Characterisation and biochemical activity testing of CE1 and GH78 family enzymes; towards improved glycosyl hydrolases for the production of biofuel and biobased products.

Stephen Christopher Austin

School of Science, Engineering and Environment

University of Salford

2021/202

Acknowledgements

The author would like to acknowledge the holy spirit for the strength and endurance to persevere through all the challenging times encountered while completing this work. Secondly, I would like to thank my mum and dad, Miss Veronica Honny and Mr. Stephen Austin for their unending love and support. Without their prayer and support, I would not have accomplished this great feat. We acknowledge Manrochem and the University of Salford for making possible this project through their generous funding contributions. Special thanks to, Drs. Natalie Ferry, Andy Clarke, Rhoderick Elder, Ian Goodhead and Shweta Yogesh Kuba for their immense contributions towards research challenges I encountered. I would also like to thank my friends Ayoola Olaweraju, Dr. Oluwasegun Iwakun, Henry Madubuike, Cecil Kwakye, Alex Aidoo-Micah and Michael Banwo. Little do they know how much their support has helped paint a bigger picture. I would like to finally thank all who contributed to making this chapter of my life a success.

Declaration

This thesis made use of metagenomic data from a study by Joynson *et al*., 2017. However, I declare that this thesis was written entirely by myself and that it has not been presented in any previous application for a degree, in whole or in part.

Chapter 1. A global problem

Abstract

Traditional chemical industries currently involved in the manufacturing and processing of high-value molecules used in the food, pharmaceutical and biofuel sector rely heavily on chemicals derived from fossil fuels such as crude oil. Fossil fuels represent a non-sustainable feedstock and although fossil fuels are used in producing high-value chemicals, their use contributes to greenhouse gas emissions consequently affecting global climate change. As a result, there is a drive to generate bio-based products and biofuel from sustainable alternatives to address the future needs of the industry as we advance towards a sustainable biobased economy. Biobased products and biofuels represent almost EUR 57 billion in annual revenue with sales expected to rise annually. The Biotechnology and Biological Sciences Research Council (BBSRC) and Lignocellulosic biorefinery network (LB Net) have put forward a list of prioritised chemicals such as levoglucosenone, 2,5-furandicarboxylic acid and itaconic acid, which can replace and improve upon petrochemicals and drive growth in the biobased economy.

Plant biomass has been described as an alternative feedstock to replace crude oil as a raw material. However, efficient, and sustainable degradation of the more resistant components of plant biomass represents a substantial technical challenge. Several studies have focused on the enzymatic hydrolysis of recalcitrant lignocellulosic biomass (cellulose, hemicellulose, and lignin) into various sugars and aromatic monomers which will form the basic units of the manufacturing process. Recent advances in metagenomic and meta transcriptomic studies have resulted in newer tools to explore previously understudied environments such as extreme environments including marine water, sediments, compost, and manures to identify novel enzymes. More recently herbivore guts have been intensively studied, including the repertoire of biomass degrading bacterial enzymes found in the gut of the black slug, *Arion ater* where a functional metagenome analysis identified 3,383 genes known to be involved

in the degradation of plant biomass. This study has screened the slug metagenomic library, selected and recombinantly expressed proteins with putative glycosyl hydrolase functionality and further tested the activity of these recombinant enzymes. Further to this, our study also looked at immobilisation techniques such as nanoparticles and cross-linked enzyme aggregates as a method of improving enzyme functionality.

1.1 Chemical industries and fossil fuels

The requirements for provision of fuel and chemicals (ethylene, methanol, and xylene) for manufacturing of plastics, textiles, and cosmetics and to meet our energy needs are continuously expanding to keep up with the growing human population. The production of these products by traditional chemical industries require the burning of fossil fuels such as natural gas, coal and crude oil (International Energy Agency, 2018; Levin, 2019). Fossil fuels are made up of dead organic matter from plants and animals accumulated over geological time frames and have been used to fuel our homes, workplaces, and drive mass industrialization. They have been a relatively cheap and abundant source of energy and chemicals because; they are reliable and readily available, with countries such as the United States of America (USA) known to have large coal and oil reserves (America's Energy Future Panel on Alternative Liquid Transportation Fuels *et al.*, 2009), and of course global oil availability has fuelled generations of geo-political conflict (Vakulchuk, Overland and Scholten, 2020).

With respect to energy production, we have relied on fossil fuels to drive human activities since the 1800's and this has resulted in observable changes in weather patterns and gradual shifts in temperatures, due mainly to the increase in the production of Greenhouse gas (GHG) emissions associated with human activities such as mining (United Nations, 2021). While some lobbyists deny global warming is induced by fossil fuel use - rather ascribing it to be a result of natural variations in the solar cycle, the scientific consensus is now beyond doubt that man-made activities are directly linked climate instability and temperature rises (The Royal Society and National Academy of Sciences, 2020). Global average surface temperature has increased by roughly a degree, approximately 0.95 – 1.20 °C, from 2011 – 2020, a rapid increase compared to temperature changes at the start of the industrial revolution, 1850 -1900 (The Intergovernmental Panel on Climate Change, 2021). The result of this increase in temperature is evident in various climate effects including a rise in sea levels, increased intensity of heat waves and strong decline in Arctic Sea ice leading

to several environmental disasters such as floods and bushfires. Detailed analyses over the last five decades have implicated the increased concentrations of CO_2 as the major culprit in the overall warming of the earth's atmosphere (Hansen *et al.*, 2010). The refining of crude oil to products for manufacturing of plastics and also energy generation and for the petrochemical industry is firstly a polluting activity itself (Hazardous Substance Research Centers/South & Southwest Outreach Program, 2003) as is the creation of new oil fields (often in pristine environments as sources become scarce) and finally burning of refined oil fossil fuel derivatives such as naphtha, petroleum, diesel, jet fuel, kerosene and other oils adds GHG to the atmosphere, particularly carbon dioxide (CO_2). Considering this, plant biomass has received a lot of attention as an alternative to fossil fuels for bioenergy and biobased product formation in the short-medium term while renewable technology sources such as solar, wind and geothermal are developed for longer-term applications. This study focuses on biomass an alternative feedstock and looks at bioconversion techniques in contrast to the use of harsh chemicals, and further investigate bioconversion as a solution to address sustainability challenges associated with chemical pretreatment of biomass.

1.2 Plant biomass (Lignocellulosic biomass) as an alternative to fossil fuels

Plant biomass or Lignocellulosic biomass (LCB) is the most abundant organic mass on earth occurring in a variety of forms with common examples such as corn and wheat stovers, sugarcane bagasse, maize straw etc. These examples also represent the tons of agricultural waste generated globally (Zhou *et al.*, 2014; Väisänen *et al.*, 2016). Renewable resources such as LCB, has attributes of near-to-zero carbon dioxide (CO_2) emissions and low-Sulphur content and has been described as the most promising feedstock to drive a gradual shift towards sustainable development (Goldemberg, 2007). As opposed to the use of fossil fuels as a feedstock to generate high value chemicals, plant biomass conversion can offer a short-medium term driver for the generation of "green chemicals" such as ethanol from the fermentation of sugars

derived from breaking down plant biomass, thus achieving a sustainable outcome. An example is illustrated in Figure 1.1, where net zero is described and as a state where the greenhouse gas emission reaching the atmosphere are balanced by removal from the atmosphere (International Energy Agency, 2021; Net Zero Climate, 2021).

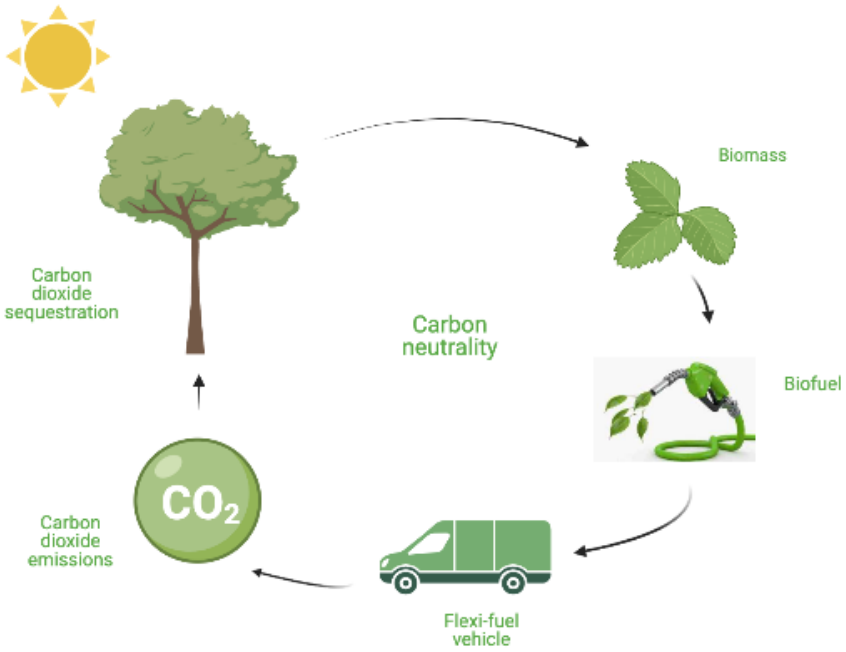


Figure 1.1. Illustration showing carbon dioxide capture and utilisation as an example of net zero emissions/carbon neutrality.

In Figure 1.1, Net Zero is illustrated and seen to involve a cycle which involves a sustainable approach to breakdown biomass which can then can be used in the production of high value chemicals such as bioethanol used in flexi fuel vehicles where the released CO_2 can be taken up by biomass, therefore continuing the cycle.

1.3 Structure of plant biomass

LCB is predominantly composed of the polysaccharides cellulose, hemicellulose, pectin and lignin which are the basic make-up of the plant cell wall superstructure

providing cell strength and structure (Guerriero *et al.*, 2016). Illustrated in Figure 1.2 is a diagram showing the various structures found in plant biomass.

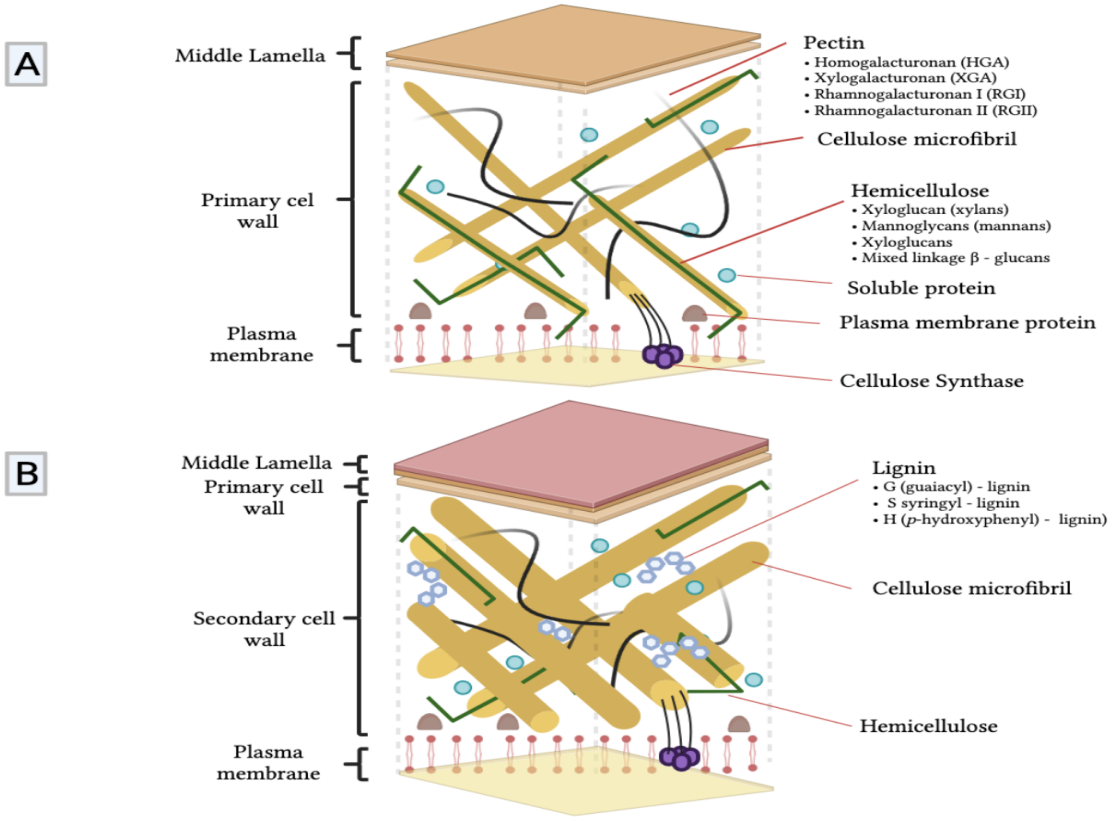


Figure 1.2 Overview of the primary and secondary plant cell wall

In Figure 1.2 (A) Primary cell wall (B) Secondary cell wall. Representation of the primary and secondary plant cell components showing various polysaccharide decorations (pectin, cellulose and hemicellulose) and aromatic component (lignin) which make up lignocellulosic biomass (Loix *et al.*, 2017).

1.3.1 Pectin

Pectin is said to be the most structurally and functionally complex of the three polysaccharides, making up approximately 2-10% of grass, 35% of primary walls in non-woody stem monocots and up to 5% of walls in woody tissue (Ridley, O'Neill and Mohnen, 2001). According to Mohnen (2008), pectin is made up of a complex mixture of galacturonic acid-rich (GalA) heteropolysaccharides including;

homogalacturonan (HG), rhamnogalacturonan I (RG-I), the substituted rhamnogalacturonan II (RG-II), apiogalacturonan (AG) and xylogalacturonan (XGA).

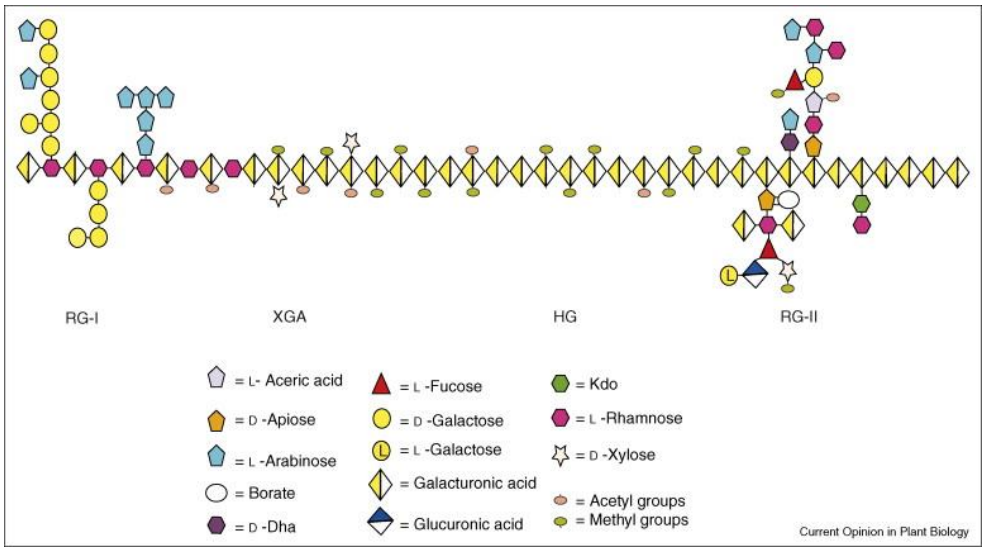


Figure 1.3. Structure of pectin showing galacturonic acid-rich (GalA) heteropolysaccharides (Mohnen, 2008).

1.3.2 Hemicellulose

Hemicellulose is non-crystalline in nature, the second most abundant component of plant biomass after cellulose and is considered to be a **potential substrate for the production of liquid fuels** and other value-added materials (Sun *et al.*, 2012). Hemicellulose represents an immense renewable resource of biopolymers with several variety of beta-linked sugar backbones and acetyl group decorations. Hemicellulose occurs in a large variety and can be divided into the general groups; xylan with a beta-1,4-D-xyl sugar backbone; mannan with b-1,4-D-Man sugar backbone and xyloglucans with a beta-1, 4-D-Glc sugar backbone. The last group represents randomly dispersed beta-1,4-Glucans and beta-1,4-Manan sugars collectively referred to as Glucomannan (Figure 1.7).

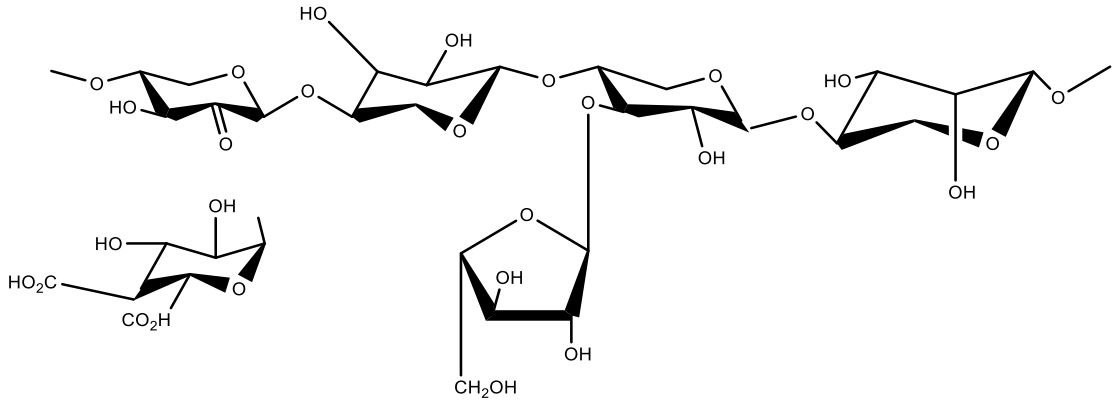


Figure 1.4. Structure of hemicellulose showing the various backbone and side chain decorations

1.3.3 Cellulose

Cellulose is the most abundant component of biomass and is considerably crystalline in nature in contrast to hemicellulose (Fontes and Gilbert, 2010). It is made up of individual glucose units which are connected via glycosidic bonds between their carbon 1 (C1) and carbon 4 (C4) positions, and the anomeric C1 carbon adopts the β-configuration (McNamara, Morgan and Zimmer, 2015). Cellulose is made up of rigid, insoluble microfibrils which has been attributed to cellulose chains having numerous intra- and intermolecular hydrogen bonds (Béguin and Aubert, 1994).

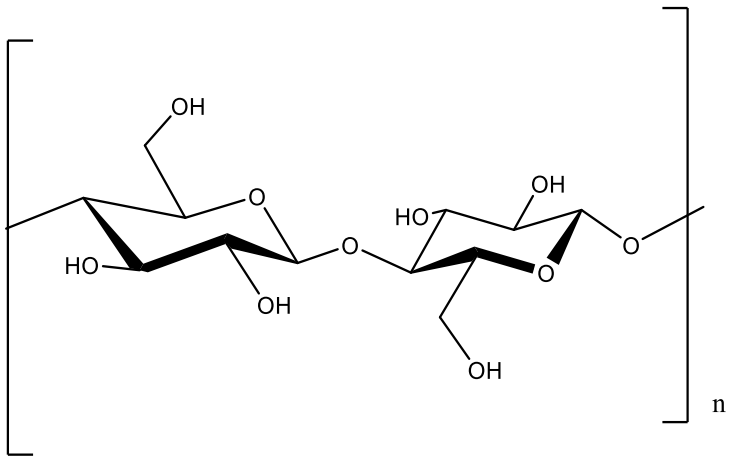


Figure 1.5. Structure of cellulose

1.3.4 Lignin

The lignin fraction of plant biomass is considered to have high carbon content and high aromaticity (Boerjan, Ralph and Baucher, 2003). It is a highly heterogeneous polymer made up of 4-hydroxyphenylpropanoid units and is embedded within the plant cell wall superstructure (de Gonzalo *et al.*, 2016). Lignin content varies depending on the source of lignin with percentages averaging (21 – 32 %) in plants such as poplar, eucalyptus and pine (Ragauskas *et al.*, 2014). Due to the various linkages and structural complexity of lignin, it could serve as an ideal renewable feedstock candidate for aromatic commodity chemicals.

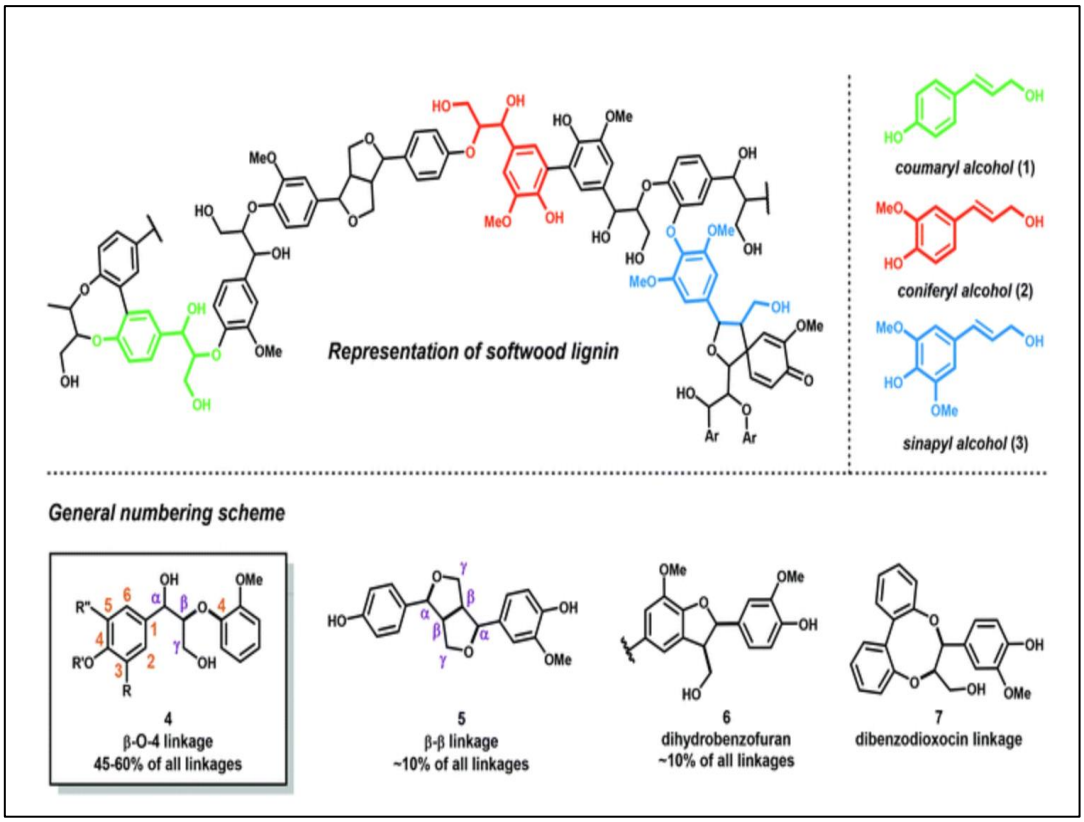


Figure 1.6 Structure of lignin and the main linkage motifs present in lignin.

In Figure 1.6 is a representation of lignin phenylpropane/monolignol units indicating various decorations of lignin with various linkages. Source (Kärkäs, 2017; Kärkäs, Matsuura, Monos, Magallanes, & Stephenson, 2016). Lignin forms ester linkages with hemicellulose, thus creating a difficulty in easy accessibility of hemicellulose and cellulose. Monolignols or phenylpropane monomers of lignin, include: guaiacyl (G, coniferyl alcohol), syringyl (S, sinapyl alcohol), which are considered to be methoxylated to various degrees and p-hydroxyphenyl (H, p-coumaryl alcohol) (Levasseur *et al.*, 2013). Lignin is a barrier to plant biomass processing and is released from cellulosic fibres as an

industrial waste of the pulp and paper industry and represents low value (Rinaldi *et al.*, 2016). Therefore, converting lignin into high value products is of high importance to chemical industries including the pulp and paper, biopharmaceuticals, and biofuel industries.

1.4 Current biomass conversion techniques

The production of bioethanol and bio-based chemicals from biomass is a promising alternative to address issues with petrochemicals derived from fossil fuels. Major conversion techniques employed technologically rely on traditional methods which involve several pre-treatment methods such as physical, physico-chemical, chemical, alkaline hydrolysis, oxidative delignification, organosolv and biological treatments (Sun and Cheng, 2002). However, pre-treatment methods are highly energy intensive and require the use of harsh chemicals to delignify and remove hemicellulose from biomass feedstocks reducing cellulose crystallinity and increasing porosity (Akin *et al.*, 1995; Sun and Cheng, 2002). This renders pre-treatment as a method for energy and chemical production economically non-viable (Chaturvedi and Verma, 2013; Rinaldi *et al.*, 2016). The main aim of pre-treatment is to improve sugar formation without degradation of carbohydrate and limit/ exclude the production of inhibitory by products that could subsequently affect hydrolysis and fermentation. If this is not achieved during the pre-treatment process, then the method is considered not to be cost effective (Chaturvedi and Verma, 2013).

1.4.1 Physical Pre-treatment

Physical treatment of LCB has existed for years and includes methods such as, chipping, grinding and milling, carried out initially during the treatment process of biomass to reduce cellulose crystallinity (McMillan, 1992). Physical treatment has been used in conjunction with pre-treatment methods as a secondary step (milling or grinding) to reduce the particle size of chipped biomass. Milling processes, particularly, vibratory ball milling and ball milling are known to increases the external surface area of biomass and increase cellulose reactivity and reduce cellulose

crystallinity resulting in complete digestion of milled biomass (Millett, Baker and Satter, 1976; Fan, Lee and Gharpuray, 1982).

1.4.2 Chemical pre-treatment

Various chemical pre-treatment methods have been used in the pre-treatment of biomass, particularly for delignification. Ozonolysis uses pressure and trioxygen (ozone) to effectively remove lignin with no toxic residues produced making it applicable for downstream process. However, large amounts of ozone are required rendering the process expensive (Vidal and Molinier, 1989). Acid hydrolysis using chemicals such as sulphuric acid (H_2SO_4) and hydrochloric acid (HCl) (Esteghlalian *et al.*, 1997); alkaline hydrolysis using dilute sodium hydroxide (NaOH) (Fan, Gharpuray and Lee, 1987); oxidative delignification using peroxidase enzyme in the presence of hydrogen peroxide (H_2O_2) (Azzam, 1989) and the organosolv process using methanol, ethanol, acetone, oxalic acid (Chum *et al.*, 1988; Ragauskas *et al.*, 2014) have also been used as pre-treatment methods. Yet, these chemical methods are not cost effective and to reduce costs and move toward a Green Chemistry approach to processing, biological pre-treatment is favoured.

1.4.3 Biological pre-treatment

Biological pre-treatment is mostly associated with the use of fungi (soft-rot, brown and white) to breakdown lignin and hemicellulose in waste materials (Jungschaffer, Schurz and Taufratzhofer, 1983). The use of fungi in the degradation of lignin and hemicellulose represents the success of microorganisms as commercial lignin and hemicellulose degraders (Davis and Perkins, 2002). According to Blanchette (1991), there are several other enzymes such as; polyphenol oxidases, laccases, H_2O_2 producing enzymes and quinone-reducing enzymes that also have ability to degrade biomass, especially the lignin superstructure creating accessibility to the sugar rich hemicellulose and cellulose fractions. The question however is, can enzyme hydrolysis compete with chemical hydrolysis?

1.4.4 Enzymatic hydrolysis vs chemical hydrolysis

The ability of for example, cellulase enzymes to carry out enzymatic hydrolysis has been studied over considerable span of time dating back to (Béguin and Aubert, 1994). Enzymatic hydrolysis has been proven to have low utility cost compared to acid or alkaline hydrolysis. This is because of the possibility of achieving enzymatic hydrolysis under milder conditions (for example, pH 4.8 and temperatures between 45-50°C) in contrast to chemical hydrolysis which require higher temperatures (pyrolysis needing ~300°C). Achieving hydrolysis under milder conditions is advantageous as the corrosion of bioreactor components can be minimised or avoided (Duff and Murray, 1996). To date, fungal enzymes used in ethanol production from lignocellulosic materials results in moderate yield based on current technologies, but the main challenges are seen with yield improvement and addressing the high cost of enzymes. Therefore, enzymatic hydrolysis as a method for energy production is rendered commercially unviable against the cost of energy production using chemicals from fossil fuel (Sun and Cheng, 2002). Notwithstanding, there is great impetus to search for novel Carbohydrate-Active enzymes (CAZymes) which will improve the commercial viability against the cost of green chemicals for energy production. With current plant biomass limited by cost and the availability of efficient enzymes that can be adapted to industrial scale processes, the search for efficient biological catalysts has intensified (Chaturvedi and Verma, 2013). To address this, several studies have focused on **enzymatic hydrolysis of biomass** as opposed to the use of traditional methods in the production of liquid biofuel for transportation and biobased chemicals for bioproduct manufacturing. In the next sections an overview of the various enzymes involved in the deconstruction of biomass is provided.

1.5 Enzymes for biomass conversion

1.5.1 Pectin degrading enzymes

Degradation of pectin components including homogalacturonan (HG), rhamnogalacturonan I (RG-I), the substituted rhamnogalacturonan II (RG-II), apiogalacturonan (AG) and xylogalacturonan (XGA), involves pectinases which catalyse random hydrolysis of 1-4-α-D-galactosiduronic linkages in pectin and other galacturonans. These enzymes operate by a common mechanism which cleaves glycosidic linkages between two neighbouring galacturonic acid monosaccharides (Abbott and Boraston, 2008). According to Hsiao *et al.*, (2008), enzymes involved in the degradation of HG include the pectin methyl esterases which de-esterify pectin making it accessible for enzymes such as; polygalacturonases (by hydrolytic cleavage) and lyases (by transelimination) to achieve degradation of polygalacturonic acid (PGA). An example of the use of pectin enzymes is seen in the hydrolysis of α-L-rhamnosyl-linkages in L-rhamnosides which include flavonoid glycosides such as naringin, rutin, quercitrin, hesperidin, dioscin, terpenyl glycosides and many other natural glycosides containing terminal α-L-rhamnose (Young, Johnston and Richards, 1989; Mutter *et al.*, 1994).

1.5.2 Hemicellulose degrading enzymes

Due to the large variety of sugars and acetyl group decorations (See Figure 1.4), several enzymes are required to achieve complete degradation of hemicellulose chains. The xylan (sugar backbone of beta-1,4-D-xyl), is majorly acted upon by acetyl xylan esterase which remove acetyl groups; glucuronidases cleave beta-D glucuronic acid residues; endoxylanases cleave the β-1,4 glycosidic linkage in the xylan backbone; arabinofuranosidases remove arabinose side chains; beta-xylosidases releases xylose from xylo-oligosaccharides and feruloyl and ferulic acid esterases remove ferulic acid residues from the xylan side chains (Dodd and Cann, 2009; Sun *et al.*, 2012). Arabinans are side chains of the complex rhamnogalacturonan I found in the hemicellulosic part of dicotyledonous cell walls also called pectin which are acted on by arabinases

(Figure 1.3). Galactomannan on the other hand is made up of polysaccharides of a mannose backbone with galactose side groups i.e., 1-6-linked alpha-D-galactopyranose and are acted on by alpha galactosidase. An example of the use of hemicellulase enzymes is seen with Acetyl xylan esterase (EC 3.1.1.72) which facilitates the action of endoxylanases by increasing accessibility to the xylan backbone by cleaving the ester bonds of acetyl groups (Sundberg and Poutanen, 1991).

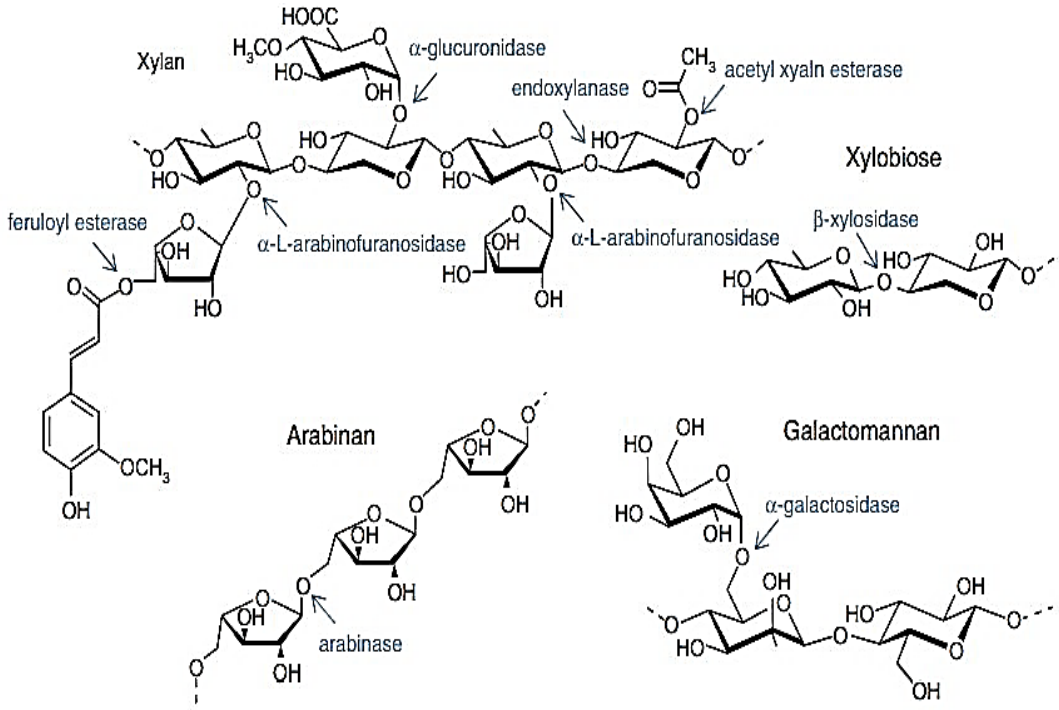


Figure 1.7 Schematic outline showing the functional coordination of hemicellulose degradation enzymes (Sun *et al.*, 2012).

Illustrated in Figure 1.7, shows hemicellulose fractions with the different substrates for hemicellulose degrading enzymes. Hemicellulose decorations mainly consists of sugars and acetyl group decorations which can be targeted by hemicellulose degrading enzymes to achieve complete degradation of hemicellulose fractions (Sun *et al.*, 2012).

1.5.3 Cellulose degrading enzymes

Cellulose breakdown involves various enzymes that lead to the final monomer product, glucose, a simple sugar with several uses to the chemical industry including sugar fermentation to bio-alcohol as a fuel blend. Cellulose degrading enzymes are

believed to work in synergy with proposed mechanisms of action involving the action of endo-acting cellulases (endo-beta-1, 4 glucanase) which create new chain ends leading to the subsequent release of cellobiose by exo-acting cellulases (cellobiohyrolases) on the non-reducing or reducing ends (Gilbert and Hazlewood, 1993; Gilbert, 2010). Cellobiose is then finally converted to glucose by beta-glucosidase. Industrial processing of cellulose is now reasonably efficient and relies on a variety of fungal and bacterial cellulases (Slaytor, 1992; Alessi *et al.*, 2018). See Figure 1.8 for a schematic representation of the conversion of cellulose into simple sugars.

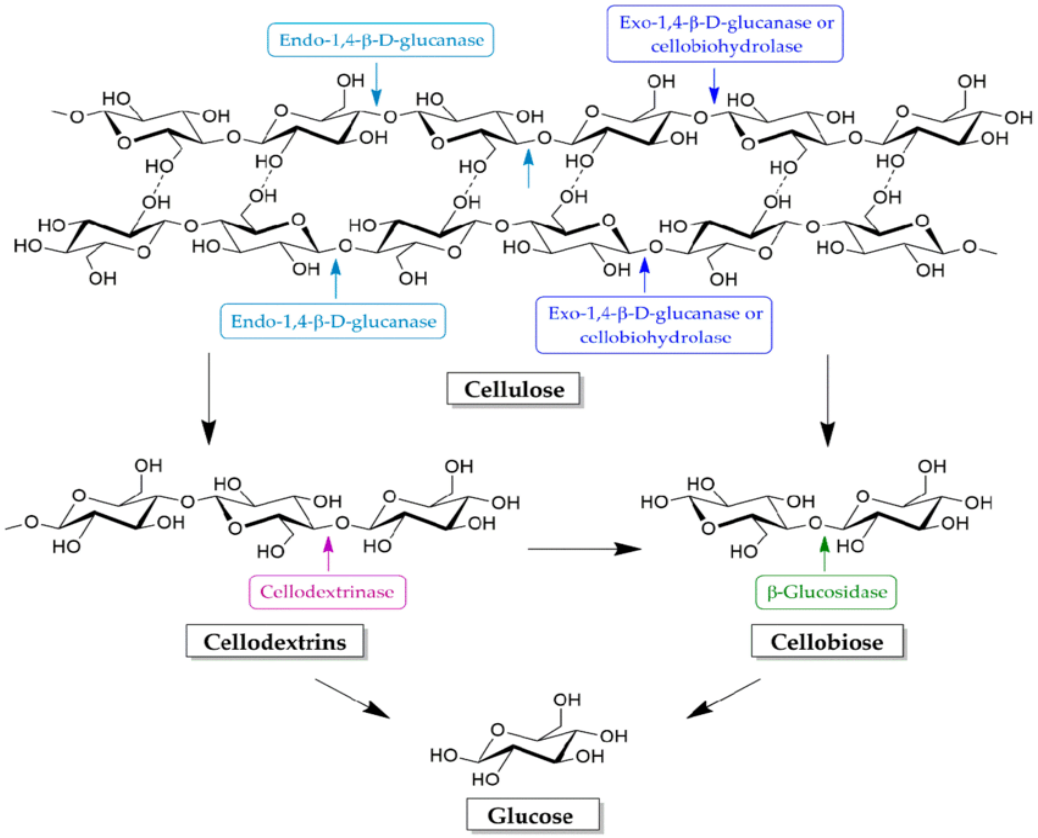


Figure 1.8 Enzymes involved in cellulose breakdown (Kumla *et al.*, 2020)

Illustrated in Figure 1.8 are the different cellulose degrading enzymes and the various target bonds that can be targeted to achieve cellulose breakdown into simpler sugar forms such as glucose which has several biotechnological applications including the fermentation to produce Bioethanol.

1.5.4 Lignin degrading enzymes

Several methods towards the improvement of lignocellulosic biomass processing, particularly focus on delignification of lignin from the biomass to enhance digestibility. So far, most lignin enzymes that are commercially available are derived from fungal sources isolated from more intensively studied environments such as soil (Davis and Perkins, 2002). Despite the resistance of lignin to degradation, nature has provided some microorganisms such as white rot fungi that can degrade lignin, although at a very slow pace (Aust, 1995; Janusz *et al.*, 2017). According to Sigoillot *et al.*, (2012), fungi are more efficient than bacteria in the breakdown of lignin. Extracellular synthesis of lignin peroxidase, manganese peroxidase, and laccase, the three primary enzymes involved in ligninolytic breakdown, is possible in these fungi (Singh Arora and Kumar Sharma, 2010). Another group of enzymes associated with delignification is the DyP-type peroxidases which belong to a novel superfamily of heme-containing peroxidases found in some fungi, bacteria as well as Archaea with functions not entirely understood. It is believed that these enzymes have oxidative activity and also hydrolytic activity making them bi-functional (Colpa, Fraaije and Van Bloois, 2014). The other type of lignin enzymes, Laccases (EC 1.10.3.2, p-diphenol: dioxygen oxidoreductases) are multi-copper proteins have been described to be involved in catalysing the reduction of oxygen to water followed by the oxidation of a substrate, typically a p-dihydroxy phenol or another phenolic compound (Solomon, Sundaram and Machonkin, 1996). There is a huge potential in the development of biological catalysts using DyP-type peroxidases and laccases for industrial applications. Since these enzymes have broad substrate specificity, broad range of compounds can be generated from their activity which can serve various applications. Some of these potential applications include; bioremediation of dye contaminated waste water, degradation of lignocellulose into various aromatic monomers, degradation of beta carotene for the food industry and even novel anti-microbial (pro) drug targets (Colpa, Fraaije and Van Bloois, 2014)

1.6 Production of Green chemicals for bioenergy and biobased products

Having described the enzymes needed for enzymatic hydrolysis of the various components of biomass, it is important to also understand the usefulness of the various sugars and aromatic monomers (Green chemicals) which can be produced as a result of these enzymatic reactions. The generation of energy and products in a sustainable approach presents challenges as well as substantial business opportunities targeted towards economic growth and creating jobs. Biologically derived products and biofuels represent billions in annual revenue with sales continuously increasing. According to a study prepared for the House of Commons Committee on exiting the European Union by the Office for Statistics (ONS), the UK chemicals sector contributed £12.1 billion to the UK economy's gross value added (GVA) and supported 99,000 employment opportunities (ONS, 2017). Although this might seem good for GVA, the major chemical contributors were petrochemicals, polymers, basic inorganics as well consumer and specialty chemicals. This means that to realise the growth of the economy, more petrochemicals are required, and this means more greenhouse gas emissions rendering the process unsustainable.

In response, LB Net-BBSRC – NIBB has carefully compiled a list of bio-based chemicals which can replace and improve upon petrochemicals and push growth of the bioeconomy further, whiles achieving sustainability (The Lignocellulosic Biorefinery Network and Biotechnology and Biological Science Research Council Network in Industrial Biotechnology and Bioenergy, 2019). Bio-based chemicals according to the European committee for standardisation (CEN, 2012), may be referred to as products generated partly or wholly from renewable sources including plants or animal-based feedstocks such as sugar, starch oils, fats and **biomass from crops and organic waste.**

Currently, Industries such as pulp and paper, textile, biomaterials, pharmaceuticals, food, cosmetics, and bioenergy are heavily dependent on harsh chemicals to

manufacture products. However, consumer product companies perspectives are evolving, and environmental pressures means that such industries need to shift focus towards the production of green chemicals (The Lignocellulosic Biorefinery Network and Biotechnology and Biological Science Research Council Network in Industrial Biotechnology and Bioenergy, 2019). For example, the pulp and paper business, which is dominated by North American, Northern European, and East Asian countries, earns billions of dollars in income each year and as such will benefit from such a shift to more sustainable processes leading to product generation (European Commission, 2013). The growth of such industries has been attributed to improved efficiencies in the process of product generation where more than half of its primary energy is generated from biorefining, a sustainable process involving the conversion of biomass into a broad spectrum of marketable bio-based products (biochemicals, materials) and bioenergy (fuels, power, heat), making the sector more energy self-sufficient and less CO_2-intensive. This is just one example of how green chemicals can contribute to environmental sustainability and as a result placing an impetus on generating and deploying more green chemicals.

1.6.1 Deployment of Green chemicals

1.6.1.1 Biofuels

To highlight further the potential use of lignocellulose biomass to generate chemicals, Celtic renewables has demonstrated the successful test drive of the world's first biofuel car running on a biofuel mix made from whisky residue (Celtic Renewables, 2021). Celtic renewables attributed the key behind this breakthrough to a bacterium from the *Clostridia* class, which could convert sugars into acetone, n-butanol, and ethanol (ABE). The ABE fermentation process has existed since 1861 when microbial fermentation to produce butanol was realised. The surge in commercial interest was triggered by the need for synthetic rubber in the 20th century (Jones and Woods, 1986). Industrial solvent production process were realised following the introduction of *Clostridium acetobutylicum* by Weizmann, and in 1915, a patent was issued for ABE

fermentation with *C. acetobutylicum* (Sauer, 2016). Acetone demand surged during the outbreak of the world war for weaponization of smokeless powder leaving butanol as a by-product (Sauer, 2016). However, the focus has shifted over the years into generating more high value products rather than its use in weaponisation. Nevertheless, significant investment is necessary to overcome processing scale issues such as toxicity of n-butanol to the culture, which potentially leads to nutrient depletion during extended fermentation times (Kujawska *et al.*, 2015). As an example, this demonstrates how bioconversions can be applied in a sustainable approach to generate base molecules for energy generation. The biggest challenge however, is the cost competitiveness of bioprocesses against chemical processes, largely due to high capital assets associated with the existing commercial processes (Thomas, DiCosimo and Nagarajan, 2002).

