37 | Glycosyl hydrolase (30) | B7Q4X6 | glucosylceramidase, putative [Ixodes scapularis] >gb\|EEC13898.1\| |
| 454limoContig8542 | 0.07 | 0.10 | Glycosyl hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig7686 | 0.32 | 0.46 | Glycosyl hydrolase (31) | Q9NFY8 | alpha glucosidase [Litopenaeus vannamei] |
| 454limoContig10250 | 0.1 | 0.14 | Glycosyl hydrolase (31) | Q9NFY8 | alpha glucosidase [Litopenaeus vannamei] |
| 454limoContig8306 | 0.08 | 0.11 | Glycosyl hydrolase (31) | Q9NFY8 | alpha glucosidase [Litopenaeus vannamei] |
| 454limoContig11638 | 0.54 | 0.77 | Glycosyl hydrolase (35) | E1ZZY3 | Beta-galactosidase [Camponotus floridanus] |
| 454limoContig2406 | 0.47 | 0.67 | Glycosyl hydrolase (35) | E1ZZY3 | Beta-galactosidase [Camponotus floridanus] |
| 454limoContig10939 | 0.19 | 0.27 | Glycosyl hydrolase (35) | B7PW84 | beta-galactosidase precursor, putative [Ixodes scapularis] |
| 454limoContig11070 | 0.38 | 0.54 | Glycosyl hydrolase (35) | P16278 | beta-galactosidase related protein precursor [Homo sapiens] |
| HC2 | 0.23 | 0.33 | Hemocyanin | D5J6V7 | - |

| | | | | | |
|---|---|---|---|---|---|
| HC1 | 0.23 | 0.33 | Hemocyanin | D5J6V6 | - |
| HC3 | 0.09 | 0.13 | Hemocyanin | D5J6V8 | - |
| HC7 | 0.04 | 0.06 | Hemocyanin | - | - |
| 454limoContig2246 | 0.16 | 0.23 | Hydrolase | B2MV88 | lysozyme [Anopheles dirus A] |
| 454limoContig7664 | 0.41 | 0.58 | Kinase | Q004B5 | arginine kinase [Litopenaeus vannamei] |
| 454limoContig10972 | 0.57 | 0.81 | Kinase | Q9NH49 | RecName: Full=Arginine kinase; Short=AK >gb\|AAF43436.1\|AF233355 |
| 454limoContig3594 | 0.23 | 0.33 | Kinase | E9CJF0 | ephrin type-B receptor 3 [Capsaspora owczarzaki ATCC 30864] |
| 454limoContig2163 | 0.23 | 0.33 | Kinase | C3VUU0 | arginine kinase [Fenneropenaeus merguiensis] |
| LRR1(SC3)_seq | 1.35 | 1.92 | LRR | - | - |
| LRR3(SC8)_seq | 0.81 | 1.15 | LRR | - | - |
| LRR2(SC5)_seq | 0.48 | 0.68 | LRR | - | - |
| LRR10(SC10)_seq | 0.91 | 1.29 | LRR | - | - |
| LRR4(SC1)_seq | 0.39 | 0.55 | LRR | - | - |
| 454limoContig8724 | 0.19 | 0.27 | Metabolic process | Q6PHI8 | transketolase-like protein 2 [Danio rerio] >gb\|AAH56536.1\| Tkt pro |
| 454limoContig705 | 0.17 | 0.24 | Metabolic process | A3EXY7 | putative succinate-CoA ligase, GDP-forming, alpha subunit [Maconel |
| 454limoContig5108 | 0.31 | 0.44 | Metabolic process | Q17CT0 | transketolase [Aedes aegypti] >gb\|EAT44192.1\| transketolase |
| 454limoContig2241 | 0.28 | 0.40 | Metabolic process | D3TM09 | transketolase [Glossina morsitans morsitans] |
| 454limoContig4741 | 0.06 | 0.09 | Metabolic process | D6WMI3 | PREDICTED: similar to adenosylhomocysteinase [Tribolium castaneum |
| 454limoContig8195 | 0.15 | 0.21 | Metabolic process | B0W8J6 | isocitrate dehydrogenase [Culex quinquefasciatus] >gb\|EDS39046.1\| |
| 454limoContig2472 | 0.15 | 0.21 | Metabolic process | P32929 | cystathionine gamma-lyase [Homo sapiens] >dbj\|BAD97096.1\| |
| 454limoContig7879 | 0.18 | 0.26 | Metabolic process | Q7ZTH6 | trifunctional enzyme subunit beta, mitochondrial [Danio rerio] |
| 454limoContig11143 | 0.1 | 0.14 | Metal binding | D3PHS5 | Beta-crystallin A1 [Lepeophtheirus salmonis] |
| 454limoContig5611 | 0.15 | 0.21 | Metal binding | D3PIA3 | Gamma-crystallin A [Lepeophtheirus salmonis] |
| 454limoContig7537 | 0.09 | 0.13 | Metal binding | D3PHC6 | Gamma-crystallin-1 [Lepeophtheirus salmonis] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig892 | 0.13 | 0.18 | Metal binding | B7Q9K4 | CDGSH iron sulfur domain-containing protein, putative [Ixodes scapu |
| 454limoContig11817 | 0.34 | 0.48 | No BLAST hits | - | No BLAST hits |
| 454limoContig11670 | 0.24 | 0.34 | No BLAST hits | - | No BLAST hits |
| 454limoContig5492 | 0.3 | 0.43 | No BLAST hits | - | No BLAST hits |
| 454limoContig910 | 0.08 | 0.11 | Oxidoreductase | C4N5V0 | Trx1 [Eriocheir sinensis] |
| 454limoContig11743 | 0.07 | 0.10 | Oxidoreductase | Q6GR65 | glutathione peroxidase 3 a precursor [Xenopus laevis] |
| 454limoContig7619 | 0.18 | 0.26 | Oxidoreductase | Q6GR65 | glutathione peroxidase 3 a precursor [Xenopus laevis] |
| 454limoContig1 | 0.34 | 0.48 | Oxidoreductase | Q16R01 | alcohol dehydrogenase [Aedes aegypti] >gb\|EAT36819.1\| |
| 454limoContig12082 | 0.41 | 0.58 | Oxidoreductase | B4M3M4 | GJ18924 [Drosophila virilis] >gb\|EDW65399.1\| GJ18924 [Drosophila |
| 454limoContig4100 | 0.2 | 0.28 | Oxidoreductase | D3TLF0 | malate dehydrogenase [Glossina morsitans morsitans] |
| 454limoContig7800 | 0.22 | 0.31 | Oxidoreductase | F5BZU2 | isocitrate dehydrogenase 1 (NADP+), soluble [Epinephelus bruneus |
| 454limoContig11726 | 0.36 | 0.51 | Oxidoreductase | D6WVW0 | PREDICTED: similar to diacetyl/L-xylulose reductase [Tribolium castan |
| 454limoContig7511 | 0.35 | 0.50 | Oxidoreductase | B7PCD2 | NADP-dependent isocitrate dehydrogenase, putative [Ixodes scapula |
| 454limoContig11711 | 0.11 | 0.16 | Oxidoreductase | Q16X08 | aldehyde dehydrogenase [Aedes aegypti] >gb\|EAT39148.1\| |
| 454limoContig5571 | 0.23 | 0.33 | Oxidoreductase | D5HRH5 | NADH dehydrogenase subunit 4 [Nicrophorus insignis] |
| 454limoContig4409 | 0.15 | 0.21 | Oxidoreductase | C6K6G3 | peroxiredoxin [Fenneropenaeus indicus] |
| 454limoContig11751 | 0.19 | 0.27 | Oxidoreductase | Q5U5D2 | dehydrogenase/reductase (SDR family) member 11 [Xenopus laevis] |
| 454limoContig10789 | 0.3 | 0.43 | Protease | Q27824 | serine collagenase 1 precursor [Celuca pugilator] |
| 454limoContig9656 | 0.2 | 0.28 | Protease | P09954 | carboxypeptidase A1 precursor [Sus scrofa] >sp\|P09954.2\|CBPA1_P |
| 454limoContig11886 | 0.23 | 0.33 | Protease | Q27824 | serine collagenase 1 precursor [Celuca pugilator] |
| 454limoContig1101 | 0.41 | 0.58 | Protease | D0VYM6 | chymotrypsin-like proteinase [Marsupenaeus japonicus] |
| 454limoContig5119 | 0.25 | 0.36 | Protease | O18488 | Chymotrypsin BII [Litopenaeus vannamei] |
| 454limoContig3839 | 0.15 | 0.21 | Protease | D2Y5C3 | trypsin 4 [Panulirus argus] |
| 454limoContig2130 | 0.1 | 0.14 | Protease | C1IHU0 | trypsin [Marsupenaeus japonicus] |
| 454limoContig2305 | 0.12 | 0.17 | Protease | P04069 | RecName: Full=Carboxypeptidase B >prf\|\|1004229A CPase B |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig7612 | 0.27 | 0.38 | Protease | E3TGS7 | cathepsin A [Eriocheir sinensis] |
| 454limoContig8267 | 0.11 | 0.16 | Protease | E3TGS7 | cathepsin A [Eriocheir sinensis] |
| 454limoContig10366 | 0.2 | 0.28 | Protease | A8Q617 | Peptidase family M13 containing protein [Brugia malayi] |
| 454limoContig4239 | 0.08 | 0.11 | Protease | Q8UUK1 | carboxypeptidase A1 preproprotein [Gallus gallus] |
| 454limoContig10008 | 0.07 | 0.10 | Protease | E3TGS7 | cathepsin A [Eriocheir sinensis] |
| 454limoContig3756 | 0.1 | 0.14 | Protease | C1IHU0 | trypsin [Marsupenaeus japonicus] |
| 454limoContig11645 | 0.13 | 0.18 | Protease | O18487 | Chymotrypsin BI [Litopenaeus vannamei] |
| 454limoContig6067 | 0.05 | 0.07 | Protease | Q1JPJ2 | xaa-Pro aminopeptidase 1 [Bos taurus] >sp\|Q1JPJ2.1\|XPP1_BOVIN |
| 454limoContig12121 | 0.15 | 0.21 | Protein binding | F4WWD9 | Endoplasmin [Acromyrmex echinatior] |
| 454limoContig6112 | 0.08 | 0.11 | Protein binding | A0PJ13 | 14-3-3-like protein [Penaeus monodon] |
| 454limoContig9834 | 0.59 | 0.84 | Protein binding | D2DSH6 | chaperonin 10 [Scylla paramamosain] |
| 454limoContig5323 | 0.15 | 0.21 | Protein binding | F5A6C1 | trap [Procambarus clarkii] |
| 454limoContig8722 | 0.21 | 0.30 | RNA binding | E9H141 | Endoribonuclease-like protein [Daphnia pulex] |
| 454limoContig7779 | 0.69 | 0.98 | RNA binding | E2AE08 | RNA-binding protein squid [Camponotus floridanus] |
| 454limoContig7558 | 0.12 | 0.17 | RNA binding | B7PYD1 | ATP-dependent RNA helicase, putative [Ixodes scapularis] |
| 454limoContig3142 | 0.27 | 0.38 | Transferase | E0WBN2 | glutathione S-transferase [Eriocheir sinensis] |
| 454limoContig2512 | 0.16 | 0.23 | Transferase | E7D176 | methyltransferase-like protein [Latrodectus hesperus] |
| 454limoContig2905 | 0.11 | 0.16 | Transferase | A9QUN5 | delta class glutathione S-transferase [Blattella germanica] |
| 454limoContig4065 | 0.17 | 0.24 | Transferase | E0W3E2 | UDP-glucuronosyltransferase 2B7 precursor, putative [Pediculus hum |
| 454limoContig5130 | 0.09 | 0.13 | Transferase | E2BKD5 | Sulfotransferase 1C4 [Harpegnathos saltator] |
| 454limoContig43 | 0.09 | 0.13 | Transferase | A9QUN5 | delta class glutathione S-transferase [Blattella germanica] |
| 454limoContig8578 | 0.16 | 0.23 | Translation | C6GBE8 | elongation factor 2 [Homarus americanus] |
| 454limoContig10078 | 0.66 | 0.94 | Translation | B3GD96 | ubiquitin/60s ribosomal protein L40 fusion [Caenorhabditis brenneri] |
| 454limoContig11853 | 0.31 | 0.44 | Translation | D2K241 | elongation factor-1 alpha [Cyamus gracilis] |
| 454limoContig7139 | 0.22 | 0.31 | Translation | Q2I3E8 | ribosomal protein L8 [Litopenaeus vannamei] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig10870 | 0.17 | 0.24 | Translation | D6R722 | 40S ribosomal protein S15 [Hypophthalmichthys molitrix] |
| 454limoContig7130 | 0.18 | 0.26 | Translation | Q86PL3 | 40S ribosomal protein [Perinereis aibuhitensis] |
| 454limoContig3328 | 0.18 | 0.26 | Translation | D2K241 | elongation factor-1 alpha [Cyamus gracilis] |
| 454limoContig2319 | 0.28 | 0.40 | Translation | A9QQ40 | 60s ribosomal protein L14 [Lycosa singoriensis] |
| 454limoContig7423 | 0.14 | 0.20 | Translation | Q9BNX9 | elongation factor-2 [Armadillidium vulgare] |
| 454limoContig10421 | 0.19 | 0.27 | Translation | F5A6E0 | ribosomal protein S11 [Procambarus clarkii] |
| 454limoContig2826 | 0.41 | 0.58 | Translation | P48149 | ribosomal protein S15Aa, isoform D [Drosophila melanogaster] |
| 454limoContig5371 | 0.34 | 0.48 | Translation | D7F2K9 | ribosomal protein S3 [Palaemonetes varians] |
| 454limoContig1792 | 0.24 | 0.34 | Translation | E3T0W9 | elongation factor 1-alpha [Litopenaeus vannamei] |
| 454limoContig8064 | 0.35 | 0.50 | Translation | A6N9Y7 | 60s ribosomal protein L18 [Ornithodoros parkeri] |
| 454limoContig2032 | 0.15 | 0.21 | Translation | B2ZPN7 | protein-disulfide isomerase [Scylla paramamosain] |
| 454limoContig10014 | 0.34 | 0.48 | Translation | B5M6E5 | 60S ribosomal protein L27 [Haplopelma schmidti] |
| 454limoContig5644 | 0.28 | 0.40 | Translation | B7QI86 | 60S ribosomal protein L7A, putative [Ixodes scapularis] >gb\|EEC18 |
| 454limoContig1706 | 0.26 | 0.37 | Translation | Q6PBK3 | 40S ribosomal protein S28 [Danio rerio] |
| 454limoContig10923 | 0.23 | 0.33 | Translation | E2J6Y5 | ribosomal protein S25 [Hyalomma marginatum rufipes] |
| 454limoContig6919 | 0.11 | 0.16 | Translation | C4PL18 | ribosomal protein L3 [Penaeus monodon] |
| 454limoContig4965 | 0.2 | 0.28 | Translation | E7DZ29 | ribosomal protein S4 [Euphydryas aurinia] |
| 454limoContig1527 | 0.22 | 0.31 | Translation | Q4PM11 | 40S ribosomal protein S13 [Ixodes scapularis] |
| 454limoContig4684 | 0.38 | 0.54 | Translation | Q70MM6 | ribosomal protein S3 [Crassostrea gigas] |
| 454limoContig2667 | 0.33 | 0.47 | Translation | F5ANJ0 | cyclophilin A [Eriocheir sinensis] |
| 454limoContig11518 | 0.22 | 0.31 | Translation | Q5MIR0 | 60S ribosomal protein L11 [Aedes albopictus] |
| 454limoContig5058 | 0.11 | 0.16 | Translation | E2BXV5 | 60S ribosomal protein L6 [Harpegnathos saltator] |
| 454limoContig11668 | 0.21 | 0.30 | Translation | B5FXQ5 | putative ribosomal protein S16 [Taeniopygia guttata] |
| 454limoContig2403 | 0.17 | 0.24 | Translation | F5A6E1 | ribosomal protein S18 [Procambarus clarkii] |
| 454limoContig1924 | 0.32 | 0.46 | Translation | B8XVP3 | 40S ribosomal protein S14 [Ochlerotatus taeniorhynchus] |
| 454limoContig8397 | 0.31 | 0.44 | Translation | F5A6E3 | ribosomal protein S6 [Procambarus clarkii] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig4104 | 0.21 | 0.30 | Translation | E2J6V8 | 60S ribosomal protein L23 [Hyalomma marginatum rufipes] |
| 454limoContig5394 | 0.17 | 0.24 | Transport | D9IG67 | voltage-dependent anion-selective channel [Eriocheir sinensis] |
| 454limoContig11077 | 0.33 | 0.47 | Transport | Q0PWR8 | putative solute carrier family 25 member 3 [Pimephales promelas] |
| 454limoContig9991 | 0.51 | 0.73 | Transport | G1FEX1 | adenine nucleotide translocase 2 [Litopenaeus vannamei] |
| 454limoContig2560 | 0.25 | 0.36 | Transport | G1FEX1 | adenine nucleotide translocase 2 [Litopenaeus vannamei] |
| 454limoContig10280 | 0.22 | 0.31 | Transport | D9IG67 | voltage-dependent anion-selective channel [Eriocheir sinensis] |
| 454limoContig11896 | 0.15 | 0.21 | Transport | Q6P5K6 | mitochondrial carnitine/acylcarnitine carrier protein [Danio rerio] |
| 454limoContig5631 | 0.12 | 0.17 | Transport | Q2KJ85 | sodium/bile acid cotransporter [Bos taurus] >gb\|AAI05472.1\| |
| 454limoContig10513 | 0.32 | 0.46 | Uncharacterised | E9GU36 | hypothetical protein DAPPUDRAFT_321849 [Daphnia pulex] |
| 454limoContig3032 | 0.09 | 0.13 | Uncharacterised | B4QMM9 | GD14140 [Drosophila simulans] >gb\|EDX09783.1\| GD14140 |
| 454limoContig65 | 0.16 | 0.23 | Uncharacterised | C3ZQG3 | hypothetical protein BRAFLDRAFT_185990 [Branchiostoma floridae] |
| 454limoContig8690 | 0.54 | 0.77 | Uncharacterised | E9H0B2 | hypothetical protein DAPPUDRAFT_199694 [Daphnia pulex] |
| 454limoContig11256 | 0.09 | 0.13 | Uncharacterised | E4XAI7 | unnamed protein product [Oikopleura dioica] |
| 454limoContig542 | 0.1 | 0.14 | Uncharacterised | Q5TMY7 | AGAP012165-PA [Anopheles gambiae str. PEST] >gb\|EAA43295.4\| |
| 454limoContig6421 | 0.28 | 0.40 | Uncharacterised | Q0C669 | unnamed protein product [Hyphomonas neptunium ATCC 15444] |
| 454limoContig11191 | 0.16 | 0.23 | Uncharacterised | C3Z4V8 | hypothetical protein BRAFLDRAFT_225607 [Branchiostoma floridae] |
| 454limoContig6376 | 0.08 | 0.11 | Uncharacterised | E9GJ71 | hypothetical protein DAPPUDRAFT_304103 [Daphnia pulex] |
| 454limoContig11239 | 0.14 | 0.20 | Uncharacterised | E9FXF2 | hypothetical protein DAPPUDRAFT_42213 [Daphnia pulex] |
| 454limoContig6720 | 0.1 | 0.14 | Uncharacterised | B3LV26 | GF16990 [Drosophila ananassae] >gb\|EDV42498.1\| GF16990 |

Summary of the label-free semi-quantitative proteomic data obtained from analysis of ***Limnoria* Gut tissue**. Protein accession values relate to the designations of contigs produced from the Limnoria HP transcriptome (68) or gene names of fully cloned genes. EMPAI, exponentially modified protein abundance index; % Molar Fraction, molar fraction percentage.

| Protein Accession | EMPAI | % Molar Fraction | Category | Uniprot ID | Top *Sensible* BLAST Hit (04/01/2012) |
|---|---|---|---|---|---|
| 454limoContig11301 | 1.62 | 2.62 | ATP metabolic process | E0W6V9 | F1F0-ATP synthase beta subunit [Marsupenaeus japonicus] |
| 454limoContig9119 | 0.46 | 0.74 | ATP metabolic process | D3W7E4 | mitochondrial ATP synthase subunit alpha precursor [Litopenaeus vannamei] |
| 454limoContig3909 | 1.24 | 2.00 | ATP metabolic process | D3W7E4 | mitochondrial ATP synthase subunit alpha precursor [Litopenaeus va |
| 454limoContig463 | 0.75 | 1.21 | ATP metabolic process | Q2F5T3 | ATP synthase [Bombyx mori] >gb\|ABD36284.1\| ATP synthase [Bomb |
| 454limoContig5235 | 0.69 | 1.11 | ATP metabolic process | A0T2T1 | mitochondrial ATP synthase subunit [Cyclorana alboguttata] |
| 454limoContig5555 | 0.47 | 0.76 | ATP metabolic process | E0W2X4 | vacuolar ATP synthase subunit E, putative [Pediculus humanus c |
| 454limoContig10536 | 0.22 | 0.36 | ATP metabolic process | D6WNL2 | PREDICTED: similar to ATP synthase delta chain, mitochondrial [Trib |
| 454limoContig4594 | 0.12 | 0.19 | ATP metabolic process | Q1HPX4 | H+ transporting ATP synthase gamma subunit [Bombyx mori] |
| 454limoContig8800 | 0.34 | 0.55 | ATP metabolic process | F4WCI9 | V-type proton ATPase catalytic subunit A [Acromyrmex echinatior] |
| 454limoContig4832 | 0.14 | 0.23 | ATP metabolic process | Q6PC35 | ATPase, H+ transporting, lysosomal, V1 subunit H [Danio rerio] >gb\| |
| 454limoContig3913 | 0.2 | 0.32 | ATP metabolic process | D7F2K3 | ATP synthase subunit g [Palaemonetes varians] |
| 454limoContig7020 | 0.11 | 0.18 | ATP metabolic process | Q0PXU6 | putative ATP synthase subunit d [Diaphorina citri] |
| 454limoContig11031 | 0.14 | 0.23 | ATP metabolic process | B5U263 | mitochondrial hydrogen-transporting ATP synthase coupling factor F |
| 454limoContig5732 | 0.16 | 0.26 | ATP metabolic process | A9LMH5 | mitochondrial F0-ATP synthase b-chain [Litopenaeus vannamei] |
| 454limoContig4114 | 0.09 | 0.15 | ATP metabolic process | E2BB53 | Vacuolar ATP synthase subunit S1 [Harpegnathos saltator] |
| 454limoContig3946 | 0.12 | 0.19 | ATP metabolic process | D6WA56 | PREDICTED: similar to H+ transporting ATP synthase O subunit |
| 454limoContig470 | 0.21 | 0.34 | ATP metabolic process | B6DDV7 | mitochondrial F1F0-ATP synthase subunit coupling factor 6 [Anopheles darlingi] |
| 454limoContig9667 | 0.43 | 0.69 | Carbohydrate metabolism | D0NA42 | xylose isomerase 1 [Phytophthora infestans T30-4] >gb\|EEY54296.1 |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig5018 | 0.62 | 1.00 | Carbohydrate metabolism | E9HGX0 | mitochondrial malate dehydrogenase [Daphnia pulex] |
| 454limoContig747 | 0.57 | 0.92 | Carbohydrate metabolism | E9HGX0 | mitochondrial malate dehydrogenase [Daphnia pulex] |
| 454limoContig9408 | 0.26 | 0.42 | Carbohydrate metabolism | B5X5I4 | malate dehydrogenase 2-1, NAD (mitochondrial) [Salmo salar] |
| 454limoContig4903 | 0.26 | 0.42 | Carbohydrate metabolism | Q1HQU8 | transaldolase [Aedes aegypti] |
| 454limoContig4759 | 0.12 | 0.19 | Carbohydrate metabolism | E2S087 | fructose 1,6-bisphosphatase [Marsupenaeus japonicus] |
| 454limoContig7445 | 0.21 | 0.34 | Carbohydrate metabolism | Q1HQU8 | transaldolase [Aedes aegypti] |
| 454limoContig3089 | 0.25 | 0.40 | Cytochromce-C oxidase | C1BMV8 | Cytochrome c oxidase subunit 5B, mitochondrial precursor [Caligus |
| 454limoContig1165 | 0.84 | 1.36 | Cytochromce-C oxidase | D3TRS8 | cytochrome c oxidase subunit VIb COX12 [Glossina morsitans morsit |
| 454limoContig1599 | 0.24 | 0.39 | Cytochromce-C oxidase | Q4FZP3 | cytochrome c oxidase subunit 6a polypeptide 1 [Xenopus laevis] |
| 454limoContig6597 | 0.11 | 0.18 | Cytochromce-C oxidase | Q1HQ98 | cytochrome c oxidase polypeptide IV [Bombyx mori] |
| 454limoContig5402 | 4.93 | 7.96 | Cytoskeleton | O96657 | actin 1 [Penaeus monodon] |
| 454limoContig11020 | 1.77 | 2.86 | Cytoskeleton | D6NTM4 | actin [Marsupenaeus japonicus] >gb\|ADH93988.1\| actin [Marsupena |
| 454limoContig2943 | 1.63 | 2.63 | Cytoskeleton | G4VXL0 | beta-actin [Onthophagus binodis] >gb\|AEI91061.1\| beta-actin [Onth |
| 454limoContig10753 | 1.24 | 2.00 | Cytoskeleton | O96657 | actin 1 [Penaeus monodon] |
| 454limoContig7621 | 0.2 | 0.32 | Cytoskeleton | C7TZK1 | Actin-2 [Schistosoma japonicum] |
| 454limoContig5067 | 1.15 | 1.86 | Cytoskeleton | Q25009 | RecName: Full=Tubulin beta-1 chain; AltName: Full=Beta-I tubulin |
| 454limoContig4628 | 0.53 | 0.86 | Cytoskeleton | D4P8F7 | myosin light chain [Marsupenaeus japonicus] |
| 454limoContig12006 | 0.61 | 0.98 | Cytoskeleton | Q8MLY5 | alpha-tubulin [Penaeus monodon] >gb\|AAM73792.1\| alpha-tubulin |
| 454limoContig7805 | 1.25 | 2.02 | Cytoskeleton | B8YDT8 | alpha-tubulin [Tetranychus cinnabarinus] |
| 454limoContig1269 | 1.55 | 2.50 | Cytoskeleton | Q9DFT2 | alpha tubulin [Notothenia coriiceps] |
| 454limoContig3259 | 0.28 | 0.45 | Cytoskeleton | Q7YZU2 | myosin heavy chain [Alpheus lottini] |
| 454limoContig3472 | 0.29 | 0.47 | Cytoskeleton | Q8I6V4 | beta-I tubulin [Callinectes sapidus] |
| 454limoContig9214 | 0.61 | 0.98 | Cytoskeleton | Q6E7L5 | slow-tonic S2 myosin heavy chain [Homarus americanus] |
| 454limoContig3054 | 0.36 | 0.58 | Cytoskeleton | P46150 | moesin/ezrin/radixin homolog [Drosophila melanogaster] |
| 454limoContig3664 | 0.26 | 0.42 | Cytoskeleton | A9CSJ9 | tropomyosin [Oratosquilla oratoria] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig311 | 0.29 | 0.47 | DNA binding | B9ENT8 | Histone H3.3 [Salmo salar] |
| 454limoContig5985 | 0.31 | 0.50 | DNA binding | Q8JHN1 | Hsp70 protein [Cetorhinus maximus] |
| 454limoContig5156 | 0.13 | 0.21 | DNA binding | Q6UZ76 | 70kD heat shock-like protein [Procambarus clarkii] |
| 454limoContig9788 | 0.32 | 0.52 | DNA binding | Q3S349 | heat shock protein 70 [Pachygrapsus marmoratus] |
| 454limoContig10216 | 0.1 | 0.16 | DNA binding | Q3T007 | HSPCA protein [Bos taurus] |
| 454limoContig663 | 0.42 | 0.68 | DNA binding | Q16MJ0 | prohibitin [Aedes aegypti] >gb\|EAT35551.1\| prohibitin |
| 454limoContig2734 | 0.31 | 0.50 | DNA binding | C5I773 | small ubiquitin-like modifier-1 [Litopenaeus vannamei] |
| 454limoContig9947 | 0.39 | 0.63 | DNA binding | C3KHE1 | ADP-ribosylation factor 1 [Anoplopoma fimbria] |
| FABP1(SC1)_seq | 0.76 | 1.23 | FABP | C8CB63 | fatty acids binding protein [Fenneropenaeus chinensis] |
| 454limoContig12146 | 0.25 | 0.40 | FABP | F1B289 | fatty acid binding protein [Eriocheir sinensis] |
| 454limoContig1071 | 0.09 | 0.15 | Ferritin | D3YHP5 | ferritin 3 [Eriocheir sinensis] |
| 454limoContig2963 | 0.48 | 0.77 | Glycosyl Hydrolase (7) | D4HRL0 | GH7 family protein [Limnoria quadripunctata] |
| GH7A | 0.14 | 0.23 | Glycosyl Hydrolase (7) | D4HRK9 | - |
| GH9A | 0.07 | 0.11 | Glycosyl Hydrolase (9) | D4HRL2 | - |
| GH9D | 0.13 | 0.21 | Glycosyl Hydrolase (9) | - | - |
| GH9B | 0.06 | 0.10 | Glycosyl Hydrolase (9) | D4HRL3 | - |
| 454limoContig9008 | 0.06 | 0.10 | Glycosyl Hydrolase (30) | B7PK21 | beta-glucocerebrosidase, putative [Ixodes scapularis] |
| 454limoContig11638 | 0.06 | 0.10 | Glycosyl Hydrolase (35) | E1ZZY3 | Beta-galactosidase [Camponotus floridanus] |
| 454limoContig2406 | 0.08 | 0.13 | Glycosyl Hydrolase (35) | E1ZZY3 | Beta-galactosidase [Camponotus floridanus] |
| 454limoContig76 | 0.96 | 1.55 | Glycolysis | D7RF64 | glyceraldehyde 3-phosphate dehydrogenase [Eriocheir sinensis] >gb\| |
| 454limoContig7325 | 0.58 | 0.94 | Glycolysis | Q6PPI0 | putative fructose 1,6-bisphosphate aldolase [Homalodisca vitripennis] |
| 454limoContig10709 | 1.38 | 2.23 | Glycolysis | D7RF64 | glyceraldehyde 3-phosphate dehydrogenase [Eriocheir sinensis] >gb |
| 454limoContig10199 | 0.2 | 0.32 | Glycolysis | A8CWB3 | putative fructose-bisphosphate aldolase [Lutzomyia longipalpis] |
| 454limoContig485 | 0.49 | 0.79 | Glycolysis | Q6QWP6 | 2-phospho-D-glycerate hydrolase [Callinectes sapidus] |
| 454limoContig2206 | 0.22 | 0.36 | Glycolysis | O96656 | phosphopyruvate hydratase [Penaeus monodon] |
| 454limoContig2063 | 0.19 | 0.31 | Glycolysis | E2C744 | L-lactate dehydrogenase [Harpegnathos saltator] |