1.6.1.2 Biobased products

The utilization of lignocellulosic biomass in the production of chemicals is not new. Several studies have been conducted looking at the biochemical conversion of biomass into chemicals that can serve as basis for industrial process such as manufacturing of plastics, biopharmaceuticals and cosmetics. These chemicals are said to be sustainable alternatives to petrochemicals, and they can be used in several applications. Examples of such chemicals include lactic acid, levoglucosenone, muconic acid, itaconic acid and levulinic acid (Zhang and Vadlani, 2015; Bruyn, 2016; Regestein *et al.*, 2018). Several such chemicals have been indicated by E4techs report as part of the top 10 chemicals that will push the growth of the UK bio-economy (Biomass Biorefinery Netweork, 2017; E4tech, 2017).

Table 1.1 The top 10 green chemicals to drive growth in UK bioeconomy. Source (Biomass Biorefinery Netweork, 2017)

01	**Lactic Acid is** used to make Polylactic Acid (PLA) that can be used to manufacture biodegradable polymers. Other applications in the cosmetics, food and pharmaceutical industries (Yadav, Chaudhari and Kothari, 2011). According reports from Zion market research (2019), lactic acid global and regional market was valued at $2.9bn and is expected to increase by 5-fold by 2025.
02	**2,5-furandicarboxylic acid (FDCA) is** a robust alternative to PET that is used to create plastic bottles, food packaging and carpets. According to a report by FiorMarkets (2017), FDCA was valued at US$ 260 million in 2017 with prospects of a 7% compound annual growth rate (CAGR).
03	**Levoglucosenone (LGO) is** a safer alternative to hazardous solvents used in pharmaceutical manufacture, as well as a flavouring and fragrance ingredient. (E4tech, 2017)
04	**5-Hydroxymethyl furfural (HMF)**: is a versatile chemical that has the potential to replace compounds used in plastics and polyesters as well as to produce high-energy biofuel. According to a Reuters market report by (Trent, 2018), 5-HMF was valued at $120 million in 2017 with expectations of increase to ~ $145 million by 2022.
05	**Muconic acid** derivatives have a substantial market potential of US$22 billion and might replace non-sustainable chemicals used in the production of PET and nylon fibres. (E4tech, 2017)
06	**Itaconic acid** is a petroleum-free alternative to acrylic acid, which is used to manufacture absorbent materials and resins for high-performance marine and automotive parts. According to market reports by Markets and Markets (2016), the itaconic market was valued at $80.8 million and is expected to grow over time.
07	**(1,3-BDO)** Pheromones, scents, pesticides, antibiotics, and synthetic rubber are all made with 1,3-butanediol as a building block. The focus is now on *Clostridium* fermentation technology, an area in which the UK excels, putting the country in a position to develop an integrated supply chain for 1,3-BDO synthesis for speciality product applications. (E4tech, 2017).
08	**Glucaric acid**: Prevents limescale and grime build-up on fabrics and dishes, making it a green alternative to phosphate-based detergents. The use of glucaric acid instead of phosphate-based reagents is a great step forward for the UK in gaining access to a multibillion-dollar global industry with substantial uses in nylon fibres and engineering polymers, polyester polyols for polyurethanes, and adipate esters for phthalate-free plasticizers. (E4tech, 2017).
09	**Levulinic acid:** Environmentally friendly herbicides, fruity flavour and perfumes and fragrance components, skin treatments, and degreasers all contain levulinic acid (E4tech, 2017)).

10	**N-butanol** is utilised in a variety of polymers and plastics, as well as in chemical and textile processes as a solvent and as a paint thinner. It has a well-established method of production via ABE fermentation and can serve as an alternative oxygenate to ethanol in blended fuels (Kolesinska *et al.*, 2019).

According to E4 tech and BBNet, these top 10 chemicals will greatly contribute to the UK bioeconomy growth due to their applications. These chemicals will serve as biobased chemicals to replace the increased demands for petrochemicals. But we must find enzymes which can achieve bioconversion of biomass feedstocks to generate bioenergy and such platform chemicals.

1.6.2 The role of microbial enzymes in green chemical production

The biorefining process of converting LCB remains economically unattractive due to an absence of biocatalysts that can mitigate cost challenges such as; cooling from high temperatures, energy burnt in the pumping or stirring of oxygen, and neutralization from acidic or basic pH (Maki, Leung and Qin, 2009). Some bacteria can resist extreme conditions (extremophiles) and hence screening, and isolation of such microbes as potential novel enzymes could address some of the cost challenges associated with the conversion of biomass into high value products. The potential identification and characterisation of novel glycoside hydrolases from eubacteria has received a wide level of attention since 2009 (Gallardo, Diaz and Pastor, 2003; L. L. Li *et al.*, 2009). The interest is due to a number of properties shared by bacteria, such as their exponential growth rate when compared to fungi, which allows for higher recombinant enzyme production in hosts such as *Escherichia coli* (Ferrer-Miralles and Villaverde, 2013). In addition, glycoside hydrolases produced from microbial sources are often expressed in multi-enzyme complexes providing increased functionality and synergy (Maki, Leung and Qin, 2009; Moraïs *et al.*, 2010). Notably, bacteria live in/on a wide variety of environmental and industrial niches, which produce cellulolytic strains that are thermophilic or cryophilic, alkaliphilic or acidophilic, and also strains that are halophilic (Doi, 2008; Alessi *et al.*, 2018). Such strains can endure extreme conditions

present during bioconversion process. If strains of bacteria can endure extreme conditions, then the rate of enzymatic hydrolysis, fermentation and product recovery will be highly improved and hence the interest in using bacteria for biofuel and biomaterial industry processing of biomass. Consequently, identification of prokaryotic enzymes from novel environmental sources using metagenomics has become a huge area of research focus due to the potential use of microbial enzymes in biodegradation of plant biomass and for the industrial production of biofuel and bio-based products.

1.7 Molecular techniques to produce microbial enzymes for green chemical production

1.7.1 Polymerase chain reaction

Prior to the cloning of target sequence and recombinant protein expression , metagenomic DNA obtained from environmental sources can be initially subjected to whole genome amplification as a crucial method of preserving the precious starting material. Whole genome amplification (WGA) is a global amplification of all DNA within a sample. Whole genome amplification offers a robust method of amplification with nanogram start materials resulting in microgram quantities allowing for downstream applications in areas such as Next generation sequencing (Babayan *et al.*, 2017), microarray (Govindarajan *et al.*, 2012), and PCR applications (Wang *et al.*, 2022). Various studies have employed WGA in the preservation of precious material from the gut of organisms such as the common black slug (Joynson *et al.*, 2017). Identified putative glycosyl hydrolases through bioinformatics can be amplified using PCR (Mullis et al, from the whole genome amplified product using oligonucleotides targeted towards specific coding DNA sequences.

1.7.2 Cloning of amplicons

Prior to recombinant protein expression of enzymes, generated amplicons can be cloned into suitable hosts which allow the recombinant expression/production of

these enzymes. The production of recombinant proteins in microbial systems following cloning of amplified genes has undoubtedly contributed to advancements in biochemistry. DNA sequences which code for proteins identified after carrying out metagenomic analysis can be cloned into expression vectors and recombinantly expressed in host strains. Cloning technologies such as the Invitrogen Gateway technology offer time saving rapid cloning reactions without the use of restriction enzymes, ligases, subcloning steps or the need for screening countless colonies. Researchers can achieve 95% cloning efficiency in as little as 5 minutes with correctly orientated fragments, making this technology one of the most successful cloning methods used in scientific studies (ThermoFisher Scientific, 2022). Gateway cloning technology used in this study involved the use of an entry vector and a destination vector. Target DNA sequences to be amplified for this method of cloning require specially designed forward primers where the 5′ end begins with CACC to facilitate entry into the vector. Positive clones can be selected at random for further by employing antibiotic resistance conferred by the antibiotic marker gene present in the sequence of the vector.

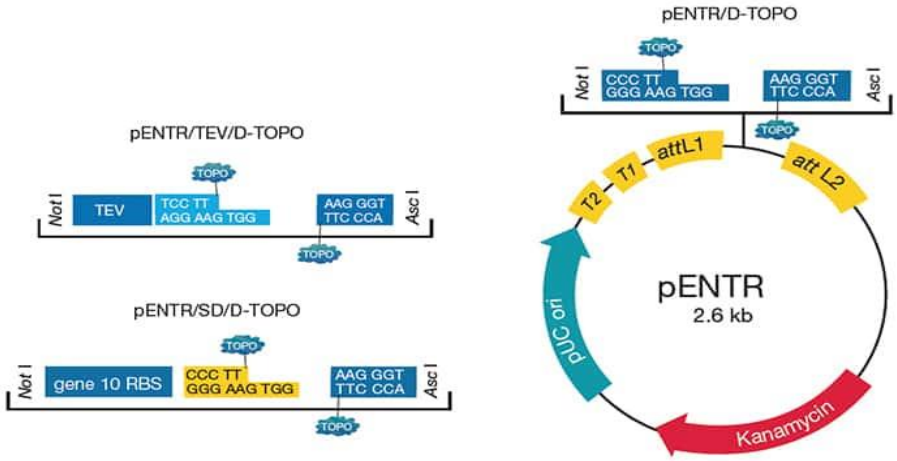


Figure 1.9 pENTR vectors available for Directional TOPO cloning and direct access to the multitude of Gateway expression vectors (Source: ThermoFisher Scientific).

In Figure 1.9 is an illustration of the various pENTR vectors available for Directional TOPO cloning. In this study the focus will be on pENTR/SD/D-TOPO as this was used to generate entry clones prior to generating expression clones for recombinant protein expression. The generated entry clones

containing the cloned gene of interest flanked by attL sequences, are then used to recombine with attR sequences found on the destination vector to create your desired expression clone.

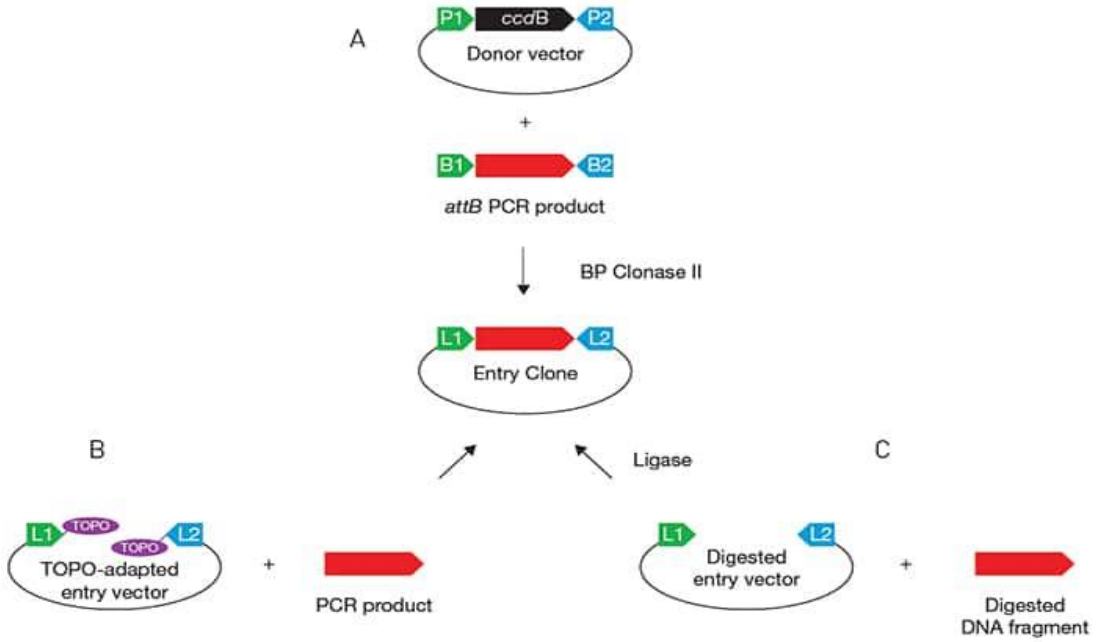


Figure 1.10 Strategies for the cloning of PCR products to generate expression clones (Source: ThermoFisher Scientific, 2022).

In Figure 1.10 is a schematic showing the strategies for the generation of expression clones through the gateway technology using an entry vector (B) and the destination vector.

1.8 Choice of expression systems

With new advances in the field of recombinant protein expression, huge amounts of animal and plant tissue extracts are no longer required to produce large amounts of a given protein. The possibility of producing large amounts of a given protein makes possible biochemical characterisation of a protein for use in several industrial applications (Rosano and Ceccarelli, 2014). Host systems from bacteria (*Escherichia coli*), yeast (*Saccharomyces cerevisiae* and *Pichia pastoris*), algae (*Phaeodactylum tricornutum*) and filamentous fungi (*Aspergillus spp*) are usually employed in the expression of recombinant proteins (Nevalainen, Te'o and Bergquist, 2005; Hempel *et*

al., 2011; Su *et al.*, 2012). Each host system has its advantages and disadvantages, and selection is subject to the application of the protein of interest. Expression systems such as mammalian, yeast, insect, bacterial, algal and cell free can be used in functional assays, structural analysis virus production, protein interactions and genetic engineering. Some of the advantages with using these systems include good scalability, low cost, simple culture conditions and high level of protein processing. However, challenges such as culture conditions, solubility of proteins, low yields of target protein, less developed technologies of some expression systems such as algal, are just a few disadvantages relating to expression systems. (Hempel *et al.*, 2011; Rosano and Ceccarelli, 2014; ThermoFisher Scientific, 2021).

Nevertheless, due to advantages with scalability, simple culture conditions and cost efficiency, prokaryotic expression systems such as *E.coli* systems have been employed in several studies for high-level production of heterologous proteins (Rosano and Ceccarelli, 2014). The use of *E. coli* host system for recombinant protein expression is considered one of the earlier and most widely used expression systems (Terpe, 2006). More than 40% of produced non-glycosylated proteins for the therapeutic market is usually produced in *E. coli* and yeast host strains. *E. coli* strains are considered less expensive, offering possibilities of manipulation to achieve rapid expression and high protein yields (Demain and Vaishnav, 2009). The production of recombinant proteins in host expression systems is theoretically easy and the steps as seen in Figure 1.11, depicts a summary of the methods which leads to the production and biochemical activity testing of proteins.

1.8.1 Why the need for protein purification and how?

Protein purification is usually performed to achieve a purity that is desirable for biochemical activity testing. Most protein purification methods are based on solubility, size, charge, and binding affinity (Berg *et al*., 2002; Janson, 2011). According to (Janson, 2011), the most widely used method of purification is the bio-affinity

method because it is considerably highly specific, occasionally resulting in highly pure proteins in a one-step purification process. Bio-specific affinities such as ligands, inhibitors and antibodies offer properties which are exploited in the isolation of target proteins from crude samples. Immobilised metal affinity chromatography (IMAC) is the most widely used binding affinity technique in research for single step purification of recombinant proteins, directed towards protein side chains, usually polyhistidine-tag (Porath, 1992; Saraswathy and Ramalingam, 2011). Purification, employing IMAC, has been described as rapid and easy allowing for single step purification achieving up to approximately 100 fold enrichment of target protein (Schmitt, Hess and Stunnenberg, 1993). Competitive interaction between imidazole and the metal-charged resins in either stepwise manner or gradient concentration of imidazole can be used to elute polyhistidine-tag proteins. After purification, proteins are usually subjected to dialysis as a last step prior to assay testing to desalt the protein and remove residual imidazole which can affect protein activity testing.

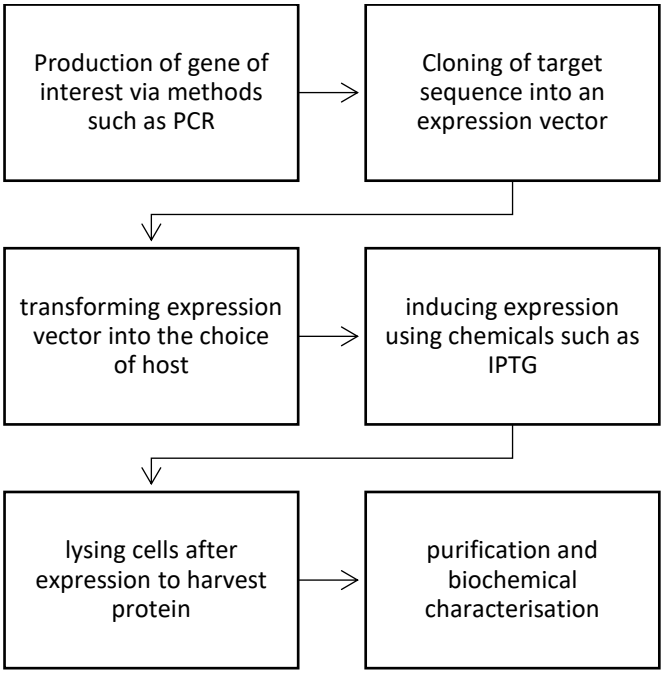


Figure 1.11 Schematic showing an example of pipeline for recombinant protein expression and characterisation.

Although the steps indicated above appear easy to follow, several challenges along the way can be expected between each recombinant protein expression study. Some of these challenges include, inclusion body formation (Singh and Panda, 2005), poor growth of host due to toxicity of the produced protein (Saida *et al.*, 2006), protein inactivity and even no protein expression observed. Addressing such challenges is on a case-by-case study and several solutions can be applied such as optimizing expression conditions such as (temperature, concentration of inducer), using strains of host which prevent basal expression to tackle toxicity, ensuring ORF with intact conserved domains are expressed and ensuring the sequence of interest is in frame with tag of choice during expression (Rosano and Ceccarelli, 2014).

1.9 Activity testing of recombinantly expressed proteins

Biochemical activity testing is carried out by reacting the protein of interest with a chosen substrate. Both natural and synthetic substrates have been used for the activity testing of glycosyl hydrolases. For example, *para*-nitrophenyl (4-nitrophenyl) linked substrates have been used in the determination of enzyme activity and described in several studies (Gallardo, Diaz and Pastor, 2003; Aspeborg *et al.*, 2012; Mewis *et al.*, 2016; Mueller *et al.*, 2018). Commercially available 4-nitrophenyl substrates with sugar moieties attached via glycosidic linkages can be purchased from manufacturers such as Sigma-Merck (UK) and Megazyme (Ireland). Enzymes specific to the sugar moiety can recognize the moiety as a substrate and the glycosidic bond broken releasing 4-nitrophenol. 4 nitrophenol is chromogenic compound observed as a yellow colour in solution and can be monitored at a wavelength between 405 - 410 nm. A standard curve for 4-nitrophenol can be constructed and the linear regression equation given as ($y = mx + c$) where the gradient (m) is the molar extinction coefficient can be used in determining the amount of nitrophenol released after activity.

Other natural substrates such as naringin or hesperidin (Izzo *et al.*, 2014) have also been used in enzyme activity testing studies of alpha rhamnosidase enzymes. For beta glucosidase enzymes, esculin and para nitrophenyl beta D-glucopyranoside has been described for activity testing (Kwon *et al.*, 1994). In another study, p-nitrophenyl acetate and xylan has been used as substrates for the activity testing of carboxyl esterases and endo 1,4-beta xylanase enzymes. In all these studies, activity testing of enzymes required a chemical reaction involving the conversion of substrate to product and determining how much product has been formed. This is known as enzyme kinetics and involves the determination of reaction rate and the investigation of the effects caused by varying the conditions of the reaction. Described in the next chapters, are the methods taken to generate amplimers, clone, express, purify, dialyse and test activity of the selected protein sequences from the metagenomic data obtained from sequencing the gut microbiota of *Arion ater*.

In the next sections we will be looking at the cloning, expression, purification and biochemical activity testing of putative glycosyl hydrolases identified from high throughput sequencing of the slug gut microbiota. For cloning, the Gateway cloning technology has been used to aid in rapid cloning of fragments achieving 95% cloning efficiency with correctly orientated fragments. For expression, our choice of host expression systems was guided by the ability of prokaryotic systems to be used in functional and structural analysis of proteins due to prokaryotic systems having good scalability, low cost, simple culture conditions and high level of protein processing. Purification of proteins was carried out using immobilised metal affinity chromatography to achieve single step purification of recombinant proteins. Finally, activity testing in our studies were performed following reactions between protein of interest and synthetic *para*-nitrophenyl (4-nitrophenyl) linked substrates. Enzyme functionality improvement was also tested employing enzyme immobilisation through nanoparticles and cross-linked enzyme aggregates.

Chapter 2. Metagenomic analysis of the gut microbiome of *Arion Ater* in search of novel enzymes for biocatalysis

Abstract

Microbial enzyme and natural product discovery pipelines were initially limited to culture dependent phenotypic screening that permitted description of a limited biological pool of organisms. Much of the remaining microbial genetic diversity (collectively termed the metagenome) thus remain unexploited. Under-explored microbial communities harbor genetic and metabolic diversity that can be actively exploited for biotechnological applications. Since the advent of next generation DNA sequencing technologies, improvements in metagenomic studies and function-based metagenomic screening in combination with advances in bioinformatics and databases have now allowed for functional screening of metagenomic libraries from under-studied environments, such as that from the gut of the black slug, *Arion ater.* Such functional studies have led to the identification of thousands of novel Carbohydrate-Active Enzymes among the wide variety of previously unexplored genetic diversity.

In this study, a metagenomic analysis approach was employed to re-assemble and re-annotate the slug gut metagenome using updated software to improve coding DNA sequences from that indicated in a previous study. Initial quality control using fastqc and trimmomatic resulted in 25795430 paired end reads with an average length of 239 bp. The total number of bases in the assembly was 209Mbp with assembled contigs having an N50 of 2.4 kbp and an L50 of 1.3 kbp. Taxonomic investigation identified bacteria as the predominant superfamily with 97% and 1% to virus, 0.7% Archaea, 0.06% *Homo sapiens*, 0.06% had no hits and 0.8% belonging to other. A final in-depth investigation into the family Enterobacteriaceae identified bacteria belonging to the genera *Enterobacter* with 21%, followed by *Citrobacter* with 17% and other notable families such as the *Klebsiella*/*Raoultella* group with 12%. Genome annotation

identified 245,605 contigs, 250Mbp, 222,541 coding DNA sequences and 92 clustered regularly interspaced short palindromic repeats (CRISPR). Our findings suggested improvement in the complete coding DNA sequences in contrast to the partial coding DNA sequences identified from that of a previous study on the same metagenomic data.

2.1 Introduction

The identification of enzymes that can act on glycosidic linkages, remove esters and act as binding modules for efficient enzymatic conversion of LCB into simple sugars and aromatic monomers will play a major part in future bio-refineries. The drive to make future bio-refineries cost efficient and sustainable has created a platform for intense research and has fuelled the expedition to find and utilise biomass degrading enzymes in contrast to the use of harsh chemicals. The majority of the commercial biomass degrading enzymes are predominantly derived from fungal communities found in environments such as soil (Patra, Das and Shin, 2018). Biomass degrading enzymes from Fungal communities have been the focus of extensive research for decades, but the focus is now shifting towards enzymes from prokaryotic communities for various reasons, including; underutilised biochemical diversity as well as more favourable characteristics for use in industry such as high exponential growth rates compared to fungi allowing for higher recombinant production of enzymes in host such as *E. coli* (Ferrer-Miralles and Villaverde, 2013).

The gut microbiota of vertebrates such as humans (Qin *et al.*, 2010; Flint *et al.*, 2012), ruminants (Stackebrandt and Hippe, 1986), and invertebrates such as termites (King *et al.*, 2010; Hess *et al.*, 2011; Brune, 2014; Franco Cairo *et al.*, 2016) and cockroaches (Bertino-Grimaldi *et al.*, 2013), have been studied and found to harbour a repertoire of microbes that assist in the breakdown of complex non-digestible carbohydrates. Also a large number of filamentous fungi including; *Neurospora crassa* found on burnt plant material are able to utilise cellulase and hemicellulases to achieve degradation of plant cell wall destruction (Davis and Perkins, 2002). Microbes present in the gut microbiota of herbivores, such as cows have also been described in having ability to achieve foregut fermentation (Brulc *et al.*, (2009). The rumen of herbivores has been studied as a provision of unique genetic resource for the discovery of plant cell wall-degrading microbial enzymes. However, the complex community of microbes in the rumen of herbivores, involved with degradation of cellulosic material are culture resistant (Hess

et al., 2011). As a result, not all microbial communities present in the gut of these herbivores are represented through culturing on selective media. This has prompted new methods for the investigation of microbial communities of the gut of organisms through metagenomic and bioinformatic approaches.

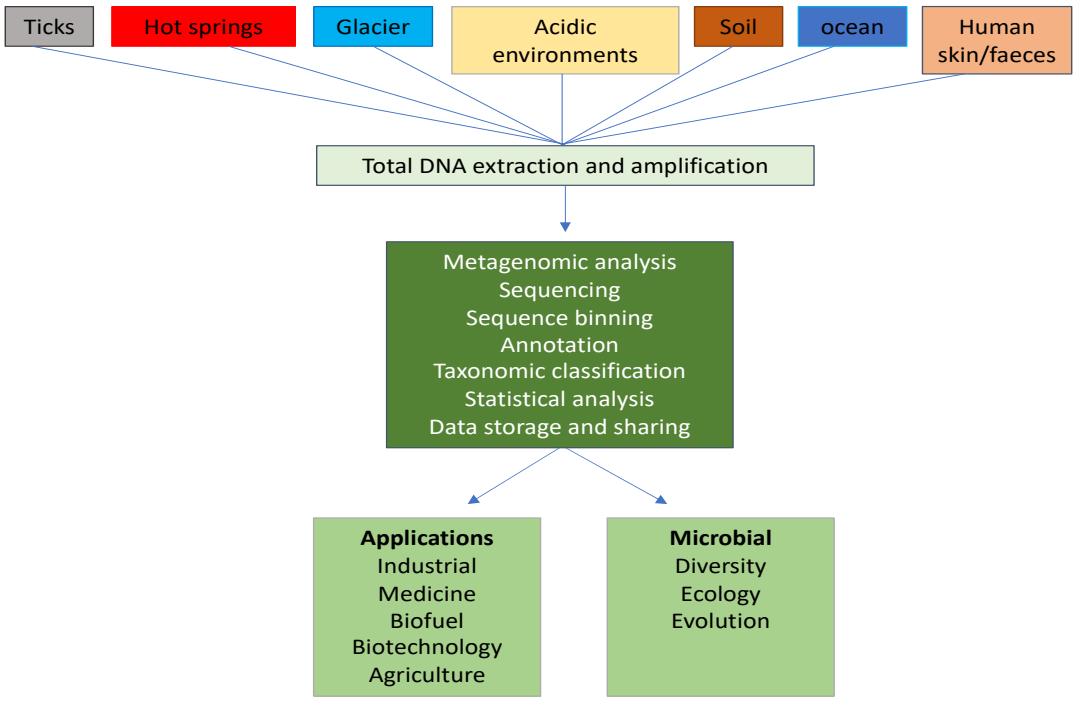


Figure 2.1 Metagenomic approach to identification of novel sequences (Neelakanta and Sultana, 2013).

A suite of aerobic and anaerobic biomass degraders can be found in a variety of ecosystems such as soil, seawater, insect and mollusc gut, compost and several other sources (Biddanda, 1988; Cardoso *et al.*, 2012; Joynson *et al.*, 2017; Patra, Das and Shin, 2018). Following identification, total DNA isolated from these ecosystems can be amplified and employing metagenomic tools and next generation sequencing and analysis, microbial communities can be profiled, and functional diversity can allow further applications. The increasing number of projects focusing on the symbiotic and mutualistic properties exhibited by microbiota from the digestive tracts of animals such as termites (*Nasutitermes sp.*), cows (*Bos taurus*), wallabies (*Macropus eugenii*), and giant pandas (*Ailuropoda melanoleuca*) demonstrates the importance of the search for

novel lignocellulose degrading enzymes and microorganism (Warnecke *et al.*, 2007; Brulc *et al.*, 2009; Pope *et al.*, 2010; Hess *et al.*, 2011).

2.1.1 Overview of successful biomass degradation by Gribbles and termites

Termites have been described as an extremely successful group of wood-degrading organisms especially as potential sources of identifying biochemical catalysts (Warnecke *et al.*, 2007). However, the exact roles played by host and symbiotic microbiota in the hindgut communities of wood feeding higher termites is poorly understood (Kambhampati and Eggleton, 2000). According to Tokuda and Watanabe (2007) the presence of intestinal microorganisms of termites led to the production of cellulases which contribute to the digestion of cellulose in host termites, suggesting that cellulose breakdown in termites is facilitated by microorganisms known as flagellates. Termites of the taxonomical order Isoptera, are known to have the ability to digest cellulose in different secretive digestive enzyme production sites including; salivary glands and predominantly in the foregut and midgut regions which were believed to be absent of microorganisms or have low amounts (Slaytor, 1992). However, Breznak and Brune (1994) provided evidence after the study by Slaytor and concluded on the presence of diverse population of microorganisms in the alimentary tract of termites. In a subsequent study, they further concluded that the degradation of lignocellulose relies on their partnership with a diverse community of bacterial, archaeal and eukaryotic symbionts (Brune, 2014). This multilateral symbiosis has led to termites being dubbed as the most successful plant decomposers according to da Costa *et al.*, (2018).

Wood boring crustacea in European waters belonging to the family *Limnoriidae*, including *Limnoria lignorum* (Borges, Merckelbach and Cragg, 2014), *Limnoria tripunctata* (Zachary and Colwell, 1979), commonly known as gribbles are known to be able to degrade wood. Limnoriids have the ability to bore into wood for protection (Menzies, 1957) and also to use wood as a primary carbon source by producing

endogenous enzymes that can digest lignocellulose (King *et al.*, 2010; Kern *et al.*, 2013). The digestive mechanism of lignocellulose and hemicelluloses digestion by Limnoriids is a subject of intensive research. The majority of this research focuses on understanding the basic process of metabolism carried out by these microorganisms, and also their use as potential candidates for plant biomass conversion into biofuels (King *et al.*, 2010). According to findings from a study carried out by El-Shanshoury *et al.*,(1994), the presence of cellulase activity and cellulases can be attributed to bacteria and fungi associated with isopods, their wood burrows and surrounding seawater. Further to this, research on enzymes from limnoriids has led to the identification of structural and functional properties of some rather unusual salt tolerant cellulase enzymes. It is hoped that the robust nature of such enzymes, which allows it to be compatible with sea water, will lower the costs of making biofuels (Kern *et al.*, 2013).

2.1.2 Microbial biodiversity and metagenomics

Microbial biodiversity is composed of interactions between many bacteria, archaea and viruses which make up mixed populations in the environment. Understanding these complex interactions between microorganisms remains a daunting task for microbiologist. For many years, morphology, growth conditions and selection profiles were the standard tools employed by microbiologist to understand mixed populations of microorganisms from the environment (Neelakanta and Sultana, 2013; Brune, 2014). Substantial contributions to our understanding of microbial classification have been made by the use of ribosomal RNA genes as molecular markers in tandem with Sanger methodology (Sanger, Nicklen and Coulson, 1977; Woese and Fox, 1977). Advances in molecular techniques now mean that PCR, cloning and sequencing of ribosomal ribonucleic acid (rRNA) genes can be developed to explore "novel uncultured world" of microorganisms (Mullis, 1990). Nevertheless, improvements in molecular techniques were not enough as metabolic and ecologic function assignment was not adequately represented via molecular techniques. To address this challenge, metagenomic tools have been introduced to discover new genes, functions, and metabolic products using high-throughput culture independent techniques, allowing

sampling of genes from diverse environments such as soil-DNA libraries (Handelsman, 2004).

Metagenomics, since its inception (Pace *et al.*, 1986), has emerged as a powerful tool in that, direct isolation of genomic DNA from an environment and cloning of genes into suitable vectors allows a possibility of transformation into suitable hosts for recombinant expression avoiding culturing limitations (Handelsman, 2004). Since its introduction, metagenomics has been used in several studies ranging from marine microorganisms (Schmidt, DeLong and Pace, 1991) to human intestinal microbiota (Maccaferri, Biagi and Brigidi, 2011). Metagenomic analysis allows the accurate taxonomic profiling and functional analysis (metaproteomic and metabolomics) of genomic data following a sequence run. However, the cost of analysis in relation to sample size during a metagenomic research, puts a limitation on metagenomics, and as such critical planning of metagenomic projects is essential in addressing the link **between phylogenetic and functional** rather than just technical research to determine the physical and biological characteristics of a sample's environment.

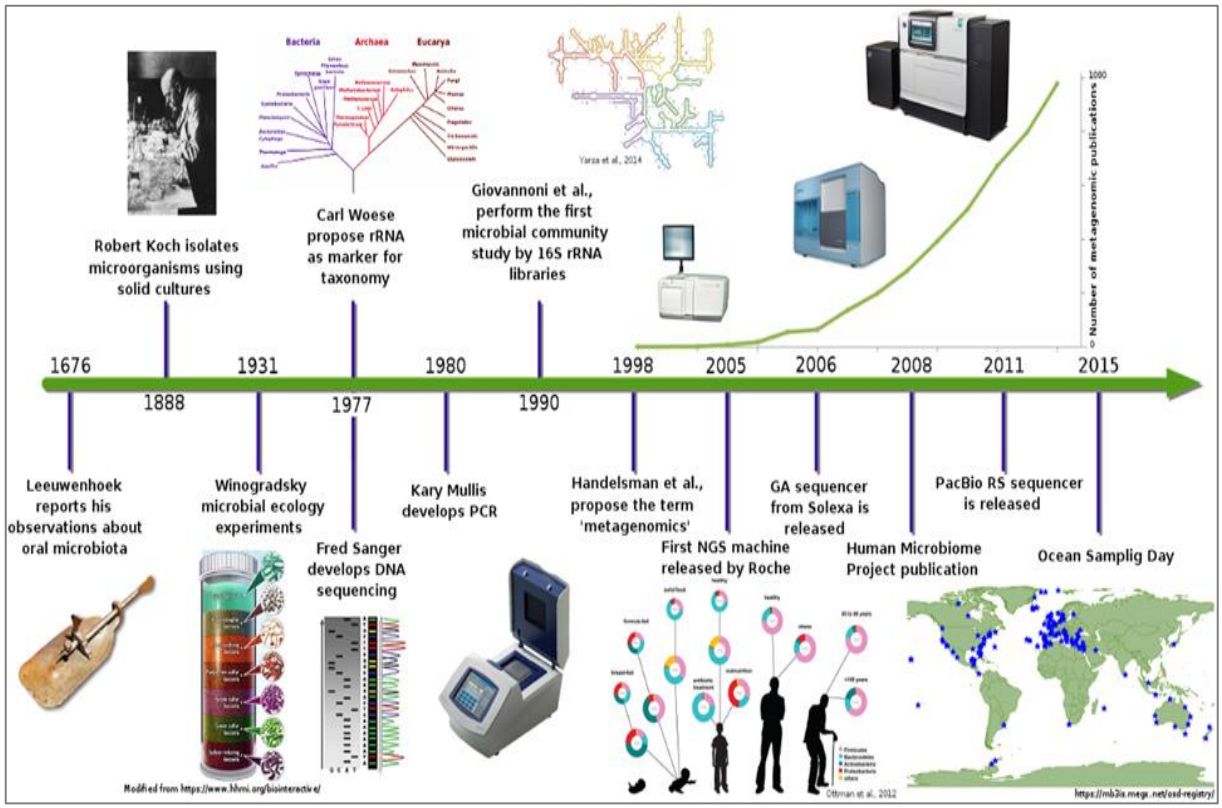


Figure 2.2 Development of molecular techniques over time.

Illustrated in Figure 2.2 indicates the advancements in molecular techniques over time, indicative of the current possibility of sequencing thousands of genes (e.g., human genome project).

The cloning and expression of genes to determine genes with novel enzyme activity is currently the focus of many research studies. It is believed that the identification of such novel enzymes from the environment, recombinantly expressing them, and carrying out biochemical tests could yield results that potentially form the basis of biological catalyst development in industrial scale processes and achieving sustainable workflows. The use of metagenomics and bioinformatics can help identify microbial genes that could assist host organisms in cellular processes including, detoxification of xenobiotics, synthesis of essential amino acids and degradation of recalcitrant biomass (Pope *et al.*, 2010; Cardoso *et al.*, 2012; Joynson *et al.*, 2017).

Function-based screening of expression libraries and sequence-based gene searches are two methods employed in identification of functional microbes through

bioprospecting of novel enzymes from metagenomes (L. L. Li *et al.*, 2009). In **function-based** screening of expression libraries, fragmented metagenomic DNA is inserted into expression vectors that are based on Cosmids, fosmids or bacterial artificial chromosomes (BACS), and gene expression examined in a suitable host (Collins and Hohn, 1978). This method of screening has been used in several studies including that described by (Yun and Ryu, 2005; Lam *et al.*, 2015; Ngara and Zhang, 2018). In **Sequence-based** gene search expression approach, an amplified fragment from metagenomic DNA is cloned into an expression system (L. L. Li *et al.*, 2009; Thomas, Gilbert and Meyer, 2012). This is a closed method and relies on known conserved sequences and thus unable to uncover non-homologous enzymes. This method of screening has been used in several studies including that described by (Loman *et al.*, 2013; Zhang *et al.*, 2021)

In Table 2.1 we compare the function-based screening and the sequence based(homology) screening.

Table 2.1 strategies for novel enzyme bioprospecting (Collins and Hohn, 1978; Yun and Ryu, 2005; Loman *et al.*, 2013; Lam *et al.*, 2015; Ngara and Zhang, 2018; Zhang *et al.*, 2021)

Strategies for novel enzyme bioprospecting	Advantages	Disadvantages
Function-based screening	• Access to previously unknown genes and their encoded enzymes • Functionally guaranteed Sequences and enzyme activities	• Due to differences such as codon usage, post-translation modification and protein folding elements, many genes from the environment may not be expressed. • Multiple vector systems must be scrutinised to arrive at suitable vectors
Homology based searching	• Regardless of gene expression and protein folding in the host, target genes can still be	• Failure to detect fundamentally different or novel genes

	identified, regardless of how comprehensive the target gene sequences are.	

2.2 Sequence-based gene search (homology search)

We focus on this method of gene search as this has been used in this study for the identification of putative function which aided in the determination or conserved domains. Protein identification and expression via sequence-based gene searches is a common approach to investigate the activity of putative proteins (L. L. Li *et al.*, 2009; Thomas, Gilbert and Meyer, 2012). Sequence-based gene search can be supported by structural alignments of known proteins using protein modelling which can act as an investigatory tool to identify active sites and ligand binding sites of proteins. For example, a protein model with highest C-score can be chosen for structural alignment. C-score gives quantitative measure of confidence of each model, where a higher value indicates a model with high confidence. Selected protein models based on highest Cscore usually have a good level of similarity to other functional characterised proteins which can be identified through biological function annotation based on structure, sequence, and protein-protein interaction. For enzyme commission predictions, COFACTOR and COACH are two good examples which have been employed in various studies. COFACTOR can be described as a method of biological function annotation based on structure, sequence, and protein-protein interaction (PPI) (Roy, Yang and Zhang, 2012; Zhang, Peter L Freddolino and Zhang, 2017) whereas, COACH is a meta-server approach to protein-ligand binding site prediction (Yang, Roy and Zhang, 2013a, 2013c).

However, not all proteins produced from sequence-based searches and structural comparisons become active. The result of this could be linked to genome assembly limitations where some essential amino acids required for activity are not present due

to base calling issues (Das and Vikalo, 2013). Nevertheless, protein expression via sequence-based gene searches has been used successfully in various studies for the identification of sequences with putative glycosyl hydrolase function identified from metagenomic data generated from sequencing organisms such as the common black slug and giant African land snail (Cardoso *et al.*, 2012; Joynson *et al.*, 2017). With the help of bioinformatic tools putative glycosyl hydrolase sequences can be queried against databases using BLASTN, BLASTP and BLASTX to aid in assigning putative functions before carrying out wet lab for the amplification, cloning and expression of these enzymes.

A previous work (Joynson *et al.*, 2017) used metagenomic approaches to characterise the bacterial diversity and functional capabilities of the gut microbiome of the *Arion ater*, a known agricultural pest. The abundance of recognised lignocellulose-degrading bacteria, as well as well-characterized bacterial plant pathogens, was discovered in gut metagenomic community sequences. This revealed a microbial community capable of degrading all components of lignocellulose, as well as a repertoire of more than 3,383 carbohydrate active enzymes (CAZymes), including several enzymes linked with lignin breakdown. This was made possible using genomic DNA sourced from the natural environment in the gut of the common black slug. DNA from the environment is limitless and bioinformatics can aid in classification of microbial diversity. Consequently, equipped with both tools, environmental niches can be studied and the quest to find novel enzymes will be made possible. This is not far-fetched and illustrated in Table 2.2 below is an overview of the number of coding sequences for various glycosyl hydrolase families found from environmental samples such as termites, wallaby, panda, snail (Cardoso *et al.*, 2012) and slug (Joynson *et al.*, 2017) using metagenomic tools.

Table 2.2 Comparison of the number of CDS representative of glycoside hydrolase profiles targeting plant structural polysaccharides in the snail, termite, giant wallaby, and human metagenomes.

Pfam group	Predominant activity	Termite	Wallaby	Panda	Snail	Slug
GH5	Cellulases	125	27	1	36	15
GH6	Endoglucanases	0	0	0	4	0
GH7	Endoglucanases	0	0	0	0	0
GH9	Endoglucanases	43	5	0	15	11
GH44	Endoglucanases	0	0	0	0	0
GH45	Endoglucanases	6	0	0	0	0
GH48	Cellobiohyrolases	0	0	0	2	0
Total		174	32	1	57	26
ENDOHEMICELLULASES						
GH8	Endoxylanases	21	2	1	46	11
GH10	Endo-1,4-β-xylanase	102	19	1	25	16
GH11	Xylanase	19	0	0	1	0
GH12	Endoglucanase & xyloglucanase	0	0	0	0	12
GH26	β-mannanase & xylanase	20	8	0	11	0
GH28	Galacturonases	15	10	0	69	6
GH53	Endo-1,4-β-galactanase	20	11	4	9	276
Total		197	50	6	161	321
XYLOGLUCANASES						
GH16	Xyloglucanase	6	6	6	12	117
GH17	1,3-β-glucosidases	0	0	0	2	60
GH81	1,3-β-glucanase	0	0	0	1	0
Total		6	6	6	15	177
DEBRANCHING ENZYMES						

GH51	α-L-arabinofuranosidase	13	19	2	22	3
GH62	α-L-arabinofuranosidase	0	0	0	2	0
GH67	α-glucuronidase	6	1	2	5	1
GH78	α-L-rhamnosidase	7	46	1	73	8
Total		26	66	5	102	12
OLIGOSACCHARIDE DEGRADING ENZYMES						
GH1	Mainly β-glucosidases	27	94	41	294	118
GH2	Mainly β-galactosidases	32	39	4	66	60
GH3	Mainly β-glucosidases	109	101	11	219	86
GH29	α-L-fucosidases	12	5	0	70	11
GH35	β-galactosidase	7	8	1	32	14
GH38	α-mannosidase	18	3	8	18	39
GH39	β-xylosidase	13	3	8	6	279
GH42	β-galactosidases	33	17	7	54	6
GH43	Arabinases & xylosidases	63	72	13	185	28
GH52	β-xylosidase	3	0	0	0	0
Total		317	342	93	944	641

Table 2.3. Comparison of the number of CDS representing Auxiliary active families (AA) of enzymes targeting plant structural polysaccharides in the common black slug.

AA FAMILY	PUTATIVE FUNCTION	Slug
AAI	Laccase / p-diphenol: oxygen oxidoreductase / ferroxidase (EC 1.10.3.2); ferroxidase (EC 1.10.3.-); Laccase-like multicopper oxidase (EC 1.10.3.-)	20
AA2	Manganese peroxidase (EC 1.11.1.13); versatile peroxidase (EC 1.11.1.16); lignin peroxidase (EC 1.11.1.14); peroxidase (EC 1.11.1.-)	5
AA3	cellobiose dehydrogenase (EC 1.1.99.18); glucose 1-oxidase (EC 1.1.3.4); aryl alcohol oxidase (EC 1.1.3.7); alcohol oxidase (EC 1.1.3.13); pyranose oxidase (EC 1.1.3.10)	148
AA4	vanillyl-alcohol oxidase (EC 1.1.3.38)	11
AA5	Oxidase with oxygen as acceptor (EC 1.1.3.-); galactose oxidase (EC 1.1.3.9); glyoxal oxidase (EC 1.2.3.15); alcohol oxidase (EC 1.1.3.13)	41
AA6	1,4-benzoquinone reductase (EC. 1.6.5.6)	62
AA7	Glucooligosaccharide oxidase (EC 1.1.3.-); chitooligosaccharide oxidase (EC 1.1.3.-)	13
AA8	Iron reductase domain	0
AA9	copper-dependent lytic polysaccharide monooxygenases (LPMOs)	0
AA10	copper-dependent lytic polysaccharide monooxygenases (LPMOs	3
AA11	copper-dependent lytic polysaccharide monooxygenases (LPMOs	0
AA12	pyrroloquinoline quinone-dependent oxidoreductase activity	0

In Table 2.2 and Table 2.3 is the number of coding sequences representative of glycosyl hydrolases and auxiliary actives targeting plant structural polysaccharides identified from successful enzyme identification studies from environmental sources. Glycosyl hydrolases from snail, termite, giant wallaby and human metagenomes have been identified using metagenomic approaches and has led to the expression/production of proteins that have been used industrially for the bioconversion of plant biomass. In the next section we look at some of the methods which have led to the identification of microbial communities and the identification of enzymes with GH functionality.

2.2.1 Metagenomic tools

Bottle necks in the analyses of sequencing data are becoming more evident with increasing throughput in metagenomics (Schoenfeld *et al.*, 2008). These bottlenecks have led to the development of new metagenomic tools capable of handling high throughput data generated from Illumina or Ion Torrent reads (Nurk *et al.*, 2013). Metagenomic tools are required in the quality filtering of reads, assembly, annotation, functional and phylogenic classification of identified sequences. FastQC has been reported as a simplified tool that can be used for quality control of high throughput sequence data from sequencing runs, thus providing a quick overview in identifying problematic areas of a sequence (Andrews, 2018). Problematic areas can then be cleaned up using flexible tools embedded in software such as trimmomatic which can perfectly handle paired end data (Bolger, Lohse and Usadel, 2014) and the resulting output files can then be assembled.

The process of assembling biological reads from sequencing into longer sequences known as contigs is known as genome assembly (Thrash, Hoffmann and Perkins, 2020). Assembly can be achieved using a variety of metagenomic assemblers and various studies have used metagenome assemblers including velvet/metavelvet (Zerbino and Birney, 2008), metaSPAdes (Nurk *et al.*, 2017), MEGA-HIT (Li *et al.*, 2015), IDBA-UD (Peng *et al.*, 2012) and Ray-Meta (Boisvert *et al.*, 2012). The assembler choice is usually based on several factors in relation to an experimental end point such as the processing capacity and scalability. The use of quality assessment tools such as Metaquast has also been described allowing for functional element prediction computing various statistics relating to the assembly such as scaffold lengths, gene prediction statistics and gene alignment statistics. Metaquast uses *Ab initio* gene identification through MetaGeneMark (Zhu, Lomsadze and Borodovsky, 2010) and reports various statistics regarding assessment of quality of reads. Significant improvement in total scaffold length has been reported when metaSPAdes was used as a benchmark against other assemblers. According to (Nurk *et al.*, 2017) total

scaffold length was significantly improved by 21% when metaSPAdes was used to assemble SOIL data set compared to IDBA-UD and 40% compared to MEGAHIT. This indicates the power of metaSPAdes as a powerful tool for the assembly of metagenomic data. In terms of the statistics associated with metagenome assemblers, the general perception is that a larger N50 value is indicative of a better assembly. However, other studies have suggested that the N50 value alone is not always a good metric for assessing contiguity. According to Salzberg *et al.*, (2012), N50 can be deceptive since contigs created by misjoins tend to be larger than they should be, inflating the N50 score leading to other statistics such as e-size being used as an assessment of contiguity.

Depending on the research, the question of "who is there?" is usually queried after data assembly. Taxonomic abundances can be estimated using software such as Krona, which enables for intuitive exploration of relative abundances and confidences within complex metagenomic categorization hierarchies (Ondov, Bergman and Phillippy, 2011). The assembled data will have to go through indexing and mapping using Burrow-Wheeler Aligner and SAMtools (Li and Durbin, 2009, 2010; Li, 2012), deduplicated to produce concordant reads with only correctly paired reads, k-mers based taxonomic labelling using kraken2 (Wood, Lu and Langmead, 2019) and re-estimation of taxa using tools such as bracken (Lu *et al.*, 2017).

Following the question of who is there, is "what do they do?" To address this, metagenome annotations have been achieved using tools such as PROKKA (Seemann, 2014) for the rapid annotation of prokaryotic genomes. To produce a thorough and trustworthy annotation of bacterial genomic sequences, Prokka coordinates several already-existing software tools such as Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm) which allows for accurate gene structure prediction, improved translation initiation site recognition, and reduced false positives (Hyatt *et al.*, 2010). Other analysis can be introduced to a metagenomics

workflow, such as variant calling if the question is to identify variants in an evolved line compared to the ancestral genome. Updated metagenomic tools offer a solution to address some of the challenges associated with assemblies such as misjoins which tend to be larger than they should be, inflating the N50 score. With error correction modules embedded in software such as metaSPAdes, such challenges are addressed leading to more improved predictions when tools such as Prokka are used.

Aims

The aim was to determine if using updated metagenomic tools which allow for accurate gene structure prediction, improved translation initiation site recognition, and reduced false positives will result in improved coding DNA sequences from that identified in a previous study by (Joynson *et al.*, 2017).

Objectives

1. To reassemble the slug metagenome using metaSPAdes and evaluate the genome assembly using Metaquast
2. To reannotate the reassembled metagenome using Prokka
3. Identify the dominant bacteria genera and determine their role as biocatalysts

2.3 Methods

2.3.1 Quality control

Read data as fastq.gz files were downloaded from the European Nucleotide Archive (ENA) Project: PRJEB21599 available via the link (*Arion ater* data). Sequence output files were assessed using FastQC version 0.11.8 (Andrews, 2018) offered via the galaxy.org servers. Initial FastQC output resulted in 25,996,846 reads passing but observing the Per base sequence quality distribution which shows quality for individual bases along x- axis and phred quality score along y-axis, some sequences had a phred quality score below 20 which is generally unacceptable quality for read lengths of 250 bp. High phred scores indicate confidence in the base call and as a result, data set was trimmed applying settings which removes bases with phred quality score below 20, resulting in 25,795,430 reads.

2.3.2 Assembly using metaSPAdes

The process of matching and combining pieces of a larger DNA sequence to recreate the original sequence is known as sequence assembly. Assembly of metagenomic data remains a challenge which limits the discovery of biological populations. Here we have used metaSPAdes to recreate the original sequence. Using the Bayes Hammer --read error correction tool, which works well on both single-cell and standard data sets, error correction was carried out through metaSPAdes using default settings on the trimmed data set.

See command line input:

Metaspades –only-error-correction -1 R1_trimmed -2 R2_trimmed -o output_directory

Following error correction, the resulting contigs.fasta output was then assembled using metaSPAdes offered via St. Petersburg genome assembler (SPAdes version 3.11.1) The resulting corrected output R1_trimmed and R2_trimmed were then assembled using --only-assembler mode

See command line input:

Metaspades.py -k 33,55,77,99,127 --only-assembler -1 R1_trimmed -2 R2_trimmed -o output_directory.

The resulting contigs.fasta file was then assessed for quality using metaquast version 5.0.2 (Mikheenko, Saveliev and Gurevich, 2016).

See command line input:

metaquast.py --output_directory contigs.fasta –max-ref-number 0

2.3.3 Read mapping

Read mapping was carried out using (BWA) Burrow-Wheeler Aligner (Li and Durbin, 2009, 2010; Li, 2012). An index to the assembled genome described in section 2.3.2 was first created using the index function of bwa. After indexing, sequence reads were mapped backed to the assembled genome using the bwa-mem function resulting in a .sam file as the output. Read pairing information and flags were then cleaned up before sorting. A compressed .bamfixmate output was then generated using samtools fixmate to enable efficient sorting (H. Li *et al.*, 2009; Li, 2011). The fixmate.bam output was then sorted into a coordinate order resulting in a sorted.bam file before duplicates were removed. Duplicate reads were then removed using the samtools markdup to mitigate the effect of PCR bias introduced during the library construction. Output sorted.deduplicated.bam files were then assessed for mapping statistics using the samtools flagstat function. Quality based selection was then employed to sub-select reads with mapping quality >= 20 as defined by Bowtie2 (Langmead *et al.*, 2019). Unmapped reads were also extracted for classification using the samtools view -b -f 4 and mapping statistics checked. Finally, fastq sequences of the mapped and unmapped bwa aligned reads were extracted for subsequent pipeline analysis.

Bwa index and bam file generation command lines

bwa index /path to contigs.fasta

bwa mem /path to scaffolds.fasta /path to read1_trimmed.fastq.gz /path to read2_trimmed.fastq.gz > arion_aligned.sam

samtools sort -n -O sam arion_aligned.sam | samtools fixmate -m -O bam - arion_aligned.fixmate.bam

```
samtools sort -O bam -o arion.sorted.bam arion_aligned.fixmate.bam
samtools markdup -r -S arion.sorted.bam arion.sorted.dedup.bam
samtools flagstat arion.sorted.dedup.bam
```

Correctly mapped reads

```
samtools view -h -b -f 3 arion.sorted.dedup.bam > arion.sorted.dedup.concordant.bam
```

extract based on mapping quality

```
samtools view -h -b -q 20 arion.sorted.dedup.bam > arion.sorted.dedup.q20.bam
```

Extract unmapped

```
samtools view -b -f 4 arion.sorted.dedup.bam > arion.sorted.dedup.unmapped.bam
```

extract fastq from correctly mapped

```
samtools fastq -1 arion.sorted.dedup.concordant.R1.fastq.gz -2 arion.sorted.dedup.concordant.R2.fastq.gz arion.sorted.dedup.concordant.bam
```

extract fastq from mapping quality

```
samtools fastq -1 arion.sorted.dedup.q20.R1.fastq.gz -2 arion.sorted.dedup.q20.R2.fastq.gz arion.sorted.dedup.q20.bam
```

extract fastq from unmapped

```
samtools fastq -1 arion.sorted.dedup.unmapped.R1.fastq.gz -2 arion.sorted.dedup.unmapped.R2.fastq.gz arion.sorted.dedup.unmapped.bam
```

2.3.4 Taxonomic classification with kraken2 and bracken

Kraken2, a taxonomic classification approach that leverages exact k-mer matches to achieve high accuracy and quick classification rates was used to classify the gathered data. Mapped DNA reads and unmapped DNA reads were investigated for taxa following assignment of taxonomic label to sequences using the minikraken2 database which was created from Refseq bacteria, archaea, and viral libraries and the GRCh38 human genome (Wood, Lu and Langmead, 2019).

Command lines for kraken on mapped and unmapped reads

Mapped_concordant reads

kraken2 --threads 4 --db minikraken2_v2_8GB_201904_UPDATE --fastq-input --report arion_mapped --gzip-compressed --paired arion.sorted.dedup.concordant.R1.fastq.gz arion.sorted.dedup.concordant.R2.fastq.gz > arion_mapped_concordant.kraken

Mapped_q20 reads

kraken2 --threads 4 --db minikraken2_v2_8GB_201904_UPDATE --fastq-input --report arion_mapped_q20 --gzip-compressed --paired arion.sorted.dedup.q20.R1.fastq.gz arion.sorted.dedup.q20.R2.fastq.gz > arion_mapped_q20.kraken

Unmapped reads

kraken2 --threads 4 --db minikraken2_v2_8GB_201904_UPDATE --fastq-input --report arion_unmapped --gzip-compressed –paired arion.sorted.dedup.unmapped.R1.fastq.gz arion.sorted.dedup.unmapped.R2.fastq.gz > arion_unmapped.kraken

From the command line, kraken2 outputs were retrieved, the report and the kraken output file which is required for visualization using tools such as krona. To improve on the kraken2 output, Bayesian Re-estimation of Abundance with kraken (bracken) was used to classify reads to the best matching location in the taxonomic tree thus producing more accurate species and genus level abundance estimates than using just kraken2. Bracken output files from our mapped and unmapped reads were then used as input for interactive taxonomic visualization using krona tools (Ondov, Bergman and Phillippy, 2011)

Command line for bracken

bracken -d minikraken2_v2_8GB_201904_UPDATE -i arion_concordant_report -o arion_concordant.bracken -l S

bracken -d minikraken2_v2_8GB_201904_UPDATE -i arion_mapped_q20 -o arion_mapped_q20.bracken -l S

bracken -d minikraken2_v2_8GB_201904_UPDATE -i arion_unmapped -o arion_unmapped.bracken -l S

krona visualisation

cat arion_concordant.bracken | cut -f 1,2 > arion_concordant.bracken.krona

cat arion_mapped_q20.bracken | cut -f 1,2 > arion_mapped_q20.bracken.krona

cat arion_unmapped.bracken | cut -f 1,2 > arion_unmapped.bracken.krona

ktImportTaxonomy arion_concordant.bracken.krona

Repeat for mappedq20 and unmapped reads

2.3.5 Prokka: rapid prokaryotic genome annotation

Genome annotation was performed using Prokka (version 1.13.3) to identify and label features of interest within the assembled metagenome (Seemann, 2014). In this study the focus was on bacteria, so we used the default settings of Prokka which uses --kingdom and annotates using Bacteria. Depending on the research question one can specify, Archaea, Virus or Mitochondria. The command line used for this was.

Prokka –outdir_prokka_report scaffolds.fasta --norrna –notrna

2.4 Results

2.4.1 Fast QC output on raw reads

Fast QC quality reports were generated following fastqc run on raw reads and quality filtered reads. Quality filtering was achieved by observing the per base sequence quality and selecting sequences with a phred quality score of 20 and above. A Phred Score of 20 indicated the likelihood of finding 1 incorrect base call among 100 bases thus indicating high confidence in the per base quality (Goswami and Sanan-Mishra, 2022).The quality per tile plots were also visualised to identify deviation from the average quality for each tile observing hot and cold regions.

Table 2.4 statistics of fastqc run on forward R1 reads

Measure	Value
Filename	R1_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12998423
Sequences flagged as poor quality	0
Sequence length	10-250
%GC	51

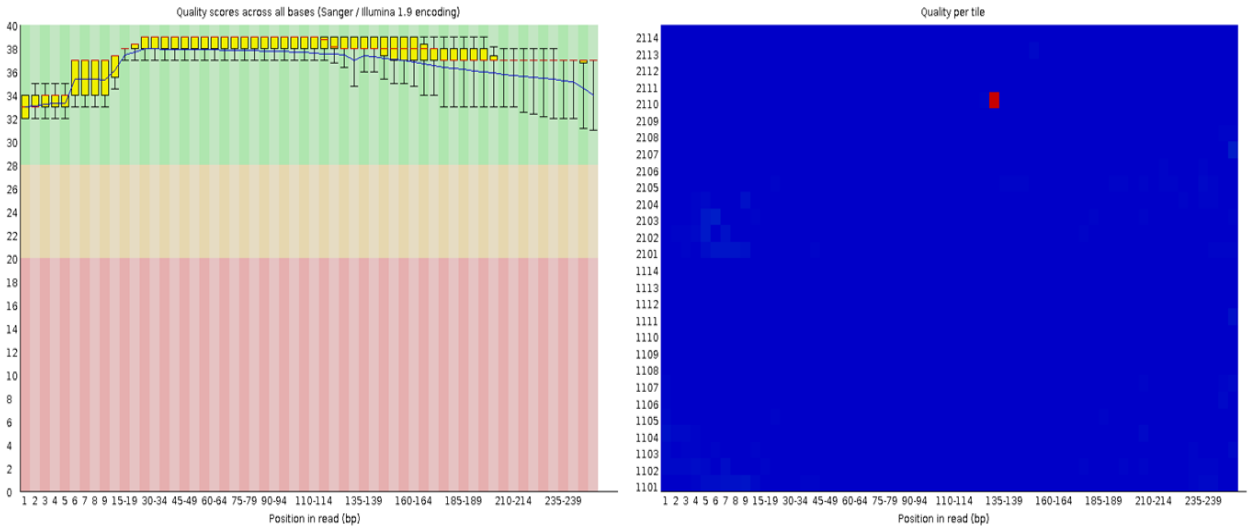


Figure 2.3. Quality score across all bases using the Sanger/Illumina 1.9 encoding

In Figure 2.3. Per base sequence quality distribution showing quality for individual bases along x- axis and phred quality score along y-axis. From the distribution plot, yellow box – interquartile range 25 - 75 percentile, red – median value, black whiskers represent 10th and 90th percentile and the blue line shows the mean quality score. Observed quality scores suggested high confidence in the sequence quality with bases all above phred quality score of 20 which describes the accuracy in the base calling of the sequencing. From the quality per tile plot, red colour was observed which perhaps could be related to a worse quality tile observed than other tiles for that base.

Table 2.5 statistics of fastqc run on reverse R2 reads

Measure	Value
Filename	R2_fastq_gz.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12998423
Sequences flagged as poor quality	0
Sequence length	10-250
%GC	51

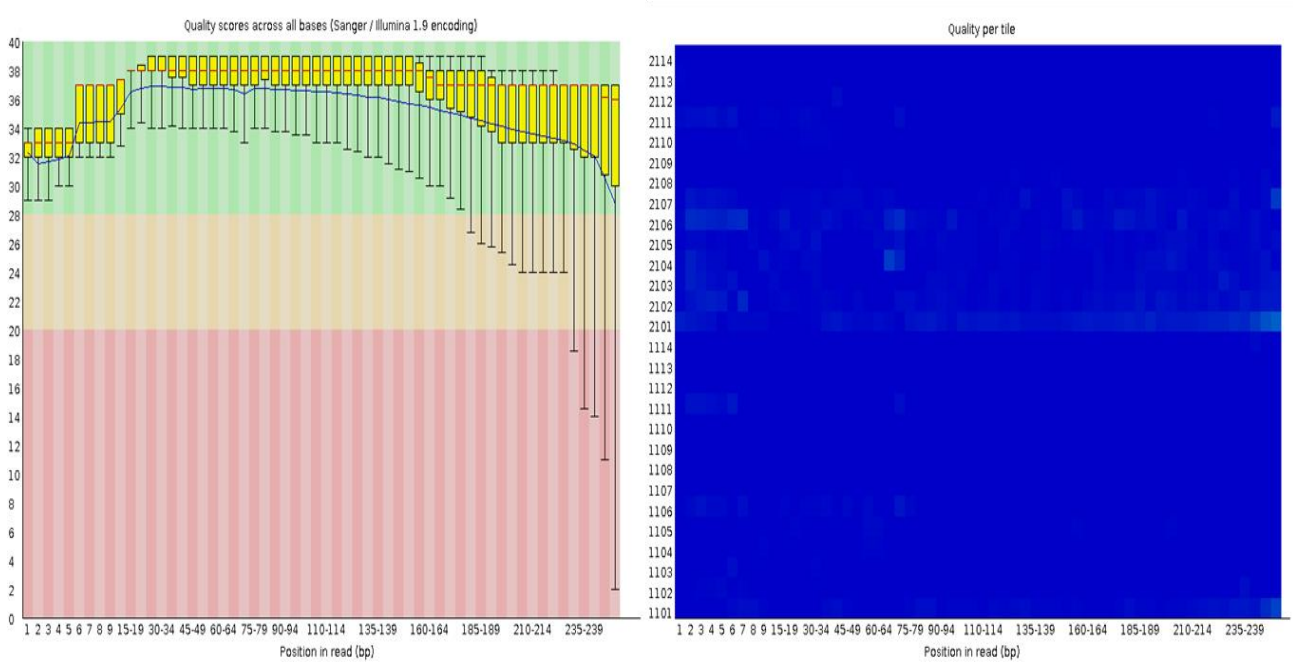


Figure 2.4 Quality score across all bases using the Sanger/Illumina 1.9 encoding

In Figure 2.4. Per base sequence quality distribution showing quality for individual bases along x- axis and phred quality score along y-axis. From the distribution plot, yellow box – interquartile range 25 - 75 percentile, red – median value, black whiskers represent 10th and 90th percentile and the blue line shows the mean quality score. From the figure above, the per base sequence mean quality score based on the phred quality score index, suggests that bases approaching positions 220 – 250 are below a phred quality score of 20 which describes inaccuracy in the base calling of the sequencing.

2.4.2 Fast QC output on trimmed reads after trimmomatic

Raw data was trimmed applying a quality filtering to select bases with phred quality score of above 20.

Table 2.6 Fastqc after dataset trimming of R1 reads

Measure	Value
Filename	Trimmomatic on R1_fastq_gz _R1 paired_.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12897715
Sequences flagged as poor quality	0
Sequence length	1-250
%GC	51

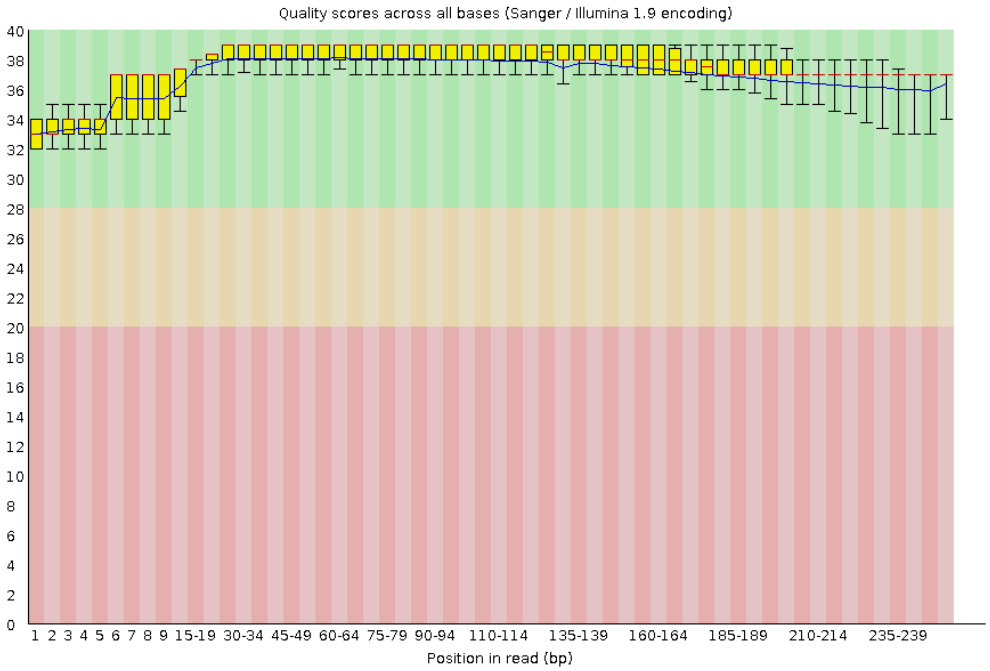


Figure 2.5 Quality score across all bases using the Sanger/Illumina 1.9 encoding

In Figure 2.5. Per base sequence quality distribution showing quality for individual bases along x- axis and phred quality score along y-axis. From the distribution plot, yellow box – interquartile range 25 - 75 percentile, red – median value, black whiskers represent 10th and 90th percentile and the blue line shows the mean quality score. Observed quality scores suggested high confidence with bases all above phred quality score of 20 which describes the accuracy in the base calling of the sequencing.

Table 2.7 Fastqc after dataset trimming of R2 reads

Measure	Value
Filename	Trimmomatic on R2_fastq_gz _R2 paired_.gz
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	12897715
Sequences flagged as poor quality	0
Sequence length	1-250
%GC	51

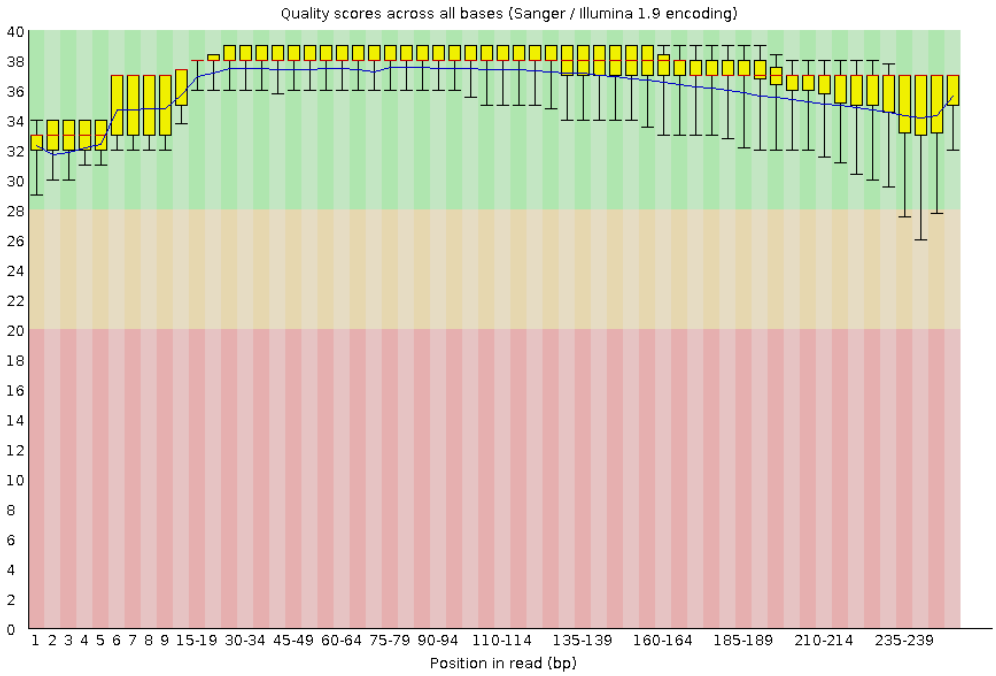


Figure 2.6 Quality score across all bases (Sanger/Illumina 1.9 encoding)

In Figure 2.6. Per base sequence quality distribution showing quality for individual bases along x- axis and phred quality score along y-axis. From the distribution plot, yellow box – interquartile range 25 - 75 percentile, red – median value, black whiskers represent 10th and 90th percentile and the blue line shows the mean quality score. Observed quality scores suggested high confidence with bases all above phred quality score of 20 after applying quality filtering compared to Figure 2.4.

2.4.3 Genome assembly

Genome assemblies were assessed for quality using metaquast. The N50, L50, %GC content and the total length of assembled sequences were compared by using the metaquast output following assembly using different kmers in increments of 22. Computed parameters have been summarized in the Table 2.8.

Table 2.8 Statistics from MetaQuast comparing quality of different kmers used in assembly

Statistics	Kmer used in assembly				
	33	55	77	99	127
Total length (Mbp)	124.3	170.3	193.5	205.5	209.5
GC%	46.64	48.38	48.89	49.07	49.11
L50	22347	25949	25208	22599	13428
N50	1211	1391	1569	1769	2353

2.4.4 Mapping results

Out of the 2360285 mapped reads, 20928948 were properly paired according to samtools flagstat results. When reads were selected based on their mapping quality of 20, 22360958 reads were mapped with 20365114 reads properly paired. Samtools flagstat was applied on the unmapped reads that were extracted and the results indicated 832436 reads with 0 properly paired reads.

2.4.5 Kraken and Bracken

Out of the deduplicated concordant reads 10464474 sequences, 6053859 sequences were classified (57.85%) with 4410615 sequences unclassified (42.15%). Kraken was also used on the Deduplicatedq20 obtained reads and resulted in 11114540 sequences with 7686728 sequences classified (69.16%) and 3427812 sequences unclassified (30.84%). For the deduplicated Unmapped sequences 402515 sequences were processed and out of this, 145968 sequences classified (36.26%) and 256547 sequences unclassified (63.74%). Bracken was then applied to re-estimate the taxonomic

abundance and improve on the kraken output. Re-estimation of the taxonomic abundance in the concordant reads resulted in bracken accurately identifying 3147 species in sample. In the q20 quality reads kraken output, 2944 species were identified accurately following estimation. Re-estimation of the unmapped reads resulted in 1981 species identified in the sample.

2.4.6 Gut microbial diversity using mapped reads

Taxonomic investigation of the concordant reads at the super kingdom level indicated that majority of the sample was dominated by bacteria with 97% and 1% to virus, 0.7% Archaea, 0.06% *Homo sapien*, 0.06% had no hits and 0.8% belonging to other root.

Figure 2.7 Gut microbial diversity of the common black slug using mapped reads

In **Figure 2.7**. An in-depth look into the relative abundance of microbial groups at the order level of the gammaproteobacterial class indicated that majority of the bacteria were associated with

Enterobacterales with 19%, Pseudomonadales with 13% and other notable order such as Alteromonadales with 6% and Xanthomonadales with 3%. A final in-depth investigation into the family Enterobacteriaceae identified bacteria belonging to the genera Enterobacter with 21%, followed by *Citrobacter* with 17% and other notable family such as *Klebsiella*/*Raoultella* group 12%.

2.4.7 Gut microbial diversity using Q20 mapped reads

Taxonomic investigation of the extracted reads using a quality mapping score of >= 20 at the super kingdom level indicated that most of the sample was dominated by bacteria with 98% and 1% to viruses, 0.4% Archaea, 0.07% Homo sapiens, 0.07% had no hits and 0.9% belonging to other root.

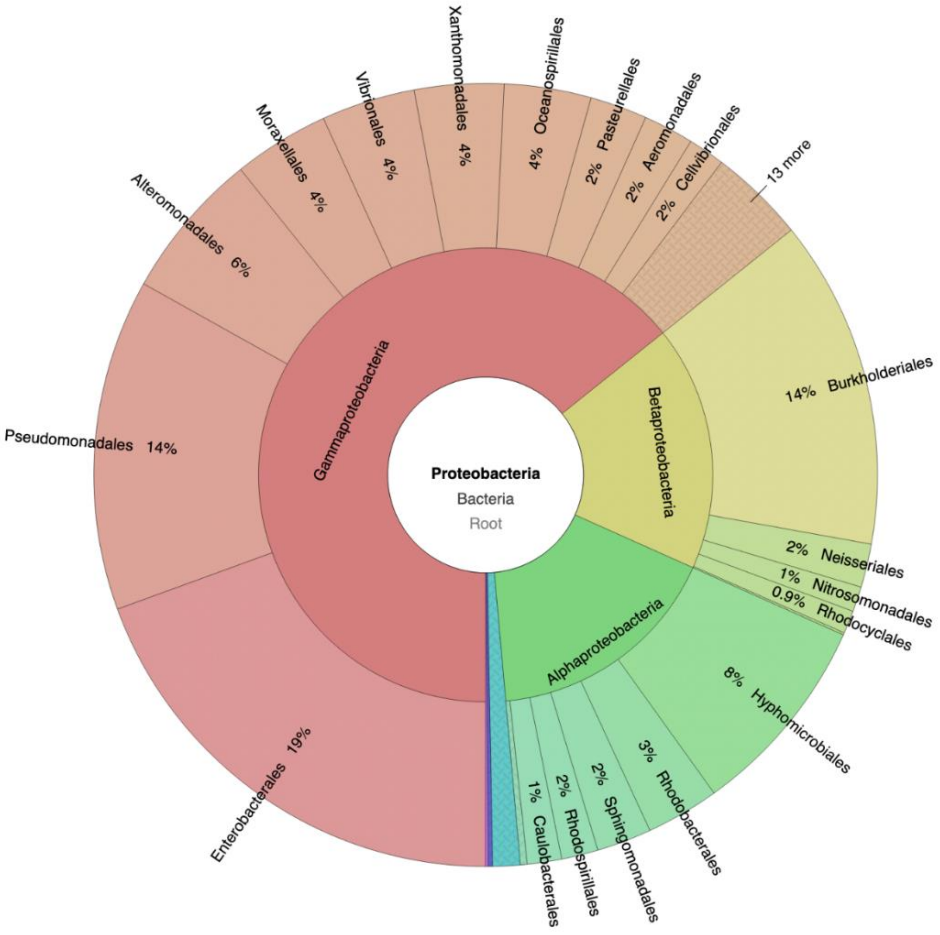


Figure 2.8 Gut microbial diversity using quality filtered mapped reads

In Figure 2.8. A deeper investigation into the relative abundance of microbial groups at the order level of the gammaproteobacterial class indicated that majority of the bacteria were associated with Enterobacterales with 19%, Pseudomonadales with 14% and other notable order such as

Alteromonadales with 6% and Xanthomonadales with 4%. A final in-depth investigation into the family Enterobacteriaceae identified bacteria belonging to the genera Enterobacter with 21%, followed by *Citrobacter* with 17% and other notable family such as *Klebsiella/Raoultella* group 12%.

2.4.8 Gut microbial diversity using unmapped reads

Taxonomic investigation of the extracted reads that could not be mapped indicated the presence of 98 % bacteria at super kingdom level, 0.7% virus, 0.2% no hits, 0.2% *Homo sapiens* and 1% covered by other.

Figure 2.9 Gut microbial diversity using unfiltered and unmapped reads

In Figure 2.9 An in-depth look into the relative abundance of microbial groups at the order level of the gammaproteobacterial class indicated that majority of the bacteria were associated with Pseudomonadales with 23%, Enterobacterales with 22%, and other notable order such as Xanthomonadales with 6%. A final in-depth investigation into the family Enterobacteriaceae identified bacteria belonging to the genera *Citrobacter* with 25% followed by *Enterobacter* with 17%, and other notable family such as *Klebsiella/Raoultella* group 19%.

2.4.9 Genome annotation

The scaffolds.fasta output was annotated successfully identifying 245,605 contigs with 250,190,610 bases, 222,541 conserved domain sequences (cds) and 92 clustered regularly interspaced short palindromic repeats (CRISPR). Resulting output prokka.faa and prokka.fnn which contain nucleotide sequence and amino acid sequence were retrieved respectively for further analysis.

2.4.10 Comparing results from the assembly and annotation in this study to a previous study by (Joynson *et al.*, 2017)

Findings from this study suggest a 2-fold increase in the number of coding DNA sequences and the improvement of predicted open reading frames (OFR) compared to that reported in a previous study carried out by (Joynson *et al.*, 2017). In Table 2.9 outputs from this study are compared to that from (Joynson *et al.*, 2017).

Table 2.9 Assembly and Annotation statistics comparison

Study	(Joynson *et al.*, 2017)	*This study*
Sequence data	81.74Mbp	209.5Mbp
N50	1.8kbp	2.4kbp
Taxa	99.4% Bacteria	97% bacteria
Coding sequences	108,691	222,541
Species found at Gamma proteobacteria level	86	696

From Table 2.9 the sequence data more than doubled when a kmer of 127 was used in the assembly for this study. In the study by (Joynson *et al.*, 2017) a kmer of 51 was used which resulted in an N50 of 1.8kbp and annotation resulting in 108,691 coding sequences. This suggests that the choice of kmer is a crucial factor in achieving better assembly and improvement in number of coding sequences.

2.4.11 Predicted open reading frames (OFR) of annotated sequences

To futher demonstrate how we have improved on the previous assembly by (Joynson *et al.*, 2017), DNA sequence of a putative esterase already annotated and expressed in a later section of this thesis was compared to the newly metaSPAdes assembled and prokka annotated sequence. We identified that approximately 17 amino acid residues were missed when assembly was performed using velvet and kmer of 51. These missed residues could contribute to the full function of a protein following expression, so it is crucial to have complete cds. Multiple sequence alignment showing this difference has been illustrated in Figure 2.10

Figure 2.10. Multiple sequence alignment showing the 17 amino acid difference at the N-terminus.

In Figure 2.10, the sequence alignments suggest metaSPAdes was able to improve on the assembly and accurately predict the open reading frame for the coding DNA sequence for a putative CE1 family enzyme as annotated by Prokka. This served as an indication that the tools embedded in the assembler, in particular Prodigal which allows for accurate gene structure prediction, improved translation initiation site recognition, and reduced false positives (Hyatt *et al.*, 2010).

2.5 Discussion

The common black slug is said to have caused economical losses with over £30 million spent each year by the UK agricultural industry on molluscide pellets (Aguiar and Wink, 2005). This economical loss prompted research into identifying the usefulness of the microbial community which enables this successful gastropod species in Europe and North America to cause such damage. Research into the repertoire of microbial community present in the common black slug has been ongoing since the 1960s, focusing on the breakdown of simple sugars and proteinase hydrolysis (Evans and Jones, 1962). In a previous study by (Joynson *et al.*, 2017), the microbial community in the gut of the British black slug and biochemical activity of various gut microbes identified were profiled through the use of metagenomic tools. Their work identified over 3,383 carbohydrate active enzymes including enzymes belonging to the auxiliary active family involved in lignin degradation. Other studies have also identified carbohydrate active enzymes from sources including a suite of aerobic and anaerobic biomass degraders found in a variety of ecosystems such as soil, seawater, insect and mollusc gut, compost and several other sources (Biddanda, 1988; Cardoso *et al.*, 2012; Joynson *et al.*, 2017; Patra, Das and Shin, 2018). This research suggested the possibility of establishing a relationship between feeding and evolutionary success and, secondly, the presence of novel enzymes with biotechnological potential that could be used in the production of base molecules to produce biopharmaceuticals, cosmetics, and liquid fuel such as biodiesel and biofuel.

In this study, the aim was to determine if using updated metagenomic tools will improve on the number of coding DNA sequences and observe improvements in the predicted open reading frames (OFR) of annotated sequences in contrast to that reported by (Joynson *et al.*, 2017). MetaSPAdes was employed to improve on assembly statistics relaying on an algorithm to identify an appropriate kmer based on sequence data. We tested this by subjecting the algorithm to carry out assembly using kmer values of 55, 77, 99 and 127. The selected appropriate kmer for our assembly was k127

as assembly output with this kmer resulted in double the sequence data, N50 of 2.4kbp and double the coding DNA sequences as seen in Table 2.9. Although some studies have identified the N50 as a good metric for assessing assembly quality, other studies have described contrary to this. According to Salzberg *et al.*, (2012), N50 can be deceptive since contigs created by misjoins tend to be larger than they should be, inflating the N50 score. They further suggest that e-size, which quantifies the size of a contig on the genome that contains a randomly chosen base as a solution to address this issue. According to the paper, this metric can assist in determining how many genes are entirely contained in contigs or scaffolds as opposed to being fragmented. Nevertheless, (Nurk *et al.*, 2017) have described total scaffold length was significantly improved by 21% when metaSPAdes was used to assemble SOIL data set compared to IDBA-UD and 40% compared to MEGAHIT, therefore indicating the dominance of metaSPAdes as a powerful tool for the assembly of metagenomic data. This was evident in our study with assembly with metaSPAdes resulting in a 48% improvement in total scaffold length compared to the use of velvet for the assembly of the slug gut metagenome confirming metaSPAdes as a dominant tool for the assembly of metagenomic data.

In terms of taxonomic classification, our study was able to not only confirm the ecological richness of the gut microbiome of the common black slug, but also improve on the ecological richness following the identification of more species under proteobacteria which have been described to be actively involved in the degradation of lignocellulosic material. Our findings identified gammaproteobacteria as the predominant class in the metagenome making up 63% of proteobacteria with 696 species in this class which is an 8-fold increase than what was previously observed. The most notable abundant genera was again no different from the genera indicated by (Joynson *et al.*, 2017) with notably 131 species of *Pseudomonas*, 33 species of *Acinobacter*, 31 species of *Vibrio*, 21 species of *Enterobacter*, 17 species of *Citrobacter* and 12 species of *Klebsiella*/*Raoultella* group. At the phylum level, proteobacteria

predominated with 67% of bacteria belonging to this phylum. The dominance of this phylum has also been reported in terrestrial land snails such as the African land snail (Cardoso *et al.*, 2012). According to Cardoso *et al*., (2012), Pseudomonas dominated at the genus level with species from this genus being widely distributed in nature, thus being able to utilise a variety of nutrients and possibly aiding host in various functions including, breakdown of lignocellulose material (Mouafo Tamnou *et al.*, 2021), detoxification of organic compounds (Sivaloganathan and Brynildsen, 2021) and synthesis of amino acids (Jagmann and Philipp, 2018)). The dominance of *Pseudomonas* has also been reported in the top 10 bacterial ranked pathogens with *Pseudomonas synringae* pathovars ranking 1st (Mansfield *et al.*, 2012).

2.6 Conclusion

The reduction in cost of next generation sequencing and the possibility of large-scale repetition is creating the platform for de novo sequencing applications towards the characterisation of novel species. Previously unexplored environments which hold thousands of unculturable microbes can now be investigated leading to taxonomic and functional annotation. However, the accurate assembly of genomes and the correct annotation of sequences remains a challenge especially when metagenomic tools are used incorrectly. The advent of metagenomic tools and updated databases seeks to address these challenges with tools allowing rapid processing times, scalability and accuracy in determining coding DNA sequences. In conclusion, we have met our aims by improving on the number of coding DNA sequences and improved on the open reading frame prediction following re-assembly and re-annotation of the metagenomic data from the common black slug using an appropriate kmer value of 127. We have demonstrated how using robust assemblers such as metaSPAdes can allow the assembly of longer contigs leading to the identification of previously unexplored bacteria species harboured amongst the repertoire of microorganisms found amongst the gut microbiota of *Arion ater*.

In this study we observed an increase in N50 and coding DNA sequences as well as an increase in the number of species under proteobacteria which increased by 8 folds compared to a previous study. Also, we show in this study how assembly can affect open reading frame and thus affect coding DNA sequences predictions. A typical example has been depicted in Figure 2.10 where 17 amino acids were missed in a previous assembly using kmer of 51. Accurate open reading frames and gene predictions are essential in testing the activity of expressed proteins from sequences retrieved from unexplored environments. Therefore, this study has contributed to knowledge by showing how a genome assembled using an inappropriate kmer can result in missed sequences for full protein functionality. The ecological richness was also observed to have improved when an appropriate kmer was used suggesting that the use of appropriate kmers can result in the identification of new species which can thus lead to identification of novel proteins towards industrial applications. The pipeline used in this study can be used in other studies where the goal is to observe improvements in previously assembled and annotated datasets.