| | | | | | |
|---|---|---|---|---|---|
| HC1 | 0.52 | 0.84 | Hemocyanin | D5J6V6 | - |
| HC2 | 0.8 | 1.29 | Hemocyanin | D5J6V7 | - |
| HC3 | 0.67 | 1.08 | Hemocyanin | D5J6V8 | - |
| HC4 | 0.72 | 1.16 | Hemocyanin | D5J6V9 | - |
| HC7 | 0.2 | 0.32 | Hemocyanin | - | - |
| 454limoContig11344 | 0.27 | 0.44 | Hydrolase | Q5U2Q3 | ester hydrolase C11orf54 homolog [Rattus norvegicus] >sp\|Q5U2Q3.1\|CK054_RAT |
| 454limoContig7664 | 0.86 | 1.39 | Kinase | Q004B5 | arginine kinase [Litopenaeus vannamei] |
| 454limoContig10972 | 0.78 | 1.26 | Kinase | Q9NH49 | RecName: Full=Arginine kinase; Short=AK >gb\|AAF43436.1\|AF233355 |
| 454limoContig2163 | 0.32 | 0.52 | Kinase | C3VUU0 | arginine kinase [Fenneropenaeus merguiensis] |
| LRR1(SC3)_seq | 0.93 | 1.50 | LRR | - | - |
| LRR2(SC5)_seq | 0.3 | 0.48 | LRR | - | - |
| LRR3(SC8)_seq | 0.3 | 0.48 | LRR | - | - |
| LRR4(SC1)_seq | 0.07 | 0.11 | LRR | - | - |
| LRR10(SC10)_seq | 0.07 | 0.11 | LRR | - | - |
| 454limoContig3724 | 0.29 | 0.47 | Metabolic process | B7PCD2 | NADP-dependent isocitrate dehydrogenase, putative [Ixodes scapula |
| 454limoContig7022 | 1.25 | 2.02 | Metabolic process | F0J8L6 | TPA: isoamyl acetate-hydrolyzing esterase [Amblyomma variegatum] |
| 454limoContig705 | 0.17 | 0.27 | Metabolic process | A3EXY7 | putative succinate-CoA ligase, GDP-forming, alpha subunit [Maconel |
| 454limoContig4499 | 0.29 | 0.47 | Metabolic process | B0WL69 | transketolase [Culex quinquefasciatus] >gb\|EDS30242.1\| transketol |
| 454limoContig6611 | 0.25 | 0.40 | Metabolic process | F5A6E9 | triosephosphate isomerase [Procambarus clarkii] |
| 454limoContig8195 | 0.32 | 0.52 | Metabolic process | B0W8J6 | isocitrate dehydrogenase [Culex quinquefasciatus] |
| 454limoContig8724 | 0.19 | 0.31 | Metabolic process | Q6PHI8 | transketolase-like protein 2 [Danio rerio] >gb\|AAH56536.1\| Tkt pro |
| 454limoContig2241 | 0.65 | 1.05 | Metabolic process | D3TM09 | transketolase [Glossina morsitans morsitans] |
| 454limoContig4861 | 0.16 | 0.26 | Metabolic process | Q17CT0 | transketolase [Aedes aegypti] >gb\|EAT44192.1\| transketolase [Aed |
| 454limoContig2761 | 0.14 | 0.23 | Metabolic process | E3TDW0 | beta-ureidopropionase [Ictalurus punctatus] >gb\|ADO28496.1 |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig9794 | 0.32 | 0.52 | Metabolic process | B0X750 | NADH-ubiquinone oxidoreductase 39 kda subunit [Culex quinquefasci |
| 454limoContig424 | 0.21 | 0.34 | No BLAST hit | - | No BLAST hit |
| 454limoContig11665 | 0.22 | 0.36 | No BLAST hit | - | No BLAST hit |
| 454limoContig6114 | 0.25 | 0.40 | No BLAST hit | - | No BLAST hit |
| 454limoContig3553 | 0.36 | 0.58 | No BLAST hit | - | No BLAST hit |
| 454limoContig12082 | 0.41 | 0.66 | Oxidoreductase | Q17L62 | oxidoreductase [Aedes aegypti] >gb\|EAT47433.1\| |
| 454limoContig529 | 0.61 | 0.98 | Oxidoreductase | Q567K5 | L-xylulose reductase [Danio rerio] (poor E value) |
| 454limoContig910 | 0.16 | 0.26 | Oxidoreductase | C4N5V0 | Trx1 [Eriocheir sinensis] |
| 454limoContig4100 | 0.2 | 0.32 | Oxidoreductase | D3TLF0 | malate dehydrogenase [Glossina morsitans morsitans] |
| 454limoContig7511 | 0.35 | 0.57 | Oxidoreductase | B7PCD2 | NADP-dependent isocitrate dehydrogenase, putative [Ixodes scapula |
| 454limoContig7800 | 0.22 | 0.36 | Oxidoreductase | F5BZU2 | isocitrate dehydrogenase 1 (NADP+), soluble [Epinephelus bruneus |
| 454limoContig11726 | 0.36 | 0.58 | Oxidoreductase | D6WVW0 | PREDICTED: similar to diacetyl/L-xylulose reductase [Tribolium castan |
| 454limoContig11005 | 0.33 | 0.53 | Oxidoreductase | C6JVX0 | mitochondrial rieske iron-sulfur protein [Tigriopus californicus] |
| 454limoContig9656 | 0.06 | 0.10 | Protease | P09954 | carboxypeptidase A1 precursor [Sus scrofa] >sp\|P09954.2\|CBPA1_P |
| 454limoContig6112 | 0.08 | 0.13 | Protein binding | A0PJ13 | 14-3-3-like protein [Penaeus monodon] |
| 454limoContig9834 | 0.59 | 0.95 | Protein binding | D2DSH6 | chaperonin 10 [Scylla paramamosain] |
| 454limoContig7318 | 0.4 | 0.65 | Regulator | E0XEM8 | luciferin regenerating enzyme [Lampyris turkestanicus] |
| 454limoContig7779 | 0.69 | 1.11 | RNA binding | E2AE08 | RNA-binding protein squid [Camponotus floridanus] |
| 454limoContig7886 | 0.51 | 0.82 | Transferase | C0SMV2 | gluthathione S-transferase mu [Takifugu obscurus] |
| 454limoContig3142 | 0.44 | 0.71 | Transferase | E0WBN2 | glutathione S-transferase [Eriocheir sinensis] |
| 454limoContig8995 | 0.38 | 0.61 | Transferase | C1C1S2 | Glutathione S-transferase Mu 3 [Caligus clemensi] |
| 454limoContig2903 | 0.39 | 0.63 | Transferase | E0WBN2 | glutathione S-transferase [Eriocheir sinensis] |
| 454limoContig10499 | 0.19 | 0.31 | Transferase | Q3B8M0 | glutathione transferase zeta 1 [Xenopus laevis] >gb\|AAI06225.1 |
| 454limoContig10078 | 0.4 | 0.65 | Translation | B3GD96 | ubiquitin/60s ribosomal protein L40 fusion [Caenorhabditis brenneri] |
| 454limoContig11853 | 0.31 | 0.50 | Translation | D2K241 | elongation factor-1 alpha [Cyamus gracilis] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig7424 | 0.1 | 0.16 | Translation | Q197G1 | tail muscle elongation factor 1 gamma [Procambarus clarkii] |
| 454limoContig1517 | 0.2 | 0.32 | Translation | D7F2K9 | ribosomal protein S3 [Palaemonetes varians] |
| 454limoContig5644 | 0.28 | 0.45 | Translation | B7QI86 | 60S ribosomal protein L7A, putative [Ixodes scapularis] >gb\|EEC18 |
| 454limoContig10421 | 0.19 | 0.31 | Translation | F5A6E0 | ribosomal protein S11 [Procambarus clarkii] |
| 454limoContig11518 | 0.22 | 0.36 | Translation | Q5MIR0 | 60S ribosomal protein L11 [Aedes albopictus] |
| 454limoContig7423 | 0.14 | 0.23 | Translation | Q9BNX9 | elongation factor-2 [Armadillidium vulgare] |
| 454limoContig3328 | 0.18 | 0.29 | Translation | D2K241 | elongation factor-1 alpha [Cyamus gracilis] |
| 454limoContig9297 | 0.2 | 0.32 | Translation | Q0PXV3 | putative elongation factor 1 beta' [Diaphorina citri] |
| 454limoContig1756 | 0.08 | 0.13 | Translation | B9ZZQ1 | elongation factor 1-alpha [Marsupenaeus japonicus] |
| 454limoContig4213 | 0.14 | 0.23 | Translation | Q7ZYQ8 | RPL18A protein [Xenopus laevis] |
| 454limoContig1924 | 0.32 | 0.52 | Translation | B8XVP3 | 40S ribosomal protein S14 [Ochlerotatus taeniorhynchus] |
| 454limoContig6899 | 0.3 | 0.48 | Translation | D2DSZ5 | ribosomal protein L19e [Scylla paramamosain] |
| 454limoContig8064 | 0.16 | 0.26 | Translation | A6N9Y7 | 60s ribosomal protein L18 [Ornithodoros parkeri] |
| 454limoContig1706 | 0.26 | 0.42 | Translation | Q6PBK3 | 40S ribosomal protein S28 [Danio rerio] |
| 454limoContig6128 | 0.3 | 0.48 | Translation | C0JBY4 | protein disulfide isomerase [Litopenaeus vannamei] |
| 454limoContig981 | 0.08 | 0.13 | Translation | Q8WQJ2 | 60S acidic ribosomal protein P0 [Spodoptera frugiperda] |
| 454limoContig2826 | 0.19 | 0.31 | Translation | P48149 | ribosomal protein S15Aa, isoform D [Drosophila melanogaster] |
| 454limoContig209 | 0.19 | 0.31 | Translation | B1N8Y0 | elongation factor 2 [Penaeus monodon] |
| 454limoContig6919 | 0.11 | 0.18 | Translation | C4PL18 | ribosomal protein L3 [Penaeus monodon] |
| 454limoContig11747 | 0.18 | 0.29 | Translation | NP_079868.1 | 60S ribosomal protein L35 [Mus musculus] >ref\|NP_997676.1 |
| 454limoContig5394 | 0.27 | 0.44 | Transport | D9IG67 | voltage-dependent anion-selective channel [Eriocheir sinensis] |
| 454limoContig9991 | 0.73 | 1.18 | Transport | G1FEX1 | adenine nucleotide translocase 2 [Litopenaeus vannamei] |
| 454limoContig2560 | 0.25 | 0.40 | Transport | G1FEX1 | adenine nucleotide translocase 2 [Litopenaeus vannamei] |
| 454limoContig10280 | 0.22 | 0.36 | Transport | D9IG67 | voltage-dependent anion-selective channel [Eriocheir sinensis] |

| | | | | | |
|---|---|---|---|---|---|
| 454limoContig11077 | 0.33 | 0.53 | Transport | Q0PWR8 | putative solute carrier family 25 member 3 [Pimephales promelas] |
| 454limoContig8530 | 0.14 | 0.23 | Transport | C1BPR3 | Transmembrane emp24 domain-containing protein 9 precursor |
| 454limoContig2870 | 0.23 | 0.37 | Transport | Q8WR15 | cellular retinoic acid/retinol binding protein [Metapenaeus ensis] |
| 454limoContig4847 | 0.14 | 0.23 | Transport | D1FPN3 | mitochondrial phosphate carrier protein [Cimex lectularius] |
| 454limoContig10513 | 0.52 | 0.84 | Uncharacterised | E9GU36 | hypothetical protein DAPPUDRAFT_321849 [Daphnia pulex] |
| 454limoContig610 | 0.21 | 0.34 | Uncharacterised | B0WNJ5 | muscle-specific protein 20 [Culex quinquefasciatus] |

# Appendix B - *Limnoria* Hemocyanin Gene Sequences

*Limnoria* HC1 gene sequence, 2037 bp.

1 60

ATGAAGGCATTCGCTTTCTTAGCTCTCTTGGCTGTGGCCTCTGCTTCCTGGATCCCTGAA
TATCAACATGATGCTCCAGTGAGTGTTGCCAAAAAGCAGCAAGATGTCCTACGTCTTCTC
TACAAGGTAACTGAACCCATTCGCAAGAGATACACAGATCTGAAAGATTCTGAACAAAAT
TTCAACCCAGTTGCTGATCTGTCAGCTTACAGCGATGGGGGTGCAGCTGCTCAAGCGCTG
ATTACTGAAGTAAATGATGGGAGAACTTTGCAACAAAAACACTGGTTTTCTCTCTTCAAT
GACCGGCATAGAGAAGAAGCTCTATTGTTATTTGATGTACTAATTCACAGCAATACTTGG
AATACTGCTGTAGCCAATGCTGCTTATTTTAGGGAGAAGCTAAATGAGGGTGTCTTTGTA
TATGCATTGTACACTGCTGTTATCCACTCCCCTCAAGGGCGTGATCTAGTTCTCCCACCT
CTTTATGAAGTAACACCACATCTGTTCGCCAACTCTGAAGTAGTGAACCAAGCTTACTCA
GCTAAAATGACACAGACACCAGGAAAATTCAGGATGACCTTTACTGGCACCAAGAAGAAT
CCTGAACAAAGAGTTGCTTACTTTGGAGAAGATATCGGTTTGAATGTTCATCACGTATCC
TGGCATATGGATTACCCATTCTGGTGGAAAGATTCCTATGGATACCATCTTGACCGCAAA
GGAGAACTTTTCTTCTGGGCTCATCACCAGCTTACTGTAAGATTTGATGCTGAAAGACTT
TCCAACAACCTAGACCTCGTTGATGAACTTTACTGGGATAGGCCAATCGTTGAAGGTTTT
GCCCCTCACACTACCTACAAATACGGCGGAGAATTCCCTGCTCGTCCAGACAATGTAAAC
TTTGAAGATGTTGATGGTGTTGCACGCATTCGTGACATGATTATTTATGAAACAAGAATT
CGAGATGCAATTGCTCATGGATACATTACTGGTGCAAATGGAGAAAAAATTGACATTAGA
AACGAAAATGGTATTGATATTCTTGGTGATGTTATTGAATCATCTGTGTACAGCCCTAAT
GCTGGATATTATGGAGCACTTCATAATCTTGCTCACATCATTCTTGGACGTCAAGGTGAC
CCACATGGAAAGTTCAATATGCCCCCAGGCGTAATGGAACATTTCGAAACTGCAACTAGA
GATCCATCTTTCTTCAGACTGCATAAATACATGGACAACATTTTCAAGGAACACAAAGAT
TCTCTTCCACCATACACAAAAGAAGAGGTTGAATTCCCAGATGTAGCTATCAATTCTGTT
GGAATAGTAGGAACACTTGAGACATTCTTTGAAGACTTCGAATTCAATCTCTTAATGGGT
GTCGATGATACTGAAGAAGTTGAAGATGTCCCCATATCTGCTATTGTTTCACGACTTAAT
CATGAGCCATTCACTCTGACAGCTGATGTTACAAACAACAATGCAGCAGATGTCTTCGCT
ACCGTCCGTATTTTCCTTTGCCCCAAATATGATGCAAATGGTATTGTATTCACCCTTGAA
GAAGGACGATGGAATTGCATTGAAGTTGATAAATTTTGGAGAACATTGACACCAGGAAGT
AACCAACTGACAAGGAAATCAACGGAAACATCAACCACTGTTCCTGATGTTCCAAGTTTC
CAATATCTCATCAACAAAGCTGATGCTGCCGTTTCATCTGGCTCCGATTTGGATCTGTCT

GCTTATTCTCGAGGATGCGGTATCCCCGACAGGCTTCTGCTCCCCAAAGGTTCAACCGAA
GGCACTGAGTTCCTGCTCGTAGTTGCTCTTACTGATGGAAGCAAGGATGCTGCTGTTGAG
GGACTAGAGAAGGACGAACGTGGAGGAAGCCATGCGCAATGCGGTATCCATGGTGAAACC
TACCCAGACAAACGACCTCTTGGATACCCAATTGATCGCATAATCCCCGATGAAAGGGTT
CTTCTTAACTTCCCTAATTTGCTTAGAACCGTTGTTACCATCAAGCACAGAGAATAA

*Limnoria* HC2 gene sequence, 2037 bp.

1 60
ATGAAGGCATTTGTGTTCTTCGGTCTCTTGGCTGTTGCCTCCGCATCTTGGCTCCCTGAA
TTTCAATCAGATTCATCTGTATCAACTGCAAAGAAGCAGCAAGATGTTTTGCGCCTTCTC
TTTAAAGTAACGGAACCAATTCGAAAAAGGTATTCCGAGCTTAAAGATGCTTCAGTAAAC
TTCCATCCCGATGCAGATCTTACAAAGTACACAGATGGAGGAGCAGCAGTGAAAGCCCTT
CTTACAGAAGTAGAAGATGGGAGGACTTTGGAACAGAAGCATTACTTCTCCCTTTTTAAT
GACCGTCATAGAGAGGAAGCCCTCATGCTCTTCGAAGTATTGATAGACAGCAAAGACTGG
GCTACAGCTGTAGCCAATGCAGCTTACTTCCGAGAGAAGTTGAACGAAGGTGTATTTGTC
TATGCATTGTATACTGCAGTAATCCATGCGCCTATAGGGAAAGATCTCGTTCTTCCACCT
CTATATGAAGTAACACCGCATTTATTTGCAAACACGGAAGTAATACACAAAGCTTATTCC
GCCAAAATGACACAGACGCCTGGTAAATTTAAAATGTCCTTCACCGGCACCAAGAAAAAC
CCTGAACAAAGAGTAGCTTACTTTGGAGAAGATATCGGTTTAAATGTACATCACGTATCC
TGGCATATGGATTACCCATTCTGGTGGAAAGATTCCTATGGATATCATCTTGACCGCAAA
GGAGAACTTTTCTTCTGGGCTCATCACCAACTCACTGTAAGATTTGATGCTGAAAGACTT
TCTAACAACCTAGACCTCGTTGATGAACTTTACTGGGACAGACCAATCGTTGAAGGATTT
GCACCTCACACTACCTACAAATACGGCGGTGAATTCCCAGCTCGTCCAGATAATGTAAAC
TTTGAAGATGTTGATGGTGTTGCTCGCATTAGAGATATGATTATTATTGAAACCAGAATT
CGTGATGCCATCGCCCATGGTTACGTCACAGCTGCTGACGGTAGTAAAATCGATATTAGA
AATGAACATGGAATTGATGTCCTTGGAGATGTGATTGAATCATCTGTGTACAGTCCTAAC
GCTGGCTATTATGGAGCACTTCACAATCTTGCTCACATTATTCTTGGACGTCAAGGAGAT
CCACATGGAAAGTTTAACATGCCCCCAGGTGTCATGGAACATTTTGAAACCGCAACTAGA
GATCCATCTTTCTTCAGACTCCATAAATACATGGATAACATTTTTAAGGAACACAAAGAC
TCTCTACCACCTTACACAAAGGAAGAGGTTGAATTCCCTGATGTCGCTATCACTGCCGTT
GGTGTTGAAGGAAAGCTTGAAACATTCTTTGAAGATTTTGAGTTTAATCTTCTTATGGGC
GTTGATGATACTGAAGAAGTTGAAGATGTTCCAATTTCAGCAATTGTTTCGCGTCTCAAT
CACGATCCTTTCACCATGACAGCCGACATTTCCAACAACAATGCAGCAGATGTTTTTGCC
ACAATACGTATTTTCATTTGTCCCAGATATGATGCAAATGGAATTCTATTTACTCTGAAT
GAAGGAAGATGGAACTGCGTTGAAATTGACAAATTTTGGAGAACACTGACACCAGGATCT

AACCACATTACCAGGAAATCAAAGGATACTTCAACCACTGTCCCAGATGTCCCAAGCTTC
CAGTCTCTCCTTGAACAAGCTGACGCTGCCGTTTCATCTGGATCTGCTTTGGACTTGTCC
GCTTATTCTCGAGGGTGCGGTATTCCCGACAGACTGTTGTTACCAAAGGGCAATACTGAA
GGCATGGAGTTTGCAGTTATTGTAGCAGTTACTGACGGTAGCAAGGATGCGGCTGTTGAA
GGCCTGGAAAAGGACGAACGTGGTGGAAGCCATGCCCAGTGTGGTATCCATGGCGAAACC
TTCCCAGACAAGAGACCTCTGGGATACCCACTTGACCGTAGGATTCTTGATGAAAGAGCT
CTGTTGAACTTCCCTAACTTGCATAAGGGTATTATTACAATTAGCCATAGAGATTAA

*Limnoria* HC3 gene sequence, 2010 bp.

1 60
ATGAAGGCCTTCGCTTTCTTGCTACTTTTGGCTGTTGCCTCCGCTTCTTGGTTACCGGAA
TTCCAATCGGATGCAGTTGTCTCAACTGCAAAAAAACAACAAGATGTACTTCGTCTTCTG
TTCAAGGTAACGGAACCTATTCGTAAGAGGTACACTGATCTCAAAGATGCATCAGTTAGC
CACAAACCTACAGGAGTCGCTGCCGAAAAACTGATCAGTGAAATCGAGGACGGTAGAACA
TTAGAACAGAAACACTACTTTTCTCTATTCAATGACCGTCACAGAGAGGAGGCACTTATG
CTTTTTGAAGTGCTAATGAGTAGCGGGGACTGGAACACAGCTGTATCCAATGCTGCATAT
TTCAGAGAAAGATTAAATGAAGGTGTCTTCGTCTATGCTATATACACTGCTGTTATCCAC
TCACCTTTGGGAAAAGATCTTGTTCTTCCTCCGCTATACGAAGTTACTCCACATTTATTT
GCCAATACCGAAGTGATCAACCAAGCGTATTCAGCTAAGATGACACAGACTCCTGGAAAA
TTCCAAATGTCCTTTACTGGAACAAAGAAGAACCCAGAACAAAGAGTTGCCTACTTTGGT
GAAGATATCGGTTTGAACGTCCATCACGTATCTTGGCATATGGATTACCCATTCTGGTGG
AAAGATTCATATGGATACCATCTTGACCGCAAAGGAGAACTTTTCTTCTGGGCTCATCAT
CAGCTCACTGTACGATTTGATGCTGAAAGGCTTTCCAACAACCTAGACCTTGTAGATGAA
CTTTACTGGGACAGACCAATCGTTGAAGGTTTTGCTCCTCATACTACCTACAAATACGGT
GGAGAATTTCCAGCTCGTCCAGATAACGTAAACTTTGAAGATGTTGATGGTGTTGCACGC
ATTCGTGACATGGTTATTTTTGAAACAAGAATTCGTGATGCTATCGCCCACGGTTACGTT
ACAGCTGCTGACGGTAGTAAAATCGATATTAGAAATGAACATGGAATTGATATCCTTGGA
GATGTAATTGAATCATCTGTGTACAGTCCTAACCCTGGCTATTATGGAGCACTTCACAAT
CTTGCTCACATTATTCTTGGACGTCAAGGAGACCCACATGGAAAGTTCAATATGCCCCCA
GGTGTTATGGAACATTTTGAAACCGCAACTAGAGATCCTTCTTTCTTCAGACTCCATAAA
TACATGGATAACATCTTCAAAGAACACAAAGACTCTCTCCCACCTTACACTAAAGAACAG
GTTGAATTCCCTGATGTTGCTGTCACTGCAGTTGGTGTTGACGGAAAGCTTGAAACATTC
TTTGAAGATTTTGAGTTTAATCTTCTTATGGGAGTTGATGACACAGAAGAAGTTGAAGAT
GTTCCAATTTCAGCAGTTGTTTCGCGTCTTAATCACGATCCTTTCACCATGACTGCCGAC
ATTAGTAACAACAATGCAGCTGATGTATTTGCCACAATACGTGTTTTCCTTTGCCCAAGA

TACGATGCGAATGGAATTCTATTCACCGTAAATGAAGGACGTTGGAACTGCGTTGAAATT
GACAAATTTTGGAGAACATTGGTACCTGGATCTAACCACATAACACGAAAATCATCGGAT
GCATCAACAACTGTACCAGATGTTCCCAGTTTCCAGTCTCTTGTTGCCCAGGCTGATGCT
GCCGTTGCATCTGGAGAAGATTTGGATTTGTCCGCTTATTCTCGAGGGTGTGGTATTCCT
GACAGGCTGCTGCTCCCTAAGGGCAATGCCCAAGGCATGGAGTTCGCCGTTATTGTAGCT
GTTACTGATGGTAGCCAGGATGCAGCTGTTGAAGGACTGGAAAAGGATGAGCGTGGTGGA
AGCCATGCCCAGTGCGGTATCCATGGTGAAACCTTCCCTGACAAACAACCCCTAGGATAC
CCACTCGACAGAAGGATTCTAGATGAAAGAGCCCTTTTGAACTTCCCTAATTTGTTCAAG
ACTATTGTGAAGATTACCCACAGAGAATAA

*Limnoria* HC4 gene sequence, 2040 bp.

1 60
ATGAAGGCTTTCGCTTTCTTGCTGCTTTTGGCTGTTGCCTCTGCTTCATGGCTACCGGAA
TTCCAATCTGATCATCATACCGTCTCAACTGCAAAGAAACAACAAGATGTACTGCGTCTT
CTGTACAAGGTAACAGAACACATCCGTAAAAGGTATACTGAACTTAAAGAAGCTTCAGTT
AACTATGACCCCGCTGGACACCCTGAAGGATGCAAGGATGGGGGAGTCGCTGTCAAAACA
TTGCTCAGTGAAATTGCGGACGGAAGAGTATTAGAACAGAAGCATTATTTCTCTCTTTTT
AATGACCGTCATAGGGAAGAGGCACTTATGCTTTTTGAGGTGCTAATGAATTGCAAGGAC
TGGCACACCGCCGTATCCAATGCTGCATATTTCAGAGAGAGGCTAAATGAAGGTGTCTTC
GTCTATGCATTGTACACTGCTGTTATCCACTCACCTTTGGGAAAAGACCTTGTTCTTCCT
CCACTATACGAAATTACACCACATTTGTTTGCTAACACCGAAGTGATTAACCGAGCATAC
TCAGCTAAGATGACACAGACTCCTGGAAAATTCCAAATGTCCTTCACTGGAACAAAGAAG
AACCCAGAACAAAGAGTTGCCTACTTTGGAGAAGATATCGGTTTGAACGTCCATCACGTA
TCCTGGCATATGGATTACCCATTCTGGTGGAAAGATTCCTATGGATACCATCTTGACCGC
AAAGGAGAACTTTTCTTCTGGGCTCATCACCAGCTCACAGTACGATTTGATGCTGAAAGG
CTTTCCAACAACCTAGACCTCGTAGATGAACTTTACTGGGACAGACCAATCGTTGAAGGT
TTTGCTCCTCATACTACCTACAAATACGGTGGAGAATTTCCTGCTCGTCCAGATAACGTT
AACTTCGAAGATGTTGATGGTGTTGCACGCATTCGTGACATGGTTATTTTTGAAACAAGA
ATTCGTGATGCTATCGCCCACGGTTACGTCACAGCTGCTGACGGTAGTAAAATCGATATT
AGAAATGAACATGGAATTGATATCCTTGGAGATGTAATTGAATCATCAGTGTACAGCCCT
AACCCTGGCTATTATGGAGCACTTCACAATCTTGCTCACATTATTCTTGGACGTCAAGGA
GATCCACATGGAAAGTTCAATATGCCCCCAGGTGTCATGGAACATTTTGAAACCGCAACT
AGAGATCCATCTTTCTTCAGACTGCATAAATACATGGATAACATTTTCAAAGAACACAAA
GACTCTCTCCCACCATACACTAAAGAAGAGGTTGAATTTCCCGATGTCGCCATCACTGCC
GTTGGTGTTGACGGAAAGCTTGAAACATTCTTTGAAGATTTTGAGTTTAATCTTCTTATG

GGAGTTGATGACACAGAAGAGGTTGAAGATGTTCCAATTTCAGCAGTTGTTTCGCGTCTT
AATCACGATCCTTTCACCATGACTGCCGACATTAGTAACAACAATGCAGCTGATGTATTT
GCCACAATACGTGTTTTCCTTTGCCCAAGATATGATGCAAATGGAATTCTATTTACTCTG
AATGAAGGAAGATGGAACTGCGTTGAAATCGACAAATTCTGGAGAACTCTTGTACCAGGT
TCTAACCACGTAACCCGAAAATCATCGGATACATCAACAACTGTACCAGATGTTCCCAGC
TTCCAGTCTCTTGTTGCCCTGGCTGATGCTGCCGTTGCATCTGGAGAGGATTTGGACTTG
TCTGCTTATTCACGAGGGTGTGGTATTCCAGACAGACTGTTGCTACCTAAGGGAAATGCC
CAAGGCATGGAGTTTGCCGTTGTTGTAGCTGTTACTGACGGTAGCAAGGATGCAGCTGTA
GAAGGACTGGAAAAGGACGAGCGAGGTGGAAGTCATGCCCAATGTGGTATCCATGGCGAA
ACCTTCCCTGACAAACGACCTCTGGGATACCCACTGGACAGAAGGATTCTAGATGAAAGA
GCTCTATTGAACTTCCCTAATTTGTTCAGGACTATTGTAAAGATTTCTCACCGAGAATAG

*Limnoria* HC7 gene sequence, 1998 bp.