Chapter 3. Cloning, Expression and Biochemical activity testing of an alpha-L-rhamnosidase from the slug gut

Abstract

Low thermal stability and selectivity restricts the application of extremely valuable enzymes such as α-L-Rhamnosidases. GH78 α-L-rhamnosidases are biotechnologically important as they can catalyse the hydrolysis of α-L-rhamnosyl linkages in L-rhamnosides, such as flavonoid glycosides including naringin, hesperidin, and rutin; polysaccharides like rhamnogalacturonan and arabinogalactan-protein; and glycolipids, thus improving bioavailability. Flavonoids exist naturally in the form of rhamnoglucosides and are key components to pharmaceutical, cosmetic and nutraceutical formulations offering beneficial effects on health such as anti-inflammatory and anti-carcinogenic properties. In this study, an intracellular alpha-L-rhamnosidase from a microbe isolated from the slug gut microbiota was recombinantly expressed in *E. coli* and purified to homogeneity using His-Tag affinity chromatography. Using SDS-PAGE and Western Blotting, a monomeric protein with a molecular mass of 64 kDa and a computed theoretical isoelectric point of 5.53 was discovered. When tested against 4 nitrophenyl rhamnopyranoside, the isolated protein showed 100 percent relative activity in selectivity with Km = 0.8562 mM 4-nitrophenyl rhamnopyranoside, Vmax = 4.675 M/min, using the Michaelis-menton model of enzyme kinetics. The enzyme was most active when the pH was between 4.95 and 5.47 at an optimum temperature between 45 and 50 degrees Celsius. The enzyme's functionality during an 8-hour period indicated continuous de-rhamnosylation, suggesting that the enzyme exhibits properties of thermally stability and therefore a potentially good candidate for industrial de-rhamnosylation. Our attempt to improve enzyme functionality through enzyme immobilisation using nanoparticles yielded inconclusive data due to enzyme-nanoparticle bio-incompatibility.

3.1 Introduction

α-L-rhamnosidase (EC 3.2.1.40), is involved in the hydrolysis of glycosidic bonds in rhamnoglucosides. GH28 and GH106 families containing α-L-rhamnosidases are involved in pectin metabolism whereas GH78 family of α-L-rhamnosidases are involved in more valuable activities in natural product synthesis (Guillotin *et al.*, 2019). According to Mutter *et al.*, (1994) GH78 α-L-rhamnosidase catalyse the hydrolysis of α-L-rhamnosyl-linkages in L-rhamnosides which include flavonoid glycosides such as naringin, rutin, quercitrin, hesperidin, dioscin, terpenyl glycosides and many other natural glycosides containing terminal α-L-rhamnose (Young, Johnston and Richards, 1989). Several industrial processes including the production of food additives (Giavasis, Harvey and McNeil, 2008), de-bittering of grapefruit juice following the hydrolysis of naringin (Puri and Banerjee, 2000; Puri, Kaur and Kennedy, 2005; Puri *et al.*, 2008), enhancement of aroma in fruit juice and wine making (Manzanares *et al.*, 2003) all involve the use of α-L-rhamnosidase, thus signifying the industrial relevance of this enzyme. The success of grape and yeast glycosidases have been described to be insufficient, prompting the need for glycosidases which can process aromatic precursors completely as part of the fermentation process in wine making (Rodriguez *et al.*, 2010). This places α-L-rhamnosidase in a good position to meet industrial de-rhamnosylation of rhamnoglucosides towards improved bioavailability of flavonoids. According to Fujimoto *et al.*, (2013), alpha-L-rhamnosidase operate through a single-step mechanism involving the inversion of stereochemistry at the anomeric location, which requires the protonation of the glycosidic oxygen by the general acid and leaving group departure concurrent with a water molecule that has been deprotonated by the general base.

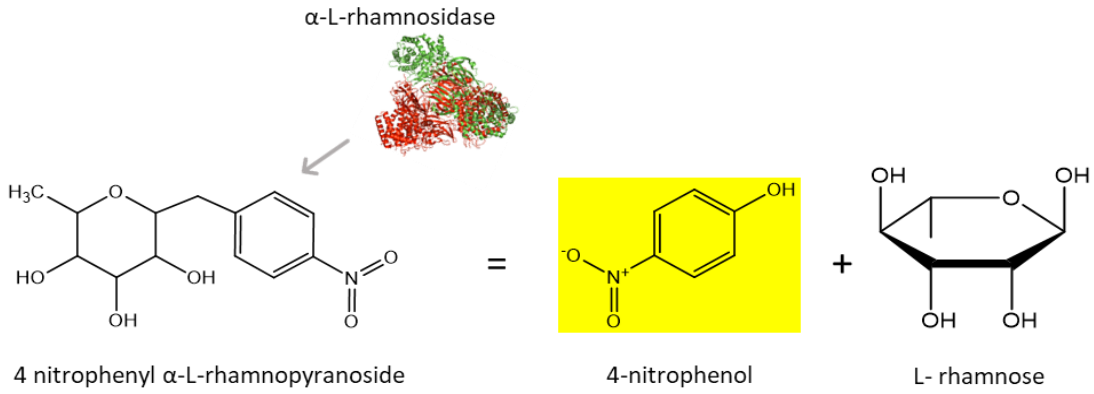


Figure 3.1 Schematic showing alpha-L-rhamnosidase activity on the synthetic substrate 4-nitrophenyl-alpha-L-rhamnopyranoside.

In Figure 3.1. Schematic representation of alpha-L-rhamnosidase acting on a synthetic substrate to produce 4-nitrophenol and a rhamnoside. Synthetic substrates such as 4-nitrophenyl α-L-rhamnoside depicted above, have been used as an easy and rapid characterisation of biochemical activity of enzymes where the 4-nitrophenol released was directly proportional to the glycoside. Such methods are used as a preliminary approach prior to testing on natural substrates.

Flavonoids are secondary plant metabolites that have a significant role in human diets (Middleton, 1998; DuPont *et al.*, 2004). Flavonoids are chemical compounds found in many vegetables and fruits as well as in products such as wine and tea (Panche, Diwan and Chandra, 2016). Flavonoids exist naturally in the form of rhamnoglucosides (Mueller *et al.*, 2018) and are considered to be key components to pharmaceutical, cosmetic and nutraceutical formulations offering beneficial effects on health such as anti-inflammatory and anti-carcinogenic properties (Panche, Diwan and Chandra, 2016; Ruprecht *et al.*, 2019). In the food industry, especially the wine and citrus fruit juice sector, de-bittering is essential in prolonging the shelf life and also improving on customer satisfactoriness (Purewal and Sandhu, 2021). De-bittering of grapefruit juices by hydrolysis of naringin has been described in a study by (Puri and Banerjee, 2000; Puri, Kaur and Kennedy, 2005; Puri *et al.*, 2008). In wineries a common current practice involves the addition of alpha-L-rhamnosidase during or after fermentation to aid in the processing of aromatic precursors (Rodriguez *et al.*, 2010). In the pharmaceutical industry, de-rhamnosylation of rutin to produce Isoquercitrin is a

vital step in the production of quercetin, a flavonoid with anti-inflammatory properties (Yadav, Yadava and Yadav, 2017). Due the natural occurrence of flavonoids as rhamnoglucosides, de-rhamnosylation is essential in increasing the bioavailability of flavonoids for downstream applications. To the best of our knowledge, majority, if not all of the alpha-L-rhamnosidases used in industry are from fungal strains most notably *Aspergillus niger* (Manzanares *et al.*, 2000; Rodriguez *et al.*, 2010) and others such as *A. nidulans* (Orejas, Ibáñez and Ramón, 1999) and *A. aculeatus* (Manzanares *et al.*, 2003). Although alpha-L-rhamnosidase from bacterial strains including thermophilic bacteria (Baudrexl *et al.*, 2019), *Dictyoglomus thermophilum* (Guillotin *et al.*, 2019) have been reported, little is mentioned about their role in industrial de-rhamnosylation.

Aims

The aim of this study was to recombinantly express an α-L-rhamnosidase from a microbe identified from the slug gut metagenome and test its thermal stability and selectivity towards de-rhamnosylation of alpha-L-rhamnosides.

Objectives

1. To amplify using PCR, a DNA sequence which encodes an α-L-rhamnosidase identified from the slug gut metagenome.
2. To clone, recombinantly express, purify and test the biochemical activity of this α-L-rhamnosidase on the synthetic substrate 4 nitrophenyl α-L-rhamnopyranoside.
3. To monitor the enzyme activity of the α-L-rhamnosidase by employing a continuous de-rhamnosylation method to determine thermal stability.

3.2 Methods

3.2.1 Bioinformatics and protein modelling

The amino acid sequence below (**Gene_ID_23621)** belonging to a putative bac rhamnosidase identified from the slug gut metagenome was queried against the NCBI protein database using the BLASTP option and selecting non-redundant protein sequences.

>Gene_ID_23621 amino acid sequence

MSQAIQQHINLNMIRDEHLLAVAEKLIPTVYSREVQPRSIVKSIEDGNVALGWRAEHIADATTFLQQSFSPGDTF
IIDFGEHCVGSLNFSCDSTGSPPDAPAHLQFIFGETLCEVSEPFSEYQGWLSSSWLQQHDEYIDILPEEKTLPRRYC
FRYIKVRVVSLSPKYTIRFTRLSATTVSSAPHSYPEYHTHDALLKKIDEVSVQTLRNCMQNVFEDGPKRDRRLWL
GDLRLQALVNDVTFAKHDLVKRCLYLFAGHRREDGMVAANIFARPEVKADDTYLFDYSLFFVDTLYNYLMAT
QDIATAKELWPVAIRQVELALERCDHTGLVRDSDDWWSFIDWNDSLNKQAPSQGVLIYCVEKAIQLAMICQP
EHGESLQKQLLTLKNAASTLFDEETGYYFSGNKYQFSYASQIWMVLAEMGDLVHRKNIMLNLLKNPPSIEMNT
PYLRHHFIDALIKCGLRNEAIQEIKQYWGAMVEFGADTFWELFDPKNPGFSPYGSKLINSYCHAWSCTPAWFIR
QFGL

The protein sequence was also queried against the protein databank (pdb) and the UniProtKB/Swiss-Prot databank using BLASTP. Multiple sequence alignments of resulting outputs were performed using Clustal OMEGA offered via the EMBL – EBI suit (https://www.ebi.ac.uk/Tools/msa/clustalo/). Protein modelling was achieved by submitting the amino acid sequence to the I-TASSER database for computation. Following bioinformatics and computation, protein models were predicted and the model with the highest Cscore value selected after clustering of decoys based on pair-wise structure similarity by SPICKER (Zhang and Skolnick, 2004). Protein alignment was performed through TM- align and biological annotations of the query protein sequence were then achieved using COFACTOR (Roy, Yang and Zhang, 2012; Zhang, Peter L. Freddolino and Zhang, 2017)and COACH (Yang, Roy and Zhang, 2013c, 2013b) based on the I-TASSER structure prediction. Protein predictions based on the EC numbers were retrieved and compared in multiple sequence alignment as well as modelling analysis using PyMol (PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

3.2.2 PCR

DNA sequence of 1570bp belonging to GH78 alpha-L-rhamnosidase was amplified using the slug gut genomic DNA as template. Amplification was achieved using a forward primer 5′-CACCATGTCTCAAGCAATCCAAC-3′ and a reverse primer 5′-GAGACCGAACTGGCGAATAAAC-3′ designed using Primer3 plus (https://www.primer3plus.com/index.html) and submitted for synthesis (Eurofins Genomics, Germany GmbH). PCR was performed in a 25 µl reaction volume using 100ng of start DNA template, 1.25 µM of each primer, 12.5 µl of Q5® High-Fidelity 2X Master Mix (NEB, UK) and supplemented with nuclease free water resulting in a 25 µl total reaction volume. The reaction mixture was incubated in a Biorad T-100 thermal cycler (Biorad, UK) at 98°C in an initial denaturation step for 30 seconds for 1 cycle, 98°C in a denaturation step for 30 seconds, 67°C in an annealing step for 30 seconds, 72°C in an extension step for 45 seconds all in a total of 35 cycles, followed by a final extension step at 72°C for 7 minutes.

3.2.3 Agarose gel electrophoresis to assess the integrity and determine yield of PCR products

Following amplification of the gene of interest, 10µl of PCR products were analysed to assess integrity of DNA on a 1% agarose gel supplemented with 1X concentrated SYBR™ Safe DNA Gel Stain (Fisher Scientific, UK). Gel images were captured using the G: Box (Syngene) with default capture settings allowing automatic exposure of areas under the UV. Once the correct size of PCR product and DNA integrity was checked, the PCR reactions were scaled up to 75 µl total reaction volumes and the amplimers cleaned up using Isolate II PCR and Gel kit (Bioline, UK). The integrity of the amplified product was finally determined by analysing the samples on a 1% agarose gel alongside the control PCR products for the zero blunt vector which was used as a quality check of blunt ended products. Following the confirmation of products by agarose gel, the products were also subjected to a further check to

determine accurate orientation of produced target sequence by sanger sequencing offered via (Eurofins genomics, Germany GMbH).

3.2.4 Cloning and transformation of PCR products into *E. coli* one shot top 10 cells

Generated blunt ended products were cloned into the Zero Blunt® TOPO® vector as a quality check (See protocol for Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen/Thermo fisher scientific, UK). Transformants grown overnight were harvested and plasmid DNA isolated using the Isolate II plasmid mini kit (Bioline, UK). Restriction endonuclease digest was performed using *EcoR I*, and the presence of cloned amplimers confirmed by analysing the digested products on 1% agarose gel supplemented with 1X SYBR™ Safe DNA Gel Stain (Fisher Scientific, UK). Following successful confirmation of blunt-end pcr products, amplimers were then cloned into the pENTR™/SD/D-TOPO® vector following manufacturer's instructions (Invitrogen™ Gateway™, Thermo Fisher scientific, UK).

3.2.4.1 Transformation and confirmation of entry clones

From the pENTR™/SD/D-TOPO® TOPO cloning reaction mix, 3µl was transferred into a vial containing 25 µl of One shot® chemically competent *E. coli* TOP10 cells and then incubated on ice for 30 minutes followed immediately by heat shocking at 42°C for 30 seconds. The reaction mixture was then allowed 2 minutes of recovery and then 125 µl of Super Optimal broth with Catabolite repression (SOC) medium pre-warmed to room temperature added. The vials were then transferred to a shaking incubator at 37°C for 1 hour with 225 revolutions per minute (rpm) before plating out 25 µl and 100 µl of the transformation reaction onto two different standard 90 mm petri dishes using a sterile L-shaped spreader (Fisher scientific, UK). Plates were then incubated at 37°C for 16 hours and positive clones were selected through antibiotic resistance offered through the kanamycin gene (Invitrogen™ Gateway™, Thermo fisher scientific, UK). Successful transformants were incubated overnight and then plasmid extraction was performed using the Isolate II plasmid mini kit (Bioline, UK). Isolated

DNA was quantified using Nanodrop 2000 spectrophotometer to determine the purity and amount of DNA isolated from cells. For the confirmation of inserts, double digests were performed using *Asc I* and *Not I* – High Fidelity (HF) restriction endonucleases and the products were analysed on a 1% agarose gel as described previously. The orientation of insert was confirmed following sanger sequencing of the isolated plasmid where multiple sequence alignments of resulting outputs were performed using Clustal OMEGA offered via the EMBL – EBI suit (https://www.ebi.ac.uk/Tools/msa/clustalo/). See appendix section 7.2

3.2.4.2 Expression vector generation via LR recombination (Entry vector + Destination vector)

The isolated plasmid DNA of a successfully transformed entry clone was combined with the expression vector pDEST 42 via LR recombination allowing site-specific recombination of entry clone with (attL) sites and destination vector with (attR) sites and the replacement of the toxic gene (ccdb) with the gene of interest, thus creating an expression clone and a by-product containing the toxic gene. The clone was transformed into 25 µl of One shot® chemically competent BL21 (DE3) *E. coli* cells and the protocol for transformation followed as per manufacturer's protocol. (See LR recombination using the pET-DEST42 Gateway™ Vector (Catalog no. 12276-010 , Invitrogen, UK). From the LR reaction transformation, 25µl and 100 µl of transformation reaction mixture was spread on a LB/ agar plate containing 100 µg/ml of the penicillin-type antibiotic (ampicillin) and plates incubated at 37°C for 16 hours. Clones which did not have the desired product for expression were killed using negative selection with the ccdB gene present in the destination plasmid vector.

3.2.5 Recombinant protein expression

After 16 hours of incubation, 3 clones were randomly selected and placed in different Universal containers with 10 ml of fresh LB supplemented with 100 µg/ml ampicillin and incubated at 37°C for 1 hour where cells reached mid log phase of 0.586 when measured at OD595nm. In an Eppendorf tube (Starlab, UK), 1 ml of culture was

collected from each tube and centrifuged at 4°C at 16,300 x g for 1 minute using Eppendorf Refrigerated Centrifuge 5427R (Fisher Scientific, UK) to achieve pellets which were stored as T = 0 (i.e., no IPTG induction). The remaining culture media were split into 2 equal volumes and 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) added to trigger transcription of the lac operon present in pDEST 42 allowing the recombinant expression of protein. Expression was carried out at two different temperatures (25°C and 37°C) to determine the optimum temperature for expression. Following IPTG induction, cells were harvested after 2 hours and then 4 hours by centrifugation at 4°C at 16,300 x g for 1 minute and stored at -20°C prior to analysis.

3.2.6 Cell lysis

The stored cell pellets from each time point were then allowed to thaw at room temperature followed by the addition of 200 µl of lysis buffer known as BugBuster® Master Mix (Merck Millipore, UK) to release soluble protein. Lysates were incubated at 25°C with gentle shaking for 1 hour and after cells were intermittently sonicated at 35% amplitude for 2 minutes on ice to achieve complete lysis of cells. Soluble proteins in the supernatant were then harvested from the cells by centrifugation at 4°C at 16,300 x g for 15 minutes.

3.2.7 SDS-PAGE and Western blotting

Harvested protein samples were prepared for SDS-PAGE and western blot analysis by mixing samples with a 2X sample loading buffer (Laemmli, 1970). SDS-PAGE and western blotting was performed using a modification from (Yang and Mahmood, 2012) and Bio-Rad (Biorad SDS-PAGE protocol) to analyse cell lysates. Protein bands were confirmed after immobilisation onto Amersham™ Hybond® P Western blotting membranes, PVDF (Merck, UK) using Monoclonal Anti-6X His tag antibody produced in mouse (Merck, UK) and Goat Anti-Mouse IgG Antibody, HRP conjugate (Merck, UK).

3.2.8 Fast Protein Liquid Chromatography (FPLC)

Protein purification was performed using HisPur™ Cobalt Chromatography cartridges (1ml) specific for his-tagged proteins (Fisher scientific, UK) through the ÄKTA start chromatography system (GE healthcare, UK). Generated chromatograms after purification runs served as guide to test fractions with peak UV absorbance at (A280nm). Fractions were analysed by SDS-PAGE and Western blotting again to confirm the presence of distinct monomeric bands. Once fractions were tested and confirmed at expected molecular weight, fractions were pulled together, and dialysis carried out using snakeskin dialysis tubing (Thermo Scientific™ SnakeSkin™ Dialysis Tubing, 3.5K MWCO, 35 mm (fisher scientific, UK). Protein samples were assessed again by SDS-PAGE and western blotting after dialysis confirming quality of purified protein. Finally, proteins were concentrated using Merck Amicon™ Ultra-2 Centrifugal Filter Units (Merck, UK) and stored at -20°C until ready to use in enzymatic assays.

3.2.9 Biochemical activity testing of an a-L-rhamnosidase GH78 enzyme isolated from the slug gut microbiota

3.2.9.1 Protein quantification by BSA microtiter plate method

Following purification, dialysis and confirmation of pure protein, protein amounts were determined by Bradford assay (Bradford, 1976) using the microtiter plate method (Biorad). The linear range of quantification was determined (1.25 μg/ml – 10 μg/ml) using different dilutions bovine serum albumin (BSA) and protein amounts were estimated using the slope of the line given as ($y = mx + c$).

3.2.9.2 4 Nitrophenol standard curve generation

For biochemical testing, aryl glycosides conjugated to the chromogenic side (4-nitrophenol) were used in enzymatic assays to determine substrate specificity of enzymes. A standard curve for 4 nitrophenol was constructed by preparing dilutions of (0.0012 – 0.02 μmol) from a 10mM stock of 4-nitrophenol using 2% Na_2CO_3 as

diluent. To a 96 well plate, 200 µl of the different dilutions were transferred in triplicates and absorbance readings at OD (405nm) measured using the Varioskan LUX Multimode Microplate Reader (ThermoFisher, UK). The fit results were retrieved applying linear regression analysis and product amounts estimated using the slope of the line (y = mx), where m represents the extinction coefficient for pNP under the conditions mentioned.

3.2.9.3 α-L-rhamnosidase activity testing

An initial stock of 10mM 4-Nitrophenyl α-L-rhamnopyranoside, pNPR (Sigma, UK) was prepared. Enzyme linearity was determined by carrying out a continuous assay method with a diluted amount of the enzyme (0.06mg/ml) and 5mM of pNPR over a period of 20 minutes recording increase in absorbance over time. A linear relationship was determined from a plot of activity (µmole of product formed) versus time. Once a linear relationship was established, substrate specificity of enzyme was determined using 5 mM concentrations of: 4-Nitrophenyl α-L rhamnopyranoside (PNPR), 4-Nitrophenol xylanopyranoside (PNPX), 4-Nitrophenyl α-L-arabinofuranoside and 4-Nitrophenyl glucuronide (PNPG). Enzyme kinetics relating to km and vmax were determined by plotting the reaction velocity against substrate concentration. Optimum conditions for enzyme activity were determined over a range of temperatures and pH and determined optimum conditions were employed in further activity tests over a period of 8 hours.

3.3 Results

3.3.1 Bioinformatics and protein modelling

Metagenomic data from a previous study by (Joynson *et al.*, 2017) was retrieved from EBI under the project number PRJEB21599 and a sequence belonging to an alpha-L-rhamnosidase was selected for bioinformatic analysis. Initial bioinformatics and protein modelling was carried out querying the retrieved sequence on The National Center for Biotechnology (NCBI) database and I-TASSER database. Highly similar sequences from our BLASTP search against the non-redundant, UniProt and PDB databases were aligned to determine homology between our selected sequence and sequences available on the databases mentioned. A table of the results used in the sequence alignment has been displayed with the accession numbers, scientific name, percentage identity and E-value which was used as a score of quality of alignment.

Table 3.1 Sequences producing significant alignment (non-redundant protein database) from BLASTP

Accession number	Percentage Identity (%)	E-Value	Scientific Name
WP_134183402.1	100.00	0.0	*Buttiauxella sp.* BIGb0552
WP_142488896.1	70.69	0.0	*Leclercia*
WP_024913484.1	69.16	0.0	*Chania multitudinisentens*

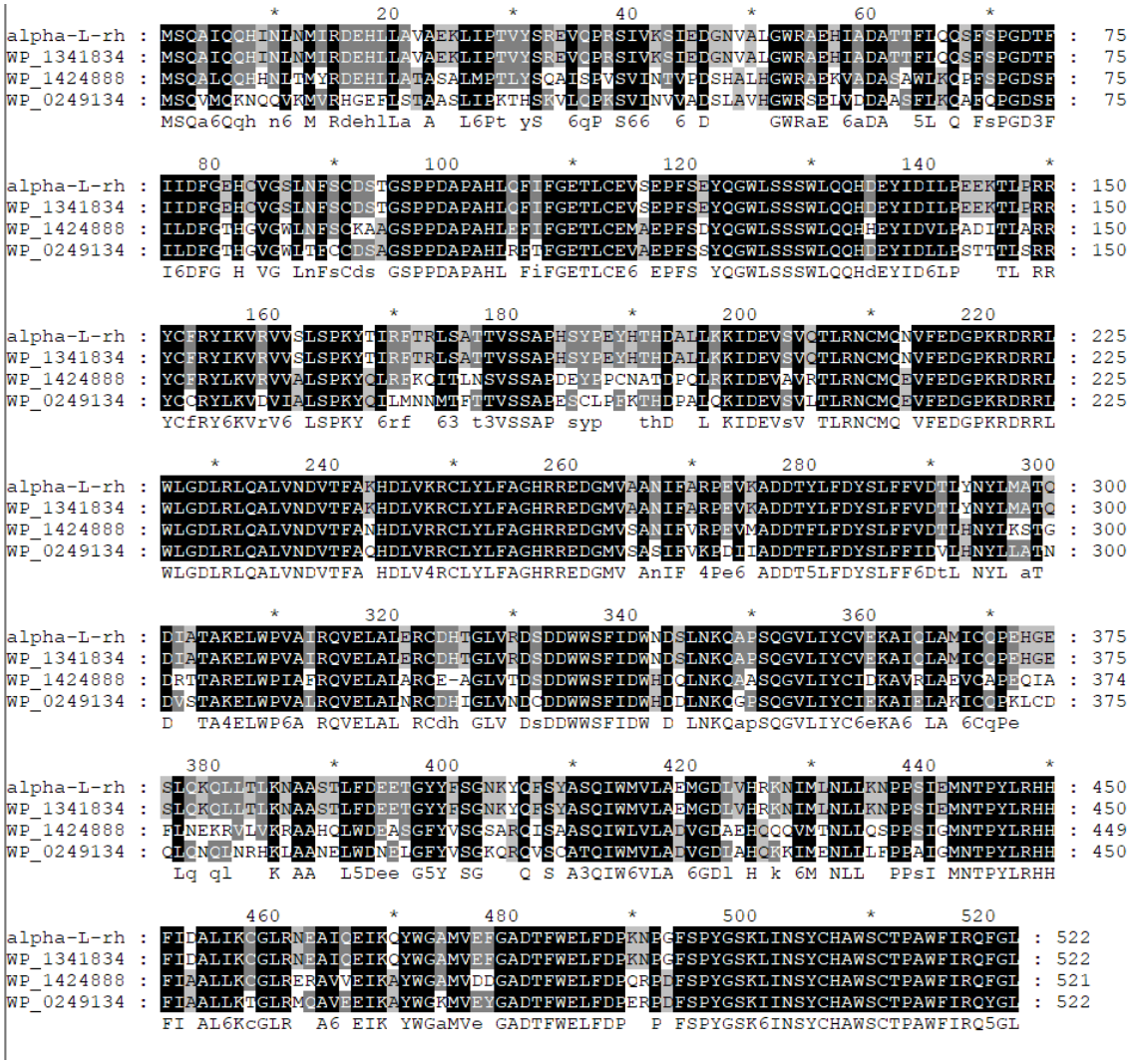


Figure 3.2 Multiple sequence alignment of the alpha-L-rhamnosidase sequence and the top 3 hits following query against non-redundant protein database

In Figure 3.2 shows multiple sequence alignment results using Clustal Omega where the output shows an alignment between our sequence and top hits from the non-redundant protein database search. Identified top hits were compared through sequence alignment where we observed that WP_134183402.1, a family 78 glycoside hydrolase catalytic domain (*Buttiauxella sp*. BlGb05) had 100 % identity and an E-value = 0.0 with our query protein sequence. As the sequence-based gene search approach was used in this study, homology was very crucial in determining putative activity of the selected target sequences. BLASTP search against the Uni-Prot database resulted in no significant similarity, so a BLASTP search was performed against the pdb database.

Table 3.2 Sequences producing significant alignment (pdb database) from BLASTP

Accession number	Percentage Identity (%)	E-Value	Scientific Name
4XHC_A	63.48	0.0	*Klebsiella oxytoca*
2OKX_A	24.78	2e-22	*Bacillus sp.GL1*
3CIH_A	23.76	2e-29	*Bacteroides thetaiotaomicron VPI-5482*

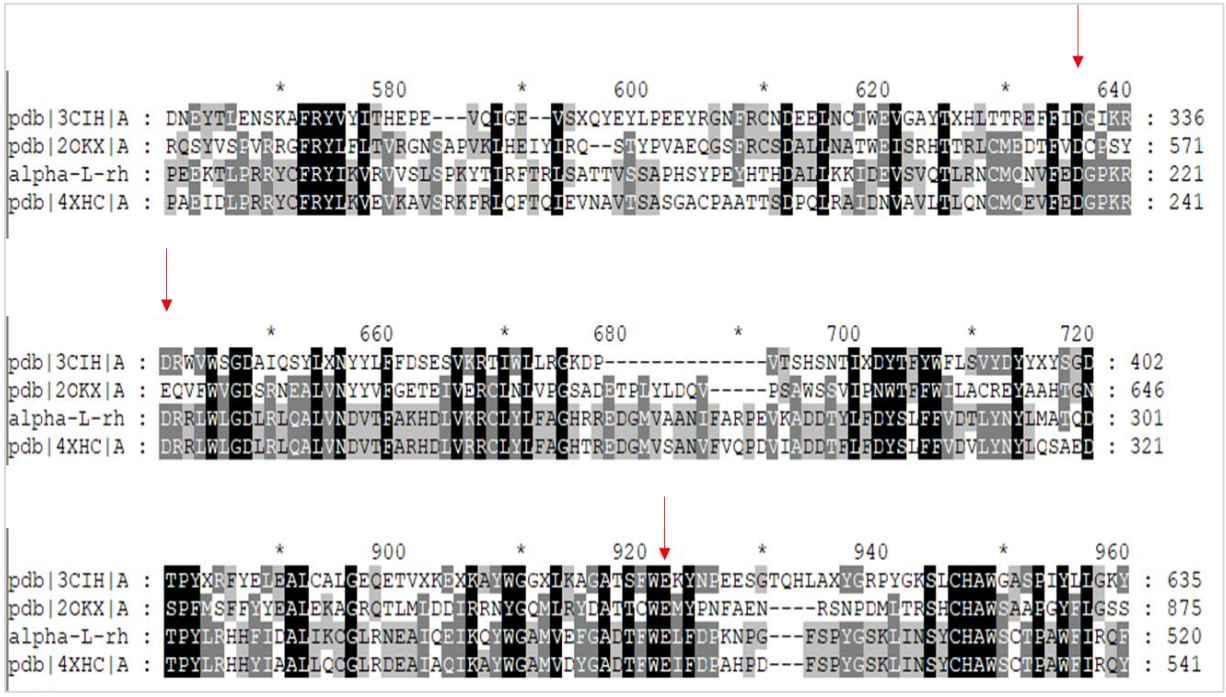


Figure 3.3 Multiple sequence alignment of our alpha-L-rhamnosidase sequence and top hits following query against pdb database

In Figure 3.3 is the multiple sequence alignment output with top pdb hits 4XHC_A, Chain A domain of an Alpha-L-rhamnosidase from *Klebsiella oxytoca* with an E-value of 0.0, 2OKX Crystal structure of GH78 family rhamnosidase of *Bacillus SP. GL1 AT 1.9 A* with an E-value of 2e-22 and 3CIH Crystal structure of a putative alpha-rhamnosidase from *Bacteroides thetaiotaomicron*. Indicate with red arrows is the position of negatively charged residues identified in pdb 2OKX_A, such as Asp567, Glu572, Asp579, and Glu841, conserved in GH family 78 enzymes known to interact with rhamnose (Cui *et al.*, 2007). These amino acid residues were conserved in the sequences that were aligned from our top hits following query against pdb database thus increasing confidence in the predicted functionality of the protein.

Table 3.3 Protein model scoring table from I-TASSER

Protein model	Cscore (measure of quality for predicted models)
Model 1	1.20
Model 2	-0.62
Model 3	-2.89

From the protein model scores from the table above, model 1 was selected due to the high Cscore which meant high quality for that particular model. The selected model was then used in a structure alignment as seen in Figure 3.4 Protein model with highest Cscore was analysed in pymol as seen illustrated in Figure 3.4. Protein models were assessed based on Cscore and the model with the highest Cscore which indicated more confidence was selected and used in structural alignment. A confidence score, or C-score, is used by I-TASSER to gauge the quality of predicted models and this aided in the selection of the protein model we used in the structural alignments.

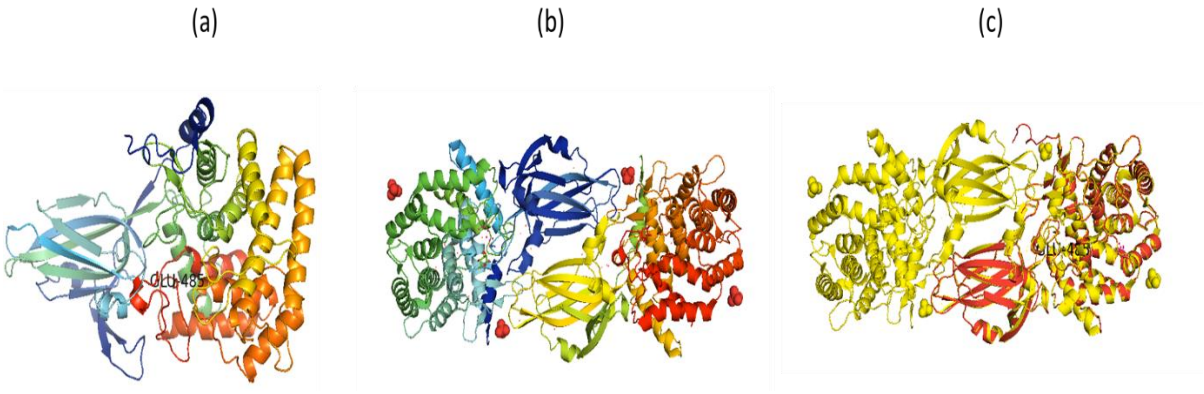


Figure 3.4 Structural alignment between protein 4xhc and our protein model

In Figure 3.4 (a) Protein model 1 of the recombinantly expressed alpha-L-rhamnosidase (b) Protein model of PDB top hit 4xhc_A. (c) Protein structural alignment using the protein model of alpha-L-rhamnosidase (red) and 4xhc_A (yellow). From the alignment close similarities were identified between some domains where the catalytic residues of alpha rhamnosidase enzymes are found albeit the larger size 4xhc_A model. The active sites of our enzyme were identified and compared to that predicted by (O'Neill *et al.*, 2015), where Asp222 was predicted to be the likely catalytic sites for rhamnosidase enzymes, thus suggesting that negative residues such as Aspartic acid are responsible for rhamnosidase activity.

3.3.2 PCR and Restriction digest analysis

Several copies of the target DNA sequence were produced using PCR and the product was cloned into our expression vector system with quality checks carried out prior to recombinant protein expression. In the agarose gel images in Figure 3.5, products generated from PCR were assessed for integrity which confirmed the presence of single discrete bands. The displayed restriction digest results were performed after transformation into entry clone and destination vector (Gateway cloning technology, ThermoFisher Scientific, UK) as a quality check before recombinant protein expression.

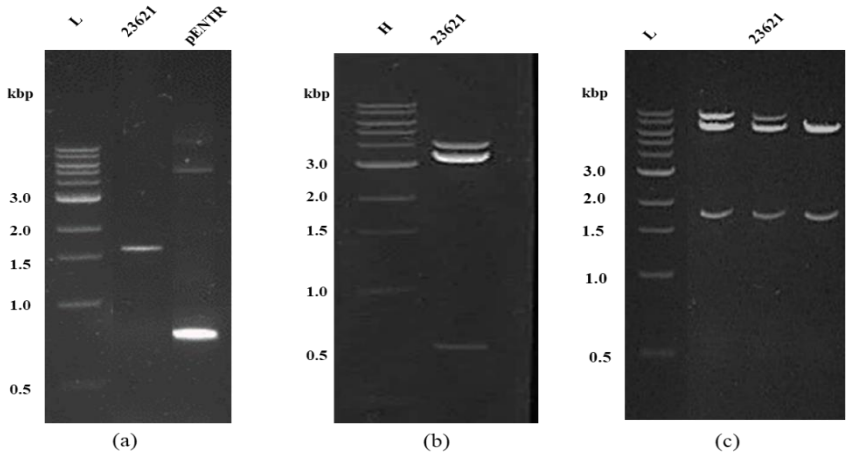


Figure 3.5 Agarose gel images of PCR and restriction digest

In Figure 3.5 (a), successful amplification of a 1570 bp fragment from the whole genome amplified (WGA) metagenomic library as seen in lane 2 and in lane 3, the successful amplification of the control template DNA for pENTR/SD/D-TOPO vector. In (b) restriction digest results confirming inserted pcr product post cloning into the entry vector. In (c) restriction digest showing successful recombination of our entry vector and the confirmation of gene of interest in the expression vector (pDEST-42). In (a, b and c) lane 1, 1 µl of (Quick-Load® Purple 1 kb Plus DNA Ladder) was loaded as DNA marker and in subsequent lanes 10 µl of samples mixed with Gel Loading Dye, Purple (6X) were loaded. Further to the successful amplification of products, sanger sequencing was also employed in this study to check for accurate orientation of the produced target sequence. See appendix section 7.2

3.3.3 Analysis of expressed a-L-rhamnosidase using SDS-PAGE and Western blot

The expression of α-L-rhamnosidase was detected in cell lysate after the expression of recombinant protein in BL21 DE3 *E. coli* cells. Recombinant protein expression was carried out at 25°C and 37°C following induction with 1mM IPTG. The supernatant of the harvested cell lysates was retrieved after centrifugation and analysed via SDS-PAGE and western blotting to determine optimum expression conditions.

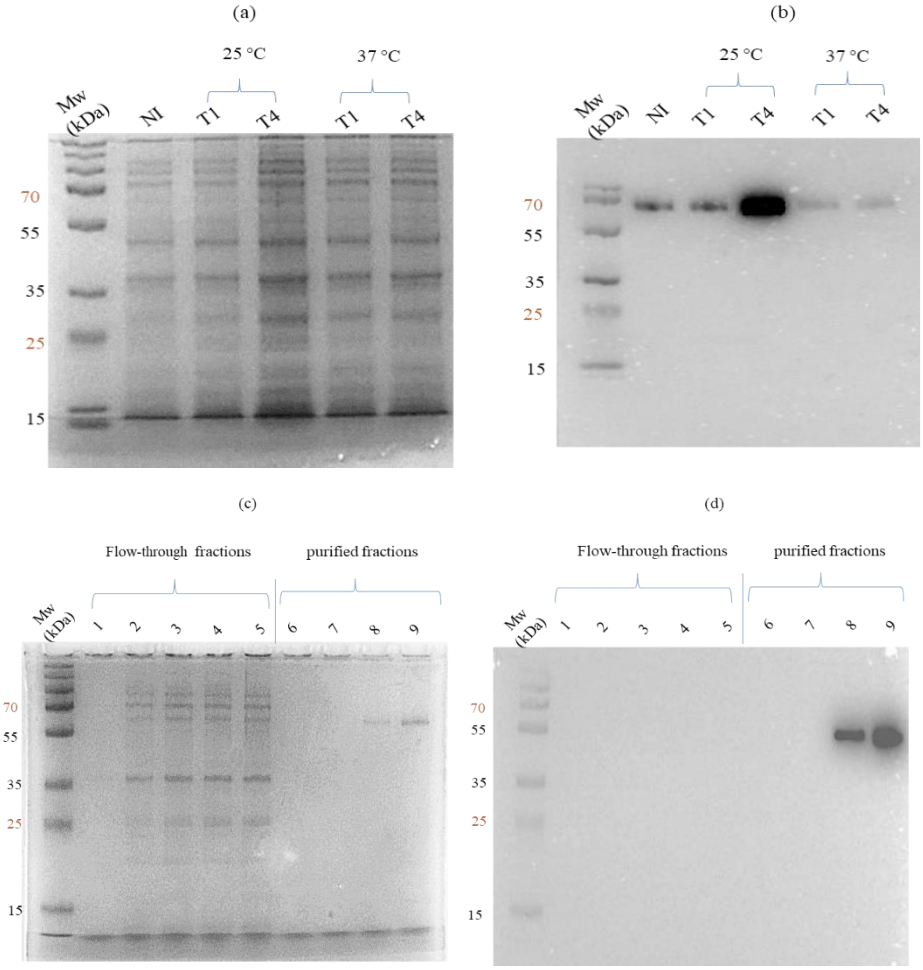


Figure 3.6 SDS-PAGE and Western blot image of expressed protein

Figure 3.6 (a) SDS-PAGE of a positive clone after time point expression using 1mM IPTG as inducer. (b) Western blot showing a monomeric protein band ~ 64 kDa (size of protein + polyhistidine tag and V5 epitope of the pDEST-42 vector which is 4 kDa). (c) SDS-PAGE after purification. (d) Western blot of purified protein samples confirming discrete monomeric protein band at ~ 64 kDa. SDS-PAGE gels were stained using InstantBlue™ Protein Stain (expedeon, UK) and the duplicate gel was analysed via western blot using specific antibodies for the his-tagged protein. For the protein marker, 2.5 µl of Pageruler plus prestained (ThermoFisher, UK) was loaded, and then in subsequent wells, 20 µl from samples mixed with Laemli sample loading buffer to achieve 1X. From the SDS-PAGE results, more

soluble protein was observed in the time point expression after 4 hours at 25°C fractions compared to 37°C . Optimal expression conditions at 25°C with 1 mM IPTG induction for 4 hours was concluded based on the SDS-PAGE and western blot results and protein dialysis was then performed to desalt protein before activity testing as seen in Figure 3.7 below.

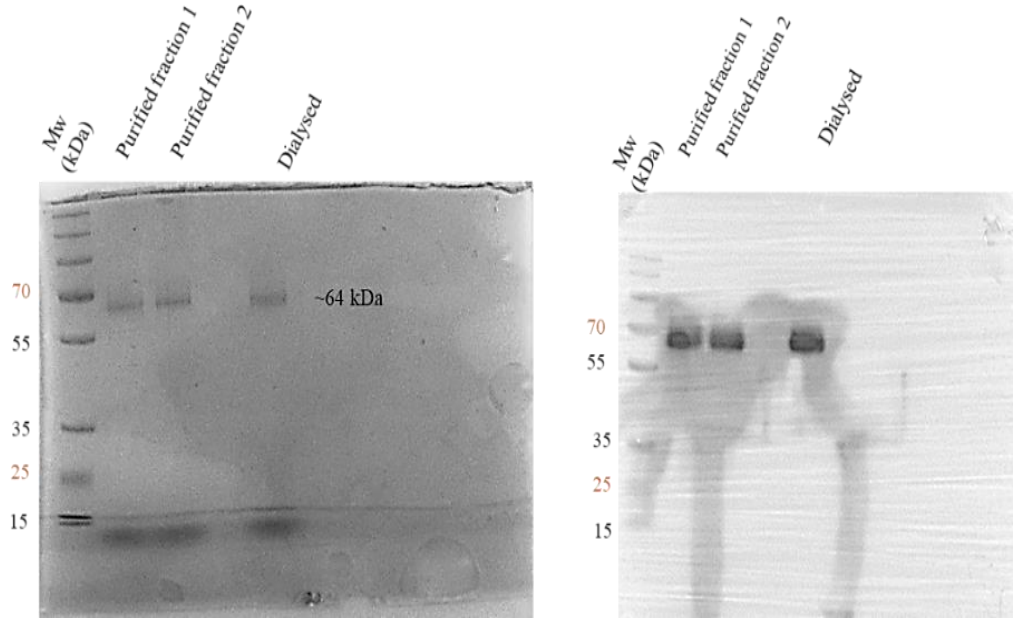


Figure 3.7 SDS-PAGE and Western blot of proteins post dialysis

In Figure 3.7 shows successful pool of protein fractions from Figure 3.6 subjected to dialysis to remove imidazole and desalt protein. Dialysed samples were run alongside aliquots from the purified fractions on SDS-PAGE and then confirmed by western blotting. From the SDS-PAGE (left) and western blot image (right) we successfully conclude the presence of purified protein after 16 hours of dialysis at 4°C.

3.3.4 Biochemical activity testing of a GH78 α-L-rhamnosidase isolated from a microbe identified in the slug gut microbiota

3.3.4.1 Protein quantification by microtiter plate assay using BSA

A standard curve was generated to estimate protein concentration within the linear working range of 0.125 mg/ml – 1 mg/ml using BSA as standard. The amount of protein following absorbance measurement at 595 nm was estimated using the linear regression equation deduced from the plot of absorbance at OD595nm versus concentration (mg/ml).

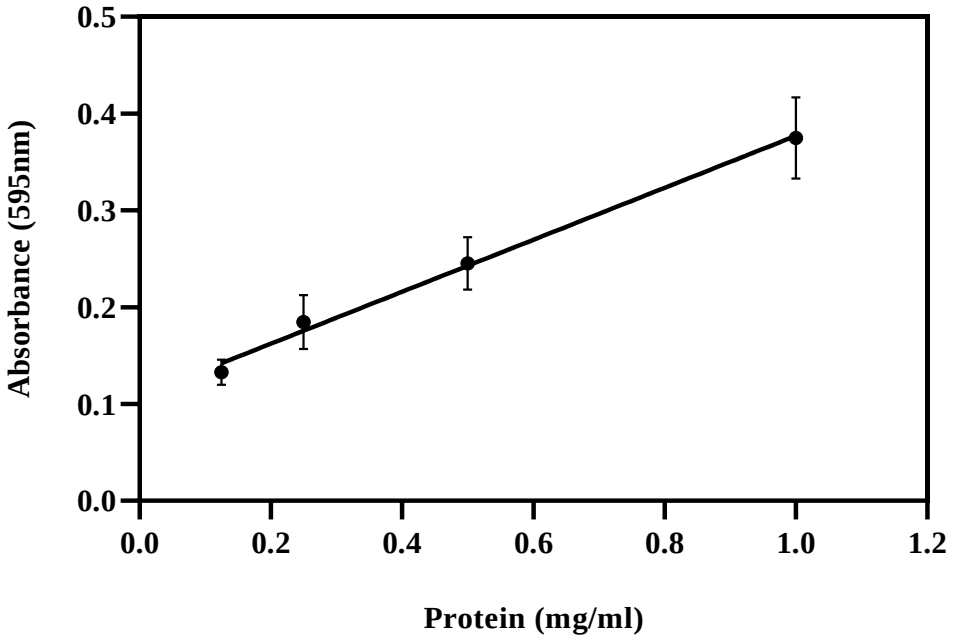


Figure 3.8 BSA standard curve generated for protein quantification

In Figure 3.8 is a standard curve used in this study for quantifying protein amounts using the standard procedure for protein quantification. The linear range which gave the line of best fit for the assay was determined between 0.125 mg/ml– 1.0 mg/ml with y= 0.0286x+0.1086 (R^2= 0.9946) using BSA as standard. Standard deviation (*n*=3) was calculated and has been indicated as positive and negative error bars.

3.3.4.2 Substrate quantification using 4-nitrophenol as a standard

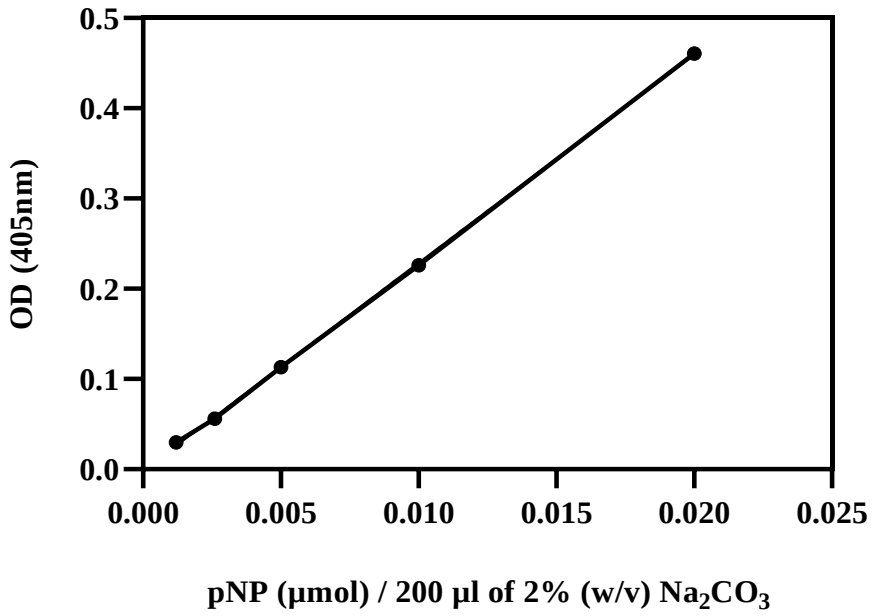


Figure 3.9 Standard curve for 4-nitrophenol

Figure 3.9 Standard curve for pNP in the linear range of (0.001 – 0.02 µmoles) in 200 µl of 2% (w/v) Na_2CO_3 goodness fit of (R^2=0.9998) using five dilutions of 4 nitrophenol. Readings were measured in triplicates, average determined and average values minus blank plotted against amount of pNP. Standard deviation (*n*=3) was calculated and has been indicated as positive and negative error bars. Since 4-nitrophenol conjugated substrates were used in this study to measure substrate conversion, the colour intensity measured was inferred to be directly proportional to product formation. Product formation calculations following enzymatic assay were achieved using (y = 23.04x) where the gradient (m= 23.04) is the molar extinction coefficient.

3.3.4.3 Alpha-L-rhamnosidase activity testing

Enzyme activity was recorded against different substrates to determine substrate specificity before determining optimum conditions for the enzyme. 4-nitrophenyl linked substrates were used where our enzyme exhibited highest substrate specificity towards 4-nitrophenyl-L-rhamnopyranoside. Model of enzyme kinetics that describes how the concentrations of the enzyme and its substrate affect the rate of an enzyme-catalysed reaction known as Michaelis-Menten was used and optimum enzyme activity conditions determined to characterise the enzyme.

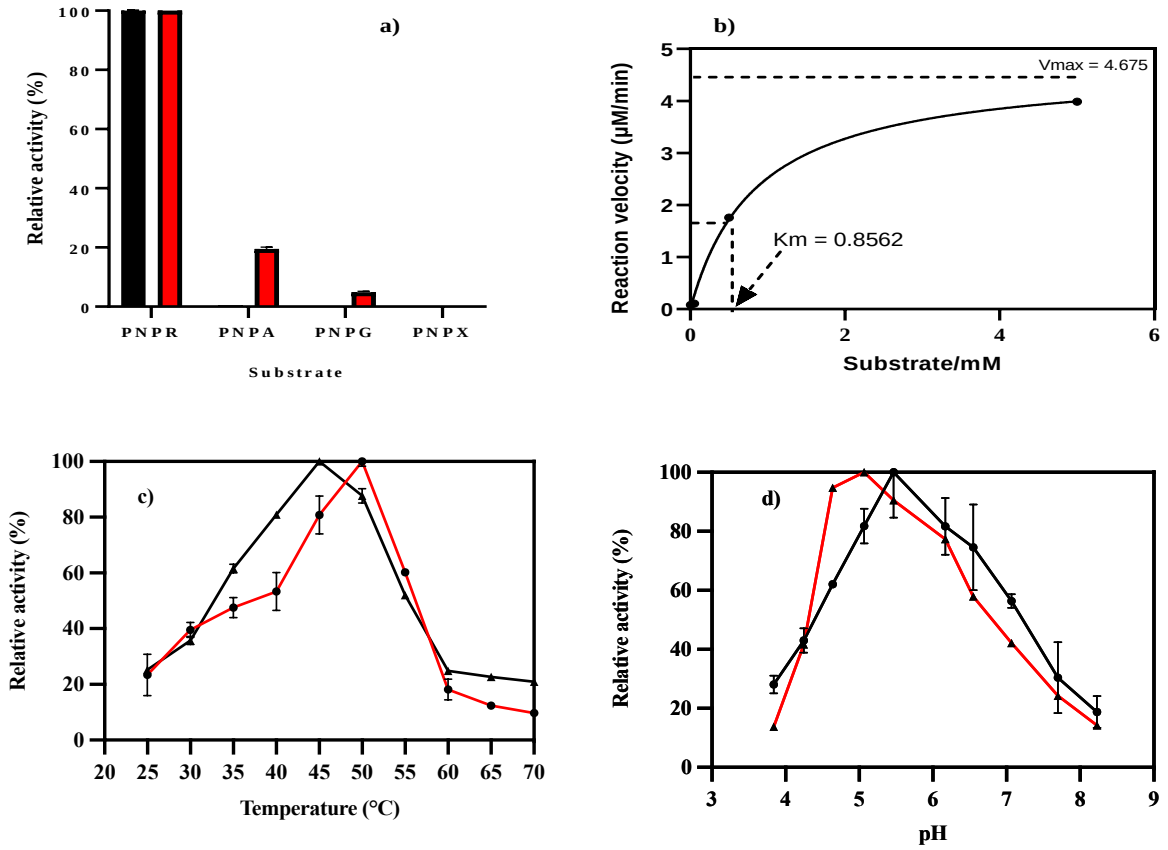


Figure 3.10 Substrate specificity of enzyme and Enzyme kinetics

In Figure 3.10, (a) Substrate specificity testing of alpha-L-rhamnosidase on 4-nitrophenylrhamnopyranoside (pNPR), 4-nitrophenyl arabinofuranoside (pNPA), 4-nitrophenyl glucopyranoside (pNPG) and 4-nitrophenyl xylanopyranoside (pNPX). (b) Enzyme kinetics showing Michaelis-Menten plot which was used in determining substrate concentration at which half of all the active sites of the enzyme are covered by substrate molecules as a measure of enzyme affinity for

substrates, given as Km = 0.8562mM. (c) Temperature profiling over 10 different temperatures to determine optimum temperature where highest enzyme activity was observed between (45 - 50°C) after 20 minutes reaction. (d) pH profiling between a range of acidic to alkaline pH (3.84 – 8.23) to determine optimum pH for enzyme activity. Optimum pH was determined to be pH 4.95 and pH 5.47 using optimum temperatures 45°C and 50°C respectively. Standard deviation (n=3) was calculated and has been indicated as positive and negative error bars. Represented with black lines, 0.06mg/ml of enzyme solution was used in assay. In red lines 0.006 mg/ml of enzyme solution was used in the assay.

Enzyme activity following optimum condition determination was then compared to enzyme activity of the same enzyme from different sources. Table 3.4 compares the specific activity of α-L-rhamnosidases from different sources under different conditions.

Table 3.4 Comparison of biochemical activity to other studies involving microbial enzyme activity

Conditions	Sample	Specific activity (U/mg)
pH 5.47, 50°C in HEPES buffer	Gene_23621 α-L-Rhamnosidase from the slug gut metagenome	~ 159
pH 6.5, 50°C in sodium phosphate buffer	α-Rhamnosidase (prokaryote) from Megazyme (megazyme_alpharhamnosidase)	~190
pH 6.5, 40°C C in sodium phosphate buffer	α-Rhamnosidase (prokaryote) from Megazyme megazyme_alpharhamnosidase	~135
pH 6.5, 65°C in citrate phosphate buffer	*Aspergillus terreus* (Ge *et al.*, 2017)	451.5 ± 12.5

In Table 3.4 the biochemical activity of our expressed α-L-rhamnosidase was compared to a commercially available prokaryotic alpha-L-rhamnosidase (MEGAZYME, Ireland) and a fungal α-L-rhamnosidase expressed from (Ge *et al.*, 2017). The specific activity of the α-L-rhamnosidase from Megazyme was stated as ~190 U/mg achieved under the conditions, pH 6.55, 50°C in sodium phosphate reaction buffer as detailed in the specifications sheet of the manufacturer. When compared to the α-L-rhamnosidase from this study we observed specific activity of ~ 159 U/mg under the conditions of, pH 5.47, 50°C and ~118 U/mg pH 6.55, 50°C when activity testing was carried out using 0.06mg/ml.

According to activity testing carried out by (Ge *et al.*, (2017), specific activity of the purified α-L-rhamnosidase from *Aspergillus terreus* was 451.5 ± 12.5 U/mg using 1mg/ml of the protein.

3.4 Discussion

Biocatalysts such as alpha-L-rhamnosidase will play a crucial role in the flavonoid industry and flavonoids in the form of dietary supplements for example, were in high demand during the COVID-19 pandemic (Mordor Intelligence, 2022). The global flavonoid market is projected to grow at a compound annual growth rate (CAGR) of 3.9% during the forecast period between (2022 -2027). The flavonoid industry is expected to be worth $1.06 billion by 2025 according to data from (Grand View Research, 2022). This signifies the important requirement of biocatalyst to meet the development expected in such industries and hence the growth in interest regarding production of alpha-L-rhamnosidase. In this study our primary aim was to recombinantly express an α-L-rhamnosidase identified from the slug gut metagenome and test its thermal stability. we successfully amplified, using PCR, the coding DNA sequences using genomic DNA obtained from WGA. Amplified sequences were validated to confirm the true existence of such sequences in nature by agarose gel electrophoresis and sanger sequencing and confirmed sequences were then cloned and proteins recombinantly expressed resulting in functional proteins which are harbored in the gut of the slug gut microbiota.

DNA sequence of 1570 bp was subjected to a search via BLAST and presented most significant BLAST hit revealed 100% similarity with an unclassified species of *Buttiauxella sp. BIGb0552* identified in *Caenorhabditis elegans* (Schoch *et al.*, 2020). A thorough search on various publication database revealed no work relating to the activity testing of this species of bacteria as a de-rhamnosylation enzyme. Also, very little research on *Buttiauxella sp.* with rhamnosidase function has been reported making this study a valuable contribution to research. From our protein models, the protein model which had the highest quality score for our predicted model was

selected for structural alignment and structural analysis revealed close similarity to 4XHC_A, Chain A domain of an Alpha-L-rhamnosidase from *Klebsiella oxytoca* with an E-value of 0.0. In addition to the structural alignment, Multiple sequence alignment with 2OKX Crystal structure of GH78 family rhamnosidase of *Bacillus SP. GL1 AT 1.9 A* with an E-value of 2e-22 and 3CIH Crystal structure of a putative alpha-rhamnosidase from *Bacteroides thetaiotaomicron and* Indicated in Figure 3.3 with red arrows is the position where negatively charged residues, such as Asp567, Glu572, Asp579, and Glu841, conserved in GH family 78 enzymes are known to interact with rhamnose (Cui *et al.*, 2007). Similar residues have been reported to be involved in the activity of rhamnose enzymes where the active site for alpha-L-rhamnosidase (Fujimoto *et al.*, 2013; O'Neill *et al.*, 2015; Guillotin *et al.*, 2019). Therefore, we were able to conclude that the alpha-L-rhamnosidase from this study possibly operates through an inverting mechanism due to the presence of Glu485 in the c-terminus of our amino acid sequence. However, this was not an aim of this study and as a result further experiments were not performed to conclude this.

Following the bioinformatics and structural alignment analysis, the DNA sequence was then successfully amplified as depicted in Figure 3.5 and quality checked through integrity assessment in agarose gels. DNA sequence was then cloned and recombinantly expressed resulting in a monomeric protein with a molecular mass of 64 kDa as observed in Figure 3.6 and Figure 3.7. Activity testing of the monomeric protein on 4-nitrophenyl alpha-L-rhamnopyranoside resulted in 100% relative activity with the enzyme exhibiting optimum activity within a ph. range of (4.97 – 5.47) and a temperature range between (45 -50°C). Enzyme kinetics revealed high substrate affinity for the substrate given by a low km of 0.8562 mM and a vmax of 4.675 μM/min, with specific activity of 67 U/mg and 159 U/mg using a different dilution of the enzyme preparation. We compared our enzyme activity with published data and with a commercial enzyme as seen in Table 3.4, to obtain information on the specific activity of the enzyme from different microorganisms.

From comparing our specific enzyme activity to a commercial α-L-rhamnosidase and an α-L-rhamnosidase from a *Aspergillus terreus*, perhaps our specific activity could have increased more than that described by (Ge *et al.*, (2017) should 1mg/ml of the protein solution have been used. This would have been interesting to see as this would have suggested that an alpha-L-rhamnosidase from prokaryotic source outperforms that from a fungal source. To validate such conclusions requires using the same substrate and enzyme amount as published in (Ge *et al.*, (2017). Nevertheless, this study has proven the possibility of recombinant expression of an alpha-L-rhamnosidase from a prokaryotic source. We have also characterised the biochemical activity of this enzyme and compared it a commercial counterpart. Our results suggests that the performance of the purified α-L-rhamnosidase could match or perhaps outperform the commercial enzyme and fungal derived α-L-rhamnosidase using our optimised reaction conditions and 1mg/ml of the pure enzyme solution.

3.5 Conclusion

This study highlights the possibility of recombinant protein expression of functional glycosyl hydrolases identified from unexplored environments such as that of the gut microbiota of the common black slug. A functional and thermally stable α-L-rhamnosidase from an uncharacterised *Buttiauxella sp. BIGb0552* was produced and tested on a synthetic rhamnoglucoside with findings suggesting that the produced enzyme was thermally stable with continuous de-rhamnosylation over a period of 8 hours. This places the expressed α-L-rhamnosidase enzyme as a good candidate for industrial bioconversion where high temperatures are required to aid in the complete breakdown of substrates towards several applications. Therefore, thermal properties of rhamnosidase to lowering industrial input is essential for its success as the activity of α-L-rhamnosidase is said to have potential use as an environmentally friendly technique to regulate the biological and pharmacological properties of flavonoid molecules such as hesperidin, naringin, diosmin and rutin (Izzo *et al.*, 2014) which can be used in several biotechnological applications in the food, pharmaceutical and

industrial process. From our research, only a limited number of prokaryotic α-L-rhamnosidase have been characterized (Manzaneres *et al.*, 2001). Most commonly available α-L-rhamnosidase are from fungal sources, but this study has revealed that bacteria represent an unexplored environment which harbours prokaryotic enzymes which can act as novel biocatalyst within industrial applications. This research offers first line of evidence towards the production of industrially relevant biocatalysts using prokaryotic expression systems described in this study.

Chapter 4. Cloning, Expression and Biochemical activity testing of a recombinant putative carboxyl esterase 1 (CE1) from the slug gut

Abstract

The recalcitrant nature of biomass hinders accessibility to sugar rich fractions which can serve as base molecules for industrial bioconversion into finished products such as biofuel. Hemicellulose forms the second most abundant component of plant biomass after cellulose with several applications linked to the use of hemicellulose breakdown products such as, nutrient additive in food and feed, drug delivery technology, bioethanol production and pulp and paper industry. To achieve this in a sustainable approach, bioconversion using biocatalysts are favoured as opposed to the use of harsh chemicals especially for pre-treatment of biomass. Hemicellulases, such as feruloyl and acetyl xylan esterases can aid in achieving such targets by targeting fractions of hemicellulose, thus creating more accessibility for other hemicellulases and subsequently leading to the production of various monomers including fermentable sugars. The use of plant biomass and biocatalysts for bioconversion is favoured over the use of harsh chemicals as this will mitigate concerns regarding global CO_2 emissions realised with current industrial technologies which produce petrochemicals used to create innumerable common things, including clothing, tyres, digital gadgets, packaging, detergents, and many more.

In this study, a gene belonging to a putative Carbohydrate Esterase 1 (CE1) was identified from the slug gut microbiota following bioinformatics and was cloned, recombinantly expressed and purified to homogeneity using affinity chromatography. SDS-PAGE and western blotting analysis revealed a monomeric protein with a molecular mass of 36 kDa having a theoretical isoelectric point of 5.80. The purified esterase exhibited highest activity when tested against 4 nitrophenyl acetate and no activity recorded on 4-nitrophenyl xylanopyranoside. Enzyme kinetics revealed a Michaelis-Menten constant, Km = 1.65 mM 4-nitrophenyl-acetate with a

Vmax = 66.81 μM/min, and the enzyme was optimally active at 40 °C, pH 7.5. We conclude from our findings; the slug gut harbours microbial enzymes with CE1 enzyme activity which can aid in the release of free sugars from hemicellulose towards subsequent fermentation into bioethanol. However, our findings suggest that substrate specificity of CE1 family enzymes is crucial for deployment towards industrial bioconversion.

4.1 Introduction

Current industrial conversion of biomass requires the use of harsh chemicals for pre-treatment which renders the process of bioconversion non-sustainable leading to global climate change concerns as a result of greenhouse gas emissions. Such pre-treatment methods including acid hydrolysis using chemicals such as sulphuric acid (H_2SO_4) and hydrochloric acid (Esteghlalian *et al.*, 1997); alkaline hydrolysis using dilute sodium hydroxide (Fan, Gharpuray and Lee, 1987); oxidative delignification using peroxidase enzyme in the presence of hydrogen peroxide (Azzam, 1989) and the organosolv process using methanol, ethanol, acetone, oxalic acid (Chum *et al.*, 1988; Ragauskas *et al.*, 2014)have also been used. Yet, these chemical methods are not cost effective and to reduce costs and move toward a green chemistry approach to processing, biological pre-treatment is favoured. This has thus prompted the need for economically friendly and renewable technology which can replace fossil fuel technologies as liquid fuel for transportation and chemicals for various industrial applications.

Biomass has been described as a very good candidate in replacing fossil fuel and therefore offering a renewable solution to address challenges in energy production particularly, greenhouse gas emissions. After cellulose, hemicellulose makes up the second-highest amount of plant biomass (Razeq *et al.*, 2018), with several applications linked to the use of hemicellulose breakdown products such as, nutrient additive in food and feed, drug delivery technology (Mishra and Malhotra, 2009) and in the pulp and paper industry (Suurnäkki *et al.*, 1997). Aside the broad application of the very

diverse hemicellulose fraction of plant biomass, it still represents an underutilised fraction of plant biomass (Ebringerová, Hromádková and Heinze, 2005). Xylan is considered the most abundant component of hemicellulose and it is believed that the deconstruction of xylan to simple sugars mainly xylose and arabinose, can be used for microbial fermentation to generate bioethanol (Saha, 2003). Bioethanol production from the fermentation of sugars is a well-established technique applied in in the biofuels industry (Sun and Cheng, 2002; Fernandes and Murray, 2010; Chaturvedi and Verma, 2013). Various enzymes are involved in this process including hemicellulases such as esterases. Esterases are part of a group of enzymes which catalyse the cleavage of ester bonds (Peng *et al.*, 2016). Carbohydrate Esterase family 1 (CE1) enzymes include acetyl xylan esterase (EC 3.1.1.72); cinnamoyl esterase (EC 3.1.1.-); feruloyl esterase (EC 3.1.1.73); carboxylesterase (EC 3.1.1.1); S-formylglutathione hydrolase (EC 3.1.2.12); diacylglycerol O-acyltransferase (EC 2.3.1.20); trehalose 6-O-mycolyltransferase (EC 2.3.1.122) that catalyse the deacetylation of xylan and xylo-oligosaccharides (Margolles-Clark *et al.*, 1996; Lombard *et al.*, 2014).

The isolation, expression and characterisation of esterases has been described in several studies including (Chen *et al.*, 2020), with extensive application in various disciplines including synthesis of cosmetics (Jeon *et al.*, 2009), bleaching (Gangwar, Tejo Prakash and Prakash, 2016), pharmaceuticals (Panda and Gowrishankar, 2005) and the pulp and paper industry (Topakas, Vafiadi and Christakopoulos, 2007). Acetyl xylan esterase (EC 3.1.1.72) is an example of hemicellulase acting enzyme which facilitates the action of endoxylanases by increasing accessibility to the xylan backbone by cleaving the ester bonds of acetyl groups (Sundberg and Poutanen, 1991). Another example is feruloyl esterase (EC 3.1.1.73) which catalyses the hydrolysis of the 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar (Faulds and Williamson, 1991). The synergistic action of endoxylanases and esterases will improve the efficient hydrolysis of xylan (Biely and Mackenzie, 1986; Blum *et al.*, 1999). As such,

xylan esterases are essential enzymes for the deconstruction of xylan backbones of hemicellulose.

Aims

In this study the aim was to recombinantly express an esterase identified from the slug gut metagenome and test its operational functionality.

Objectives

1. To amplify using PCR, a DNA sequence which encodes a putative Carbohydrate Esterase family 1 enzyme identified from the slug gut metagenome.
2. To clone, recombinantly express, purify and test the biochemical activity of this esterase on the synthetic substrate 4-nitrophenyl acetate and 4-nitrophenyl xylanopyranoside and determine enzyme specificity.
3. To monitor the activity of the expressed esterase by employing a continuous assay method and determine thermal stability of the expressed esterase.

4.2 Methods

4.2.1 Bioinformatics and protein modelling

The amino acid sequence below belonging to a putative CE1 was queried against the NCBI protein database using the BLASTP option and selecting non-redundant protein sequences.

> Gene_ID_1463 amino acid sequence

MTRDETGVWSWTSTQLQPNLYEYYFDVDGFRSIDTGSRFTKPQRQVNTSLILVPGGILDERAVPHGE
LRTLTYHSSALQAERQVYVWTPPGYKPGGEPLPVLYFYHGFGDTGLSGITQGRIPQIMDNLLAEGKI
KPMLVVVPDTETDIKQAVAENFPPVERRKDFYPLNAAAADKELMEDIIPLISNRFTVRKDAAGRAL
AGLSQGGYQALVSGMSHLDSFGWLATFSGVTTTTVPNERVSRQLAQPQAINQQLRNFTLVVGEKD
GVTGKDIAGLKSQLEQQGVKFSYTSYPGLGHEMDVWRPAYAEFVQR

The protein sequence was also queried against the protein databank (pdb) and the UniProtKB/Swiss-Prot databank using BLASTP option. Protein modelling was achieved by submitting the amino acid sequence to the I-TASSER database for computation. Following bioinformatics and computation, protein models were predicted and the model with the highest Cscore value selected after clustering of decoys based on pair-wise structure similarity by SPICKER (Zhang and Skolnick, 2004). Protein alignment was performed through TM- align and biological annotations of the query protein sequence were then achieved using COFACTOR (Roy, Yang and Zhang, 2012; Zhang, Peter L. Freddolino and Zhang, 2017)and COACH (Yang, Roy and Zhang, 2013c, 2013b) based on the I-TASSER structure prediction. Protein predictions based on the EC numbers were retrieved and compared in multiple sequence alignment as well as modelling analysis using PyMol (PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

4.2.2 PCR

DNA sequence of 936bp was amplified from the slug gut metagenome data using Q5® High-Fidelity 2X Master Mix (NEB, UK). Amplification was achieved using forward primer 5′-CACCATGACTCGCGACGAAAC-3′ and reverse primer 5′-TTACCGCTGAACAAACTCTGC-3′ designed using Primer3 plus (https://www.primer3plus.com/index.html) and submitted for synthesis (Eurofins Genomics, Germany GmbH). PCR was performed in a 25 µl reaction volume using 100ng of start DNA template, 1.25 µM of each primer, 12.5 µl of Q5® High-Fidelity 2X Master Mix (NEB, UK) and supplemented with nuclease free water to achieve a 25 µl total reaction volume. The reaction mixture was incubated in a Biorad T-100 thermal cycler (Biorad, UK) at 98°C in an initial denaturation step for 30 seconds for 1 cycle, 98°C in a denaturation step for 30 seconds, 66/67°C in a gradient annealing step for 30 seconds, 72°C in an extension step for 28 seconds all in a total of 35 cycles followed by a final extension step at 72°C for 7 minutes.

4.2.3 Plasmid construct synthesis

After confirmation of amplimer by agarose gel electrophoresis, a plasmid construct was designed via the GeneArt Instant designer (ThermoFisher Scientific, UK) generating a plasmid construct of 6,693 bp in pET151/D-TOPO vector. The synthesised plasmid construct was then quality checked by restriction digest using *Sac*I and *Apo*I following plasmid isolation using the isolate II plasmid mini kit (Bioline, UK). After successful QC checks, plasmid construct was transformed into BL21 DE3 cells for pilot expression studies as described in earlier sections of this study (3.2.5).

4.2.4 Recombinant protein expression

After 16 hours of incubation, three clones from our transformation plate were selected and placed in different universal containers with 10 ml of fresh LB supplemented with 50 μg/ml carbenicillin. The three positive transformants were incubated at 37°C for 1 hour where cells reached mid log phase of 0.485 when measured at an OD595nm. Into an Eppendorf tube (Starlab, UK), 1 ml of culture was collected and centrifuged at 4°C at 16,300 x g for 1 minute and the pellets stored as T = 0 (i.e., no IPTG induction). The remaining culture media were split into 2 equal volumes and 0.7 mM IPTG added to the tube labelled as Induced (I). Following IPTG induction, cells were harvested after 2 hours and then 4 hours by centrifugation at 4°C at 16,300 x g for 1 minute and stored at -20°C prior to analysis.

4.2.5 Cell lysis

Stored pellets from the time point expression were retrieved and allowed to thaw at room temperature. To each pellet 500 μl of lysis buffer (7.7.2) was added and a -80°C freeze thaw cycle was repeated 3 times, placing tubes on dry ice for 5 minutes and then incubating at 42°C for 5 minutes. After this, sonication at 35% amplitude for 2 minutes on ice was performed to achieve complete lysis.

4.2.6 SDS-PAGE and Western blotting

Harvested protein samples were prepared for SDS-PAGE and western blotting analysis by mixing samples with sample loading buffer (Laemmli, 1970) to achieve a 1X sample. Cell lysates were analyses via SDS-PAGE and western blotting using a modification from (Yang and Mahmood, 2012) and Bio-Rad (Biorad SDS-PAGE protocol). Protein bands were confirmed after immobilisation onto Amersham™ Hybond® P Western blotting membranes, PVDF (Merck, UK) using Monoclonal Anti-6X His tag antibody produced in mouse (Merck, UK) and Goat Anti-Mouse IgG Antibody, HRP conjugate (Merck, UK).

4.2.7 Protein purification

Protein purification was achieved using His Pur gravity columns (Cytiva, Fisher, UK) varying the concentration of imidazole in the elution buffers. Eluted fractions were then analysed by SDS-PAGE and western blotting to confirm purified fractions at expected molecular weight. Once fractions were tested and confirmed, fractions were pulled together, dialysis carried out using snakeskin dialysis tubing (Thermo Scientific™ SnakeSkin™ Dialysis Tubing, 3.5K MWCO, 35 mm (fisher scientific, UK) and proteins concentrated using Merck Amicon™ Ultra-2 Centrifugal Filter Units.

4.2.8 Biochemical activity testing

4.2.8.1 Protein quantification by BSA microtiter plate method

Following purification, dialysis and confirmation of pure protein, protein amounts were determined by Bradford assay (Bradford, 1976) using the microtiter plate method (Biorad). The linear range of quantification was determined (1.25 μg/ml – 10 μg/ml) using different dilutions bovine serum albumin (BSA) and protein amounts were estimated using the slope of the line given as y =mx + c. Absorbance measurements were performed using the FlUOstar omega plate reader (BMG LABTECH, UK).

4.2.8.2 4 Nitrophenol standard curve generation

A standard curve for 4 nitrophenol was constructed according to assay conditions adopted from (Franka Ganske and BMG LABTECH, 2009). Eight different amounts (0.0025 – 0.10 μmoles) from a 10 mM stock solution of 4 nitrophenol (Sigma, UK) were prepared in 200 μl of reaction buffer (50mM sodium phosphate, pH 7.5). To the reaction buffer, 40 μl of DMSO was added to mimic substrate amounts that will be used in the assay. Absorbance readings (OD 410 nm) were recorded using the FLUOstar omega plate reader (BMG LABTECH, UK) and blank corrected average readings were plotted against the amounts of pNP in 240 μl total assay volume. The fit results were retrieved applying linear regression analysis and product amounts estimated using slope of the line (y = mx), where m represents the extinction coefficient for pNP under the conditions mentioned.

4.2.8.3 Enzyme activity testing

From a 10 mM stock solution of 4-nitrophenyl-acetate (pNPA) solved in DMSO, different dilutions of pNPA were prepared. In each well 190 μl sodium phosphate buffer (50 mM, pH 7.5) and 10 μl enzyme preparation was pipetted. Different amounts of pNPA ranging from (0.01 μmol to 0.2 μmol) in 40 μl DMSO were added to the reaction mixture achieving a final volume of 240 μl. Enzymatic reaction was initiated after addition of the substrate and, absorbance measurements at 410 nm over a period of 120 seconds at 37°C were recorded using Varioskan LUX Multimode Microplate Reader. Absorbance values were adjusted based on the blank which represented the autohydrolysis of pNPA without enzyme. In addition to the blank, a pNPA-free negative control (NC) made up of only of buffer and enzyme was included in our assay. Enzyme kinetics were performed to determine the km and vmax under optimum assay conditions, and to validate our assay, commercial acetyl xylan esterase purchased from (Sigma, UK) was used as a positive control for activity testing.

4.3 Results

From the metagenomic data, a sequence belonging to putative CE1 was selected for investigation. Initial bioinformatics and protein modelling was carried out employing the National Center for Biotechnology (NCBI) database and I-TASSER database. Multiple sequences from our BLASTP search against the non-redundant, UniProt and PDB databases were aligned. A table of the results used in the sequence alignment has been displayed with the accession numbers, scientific name, percentage identity and E-value which was used as a score of quality of alignment.

4.3.1 Bioinformatics and protein modelling results

BLASTP query using the amino acid sequence from the metagenomic library against the non-redundant, UniProt and PDB databases resulted in top hits, out which the top 3 were selected. The selected top hits and their accession numbers, scientific name, percentage identity and E-value were tabulated as seen in Table 4.1. Retrieved FASTA sequences from the 3 top hits were aligned in a multiple sequence alignment to identify sequence similarity.

Table 4.1 Sequence producing significant alignment (non – redundant) from BLASTP

Accession number	Percentage identity (%)	E-value	Scientific name
WP_213794851.1	81.67	0.0	*Klebsiella aerogenes*
WP_064548457.1	92.28	0.0	*Buttiauxella ferragutiae*
WP_151992792.1	93.18	0.0	*Buttiauxella massiliensis*

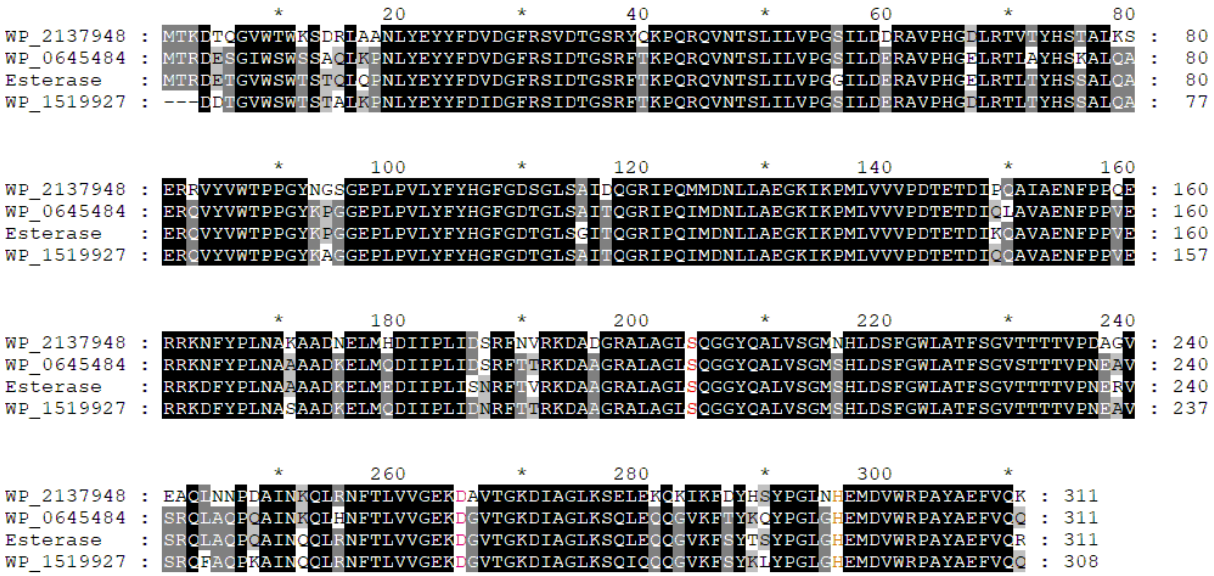


Figure 4.1 Multiple sequence alignment of our esterase sequence and the top 3 hits following query against non-redundant protein database

Figure 4.1. Multiple sequence alignment performed using Clustal Omega where the output shows an alignment between our sequence and top hits from the non-redundant protein database search. We observed high level of similarity between our query sequence and WP_151992792.1 with 93.18% identity and an E-value = 0.0 suggesting good hit for homology. As the sequence-based gene search approach was used in this study, homology was very crucial in determining putative activity of the selected target sequences.

Table 4.2 Sequence producing significant alignment (Uni-Prot)

Accession number	Percentage identity (%)	E-value	Scientific name
P10478.3	30.28	2e-26	*Acetivibrio thermocellus*
D5EY13.1	28.26	2e-23	*Prevotella ruminicola 23*
D5EXZ4.1	29.60	3e-33	*Prevotella ruminicola 23*

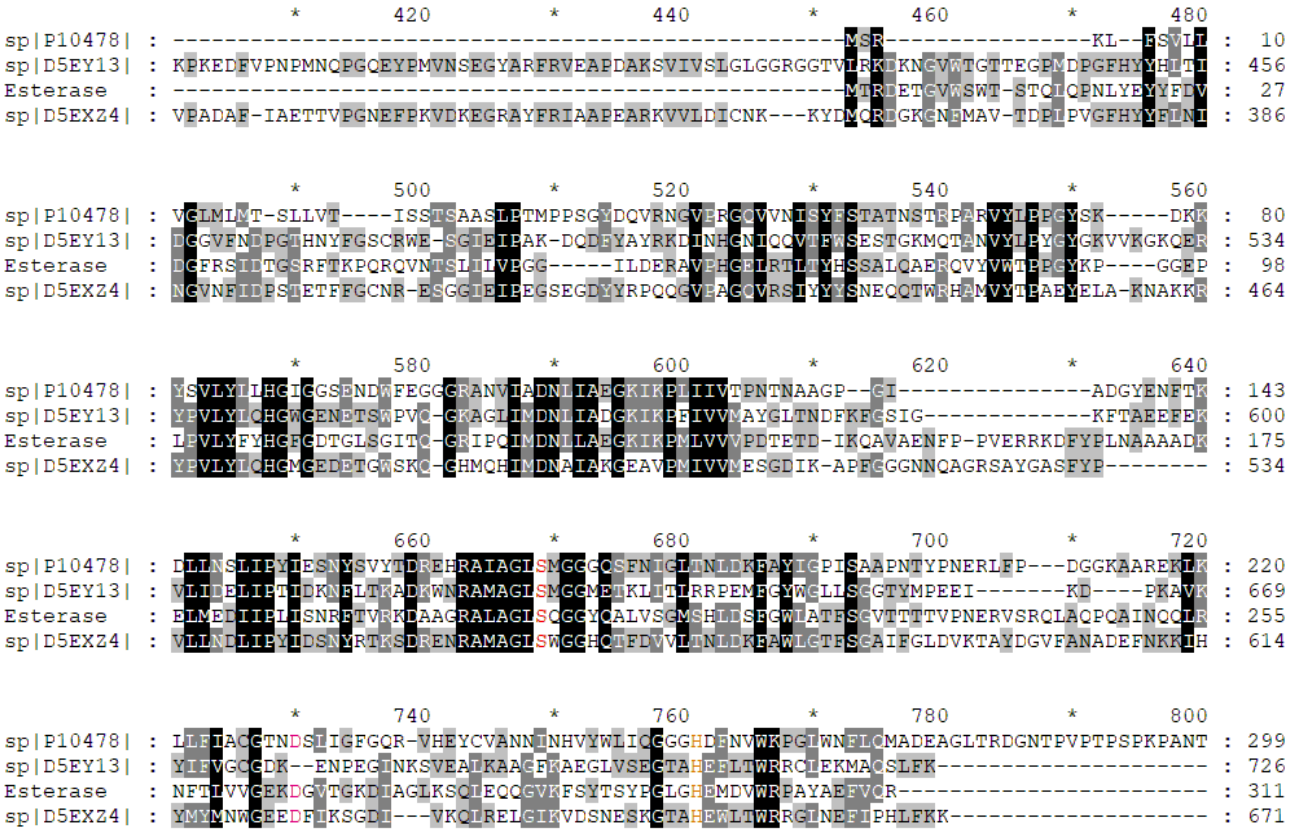


Figure 4.2 Multiple sequence alignment using top hits from blast p search against UniProt database proteins

Figure 4.2 Multiple sequence alignment of top hits from the UniProt protein database search and query amino acid sequence revealed 30.28% identity to accession P10478.3 and E-value of 2e-26 which shows a good hit for homology. As the sequence-based gene search approach was used in this study, homology was very crucial in determining putative activity of the selected target sequences.

Table 4.3 Sequence producing significant alignment (PDB)

Accession number	Percentage identity (%)	E-value	Scientific name
7B5V_A	29.02%	1e-39	*Dysgonomonas mossii DSM 22836*
1JJFA	30.16	1e-28	*Acetivibrio thermocellus*
6RZO	30.06	1e-30	uncultured bacterium

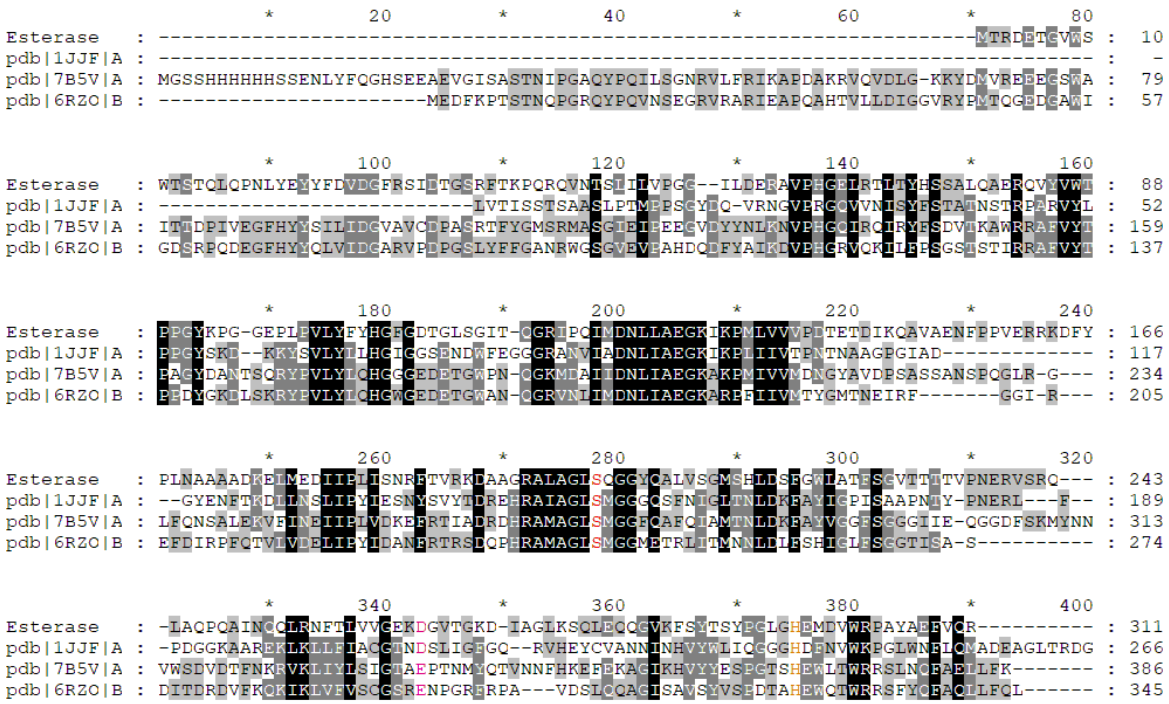


Figure 4.3 Multiple sequence alignment using top hits from blast p search against pdb database proteins

From Figure 4.3 Multiple sequence alignment of top hits from the PDB database search and query amino acid sequence revealed 30.16% identity to accession 1JJFA and E-value of 2e-26 which shows a good hit for homology. Highlighted in red = Serine, pink = Aspartic acid and orange=Histidine which collectively make up the catalytic triad well-known to be found in esterase enzymes.

Following I-TASSER analysis, we identified protein models and scoring relating to the various protein models generated following analysis of our amino acid sequence. Summarised in the table below is the protein model scoring based on the highest Cscore which determines the quality of confidence in the protein model. The protein structure model with the highest Cscore was selected and used in structural alignment against known protein structures of similar proteins.