1 60
ATGAAGCTTTTAGTCCTCTTCGCCGTTGTAGCTTTCGCCTCAGCTGCCTCTGGAGAGGTC
AGCAATGCTTACAAGCAGCAACAAGTTAACAGACTTCTATATAGAATTACTGAAGATGTA
AGACCAGAGTTCACTGAAATCTACCAGGCTCAGCAAAGTTACCATCCTGAAGAGCACCTT
CCAGCATGTTTGGATGCAGCCAAGGAAATTATAACCTTGTTGAAGGATATTGAAGGTGGT
CGTCTTTTAGAGAAAAAACATTACTTCTCCCTCTTTAACGATAAACACAGAAGGGATGCT
CTTTTGCTGGTTAAAGTTCTTCTTAGTTGTCAAGATTTCGACAGTGCTCTAAGTAACGCG
GCATATTTCCGTCAATATCTTAACGAAGGAGCTTTCATTTACGCTATTTACATAGCTGTC
ACTCACTCGGACCTTACCAGATCAGTTCAGCTTCCTCCGTTGTATGAAATTACACCACAC
ATGTTCATCAACACAGAAGTTGTACAAAAGGCTTATACAGCCCAGATGACTCAAACGCCA
AGCAAAATATCTATGGGCTTCACTGGATCCAAGAAAAATAAGGAACAGAGAGTTGCATAC
TTCGGTGAGGATGTTGGCATGAACTCTCACCACGTTCACTGGCATATGGACTTCCCATTC
TGGTGGGATGGATACAAAATGGACCGAAAGGGAGAACTTTTCTTCTGGATTCATCACCAA
TTGACAGCAAGATTTGACGCTGAAAGATTATCAAATAAGCTTCCAGTTGTTGACGAGCTT
TATTGGGATAAACCTATCTACGAAGGATTTGCTCCTCATACCACATACAGATACGGTGGT
GAGTTCCCGGCTCGTCCAGATAACAAGAAATTCGAGGATGTTGCTGGAGTTGCAAGAGTT
AGAGACATGAAGGTCACAGAAGACAGAATCCGTGATGCTATTGCTCATGGATACATTCAA
GATATTCATGGAACTGATATTCCACTAGATGAAGAACACGGCATTGATATCTTAGGAGAC
ATCATTGAATCCTCTACAAGCAGCCCAAACCAATTTTACTACGGATCCCTTCACAATACA
GCTCACGTAATCTTGGGACGACAAGCTGACCCCAAAGGAAAGTTCAACATGCCTCCAGGT
GTGATGGAACATTTCGAAACAGCAACAAGAGATCCTTCATTCTTCAGACTGCACAAGTAT
ATGGACAACATTTTCAAAGAACATAAAGATACACTTACACCTTATACCAAGGAAGATCTC

GAGTACACAAACGTTGAAATCACCGATATTACAGTAGGTGGAGATCTCGTTACATTCTTC
GAAGAATTTGAAATTGATCTTCTTAATGGAATTGATACCAGTGAAAATATTCCTGAGGTG
CCAATTTCTGCATTAGTTACAAGATTGAACCACAAACCTTTCTCATACACCATTGACGTT
ACAGCAGACCATTATGATAAGGCAACAGTGCGTATCTACCTTTGCCCAAAATACGATTAT
AATGGAATTGAATACACTCTTAATGAAGGACGTTGGGGCTGCGTTGAAATGGACAAATTC
TGGGTGATTGTAAATGAAGGAACAAACACCATTACTCGTGATTCCTCCGAGTCATCAGTG
ACCATTCCCGACAGGACATCTTTCCAGACTTTGATCGAGAAGGCAGATGCGGCCGTATCA
AGTGGAGGTGACATAACTGTTTCGGATATTCGAGCTTGTGGACAACCCAACCGTCTTTTG
CTACCTAAAGGAAGCGCACAAGGAGTCGACTTCGAACTTCTAGTTGCTGTTACTTCGGGA
GATGACGCTGTCTACAGAGGCTCCATTGATGACGAGTTTGGTGGTAACCACGCCTACTGC
GGAGCCAAGGGCAAGAAATACCCTGACAACAGGCCAATGGGATATCCTCTTGACAGAAAT
ATTCCAGATGACAGAGTCTTCAAGGTCCCGAACATTCGCTGGGTCACTGTTAATATCAAG
CATGATCCTTCTTTGTAA

*Limnoria* HC1 E. coli codon optimised gene sequence without signal peptide sequence, 1992 bp.

1 60

AGCTGGATTCCTGAATATCAGCATGATGCACCGGTTAGCGTTGCAAAAAAACAGCAGGAT
GTTCTGCGTCTGCTGTATAAAGTTACCGAACCGATTCGTAAACGTTACACCGATCTGAAA
GATAGCGAACAGAATTTTAACCCGGTTGCAGATCTGAGCGCATATAGTGATGGTGGTGCA
GCAGCACAGGCACTGATTACCGAAGTTAATGATGGTCGTACCCTGCAGCAGAAACATTGG
TTTAGCCTGTTTAATGATCGCCATCGTGAAGAAGCACTGCTGCTGTTTGATGTTCTGATT
CATAGCAATACCTGGAATACCGCAGTTGCAAATGCAGCATATTTTCGCGAGAAACTGAAT
GAAGGTGTTTTTGTGTATGCACTGTATACGGCAGTTATTCATAGTCCGCAGGGTCGTGAT
CTGGTTCTGCCACCGCTGTATGAAGTGACACCGCACCTGTTTGCAAATAGCGAAGTTGTT
AATCAGGCATACAGCGCAAAAATGACCCAGACACCGGGTAAATTTCGTATGACCTTTACC
GGCACCAAAAAAAACCCGGAACAGCGTGTTGCATATTTCGGTGAAGATATTGGTCTGAAT
GTTCATCATGTTAGCTGGCATATGGATTATCCGTTTTGGTGGAAAGACAGCTATGGTTAT
CATCTGGATCGTAAAGGCGAACTGTTTTTCTGGGCACATCATCAGCTGACCGTTCGTTTT
GATGCAGAACGTCTGAGCAATAATCTGGATCTGGTTGATGAACTGTATTGGGATCGTCCG
ATTGTTGAAGGTTTTGCACCGCATACCACCTATAAATACGGTGGTGAATTTCCGGCACGT
CCGGATAATGTTAATTTTGAAGATGTTGATGGCGTTGCCCGTATTCGTGATATGATTATC
TATGAAACACGCATTCGTGATGCAATTGCCCATGGTTATATTACCGGTGCAAATGGCGAA
AAAATTGATATCCGCAATGAAAACGGCATCGATATTCTGGGTGATGTTATTGAAAGCAGC
GTTTATAGCCCGAATGCAGGTTATTATGGTGCACTGCATAATCTGGCACATATCATTCTG
GGTCGTCAGGGTGATCCGCATGGCAAATTCAATATGCCTCCGGGTGTTATGGAACATTTT
GAAACCGCAACCCGTGATCCGAGCTTTTTTCGTCTGCACAAATATATGGATAACATTTTC
AAAGAACATAAAGATAGCCTGCCTCCGTATACCAAAGAAGAGGTTGAATTTCCTGATGTG
GCCATTAATAGCGTTGGTATTGTTGGCACCCTGGAAACCTTTTTTGAGGATTTTGAATTT
AACCTGCTGATGGGTGTTGATGATACCGAAGAGGTGGAAGATGTTCCGATTAGCGCAATT
GTTAGCCGTCTGAATCATGAACCGTTTACCCTGACCGCAGATGTTACCAATAATAACGCA
GCAGATGTTTTTGCCACCGTGCGTATTTTTCTGTGCCCTAAATATGATGCCAACGGCATT
GTTTTTACCCTGGAAGAAGGTCGTTGGAATTGTATTGAAGTGGATAAATTCTGGCGTACC
CTGACTCCGGGTAGCAACCAGCTGACCCGTAAAAGCACCGAAACCAGCACCACCGTTCCG
GATGTGCCGAGCTTTCAGTATCTGATCAATAAAGCAGATGCAGCAGTTAGCAGCGGTAGC
GATCTGGACCTGTCAGCATATAGCCGTGGTTGTGGTATTCCGGATCGTCTGCTGCTGCCG
AAAGGTAGCACCGAAGGCACCGAATTTCTGCTGGTTGTTGCACTGACCGATGGTAGCAAA
GATGCAGCCGTTGAAGGTCTGGAAAAAGATGAACGTGGTGGTAGCCATGCACAGTGTGGC
ATTCATGGTGAAACCTATCCTGATAAACGTCCGCTGGGTTATCCGATTGATCGTATTATC
CCGGATGAACGCGTGCTGCTGAATTTTCCGAACCTGCTGCGTACCGTTGTTACCATTAAA

CATCGTGAATAA

# Appendix C – Summary of *E. coli* Heterologous Hemocyanin Expression Experiments

Abbreviations used for all tables in this appendix: $OD_{600}$, optical density at 600 nm; IPTG, isopropyl β-D-1-thiogalactopyranoside; glucose added, was glucose added to the growth media at all stages of the experiment?; soluble, was the target protein observed to be expressed in a soluble form?; insoluble, was the target protein observed to be expressed in an insoluble form?; Western blot; was expression of the target protein confirmed by Western blot?; protein ID; was the identity of the target protein confirmed by gel band identification?; purity (%), an estimate of the purity level achieved; RG-2, Rosetta-gami 2 *E. coli* cells; 2YT, 2x YT medium; LB, Luria-Bertani broth; AIM, auto-induction medium; PTU, phenylthiourea.

Summary of experiments to heterologously express the *Limnoria* HC3 gene using the **pET28a(+)** vector.

| Expression Strain | Culture Scale | $OD_{600}$ at Induction | IPTG Conc. (mM) | Time Post Induction (hours) | Expression Temperature | Glucose Added | Media | Media Additives | Soluble | Insoluble | Western Blot | Protein ID | Purification Method | Purity (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RG-2 | 20 ml | 0.6 | 1.0 | 4 | 20 ℃ | N | 2YT | ±1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 20 ml | 0.6 | 1.0 | 4 | 30 ℃ | N | 2YT | ±1 mM $CuCl_2$ | N | Y | Y | - | - | - |
| RG-2 | 20 ml | 0.6 | 1.0 | 4 | 37 ℃ | N | 2YT | ±1 mM $CuCl_2$ | N | Y | Y | - | - | - |
| RG-2 | 20 ml | 0.6 | 1.0 | 21 | 20 ℃ | N | 2YT | ±1 mM $CuCl_2$ | N | N | Y | - | - | - |
| RG-2 | 20 ml | 0.6 | 1.0 | 21 | 30 ℃ | N | 2YT | ±1 mM $CuCl_2$ | N | N | Y | - | - | - |
| RG-2 | 20 ml | 0.6 | 1.0 | 21 | 37 ℃ | N | 2YT | ±1 mM $CuCl_2$ | N | Y | Y | - | - | - |
| RG-2 | 30 ml | 0.5 | 0.1 | 3, 4, 5, 6, 8 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 0.5 | 0.1 | 3, 4, 5, 6, 8 | 20 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 0.5 | 1.0 | 3, 4, 5, 6, 8 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 0.5 | 1.0 | 3, 4, 5, 6, 8 | 20 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 1.0 | 0.1 | 3, 4, 5, 6, 8 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 1.0 | 0.1 | 3, 4, 5, 6, 8 | 20 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 1.0 | 1.0 | 3, 4, 5, 6, 8 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 30 ml | 1.0 | 1.0 | 3, 4, 5, 6, 8 | 20 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | - | - |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RG-2 | 125 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | N | Y | Y | - | - | - |
| RG-2 | 125 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | LB | 1 mM $CuCl_2$ | N | Y | Y | - | - | - |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | N | Y | - | Ni-IMAC, 1 ml HisTrap HP | 50 |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | N | Y | - | Ni-IMAC, 5 ml HisTrap HP | 50 |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | Ni-IMAC, 1 ml HisTrap HP | 50 |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuSO_4$ | Y | N | Y | - | Ni-IMAC, 1 ml HisTrap HP | 50 |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 50 µM PTU | Y | N | Y | - | Ni-IMAC, 1 ml HisTrap HP | 50 |
| RG-2 | 10 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | Ni-IMAC, 1 ml HisTrap HP GF HiLoad 16/60 Superdex 200 | |
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | Ni-IMAC, 5 ml HisTrap FF Crude | 80 |
| RG-2 | 4 x 500 ml | 1.0 | 0.1 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | Y | Ni-IMAC, 5 ml HisTrap HP | 60 |

Summary of experiments to heterologously express the *Limnoria* HC3 gene using the **pET52b(+)** vector.

| Expression Strain | Culture Scale | $OD_{600}$ at Induction | IPTG Conc. (mM) | Time Post Induction (hours) | Expression Temperature | Glucose Added | Media | Media Additives | Soluble | Insoluble | Western Blot | Protein ID | Purification Method | Purity (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RG-2 | 10 ml | 0.8 | 1.0 | 3 | 15 ℃ | Y | 2YT | | Y | Y | Y | - | - | - |
| RG-2 | 10 ml | 1.0 | 1.0 | 3 | 15 ℃ | Y | 2YT | | Y | Y | Y | - | - | - |
| BL21 pLysS | 10 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | | Y | Y | Y | - | - | - |
| BL21 * | 10 ml | 1.0 | 1.0 | 3 | 15 ℃ | Y | 2YT | | N | Y | Y | - | - | - |
| RG-2 | 2 x 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | Y | Strep II tag, 1 ml Strep Tactin® Superflow™ | 90 |
| RG-2 | 4 x 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | Y | - | Strep II tag, 5 ml StrepTrap HP | 90 |
| RG-2 | 2 x 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | N | - | Strep II tag, 1 ml StrepTrap HP | 70 |
| RG-2 | 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | Y | 2YT | 1 mM $CuCl_2$ | N | Y | Y | - | Strep II tag, 1 ml StrepTrap HP (with avidin and high salt wash) | - |
| RG-2 | 500 ml | N/A | - | 21 | 30 ℃ | Y | AIM | - | N | N | Y | - | Strep II tag, 1 ml StrepTrap HP (with avidin and high salt wash) | - |
| RG-2 | 4 x 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | N | - | Strep II tag, 1 ml StrepTrap HP (with avidin and high salt wash) | 90 |
| RG-2 | 12 x 500 ml | 1.0 | 1.0 | 3 | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | Y | N | - | Strep II tag, 1 ml StrepTrap HP (with avidin and high salt wash) | 80 |

Summary of experiments to heterologously express the *Limnoria* HC3 gene using the **pET26b(+)** vector. * did not bind to column; no soluble protein eluted from the purification.

| Expression Strain | Culture Scale | $OD_{600}$ at Induction | IPTG Conc. (mM) | Time Post Induction (hours) | Expression Temperature | Glucose Added | Media | Media Additives | Soluble | Insoluble | Western Blot | Protein ID | Purification Method | Purity (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RG-2 | 10 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | Y | Y | Y | - | - | - |
| RG-2 | 10 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | N | Y | Y | - | - | - |
| RG-2 | 100 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | N | Y | Y | - | - | - |
| RG-2 | 250 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | Y | Y | Y | - | - | - |
| RG-2 | 500 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | N | Y | Y | - | - | - |
| RG-2 | 2 x 250 ml | 1.0 | 1.0 | 24 hours | 15 ℃ | Y | 2YT | - | Y | Y | Y | - | S trep II tag, 1 ml StrepTrap HP (with avidin and high salt wash) | * |

Summary of experiments to heterologously express the *Limnoria* HC1 (*E. coli* codon optimised) gene using the **pET28a(+)** vector.

| Expression Strain | Culture Scale | $OD_{600}$ at Induction | IPTG Conc. (mM) | Time Post Induction (hours) | Expression Temperature | Glucose Added | Media | Media Additives | Soluble | Insoluble | Western Blot | Protein ID | Purification Method | Purity (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RG-2 | 2 x 500 ml | 1.0 | 0.1 | 3 hours | 15 ℃ | N | 2YT | 1 mM $CuCl_2$ | Y | N | Y | Y | - | 50 |

# Appendix D – Associated Publication

The proteomic data detailed in this thesis contributed towards the work described in the below publication.

**Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance.**

*Proceedings of the National Academy of Sciences of the United States of America (2013).*

Marcelo Kern, John E. McGeehan, Simon D. Streeter, Richard N. A. Martin, Katrin Besser, Luisa Elias, **William Eborall**, Graham P. Malyon, Christina M. Payne, Michael E. Himmel, Kirk Schnorr, Gregg T. Beckham, Simon M. Cragg, Neil C. Bruce, and Simon J. McQueen-Mason.

# Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance

Marcelo Kern[a,1], John E. McGeehan[b,1], Simon D. Streeter[b], Richard N. A. Martin[b], Katrin Besser[a], Luisa Elias[a], Will Eborall[a], Graham P. Malyon[b], Christina M. Payne[c,d], Michael E. Himmel[d], Kirk Schnorr[e], Gregg T. Beckham[f,2], Simon M. Cragg[b,2], Neil C. Bruce[a,2], and Simon J. McQueen-Mason[a,2]

[a]Centre for Novel Agricultural Products, Department of Biology, University of York, York YO10 5DD, United Kingdom; [b]School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DY, United Kingdom; [c]Department of Chemical and Materials Engineering, University of Kentucky, Lexington, KY 40506; [d]Biosciences Center and [f]National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401; and [e]Novozymes A/S, 2880 Bagsvaerd, Denmark

Edited by Alexis T. Bell, University of California, Berkeley, CA, and approved May 8, 2013 (received for review January 24, 2013)

**Nature uses a diversity of glycoside hydrolase (GH) enzymes to convert polysaccharides to sugars. As lignocellulosic biomass deconstruction for biofuel production remains costly, natural GH diversity offers a starting point for developing industrial enzymes, and fungal GH family 7 (GH7) cellobiohydrolases, in particular, provide significant hydrolytic potential in industrial mixtures. Recently, GH7 enzymes have been found in other kingdoms of life besides fungi, including in animals and protists. Here, we describe the in vivo spatial expression distribution, properties, and structure of a unique endogenous GH7 cellulase from an animal, the marine wood borer *Limnoria quadripunctata* (LqCel7B). RT-quantitative PCR and Western blot studies show that LqCel7B is expressed in the hepatopancreas and secreted into the gut for wood degradation. We produced recombinant LqCel7B, with which we demonstrate that LqCel7B is a cellobiohydrolase and obtained four high-resolution crystal structures. Based on a crystallographic and computational comparison of LqCel7B to the well-characterized *Hypocrea jecorina* GH7 cellobiohydrolase, LqCel7B exhibits an extended substrate-binding motif at the tunnel entrance, which may aid in substrate acquisition and processivity. Interestingly, LqCel7B exhibits striking surface charges relative to fungal GH7 enzymes, which likely results from evolution in marine environments. We demonstrate that LqCel7B stability and activity remain unchanged, or increase at high salt concentration, and that the *L. quadripunctata* GH mixture generally contains cellulolytic enzymes with highly acidic surface charge compared with enzymes derived from terrestrial microbes. Overall, this study suggests that marine cellulases offer significant potential for utilization in high-solids industrial biomass conversion processes.**

gribble | carbohydrate degrading enzymes

Increasing worldwide demand for energy and concerns over greenhouse gas emissions have driven the search for renewable alternatives to transportation fuels from fossil origins, and liquid biofuel production from cellulose, the most abundant biopolymer on Earth, has emerged as a promising option. Cellulose comprises the main structural component of plant biomass and is composed entirely of glucose, but the crystalline nature of this material makes its cost-effective conversion into fermentable sugars challenging (1–3). Cellulose is arranged as crystalline microfibrils cross-linked to hemicellulose and lignin to form a recalcitrant biopolymer matrix (4, 5). Improving methods of sugar release from lignocellulose by manipulation of biomass composition and new chemical treatments are areas of intense research (6); additionally, a key barrier to large-scale conversion of lignocellulose to biofuels is the limited availability of low-cost enzymes to efficiently depolymerize biomass from various sources while maximizing sugar release. Typically, both natural enzyme mixtures and industrial enzyme preparations for biomass deconstruction consist of a mixture of glycoside hydrolases (GHs) that exhibit synergistic properties (7). Filamentous fungi in particular have traditionally been used as a source of GHs, due to their ability to produce high titers of secreted, active cellulolytic enzymes (8, 9). In addition to fungi, bacteria and arthropods, including termites, cockroaches, and beetles, are also recognized for their lignocellulose hydrolysis capabilities (10–12). More recently, the hydrolytic capacity of the wood-boring crustacean *Limnoria quadripunctata*, which thrives on a diet of wood, was revealed. Unusually, *Limnoria* do not rely on gut microbes for wood digestion, but instead have a digestive system free of microbial populations, and digest lignocellulose using endogenous enzymes (13). At least three classes of GHs are necessary to deconstruct cellulose (7), and analysis of a cDNA library constructed from the *L. quadripunctata* digestive gland revealed candidate transcripts that putatively represent all three classes of hydrolases, namely cellobiohydrolases, endoglucanases, and β-glucosidases. Specifically, endoglucanases cleave internal points of cellulose chains in accessible regions to create new, free chain ends; cellobiohydrolases processively hydrolyze cellulose chains from either the reducing or nonreducing end and release cellobiose as the major product; and β-glucosidases hydrolyze cellobiose to glucose. The exposure of chain ends by endoglucanases for cellobiohydrolase attachment and detachment is an important synergism between these two cellulase classes, and the hydrolysis of cellobiose by β-glucosidases mitigates cellobiose inhibition (7, 14–16). The *L. quadripunctata* transcriptome was dominated by a few GH families, with transcripts representing GH7s being particularly abundant, representing ∼12% of total ESTs sequenced. This was unexpected as GH7s were previously only reported from fungi and symbiont protists from termites, but not from animals (13).

GH family 7 cellobiohydrolases are of particular importance, as they typically represent the most significant hydrolytic potential

Author contributions: M.K., J.E.M., K.S., G.T.B., N.C.B., and S.J.M.-M. designed research; M.K., J.E.M., S.D.S., R.N.A.M., K.B., L.E., W.E., G.P.M., C.M.P., K.S., and G.T.B. performed research; M.K., J.E.M., S.D.S., C.M.P., M.E.H., G.T.B., S.M.C., N.C.B., and S.J.M.-M. analyzed data; and M.K., J.E.M., C.M.P., G.T.B., S.M.C., N.C.B., and S.J.M.-M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4GWA, 4HAP, 4HAQ, and 4IPM) and the sequence reported in this paper has been deposited in the GenBank database (accession no. KC776193).

[1]M.K. and J.E.M. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: simon.mcqueenmason@york.ac.uk, gregg.beckham@nrel.gov, Simon.Cragg@port.ac.uk, or neil.bruce@york.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1301502110/-/DCSupplemental.

in many natural and industrial enzyme mixtures (7), and most GH7s studied to date originate from fungi. Along with the discovery of complementary enzymes (17–21), prospecting and engineering GH7s for higher activity represents one of the most important research activities to improve enzyme mixtures, which will in turn significantly aid in the commercialization of biofuel processes based on enzymatic depolymerization of polysaccharides (2, 7, 8). To that end, here, we examine multiple aspects of lignocellulose breakdown by a GH7 cellulase from *L. quadripunctata* (LqCel7B). We use reverse transcription, quantitative PCR (RT-qPCR) and Western blot analysis to demonstrate that LqCel7B is expressed within the *L. quadripunctata* hepatopancreas and subsequently secreted into the gut for biomass digestion. To understand the ability of LqCel7B to operate in saline conditions, we use X-ray crystallography to solve four high-resolution LqCel7B structures and then compare the LqCel7B structure to the well-characterized *Hypocrea jecorina* Cel7A (HjCel7A) structure using molecular dynamics (MD) simulations. The structures overall exhibit several key differences to HjCel7A, including a strikingly acidic surface charge and a tyrosine residue at the tunnel entrance, which likely plays a role in substrate acquisition and processivity. We also demonstrate that LqCel7B is a cellobiohydrolase rather than an endoglucanase due to its structure and activity profile on several model cellulose substrates, and that LqCel7B is active at high salt concentrations, likely due to its high surface charge distribution. Based on the results presented here for LqCel7B, we examine the sequences of known enzymes in the cellulolytic mixture of *L. quadripunctata*, which suggest that the additional cellulolytic enzymes from this organism all exhibit high surface charge compared with GHs from terrestrial and freshwater organisms. This study suggests that *L. quadripunctata* and other marine organisms represent interesting targets for prospecting cellulolytic enzymes for industrial biomass conversion and for understanding cellulose conversion in marine environments.

## Results

**Distribution of LqCel7B in Digestive System Organs.** Limnoriids have a bipartite digestive system in which enzymes are produced in the hepatopancreas, and subsequently secreted into the lumen of this gland and then transferred to the hindgut, their presumptive site of action. The distribution and abundance of LqCel7B in the *L. quadripunctata* digestive system was investigated through RT-qPCR and Western blot analysis (Fig. 1*A*). RT-qPCR studies demonstrated that *LqCel7B* transcripts were abundant in the hepatopancreas, but low or absent from the hindgut and rest of the body. In contrast, probing protein extracts with anti-LqCel7B antibodies revealed a unique 55-kDa band in both gut and hepatopancreas extracts (Fig. 1*A*). SEM of the gut tissue showed the hindgut lumen tightly packed with fine wood particles (Fig. 1*B*), indicating that LqCel7B functions in a high-solids loading environment. These studies not only provide direct evidence that *LqCel7B* gene expression is confined to the hepatopancreas, as previously suggested by in situ hybridization (13), but that the protein is subsequently translocated into the hindgut.

**Protein Production, Physical Properties, and Enzyme Kinetics.** LqCel7B, a single domain protein lacking a carbohydrate-binding module (CBM) and linker sometimes observed in GH7 enzymes, was expressed in *Aspergillus oryzae* and *Aspergillus niger* and subsequently purified using anion exchange and size exclusion chromatography (Fig. S1 *A* and *B*). Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) analyses of purified samples confirmed all peptides detected matched the LqCel7B sequence (*SI Results, LC-MS/MS Analysis*). The protein behaves as a monodisperse monomer with a molecular weight of 45 kDa from size exclusion chromatography multiangle laser light scattering, corresponding closely to the 46.5 kDa predicted from the sequence of the signal-peptide cleaved protein (Fig. S1*C*). Velocity analytical ultracentrifugation (AUC) confirmed the monomeric nature in solution and yielded a sedimentation coefficient of 4.45 S (Fig. S1*D*). pH and temperature optima were determined using the chromogenic substrates *p*-nitrophenyl-β-D-cellotrioside (*p*NP-G3), and *p*-nitrophenyl-β-D-cellopentaoside (*p*NP-G5) (Fig. S2 *A* and *B*). LqCel7B exhibits a variable pH optima range (Fig. S2*A*): pH 4–6.5 for *p*NP-G3 and pH 4–8 for *p*NP-G5. High activity was maintained over a wide temperature range (Fig. S2*B*). Furthermore, the kinetic properties of LqCel7B were determined using *p*NP-G3 and *p*NP-G5 (Fig. S2*C*) due to very low activity on *p*-nitrophenyl-β-D-cellobioside (*p*NP-G2) (Fig. S2*D*). The higher $V_{max}$ and $k_{cat}$ together with lower $K_m$ and higher catalytic efficiency ($k_{cat}/K_m$) indicate that *p*NP-G5 is a better substrate. The LqCel7B $K_m$ value for *p*NP-G5 is within the range reported for other GH7 cellobiohydrolases (22–26). The inhibitory effect of cellobiose, which is commonly observed in GH7 enzymes from fungi (15, 16, 26–28), on the hydrolytic activity of LqCel7B was investigated using *p*NP-G5 (Fig. S2*F*), and a $K_i$ value of 1.40 mM was measured.