Table 4.4 Protein model scoring table from I-TASSER

Protein model	Cscore
Model 1	0.02
Model 2	-1.22
Model 3	-2.40

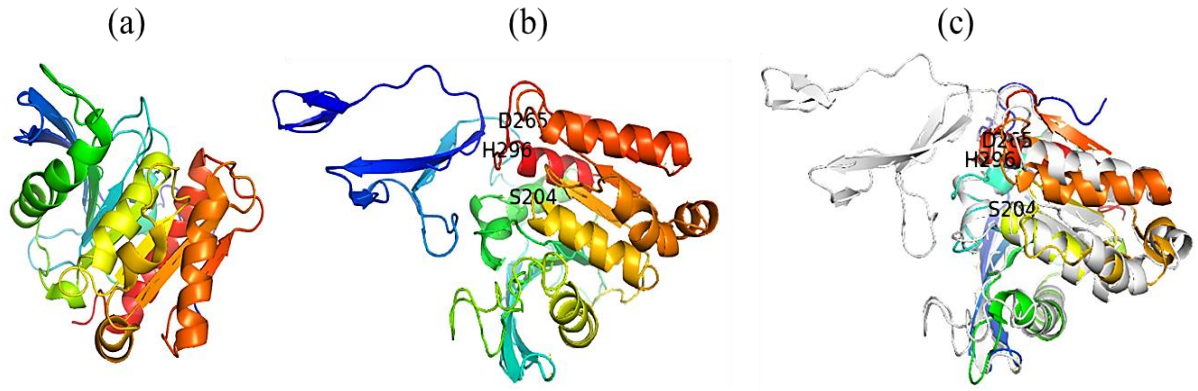


Figure 4.4 I-TASSER generated protein model results showing active sites compared to 1JJFA

In Figure 4.4 (a) Chain A, Endo-1,4-beta-xylanase Z [*Acetivibrio thermocellus*] (b) Predicted protein model 1 of our protein showing active sites. (c) protein structure alignment between our predicted model and model 1JJFA. The catalytic triad involving serine located in a consensus sequence Gly-X-Ser-X-Gly for an esterase activity was successfully identified as indicated in the multiple sequence alignment in Figure 4.3. We also identified the well-known Histidine and Aspartic acid residues commonly found in esterase enzymes.

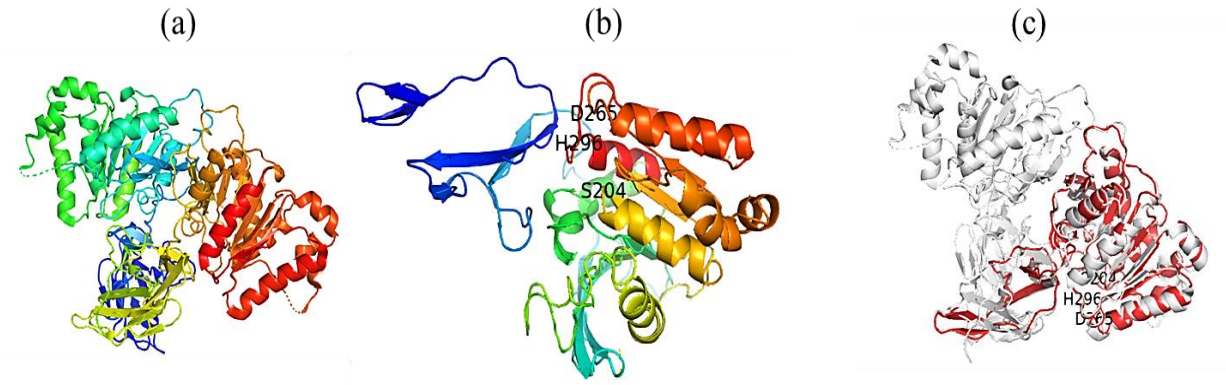


Figure 4.5 I-TASSER generated protein model results showing active sites compared to 7B5V_A

In Figure 4.5 (a) The carbohydrate binding module family 48 (CBM48) and carboxy-terminal carbohydrate esterase family 1 (CE1) domains of the multidomain esterase DmCE1B from *Dysgonomonas mossii* (b) final predicted model of protein showing active sites. (c) protein structure alignment between our predicted model and model 1JJFA. Observations from the structural alignment in Figure 4.4 (c) and Figure 4.5 (c), we identified higher level of similarity between our sequence and

the sequence belonging to carbohydrate binding module family 48 (CBM48) and carboxy-terminal carbohydrate esterase family 1 (CE1) domains of the multidomain esterase DmCE1B from *Dysgonomonas mossii* suggesting esterase activity as the putative function.

4.3.2 PCR and Restriction digest analysis

In the agarose gel images in Figure 4.6, amplimers generated from PCR were assessed for integrity using a 1% agarose gel which confirmed the presence of single discrete bands after performing a gradient PCR. The displayed restriction digest results were performed as quality check following plasmid construct synthesis to assess the quality of plasmid construct before transformation and recombinant expression.

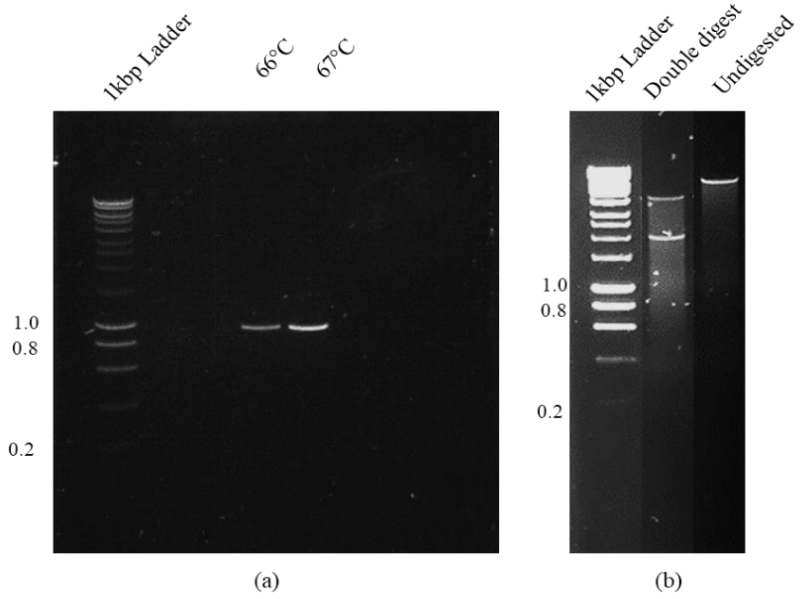


Figure 4.6 Agarose gel images of PCR and restriction digest

In Figure 4.6 (a) The successful amplification of a 936 bp fragment from the WGA metagenomic library as seen in lanes 4 and 5 using annealing temperatures of 66 and 67 respectively. In lane 1, 1 µl of (Quick-Load® Purple 1 kb Plus DNA Ladder) was loaded as DNA marker and in subsequent lanes 10 µl of PCR product (7 µl of PCR product + 3 µl of purple loading dye were loaded. In **(b)** we confirm the presence of gene of interest after our in-house QC. In lane 1, 1 µl of (Quick-Load® Purple 1 kb DNA ladder, NEB) was loaded as DNA marker and in subsequent lanes 10 µl of restriction digest product was loaded. In lane 2 (double digest performed using *Sac I* and *Apo I*) and in lane 3 (undigested plasmid) used as control for our digest.

4.3.3 SDS-PAGE and Western blot analysis

The expression of esterase enzyme was detected in cell lysate following recombinant protein expression in BL21 DE3 *E. coli* cells. Soluble protein in the supernatants from the different time point expressions were harvested and analysed by SDS-PAGE and western blotting to determine optimum expression conditions. See Figure 4.7.

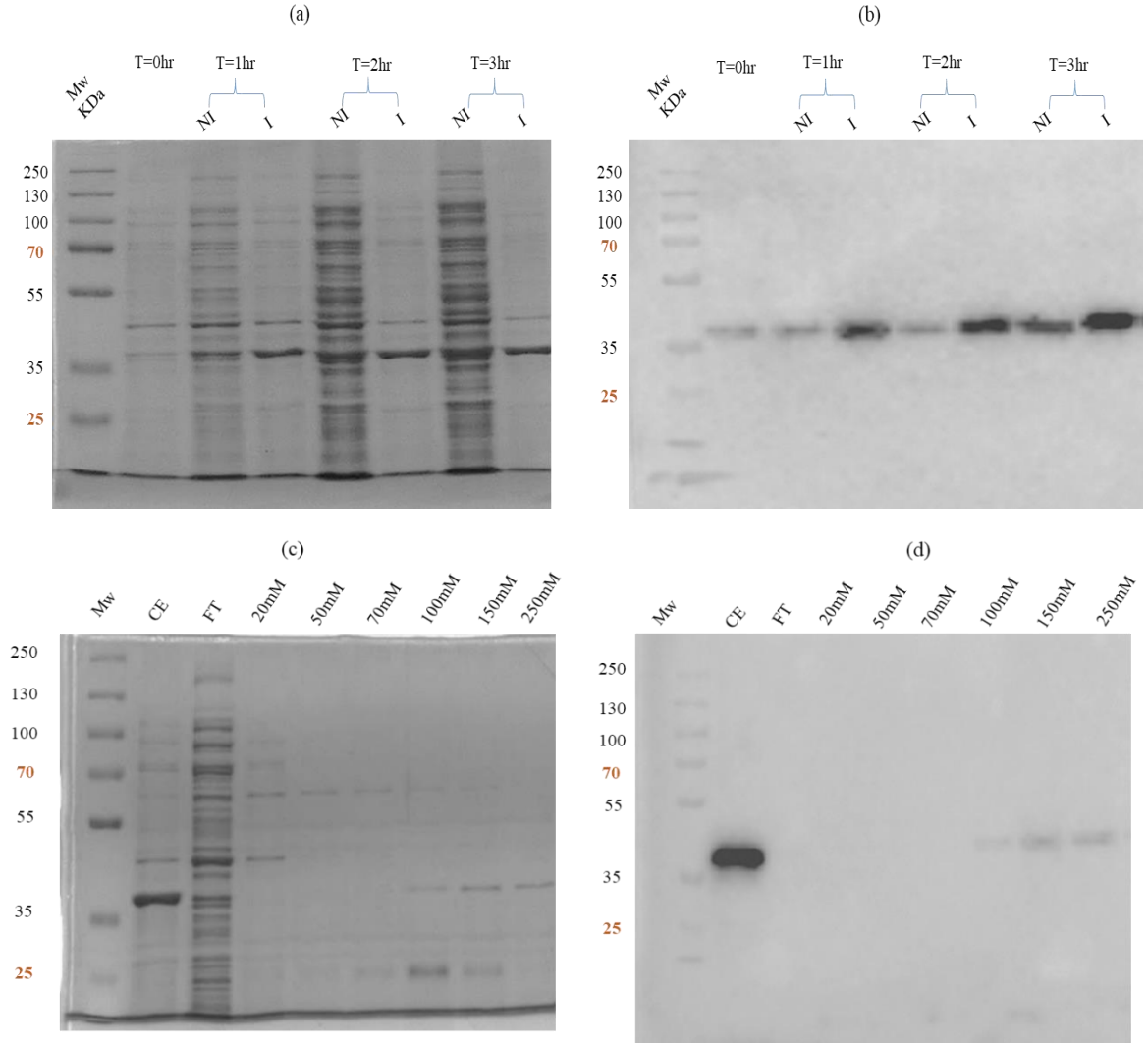


Figure 4.7 SDS-PAGE and Western blot image of expressed protein (1463) showing expression at different time points.

Figure 4.7 (a) SDS-PAGE of positive clone time point expression after 0.7mM IPTG. (b) Western blot confirming expressed histidine tagged protein. (c) SDS-PAGE of eluted fractions after purification (d) Western blot of purified histidine tagged protein. SDS-PAGE gels were stained with InstantBlue™ Protein Stain (expedeon) and immobilised proteins via western blot were detected using specific

antibodies for the his-tagged protein. For the protein marker, 2.5 µl of Pageruler plus was loaded, and then in subsequent wells, 20 µl from samples mixed with Laemli sample loading buffer to achieve 1x. From the SDS and western blot, a monomeric protein band ~36 kDa was detected. We also observed from the SDS-PAGE results that time point expression after 4 hours resulted in an increase in soluble protein after lysis of cell. Thus, we concluded optimal expression conditions at 37°C for 4-hours post induction with 0.7mM IPTG. Proteins were then dialysed, and activity testing performed using desalted proteins.

4.3.4 Biochemical activity testing results

4.3.4.1 4-nitrophenol standard curve generation

For biochemical testing, aryl glycoside conjugated to the chromogenic side (4-nitrophenol) was used as substrate for enzymatic assays. Since 4-nitrophenol released from the conjugated substrates is directly proportional to product formation, a standard curve of 4 nitrophenol was developed to help in the estimation of product formed.

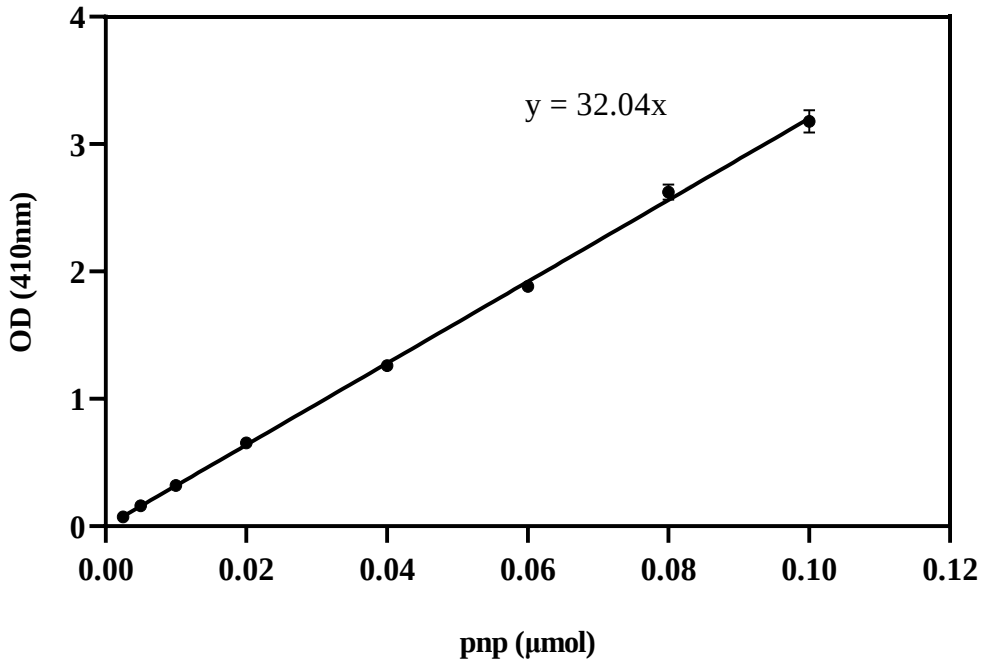


Figure 4.8 pNP standard curve using 240 µl volume with 16.6% DMSO

In Figure 4.8 is a standard curve for 4-nitrophenol in the linear range of 0.0025 – 0.10 µmoles with (R^2=0.9993) using 8 dilutions of 4 nitrophenol. Readings were measured in triplicates, average determined and average values minus blank plotted against amount of pNP. Standard deviation (*n*=3)

was calculated and has been indicated as positive and negative error bars. The slope of the curve was indicated as 32.04 which stands for the molar extinction coefficient for pNP under our assay conditions.

4.3.4.2 Signal curve of esterase catalysed reactions varying concentrations of substrate

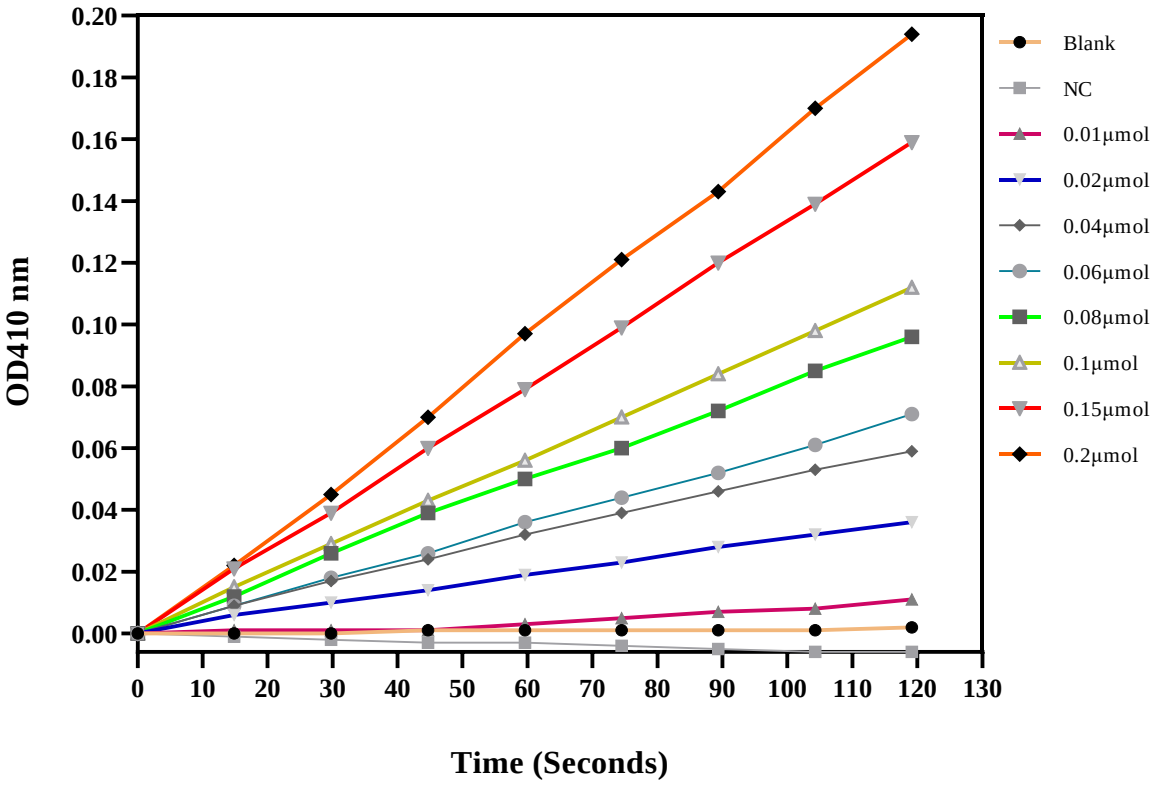


Figure 4.9 Signal curves of esterase-catalyzed reactions using different concentrations of the substrate pNPA (pNPA concentrations range from 0.01 µmol to 0.2 µmol / 240 µl).

In Figure 4.9 is a Signal curve for esterase using 8 different amounts of pNPA solved in 40 µl DMSO added to a reaction mixture consisting of 10 µl of enzyme and 190 µl of reaction buffer (50 mM Sodium phosphate, pH 7.5) and incubated at 37°C for a period of 120 seconds. Average OD 410 nm readings were corrected from blank readings and negative control (NC) represents reaction buffer + enzyme. From the graph a good signal can be seen with increasing amounts of pNP product formed showing an increase in enzyme reaction rate over time.

4.3.4.3 Enzyme kinetics

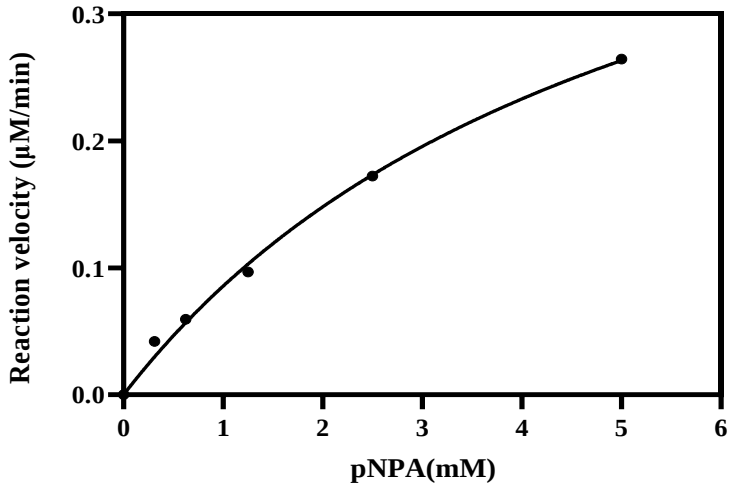


Figure 4.10 Michaelis-Menten plot of the reaction velocity, obtained with the esterase enzyme preparation, depending on the substrate concentration.

Figure 4.10. Enzyme kinetics showing Michaelis-Menten plot which was used in determining substrate concentration at which half of all the active sites of the enzyme are covered by substrate molecules as a measure of enzyme affinity for substrates, given as Km = 5.404. From our Michaelis-Menten plot, we observed high Km = 5.404 indicative of low enzyme affinity for the substrate as more substrate concentration is required to occupy half of the active sites and reach Vmax = 0.547781µM/min.

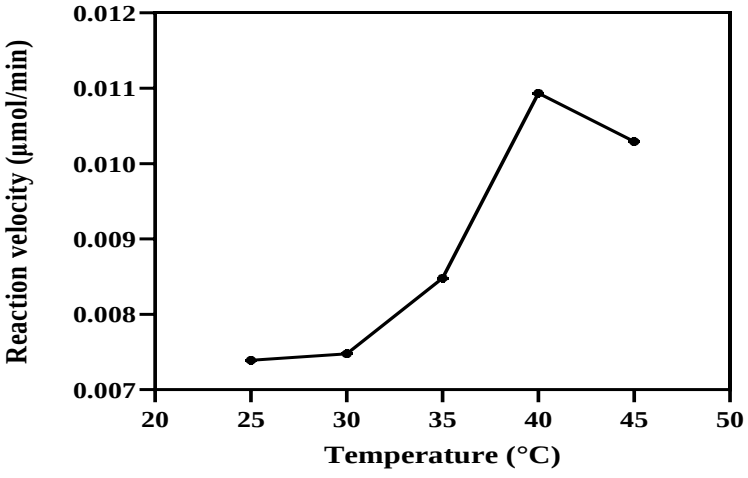


Figure 4.11 Temperature profile for esterase enzyme activity

Figure 4.11. Temperature profile of pNP Acetate over a range of temperatures (25 – 45°C). From the results we observed a decrease in activity as the enzyme approaches 50°C, suggesting optimum

temperature for activity observed for this esterase to be 40°C. For optimum pH determination, the substrate pNPA was found to be very sensitive to pH changes and thus affected the authenticity of our results, not indicated in this study.

4.4 Discussion

Accessibility of the sugar rich fractions of biomass will play a crucial role as such monomers will serve as base molecules for industrial applications such as food and feed, drug delivery technology and in the pulp and paper industry. Hemicellulose represents a challenge due to the recalcitrant nature of biomass. Biomass has been dubbed a good alternative feedstock to replace fossil fuels used in the production of petrochemicals. Nevertheless, achieving sustainable bioconversion methods applicable on an industrial scale is paramount to success of such bioconversion technologies as well as mitigating global GHG emissions. Carbohydrate Esterases 1 family enzymes exhibit proprieties which makes the application of these enzymes industrially relevant. The cosmetics (Jeon *et al.*, 2009), bleaching (Gangwar, Tejo Prakash and Prakash, 2016), pharmaceuticals (Panda and Gowrishankar, 2005), pulp and paper industry (Topakas, Vafiadi and Christakopoulos, 2007) and the biofuel industry (de Oliveira *et al.*, 2018) have all been named to benefit from the extensive application of esterases. According to Turati *et al.*, (2019) esterases are capable of catalysing reactions including esterification, transesterification and ester hydrolysis achieving good enantioselectivity resulting in the production of stereoisomeric products with several industrial applications. Therefore, esterases have a varied application which could meet industrial demands for bioconversion of biomass.

In this study, the aim was to recombinantly express an esterase identified from the slug gut metagenome through the power of metagenomics and test its thermal stability. Using a metagenomic approach we were able to assemble and annotate sequences from the slug gut metagenome with putative glycosyl hydrolase activity. The selected putative glycosyl hydrolase sequence for this study through

bioinformatics identified conserved domains similar to that of carbohydrate esterase family 1 enzymes involved in hemicellulose breakdown which was further validated through our biochemical activity testing. We also set out to apply a method of enzyme immobilisation through Cross Linked Enzyme Aggregates (CLEA's) to improve the functionality of the enzyme. Findings from this study suggested the successful amplification of a discrete 936bp target sequence from the slug gut metagenome as seen in Figure 4.6 (a) with esterase activity testing following expression. Bioinformatics carried out using multiple sequence alignment offered through Clustal Omega and structural alignments using Pymol revealed the presence of the catalytic active sites responsible for esterase activity observed through our structural alignment to known esterases and endo-1,4-beta-xylanse enzymes depicted in in Figure 4.4 (c) and Figure 4.5 (c). The amino acid sequence consisting of 311 amino acids was subjected to multiple sequence alignment analysis using top BLASTP hits. FASTA sequences from accession numbers from our top hit query searches were selected and used in the multiple sequence alignments which identified good hits for homology with approximately 30% identity when queried against the PDB and UniProt database. Homology was essential to this study as we adopted the sequence-based gene search approach. When the sequence was queried against the non-redundant protein database, 93.18% identity was confirmed to accession WP_151992792.1 with an e-value of 0.0 which is good hit for homology. Based on this percentage identity and e-value, it was deduced that the target sequence identified as a CE1 family enzyme could be from *Buttiauxella massiliensis*. See Figure 4.1 and Table 4.1.

As well as homology searchers, protein structural alignment was performed using the predicted protein model from our protein prediction results through I-TASSER. Protein model which had the highest C-score was chosen for structural alignment. C-score gives quantitative measure of confidence of each model, where a higher value indicates a model with high confidence. Therefore, the selected protein model used in this study had a high Cscore value of 0.02 compared to other models which had

negative Cscore values as indicated in Table 4.4. Structural alignments pointed in the direction of the carbohydrate binding module family 48 (CBM48) and carboxy-terminal carbohydrate esterase family 1 (CE1) domains of the multidomain esterase DmCE1B from *Dysgonomonas mossii* (PDB accession 7b5v) (Kmezik *et al.*, 2021) and Crystal structure of the N-terminal carbohydrate binding module family 48 and ferulic acid esterase from the multi-enzyme CE1-GH62-GH10 (PDB accession 6RZO) (Holck *et al.*, 2019). Also, structural alignment revealed the presence of active sites found in enzymes belonging to EC 3.1.1.73 (Holck *et al.*, 2019; Kmezik *et al.*, 2021) and three active sites related to active sites found in enzymes belonging to EC 3.1.1.72 (Poutanen *et al.*, 1990; Sundberg and Poutanen, 1991) known to exhibit properties of carboxyl esterase enzymes. The presence of the catalytic triad consensus sequence (Gly-X-S-X-Gly) indicated more confidence in our protein prediction results and thus led to the production of the said protein. Through this alignment we were able to confirm the presence of catalytic sites including Serine, Aspartic acid and histidine, responsible for the nucleophilic attack of the active site serine on the carbonyl carbon atom of the ester (Lockridge and Quinn, 2010).

Following the confidence in our protein modelling, experimental work was performed to produce a single discrete amplimer of 936 bp which was successfully generated through a gradient PCR and used in a plasmid construct synthesis. The plasmid construct was successfully transformed into BL21 DE3 *E. coli* cells with recombinant protein expression successfully achieved as displayed Figure 4.7 showing a monomeric protein band at 36 kDa. The successfully expressed and purified CE1 enzyme was then subjected to activity testing on 4 nitrophenyl acetate, a chromogenic substrate which has been used in the activity testing of esterase enzymes (Blum *et al.*, 1999). For product determination after enzyme activity, we successfully developed a standard curve based on the chromogenic 4-nitrophenol where the slope of the line given as 32.04 OD/μmol, was used as the molar extinction determining product formed as this was directly proportional to the product released (See Figure 4.8). A

signal curve was also successfully generated indicating a strong signal of activity varying the amounts of pNPA used in the reaction as observed in Figure 4.9.

However, data from the enzyme kinetics suggested that more substrate concentration was required to occupy all the active sites of the enzyme given by the Michaelis Menten Km = 5.404 mM of pNPA with a reaction velocity max of 1.964 μM/min of pNPA under optimum conditions. Based on the high Km value observed in Figure 4.10 we believe that the choice of substrate had a major effect on the enzyme activity. Although 4-nitrophenyl acetate has been used as a substrate to test for the activity of CE1 enzymes, feruloyl esterases in particular have a poor affinity for 4-nitrophenyl acetate (Faulds and Williamson, 1991). Therefore, having a look at our structural alignment and our data from our Michaelis Menten plot, we strongly believe that the CE1 enzyme that was recombinantly expressed will demonstrate high substrate specificity on 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar (Faulds and Williamson, 1991, 1994; PA, CB and Williamson, 1996).

4.5 Conclusion

With this study, we have validated the power of metagenomics in identifying a novel coding DNA sequence with putative glycosyl hydrolase function. We have successfully produced a carbohydrate esterase family 1 enzyme produced by microbes identified from the slug gut microbiota through metagenomics. Biochemical activity testing suggested activity on nitrophenyl esters with the release of acetic acid as monitored via our enzyme assays. However, the high Km value from the enzyme kinetics prompted an in-depth functional investigation which we finally conclude in the possibility that the enzyme could be more of a Feruloyl esterase than an acetyl esterase. However, time did not allow such testing to proceed, so our recommendation for future experiments should focus on testing the activity of this enzyme on 4-hydroxy-3-methoxycinnamoyl to make such valid conclusions (Faulds and Williamson, 1991, 1994).

Chapter 5. Enzyme Immobilisation towards improved operational functionality

Abstract

Biodegradable and biocompatible enzymes used in bioconversions usually demonstrate adequate activity but are affected by various conditions and components in a bioreactor mix that might inhibit the enzymes performance and consequently reduce product yield. Enzyme immobilisation has been described as a technique with potential improvement to the operational functionality of enzymes such as thermal stability and reusability. However, the application of such technologies is not without advantages and disadvantages. Also creating such biocatalysts will have to match or outperform the use of chemicals on cost-to-benefit ratio economically. Enzyme immobilisation techniques such as nanoparticles, hydrogels and cross-linked enzyme aggregates have been employed successfully in industrial applications such as hydrolysis of pectin, interesterification of food fats and oils, production of biodiesel, CO_2 capture and several other applications. As a result, enzyme immobilisation is instrumental in biocatalyst development.

In this study, 15nm nanoparticles were synthesised as alpha-L-rhamnosidase immobilisation supports and successfully characterised through UV-vis and differential centrifugation sedimentation. Successfully synthesised and characterised nanoparticle-enzyme conjugates were employed in a residual activity testing over a period of 8 hours observing product formation. From this study, we found that the activity of the recombinantly expressed free α-L-rhamnosidase was maintained at higher level compared to the nanoparticle-enzyme conjugate, suggesting that the technique of immobilisation did not improve enzyme functionality in this study.

We also developed Cross-linked Enzyme Aggregates using a putative CE1 enzyme achieved using 0.05 % and 0.1 % glutaraldehyde and characterised the degree of cross linking over time as seen in colour development profile and UV-vis absorbance

spectrum. Cross linked enzymes exhibited a significant increase in activity per mg of protein compared to non-cross-linked enzymes indicated from the results of our Mann Whitney test with p = 0.0079 suggesting that Cross-linked Enzyme Aggregates do indeed improve functionality oof enzymes

5.1 Introduction

Enzymes are biodegradable, biocompatible, biological catalysts derived from renewable resources (Sheldon and van Pelt, 2013). Enzymes are able to catalyse chemical processes or reactions and have several uses in industries, research and clinical diagnostics (Arsalan and Younus, 2018). Achieving biocatalysis through bioconversion will serve as a more sustainable alternative to the use of harsh chemicals in industry. Enzymes are naturally reactive and selective in their natural environments but may suffer from degradation when applied to processes outside of their natural environments such as in industrial applications (Bonzom *et al.*, 2018). Enzymatic reactions are usually carried out under milder conditions such as ambient temperatures, physiological pH in H_2O and atmospheric pressure achieving high rates of activity and selectivity. As such, production of enzymes which can work under these conditions at an industrial scale is preferred for efficient industrial/ commercial applications. Recombinantly expressed enzymes in the laboratory can serve as biological catalyst and can be used in the conversion of natural substrates such as birchwood xylan, starch, cellulose, and other sugar rich biomass examples.

Biocatalysis applications using enzymes such as proteases, lipases or xylanases are extensive and have been documented in their use in for example, food industries for baking, starch conversion and beverage processing (Dekker, 1994). Enzymes such as amylases have been used in the textile industry and the pulp and paper industry achieving end products with high value (Raveendran *et al.*, 2018). Pharmaceutical industries, waste management and biofuel companies also employ the use of enzymes in the generation of high value molecules (Béguin and Aubert, 1994; Kuhad *et al.*, 2011;

Eman Talukder *et al.*, 2017). Nevertheless, enzyme stability, shelf life, recovery and the reusability adversely limits the usability of enzymes in large scale industrial processes (Chapman, Ismail and Dinu, 2018). Addressing such limitations by employing enzyme immobilisation might hold the key to the improvement in enzyme applications in industries as we strive to achieve sustainable industrial methods. The transformation from laboratory scale to industrial scale is one of the most daunting tasks but an exciting goal in biotechnology.

5.1.1 Types of enzyme Immobilisation

Enzyme immobilisation has been in existence as early as 1916 when Nelson and Griffin described the adsorption of invertase on charcoal and the effect charcoal had on enzyme activity (Nelson and Griffin, 1916). Several studies on enzyme immobilisation have since been carried out over the years but the major advances in enzyme immobilisation can be observed in industrial applications such as hydrolysis of pectin, interesterification of food fats and oils, production of biodiesel, CO_2 capture and several other applications (DiCosimo *et al.*, 2013). The growing interest in the use of immobilised enzymes in industry stems from the attractive properties such as ability to resist environmental changes, such as; pH and high temperatures (Cherry and Fidantsef, 2003). In reactions, immobilised enzymes have been described to be more stable than free enzymes with good recoverability allowing easier separation of catalysts from the product post reaction, particularly essential in pharmaceutical and food industries where contamination is major setback (Chapman, Ismail and Dinu, 2018; Patra, Das and Shin, 2018). Additionally immobilisation is said to improve enzyme performance in organic solvents whiles maintaining functional stability (Ahmad and Sardar, 2015). In other studies, structural rigidity and stabilisation is said to prevent dissociation-related inactivation and the re-use of an enzyme over a period of time avoiding the replication of tedious extraction and purification processes (Rodrigues *et al.*, 2013; Guzik, Hupert-Kocurek and Wojcieszyńska, 2014).

Several immobilisation methods have been described which are involved in alterations of an enzyme's stability and functional properties (Brena, González-Pombo and Batista-Viera, 2013). These methods of immobilisation can be broadly categorised into physical and chemical methods. Physical methods of immobilisation are usually achieved via weak interactions whereas chemical methods of immobilisation are achieved via covalent bonding. In both methods, the primary objective is to maximise the advantages of enzyme catalysis whiles achieving low synthesis cost but rendering high binding capacity (Liese and Hilterhaus, 2013). Several methods of immobilisation including adsorption, covalent bonding, entrapment, and cross-linking have been employed. See Figure 5.1 for a diagrammatic representation of examples of immobilisation methods.

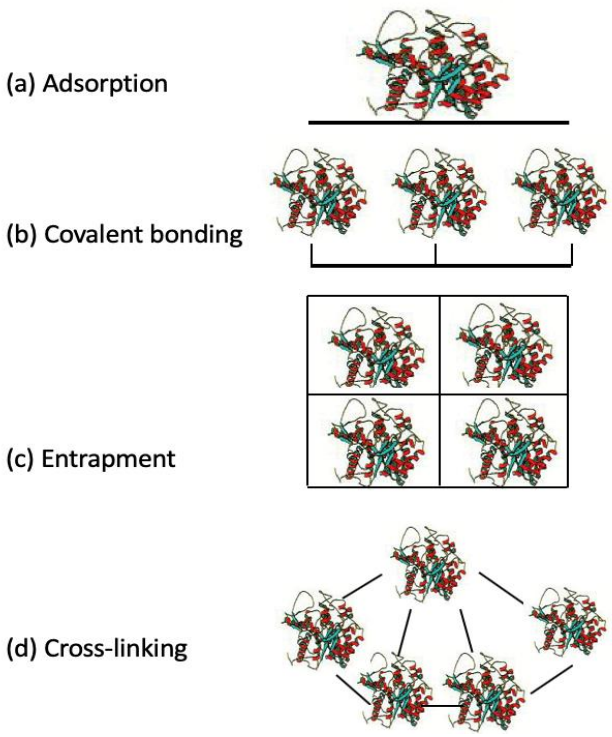


Figure 5.1 Diagrammatic representation of various methods of enzyme immobilisation

In Figure 5.1 schematic of techniques described for the immobilisation of enzymes (a) Adsorption is the simplest and oldest method of immobilisation involving the incubation of an enzyme with an adsorbent under appropriate conditions. (b) Covalent bonding involves the formation of a covalent bond between the enzyme and supporting matrix and is dependent on functional group interactions. (c) Entrapment could either involve physical caging or covalent bonding and involves the use of natural

polymers such as agarose and chitosan (Krajewska, 2004) and the use of inorganic polymers such as silica sol gels and hydrogels (Sheldon and van Pelt, 2013) and (d) Crosslinking employs bifunctional or multifunctional reagents of ligands in the attachment of biocatalyst to each other (Ahmad and Sardar, 2015).

5.1.2 Nanoparticles

The choice of supporting matrix for enzyme immobilisation is usually made after consideration of properties such as; biocompatibility, resistance to microbial attacks, cost, resistance to compression to list a few (Brodelius and Mosbach, 1987). These supporting matrixes can be in the form of either organic (natural or synthetic polymers) or inorganic (natural and processed minerals) depending on the chemical composition. Nanoparticles made through materials such as, gold (Au) and silver (Ag) have been used in the immobilisation of enzymes with success (Bailes *et al.*, 2012; Petkova *et al.*, 2012). Bioconjugation of glucose oxidase using gold nanoparticles (AuNPs) has been described by (Li *et al.*, 2007), bioconjugation of histidine (HIS-6) tagged acetylcholinesterase on nickel nanoparticles for a detection system has been described by (Ganesana *et al.*, 2011). In another study (Khoshnevisan *et al.*, 2011) achieved the bioconjugation of cellulase on superparamagnetic nanoparticle. The list of applications is exhaustive but, in all applications, the use of nanoparticles relies on the unique properties of the chosen nanoparticle. Nanoparticles offer robust support for immobilisation because of their capability to maintain factors that affect the efficiency of biocatalysts. Enzymes conjugated to nanoparticles are said to exhibit Brownian movement, a phenomenon which allows microscopic particles such as nanoparticles to remain suspended in liquid or gas, making possible enhancement of activity, compared to unconjugated enzymes (Gupta *et al.*, 2011; Ahmad and Sardar, 2015). In other applications, nanoparticles with magnetic properties (superparamagnetic) have been separated from reaction mixtures employing a magnetic field and thus eliminating/reducing enzyme contamination issues (Gupta *et al.*, 2011). The are several methods of linking a protein to a nanoparticle as a solid

support to enhance activity, improve stability or add any additional properties. Below is a diagrammatic representation of the more detailed approaches widely used in the immobilisation of enzymes /protein to nanoparticle.

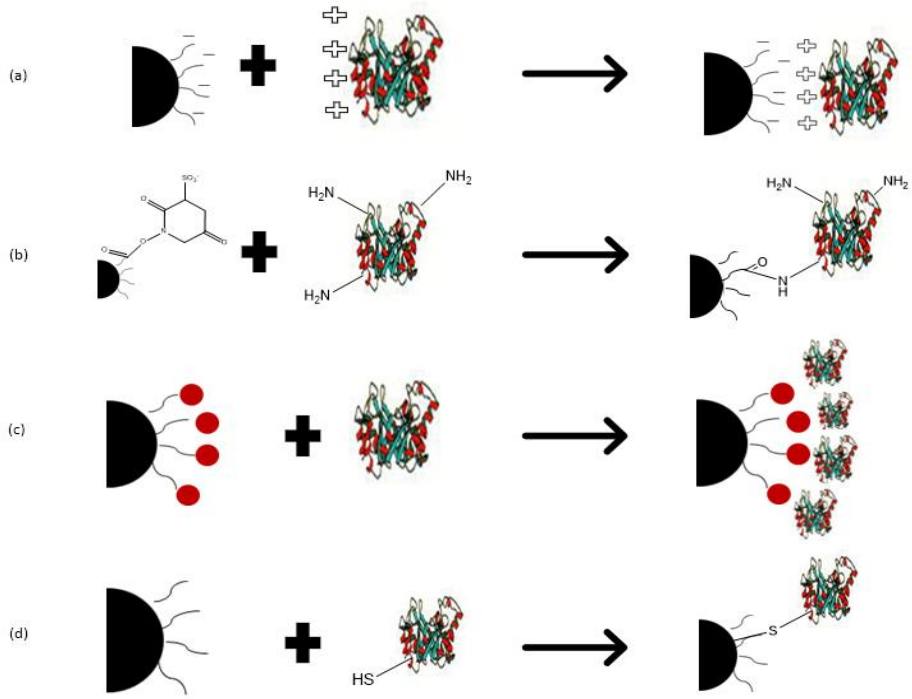


Figure 5.2 Methods of linking protein to nanoparticle

In Figure 5.2. Represented above is a diagrammatic representation of different enzyme -nanoparticle linkage approaches (a) Electrostatic adsorption (b) Covalent attachment of enzyme to nanoparticle ligand complex (c) Conjugation via protein specific affinity (d) Direct linkage to surface of nanoparticle.

5.1.2.1 Covalent attachment

For the purpose of this study, particular focus is on the covalent attachment of enzyme to synthesised nanoparticles employing different surface modifications such as the binding of primary amines with sulfo-NHS esters or R-COOH groups via reaction with EDC (Krpetić *et al.*, 2012). EDC/NHS coupled reactions involve EDC, a water soluble carbodiimide and N-hydroxysuccinimide (NHS) or its water-soluble analog (Sulfo-NHS) used to improve efficiency or create dry-stable (amine-reactive) intermediates (Pandey, Demchenko and Stine, 2012). Such synthesised nanoparticles can be of different sizes and shapes such as nanorods, nanostars and spherical

conformations. However, there are drawbacks in the application of this method of immobilisation including biocompatibility, cytotoxicity , loading capacity and the fact that nanoparticles are non-biodegradable (Roma-Rodrigues *et al.*, 2020). As such other immobilisation methods have been discussed including carrier-free immobilisation.

5.1.3 Carrier-free Immobilisation by cross linking

Enzyme cross-linking has been possible since the 1960s, when solid phase protein chemistry enabled cross-linking of enzymes by reacting surface amine (NH2) groups with glutaraldehyde, a bi-functional chemical cross-linker. Insoluble cross-linked enzymes (CLEs) with retained catalytic activity are formed as a result of this process. However, the retention in activity using this method was low, the method was hardly reproducible and other challenges with handling gelatinous CLEs were realised. As a result, around the 1990s, the emphasis turned to the usage of carrier-bound enzymes, resulting in the development of cross-linked enzyme crystals (CLECs) and cross-linked enzyme aggregates (CLEAS®).