**LqCel7B Substrate Specificity.** LqCel7B hydrolytic activity was tested on phosphoric acid swollen cellulose (PASC) and Avicel, respectively, at low [0.1% (wt/vol) PASC or Avicel] and high [1% (wt/vol) PASC or 2% (wt/vol) Avicel] solids loadings (Table 1). The highest rate of substrate conversion was detected in the high-solids loading assays, possibly because of the lack of a CBM, which results in more association events due to reduced diffusion limitations (14, 29). A similar value for PASC hydrolysis ratio was reported for HjCel7A and *Thermoascus aurantiacus* Cel7A (assayed with CBMs) as well as *Melanocarpus albomyces* Cel7B (which naturally lacks a CBM), whereas LqCel7B showed lower Avicel conversion ratios compared with the same cellobiohydrolases without CBMs (30, 31). Also, the main product from PASC and Avicel hydrolysis was cellobiose (G2), followed by glucose (G1) and trace cellotriose (G3). G1 and G2 were also the main reaction products detected by high performance anion exchange chromatography (HPAEC) analysis on hydrolysis of

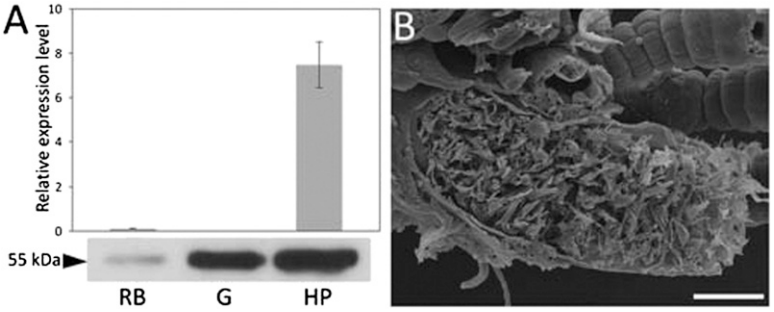



**Fig. 1.** Abundance and functional environment of the cellobiohydrolase LqCel7B in the *L. quadripunctata* digestive system. (*A*) Reverse transcription, quantitative PCR (RT-qPCR) using RNA extracted from rest of the body (RB), hindgut (G), and hepatopancreas (HP). Expression levels presented are normalized to ubiquitin (*Upper*). Western blot of protein extracts prepared from these organs probed with purified anti-LqCel7B (*Lower*). (*B*) Scanning electron micrograph showing an obliquely sectioned hindgut packed with wood particles. (Scale bar, 50 μm.)

**Table 1. Substrate conversion ratio and processivity estimated for LqCel7B**

| | | PASC | | Avicel | |
|---|---|---|---|---|---|
| [S] | Time (h) | Conversion (%) | Processivity | Conversion (%) | Processivity |
| Low loading | 1 | 3.69 ± 0.42 | 9.20 ± 0.12 | 0.30 ± 0.02 | 4.74 ± 0.26 |
| | 24 | 5.02 ± 0.68 | 10.65 ± 0.71 | 0.43 ± 0.05 | 6.95 ± 0.96 |
| High loading | 1 | 5.21 ± 0.72 | 6.62 ± 0.07 | 0.22 ± 0.01 | 4.63 ± 0.15 |
| | 24 | 10.34 ± 1.14 | 3.88 ± 0.14 | 0.66 ± 0.11 | 3.11 ± 0.04 |

Low loading: 0.1% (wt/vol) PASC or Avicel; high loading: 1% (wt/vol) PASC; 2% (wt/vol) Avicel. Processivity = G2/(G1 + G3). Conversion rates shown are a product mass percentage relative to substrate mass. [S] denotes substrate concentration. All values were determined by HPAEC and represent mean ± SE.

cellooligosaccharides (Fig. S2*G*). As G2 is produced primarily through processive hydrolysis, whereas G1 and G3 are the main products released upon initial hydrolysis events (32), processivity was semiquantitatively estimated as the ratio G2/(G1 + G3) (33, 34). Processivity ratios above 7 were achieved in almost all reactions containing PASC and in extended reactions containing Avicel at a low-solids loading, which are within the range measured for *Phanerochaete chrysosporium* Cel7D and HjCel7A (33). Additionally, LqCel7B was able to hydrolyze carboxymethyl cellulose (CMC), albeit at much lower conversions than PASC (Fig. S2*E*), indicating that LqCel7B can conduct endoinitiation, which is a common feature of most GH7 cellobiohydrolases (35, 36). We note that more "open" GH7 cellobiohydrolases also exhibit low activity on CMC (37). These results suggest that LqCel7B is a cellobiohydrolase, which is supported by the structural characteristics described below.

**Structural and Simulation Studies.** The structural characterization of LqCel7B is based on the solution of four X-ray structures, one apo form at 1.6-Å resolution, and three ligand bound complexes ranging from 1.9 Å to 1.14 Å. The lattermost structure of LqCel7B at 1.14-Å resolution in complex with thiocellobiose reveals close to atomistic detail from the high-quality electron density maps (Fig. S3). Crystallography statistics of the four models are presented in Table S1. The apo structure is shown in Fig. 2*A* with the cellononaose ligand from the HjCel7A structure (PDB ID code 8CEL) docked into the tunnel. Four highly conserved tryptophan residues in GH7 cellobiohydrolases are in yellow, and the catalytic residues are in blue. The overall fold is homologous to other GH7 cellobiohydrolases, comprising a globular and equatorially elongated protein, but with several interesting features particularly in the surface loop regions (Fig. 2*B*) (33, 38–41). The active site tunnel, formed by four pairs of curved antiparallel β-sheets packed face to face, extends from the −7 binding site at Trp57 to the +2 binding site near Phe362, representing a distance of ∼48 Å between the tunnel entrance and exit. The conserved triad of catalytic residues (EXDXXE; Fig. S4*A*), Glu233, Asp235, and Glu238 use a two-step, retaining mechanism for glycosidic bond hydrolysis (11, 33, 38–43).

Three holo crystal structures not only provide reference sites for docking of the full cellononaose ligand, but a detailed picture of the hydrogen bonding networks at both the product and entrance sites. Cocrystallization of LqCel7B with cellopentaose resulted in a structure with cellobiose bound in the +1 and +2 product sites, anchored through a combination of hydrogen bonds between the hydroxyl groups of cellobiose and the glycosyl binding residues Arg271, His249, Arg417 and the catalytic residues Asp235 and Glu238 as well as hydrophobic stacking between Trp401 and the pyranose rings of cellobiose (Fig. S4*B*). The absence of cellopentaose and presence of cellobiose indicates that activity remained intact during crystallization. Additionally, thiocellobiose was cocrystallized with LqCel7B crystals in an alternative $P2\ 2_1\ 2_1$ space group by seeding trials (Fig. S3). Thiocellobiose is located in the same position as the cellobiose in the previous structure, potentially contributing a stabilizing role in the crystal and aiding data collection at high resolution. Further detailed features were apparent in these electron density maps, including dual conformation of the proton donor Glu238 in the active site, an observation that correlated directly with MD simulations (Fig. S5*F*). Additionally, soaking native crystals of the apo form of LqCel7B with cellotriose, revealed both cellobiose in the product site as before, with the addition of cellotriose in the active site entrance from the −7 to −5 sites, stabilized via hydrogen bonding occurring between Asp124 and Lys202 and hydrophobic stacking between Trp55 and Trp57 and the pyranose rings (Fig. S4*C*).

Fig. 2*B* shows the structural alignment of LqCel7B and the HjCel7A 8CEL crystal structure (38) and Fig. S6 *B* and *C* show the surface representation of the same structures. The LqCel7B active site tunnel is mostly enclosed, as opposed to the more exposed cleft of the endoglucanase HjCel7B (PDB ID code 1EG1; Fig. S6*A*), and it is, therefore, similar to other known structures of GH7 cellobiohydrolases (38–41), suggesting that LqCel7B is a cellobiohydrolase as indicated by the activity measurements described above (35). Structurally, the most significant differences in the active site tunnels between LqCel7B and HjCel7A are observed at the −7 subsite, the exo loop (33), and the product sites, which are discussed below in turn.

As shown in Fig. 2*C*, the −7 subsite in LqCel7B exhibits a potential, additional aromatic-carbohydrate interaction site due to Tyr121. In the apo structure of LqCel7B, Tyr121 is significantly displaced from the −7 subsite (pink) when the ligand is not present. However, in the LqCel7B structure with the cellotriose ligand bound in the −7 to −5 sites, Tyr121 is bound directly to the glucose residue in the −7 site (gray). Additionally, rapid binding of Tyr121 to the −7 glucose unit is observed (Fig. 2*C*, in blue) in MD simulations constructed from the LqCel7B apo structure with the 8CEL ligand docked into the active site tunnel. As shown in Fig. 2*C*, excellent agreement is observed between MD simulations and the structural data with Tyr121-glucosyl binding at the −7 site. Further quantitative characterization of these results is provided in Fig. S5 *A–D*. Lastly, the Tyr121 residue is observed in the MD simulations to undergo a dihedral rotation about the $\chi_1$ dihedral such that the aromatic ring of Tyr121 binds to the sugar in the −6 site (Fig. S5*E* and Movie S1). This dihedral rotation in turn induces larger fluctuations in the ligand, as shown in Fig. 2 *D* and *E*. This additional aromatic binding platform at the entrance of a GH7 tunnel has not been observed previously in contact with a ligand, although it has been suggested to play a role in guiding a cellulose chain into the tunnel in previous structural studies of the apo structure of *M. albomyces* (40) and in a recent computational study of the *Heterobasidion irregulare* cellobiohydrolase Cel7A (44). These results suggest that an additional aromatic residue at the tunnel entrance may play a role in substrate acquisition and processivity.

Additionally, Fig. 2*D* shows that the exo loop (33) in LqCel7B is able to undergo significant conformational changes relative to the exo loop in HjCel7A. Fig. 2*F* shows the distance between the exo loops and the corresponding loop across the enzyme tunnel as a function of time. As shown, the HjCel7A loop does not open during the 250-ns MD simulation (44), whereas the LqCel7B exo loop undergoes a conformational change between a closed and open state quite readily on the nanosecond time scale. This difference in exo-loop behavior is likely to increase the propensity of LqCel7B to engage in endotype initiation on the cellulose surface relative to HjCel7A (35), as suggested in previous studies of *P. chrysosporium* Cel7D and HjCel7A (35, 44). Lastly, we note that there are significant protein fluctuations at the product site

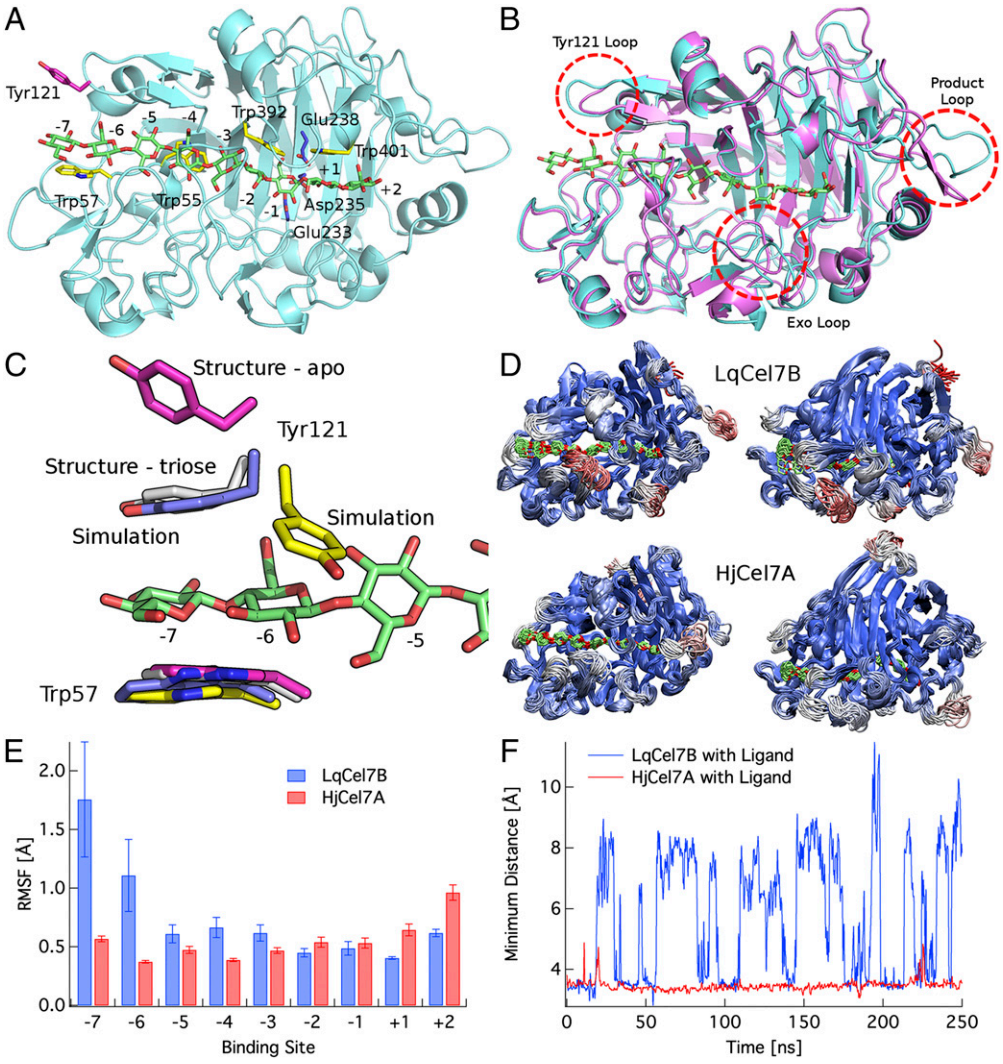



**Fig. 2.** Structural and computational analysis of LqCel7B. (*A*) Overall fold of the apo LqCel7B structure with the catalytic residues highlighted in blue and a selection of the planar Trp residues that line the ligand binding tunnel in yellow. Tyr121 is shown in pink. The cellononaose ligand is docked into the LqCel7B structure from the HjCel7A structure and is shown in green (38). (*B*) Structural comparison of the apo LqCel7B structure and HjCel7A, with the ligand from 8CEL (38). (*C*) View of the −7 to −5 binding sites in two crystallographic and two simulation conformations, including the ligand from 8CEL (38) in green. Tyr121 binds directly to the ligand in the −7 glucose binding site. The color coding for Tyr121 and Trp57 is as follows: pink, apo structure conformations; gray, cellotriose LqCel7B structure; blue, MD simulation snapshot of LqCel7B with Tyr121 stacking on the −7 glucose unit; and yellow, MD simulation snapshot of LqCel7B with Tyr121 binding to the −6 glucose unit. Fig. S5*E* shows the cluster view of Tyr121 over the 250-ns MD simulation. (*D*) Cluster representations of LqCel7B and HjCel7A over a 250-ns MD simulation. HjCel7A data are taken from ref. 44, and the enzymes are colored by root mean square fluctuations (RMSFs), wherein red represents fluctuations up to 4 Å and blue represents the minimum fluctuations. The Root square mean deviation (RSMD) and RMSF of both enzymes are shown in Fig. S5 *A*–*C*. (*E*) RMSF of LqCel7B and HjCel7A (44) ligands. Error bars represent one SD from the average value measured by block averaging. (*F*) Time series of the exo-loop opening and closing events in LqCel7B and HjCel7A MD simulations over 250 ns with a cellononaose ligand present (33, 38). The exo-loop motions are the minimum distance between the exo loop and the corresponding loops on the other side of the tunnel: for LqCel7B, the minimum distance from the exo loop (residues 265–274) to residues 394–399 and for HjCel7A, the minimum distance from the exo loop (residues 244–253) to residues 369–373. A histogram of this time series is shown in Fig. S5*D*.

of LqCel7B not observed in HjCel7A. This, however, does not affect ligand fluctuations (Fig. 2*E*), but may impact product inhibition, which will be examined in future work. An illustration of the LqCel7B bound to cellulose is shown in Movie S2.

**LqCel7B Salt Tolerance.** Analysis of the surface properties of LqCel7B reveals a striking predominance of negatively charged residues as shown in Fig. 3*A*. Homology modeling and surface charge analysis suggest that the GH7 enzyme from the marine crustacean *Chelura terebrans* exhibits negatively charged residues that create an almost continuous acidic surface, as shown in Fig. 3*B*. Conversely, the homology model constructed for the GH7 ortholog from the related freshwater crustacean *Daphnia magna* exhibits a more neutral composition with both acidic and basic residues less numerous and more evenly distributed in common with the fungal HjCel7A (Fig. 3 *C* and *D*). The high number of acidic surface residues also endows the enzyme with a low isoelectric point (pI), far below that calculated for the vast majority of other identified GH7s as shown through a comparison of pI values across GH7 family members (Fig. 3*E*). The acidic nature of *Limnoria* and *Chelura* GH7s may be a consequence of their marine environment and the physiological need of the enzymes to remain active in the gut of these crustaceans, which likely contain seawater. This hypothesis was further corroborated by the construction of homology models of the entire cellulolytic mixture of *L. quadripunctata*, which suggests that the additional known GH7, GH5, and GH9 enzymes from this organism all have high negative surface charges compared with terrestrial microbial equivalents (Fig. S6 *F*–*H*). Highly negative electrostatic potential surface is one of the hallmarks of halophilic proteins, which not only require salts for optimum activity, but also remain active at high ionic strength values (45). We also examined the hydrolytic activity of LqCel7B in high salt conditions using PASC and Avicel as substrates (Fig. 3*F*). Starting at 0 M NaCl, then 0.5 M to reflect the ionic strength of seawater, then progressively increasing concentrations up to 4 M, NaCl does not inhibit enzyme activity, and activity increases significantly at some salt concentrations.

## Discussion and Conclusions

LqCel7B is a uniquely and extensively characterized nonfungal GH7 cellulase, and it exhibits a number of features that differ from well-characterized fungal examples of this important cellulase family. From a protein sequence standpoint, previous work showed that the Limnoriid GH7s formed a well-separated

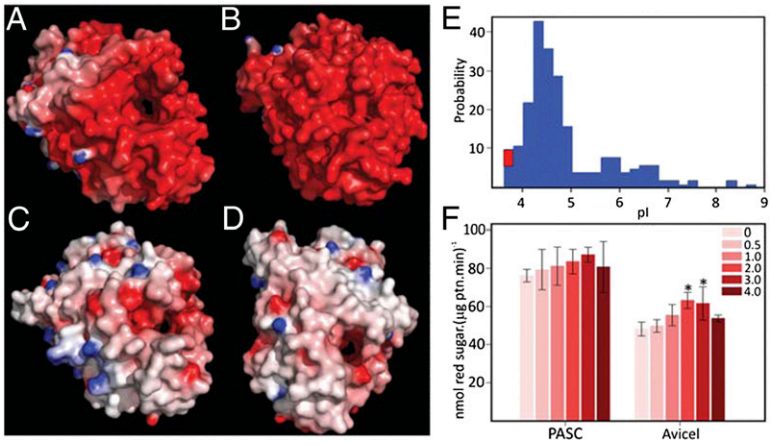



**Fig. 3.** LqCel7B salt tolerance. Electrostatic potential distribution on the solvent accessible surface of *Limnoria quadripunctata* LqCel7B (*A*), *Chelura terebrans* GH7 (*B*), *Daphnia magna* GH7 (*C*), and *Hypocrea jecorina* Cel7A (*D*). Electrostatic potential between −7 kT/e and 7 kT/e was shown as a colored gradient from red (acidic) to blue (basic). (*E*) pI clusters showing GH7s from three GH7 *L. quadripunctata* enzymes grouped together with the *C. terebrans* ortholog (red square) as a unique cluster at low pI value (3.8 < pI < 3.9). (*F*) Hydrolytic activity of LqCel7B as a function of sodium chloride concentration expressed as reducing sugar released from PASC and Avicel. Asterisks represent statistically significant values following one-way ANOVA DSF (*$P < 0.01$).

branch from fungal and protist equivalents in phylogenetic analysis of protein sequences (13). The lack of a CBM is intriguing and together with the high abundance of LqCel7B in the gut lumen (Fig. 1) could reflect an evolutionary adaptation to the gut environment, which is densely packed with wood particles (13). CBMs are known to increase the concentration of the enzymes in the vicinity of the substrate, but because of the high substrate loading in the gut (Fig. 1*B*), the need for a CBM to bind the enzyme to the cellulose surface has presumably been made redundant (Fig. S6 *D* and *E*) (14, 29). The absence of a CBM may also permit the retention of the enzyme after wood digestion by the animal rather than being lost with the excreted, processed biomass. Despite not possessing a CBM, the kinetic properties of the LqCel7B and its extent of processivity are broadly similar to other GH7 cellobiohydrolases.

An intriguing feature of LqCel7B is its ability to retain activity at high salt concentrations, a property that could lend itself to saccharification following or during biomass pretreatment with ionic liquids (ILs), which are strong enzyme-denaturing agents (46). For this reason, IL pretreated biomass requires rinsing in water before enzymatic hydrolysis. Alternatively, direct saccharification of IL pretreated biomass without separate cellulose recovery and wash has also been attempted but enzyme activity was dependent on IL dilution (47). It has been shown that a GH5 enzyme from a halophilic archeon, with a synonymous pattern of acidic surface charges to LqCel7A, demonstrated significant IL compatibility (48). In either approach, LqCel7B and other marine cellulases could represent promising candidate enzymes. In summary, LqCel7B is a unique functionally characterized animal family GH7 enzyme, and the data presented in this report demonstrate unusual features that likely reflect the environment within which the enzyme operates.

## Methods

Methods on protein extraction from the *Limnoria* digestive system and Western blot analysis, RNA extraction and reverse transcription quantitative PCR, biophysical analysis, enzyme kinetics, and HPAEC are discussed in *SI Methods*.

**Expression and Purification.** LqCel7B was expressed in *A. oryzae* (kindly provided by Novozymes, Bagsvaerd, Denmark) and in *A. niger* culture supernantants. One hundred milliliters of the *A. oryzae* culture supernatant was sterile filtered and concentrated to 10 mL using a polyethersulfone Ultraconcentrator MWCO 10 kDa (Novagen) by centrifugation at 3,450 × *g* rpm at 4 °C. The concentrated fraction was then desalted and buffer exchanged into 20 mM Tris pH 8.0 containing 50 mM NaCl using a Zeba Desalt Spin column (Pierce Biotechnology). The desalted supernatant was purified through a 5-mL anion exchange Q-Trap column (GE Healthcare) using an Äkta 100 (GE Healthcare), equilibrated in 20 mM Tris pH 8.0, 50 mM NaCl. The enzyme was eluted with a 0.05–1 M NaCl gradient in the same buffer at a flow rate of 1 mL/min and the elution was monitored through absorbance at 280 nm. Eluted fractions containing absorbance peaks were analyzed by SDS/PAGE to confirm the presence of the recombinant protein. *A. niger* protein production protocol is described in *SI Methods*.

**Enzyme Assays.** The standard concentration of the polymeric substrates in the reactions was 0.1% and 1% (wt/vol) PASC or CMC and 0.1% and 2% (wt/vol) Avicel. All reactions containing oligosaccharides were carried out at 30 °C for 4 h or, alternatively, for 1 h or 24 h for polysaccharides. All reactions were conducted in 100 mM citrate buffer pH 4.5 (for *p*NP- and cellooligomer substrates) or 5.0 (for polysaccharide substrates), 500 mM NaCl, and 1 μg (*p*NP- and cellooligomer reactions) or 10 μg (polysaccharide reactions) of purified LqCel7B. Enzyme activity was measured according to release of *p*NP ions at 405 nm or release of reducing sugars at 620 nm (49).

**Crystallography.** Purified LqCel7B was concentrated to 15 mg/mL and crystallized by vapor diffusion in 0.2 M $MgCl_2$, 0.1 M Bis·Tris pH 5.0 and 25% (wt/vol) PEG 3350 at 289 K. The crystals were cryoprotected in a solution of 25% (vol/vol) glycerol and cryocooled in liquid nitrogen before data collection. Cocrystallization was performed by incubating LqCel7B in a 1:1.5 molar ratio with cellopentaose. The crystals were grown by vapor diffusion in 0.1 M $CH_3CO_2Na$ pH 4.5, 15% (wt/vol) PEG 6000 and 0.5 M $CaCl_2$ at 289 K. The crystals were transferred to a cryoprotectant containing 6.25% (vol/vol) diethylene glycol, 6.25% (wt/vol) 2-Methyl-2,4-pentanediol (MPD), 18.75% (vol/vol) 1,2-propanediol and 6.25% (vol/vol) DMSO, and cryocooled in liquid nitrogen. Free protein crystals were soaked with 1:10 molar ratio of cellotriose for 90 min before transfer to the same cryoprotectant for cryocooling. Cocrystallization was performed in buffer containing up to a 20-fold molar excess of thiocellobiose with LqCel7B at a concentration of 15 mg/mL. Several rounds of micro- and macroseeding were used to generate larger crystals. The cryoprotectant was 6.25% (vol/vol) diethylene glycol, 6.25% (vol/vol) MPD, 6.25% (vol/vol) 1,2-propanediol, 12.5% (vol/vol) DMSO, 6.25% (vol/vol) glycerol and 6 mM 3-(1-pyridinio)-1-propanesulfonate and contained 5 mM thiocellobiose. Data were processed using either XDS and XSCALE (50) or MosFlm (51) and SCALA (52) and molecular replacement was carried out with Phaser (53) using the *Talaromyces emersonii* cellobiohydrolase IB (CBH IB) structure (PDB accession code: 1Q9H). Reiterative cycles of model building and refinement were carried out using COOT (54) and REFMAC (55) or Phenix (56), respectively. Figures were produced with the program PyMol with electrostatic surfaces calculated using DELPHI (57).

**MD Simulations.** The 1.6-Å resolution apo LqCel7B structure was used to build the LqCel7B simulation input, and the HjCel7A 8CEL ligand was threaded into the active site via protein backbone alignment (38). The simulation setup followed the protocols outlined by Bu et al. (58) and Momeni et al. (44). MD simulations were conducted as described in detail in *SI Methods*. Briefly, the TIP3P potential was used for water (59) and the CHARMM force fields were used for the protein and carbohydrates (60–62). Periodic boundary conditions were applied and the particle mesh Ewald method was used to treat long-range electrostatics (63). MD simulations were conducted for 250 ns with a 2-fs time step.

**ACKNOWLEDGMENTS.** We thank the staff at the Diamond Light Source for their assistance, Jo Diamond for SEM work, Clare Steele-King for critical reading of the manuscript, and Jerry Ståhlberg and Mats Sandgren for helpful

discussions. This work was funded by Biotechnology and Biological Sciences Research Council (BBSRC) Grant BB/G016178/1. Visits between the research teams were supported by BBSRC–US Partnering Award BB/H531543/1. G.T.B., M.E.H., and C.M.P. acknowledge the US Department of Energy's BioEnergy Technologies Office for funding. Computer time for this research was provided by the National Institute for Computational Sciences Kraken cluster under the National Science Foundation XSEDE Grant MCB090159 and by the National Renewable Energy Laboratory Computational Sciences Center supported by the Department of Energy, Energy Efficiency and Renewable Energy under Contract DE-AC36-08GO28308.

1. Somerville C (2006) Cellulose synthesis in higher plants. *Annu Rev Cell Dev Biol* 22:53–78.
2. Himmel ME, et al. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* 315(5813):804–807.
3. Carroll A, Somerville C (2009) Cellulosic biofuels. *Annu Rev Plant Biol* 60:165–182.
4. Burton RA, Gidley MJ, Fincher GB (2010) Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat Chem Biol* 6(10):724–732.
5. Chundawat SPS, Beckham GT, Himmel ME, Dale BE (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu Rev Chem Biomol Eng* 2:121–145.
6. Geddes CC, Nieves IU, Ingram LO (2011) Advances in ethanol production. *Curr Opin Biotechnol* 22(3):312–319.
7. Percival Zhang YH, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol Adv* 24(5):452–481.
8. Wilson DB (2011) Microbial diversity of cellulose hydrolysis. *Curr Opin Microbiol* 14(3): 259–263.
9. Baldrian P, Valásková V (2008) Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiol Rev* 32(3):501–521.
10. Geib SM, et al. (2008) Lignin degradation in wood-feeding insects. *Proc Natl Acad Sci USA* 105(35):12932–12937.
11. Watanabe H, Tokuda G (2010) Cellulolytic systems in insects. *Annu Rev Entomol* 55: 609–632.
12. Bayer EA, Lamed R, White BA, Flint HJ (2008) From cellulosomes to cellulosomics. *Chem Rec* 8(6):364–377.
13. King AJ, et al. (2010) Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci USA* 107(12):5345–5350.
14. Jalak J, Kurašin M, Teugjas H, Väljamäe P (2012) Endo-exo synergism in cellulose hydrolysis revisited. *J Biol Chem* 287(34):28802–28815.
15. Gruno M, Väljamäe P, Pettersson G, Johansson G (2004) Inhibition of the *Trichoderma reesei* cellulases by cellobiose is strongly dependent on the nature of the substrate. *Biotechnol Bioeng* 86(5):503–511.
16. Schülein M (1997) Enzymatic properties of cellulases from *Humicola insolens*. *J Biotechnol* 57(1-3):71–81.
17. Vaaje-Kolstad G, et al. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330(6001):219–222.
18. Quinlan RJ, et al. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA* 108(37):15079–15084.
19. Phillips CM, Beeson WT, Cate JH, Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol* 6(12):1399–1406.
20. Westereng B, et al. (2011) The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PLoS ONE* 6(11):e27807.
21. Selig MJ, Knoshaug EP, Adney WS, Himmel ME, Decker SR (2008) Synergistic enhancement of cellobiohydrolase performance on pretreated corn stover by addition of xylanase and esterase activities. *Bioresour Technol* 99(11):4997–5005.
22. Igarashi K, Samejima M, Eriksson KEL (1998) Cellobiose dehydrogenase enhances *Phanerochaete chrysosporium* cellobiohydrolase I activity by relieving product inhibition. *Eur J Biochem* 253(1):101–106.
23. Lahjouji K, et al. (2007) Biochemical and molecular characterization of a cellobiohydrolase from *Trametes versicolor*. *Appl Microbiol Biotechnol* 75(2):337–346.
24. Lee K-M, et al. (2011) Production and characterization of cellobiohydrolase from a novel strain of *Penicillium purpurogenum* KJS506. *Appl Biochem Biotechnol* 163(1): 25–39.
25. Ståhlberg J, et al. (1996) Activity studies and crystal structures of catalytically deficient mutants of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* 264(2):337–349.
26. Tuohy MG, et al. (2002) Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii*. *Biochim Biophys Acta* 1596(2):366–380.
27. Lee K-M, et al. (2011) Characterization of cellobiohydrolase from a newly isolated strain of *Agaricus arvencis*. *J Microbiol Biotechnol* 21(7):711–718.
28. Shin K, Kim YH, Jeya M, Lee J-K, Kim Y-S (2010) Purification and characterization of a thermostable cellobiohydrolase from *Fomitopsis pinicola*. *J Microbiol Biotechnol* 20(12):1681–1688.
29. Zhang J, Moilanen U, Tang M, Viikari L (2013) The carbohydrate-binding module of xylanase from *Nonomuraea flexuosa* decreases its non-productive adsorption on lignin. *Biotechnol Biofuels* 6(1):18.
30. Voutilainen SP, et al. (2008) Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases. *Biotechnol Bioeng* 101(3):515–528.
31. Szijártó N, et al. (2008) Hydrolysis of amorphous and crystalline cellulose by heterologously produced cellulases of *Melanocarpus albomyces*. *J Biotechnol* 136(3-4):140–147.
32. Fox JM, Levine SE, Clark DS, Blanch HW (2012) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry* 51(1):442–452.
33. von Ossowski I, et al. (2003) Engineering the exo-loop of Trichoderma reesei cellobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D. *J Mol Biol* 333(4):817–829.
34. Horn SJ, et al. (2006) Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc Natl Acad Sci USA* 103(48):18089–18094.
35. Kurašin M, Väljamäe P (2011) Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* 286(1):169–177.
36. Ståhlberg J, Johansson G, Pettersson G (1993) *Trichoderma reesei* has no true exocellulase: All intact and truncated cellulases produce new reducing end groups on cellulose. *Biochim Biophys Acta* 1157(1):107–113.
37. Uzcategui E, Ruiz A, Montesino R, Johansson G, Pettersson G (1991) The 1,4-beta-D-glucan cellobiohydrolases from Phanerochaete chrysosporium. I. A system of synergistically acting enzymes homologous to Trichoderma reesei. *J Biotechnol* 19(2-3):271–285.
38. Divne C, Ståhlberg J, Teeri TT, Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from Trichoderma reesei. *J Mol Biol* 275(2):309–325.
39. Grassick A, et al. (2004) Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii*. *Eur J Biochem* 271(22):4495–4506.
40. Parkkinen T, Koivula A, Vehmaanperä J, Rouvinen J (2008) Crystal structures of Melanocarpus albomyces cellobiohydrolase Cel7B in complex with cello-oligomers show high flexibility in the substrate binding. *Protein Sci* 17(8):1383–1394.
41. Ubhayasekera W, Muñoz IG, Vasella A, Ståhlberg J, Mowbray SL (2005) Structures of *Phanerochaete chrysosporium* Cel7D in complex with product and inhibitors. *FEBS J* 272(8):1952–1964.
42. Davies G, Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3(9):853–859.
43. White A, Rose DR (1997) Mechanism of catalysis by retaining beta-glycosyl hydrolases. *Curr Opin Struct Biol* 7(5):645–651.
44. Momeni MH, et al. (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregulare. *J Biol Chem* 288(8):5861–5872.
45. Delgado-García M, Valdivia-Urdiales B, Aguilar-González CN, Contreras-Esquivel JC, Rodríguez-Herrera R (2012) Halophilic hydrolases as a new tool for the biotechnological industries. *J Sci Food Agric* 92(13):2575–2580.
46. Vancov T, Alston AS, Brown T, McIntosh S (2012) Use of ionic liquids in converting lignocellulosic material to biofuels. *Renew Energy* 45:1–6.
47. Kamiya N, et al. (2008) Enzymatic in situ saccharification of cellulose in aqueous-ionic liquid media. *Biotechnol Lett* 30(6):1037–1040.
48. Zhang T, et al. (2011) Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. *Green Chem* 13(8):2083–2090.
49. Jue CK, Lipke PN (1985) Determination of reducing sugars in the nanomole range with tetrazolium blue. *J Biochem Biophys Methods* 11(2-3):109–115.
50. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
51. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, Vol 26.
52. Bailey S; Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
53. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61(Pt 4):458–464.
54. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
55. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3): 240–255.
56. Afonine PV, et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68(Pt 4):352–367.
57. Rocchia W, Alexov E, Honig B (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem B* 105(28):6507–6514.
58. Bu L, et al. (2011) Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. *J Biol Chem* 286(20):18161–18169.
59. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2): 926–935.
60. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616.
61. Guvench O, et al. (2011) CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J Chem Theory Comput* 7(10):3162–3180.
62. Guvench O, et al. (2008) Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem* 29(15):2543–2564.
63. Essmann U, et al. (1995) A smooth particle mesh ewald method. *J Chem Phys* 103(19): 8577–8593.

# Supporting Information

## Kern et al. 10.1073/pnas.1301502110

### SI Methods

**Velocity Analytical Ultracentrifugation.** Sedimentation velocity experiments were performed in a Beckman XL-A analytical ultracentrifuge equipped with an An50-Ti rotor. Double-sector Epon cells with path lengths of 1.2 cm were used with quartz window assemblies. The final protein concentrations were in the range 12.5 μM in a buffer containing 20 mM Tris·HCl pH 8 and 50 mM NaCl. Samples were equilibrated at 21 °C and then accelerated to 40,000 rpm. Radial scans were performed at 5-min intervals at 280 nm. The partial specific volume for the cellobiohydrolase 7B from *Limnoria quadripunctata* (LqCel7B) was calculated from the amino acid composition using SEDNTERP (1) at 0.7062 mL/g, with a buffer density of 1.00087 g/mL and a viscosity of 0.010126 P. Analysis of the scans was performed using the program SEDFIT (2).

**Size Exclusion Chromatography Multiangle Laser Light Scattering.** Size exclusion chromatography multiangle laser light scattering (SEC-MALLS) was performed using a 5-μm analytic size exclusion column (Wyatt Technology; WTC-MP030S5) in a buffer containing 20 mM Tris·HCl pH 8.0 and 50 mM NaCl. The scattered light intensity and protein concentration of the column eluate were recorded using a DAWN-HELEOS laser photometer and an OPTILAB-rEX differential refractometer (dRI), respectively. The weight-averaged molecular mass of the sample peaks in the eluate was determined from the combined data from both detectors using ASTRA software version 5.1 (Wyatt Technology), using a value of refractive index increment (dn/dc) = 0.186 mL/g.

**Liquid Chromatography Coupled to Tandem Mass Spectrometry Analysis.** Q-Trap purified LqCel7B protein solutions were diluted to 50 μL with aqueous 50 mM ammonium bicarbonate before reducing with 5 μL aqueous 50 mM Tris(2-carboxyethyl)phosphine (TCEP) and incubated at 60 °C for 1 h. Postreduction, proteins were alkylated with 2.5 μL 200 mM methylmethanethiosulfonate (MMTS), with incubation for 10 min before tryptic digestion. Sequencing-grade, modified porcine trypsin (Promega) was dissolved in the 50 mM acetic acid supplied by the manufacturer and then diluted fivefold by adding 50 mM ammonium bicarbonate to give a final trypsin concentration of 0.02 μg/μL. A 10-μL aliquot of trypsin solution was added to each sample, before overnight incubation at 37 °C. Samples were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry $C_{18}$, 5-μm trap (180 μm × 20 mm) and a nanoAcquity BEH130 1.7-μm $C_{18}$ capillary column (75 μm × 250 mm). The trap wash solvent was 0.1% (vol/vol) aqueous formic acid and the trapping flow rate was 10 μL/min. The trap was washed for 5 min before switching flow to the capillary column. The separation used a gradient elution of two solvents [solvent A: 0.1% (vol/vol) formic acid; solvent B: acetonitrile containing 0.1% (vol/vol) formic acid]. The flow rate for the capillary column was 300 nL/min, column temperature was 60 °C, and the gradient profile was as follows: initial conditions 5% solvent B (2 min), followed by a linear gradient to 35% solvent B over 20 min and then a wash with 95% solvent B for 2.5 min. The column was returned to initial conditions and reequilibrated for 25 min before subsequent injections. The nanoLC system was interfaced with a maXis liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) system (Bruker Daltonics) with a nanoelectrospray source fitted with a steel emitter needle (180 μm o.d. × 30 μm i.d.; Proxeon). Positive electron spray ionization (ESI)-MS and MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition, and processing were performed using Compass 1.3 SP1 software (microTOF control, HyStar, and DataAnalysis software; Bruker Daltonics). The following instrument settings were used: ion spray voltage = 1,400 V; dry gas 4 L/min; dry gas temperature = 160 °C and ion acquisition range *m/z* 50-2,200. AutoMSMS settings were as follows: MS = 0.5 s (acquisition of survey spectrum); MS/MS [collision induced dissociation (CID) with $N_2$ as collision gas]; ion acquisition range, *m/z* = 350-1,400; 0.1-s acquisition for precursor intensities above 100,000 counts; for signals of lower intensities down to 1,000 counts acquisition time increased linear to 1.5 s; the collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table; 3 precursor ions, absolute threshold 1,000 counts, preferred charge states, 2–4; singly charged ions excluded. Two MS/MS spectra were acquired for each precursor and former target ions were excluded for 60 s.

**RNA Extraction, Reverse Transcription, and RT-Quantitative PCR Analysis.** The experiments described in this paper involving animals were approved by the Ethical Review Committee of the University of Portsmouth. Material from 50 *Limnoria quadripunctata* specimens feeding on Balau wood was dissected into the hepatopancreas, the gut, and the rest of the body and flash frozen in liquid nitrogen before extracting total RNA with TRIzol reagent (Invitrogen) according to the manufacturer's protocol. The RNA concentration was determined spectrophotometrically (NanoDrop 2000; Thermo Scientific) and its integrity evaluated using the Bioanalyzer (Agilent). Total RNA (400 ng) was treated with RQ1 DNase (Promega) to remove genomic DNA and subjected to reverse transcription using oligo dT primers and SuperScriptII RT (Invitrogen) according to the manufacturer's instructions. A control reaction with no enzyme (no reverse transcription, RT) was carried out in parallel to confirm removal of genomic DNA contamination. Quantitative PCR of diluted cDNA and no RT controls (equivalent to 4 ng total RNA per reaction) was performed with Power Sybr Green (Applied Biosystems) in a 25-μL reaction according to the manufacturer's protocol on an ABI Prism 7000 sequence detector (Applied Biosystems). Specific primers for *L. quadripunctata* genes coding for a glycosyl hydrolase (GH) family 7 LqCel7B (GenBank: FJ940757.1), ubiquitin (Ubi), and GAPDH were used: GH7b1_F 5′-TTGCTGGCAAAGCTAATTCTGAT-3′, GH7b1_R 5′-GCAGCAGGCTCCCATTTGT-3′, Ubi1_F 5′-GGTTGATCTTTGCCGGAAAG-3′, Ubi1_R 5′-TCTCAAAACGAGGTGAAGTGTTG-3′, GAPDH1_F 5′-TGTAATTTTCCTTCCATCGACAAC-3′, GAPDH1_R 5′-CTCCACACACGGTCGCTACA-3′, GAPDH2_F 5′-CTCTACCTCCGCGCCAATC-3′, GAPDH2_R 5′-CGCTGTAACGGCTACTCAGAAGA-3′. Specificity of PCR products has been verified by sequencing and analysis of dissociation curves and no template control PCRs (cDNA omitted) were done for each primer pair to rule out any contamination in the PCR mix. The amplification ratio for the 5′ and 3′ regions for GAPDH (primers 1/2) was used to assess the consistency of cDNA synthesis in different samples. PCR efficiencies [$E = 10^{(-1/\text{slope})}$] were calculated from a standard curve with four technical replicates using a twofold dilution series of hepatopancreas cDNA. The values for the cycles of quantification (Cq) were adjusted with corresponding PCR efficiencies using the efficiency calibrated mathematical method for the relative expression ratio in real-time PCR [$(E^{Cq})_{ref}/(E^{Cq})_{tar}$] (3). Ubiquitin was found to be one of the most stably expressed genes in the samples out of eight potential reference genes, which were evaluated using the programs geNorm, BestKeeper, and NormFinder. Gene expression levels are shown as relative ubiquitin-normalized efficiency-corrected values of the mean from three technical replicates.

**Expression of Native LqCel7B in *Aspergillus niger*.** *LqCel7B* (lacking the signal peptide sequence), was codon optimized by GeneArt (Invitrogen) for expression in *A. niger*. A gel-purified and HpaI/XbaI-digested DNA fragment containing the *LqCel7B* gene was cloned into the HpaI/XbaI digested pIGF expression vector containing the *pyrG* selection marker gene encoding the orotidine 5′-phosphate carboxylase involved in the biosynthesis of uracil (kindly provided by Prof David Archer, The University of Nottingham, Nottingham, UK). *A. niger* strain D15 (kindly provided by Peter Punt, TNO, Zeist, the Netherlands) protoplasts were transformed with pIGF-*pyrG::LqCel7B* expression construct following the polyethylene glycol/$CaCl_2$ method as previously described (4). For protein production, 200 mL of ACM medium (0.52 g/L KCl, 1.52 g/L $KH_2PO_4$, 2 g/L peptone, 1.5 g/L bactocasamino acids, 5 g/L $NH_4Cl$, 0.52 g/L $MgSO_4$, 43.8 g/L Tri-Na citrate), supplemented with trace elements (76.5 μM $ZnSO_4$ $7H_2O$, 178 μM $H_3BO_3$, 25.3 μM $MnCl_2$ 4 $H_2O$, 18 μM $FeSO_4$ 7 $H_2O$, 7.15 μM $CoCl_2$ 6 $H_2O$, 6.41 μM $CuSO_4$ 5 $H_2O$, 6.2 μM $NaMoO_4$ 2 $H_2O$, 134 μM $Na_2$ EDTA), and vitamins (26.5 μM riboflavin, 29.6 μM thiamine·HCl, 81.9 μM nicotinamide, 24.3 μM pyridoxin·HCl, 4.20 μM pantothenic acid, 81.9 nM biotin), and containing 2% (wt/vol) soluble polyvinylpyrrolidone (PVP) and 150 g/L maltose were inoculated with $1 \times 10^6$/mL spores and grown in shake flasks at 20 °C for 3 d at 150 rpm. Culture supernatant was separated from the mycelium through filtration through Miracloth (Calbiochem/EmdMillipore), centrifuged at 12,000 × *g* for 20 min, and clarified through filtration. LqCel7B was purified by anion exchange chromatography on a 5-mL HiTrap Q as described in *Methods*, *Expression and Purification*.

**Sequences.** LqCel7B lacking the native signal peptide sequence was expressed and crystallized from *A. niger* and *Aspergillus oryzae* recombinant culture supernatants: XQAGTETEEYHLPLTWERDGSSVSASVVIDSNWRWTHSTEDTTNCYDGNEWDSTLCPDADTCTENCAIDGVDQGTWGDTYGITASGSKLTLSFVTEGEYSTDIGSRVFLMADDDNYEIFNLLDKEFSFDVDASNLPCGLNGALYFVSMDEDGGTSKYSTNTAGAKYGTGYCDAQCPHDMKFIAGKANSDGWTPSDNDQNAGTGEMGACCHEMDIWEANSQAQSYTAHVCSVDGYTPCTGTDCGDNGDDRYKGVCDKDGCDYAAYRLGQHDFYGEGGTVDSGSTLTVITQFITGGGGLNEIRRIYVQDGGQTIQNAAVNFPGDVDPYDSITEDFCVDIKRYFGDTNDFDAKGGMSGMSNALKKGMVLVMSLWDDHYANMLWLDATYPVDSTEPGALRGPCSTDSGDPADVEANFPGSTVTFSNIKIGPIQSYD. Following cleavage of the signal peptide, the N-terminal glutamine is modified by cyclization to generate pyrrolidone carboxylic acid (PCA, denoted as X). This is the sequence deposited in the Protein Data Bank (PDB ID code 4GWA). The gene encoding LqCel7B was designated *LqGH7B* (FJ940757), as previously published (5). A *Chelura terebrans* GH7 full-length sequence has been deposited in GenBank (www.ncbi.nlm.nih.gov/GenBank) under accession no. KC776193.

**Antibody Production and Purification.** The sequence encoding *LqCel7B* was inserted into the NheI/EcoRI sites of the vector pET28A as a C-terminal 6× His tagged-fusion. The resulting construct was transformed into BL21 *Escherichia coli* strain and fresh colonies were picked from LB plates containing chloramphenicol 42 mg/L and Kan 50 mg/L. A 5-mL culture of three colonies was set up in the same medium and grown overnight at 37 °C under agitation. Two milliliters of this culture was used to inoculate 200 mL of 2YT medium and grown to early exponential phase ($OD_{600}$ = 0.5–0.6) and then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown overnight at 30 °C. Cells were harvested by spinning cultures down at 3,000 × *g* for 5 min and resuspended in 100 mM $NaH_2PO4$, 10 mM Tris·Cl, 8 M urea following manufacturer instructions (Qiagen). Preparation of clarified cell extracts and purification of LqCel7B under denaturing conditions were performed using a Nickel-NTA column (Qiagen). Preparation purity was assessed by SDS/PAGE and sample concentration was determined by Bradford (Pierce). About 1 mg of recombinant protein was used to raise polyclonal antibodies in sheep (Scottish National Blood Transfusion Service, Penicuik, UK). To enrich the serum for LqCel7B antibodies, an affinity column was made using LqCel7B purified from *A. oryzae* culture supernatants as described in *Methods Expression and Purification*. A total of 2.7 mg LqCel7B was coupled to 2 mL of rehydrated Fast Flow CNBr-activated Sepharose (GE Healthcare) and binding of the anti-LqGH7 serum, column wash, elution of affinity purified antibody fractions, and column regeneration were performed according to the resin manufacturer instructions. Purified antibody fractions were then characterized for their affinity to LqCel7B by dot blot and Western blot using both recombinant LqCel7B and *Limnoria quadripunctata* whole body extracts. Fractions showing the highest titer and no unspecific binding were selected for Western Blot experiments.

**Protein Extraction and Western Blot Analysis.** Material from 20 *L. quadripunctata* specimens feeding on willow wood shavings was dissected into the hepatopancreas, the gut, and the rest of the body and flash frozen in liquid nitrogen before extracting proteins with TRIzol reagent (Invitrogen) according to the manufacturer's protocol. The protein concentration was determined using Bradford reagent (Pierce). After separation of 1.3 μg protein on a 10% gel in SDS/PAGE under denaturing conditions, proteins were either visualized by Coomassie stain with InstantBlue (Expedeon) to verify equal loading or transferred onto Protran85 nitrocellulose membranes (Whatman) by semidry blotting in Towbin buffer for Western blot analysis. Primary polyclonal antibodies for detection of LqGH7 protein were raised in sheep and detected by secondary antisheep horseradish peroxidase-conjugated antibodies (Sigma). SuperSignal West Pico Chemiluminescent Substrate (Pierce) was added before exposure on ECL Hyperfilm (GE Healthcare) to visualize protein accumulation.

**Scanning Electron Microscopy Specimen Preparation and Imaging.** Specimens of *L. quadripunctata* Holthuis were fixed in 4% (vol/vol) glutaraldehyde in 0.2 M sodium cacodylate with 0.3 M sodium chloride and 2 mM calcium chloride adjusted to pH 7.8 for 1 h at room temperature, rinsed in cacodylate buffer, postfixed for 1 h in 1% (wt/vol) aqueous osmium tetroxide in cacodylate buffer, and then rinsed in buffer. Specimens were dehydrated through a graded series of ethanol solutions (40%, 50%, 75%, 90%, 95%, and then 100%), transferred to xylene and then paraffin wax/xylene 1:1 before embedding in wax. The embedded specimens were sectioned until the hindgut was revealed; then the wax was removed using several changes of xylene at 60 °C. From xylene the specimens were transferred to ethanol and then hexamethyldisilazane, which was allowed to evaporate overnight. The dry specimens were mounted on carbon tabs, sputter coated with gold and palladium, and imaged at 15 kV acceleration voltage with a JEOL 6060LV SEM operating in secondary electron mode.

**Substrate Preparation.** To comprehensively characterize the activity of LqCel7B both oligosaccharide and polysaccharide substrates were tested. Substrates included *p*-nitrophenyl-β-D-glucopyranoside (*p*NP-G1), *p*-nitrophenyl-β-D-cellobioside (*p*NP-G2), *p*-nitrophenyl-β-D-cellotrioside (*p*NP-G3), *p*-nitrophenyl-β-D-cellotetraoside (*p*NP-G4), *p*-nitrophenyl-β-D-cellopentaoside (*p*NP-G5), *p*-nitrophenyl-β-D-xylopyranoside (*p*NP-X), *p*-nitrophenyl-β-D-galactopyranoside (*p*NP-Gal), and *p*-nitrophenyl-β-D-mannopyranoside (*p*NP-M), all purchased from Sigma-Aldrich. Substrates cellobiose, cellotriose, cellotetraose, cellopentaose, cellohexaose, and celloheptaose were supplied by Carbosynth. Polymeric substrates assayed included phosphoric acid swollen cellulose (PASC) and Avicel. For preparation of PASC, 5 g Avicel was moistened with water and 150 mL ice-cold 85% phosphoric acid was added

and the suspension stirred on an icebath for 1 h followed by the addition of 500 mL cold acetone under continuous stirring. The swollen Avicel (henceforth referred to as PASC) is filtered on a glass-filter funnel and washed three times with 100 mL ice-cold acetone and twice with 500 mL water. PASC is then suspended in 500 mL water and blended to homogeneity using an Ultra Thorax homogenizer. Following homogenization, substrates were washed in 80% (vol/vol) ethanol and spun down for 5 min at 4,000 × *g* in a benchtop centrifuge. The supernatant was discarded and the pelleted polysaccharides were then washed once more as described, residual ethanol was removed, and the pellets were allowed to dry at room temperature and were resuspended in water.

**Enzyme Assays.** The standard concentration of *p*NP esters and cellooligosaccharides was 0.5 mM unless otherwise stated. All reactions containing cellooligosaccharides or *p*NP esters were carried out at 30 °C for 4 h in 100 mM citrate buffer pH 4.5, 500 mM NaCl, and 1 μg of purified LqCel7B. Enzyme activity was measured according to release of *p*NP ions at 405 nm for *p*NP esters and products of reactions containing cellooligosaccharides were extracted and analyzed following the method described for high performance anion exchange chromatography (HPAEC) analysis.

**Kinetic Studies.** For substrates *p*NP-G3 and *p*NP-G5 (Carbosynth), reactions were set up as described in *Methods* and the relevant substrate was added to the following final concentrations 0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, and 3 mM. Reactions were incubated at 30 °C and after all reaction components were mixed together, reactions were sampled at different time points (0, 15 min, 30 min, 1 h, 2 h, and 4 h) and immediately stopped at each given time point by the addition of $Na_2CO_3$ to a final concentration of 500 mM. *p*NP release was quantified as above. All reactions were carried out in triplicates and the results presented were generated from two independent experiments. Enzyme kinetics data were analyzed through construction of Michaelis–Menten plots by graphing reaction rates and substrate concentration values, which were then fitted to nonlinear regression using SigmaPlot V.11 (Systat Software) to yield values for the kinetic constants $V_{max}$, $K_m$, $k_{cat}$, and $k_{cat}/K_m$.

**Cellobiose Inhibition.** *p*NP-G5 was chosen as a substrate and reactions containing this substrate at concentrations of 0.1, 0.25, 0.5, 0.75, 1, 1.5, and 2 mM were performed with/without the addition of cellobiose (Sigma-Aldrich) to the final concentrations of 5, 10, 25, 50, and 100 mM. Reactions were incubated at 30 °C and time points were harvested from each reaction at 5 min, 30 min, 1 h, 1.5 h, and 2 h following reaction start. Michaelis–Menten plots were constructed comparing the reaction rates obtained at different substrate concentrations in absence and presence of increasing concentrations of inhibitor. To estimate $K_i$ values, $K_m$ and $V_{max}$ were calculated for each inhibitor concentration using nonlinear regression as described above and the respective Lineweaver–Burke double reciprocal plots.

**Determination of pH and Temperature Optima.** *p*NP-G3 and *p*NP-G5 were used to determine the pH optimum of LqCel7B. *p*NP-ester reactions were set up containing 500 μM of *p*NP-G3 or *p*NP-G5 and 1 μg purified LqCel7B. All reactions contained 500 mM NaCl and 100 mM of the appropriate buffer. Various buffers were used to cover the broad pH range shown: citrate buffer pH 3.5–4.5, acetate buffer pH 5–6.5, phosphate buffer pH 7–8, Tris·HCl buffer pH 8.5–9, and glycine buffer pH 9.5–10. All reactions were incubated for 4 h at 30 °C and carried out in triplicates. For temperature optimum determination, reactions were set up as above in 100 mM sodium acetate buffer pH 4.5 and incubated at temperatures from 10 °C to 50 °C using DNA Engine Tetrad 2 Thermal Cycler (Bio-Rad). All reactions were incubated for 4 h at respective temperatures and carried out in triplicates.

**High Performance Anion Exchange Chromatography.** Products of reactions set up as above were extracted in 80% (vol/vol) ethanol, spun down for 10 min at 10,000 × *g* at 4 °C and the alcoholic supernatant was recovered, evaporated in vacuum using a speed vac (Savant SPD131DDA, refrigerated vapor trap RVT4104; Thermo Scientific), and the remaining pellet was resuspended in water and filtered through a 0.45-μm syringe filter (Milex; Millipore). Reaction products were then separated using a Dionex ICS3000 instrument connected to a CarboPac PA20 3 × 150 mm column (Dionex). Reaction products were loaded onto the column preequilibrated with 100 mM NaOH for 1 min at a flow rate of 0.5 mL/min, and then resolved through a gradient from 100 mM NaOH/0 mM NaAc to 80 mM NaOH/100 mM NaAc over 30 min at a flow rate of 0.5 mL/min followed by a final column wash in 100 mM NaOH for 5 min. Mono/polysaccharides were identified by comparison of their retention times with standards.

**Obtaining Atomic Resolution X-Ray Crystal Structure of LqCel7B.** To obtain higher resolution structural data of LqCel7B, an approach of microseeding and subsequent macroseeding was used. Crystals were grown in the presence of thiocellobiose in conditions identical to the cocrystal structure (PDB ID code 4HAP) (Fig. S3*A*). These crystals presented an identical morphology to previous cocrystal complexes and diffracted in a similar manner in space group *P1* with identical cell dimensions. These crystals were then crushed with an acupuncture needle and the whole drop was mixed with fresh protein, ligand, and buffer. This produced crystals with an alternative morphology but they were too small for diffraction analysis (Fig. S3*B*). The original *P1* crystals were again crushed with a needle, which was then used to streak seed multiple drops, each with a much lower protein concentration but retaining a high thiocellobiose concentration. These crystals were isolated, and although small, diffracted to beyond 2 Å in an orthorhombic space group (Fig. S3*C*). These crystals were then individually placed into fresh crystallization drops containing a high protein concentration and left for several days before cryocooling in liquid nitrogen (Fig. S3 *D*–*G*). The resulting crystals diffracted to beyond 1 Å at beamline I04 at the Diamond Light Source (Didcot, UK). (Fig. S3*H*). The resulting electron density showed all 420 amino acids, several metal ions, the thiocellobiose ligand, and hundreds of ordered water molecules (Fig. S3*I*).

**Protein Setup for Molecular Dymanics Simulations.** The *Hypocrea jecorina* Cel7A (HjCel7A) enzyme was taken from the structure (PDB ID code 8CEL) with a cellononaose chain spanning the active site (6). The *L. quadripunctata* Cel7B (LqCel7B) structure was taken from the structure solved in this study. To dock a ligand into the LqCel7B structure, HjCel7A and LqCel7B structures were aligned in Pymol (7), and the resulting coordinates were saved with a ligand in the enzyme tunnels and includes the crystallographic waters. The protonation states of LqCel7B were determined with H++ at a salinity of 0.6 mM and pH 5 (8–11). The protonation states of the catalytic residues in each enzyme were assigned based on the proposed catalytic mechanism for *Trichoderma reesei* (*H. jecorina*) Cel7A (6). The apo structure of LqCel7B was also constructed with matching protonation states and crystallographic water placement. After determining the protonation states, the structures were solvated with water in 84-Å cubic boxes. Sodium cations were added to ensure charge neutrality. In each, the systems were ~59,000 atoms.