5.1.3.1 Cross-linked enzyme crystals (CLECs)

CLECs are robust immobilisates typically (1 – 100 μm) in particle size made by enabling an enzyme to crystallise from an aqueous buffer at the optimal pH and then crosslinking the crystals with a bifunctional reagent, commonly glutaraldehyde. CLECs were found to be much more resistant to denaturation by heat, organic solvents, and proteolysis than soluble enzymes or lyophilized powders. This led to the use of CLECs in industrial applications with Altus Biologics commercialising this method of biocatalysis (Margolin and Navia, 2001). The use of CLECs industrially, has many advantages such as recyclability, functional stability, and increased activity. However, the process of making CLECs is often laborious requiring a series of enzyme purification steps which uses expensive columns before they are even crystallised. Due to this limitation, an alternative to CLECs was created known as CLEAs.

5.1.3.2 Cross-linked enzyme aggregates (CLEAS®)

CLEAS® employs a much simpler and less expensive method which does not require crystallisation. In this type of immobilisation, enzymes are precipitated by the addition of salts such as ammonium sulphate, water miscible organic solvents such as acetone or non-ionic polymers such as Polyethylene glycol (PEG) with resulting tertiary structure intact (Sheldon and van Pelt, 2013). The use of ammonium sulphate or polyethylene glycol for this method of immobilisation achieves both purification and immobilisation in one unit, suggesting CLEAS® can be employed in the immobilisation of enzymes directly from cell lysates or fermentation broth (Sheldon and van Pelt, 2013). After precipitation, enzymes can be conjugated by the introduction of glutaraldehyde, siloxane or cross linked to water-in-oil emulsions of lipases (Brady and Jordaan, 2009). CLEAs has several advantages that makes it a good candidate for industrial applications. Enzymes from crude extracts can be used with high recovery, improved stability, and enhanced activity. This method has been used for the immobilisation of enzymes of industrial importance such as hydrolases, oxidoreductases and lyases (Fernandez-Lafuente, 2009; Sheldon, 2010, 2011). A more advanced version of CLEAs is the Combi-CLEAs which can be achieved by the combination and cross-linking of two or more enzymes towards a one-pot catalytic conversion. This could see many potential benefits towards industrial application as fewer units of operation, less reactor volume, shorter cycle times and less waste generation among other advantages can be realised with this type of immobilisation (Sheldon and van Pelt, 2013).

5.1.4 Hydrogels

Another enzyme immobilisation method involving the use of hydrogels, has been documented in several studies relating to applications in diverse areas such as bioengineering, tissue engineering and regenerative medicine to wastewater treatment and soft robotics (Wang *et al.*, 2019). Hydrogels can be described as three-dimensional (3D) cross-linked polymer networks, with the ability to absorb and retain

large amount of water (Wang *et al.*, 2019). Enzymes can be immobilised in hydrogels that exhibit natural, synthetic or cryogel properties and can have several applications, for example, the use of polyvinyl alcohol (PVA) cryogels for the immobilisation of whole cells (Sheldon and van Pelt, 2013; Lozinsky, 2018). Whole cells/cells immobilisation methods have been in use since the 1970s as an effective method to improve the performance and economics of many fermentation processes (Zhu, 2007). The process of cell immobilisation employing hydrogels offers advantages such as prolonged stability of bioreactors, easy product recovery and increased yield, reaction selectivity, easy separation of biomass and specific metabolic improvements such as resistance to forces imparted by environmental stresses such as, pH, temperature, organic solvents, salts, inhibiting substrates and products, poisons, self-destruction (Zhu, 2007; Nemati and Webb, 2011). This has maintained the interest in hydrogels as an immobilisation method for enzyme in diverse applications.

5.1.5 Industrial applications of immobilisation techniques

Current research is heavily focused on improving the efficiency of biomass processing enzymes to meet medium-term demands of producing sugars that can be fermented for bioethanol production or aromatic monomers for bio-based products. The application of enzyme immobilisation methods towards biocatalysis will aid in the formation of biocatalysts with enhanced properties for industrial scale processing of biomass into chemical building blocks (green chemicals). In the case of biocatalytic reactions, enzyme immobilisation will allow for biocatalyst separation, recycling, and application in various reactor ideas (Schmieg *et al.*, 2019). Biocatalysts with improved operational functionality can lead to high product yield formation. The production of **green chemicals** from the bioconversion of lignocellulosic biomass using immobilised enzymes is attractive as an approach to achieve sustainability as we shift towards the generation of sustainable chemicals to replace petrochemicals.

Aims

- To improve thermal stability of expressed alpha-L-rhamnosidase.
- To improve thermal stability of expressed carbohydrate esterase family 1 enzyme.

Objectives

- Synthesis and characterization of 15 nm citrate stabilised, and EDC/NHS coupled nanoparticles with functionalized Nickel -NTA linkers .
- Develop an immobilisation concept using glucose oxidase and demonstrated the possibility of enzyme immobilisation.
- Develop alpha-L-rhamnosidase nanoparticle conjugates and test for improved functionality.
- Develop immobilisation technique using Cross linked enzyme aggregates achieved following the rapid introduction of glutaraldehyde to precipitated protein solutions.
- Compare the activity of free enzyme to cross linked enzyme to assess improved functionality following immobilisation.

5.2 Methods

5.2.1 Synthesis of Gold nanoparticles

The method of Gold nanoparticle (AuNP) synthesis was adopted from a modified protocol described by (Turkevich, Stevenson and Hillier, 1951; Krpetić *et al.*, 2013). Nanoparticles were synthesised following the introduction of 760 µl of 0.1M Gold (III) Chloride Trihydrate solution ($HAuCl_4$) into a one neck round bottom flask containing 299.24ml of milli Q water under heating at 100°C and magnetic stirring. After this, 9ml of a 34 mM trisodium citrate solution ($Na_3C_6H_5O_7$) which had been heated to ~ 60 – 70°C was then rapidly introduced into the HAuCl4 mixture and allowed a reflux time of approximately 30 mins observing the colour changes below.

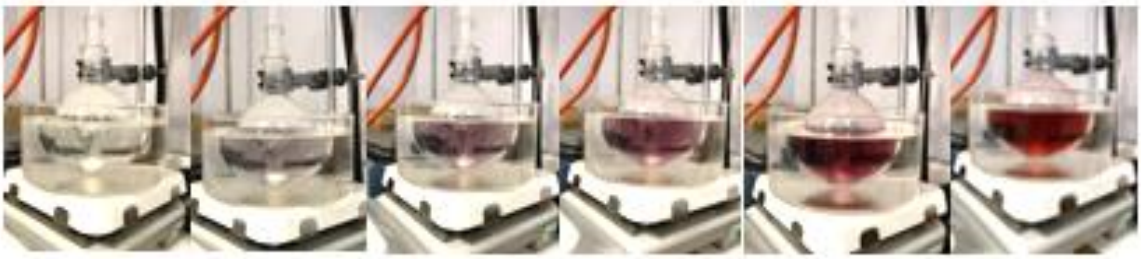


Figure 5.3 Illustration of nanoparticle synthesis.

Illustrated in Figure 5.3 is Time lapse illustration showing synthesis of citrate nanoparticles observed through colour development from left, start of reaction to right end of synthesis.

After synthesis, the now citrate gold solution, was incubated at room temperature overnight whiles under gentle stirring. After overnight incubation, synthesised nanoparticles were filtered through a 0.22 μm filter (Starlab, UK) and aliquots obtained for UV-Vis and Differential Centrifugation Sedimentation (DCS) to check monodispersing of nanoparticles before functionalisation.

5.2.2 Characterisation of Synthesised nanoparticles

A 0.5 % agarose gel was prepared to confirm successful immobilisation with regards to gel retardation of samples. Here, we used the properties of the synthesised nanoparticles and observed the migration patterns as an indication of successfully

synthesised nanoparticles. Synthesised PEG-COOH nanoparticles generally have negative charge compared to naked nanoparticles and thus the migration pattern will differ when observed in an agarose gel.

5.2.2.1 UV – vis spectroscopy

For UV – vis spectroscopy, 1ml aliquot of synthesised nano particle samples were transferred to quartz cuvettes with a path length of 1 cm and absorbance spectrum measurements were carried out scanning spectra of (400-800nm) in the Varian Cary 100 Bio UV-Visible Spectrophotometer (Agilent, UK).

5.2.2.2 DCS

Particle size distribution characterisation was achieved using CPS disc centrifuge DC24000 (CPS Instruments Inc. UK). For DCS analysis, freshly prepared 8 - 24 wt. % sucrose solution in Milli-Q water was used as the gradient fluid. The gradient fluid was loaded instantaneously in nine successions into the disc at 24,000 rpm. Calibration was performed using 0.476 μm, polyvinyl chloride (PVC) particles (Analytik Ltd. UK) as calibration standard before measurement of nanoparticles within the range of 0 – 100 nm (Krpetić *et al.*, 2013).

5.2.3 Surface modification and characterisation of synthesised nanoparticles

After UV-Vis and DCS results showed monodispersing with no alterations to surface plasmon resonance accurate size estimates, 10 ml aliquots of the citrate nanoparticles were transferred to two different tubes for functionalisation. Carboxyl terminated polyethylene glycol (PEGCOOH) functionalisation was achieved by the addition of 25µl of 0.05M PEG-COOH solution to tube (A) and to tube B, 20µl of an equal mixture of (0.05M) PEG-COOH and (0.1 M) PEG-OH solution and then incubated overnight at room temperature with gentle stirring. From tube (A) and (B), the tubes were then centrifuged at 13,000 rpm at 10°C for 10 mins or until clear supernatant was observed. The supernatant was then discarded, and nanoparticles resuspended in 10 ml of Milli-

Q H_2O (Merck, UK). Aliquots from both tubes were characterised by UV-Vis and DCS methods detailed above.

5.2.4 Carboxyl terminated coupling and characterisation

For the carboxyl terminated coupling, 25 µl of a 6mM EDC/NHS solution was added to tubes (A and B) from above preparation. Prepared reaction mixtures were wrapped in foil and incubated overnight in the dark at room temperature followed by the addition of 2.5 µl of a 6mM NTA solution added to each tube. The tubes were again incubated in the dark overnight and characterised by UV-vis and DCS as detailed above. For the final functionalisation of the nanoparticles, 2.5 µl of a 6 mM Nickel (II) chloride was mixed with 1ml aliquots of the now EDC/NHS functionalised nanoparticles to generate a final functionalised nanoparticle with EDC/NHS-Nickel ligand complex. UV-Vis and DCS were then employed in characterising the ligand complex prior to immobilisation.

5.2.5 A colorimetric protein assay that is immediate and based on the spontaneous development of a protein corona on gold nanoparticles.

Determination of the amount of protein loaded onto nanoparticles was essential and thus, we explored a method by (Ho *et al.*, 2015) for BSA protein quantification after immobilisation. Different concentrations of NaCl ranging from (1M to 5M) were used to test aggregation of nanoparticles before enzyme immobilisation. For the standard curve generation, nanoparticles were incubated with different concentrations of BSA from a stock of 10 mg/ml and incubated this at 25°C for 30 mins. After 30 mins, aggregation was induced by adding 1M NaCl to all nanoparticle enzyme dilutions and the UV-vis spectra (400-800nm) determined. Absorbance measurements were recorded at 520 and normalised to 580 nm which is the wavelength indicative of protein aggregation. The A520/580 ratio was plotted against the amounts of BSA in order to generate a standard curve which was used in the quantification of immobilised proteins. The BSA-nanoparticle solution was also analysed using a 0.5%

agarose gel where the migration of nanoparticles in the agarose gel was monitored to determine any retardation indicative of protein corona formation.

5.2.6 Immobilisation of glucose oxidase (GO)

In this study we have used glucose oxidase as this is a well-studied enzyme used in previous bioconjugation studies (Zhanfang and Teng, 2009). Enzyme-nanoparticle conjugates were prepared by mixing equimolar concentrations of nanoparticles and enzymes. Reaction mixtures were incubated overnight to allow GO immobilisation onto citrate stabilised nanoparticles with functionalised PEG-COOH. After overnight incubation, the reaction mixture was centrifuged at 11,000 xg for 2 mins and supernatant transferred into another tube. The nanoparticle (protein corona) that was formed was resuspended in the enzyme reaction buffer and used in subsequent enzyme reactions.

5.2.6.1 Glucose oxidase (GO) assay

GO assay was performed referring to the protocol described in Enzymatic Assay of Glucose Oxidase (Sigma, UK). Glucose oxidase from *Aspergillus niger* (G7141-10KU), Peroxidase from horseradish (P8250-5KU) and *o*-Dianisidine dihydrochloride (F5803-50MG) were purchased from Sigma, UK for our experiments.

5.2.6.2 GO-nanoparticle assays

For the assay using immobilised GO, 1ml of citrate stabilised nanoparticle with terminal carboxyl groups were bonded with side-chain amino groups of protein surface using ~0.2 mg/ml of GO through EDC/NHS coupling reaction. A plot of enzyme activity versus time was plotted to determine enzyme rate for different dilutions of enzyme. For the enzymatic assay of glucose oxidase, the unit definition was given as; 1 unit oxidizes 1.0 μmole of β-D-glucose to D-gluconolactone and H_2O_2 per minute at pH 5.1 at 35ºC (equivalent to an O_2 uptake of 22.4 μl per minute) in a

reaction mix saturated with oxygen for 1 hour. See the schematic for the enzymatic reaction in Figure 5.4.

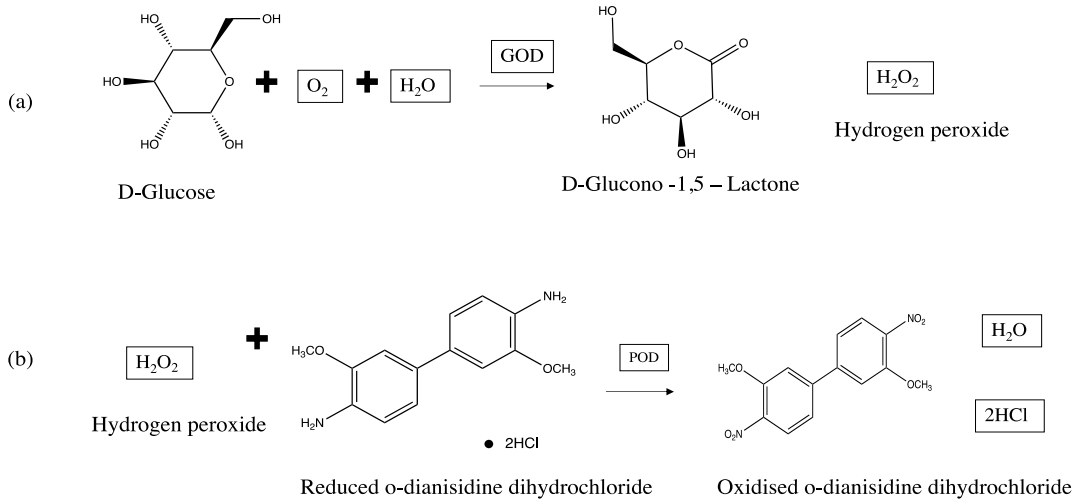


Figure 5.4 schematic showing the enzymatic conversion by glucose oxidase

In Figure 5.4 (a) Coupled enzymatic reaction involving the conversion of D-glucose to D-glucono – 1,5 lactone and hydrogen peroxide by glucose oxidase (GO) in the presence of oxygen and water. Represented in (b) Oxidation of O-dianisidine dihydrochloride (ODDH) by peroxidase (POD) in the presence of peroxide, leading to the formation of oxidised ODDH which has an observable colour of brownish to orange, which was measured spectrophotometrically to determine the amount of glucose converted.

5.2.7 Enzyme immobilisation and activity testing of alpha-L-rhamnosidase

Alpha-L-rhamnosidase nanoparticle conjugates were formed by reacting protein solutions with EDC/NHS-Nickel ligand complex. We allowed 16 hours incubation time to achieve efficient covalent bonding between proteins and EDC/NHS-Nickel ligand complex. After 16 hours of incubation, protein-EDC/NHS-Nickel ligand complex was centrifuged at 6,000 x g for 2 minutes and the supernatant transferred to another tube. The protein - EDC/NHS-Nickel ligand complex was then resuspended in 20mM HEPES, pH 5.47, before activity testing of the protein-EDC/NHS-Nickel ligand complex. Activity testing was carried out using optimum temperature and optimum pH as displayed in Figure 5.11 Enzyme activity testing with nanoparticles.

Reactions were carried out at 50°C, pH 5.95, over a period of 8 hours and the activity after every 30 mins calculated.

5.2.8 Cross Linked Enzyme Aggregates (CLEAS) of acetyl esterase

Proteins were precipitated using ammonium sulphate achieving 45% saturation before enzyme cross linking. Precipitation was initially performed at room temperature for 2 hours then transferred to a cold room at 4°C overnight while mixing at 150 rpm. After 16 hours of precipitation, proteins were harvested by centrifugation at 4°C at 16,000 x g for 10 mins followed by resuspension in 50mM sodium phosphate buffer, pH 7.5.

5.2.9 Glutaraldehyde standard curve using BSA

A glutaraldehyde standard curve using BSA of known amounts within the range (0.2 – 0.8 mg/ml) was developed to quantify proteins after cross-linking. Cross-linked enzyme aggregates were achieved following the introduction of glutaraldehyde at (0.05% and 0.1%) to known amounts of protein and incubation at room temperature overnight. Activity testing of the cross-linked enzyme aggregates followed the same protocol of activity testing as described in section 4.2.8.3.

5.2.10 Enzyme assay using cross-linked enzymes

Activity testing of enzymes after cross-linking was the same as described in section 4.2.8.3 with minor changes. Here testing was carried out at 40°C which was the determined optimum temperature for activity. Reactions were carried out for 5 minutes and activity calculated per mg of protein amount in the enzyme reaction mix.

5.3 Results

5.3.1 Characterisation of nanoparticles by agarose gel and SDS -PAGE

Synthesised PEG-COOH nanoparticles have an overall negative charge compared to naked nanoparticles and ass seen in the migration pattern observed in Figure 5.5(a).

In the SDS – PAGE we observed band smearing in the wells where protein coronas were loaded, and no band smearing observed in the wells which had only nanoparticles loaded.

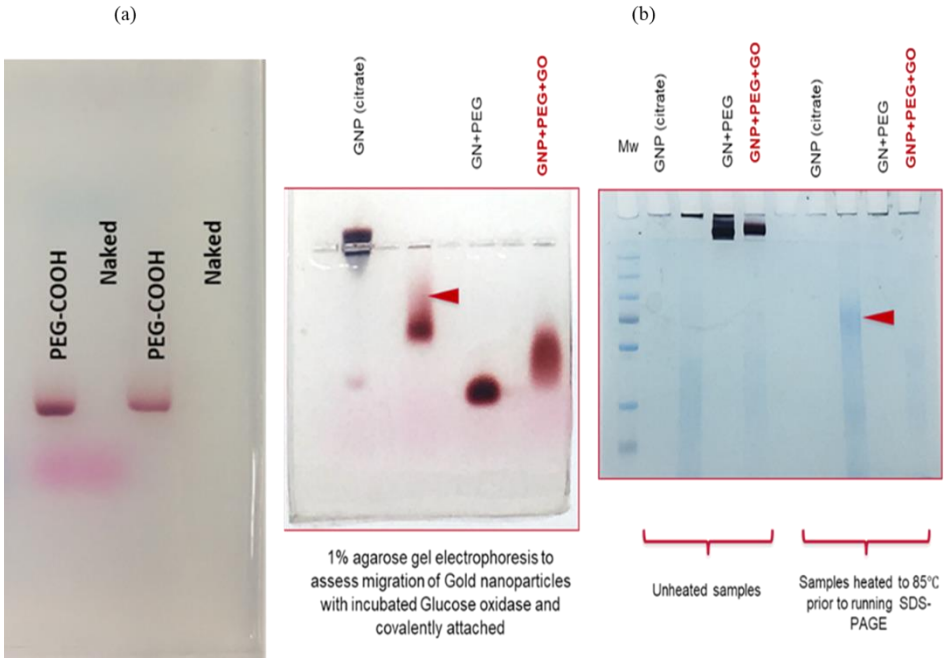


Figure 5.5 Agarose gel electrophoresis and Native-PAGE

In Figure 5.5 (a) 0.5% agarose gel showing naked nanoparticles and PEG-COOH functionalised nanoparticles (b) 0.5% agarose gel showing nanoparticles with immobilised glucose oxidase alongside 7.5% Native-PAGE. Nanoparticle samples were loaded directly onto gels without loading dye but for the Native-PAGE samples were mixed with 2X Laemmli buffer without SDS and run under non-denaturing conditions. Pageruler plus pre stained marker was loaded as reference and staining of Native gel was achieved using instant blue (Expedeon, UK). **Red arrow** indicates lane with unconjugated Glucose oxidase and from the Native-PAGE successful immobilisation of glucose oxidase onto nanoparticles can be seen.

5.3.2 Characterisation of nanoparticles by UV – vis

Gold colloidal dispersion colour changes can be seen with the naked eye and are linked to changes in SPR. During the formation and aggregation features of nanoparticles, changes in SPR were detected by observing shifts in absorbance peaks at OD 520nm, which is a specified wavelength of 15 nm nanoparticles and used as an

indication for changes in particle size, surface chemistry, or interparticle distances (Krpetić *et al.*, 2012).

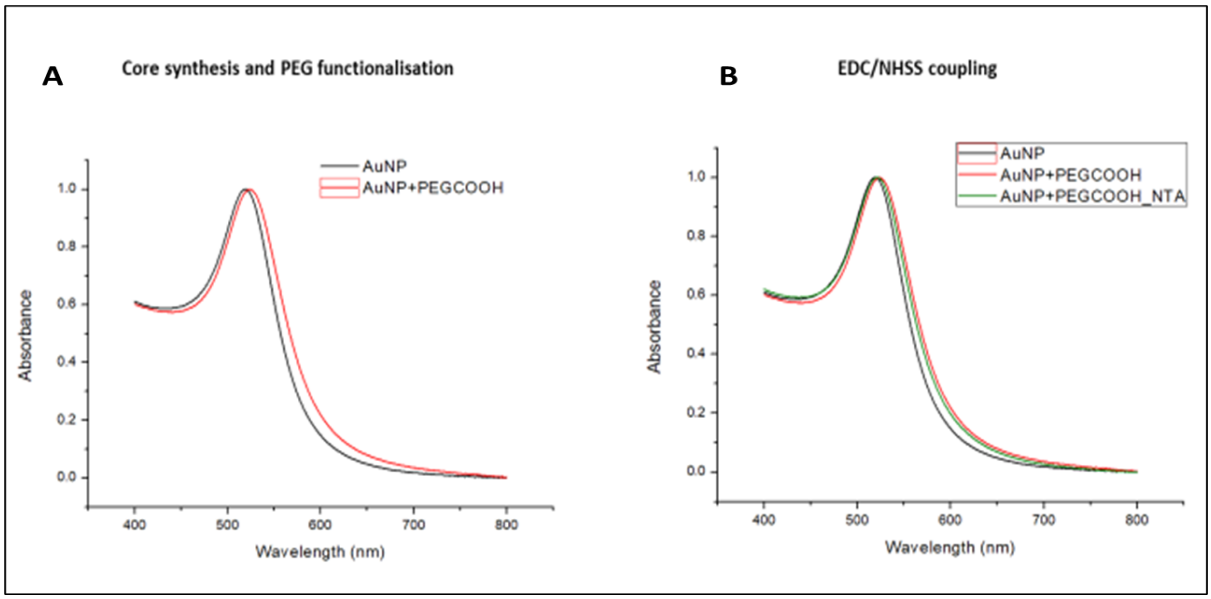


Figure 5.6 UV-vis characterisation of citrate stabilised nanoparticles and functionalised nanoparticles

Key: Core_citrate_AuNP = , PEG_COOH functionalised = , PEG_COOH functionalised with NTA linker = , PEG_COOH functionalised with NTA linker and Nickel ligand =

In Figure 5.6 (A) UV-vis image shows peaks with no disturbances detected, indicating no observable alterations of surface plasmon resonance. An observable peak is seen at ~520nm which is the peak SPR wavelength for 15 nm AuNPs. The absence of no alterations to surface plasmon resonance implied nanoparticles no aggregation was observed in samples. In (B) the citrate stabilised nanoparticle with

functional ligand attachment. In (C) we show functionalised citrate stabilised nanoparticles following coupling using EDC/NHSS. In (D) we depict the alterations of SPR following stabilised, functionalised, coupled nanoparticles with Ni to allow binding of his-tagged proteins.

5.3.3 DCS characterisation of particle size distribution

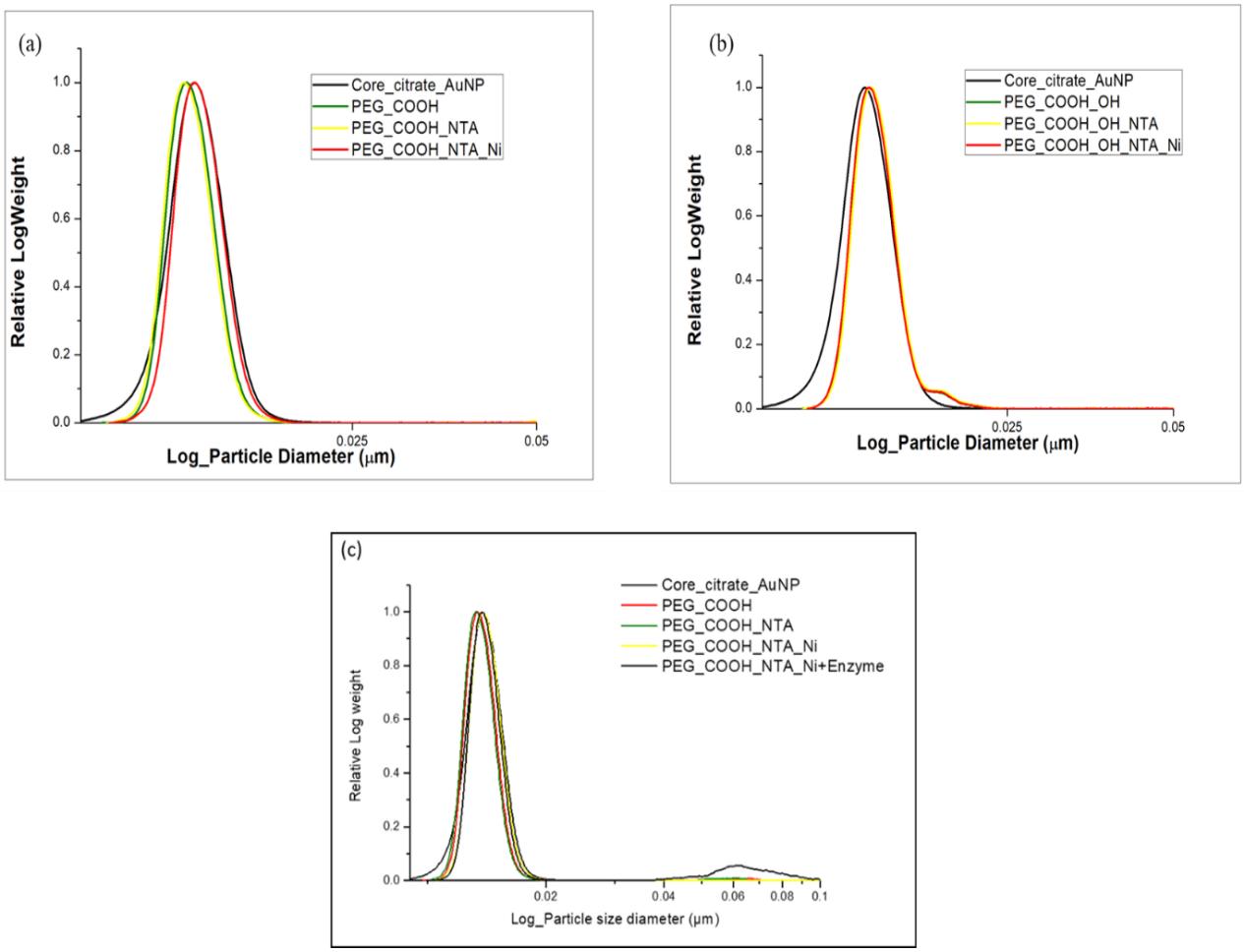


Figure 5.7 DCS graph plots for particle size distribution.

Key: Core_citrate_AuNP = , PEG_COOH functionalised = , PEG_COOH functionalised with NTA linker = , PEG_COOH functionalised with NTA linker and Nickel ligand =

In Figure 5.7, DCS plots of particle size distribution (a) Shifts to the left of the Log particle diameter scale observed following functionalisation and ligand attachment with a shift back to the right following Nickel attachment . (b) 50%v/v mixture of PEG_COOH and PEG_COOH_OH functionalised citrate stabilised nanoparticle with the NTA linker and Nickel ligand showed a shift to the right following functionalisation , ligand attachment and nickel attachment. (c) we show the successful attachment of histidine tagged protein to a PEG_COOH functionalised with NTA linker and Nickel

ligand (nickel/EDC/NHS) nanoparticle complex. However, we observed an alteration around 58 μm, the bump observed in (c) which could be due to sample loading error or some aggregation. Data plot from (a) provides information of the increase in particle diameter following functionalisation but data plot from (b) requires a repeat to understand why the shift was to the right suggesting of increase in particle diameter following PEGCOOH_OH functionalisation and coupling with EDC/NHSS_NTA.

5.3.4 A colorimetric protein assay that is immediate and based on the spontaneous development of a protein corona on gold nanoparticles.

A known amount of BSA was immobilised onto nanoparticle support to test the concept of protein corona formation and how concentration of salts affect protein corona formation. Changes in SPR were observed following protein corona formation by the addition of different concentrations of NaCl ranging from (0 – 5 M). Successfully formed protein coronas showed no aggregation in the presence of salts as depicted in the images below.

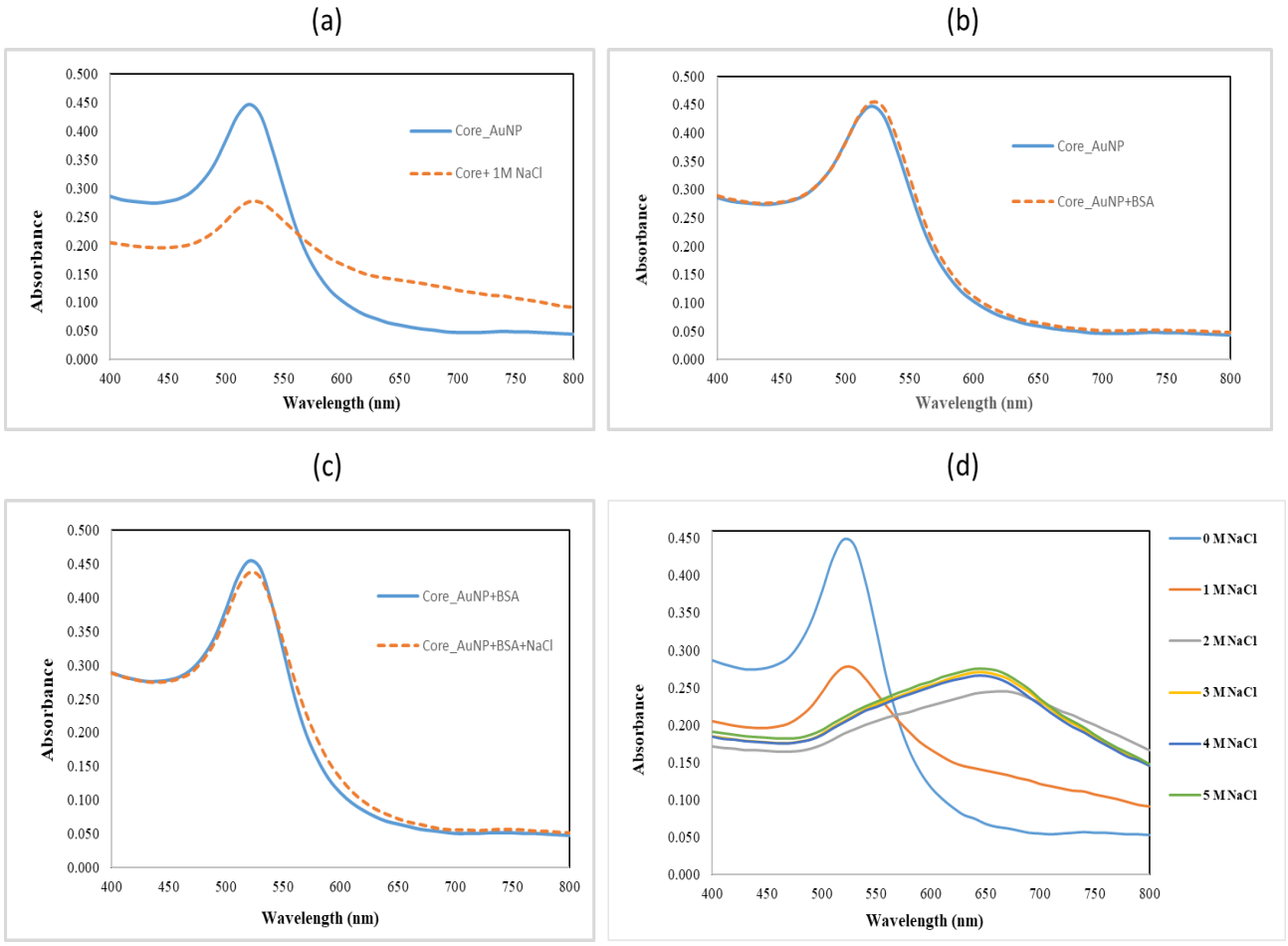


Figure 5.8 UV-Vis absorption spectrum for colorimetric protein assay

In Figure 5.8 (a)UV-Vis absorption spectrum (400 -800nm) showing colloidal stability using 1M NaCl. (b) UV-Vis absorption spectrum (400 -800nm) of nanoparticles after addition of BSA. (c) UV-Vis spectrum (400 -800 nm) showing colloidal stability of naked nanoparticles incubated with 30 µg of BSA. (d) UV-Vis spectrum showing the degree naked nanoparticles stability following the addition of different concentrations of salt. In (c) we observed no aggregation of nanoparticles as a result of BSA attachment to nanoparticles demonstrating the colloidal stability of nanoparticles with formation of a protein corona.

5.3.5 Protein standard curve

The A520/A580 nm measurements were subsequently used to plot a concentration calibration curve using BSA as standard. Different concentrations of BSA were immobilised onto nanoparticles and different amounts of BSA protein solution were

measured at A520/A580 nm and calculations determined as a function of protein amount.

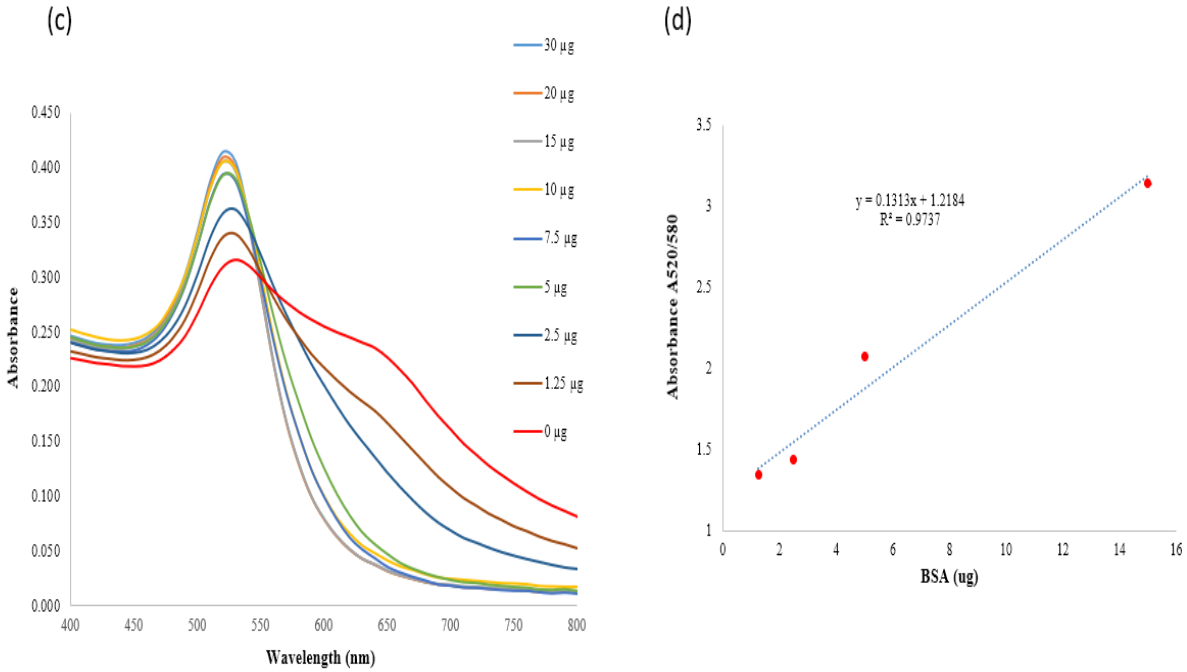


Figure 5.9 Protein amount calibration curve of BSA obtained via sequential approach

In Figure 5.9 (c) UV-Vis absorbance spectrum measurement with increasing amount of BSA where absorbance values at peak 520 nm were seen to increase with increasing BSA amounts. (d) determined linear range of the protein assay for the quantification of protein amount between 1.25 µg – 15 µg of BSA. The linear regression given by $y=0.1313x+1.2184$ and $R^2= 0.9737$, was used in the estimation of protein amounts prior to testing. Standard deviation errors were calculated from triplicate measurements ($n=3$). The results suggest that the sequential approach of BSA protein corona formation can be used successfully in the generation of a standard curve for the quantification of amount of protein loaded on gold nanoparticles.

5.3.6 GO assay (Proof of enzyme immobilisation concept)

Free GO was tested to determine the linearity of the assay using two different concentrations of the free enzyme where 0.047 mg/ml resulted in a linear assay over a period of about 7 minutes as seen in Figure 5.10.

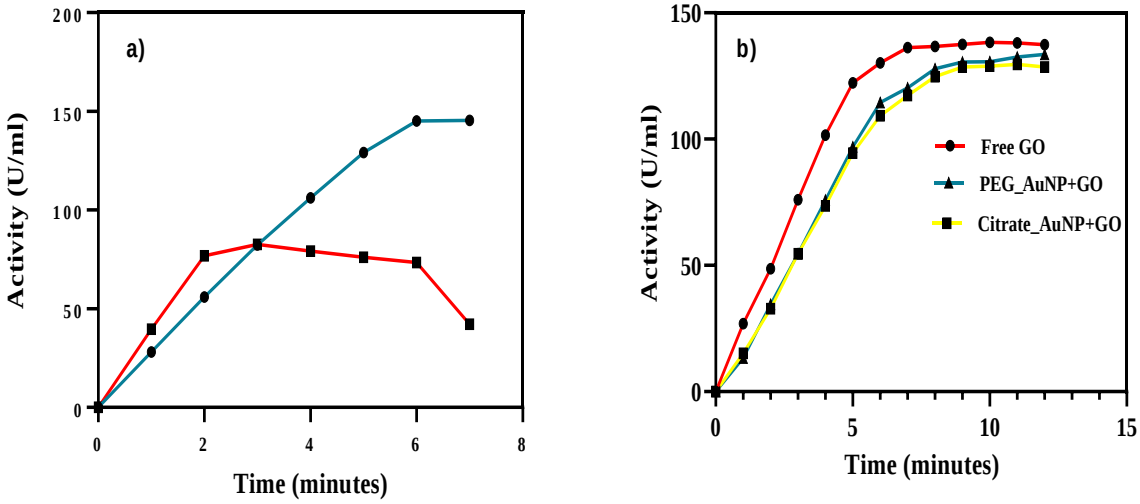


Figure 5.10 Enzyme activity assay plot for Free-GO and GO + AuNP

In Figure 5.10 we demonstrate the linearity of glucose oxidase assay using (a) 0.095mg/ml of glucose oxidase (red) and 0.047mg/ml (blue) over 7 different time points. (b) Enzyme activity plot comparing performance of glucose oxidase, Citrate stabilised nanoparticle with glucose oxidase and PEG functionalised nanoparticle with glucose oxidase. From the Activity graphs plotted, in (a) reduction in enzyme activity was observed after 3 mins using 0.095 mg/ml suggesting high substrate conversion rate compared to 0.047 mg/ml where activity still increased over a period of after 7 mins. Therefore, determining enzyme activity with 0.047 mg/ml of free enzyme allowed appropriate estimation of activity and calculation of kinetic parameters. From the Activity graphs plotted in (b), the results suggest that free enzyme activity was ~ 5U higher than PEGCOOH_EDC/NHS which was also 5U higher than citrate stabilised nanoparticles after 12 minutes of incubation. The results therefore demonstrate successful immobilisation and activity testing of glucose oxidase as a concept development step prior to immobilisation of our expressed protein.

5.3.7 Activity test of free alpha -L-rhamnosidase and alpha -L-rhamnosidase-EDC/NHS-Nickel ligand complex.

After the success of the immobilisation concept development, expressed alpha -L – rhamnosidase nanoparticle ligand complex was tested for activity using optimum temperature and optimum pH as displayed in Figure 3.10. Reactions were carried out at 50°C, pH 5.95, over a period of 8 hours and the relative activity after every 30 mins has been plotted against time as displayed below. Free enzymes activity was also performed to determine the effect of nanoparticle on enzyme activity.

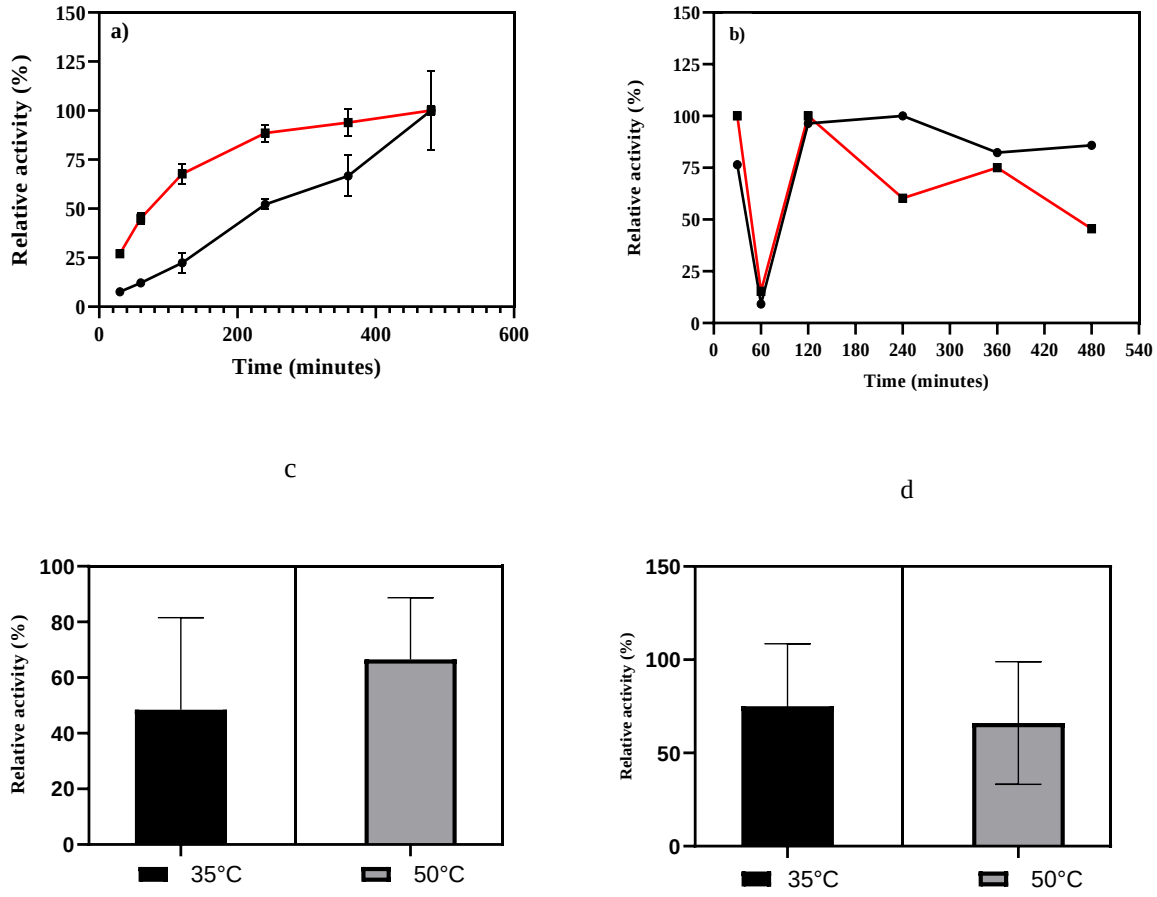


Figure 5.11 Enzyme activity testing with nanoparticles

In Figure 5.11 (a) Activity testing over a period of 8 hours using free enzymes. (b) Activity testing over a period of 8 hours using enzyme-nanoparticle complex. In both (a) and (b) activity testing was carried out at (red = 50°C, black = 35°C). In the free enzyme plot, activity is seen to increase steadily over the 8-hour period with activity at 50°C being higher than 35°C as seen in (c). However, in the nanoparticle assay plot, a vertical reduction in activity was seen after 1 hour followed by a sharp increase in activity

back to 100% after 2 hours. This could possibly be linked to nanoparticle aggregation, as settled nanoparticles were observed in the microtiter plates post reaction. We also believe the sharp increase in activity after 2 hours could be linked to enzyme dissociation due to poor compatibility. Based on this we could not conclude on the improvement in enzyme functionality. Error bars indicated on graph plots were calculated percentage errors showing the percentage variation in each data point.