**Simulation Protocol.** All simulations were set up, minimized, heated, and density was equilibrated using CHARMM (12). Following setup, each system was minimized according to the following protocol: (*i*) 1,000 steps of steepest descent minimization with the protein and ligand (where applicable) fixed, (*ii*) 1,000 steps of steepest descent minimization with the ligand (where applicable) fixed, and (*iii*) 1,000 steps of steepest descent

minimization with the entire system unrestrained. The systems were then heated from 100 K to 300 K in increments of 50 K every 4 ps. From there, the density was equilibrated with the Nosé–Hoover thermostat and barostat at 300 K and 1 atm (13, 14). Heating, equilibration, and production simulation uniformly applied the following parameters: The CHARMM36 force field including the CMAP correction was used to describe the proteins (15, 16). Modified TIP3P was used to describe the water molecules (17, 18), and the CHARMM36 carbohydrate force field was used to describe the cellooligomers (19–21). The particle mesh Ewald method (22) was used to treat long-range electrostatics with a sixth-order b-spline interpolation, a Gaussian distribution width of 0.34 Å, and a grid size of 80 × 80 × 80. A nonbonded cutoff distance of 10 Å was applied. The SHAKE algorithm was used to fix the bond distances to hydrogen atoms (23). Production simulations were performed using NAMD for a total of 250 ns with a 2-fs time step in the NVT ensemble at 300 K (24).

## SI Results

**LC-MS/MS Analysis.** Purity of Q-Trap purified fractions was assessed by LC-MS/MS to ensure that no other contaminant proteins/enzymes were present in the samples used for enzyme characterization. Nine peptides matching the LqCel7B sequence were detected in the sample, all of which showed $E$ values of at least $10^{-5}$, indicating a very low probability of finding a random peptide match unspecific to the LqCel7B sequence. Peptides and their respective $E$ values (shown in parentheses) are listed as follows: R.LGQHDFYGEGGTVDSGSTLTVITQFITGGGGLNEIR.R ($1.90 \times 10^{-15}$), K.LTLSFVTEGEYSTDIGSR.V ($2.30 \times 10^{-12}$), R.VFLMADDDNYEIFNLLDK.E ($2.10 \times 10^{-09}$), R.DGSSVSASVVIDSNWR.W ($1.90 \times 10^{-08}$), K.GMVLVMSLWDDHYANMLWLDATYPVDSTEPGALR.G ($2.40 \times 10^{-06}$), K.GGMSGMSNALK.K ($2.70 \times 10^{-06}$), R.YFGDTNDFDAK.G ($5.50 \times 10^{-06}$), K.GGMSGMSNALKK.G ($1.30 \times 10^{-05}$), and K.RYFGDTNDFDAK.G ($7.40 \times 10^{-05}$). For all peptides, theoretical molecular weight and experimental molecular weight were in very close agreement.

**Molecular Dynamics Simulation Analysis.** Beyond the results presented in the manuscript, molecular dynamics (MD) simulation analysis included examination of the root-mean-square deviation (RMSD) of the protein backbone, as shown in Fig. S5. Comparing LqCel7B simulations, with and without a ligand, we see little deviation in the backbone movements over the course of the 250-ns simulations. In general, the RMSD of the protein backbones indicate relatively well-equilibrated simulations.

Root-mean-square fluctuation (RMSF) of the protein backbone was also calculated for LqCel7B with and without a ligand and for the ligand-bound HjCel7A simulation. The RMSF as a function of residue number is shown in Fig. S5*A*, comparing LqCel7B with and without a ligand. Fig. S5*B* compares the RMSF of the ligand-bound *v* with HjCel7A. From Fig. S5*A*, we see the presence of the cellononaose ligand bound in the active site has little effect on protein dynamics.

Comparing the protein backbone RMSF of the ligand-bound LqCel7B and HjCel7A (Fig. S5*B*), we see LqCel7B exhibits increased fluctuation primarily in three loop regions relative to HjCel7A. The first loop region (residues 215–222 in LqCel7B and 194–201 in HjCel7A) is adjacent to the putative exo loop. In LqCel7B, this loop predominately hydrogen bonds with the exo loop, whereas in HjCel7A the exo loop hydrogen bonds preferentially with a different loop across the active site tunnel (residues 369–373).

The second loop of interest (residues 266–273 in LqCel7B and 246–253 in HjCel7A) corresponds to the exo loop. As mentioned above for HjCel7B, this loop hydrogen bonds with two adjacent loops. In HjCel7A, the exo loop almost exclusively hydrogen bonds to a loop across the active site tunnel through the Tyr-396/Asn-267 interaction.

The third loop exhibiting increased fluctuations in LqCel7B (residues 336–346 in LqCel7B and 314–323 in HjCel7A) corresponds to the product loop labeled in Fig. 2*B*. The product loop in LqCel7B exhibits lower hydrogen bonding to the remainder of the protein relative to HjCel7A. This is likely due to the replacement of a Tyr residue in HjCel7A with a Phe in LqCel7B as well as the LqCel7B loop having fewer polar residues on this loop by comparison.

The two regions in HjCel7A exhibiting increased fluctuations over LqCel7B (residues 22–26 and residues 386–388) correspond to the apex of the HjCel7A protein near the N terminus and a product side loop addition. Interestingly, these loops represent deletions in the LqCel7B protein sequence.

Related to the RMSF analysis, the minimum distance between the putative exo loop (residues 265–274 in LqCel7B and 244–253 in HjCel7A) and the loop across the active site tunnel to which the exo loop is thought to hydrogen bond (residues 394–398 in LqCel7B and residues 369 and 373 in HjCel7A) has been reported (Fig. 2). Histograms of the binned data are shown in Fig. S5*D*. We see one distinct state for the minimum distance of these two loops in HjCel7A. However, LqCel7B illustrates the findings alluded to in the RMSF analysis above. The exo-loop hydrogen bonds to two different adjacent loops, and thus exhibits two minimum distance states.

Snapshots from the LqCel7B apo MD simulation, Fig. S5*E*, illustrate a significant degree of conformational freedom of the Tyr-121 residue illustrated in Fig. 2*C*. From the snapshots we observe occupation of all four Tyr-121 conformations illustrated by structural studies and the bound MD simulations.


1. Laue TM, Shah BD, Ridgeway TM, Pelletier SL (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, eds Harding SE, Roew AJ, Horton JC (The Royal Society of Chemistry, Cambridge, UK), pp 90–125.
2. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* 78(3):1606–1619.
3. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9):e45.
4. van Hartingsveldt W, Mattern IE, van Zeijl CMJ, Pouwels PH, van den Hondel CA (1987) Development of a homologous transformation system for *Aspergillus niger* based on the pyrG gene. *Mol Gen Genet* 206(1):71–75.
5. King AJ, et al. (2010) Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proc Natl Acad Sci USA* 107(12):5345–5350.
6. Divne C, Ståhlberg J, Teeri TT, Jones TA (1998) High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol* 275(2):309–325.
7. Kurasin M, Väljamäe P (2011) Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* 286(1):169–177.
8. Kim DW, Hong YG (2000) Ionic strength effect on adsorption of cellobiohydrolases I and II on microcrystalline cellulose. *Biotechnol Lett* 22(16):1337–1342.
9. Gordon JC, et al. (2005) H++: A server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res* 33(Web Server issue):W368–W371.
10. Myers J, Grothaus G, Narayanan S, Onufriev A (2006) A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* 63(4):928–938.
11. Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 40(Web Server issue):W537–W541.
12. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30(10):1545–1614.
13. Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Mol Phys* 50(5):1055–1076.
14. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A* 31(3):1695–1697.
15. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102(18):3586–3616.
16. Mackerell AD, Jr., Feig M, Brooks CL, 3rd (2004) Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415.
17. Durell SR, Brooks BR, Ben-Naim A (1994) Solvent-induced forces between 2 hydrophilic groups. *J Phys Chem* 98:2198–2202.
18. Jorgensen WL, Chandrasekhar J, Madura JD (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
19. Guvench O, et al. (2008) Additive empirical force field for hexopyranose monosaccharides. *J Comput Chem* 29(15):2543–2564.

20. Guvench O, Hatcher ER, Venable RM, Pastor RW, Mackerell AD (2009) CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J Chem Theory Comput* 5(9):2353–2370.
21. Guvench O, et al. (2011) CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling. *J Chem Theory Comput* 7(10):3162–3180.
22. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8857.
23. Ryckaert J, Ciccotti G, Berendsen H (1977) Numerical integration of Cartesian equations of motion of a system with constraints: Molecular-dynamics of N-alkanes. *J Comput Phys* 23: 327–341.
24. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

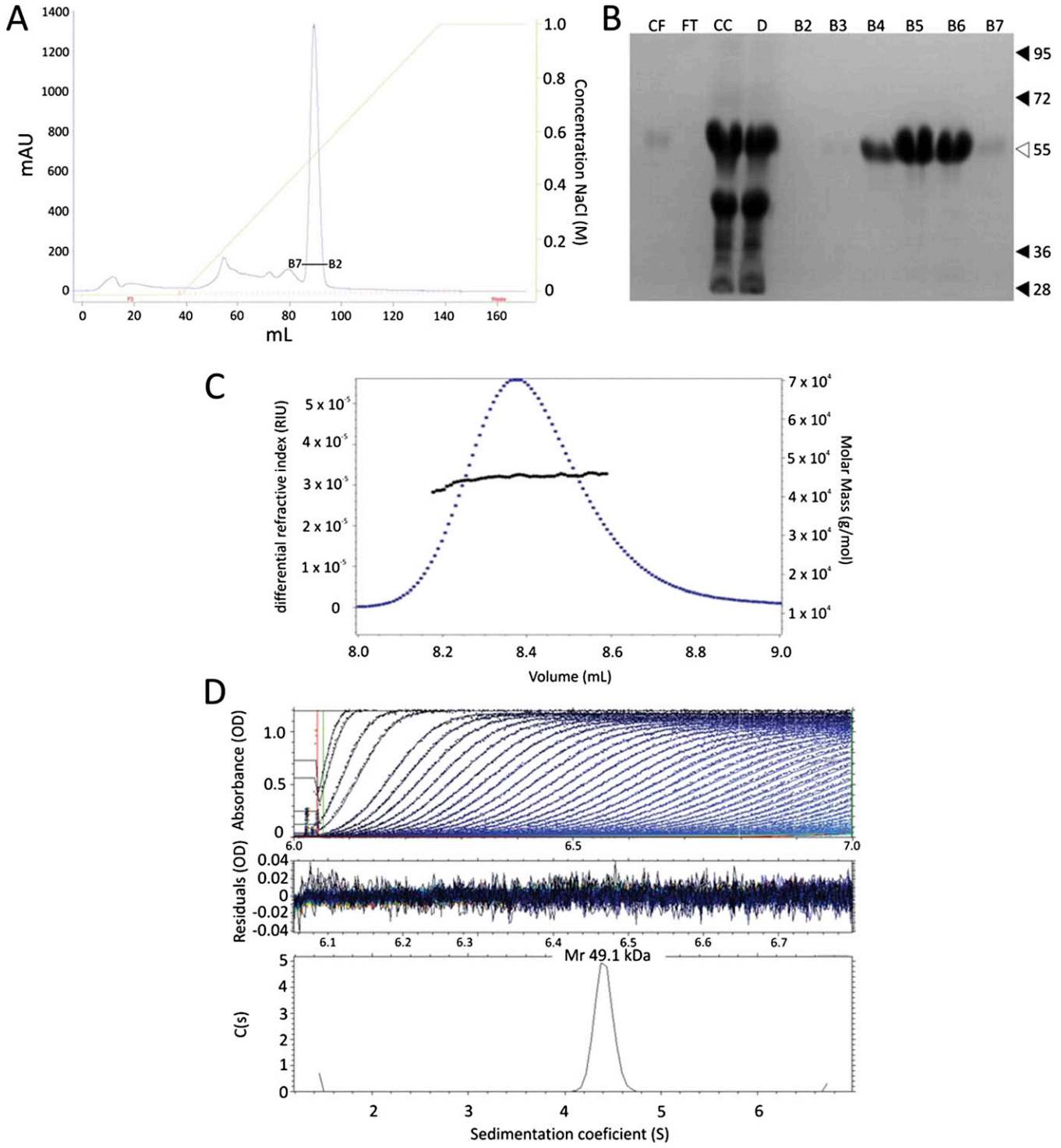
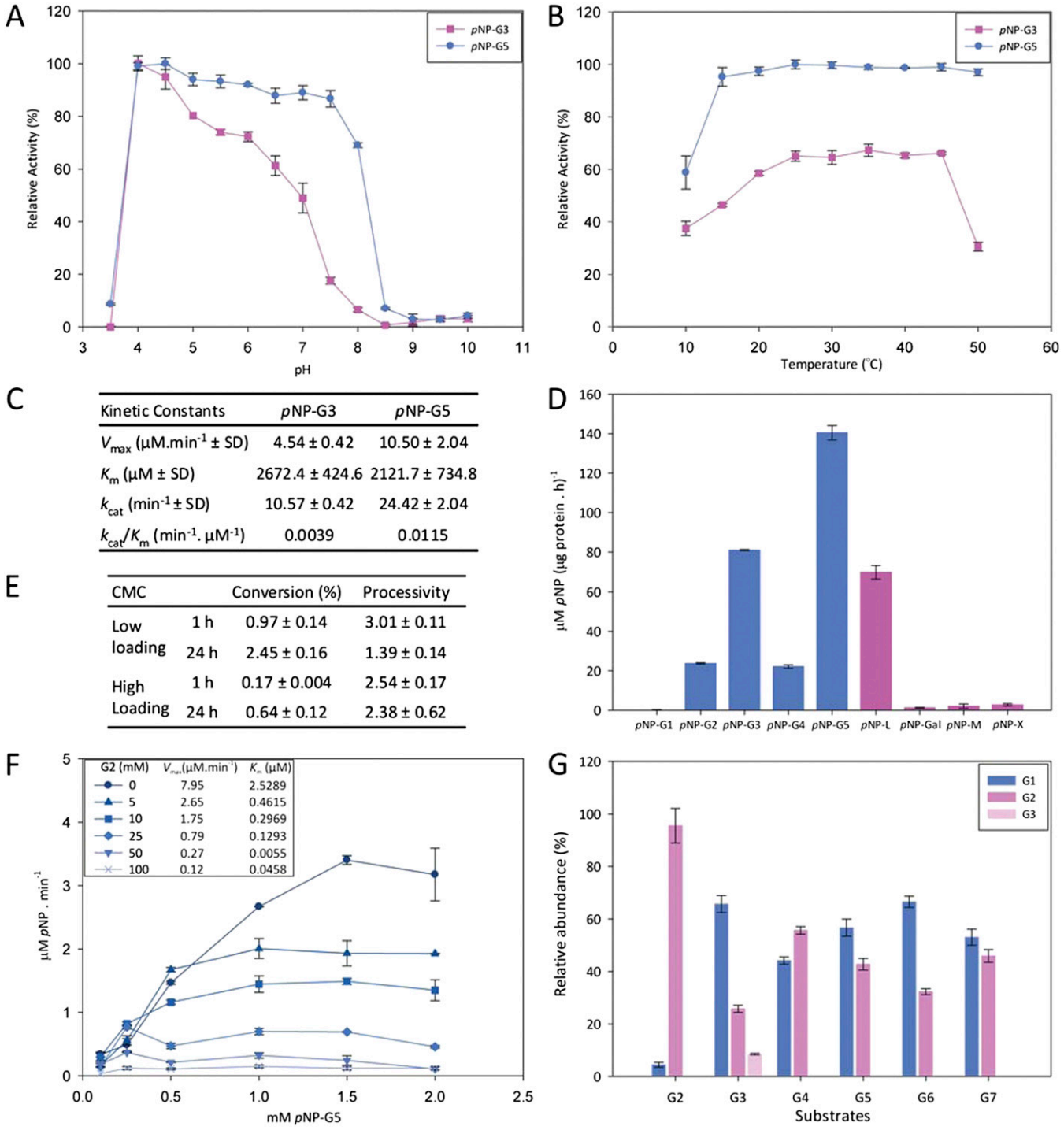



**Fig. S1.** Purification and biophysical characterization of LqCel7B. (*A*) Elution profile of LqCel7B purified on a Q-Trap column through a linear gradient of NaCl (0.05–1 M). Absorbance at 280 nm was used to monitor enzyme elution and a large peak was produced and collected in fractions B2 to B7. (*B*) Coomassie blue stained SDS/PAGE showing the electrophoretic profile of samples harvested throughout the purification protocol. CF, culture filtrate; FT, PES ultrafilter flow through; CC, PES ultrafilter-concentrated culture supernatant; D, desalted concentrated culture supernatant; B2–B7, eluted fractions. LqCel7B is eluted as a 55-kDa protein in fractions B4–B7 (white arrowhead). (*C*) Size exclusion plot is shown (blue) with the corresponding molar mass values from multiangle laser light scattering across the peak (black). (*D*) Velocity analytical ultracentrifugation is depicted as the experimental data points with lines of best fit (*Top*), residuals to those fits (*Middle*), and C(S) plot (*Bottom*).




C

| Kinetic Constants | *p*NP-G3 | *p*NP-G5 |
|---|---|---|
| $V_{max}$ (μM.min$^{-1}$ ± SD) | 4.54 ± 0.42 | 10.50 ± 2.04 |
| $K_m$ (μM ± SD) | 2672.4 ± 424.6 | 2121.7 ± 734.8 |
| $k_{cat}$ (min$^{-1}$ ± SD) | 10.57 ± 0.42 | 24.42 ± 2.04 |
| $k_{cat}/K_m$ (min$^{-1}$. μM$^{-1}$) | 0.0039 | 0.0115 |




E

| CMC | | Conversion (%) | Processivity |
|---|---|---|---|
| Low loading | 1 h | 0.97 ± 0.14 | 3.01 ± 0.11 |
| | 24 h | 2.45 ± 0.16 | 1.39 ± 0.14 |
| High Loading | 1 h | 0.17 ± 0.004 | 2.54 ± 0.17 |
| | 24 h | 0.64 ± 0.12 | 2.38 ± 0.62 |




**Fig. S2.** LqCel7B enzymatic properties. pH (*A*) and temperature (*B*) optima determination were carried out using the chromogenic substrates *p*NP-β-D-cellotrioside (■, *p*NP-G3) and *p*NP-β-D-cellopentaoside (●, *p*NP-G5). (*C*) LqCel7B kinetic constants calculated using *p*NP-G3 and *p*NP-G5. (*D*) LqCel7B activity on *p*-nitrophenol esters. Substrates: *p*NP-β-D-glucoside (*p*NP-G1); *p*NP-β-D-cellobioside (*p*NP-G2); *p*NP-β-D-cellotrioside (*p*NP-G3); *p*NP-β-D-cellotetraoside (*p*NP-G4); *p*NP-β-D-cellopentaoside (*p*NP-G5); *p*NP-β-D-lactoside (*p*NP-L); *p*NP-β-D-galactoside (*p*NP-Gal); pNP-β-D-mannoside (*p*NP-M); and *p*NP-β-D-xyloside (*p*NP-X). (*E*) Carboxymethyl cellulose (CMC) conversion ratio and processivity estimated for LqCel7B at low- and high-solids loadings [CMC 0.1% and 1% (wt/vol), respectively]. Conversion ratios shown are a product mass percentage relative to substrate mass. Processivity = G2/(G1 + G3). Reaction products were resolved by HPAEC. (*F*) Michaelis–Menten plot showing progressive reaction rate reduction (μM *p*NP·min$^{-1}$) followed by a decrease in $V_{max}$ and $K_m$ values at increasing concentrations of cellobiose (as shown in *Inset*). $V_{max}$ and $K_m$ values were calculated based on the Michaelis–Menten data fitted to nonlinear regression as described in *Cellobiose Inhibition* in *SI Methods*. (*G*) HPAEC analysis of reaction products released from hydrolysis of oligosaccharides cellobiose (G2), cellotriose (G3), cellotetraose (G4), cellopentaose (G5), cellohexaose (G6), and celloheptaose (G7). Products (shown in *Inset*): G1, glucose; G2, cellobiose; and G3, cellotriose. All values represent mean ± SE and each measurement was carried out in triplicate. Repeat experiment yielded similar results.

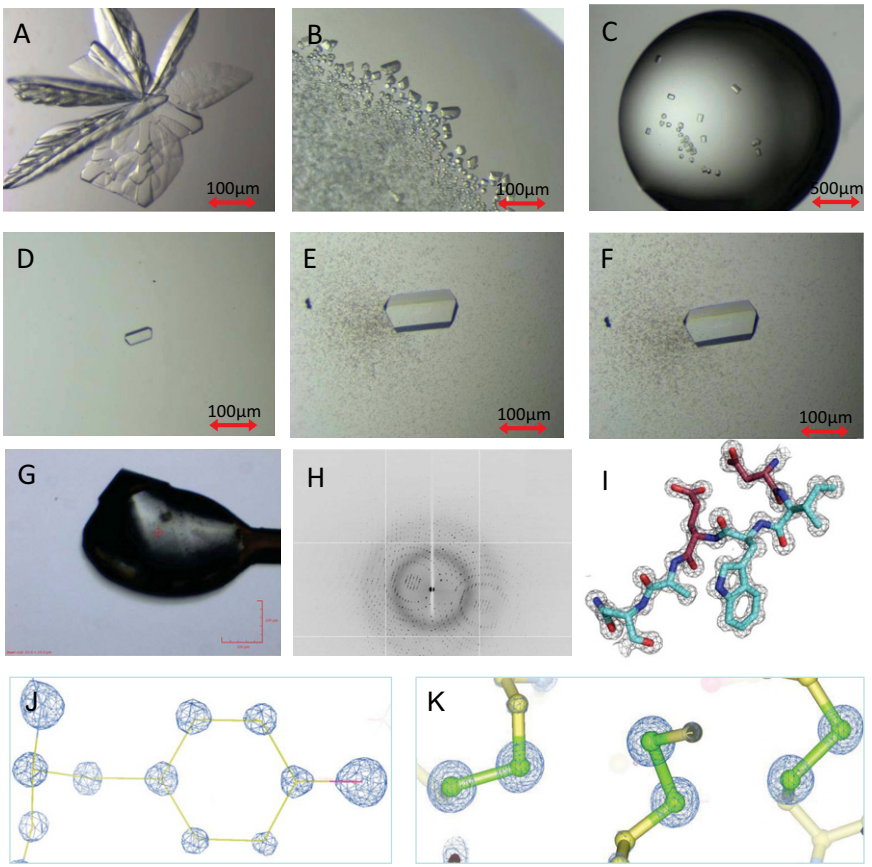



**Fig. S3.** High-resolution crystallography process for the thiocellobiose cocrystal complex. (*A*) Original LqCel7B and thiocellobiose cocrystals. (*B*) Results of microseeding experiments. (*C*) Results of streak-seeding experiments into low protein concentration. (*D*) Macroseeding, day 1. (*E*) Macroseeding day 2. (*F*) Macroseeding day 3. (*G*) Macroseeded crystal (day 5) cryocooled and mounted at beamline I04 at the Diamond Light Source. (*H*) Resulting diffraction image, top edge of detector at 1.08 Å. (*I*) resulting electron density for residue range 235–240 ($2F_o$–$F_c$ map contoured at 2σ). (*J*) High-quality electron density map around residue Tyr166 ($2F_o$–$F_c$ contoured at 6σ). (*K*) Despite the high X-ray radiation dose absorbed by the crystal, the multiple disulphide bridges in the structure remain largely intact. Here we see three sulfur atom pairs (251:276, 197:230, and 259:264) with corresponding electron density ($2F_o$–$F_c$ contoured at 8σ).

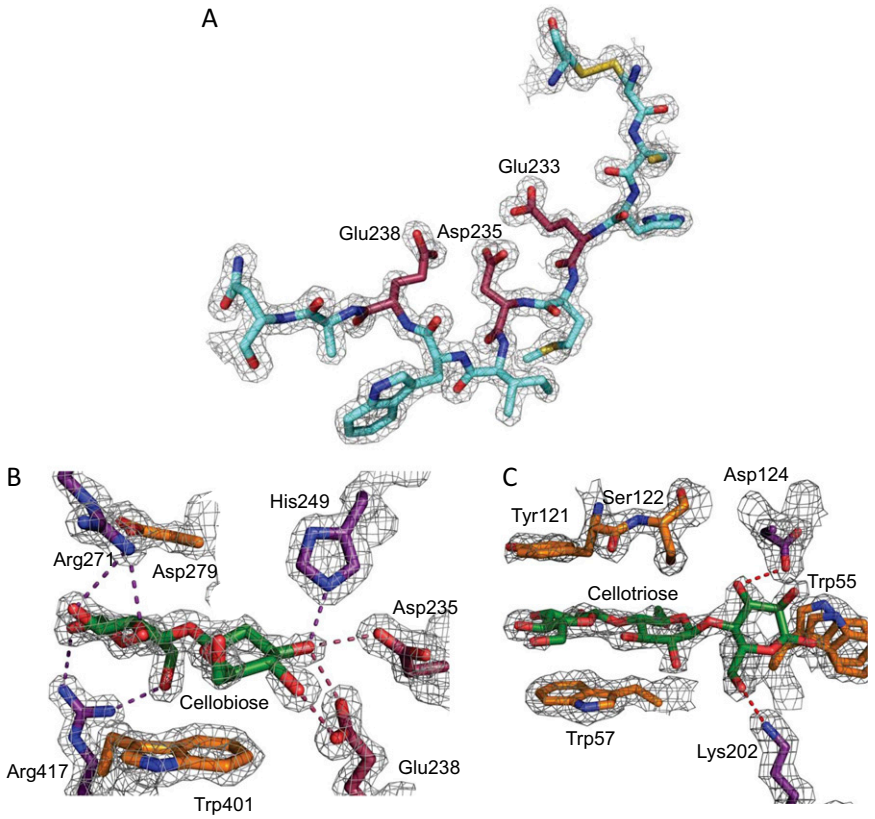



**Fig. S4.** Structural analysis of LqCel7B and its interaction with ligands. Red, acidic-cleavage residues; purple, residues that directly hydrogen bond to ligand molecules; orange, resides that π-π electron stack with ligand molecules. (*A*) $2F_o–F_c$ map of residues 230–240 (plus Cys197 to complete the disulphide bridge to the *Upper Right*) showing the high quality of the data obtained. (*B*) The binding site of cellobiose in the postcleavage portion of the ligand binding tunnel from the cocrystal structure. In this position cellobiose is binding in an inhibitory fashion. Several other, through water and backbone hydrogen bonds are formed with cellobiose in this position. (*C*) The binding site of cellotriose at the entrance to the ligand binding tunnel from the ligand soaked structure. Images rendered in PyMol with the $2F_o–F_c$ map contoured to 1.6σ.

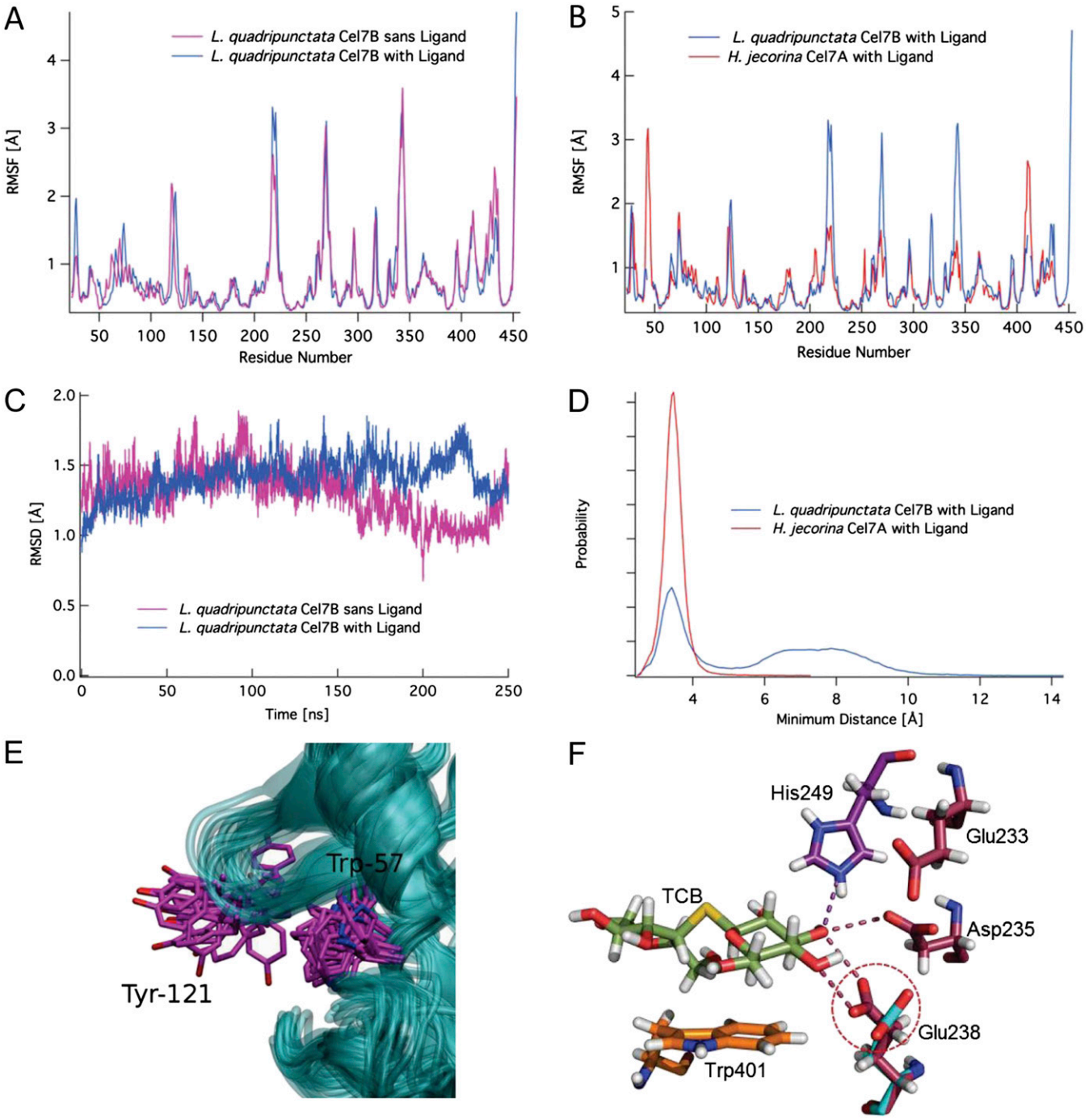



**Fig. S5.** Ligand-induced structural dynamics. (*A*) Root-mean-square fluctuation (RMSF) of the protein backbone as a function of residue number from the 250 ns MD simulations of LqCel7B with and without the bound cellanonaose ligand. (*B*) RMSF of the protein backbones of the cellanonaose-bound LqCel7B and HjCel7A from the 250-ns MD simulations. The residue numbering along the *x* axis is according to the LqCel7B numbering scheme (starting at residue 23 and ending at residue 453). HjCel7A residue numbering has been scaled accordingly. (*C*) Root-mean-square deviation (RMSD) of LqCel7B with and without a bound cellononaose ligand over the course of 250-ns MD simulations. (*D*) Binned and normalized representation of the minimum distance between the putative exo loop and the adjacent hydrogen-bonding loop. (*E*) Snapshots every 12.5 ns from the 250-ns simulation of LqCel7B apo structure. Tyr121 and Trp57 are shown in magenta and illustrate the significant conformational freedom of the Tyr-121 in absence of a bound ligand. (*F*) The LqCel7B structure in complex with thiocellobiose (TCB) to 1.14 Å revealed a dual conformation for the active site residue and proton donor, Glu238 (red circle). This refined to an occupancy of 70% in a standard ligand-bound position and 30% as an alternative rotomer. This is corroborated by MD simulation where the dihedral position from the carboxylic acid carbon to the alpha carbon of Glu238 shows a conformational fluctuation between approximately −180° and −90°.

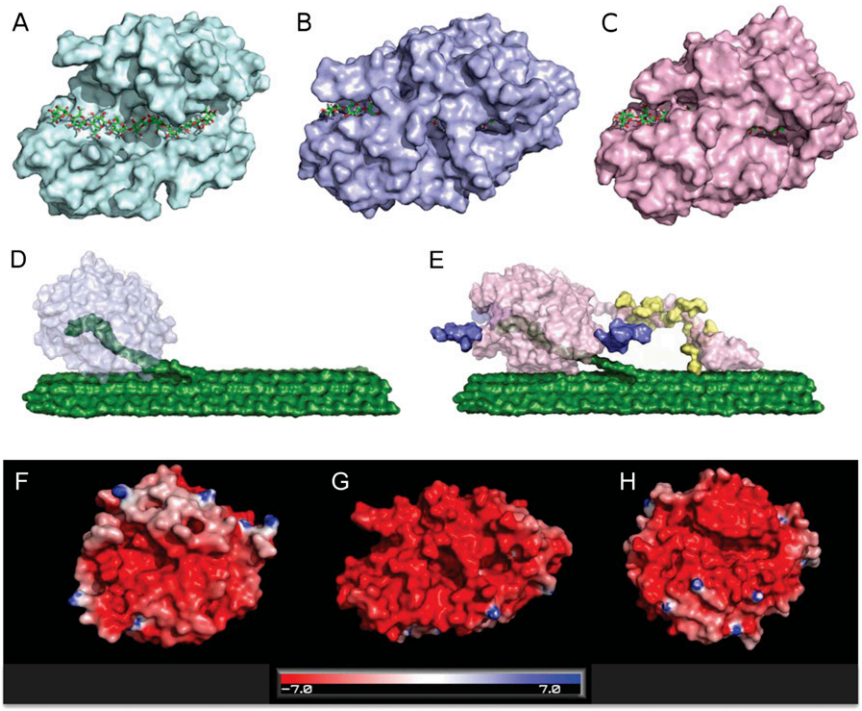



**Fig. S6.** Structure-based comparisons. Surface representations are shown for the (*A*) *H. jecorina* GH7 endoglucanase (1EG1), denoted HjCel7B; (*B*) LqCel7B; and (*C*) HjCel7A (8CEL) crystal structures with the bound ligand from 8CEL modeled in green. The LqCel7B active site tunnel displays a fold that is more similar to the enclosed structure of cellobiohydrolases such as HjCel7A. Notable differences between LqCel7B and HjCel7A include an additional opening of the central region of the tunnel, a bulky extension to the −7 subsite (*Upper Left* region), and an overall elongation of the molecule partly due to a modification of the product loop (*Far Right* area). Two model illustrations of GH7 binding to cellulose with and without a carbohydrate-binding module (CBM) are shown in *D*: LqCel7B and *E*: HjCel7A. HjCel7A is shown with *N*-glycans (blue) and O-glycans (yellow) with cellulose is shown in green. Electrostatic potential surface maps of different Carbohydrate-Active enZYmes Database (CAZy) members found in *L. quadripunctata* are shown for (*F*) LqGH5, (*G*) LqCel7B, and (*H*) LqGH9. Surface maps were built using either homology models (LqGH5 and LqGH9) or crystallographic data (LqCel7B). Electrostatic potential between −7 *kT*/e and 7 *kT*/e is shown as colored gradient from red to blue.

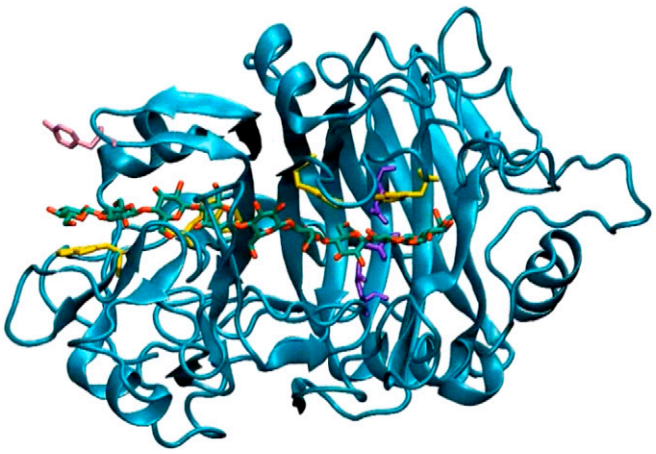
**Table S1. Crystallographic data collection and refinement statistics of free and ligand-bound LqCel7B**

| Crystal parameters | Free protein | Cellobiose | Cellobiose and cellotriose | Thiocellobiose |
|---|---|---|---|---|
| PDB accession codes | 4GWA | 4HAP | 4HAQ | 4IPM |
| Space group | *P1* | *P1* | *P1* | $P22_12_1$ |
| Cell dimensions: a, b, c; Å | 51.92, 54.05, 71.97 | 52.25, 54.19, 72.45 | 52.39, 54.27, 72.67 | 47.49, 79.78, 105.16 |
| α, β, γ; ° | 94.13, 95.70, 94.64 | 94.26, 96.02, 94.99 | 94.39, 96.33, 94.99 | 90, 90, 90 |
| Solvent content, % | 34.4 | 38.5 | 38.5 | 45.7 |
| Protein molecules in a.u. | 2 | 2 | 2 | 1 |
| Data collection | | | | |
| Beamline | DLS-I04-1 | DLS-I02 | DLS-I02 | DLS-I04 |
| Wavelength, Å | 0.9173 | 0. 9795 | 0. 9795 | 0.9795 |
| Resolution, Å | 50–1.6 | 39.9–1.6 | 44.6–1.9 | 23.2–1.14 |
| No. measured reflections | 336,757 | 316,154 | 201,188 | 516,774 |
| No. unique reflections | 100,073 | 96,748 | 59,787 | 140,362 |
| Completeness, % | 98.0 (96.3) | 93.4 (91.9) | 96.03 (88.3) | 96.4 (72.7) |
| $\langle I/\sigma(I)\rangle$ | 10.6 (2.34) | 7.6 (1.2) | 6.7 (1.3) | 8.4 (1.3) |
| Multiplicity | 3.4 (3.0) | 3.3 (3.1) | 3.4 (3.5) | 3.7 (2.3) |
| $R_{merge}$* | 7.9 (47.9) | 11.3 (59.5) | 14.6 (51.2) | 7.4 (57.9) |
| Refinement parameters | | | | |
| $R_{work}/R_{free}$ | 14.8/18.4 | 14.4/19.1 | 16.5/21.7 | 12.2/15.5 |
| No. atoms/B-factors | | | | |
| Protein | 6,559/14.6 | 6,559/11.9 | 6,559/12.6 | 6,286/10.5 |
| Water | 989/29.3 | 1,276/26.5 | 565/19.6 | 687/32.1 |
| Ligands | N/A | 46/21.3 | 81/31.8 | 53/23.17 |
| R.m.s. deviations | | | | |
| Bond lengths, Å | 0.018 | 0.021 | 0.019 | 0.022 |
| Bond angles, ° | 1.787 | 2.009 | 2.043 | 2.025 |

Values in parentheses are for the highest resolution shell, 1.60/1.60/1.90/1.14 Å, respectively.

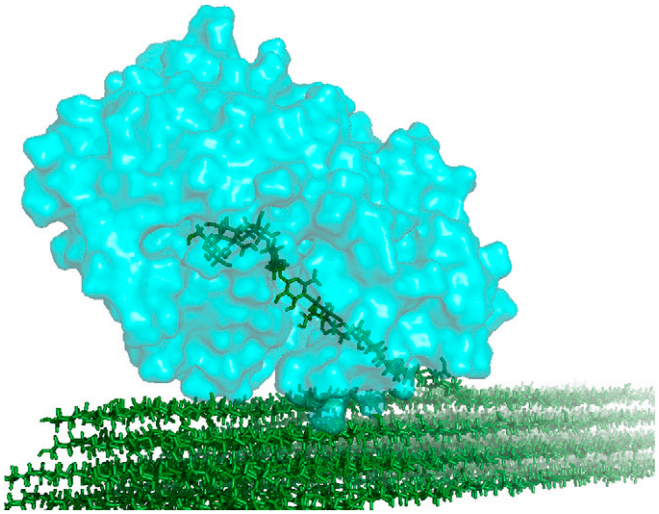
*$R_{merge} = \sum_h \sum_i |(I_{hi} - I_h)| / \sum_h \sum_i (I_{hi})$.

where $I_h$ is the mean intensity of the scaled observations $I_{hi}$.



**Movie S1.** A snapshot of the MD simulation constructed from the LqCel7B structure with the 8CEL ligand docked into the active site tunnel. In addition to the general loop flexibility, the dynamic movement of residue Tyr121 (pink) can be observed between the −7 and −6 sites.

Movie S1



**Movie S2.** The general topology of LqGH7B is rendered as both a surface and cartoon model (cyan) bound to an illustration of bulk cellulose (green). Residues of the catalytic triad are colored in blue; highly conserved tryptophan residues in the active site tunnel are colored in yellow.

Movie S2

# Abbreviations

| | |
|---|---|
| 2D-PAGE | Two dimensional – poly acrylamide gel electrophoresis |
| 2YT | 2x YT broth |
| ABC | Ammonium bicarbonate |
| AEBSF | 4-(2-Aminoethyl)benzenesulfonyl fluoride hydrochloride |
| AIM | Auto-induction media |
| ATP | Adenosine tri-phosphate |
| BAC | Bacterial artificial chromosome |
| BLAST | Basic local alignment search tool |
| BSA | Bovine serum albumin |
| C.S. | Crude serum |
| CAZy | Carbohydrate-Active enZYmes Database |
| CBM | Carbohydrate binding module |
| CESA | Cellulose synthase |
| CHP | Combined heat and power |
| Contig | Contiguous sequence |
| Ct | Cycle threshold |
| CV | Column volume |
| DI | Domain interface |
| DMSO | Dimethyl sulphoxide |
| DNA | Deoxyribonucleic acid |
| dNTPs | Deoxy nucleotide tri-phosphates |
| DPO | Di-phenoloxidase |
| DPO | Di-phenoloxidase |
| DTE | 1,4-Dithioerythritol |
| DTT | DL-Dithiothreitol |
| ECL | Enhanced chemiluminescence |
| EDTA | Ethylenediaminetetraacetic acid |
| eGFP | Enhanced green fluorescent protein |
| emPAI | Exponentially modified protein abundance index |
| ESI | Electrospray ionisation |
| EST | Expressed sequence tag |
| ETL | Early-to-late promoter |
| FABP | Fatty acid binding protein |
| FCS | Foetal calf serum |

| | |
|---|---|
| FPLC | Fast protein liquid chromatography |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GeLC-MS | Gel-based liquid chromatography-mass spectrometry |
| GH | Glycosyl hydrolase |
| GOI | Gene of interest |
| HC | Hemocyanin |
| $His_6$ | Hexa-histidine polypeptide tag |
| HMW | High molecular weight |
| HP | Hepatopancreas |
| HQ | Hydroxyquinone |
| HRP | Horse radish peroxidase |
| HRV | Human rhinovirus 3C protease |
| IAA | Iodoacetamide |
| IEF | Isoelectric focussing |
| Ig | Immunoglobulin |
| IMAC | Immobilised metal affinity chromatography |
| IPG | Immobilised pH gradient |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| IR | Iron reductase |
| LB | Luria-Bertani broth, Miller |
| LBA | Luria-Bertani broth agar |
| LC-MS | Liquid chromatography – mass spectrometry |
| LIC | Ligation independent cloning |
| LiP | Lignin peroxidase |
| LRR | Leucine rich repeat |
| MALDI | Matrix assisted laser desorption ionisation |
| MALDI-TOF/TOF | Matrix assisted laser desorption ionisation – tandem time of flight mass spectrometry |
| MBTH | 3-Methyl-2-benzothiazolinone hydrazone hydrochloride hydrate |
| MeCN | Acetonitrile |
| MFP | Molar percentage fraction |
| MnP | Manganese peroxidase |
| MOI | Multiplicity of infection |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| MSA | Multiple sequence alignment |
| MWCO | Molecular weight cut off |

| | |
|---|---|
| NAD | β-Nicotinamide adenine dinucleotide |
| NCBI | National Center for Biotechnology Information |
| NTD | N-terminal domain |
| OB | Occlusion bodies |
| ODV | Occlusion derived virus |
| ODx | Optical density, x = wavelength (nm) |
| ORF | Open reading frame |
| P.I. | Pre-immune serum |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PGIP | Polygalacturonase-inhibiting protein |
| PMO | Polysaccharide monooxygenase |
| PO | Phenoloxidase |
| polh | Polyhedrin |
| PPO | Pro-phenoloxidase |
| PTU | Phenylthiourea |
| qPCR | Reverse transcription quantitative polymerase chain reaction |
| RNA | Ribonucleic acid |
| RoB | Rest of body |
| ROS | Reactive oxygen species |
| RT | Room temperature |
| S:N | Signal to noise ratio |
| SDS | Sodium dodecyl sulphate |
| SDS-PAGE | Sodium dodecyl sulphate – polyacrylamide gel electrophoresis |
| SFM | Serum free media |
| SLH | S-layer homology |
| SNAP | Sophisticated numerical annotation procedure |
| TBE | Tris-borate EDTA |
| TBS | Tris buffered saline |
| TBST | Tris buffered saline with TWEEN 20 |
| Tm | Melting temperature |
| UPLC | Ultra performance liquid chromatography |
| UTR | Un-translated region |
| VP | Versatile peroxidase |
| X-Gal | 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside |

# References


1. EIA (2013) International Energy Outlook 2013. (U.S. Energy Information Administration).
2. IEA (2012) $CO_2$ Emissions From Fuel Combustion. (International Energy Agency).
3. Naik SN, Goud VV, Rout PK, & Dalai AK (2010) Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14(2):578-597.
4. Hochman G, Rajagopal D, & Zilberman D (2010) Are Biofuels the Culprit? OPEC, Food, and Fuel. *The American Economic Review* 100(2):183-187.
5. Poudel BN, Paudel KP, Timilsina G, & Zilberman D (2012) Providing Numbers for a Food versus Fuel Debate: An Analysis of a Future Biofuel Production Scenario. *Applied Economic Perspectives and Policy* 34(4):637-668.
6. Leung DYC, Wu X, & Leung MKH (2010) A review on biodiesel production using catalyzed transesterification. *Applied Energy* 87(4):1083-1095.
7. Singhabhandhu A & Tezuka T (2010) A perspective on incorporation of glycerin purification process in biodiesel plants using waste cooking oil as feedstock. *Energy* 35(6):2493-2504.
8. Ayoub M & Abdullah AZ (2012) Critical review on the current scenario and significance of crude glycerol resulting from biodiesel industry towards more sustainable renewable energy industry. *Renewable and Sustainable Energy Reviews* 16(5):2671-2686.
9. Behr A, Eilting J, Irawadi K, Leschinski J, & Lindner F (2008) Improved utilisation of renewable resources: New important derivatives of glycerol. *Green Chemistry* 10(1):13-30.
10. Weiland P (2010) Biogas production: current state and perspectives. *Applied Microbiology and Biotechnology* 85(4):849-860.
11. Braun R, Weiland P, & Wellinger A (2008) Biogas from energy crop digestion. *IEA Bioenergy, Task*.
12. Berglund M & Börjesson P (2006) Assessment of energy performance in the life-cycle of biogas production. *Biomass and Bioenergy* 30(3):254-266.
13. Buswell AM (1930) Production of fuel gas by anaerobic fermentations. *Industrial & Engineering Chemistry* 22(11):1168-1172.
14. Holm-Nielsen JB, Al Seadi T, & Oleskowicz-Popiel P (2009) The future of anaerobic digestion and biogas utilization. *Bioresource Technology* 100(22):5478-5484.
15. Perlack RD*, et al.* (2005) Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. (DTIC Document).
16. Quaye AK & Volk TA (Biomass production and soil nutrients in organic and inorganic fertilized willow biomass production systems. *Biomass and Bioenergy* (0).
17. Cadoux S, Riche AB, Yates NE, & Machet J-M (2012) Nutrient requirements of Miscanthus x giganteus: Conclusions from a review of published studies. *Biomass and Bioenergy* 38(0):14-22.
18. Cai X, Zhang X, & Wang D (2010) Land Availability for Biofuel Production. *Environmental Science & Technology* 45(1):334-339.
19. Lee J (1997) Biological conversion of lignocellulosic biomass to ethanol. *Journal of Biotechnology* 56(1):1-24.
20. DOE U (2006) Breaking the biological barriers to cellulosic ethanol: a joint research agenda. *based on a workshop (Rockville, MD, December 7 to 9, 2005)*.

21. Somerville C*, et al.* (2004) Toward a Systems Approach to Understanding Plant Cell Walls. *Science* 306(5705):2206-2211.
22. Cosgrove DJ (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6(11):850-861.
23. Zhong R, Lee C, & Ye Z-H (2010) Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trends in Plant Science* 15(11):625-632.
24. Wingren A, Galbe M, & Zacchi G (2003) Techno-Economic Evaluation of Producing Ethanol from Softwood: Comparison of SSF and SHF and Identification of Bottlenecks. *Biotechnology Progress* 19(4):1109-1117.
25. Gomez LD, Steele-King CG, & McQueen-Mason SJ (2008) Sustainable liquid biofuels from biomass: the writing's on the walls. *New Phytologist* 178(3):473-485.
26. Pauly M & Keegstra K (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *The Plant Journal* 54(4):559-568.
27. Scheller HV & Ulvskov P (2010) Hemicelluloses. *Annual Review of Plant Biology* 61(1):263-289.
28. McNeil M, Darvill AG, Fry SC, & Albersheim P (1984) Structure and function of the primary cell walls of plants. *Annual review of biochemistry* 53(1):625-663.
29. Boerjan W, Ralph J, & Baucher M (2003) Lignin Biosynthesis. *Annual Review of Plant Biology* 54(1):519-546.
30. Vanholme R, Demedts B, Morreel K, Ralph J, & Boerjan W (2010) Lignin Biosynthesis and Structure. *Plant Physiology* 153(3):895-905.
31. Baucher M, Monties B, Montagu MV, & Boerjan W (1998) Biosynthesis and Genetic Engineering of Lignin. *Critical Reviews in Plant Sciences* 17(2):125-197.
32. Bugg TDH, Ahmad M, Hardiman EM, & Rahmanpour R (2011) Pathways for degradation of lignin in bacteria and fungi. *Natural Product Reports* 28(12):1883-1896.
33. Stewart JJ, Akiyama T, Chapple C, Ralph J, & Mansfield SD (2009) The Effects on Lignin Structure of Overexpression of Ferulate 5-Hydroxylase in Hybrid Poplar1. *Plant Physiology* 150(2):621-635.
34. Vanholme R, Morreel K, Ralph J, & Boerjan W (2008) Lignin engineering. *Current Opinion in Plant Biology* 11(3):278-285.
35. Mohan D, Pittman CU, & Steele PH (2006) Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. *Energy & Fuels* 20(3):848-889.
36. Zhang Q, Chang J, Wang T, & Xu Y (2007) Review of biomass pyrolysis oil properties and upgrading research. *Energy Conversion and Management* 48(1):87-92.
37. Lappas AA, Samolada MC, Iatridis DK, Voutetakis SS, & Vasalos IA (2002) Biomass pyrolysis in a circulating fluid bed reactor for the production of fuels and chemicals. *Fuel* 81(16):2087-2095.
38. Puig-Arnavat M, Bruno JC, & Coronas A (2010) Review and analysis of biomass gasification models. *Renewable and Sustainable Energy Reviews* 14(9):2841-2851.
39. Tijmensen MJA, Faaij APC, Hamelinck CN, & van Hardeveld MRM (2002) Exploration of the possibilities for production of Fischer Tropsch liquids and power via biomass gasification. *Biomass and Bioenergy* 23(2):129-152.
40. Kirkels AF & Verbong GPJ (2011) Biomass gasification: Still promising? A 30-year global overview. *Renewable and Sustainable Energy Reviews* 15(1):471-481.
41. Asche F, Oglend A, & Osmundsen P (2012) Gas versus oil prices the impact of shale gas. *Energy Policy* 47(0):117-124.

42. Henstra AM, Sipma J, Rinzema A, & Stams AJM (2007) Microbiology of synthesis gas fermentation for biofuel production. *Current Opinion in Biotechnology* 18(3):200-206.
43. Henstra AM & Stams AJM (2011) Deep Conversion of Carbon Monoxide to Hydrogen and Formation of Acetate by the Anaerobic Thermophile Carboxydothermus hydrogenoformans. *International Journal of Microbiology* 2011.
44. Sipma J*, et al.* (2006) Microbial CO Conversions with Applications in Synthesis Gas Purification and Bio-Desulfurization. *Critical Reviews in Biotechnology* 26(1):41-65.
45. Galbe M & Zacchi G (2007) Pretreatment of Lignocellulosic Materials for Efficient Bioethanol Production. *Biofuels,* Advances in Biochemical Engineering/Biotechnology, ed Olsson L (Springer Berlin Heidelberg), Vol 108, pp 41-65.
46. da Costa Sousa L, Chundawat SPS, Balan V, & Dale BE (2009) 'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies. *Current Opinion in Biotechnology* 20(3):339-347.
47. Tassinari T, Macy C, Spano L, & Ryu DDY (1980) Energy requirements and process design considerations in compression-milling pretreatment of cellulosic wastes for enzymatic hydrolysis. *Biotechnology and Bioengineering* 22(8):1689-1705.
48. Wyman CE*, et al.* (2005) Coordinated development of leading biomass pretreatment technologies. *Bioresource Technology* 96(18):1959-1966.
49. Teymouri F, Laureano-Perez L, Alizadeh H, & Dale BE (2005) Optimization of the ammonia fiber explosion (AFEX) treatment parameters for enzymatic hydrolysis of corn stover. *Bioresource Technology* 96(18):2014-2018.
50. Lee SH, Doherty TV, Linhardt RJ, & Dordick JS (2009) Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. *Biotechnology and Bioengineering* 102(5):1368-1376.
51. Dadi AP, Varanasi S, & Schall CA (2006) Enhancement of cellulose saccharification kinetics using an ionic liquid pretreatment step. *Biotechnology and Bioengineering* 95(5):904-910.
52. Mora-Pale M, Meli L, Doherty TV, Linhardt RJ, & Dordick JS (2011) Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass. *Biotechnology and Bioengineering* 108(6):1229-1245.
53. Galbe M, Sassner P, Wingren A, & Zacchi G (2007) Process Engineering Economics of Bioethanol Production. *Biofuels,* Advances in Biochemical Engineering/Biotechnology, ed Olsson L (Springer Berlin Heidelberg), Vol 108, pp 303-327.
54. Yang B & Wyman CE (2008) Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining* 2(1):26-40.
55. Gilbert HJ (2010) The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction. *Plant Physiology* 153(2):444-455.
56. Jørgensen H, Kristensen JB, & Felby C (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining* 1(2):119-134.
57. Gao D*, et al.* (2011) Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnology for Biofuels* 4(1):1-11.
58. Watanabe H & Tokuda G (2010) Cellulolytic Systems in Insects. *Annual Review of Entomology* 55(1):609-632.
59. Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3(9):853-859.
60. Rye CS & Withers SG (2000) Glycosidase mechanisms. *Current Opinion in Chemical Biology* 4(5):573-580.

61. Kern M*, et al.* (2013) Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. *Proceedings of the National Academy of Sciences*.
62. Boraston AB, Bolam D, Gilbert H, & Davies GJGJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J* 382:769-781.
63. Tormo J*, et al.* (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *The EMBO journal* 15(21):5739-5751.
64. Boraston AB*, et al.* (2003) Structure and Ligand Binding of Carbohydrate-binding Module CsCBM6-3 Reveals Similarities with Fucose-specific Lectins and "Galactose-binding" Domains. *Journal of Molecular Biology* 327(3):659-669.
65. Cunniff J & Cerasuolo M (2011) Lighting the way to willow biomass production. *Journal of the Science of Food and Agriculture* 91(10):1733-1736.
66. Yang F*, et al.* (2013) Engineering secondary cell wall deposition in plants. *Plant Biotechnology Journal* 11(3):325-335.
67. Mortimer JC*, et al.* (2010) Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. *Proceedings of the National Academy of Sciences* 107(40):17409-17414.
68. King AJ*, et al.* (2010) Molecular insight into lignocellulose digestion by a marine isopod in the absence of gut microbes. *Proceedings of the National Academy of Sciences* 107(12):5345.
69. Lomascolo A, Stentelaire C, Asther M, & Lesage-Meessen L (1999) Basidiomycetes as new biotechnological tools to generate natural aromatic flavours for the food industry. *Trends in Biotechnology* 17(7):282-289.
70. Carriquiry MA, Du X, & Timilsina GR (2011) Second generation biofuels: Economics and policies. *Energy Policy* 39(7):4222-4234.
71. Sims REH, Mabee W, Saddler JN, & Taylor M (2010) An overview of second generation biofuel technologies. *Bioresource Technology* 101(6):1570-1580.
72. Luo C, Brink DL, & Blanch HW (2002) Identification of potential fermentation inhibitors in conversion of hybrid poplar hydrolyzate to ethanol. *Biomass and Bioenergy* 22(2):125-138.
73. Palmqvist E, Hahn-Hägerdal B, Galbe M, & Zacchi G (1996) The effect of water-soluble inhibitors from steam-pretreated willow on enzymatic hydrolysis and ethanol fermentation. *Enzyme and Microbial Technology* 19(6):470-476.
74. Palmqvist E & Hahn-Hägerdal B (2000) Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresource Technology* 74(1):25-33.
75. Delgenes JP, Moletta R, & Navarro JM (1996) Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by Saccharomyces cerevisiae, Zymomonas mobilis, Pichia stipitis, and Candida shehatae. *Enzyme and Microbial Technology* 19(3):220-225.
76. Jeffries TW & Jin YS (2004) Metabolic engineering for improved fermentation of pentoses by yeasts. *Applied Microbiology and Biotechnology* 63(5):495-509.
77. Tao H*, et al.* (2001) Engineering a Homo-Ethanol Pathway inEscherichia coli: Increased Glycolytic Flux and Levels of Expression of Glycolytic Genes during Xylose Fermentation. *Journal of Bacteriology* 183(10):2979-2988.
78. Sedlak M, Edenberg HJ, & Ho NWY (2003) DNA microarray analysis of the expression of the genes encoding the major enzymes in ethanol production during glucose and xylose co-fermentation by metabolically engineered Saccharomyces yeast. *Enzyme and Microbial Technology* 33(1):19-28.
79. Van Vleet JH & Jeffries TW (2009) Yeast metabolic engineering for hemicellulosic ethanol production. *Current Opinion in Biotechnology* 20(3):300-306.

80. Nielsen J, Larsson C, van Maris A, & Pronk J (2013) Metabolic engineering of yeast for production of fuels and chemicals. *Current Opinion in Biotechnology* 24(3):398-404.
81. Nakamura CE & Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* 14(5):454-459.
82. Ballesteros M, Oliva JM, Negro MJ, Manzanares P, & Ballesteros I (2004) Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SFS) with Kluyveromyces marxianus CECT 10875. *Process Biochemistry* 39(12):1843-1848.
83. Hari Krishna S, Janardhan Reddy T, & Chowdary GV (2001) Simultaneous saccharification and fermentation of lignocellulosic wastes to ethanol using a thermotolerant yeast. *Bioresource Technology* 77(2):193-196.
84. Saha BC, Nichols NN, Qureshi N, & Cotta MA (2011) Comparison of separate hydrolysis and fermentation and simultaneous saccharification and fermentation processes for ethanol production from wheat straw by recombinant Escherichia coli strain FBR5. *Applied Microbiology and Biotechnology* 92(4):865-874.
85. Mamma D*, et al.* (1996) Bioethanol from sweet sorghum: Simultaneous saccharification and fermentation of carbohydrates by a mixed microbial culture. *Process Biochemistry* 31(4):377-381.
86. Antón J, Rosselló-Mora R, Rodríguez-Valera F, & Amann R (2000) Extremely Halophilic Bacteria in Crystallizer Ponds from Solar Salterns. *Applied and Environmental Microbiology* 66(7):3052-3057.
87. Slobodkin AI*, et al.* (2012) Thermosulfurimonas dismutans gen. nov., sp. nov., an extremely thermophilic sulfur-disproportionating bacterium from a deep-sea hydrothermal vent. *International Journal of Systematic and Evolutionary Microbiology* 62(Pt 11):2565-2571.
88. Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycological Research* 105(12):1422-1432.
89. Bayer EA, Lamed R, White BA, & Flint HJ (2008) From cellulosomes to cellulosomics. *The Chemical Record* 8(6):364-377.
90. Vicuña R (1988) Bacterial degradation of lignin. *Enzyme and Microbial Technology* 10(11):646-655.
91. Zimmermann W (1990) Degradation of lignin by bacteria. *Journal of Biotechnology* 13(2–3):119-130.
92. Kirk TK & Farrell RL (1987) Enzymatic "Combustion": The Microbial Degradation of Lignin. *Annual Review of Microbiology* 41(1):465-501.
93. ten Have R & Teunissen PJM (2001) Oxidative Mechanisms Involved in Lignin Degradation by White-Rot Fungi. *Chemical Reviews* 101(11):3397-3414.
94. Martinez D*, et al.* (2009) Genome, transcriptome, and secretome analysis of wood decay fungus Postia placenta supports unique mechanisms of lignocellulose conversion. *Proceedings of the National Academy of Sciences* 106(6):1954-1959.
95. Yelle DJ, Ralph J, Lu F, & Hammel KE (2008) Evidence for cleavage of lignin by a brown rot basidiomycete. *Environmental Microbiology* 10(7):1844-1849.
96. Ramachandra M, Crawford DL, & Hertel G (1988) Characterization of an extracellular lignin peroxidase of the lignocellulolytic actinomycete Streptomyces viridosporus. *Applied and Environmental Microbiology* 54(12):3057-3063.
97. Camarero S, Sarkar S, Ruiz-Dueñas FJ, Martínez MaJ, & Martínez ÁT (1999) Description of a Versatile Peroxidase Involved in the Natural Degradation of Lignin That Has Both Manganese Peroxidase and Lignin Peroxidase Substrate Interaction Sites. *Journal of Biological Chemistry* 274(15):10324-10330.
98. Ruiz-Dueñas FJ, Camarero S, Pérez-Boada M, Martínez MJ, & Martínez AT (2001) A new versatile peroxidase from Pleurotus. *Biochemical Society transactions* 29(Pt 2):116-122.

99. Paszczyński A, Huynh V-B, & Crawford R (1985) Enzymatic activities of an extracellular, manganese-dependent peroxidase from Phanerochaete chrysosporium. *FEMS Microbiology Letters* 29(1-2):37-41.
100. Kuwahara M, Glenn JK, Morgan MA, & Gold MH (1984) Separation and characterization of two extracelluar H2O2-dependent oxidases from ligninolytic cultures of Phanerochaete chrysosporium. *FEBS Letters* 169(2):247-250.
101. Ahmad M*, et al.* (2011) Identification of DypB from Rhodococcus jostii RHA1 as a Lignin Peroxidase. *Biochemistry* 50(23):5096-5107.
102. Roberts JN*, et al.* (2011) Characterization of Dye-Decolorizing Peroxidases from Rhodococcus jostii RHA1. *Biochemistry* 50(23):5108-5119.
103. Hofrichter M (2002) Review: lignin conversion by manganese peroxidase (MnP). *Enzyme and Microbial Technology* 30(4):454-466.
104. Hammel KE & Cullen D (2008) Role of fungal peroxidases in biological ligninolysis. *Current Opinion in Plant Biology* 11(3):349-355.
105. Khindaria A, Yamazaki I, & Aust SD (1996) Stabilization of the Veratryl Alcohol Cation Radical by Lignin Peroxidase. *Biochemistry* 35(20):6418-6424.
106. Sollewijn Gelpke MD, Lee J, & Gold MH (2002) Lignin Peroxidase Oxidation of Veratryl Alcohol: Effects of the Mutants H82A, Q222A, W171A, and F267L. *Biochemistry* 41(10):3498-3506.
107. Tuor U, Wariishi H, Schoemaker HE, & Gold MH (1992) Oxidation of phenolic arylglycerol beta-aryl ether lignin model compounds by manganese peroxidase from Phanerochaete chrysosporium: oxidative cleavage of an alpha-carbonyl model compound. *Biochemistry* 31(21):4986-4995.
108. Hammel KE, Tien M, Kalyanaraman B, & Kirk TK (1985) Mechanism of oxidative C alpha-C beta cleavage of a lignin model dimer by Phanerochaete chrysosporium ligninase. Stoichiometry and involvement of free radicals. *Journal of Biological Chemistry* 260(14):8348-8353.
109. Enoki A, Goldsby GP, & Gold MH (1980) Metabolism of the lignin model compounds veratrylglycerol-β-guaiacyl ether and 4-ethoxy-3-methoxyphenylglycerol-β-guaiacyl ether by Phanerochaete chrysosporium. *Archives of Microbiology* 125(3):227-231.
110. Enoki A & Gold MH (1982) Degradation of the diarylpropane lignin model compound 1-(3′,4′-diethoxyphenyl)-1,3-dihydroxy-2-(4″-methoxyphenyl)-propane and derivatives by the basidiomycete Phanerochaete chrysosporium. *Archives of Microbiology* 132(2):123-130.
111. Ohta M, Higuchi T, & Iwahara S (1979) Microbial degradation of dehydrodiconiferyl alcohol, a lignin substructure model. *Archives of Microbiology* 121(1):23-28.
112. Nakatsubo F, Kirk TK, Shimada M, & Higuchi T (1981) Metabolism of a phenylcoumaran substructure lignin model compound in ligninolytic cultures of Phanerochaete chrysosporium. *Archives of Microbiology* 128(4):416-420.
113. Masai E*, et al.* (1993) Characterization of the Cα-Dehydrogenase Gene Involved in the Cleavage of β-Aryl Ether by Pseudomonas paucimobilis. *Bioscience, Biotechnology, and Biochemistry* 57(10):1655-1659.
114. Jensen KA, Houtman CJ, Ryan ZC, & Hammel KE (2001) Pathways for Extracellular Fenton Chemistry in the Brown Rot Basidiomycete Gloeophyllum trabeum. *Applied and Environmental Microbiology* 67(6):2705-2711.
115. Gómez-Toribio V, García-Martín AB, Martínez MJ, Martínez ÁT, & Guillén F (2009) Induction of Extracellular Hydroxyl Radical Production by White-Rot Fungi through Quinone Redox Cycling. *Applied and Environmental Microbiology* 75(12):3944-3953.
116. Eastwood DC*, et al.* (2011) The Plant Cell Wall–Decomposing Machinery Underlies the Functional Diversity of Forest Fungi. *Science* 333(6043):762-765.

117. Harris PV*, et al.* (2010) Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. *Biochemistry* 49(15):3305-3316.
118. Quinlan RJ*, et al.* (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences* 108(37):15079-15084.
119. Phillips CM, Beeson WT, Cate JH, & Marletta MA (2011) Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa. *ACS Chemical Biology* 6(12):1399-1406.
120. Karkehabadi S*, et al.* (2008) The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from Hypocrea jecorina, at 1.6 Å Resolution. *Journal of Molecular Biology* 383(1):144-154.
121. Westereng B*, et al.* (2011) The Putative Endoglucanase PcGH61D from Phanerochaete chrysosporium Is a Metal-Dependent Oxidative Enzyme that Cleaves Cellulose. *PLoS ONE* 6(11):e27807.
122. Beeson WT, Phillips CM, Cate JHD, & Marletta MA (2011) Oxidative Cleavage of Cellulose by Fungal Copper-Dependent Polysaccharide Monooxygenases. *Journal of the American Chemical Society* 134(2):890-892.
123. Geib SM*, et al.* (2008) Lignin degradation in wood-feeding insects. *Proceedings of the National Academy of Sciences* 105(35):12932-12937.
124. Tartar A*, et al.* (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes. *Biotechnology for Biofuels* 2(1):1-19.
125. Hess M*, et al.* (2011) Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science* 331(6016):463-467.
126. Zimmer M & Topp W (1998) Microorganisms and Cellulose Digestion in the Gut of the Woodlouse Porcellio scaber. *Journal of Chemical Ecology* 24(8):1397-1408.
127. Kudo T (2009) Termite-Microbe Symbiotic System and Its Efficient Degradation of Lignocellulose. *Bioscience, Biotechnology, and Biochemistry* 73(12):2561-2567.
128. Franco Cairo JP*, et al.* (2011) Functional characterization and target discovery of glycoside hydrolases from the digestome of the lower termite Coptotermes gestroi. *Biotechnology for Biofuels* 4(1):50.
129. Boyle PJ & Mitchell R (1978) Absence of Microorganisms in Crustacean Digestive Tracts. *Science* 200(4346):1157-1159.
130. Brunet M, Arnaud J, & Mazza J (1994) Gut structure and digestive cellular processes in marine Crustacea. *Oceanography and Marine Biology: An Annual Review* 32:335-367.
131. Tupper C (1999) The functional anatomy of the gut and feeding related studies of the marine wood boring isopod Limnoria sp. (Limnoriidae: Isopoda). (Buckinghamshire New University).
132. Dymond J (2005) The functional anatomy of the digestive system of the marine wood-boring isopod Limnoria quadripunctata, with evidence of endogenous cellulase production. PhD (University of Portsmouth).
133. Humes AG (1992) *Microscopic Anatomy of Invertebrates, Crustacea* (Wiley).
134. Lešer V*, et al.* (2008) Epithelial thickness and lipid droplets in the hepatopancreas of Porcellio scaber (Crustacea: Isopoda) in different physiological conditions. *Zoology* 111(6):419-432.
135. Clifford B & Witkus ER (1971) The fine structure of the hepatopancreas of the woodlouse, Oniscus ascellus. *Journal of Morphology* 135(3):335-349.
136. Cantarel BL*, et al.* (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* 37(suppl 1):D233-D238.

137. Decker H*, et al.* (2007) Similar enzyme activation and catalysis in hemocyanins and tyrosinases. *Gene* 398(1-2):183-191.
138. Laemmli UK (1970) Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 227(5259):680-685.
139. Mason HS (1948) The chemistry of melanin: III. Mechanism of the oxidation of dihydroxyphenylalanine by tyrosinase. *Journal of Biological Chemistry* 172(1):83-99.
140. Winder AJ & Harris H (1991) New assays for the tyrosine hydroxylase and dopa oxidase activities of tyrosinase. *European Journal of Biochemistry* 198(2):317-326.
141. Savitzky A & Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36(8):1627-1639.
142. Wulfkuhle JD, Liotta LA, & Petricoin EF (2003) Proteomic applications for the early detection of cancer. *Nat Rev Cancer* 3(4):267-275.
143. Wulfkuhle JD*, et al.* (2001) New approaches to proteomic analysis of breast cancer. *PROTEOMICS* 1(10):1205-1215.
144. Bichsel VE, Liotta LA, & Petricoin EF (2001) Cancer proteomics: from biomarker discovery to signal pathway profiling. *Cancer journal (Sudbury, Mass.)* 7(1):69-78.
145. Emmert-Buck MR*, et al.* (2000) An approach to proteomic analysis of human tumors. *Molecular Carcinogenesis* 27(3):158-165.
146. Pan J, Chen H-Q, Sun Y-H, Zhang J-H, & Luo X-Y (2008) Comparative Proteomic Analysis of Non-small-cell Lung Cancer and Normal Controls Using Serum Label-Free Quantitative Shotgun Technology. *Lung* 186(4):255-261.
147. Xu D*, et al.* (2008) Novel MMP-9 Substrates in Cancer Cells Revealed by a Label-free Quantitative Proteomics Approach. *Molecular & Cellular Proteomics* 7(11):2215-2228.
148. Grønborg M*, et al.* (2006) Biomarker Discovery from Pancreatic Cancer Secretome Using a Differential Proteomic Approach. *Molecular & Cellular Proteomics* 5(1):157-171.
149. Wiśniewski JR, Zougman A, Nagaraj N, & Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Meth* 6(5):359-362.
150. Ishihama Y*, et al.* (2005) Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics* 4(9):1265-1272.
151. Shinoda K, Tomita M, & Ishihama Y (2010) emPAI Calc—for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* 26(4):576-577.
152. Pettersen Roger C (1984) The Chemical Composition of Wood. *The Chemistry of Solid Wood,* Advances in Chemistry, (American Chemical Society), Vol 207, pp 57-126.
153. Malyon G (2011) Insight into the digestive processes of the wood-boring marine crustacean Limnoria quadripunctata. PhD (University of Portsmouth).
154. Mann S, Bannister JV, & Williams RJP (1986) Structure and composition of ferritin cores isolated from human spleen, limpet (Patella vulgata) hemolymph and bacterial (Pseudomonas aeruginosa) cells. *Journal of Molecular Biology* 188(2):225-232.
155. Theil EC (1987) Ferritin: Structure, Gene Regulation, and Cellular Function in Animals, Plants, and Microorganisms. *Annual Review of Biochemistry* 56(1):289-315.
156. Truchot JP (1992) Respiratory Function of Arthropod Hemocyanins. *Blood and Tissue Oxygen Carriers,* Advances in Comparative and Environmental Physiology, ed Mangum C (Springer Berlin Heidelberg), Vol 13, pp 377-410.

157. Di Matteo A*, et al.* (2003) The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proceedings of the National Academy of Sciences* 100(17):10124-10128.
158. Volbeda A & Hol WGJ (1989) Crystal structure of hexameric haemocyanin from Panulirus interruptus refined at 3.2 Å resolution. *Journal of Molecular Biology* 209(2):249-279.
159. Magnus K & Ton-That H (1992) The crystal structure of the oxygenated form of Limulus polyphemus subunit II hemocyanin. *Journal of Inorganic Biochemistry* 47(3-4):20-20.
160. van Holde KE & Miller KI (1995) Hemocyanins. *Advances in Protein Chemistry*, eds C.B. Anfinsen FMRJTE & David SE (Academic Press), Vol Volume 47, pp 1-81.
161. van Holde KE, Miller KI, & Decker H (2001) Hemocyanins and Invertebrate Evolution. *Journal of Biological Chemistry* 276(19):15563-15566.
162. Cuff ME, Miller KI, van Holde KE, & Hendrickson WA (1998) Crystal structure of a functional unit from Octopus hemocyanin. *Journal of Molecular Biology* 278(4):855-870.
163. Kitajima N*, et al.* (1992) A new model for dioxygen binding in hemocyanin. Synthesis, characterization, and molecular structure of the .mu.-.eta.2:.eta.2 peroxo dinuclear copper (II) complexes,[Cu (HB (3, 5-R2pz) 3)] 2 (O2)(R= isopropyl and Ph). *Journal of the American Chemical Society* 114(4):1277-1291.
164. Mangum CP, Greaves J, & Rainer JS (1991) Oligomer Composition and Oxygen Binding of the Hemocyanin of the Blue Crab Callinectes sapidus. *Biol Bull* 181(3):453-458.
165. Hellmann N, Paoli M, Giomi F, & Beltramini M (2010) Unusual oxygen binding behavior of a 24-meric crustacean hemocyanin. *Archives of biochemistry and biophysics* 495(2):112-121.
166. Bridges CR (2001) Modulation of haemocyanin oxygen affinity: properties and physiological implications in a changing world. *The Journal of experimental biology* 204(Pt 5):1021.
167. Decker H & Sterner R (1990) Nested allostery of arthropodan hemocyanin (Eurypelma californicum and Homarus americanus) : The role of protons. *Journal of Molecular Biology* 211(1):281-293.
168. Kuiper HA*, et al.* (1975) Subunit composition, X-ray diffraction, amino acid analysis and oxygen binding behaviour of Panulirus interruptus hemocyanin. *Journal of Molecular Biology* 99(4):619-629.
169. Terwilliger NB, Terwilliger RC, Applestein M, Bonaventura C, & Bonaventura J (1979) Subunit structure and oxygen binding by hemocyanin of the isopod Ligia exotica. *Biochemistry* 18(1):102-108.
170. Decker H & Jaenicke E (2004) Recent findings on phenoloxidase activity and antimicrobial activity of hemocyanins. *Developmental and comparative immunology* 28(7-8):673-687.
171. Nagai T, Osaki T, & Kawabata S (2001) Functional conversion of hemocyanin to phenoloxidase by horseshoe crab antimicrobial peptides. *Journal of Biological Chemistry* 276(29):27166.
172. Decker H & Rimke T (1998) Tarantula hemocyanin shows phenoloxidase activity. *Journal of Biological Chemistry* 273(40):25889.
173. Lee SY, Lee BL, & Söderhäll K (2004) Processing of crayfish hemocyanin subunits into phenoloxidase. *Biochemical and biophysical research communications* 322(2):490-496.
174. Nillius D, Jaenicke E, & Decker H (2008) Switch between tyrosinase and catecholoxidase activity of scorpion hemocyanin by allosteric effectors. *FEBS letters* 582(5):749-754.

175. Fujieda N, Yakiyama A, & Itoh S (2010) Catalytic oxygenation of phenols by arthropod hemocyanin, an oxygen carrier protein, from Portunus trituberculatus. *Dalton Transactions* 39(12):3083-3092.
176. Adachi K, Hirata T, Nishioka T, & Sakaguchi M (2003) Hemocyte components in crustaceans convert hemocyanin into a phenoloxidase-like enzyme. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 134(1):135-141.
177. Decker H, Ryan M, Jaenicke E, & Terwilliger N (2001) SDS-induced Phenoloxidase Activity of Hemocyanins from Limulus polyphemus, Eurypelma californicum, and Cancer magister. *Journal of Biological Chemistry* 276(21):17796.
178. Decker H & Tuczek F (2000) Tyrosinase/catecholoxidase activity of hemocyanins: structural basis and molecular mechanism. *Trends in Biochemical Sciences* 25(8):392-397.
179. Sugumaran M (2009) Complexities of cuticular pigmentation in insects. *Pigment Cell & Melanoma Research* 22(5):523-525.
180. Pye AE (1974) Microbial activation of prophenoloxidase from immune insect larvae. *Nature* 251(5476):610-613.
181. Cerenius L, Lee BL, & Söderhäll K (2008) The proPO-system: pros and cons for its role in invertebrate immunity. *Trends in immunology* 29(6):263-271.
182. Moussian B (2010) Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochemistry and Molecular Biology*.
183. Sugumaran M (1998) Unified mechanism for sclerotization of insect cuticle. *Advances in Insect Physiology* 27:229-334.
184. Sritunyalucksana K (2000) The proPO and clotting system in crustaceans. *Aquaculture*.
185. Sugumaran M (2002) Comparative biochemistry of eumelanogenesis and the protective roles of phenoloxidase and melanin in insects. *Pigment Cell Research* 15(1):2-9.
186. Jiang N, Tan NS, Ho B, & Ding JL (2007) Respiratory protein–generated reactive oxygen species as an antimicrobial strategy. *Nature Immunology* 8(10):1114-1122.
187. Thalhimer W & Palmer B (1911) A comparison of the bactericidal action of quinone with that of some of the commoner disinfectants. *The Journal of Infectious Diseases* 9(2):181-189.
188. Bolton JL, Trush MA, Penning TM, Dryhurst G, & Monks TJ (2000) Role of Quinones in Toxicology. *Chem. Res. Toxicol* 13(3):135-160.
189. O'Brien PJ (1991) Molecular mechanisms of quinone cytotoxicity. *Chemico-biological interactions* 80(1):1-41.
190. Wardman P & Candeias LP (1996) Fenton chemistry: an introduction. *Radiation research* 145(5):523-531.
191. Monks TJ, Hanzlik RP, Cohen GM, Ross D, & Graham DG (1992) Quinone chemistry and toxicity. *Toxicology and Applied Pharmacology* 112(1):2-16.
192. Walling C (1975) Fenton's reagent revisited. *Accounts of Chemical Research* 8(4):125-131.
193. Kirk TK, Ibach R, Mozuch MD, Conner AH, & Highley L (1991) Characteristics of cotton cellulose depolymerized by a brown-rot fungus, by acid, or by chemical oxidants. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood* 45(4):239-244.
194. Baldrian P & Valášková V (2008) Degradation of cellulose by basidiomycetous fungi. *FEMS Microbiology Reviews* 32(3):501-521.
195. Cohen R & Jensen KA (2002) Significant levels of extracellular reactive oxygen species produced by brown rot basidiomycetes on cellulose. *FEBS letters* 531(3):483-488.

196. Ander P & Eriksson KE (1976) The importance of phenol oxidase activity in lignin degradation by the white-rot fungus Sporotrichum pulverulentum. *Archives of Microbiology* 109(1):1-8.
197. Faison BD & Kirk TK (1983) Relationship between lignin degradation and production of reduced oxygen species by Phanerochaete chrysosporium. *Applied and environmental microbiology* 46(5):1140.
198. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Research* 29(9):e45-e45.
199. Zhou M, Diwu Z, Panchuk-Voloshina N, & Haugland RP (1997) A Stable Nonfluorescent Derivative of Resorufin for the Fluorometric Determination of Trace Hydrogen Peroxide: Applications in Detecting the Activity of Phagocyte NADPH Oxidase and Other Oxidases. *Analytical Biochemistry* 253(2):162-168.
200. Bhattacharya A*, et al.* (2009) Denervation Induces Cytosolic Phospholipase A2-mediated Fatty Acid Hydroperoxide Generation by Muscle Mitochondria. *Journal of Biological Chemistry* 284(1):46-55.
201. Lombardi A*, et al.* (2010) UCP3 Translocates Lipid Hydroperoxide and Mediates Lipid Hydroperoxide-dependent Mitochondrial Uncoupling. *Journal of Biological Chemistry* 285(22):16599-16605.
202. Klabunde T, Eicken C, Sacchettini JC, & Krebs B (1998) Crystal structure of a plant catechol oxidase containing a dicopper center. *Nat Struct Mol Biol* 5(12):1084-1090.
203. Gerdemann C, Eicken C, & Krebs B (2002) The Crystal Structure of Catechol Oxidase: New Insight into the Function of Type-3 Copper Proteins. *Accounts of Chemical Research* 35(3):183-191.
204. DuBois KP & Erway WF (1946) Studies on the mechanism of action of thiourea and related compounds: II. Inhibition of oxidative enzymes and oxidations catalyzed by copper. *Journal of Biological Chemistry* 165(2):711-721.
205. Bustin SA*, et al.* (2009) The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clinical Chemistry* 55(4):611-622.
206. Nolan T, Hands RE, & Bustin SA (2006) Quantification of mRNA using real-time RT-PCR. *Nat. Protocols* 1(3):1559-1582.
207. Janson J-C (2012) *Protein purification: principles, high resolution methods, and applications* (Wiley. com).
208. O'Reilly DR, Miller LK, & Luckow VA (1994) *The baculovirus expression vectors: a laboratory manual* (Oxford University Press) p 360.
209. Kost TA, Condreay JP, & Jarvis DL (2005) Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotech* 23(5):567-575.
210. Sano K-I, Maeda K, Oki M, & Maéda Y (2002) Enhancement of protein expression in insect cells by a lobster tropomyosin cDNA leader sequence. *FEBS Letters* 532(1–2):143-146.
211. Zhao Y, Chapman DAG, & Jones IM (2003) Improving baculovirus recombination. *Nucleic Acids Research* 31(2):e6-e6.
212. Higgins MK, Demir M, & Tate CG (2003) Calnexin co-expression and the use of weaker promoters increase the expression of correctly assembled Shaker potassium channel in insect cells. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1610(1):124-132.
213. Makrides SC (1996) Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews* 60(3):512-538.
214. Baneyx F (1999) Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology* 10(5):411-421.
215. Sørensen HP & Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology* 115(2):113-128.

216. Assenberg R, Wan PT, Geisse S, & Mayr LM (2013) Advances in recombinant protein expression for use in pharmaceutical research. *Current Opinion in Structural Biology* 23(3):393-402.
217. Studier FW & Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189(1):113-130.
218. Rosenberg AH*, et al.* (1987) Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene* 56(1):125-135.
219. William Studier F, Rosenberg AH, Dunn JJ, & Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzymology*, ed David VG (Academic Press), Vol Volume 185, pp 60-89.
220. Sorensen H & Mortensen K (2005) Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories* 4(1):1.
221. Hannig G & Makrides SC (1998) Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in Biotechnology* 16(2):54-60.
222. Weickert MJ, Doherty DH, Best EA, & Olins PO (1996) Optimization of heterologous protein production in Escherichia coli. *Curr Opin Biotechnol* 7:494 - 499.
223. Strober W (2001) Trypan Blue Exclusion Test of Cell Viability. *Current Protocols in Immunology*, (John Wiley & Sons, Inc.).
224. de Marco A, Volrath S, Bruyere T, Law M, & Fonné-Pfister R (2000) Recombinant Maize Protoporphyrinogen IX Oxidase Expressed in Escherichia coli Forms Complexes with GroEL and DnaK Chaperones. *Protein Expression and Purification* 20(1):81-86.
225. Thain A, Gaston K, Jenkins O, & Clarke AR (1996) A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends in Genetics* 12(6):209-210.
226. Studier FW (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology* 219(1):37-44.
227. Grunberg-Manago M (1999) Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annual Review of Genetics* 33(1):193-227.
228. Lopez PJ, Marchand I, Joyce SA, & Dreyfus M (1999) The C-terminal half of RNase E, which organizes the Escherichia coli degradosome, participates in mRNA degradation but not rRNA processing in vivo. *Molecular Microbiology* 33(1):188-199.
229. Stewart EJ, Aslund F, & Beckwith J (1998) Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *EMBO J* 17:5543 - 5550.
230. Bessette PH, Aslund F, Beckwith J, & Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc Natl Acad Sci U S A* 96:13703 - 13708.
231. Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Current Opinion in Biotechnology* 6(5):494-500.
232. Le Calvez H, Green JM, & Baty D (1996) Increased efficiency of alkaline phosphatase production levels in Escherichia coli using a degenerate Pe1B signal sequence. *Gene* 170(1):51-55.
233. Schein CH (1989) Production of soluble recombinant proteins in bacteria. *BioTechnology* 7:1141 - 1148.
234. Grossman TH, Kawasaki ES, Punreddy SR, & Osburne MS (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* 209(1–2):95-103.

235. Mendoza JA, Dulin P, & Warren T (2000) The Lower Hydrolysis of ATP by the Stress Protein GroEL Is a Major Factor Responsible for the Diminished Chaperonin Activity at Low Temperature. *Cryobiology* 41(4):319-323.
236. Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN, & Timmis KN (2003) Chaperonins govern growth of Escherichia coli at low temperatures. *Nat Biotech* 21(11):1266-1267.
237. Cazemier A*, et al.* (2003) Promicromonospora pachnodae sp. nov., a member of the (hemi)cellulolytic hindgut flora of larvae of the scarab beetle Pachnoda marginata. *Antonie van Leeuwenhoek* 83(2):135-148.
238. Breznak JA & Brune A (1994) Role of Microorganisms in the Digestion of Lignocellulose by Termites. *Annual Review of Entomology* 39(1):453-487.
239. Smith DB & Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* 67(1):31-40.
240. di Guana C, Lib P, Riggsa PD, & Inouyeb H (1988) Vectors that facilitate the expression and purification of foreign peptides in Escherichia coli by fusion to maltose-binding protein. *Gene* 67(1):21-30.
241. LaVallie ER*, et al.* (1993) A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the E. coli Cytoplasm. *Nat Biotech* 11(2):187-193.
242. Bardwell JCA, McGovern K, & Beckwith J (1991) Identification of a protein required for disulfide bond formation in vivo. *Cell* 67(3):581-589.
243. Zhang Z*, et al.* (2002) Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in Escherichia coli. *Protein Expression and Purification* 26(2):218-228.
244. Iiyama K & Wallis AFA (1988) An improved acetyl bromide procedure for determining lignin in woods and wood pulps. *Wood Science and Technology* 22(3):271-280.
245. Ahmad M*, et al.* (2010) Development of novel assays for lignin degradation: comparative analysis of bacterial and fungal lignin degraders. *Molecular BioSystems* 6(5):815-821.