5.3.8 Biochemical activity testing of cross-linked esterase enzyme

5.3.8.1 Characterisation of esterase enzyme cross-links

Here we report the results from our activity testing employing the technique of enzyme cross linking to form CLEAs. Cross-linked enzymes were characterised through visual inspection after incubation at different time points. After 13 hours of incubation, we characterised the final cross-linked enzymes using UV-vis spectroscopy and observed alterations to protein peaks as observed in the peak shifts and disturbances at 280 nm represented in Figure 5.12.

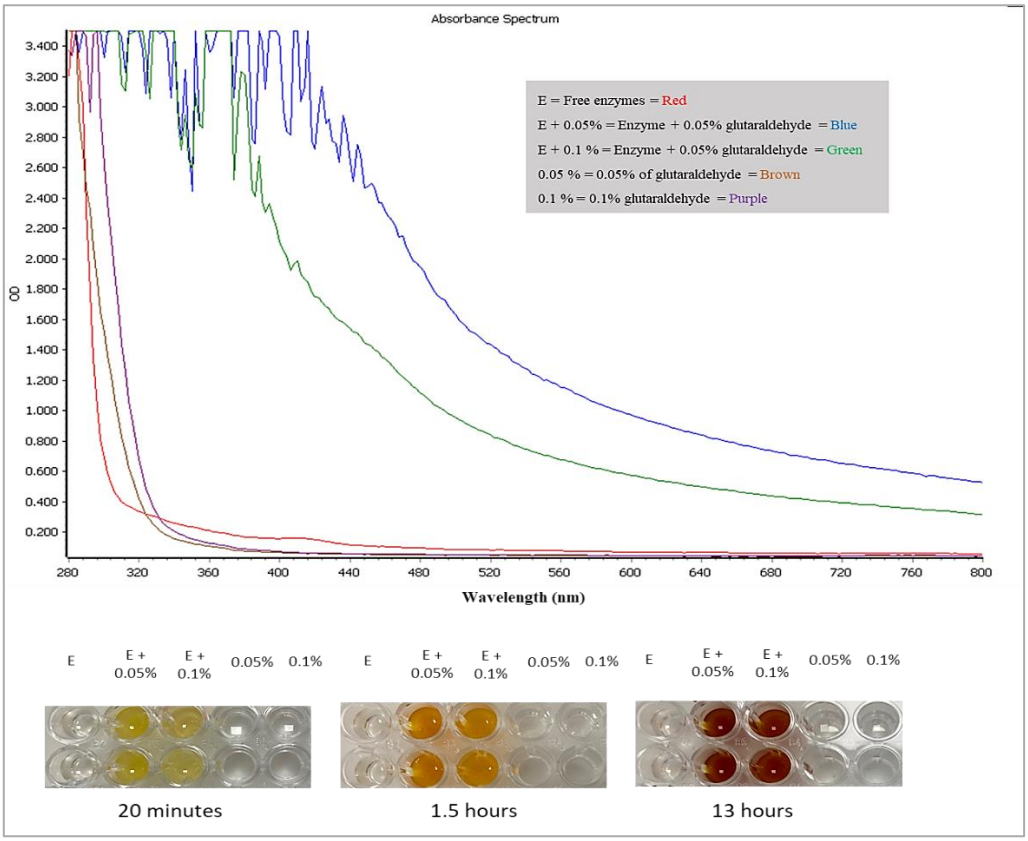


Figure 5.12 Absorbance spectrum of crosslinked enzymes vs glutaraldehyde blanks

In Figure 5.12, we observed following comparison, a shift in peaks to the right of the spectrum in addition to high degree of disturbance following cross linking of our enzyme. The observable shift in peaks served as a confirmation of successful cross-linking following 13 hours incubation of enzyme in 0.05% and 0.1% glutaraldehyde. Observable colour spectra can be seen with the naked eye in the microplate images after incubation at various times indicating that colour changes following glutaraldehyde addition can also serve as an indication of successful enzyme cross-link formation.

5.3.8.2 Instantaneous colorimetric assay for protein quantification with glutaraldehyde

In order to measure equal amount of protein used in both free enzyme and cross-linked enzyme assays, a standard curve using BSA, and 0.05% glutaraldehyde was developed. The linear range of the glutaraldehyde-BSA standard curve was compared to a BSA only standard curve to validate our methods before using the glutaraldehyde-BSA standard curve for protein quantification as seen in Figure 5.13.

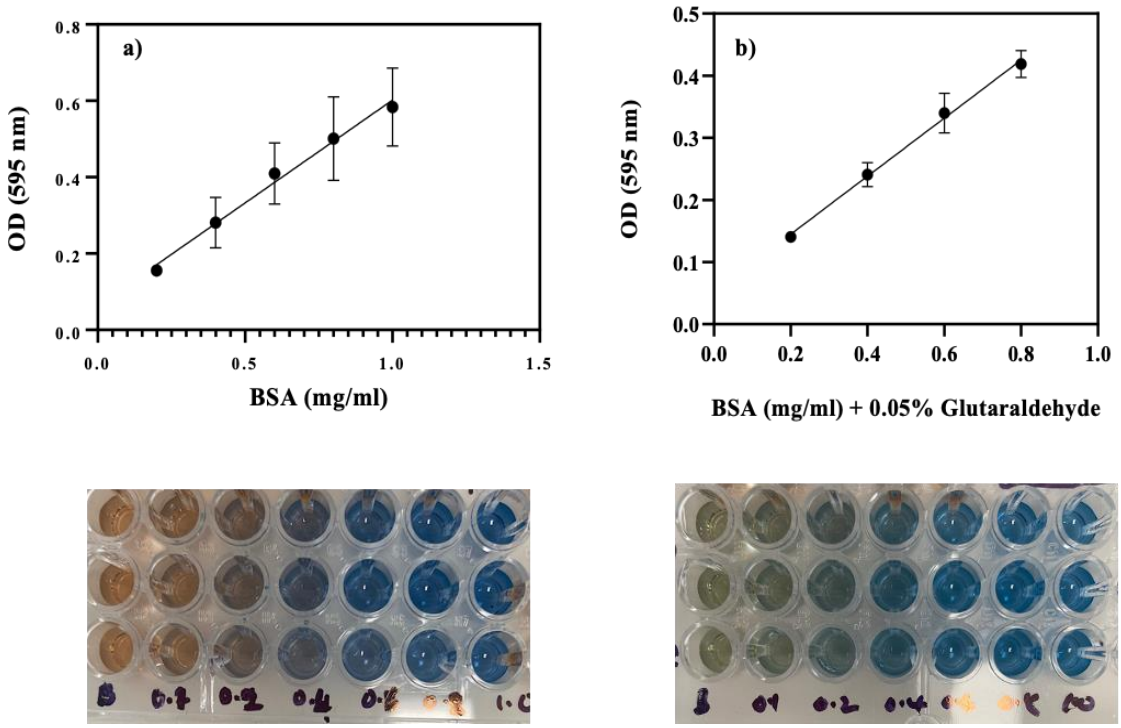


Figure 5.13 Standard curve of BSA and BSA + 0.05% Glutaraldehyde

In Figure 5.13 (a) Standard curve for BSA-only with determined linearity between 0.2 – 1.0 mg/ml and R^2 = 0.9900 via the microplate method. (b) Standard curve for BSA + 0.05 % glutaraldehyde with

determined linearity between 0.2 -0.8 mg/ml and $R^2 = 0.9972$ via the microplate method. Standard errors have been calculated and depicted as error bars where n= 3.

5.3.8.3 Comparison of free enzyme and cross-linked esterase enzyme activity

The results from our activity testing using free enzyme and cross-linked enzymes were compared and analysed. Activity of the commercial enzyme purchased was also performed and compared to the activity of our enzyme following cross linking. Statistical tests were used to determine significant differences in our activity testing and depicted in the graphs in Figure 5.14 is the output of our analysis.

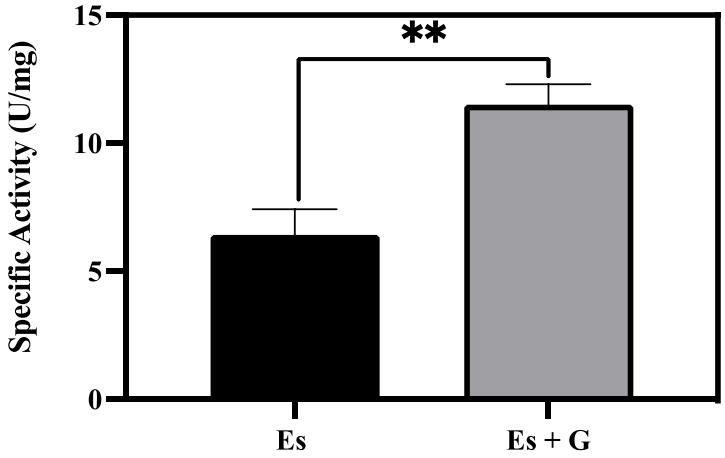


Figure 5.14 Activity plots of free and cross-linked esterase enzyme

Key
Es = Free enzymes
Es + G = Enzyme + 0.05% glutaraldehyde

Figure 5.14. Specific activity plots from measuring the rate of reaction catalysed by cross-linked enzymes expressed as micromoles of substrate transformed (or product formed) per minute per mg/ml of enzyme used.

In Figure 5.14, specific activity of free and cross-linked expressed esterase enzyme activity at 40°C is compared using 5mM pNPA in a total reaction volume of 240 µl. (a) Es activity was compared to ES + G and analysis using the Mann Whitney test indicated a significant difference in observed medians with p = 0.0079 (**) suggesting an improvement in enzyme functionality when cross-linked.

5.4 Discussion

The use of enzyme immobilization techniques such as nanoparticles (Gupta *et al.*, 2011), hydrogels (Lozinsky, 2018; Wang *et al.*, 2019) and cross linked enzyme aggregates (Sheldon and van Pelt, 2013) are said to improve the functionality of enzymes. Nanoparticles have been reported to reduce contamination issues when superparamagnetic nanoparticles were used (Gupta *et al.*, 2011). Hydrogels have been employed in the immobilisation of whole cells since the 1970s as an effective method to improve the performance and economics of many fermentation processes (Zhu, 2007). Cross linked enzyme aggregates and cross linked enzyme crystals have also been described in developing biocatalyst with improved functionality, recyclability and thermal stability using both purified and cell lysate proteins (Margolin and Navia, 2001; Fernandez-Lafuente, 2009; Sheldon, 2010, 2011).This among many other advantages exhibited using enzyme immobilisation has led to research into using immobilised enzymes as biocatalysts with improved functionality towards industrial applications. It is with these same reasons why this study aimed to test two immobilization techniques to improve the functionality of expressed proteins by improving thermal stability of an expressed alpha-L-rhamnosidase and improving thermal stability of an expressed carbohydrate esterase family 1 enzyme.

5.4.1 Thermal stability improvement of an α-L-Rhamnosidase using gold nanoparticle complex

α-L-Rhamnosidases are extremely valuable industrial catalysts, however, their low thermal stability and selectivity still restrict some of their applications (Guillotin *et al.*, 2019). The ability to function under high temperatures during pre-treatment of biomass will allow the efficient bioconversion of biomass into bioproducts. Therefore, thermal properties conferred by enzyme immobilisation onto supports such as nanoparticles can aid rhamnosidase in lowering industrial input and this is essential for its success and potential use as an environmentally friendly technique to regulate

the biological and pharmacological properties of flavonoid molecules such as hesperidin, naringin, diosmin and rutin.

In this study, an alpha-L-rhamnosidase-EDC/NHS-Nickel ligand complex was successfully developed through the synthesis of 15 nm nanoparticles in accordance with modified protocols from (Turkevich, Stevenson and Hillier, 1951; Krpetić *et al.*, 2013). Characterisation of the synthesised nanoparticles confirmed no alterations to surface plasmon resonance and particle size distribution was accurate for 15 nm nanoparticles as indicated in the UV-Vis and DCS results displayed in Figure 5.6 and Figure 5.7 respectively. Protein quantification after enzyme immobilization was successfully achieved using BSA Immobilisation onto nanoparticles as described in the study by (Ho *et al.*, 2015). The expressed alpha–L-rhamnosidase was successfully immobilised onto 15nm functionalized nanoparticles and activity testing performed over a period of 8 hours measuring the rate over time. In Figure 5.11 (a) activity testing results indicated significant increase in relative activity when free enzyme was incubated at 35°C and 45°C over a period of 8 hours. However, this was not the same for our enzyme-nanoparticle complex. We observed mixed results where after 1 hour, a sharp decline in activity was observed followed by an increase in activity after incubation at 35°C and 45°C. The possible reason for this observation could be linked to nanoparticles settling at the bottom of the microplate and affecting the absorbance readings. This was surprising as enzymes conjugated to nanoparticles are said to exhibit Brownian movement suggesting that nanoparticles remain suspended in solution (Gupta *et al.*, 2011; Ahmad and Sardar, 2015). It is unclear why this aggregation was observed because our colloidal stability testing indicated that successfully formed protein corona shows no aggregation when high concentration of salts such as NaCl are added as seen in Figure 5.8. As a result, small traces of buffer salts were not expected to cause aggregation of nanoparticles. On this basis, our enzyme nanoparticle complex was perhaps not biocompatible, therefore why after 1 hour dissociation occurred. Biocompatibility has been described to be an issue with

nanoparticles (Brodelius and Mosbach, 1987; Rodriguez *et al.*, 2010) so it is not surprising that such results were observed in this study.

5.4.2 Operational functionality improvement of carbohydrate esterase family 1 enzyme using glutaraldehyde cross linked enzyme aggregates

To address challenges such as biocompatibility when using nanoparticles, cross linked enzyme aggregates were formed by the addition of glutaraldehyde. Glutaraldehyde has been used as a bifunctional agent to achieve cross linking in various applications improving the functionality of enzymes (Nimni *et al.*, 1987). The result from this study shows the successful formation of cross-linked enzyme aggregates as displayed in Figure 5.12 showing the incubation times of the glutaraldehyde cross-linking and absorption spectrum showing changes in protein peaks at 280 nm. Cross-linked enzyme aggregates were successfully quantified through a colorimetric assay by cross linking BSA as seen in Figure 5.13. Our standard curve with glutaraldehyde-BSA was comparable to the standard curve of BSA alone suggesting confidence in reliable protein quantification results given by $R^2 = 0.9972$ and $R^2 = 0.9900$ respectively as seen in Figure 5.13. Quantified cross-linked enzymes were then used in an assay where the enzyme activity per mg of protein was calculated and plotted against time. The difference in activity per mg of protein in free enzyme and cross-linked enzymes was compared through a paired T-test, assuming non-parametric gaussian distribution based on the median between free enzyme and cross-linked enzymes. The Mann Whitney test was used for this analysis and the resulting statistical analysis indicated significant difference between free enzyme and cross-linked enzyme activity given by p value = 0.0079 (**). Therefore, our attempt to improve the functionality of the expressed esterase was successful.

Table 5.1 Summary of pros and cons realised using immobilisation techniques in this study

Immobilisation technique	Pros	Cons	Comment
Nanoparticles	None realised from our study	• Expensive • Requires purified proteins • Tedious process involved in producing proteins and then purifying proteins before conjugation	• Time consuming, • Issues with protein load quantification • aggregate formation during assay at high temperatures • biocompatibility issues Further improvements needed to develop assays.
Cross-linked enzyme aggregates	• Less materials required • Easy technique requiring little expertise • Not expensive • Cross linking can be achieved rapidly	• Determining appropriate time for 100% cross-linking • Free enzymes might still be in solution	• Less expensive technique requiring little expertise. • Less tedious as cross-linked enzyme aggerates can be achieved using cell lysates.

5.5 Conclusion

As a conclusion to our study on improving the operational functionality of biocatalysts identified from the slug gut metagenome using enzyme immobilisation techniques, our findings suggest that, although enzyme immobilisation onto nanoparticles is feasible, enzyme load quantification and requirements such as tedious processes in the production of pure enzymes and synthesis of nanoparticles, creates a massive drawback in its application. Following synthesis of said nanoparticles, there are uncertainties regarding enzyme-nanoparticle compatibility and improvement in the functionality of the immobilised enzyme. From this study, thermal stability testing results were inconclusive when the expressed α-L-Rhamnosidases was immobilised onto nanoparticles and tested in a continuous de-rhamnosylation method. Biocompatibility was perhaps the most concerning issue with this method of immobilisation, especially when enzyme-nanoparticle conjugates were incubated at optimum enzyme activity temperature and ph.

To address drawbacks with the use of nanoparticles, cross linked enzyme aggregates were produced achieving cross links in less than 10 minutes. The process was relatively easy, less expensive and the possibility of using cell lysates made this process more appealing than the use of purified enzymes and nanoparticles. Glutaraldehyde induced cross-linked enzyme aggregates used in this study showed significant improvement in enzyme activity seen in the increase in specific activity when the same amount of non-cross-linked enzyme and cross-linked enzyme was used. On this basis our findings suggests that cross linked enzyme aggregates offer an easier, rapid, less tedious process to achieve biocatalysts with potential industrial applications in targeting underutilised hemicellulose rich fraction of biomass to aid other hemicellulases gain accessibility to sugar rich fractions. Interestingly, multiple enzymes can be linked by enzyme cross links forming COMBI-CLEAS which can be used in a one pot enzymatic reaction. We recommend future work using the produced enzymes from this study to form COMBI-CLEAS and perform activity testing using optimised conditions. Other studies have also demonstrated that directed site mutagenesis could improve enzyme thermostability by introducing or deleting sequences which improve the function of the recombinantly expressed enzyme when activity testing was carried out(Zhang *et al.*, 2014; Sukul, Lupilov and Leichert, 2018).

Chapter 6. Contribution to knowledge and future research

6.1 New genome assembly

With the current advances in metagenomics, the exploration of previously unculturable diverse environments is now possible including soil (Patra, Das and Shin, 2018), guts of humans (Qin *et al.*, 2010; Flint *et al.*, 2012), ruminants (Stackebrandt and Hippe, 1986), and invertebrates such as termites (King *et al.*, 2010; Hess *et al.*, 2011; Brune, 2014; Franco Cairo *et al.*, 2016) and cockroaches (Bertino-Grimaldi *et al.*, 2013). The possibility of this exploration has led to expeditions where novel sequences with glycosyl hydrolase activity towards biotechnological applications were identified (Rees *et al.*, 2003; L. L. Li *et al.*, 2009; Cardoso *et al.*, 2012; Izzo *et al.*, 2014; Joynson *et al.*, 2017). Our initial aim was to clone, express and test protein sequences belonging to glycosyl hydrolases from a previous study by (Joynson *et al.,* 2017) . This was successfully achieved by validation through PCR, cloning these genes and recombinantly expressing proteins towards activity testing. Another aim from this study was to improve on the partial sequences that were produced from the previous assembly by (Joynson *et al.,* 2017), employing different assemblers and increasing kmer size for assembly. This improvement in assembly was successfully achieved using higher kmer values allowing better contig assembly. Our reason for selecting higher kmer was due to the mere fact that sequences generated form metagenomic reads were greater than 65 bp which requires kmer sizers higher than 51 for good assembly output (Zerbino and Birney, 2008). MetaSPAdes was chosen because of its capacity to solve a variety of issues in metagenomic assembly by leveraging computational approaches that have previously proven beneficial in single-cell assemblies and highly polymorphic diploid genome assemblies. From our results we can confidently conclude that metaSPAdes did improve the metagenome assembly than was previously described with more coding DNA sequences identified following prokka metagenome annotation using metaSPAdes scaffolds. This improvement will allow further work investigating the slug gut metagenome for glycosyl hydrolases,

get access to complete sequences which perhaps will encode full functional enzymes. We believe our pipeline can also be used for other studies focusing on identifying complete sequences which will lead to the production of functional enzymes following recombinant protein expression.

6.1.1 Enzymes for the future

Enzymes capable of hydrolysing hemicellulose, lignin, pectin and the cellulose fraction of the plant cell wall have been produced from our work. An α-L-rhamnosidase and an esterase have been successfully expressed and tested with enhanced enzyme activity characterisation achieved in this study for the esterase enzyme. Also expressed is a beta glucosidase, an auxiliary active family enzyme and an endo 1,4-beta xylanase which have been handed over to be used in future experiments. Following our success with expression and purification of enzymes, we can confidently confirm the true existence of carbohydrate active enzymes in the gut of *Arion ater* and recommend future work in testing activity of produced enzymes to proceed. Modifications such as the addition of chemicals to enzymes could improve the enzyme functionality. For example, in a work by (Ge *et al.*, 2017), a sugar alcohol commonly known as sorbitol, was added to an expressed α-L-rhamnosidase resulting in enhanced stability. Site-directed mutagenesis has also been described to improve the activity of esterase enzymes (Zhang *et al.*, 2014).

In addition, we demonstrated the possibility of biocatalyst development through enzyme immobilisation. The use of immobilized enzymes has been described as an attractive approach. However, the technique still needs optimisation to meet industrial scale bioconversion demands. This will allow a more sustainable approach to achieving conversion of biomass into sugars and aromatic monomers considered as high value chemicals for industries such as, food and drugs, biofuel, and specialty chemicals. We recommend future work to focus on using Cross-linked enzyme aggregates (CLEAS) as our study found various advantages and recorded an increase

in enzyme function when such technique was employed. CLEAS are easy to handle, cheap to make and require less time in the formation making this technique a feasible method of enzyme immobilisation. Cross linked enzyme aggregates can also be used to conjugate multiple enzymes to form Combi-CLEAS to achieve one pot enzymatic reactions. Expressed proteins from this study could be linked together forming multifunctional CLEAS and activity testing can be carried out with optimised conditions from this study.

6.2 Potential industrial applications of α-L-rhamnosidase

α-L-rhamnosidase (EC 3.2.1.40) have been implicated in several industrial applications with a significant interest in its application as biocatalysts in a variety of food, pharmaceutical and chemical industrial processes (Manzanares *et al.*, 2007). In the food industry, α-L-rhamnosidase is used in the debittering of citrus fruits by breaking down naringin, the principle bitter flavanone glycoside found in citrus fruits (Puri and Banerjee, 2000; Raveendran *et al.*, 2018). In other works, it has been described in the enhancement of wine and grape juice aroma (Caldini *et al.*, 1994), studying plant structure and bacterial polysaccharides (Michon *et al.*, 1987). α-L-rhamnosidase has been characterised in their usage for de-rhamnosylation of natural chemicals as well as reverse rhamnosylation of a variety of minor organic molecules. (De Winter *et al.*, 2013). The reason for the interest is linked to the use of Isoquercitrin, a de-rhamnosylation product of rutin that has been linked to a variety of biological benefits, including anti-mutagenesis, anti-virus, anti-hypertensive, anti-proliferative effects, lipid peroxidation, oxidative-stress protection, and other pharmacological effects. (Walle, 2004; Amado *et al.*, 2009; Ju You, Jin Ahn and Ji, 2010; Valentová *et al.*, 2014). Finally we recommend enzyme cross linking of an alpha-L-rhamnosidase with the beta glucosidase expressed also from the slug gut to achieve a one pot reaction which

can lead to the production of quercetin which has huge pharmaceutical applications as an antioxidant and anti-inflammatory (Yadav, Chaudhari and Kothari, 2011).

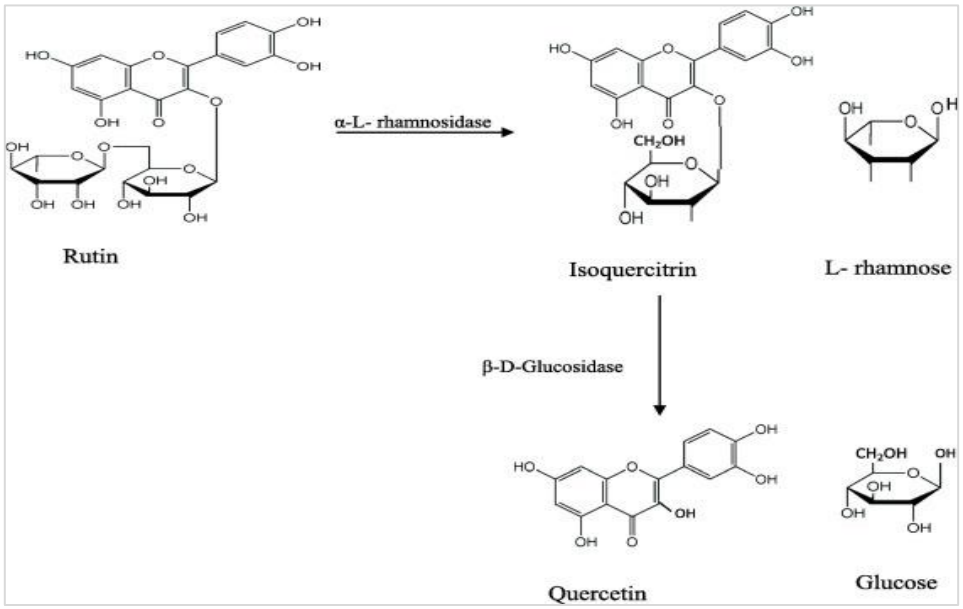


Figure 6.1. Illustration showing potential application of alpha-L-rhamnosidase enzyme in conjunction with beta-D-glucosidase enzymes towards generation of pharmaceuticals.

In Figure 6.1 is an illustration showing the potential application of alpha-L-rhamnosidase in the pathway leading to the production of Quercetin which has been described to have medicinal properties targeting diabetes/obesity and circulatory dysfunctions, inflammation and mood disorders or as an adaptogen (Chen *et al.*, 2016; Deepika and Maurya, 2022).

6.3 Potential industrial applications of the putative CE1

Acetyl xylan esterase (EC 3.1.1.72) and feruloyl esterase (EC 3.1.1.73) are examples of hemicellulase acting enzyme which facilitates the action of endoxylanases by increasing accessibility to the xylan backbone. This is achieved by cleaving the ester bonds of acetyl groups and the hydrolysis of 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar respectively (Sundberg and Poutanen, 1991) (Faulds and Williamson, 1991). Esterases have also been described to achieve ester hydrolysis and transesterification with very resulting in good enantioselectivity

(Turati *et al.*, 2019). This suggests that esterases will play an essential role in the release of sugars from xylan backbones which can be used in fermentation towards the production of bioethanol. Transesterification of short chain fatty acid such as Fatty acid methyl esters (FAME) have potential applications leading to the production of SCEE biodiesel (de Oliveira *et al.*, 2018). Therefore, the production of esterases could meet industrial demands to produce bioethanol through fermentable sugars as well as SCEE for biodiesel production.

Chapter 7. Appendices

7.1 Nucleotide sequences of genes used in this study

Nucleotide sequences retrieved from metagenomic library that were used in this study have been detailed below. Highlighted in green is position of the tag (CACC) that was introduced during PCR to facilitate cloning into pENTR. Highlighted in red is the forward primer site of the sequences are the regions targeted for PCR primer designs.

Gene_id_23621

Gene_ID_23621 (1,570 bp) sugar hydrolase [Chania multitudinisentens]

CACCATGTCTCAAGCAATCCAACAGCATATAAATCTCAACATGATCAGAGATGAGCATTTACTTGCTGTGGC
AGAGAAACTCATTCCAACAGTCTACAGTCGCGAAGTACAACCGCGTTCAATTGTTAAAAGTATTGAGGACGGA
AATGTAGCGCTTGGCTGGCGTGCCGAGCACATTGCAGATGCAACGACTTTTCTACAGCAGTCATTTTCACCGG
GTGACACCTTTATCATTGATTTTGGTGAGCATTGTGTCGGTTCACTCAATTTTTCATGTGATTCCACCGGGAGCC
CTCCTGATGCGCCAGCACATCTGCAGTTCATCTTTGGCGAAACGCTGTGTGAAGTTTCCGAGCCTTTCAGCGA
GTACCAGGGCTGGTTAAGTAGCAGTTGGCTGCAACAACATGATGAATACATCGATATCTTACCTGAAGAGAAA
ACGTTGCCCAGACGTTACTGTTTCCGCTATATCAAAGTGCGAGTCGTATCACTTTCACCGAAATATACCATTCG
ATTTACTCGCTTGAGCGCCACCACAGTGAGCTCAGCTCCGCACAGTTATCCTGAGTACCATACGCACGATGCA
CTGTTAAAAAAGATAGATGAGGTTTCGGTTCAAACCTTGCGCAATTGTATGCAGAACGTTTTTGAAGATGGAC
CTAAGCGTGACCGTCGTCTTTGGCTAGGCGATCTCCGCCTTCAGGCGTTGGTCAATGATGTCACCTTCGCGAA
ACATGATCTTGTAAAACGCTGTCTTTATCTCTTTGCAGGCCATAGAAGAGAAGACGGAATGGTTGCTGCTAATA
TTTTTGCCCGTCCTGAAGTAAAAGCAGACGATACCTATCTGTTCGATTATTCATTATTTTTTGTCGACACTTTATA
CAACTATTTAATGGCGACGCAGGATATAGCAACAGCAAAAGAGTTGTGGCCTGTAGCTATTCGCCAGGTTGA
ATTGGCACTGGAGCGATGTGATCACACCGGTCTGGTGAGAGACAGTGACGACTGGTGGTCATTTATTGACTG
GAATGATTCCCTTAATAAACAAGCTCCCTCACAAGGGGTTCTGATTTATTGCGTTGAAAAAGCCATTCAGCTTG
CCATGATTTGTCAACCGGAACATGGCGAGAGTTTGCAAAAGCAATTGCTGACGCTTAAAAATGCAGCATCAAC
TTTATTTGATGAAGAAACTGGATACTATTTTAGCGGGAATAAATATCAATTTAGTTATGCATCGCAAATATGGA
TGGTCTTGGCTGAAATGGGCGATTTAGTCCATCGGAAAAATATCATGCTAAACCTACTCAAAAATCCACCAAG
TATCGAAATGAATACCCCTTATTTACGCCATCATTTCATTGATGCCCTCATTAAATGCGGTTTGCGTAATGAAGC
CATTCAAGAAATTAAACAGTATTGGGGGGCAATGGTTGAATTTGGTGCTGATACATTCTGGGAGTTATTTGATC
CTAAGAACCCTGGTTTTTCTCCATATGGTAGTAAATTAATTAATAGTTACTGCCACGCCTGGAGTTGCACACCG
GCCTGGTTTATTCGCCAGTTCGGTCTCTAA

>Gene_ID_23621 amino acid sequence

MSQAIQQHINLNMIRDEHLLAVAEKLIPTVYSREVQPRSIVKSIEDGNVALGWRAEHIADATTFLQQSFSPGDTF
IIDFGEHCVGSLNFSCDSTGSPPDAPAHLQFIFGETLCEVSEPFSEYQGWLSSSWLQQHDEYIDILPEEKTLPRRYC
FRYIKVRVVSLSPKYTIRFTRLSATTVSSAPHSYPEYHTHDALLKKIDEVSVQTLRNCMQNVFEDGPKRDRRLWL
GDLRLQALVNDVTFAKHDLVKRCLYLFAGHRREDGMVAANIFARPEVKADDTYLFDYSLFFVDTLYNYLMAT
QDIATAKELWPVAIRQVELALERCDHTGLVRDSDDWWSFIDWNDSLNKQAPSQGVLIYCVEKAIQLAMICQP
EHGESLQKQLLTLKNAASTLFDEETGYYFSGNKYQFSYASQIWMVLAEMGDLVHRKNIMLNLLKNPPSIEMNT
PYLRHHFIDALIKCGLRNEAIQEIKQYWGAMVEFGADTFWELFDPKNPGFSPYGSKLINSYCHAWSCTPAWFIR
QFGL

Number of amino acids: 522

Molecular weight: 60074.35 Da

Theoretical pI: 5.53

7.2 Gene_id_23621 sequence alignment (pENTR cloning)

```
>>233621_seq (1150 nt)
 Waterman-Eggert score: 5502; 529.3 bits; E (1) < 2.2e-153
98.1% identity (98.1% similar) in 1137 nt overlap (577-1710:10-1146)

       580       590       600       610       620       630
23621_ GATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 GATTTTATTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTT
       10        20        30        40        50        60

       640       650       660       670       680       690
23621_ TATAATGCCAACTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCTTGTTTAACTTTAAGAA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 TATAATGCCAACTTTGTACAAAAAAGCAGGCTCCGCGGCCGCCTTGTTTAACTTTAAGAA
       70        80        90       100       110       120

       700       710       720       730       740       750
23621_ GGAGCCCTTCACCATGTCTCAAGCAATCCAACAGCATATAAATCTCAACATGATCAGAGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 GGAGCCCTTCACCATGTCTCAAGCAATCCAACAGCATATAAATCTCAACATGATCAGAGA
      130       140       150       160       170       180

       760       770       780       790       800       810
23621_ TGAGCATTTACTTGCTGTGGCAGAGAAACTCATTCCAACAGTCTACAGTCGCGAAGTACA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 TGAGCATTTACTTGCTGTGGCAGAGAAACTCATTCCAACAGTCTACAGTCGCGAAGTACA
      190       200       210       220       230       240

       820       830       840       850       860       870
23621_ ACCGCGTTCAATTGTTAAAAGTATTGAGGACGGAAATGTAGCGCTTGGCTGGCGTGCCGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 ACCGCGTTCAATTGTTAAAAGTATTGAGGACGGAAATGTAGCGCTTGGCTGGCGTGCCGA
      250       260       270       280       290       300

       880       890       900       910       920       930
23621_ GCACATTGCAGATGCAACGACTTTTCTACAGCAGTCATTTTCACCGGGTGACACCTTTAT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 GCACATTGCAGATGCAACGACTTTTCTACAGCAGTCATTTTCACCGGGTGACACCTTTAT
      310       320       330       340       350       360

       940       950       960       970       980       990
23621_ CATTGATTTTGGTGAGCATTGTGTCGGTTCACTCAATTTTTCATGTGATTCCACCGGGAG
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CATTGATTTTGGTGAGCATTGTGTCGGTTCACTCAATTTTTCATGTGATTCCACCGGGAG
      370       380       390       400       410       420

      1000      1010      1020      1030      1040      1050
23621_ CCCTCCTGATGCGCCAGCACATCTGCAGTTCATCTTTGGCGAAACGCTGTGTGAAGTTTC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CCCTCCTGATGCGCCAGCACATCTGCAGTTCATCTTTGGCGAAACGCTGTGTGAAGTTTC
      430       440       450       460       470       480
```

```
    1060     1070     1080     1090     1100     1110
23621_ CGAGCCTTTCAGCGAGTACCAGGGCTGGTTAAGTAGCAGTTGGCTGCAACAACATGATGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CGAGCCTTTCAGCGAGTACCAGGGCTGGTTAAGTAGCAGTTGGCTGCAACAACATGATGA
   490      500      510      520      530      540

    1120     1130     1140     1150     1160     1170
23621_ ATACATCGATATCTTACCTGAAGAGAAAACGTTGCCCAGACGTTACTGTTTCCGCTATAT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 ATACATCGATATCTTACCTGAAGAGAAAACGTTGCCCAGACGTTACTGTTTCCGCTATAT
   550      560      570      580      590      600

    1180     1190     1200     1210     1220     1230
23621_ CAAAGTGCGAGTCGTATCACTTTCACCGAAATATACCATTCGATTTACTCGCTTGAGCGC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CAAAGTGCGAGTCGTATCACTTTCACCGAAATATACCATTCGATTTACTCGCTTGAGCGC
   610      620      630      640      650      660

    1240     1250     1260     1270     1280     1290
23621_ CACCACAGTGAGCTCAGCTCCGCACAGTTATCCTGAGTACCATACGCACGATGCACTGTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CACCACAGTGAGCTCAGCTCCGCACAGTTATCCTGAGTACCATACGCACGATGCACTGTT
   670      680      690      700      710      720

    1300     1310     1320     1330     1340     1350
23621_ AAAAAAGATAGATGAGGTTTCGGTTCAAACCTTGCGCAATTGTATGCAGAACGTTTTTGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 AAAAAAGATAGATGAGGTTTCGGTTCAAACCTTGCGCAATTGTATGCAGAACGTTTTTGA
   730      740      750      760      770      780

    1360     1370     1380     1390     1400     1410
23621_ AGATGGACCTAAGCGTGACCGTCGTCTTTGGCTAGGCGATCTCCGCCTTCAGGCGTTGGT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 AGATGGACCTAAGCGTGACCGTCGTCTTTGGCTAGGCGATCTCCGCCTTCAGGCGTTGGT
   790      800      810      820      830      840

    1420     1430     1440     1450     1460     1470
23621_ CAATGATGTCACCTTCGCGAAACATGATCTTGTAAAACGCTGTCTTTATCTCTTTGCAGG
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CAATGATGTCACCTTCGCGAAACATGATCTTGTAAAACGCTGTCTTTATCTCTTTGCAGG
   850      860      870      880      890      900

    1480     1490     1500     1510     1520     1530

23621_ CCATAGAAGAGAAGACGGAATGGTTGCTGCTAATATTTTTGCCCGTCCTGAAGTAAAAGC
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 CCATAGAAGAGAAGACGGAATGGTTGCTGCTAATATTTTTGCCCGTCCTGAAGTAAAAGC
   910      920      930      940      950      960

    1540     1550     1560     1570     1580     1590
23621_ AGACGATACCTATCTGTTCGATTATTCATTATTTTTTGTCGACACTTTATACAACTATTT
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
233621 AGACGATACCTATCTGTTCGATTATTCATTATTTTTTGTCGACACTTTATACAACTATTT
   970      980      990     1000     1010     1020

    1600     1610     1620     1630     1640     1650
23621_ AATGGCGACGCAGGATATAGCAACAGCAAAAGAGTTGTGGCCTGTAGCTATTCGCCAGGT
```

```
      ::::::::::::::  ::::::::::::: ::::::::::::::: ::::  ::::::
233621 AATGGCGACGCAGGANNTAGCAACAGCAAANGAGTTGTGGCCTGTAGNTATTNNCCAGGT
   1030     1040     1050     1060     1070     1080

      1660     1670     1680     1690     1700     1710
23621_ TGAATTGGCACTGG-AGCGATGTGATCACACCGG-TCTGG-TGAGAGACAGTGACGA
      ::::::::: :::   :::::::::: :   :: ::::: ::  :  ::::::: :
233621 TGAATTGGCCNTGGGNNCGATGTGATCCCCNNGGNTCTGGNTGNNAANCAGTGACNA
   1090     1100     1110     1120     1130     1140
```

Gene_ ID_1463 (putative CE1 enzyme

> Gene_ID_1463 nucleotide sequence

ATGACTCGCGACGAAACCGGTGTCTGGTCGTGGACCAGCACACAACTTCAGCCAAACCTT
TACGAATACTATTTCGATGTGGATGGTTTTCGCAGTATCGACACGGGTAGCCGTTTCACTA
AACCGCAGCGCCAGGTCAATACCAGTCTGATTTTGGTTCCCGGCGGTATTTTAGATGAGCG
CGCTGTACCGCACGGTGAACTGCGAACGTTGACCTATCACTCAAGTGCTTTGCAGGCTGAG
CGGCAGGTTTATGTCTGGACTCCACCGGGCTACAAACCCGGTGGTGAGCCACTGCCAGTA
CTCTATTTCTATCATGGGTTTGGCGATACAGGCTTGTCGGGCATTACGCAGGGACGCATTCC
ACAAATCATGGATAACCTCCTGGCAGAAGGCAAAATTAAGCCAATGCTGGTAGTGGTTCC
GGATACTGAAACCGACATCAAACAAGCGGTGGCAGAAAACTTCCCGCCGGTTGAACGCC
GCAAAGATTTCTATCCTCTCAACGCTGCAGCGGCTGATAAAGAGTTAATGGAAGATATTAT
TCCGTTGATCAGCAACCGTTTCACGGTGCGTAAAGATGCGGCAGGCAGGGCGTTAGCGGG
TTTATCGCAAGGTGGTTATCAGGCGCTGGTTTCGGGTATGAGTCATCTGGATAGCTTTGGCT
GGCTGGCGACATTCAGCGGCGTGACCACGACTACCGTACCGAATGAACGCGTTTCTCGCC
AGCTTGCTCAACCGCAGGCGATTAATCAGCAATTGCGTAATTTTACCCTGGTTGTGGGTGA
AAAAGACGGTGTAACGGGCAAGGACATCGCTGGTCTGAAAAGTCAGTTAGAACAGCAGG
GGGTGAAGTTTAGCTACACTTCTTACCCGGGTCTTGGGCATGAAATGGATGTCTGGCGGCC
AGCCTATGCAGAGTTTGTTCAGCGGTAA

> **Gene_ID_1463 amino acid sequence**

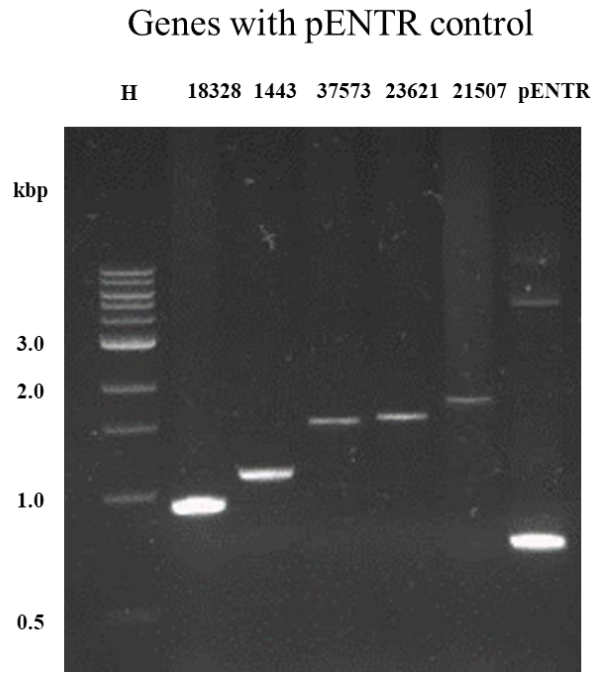
MTRDETGVWSWTSTQLQPNLYEYYFDVDGFRSIDTGSRFTKPQRQVNTSLILVPGGILDER
AVPHGELRTLTYHSSALQAERQVYVWTPPGYKPGGEPLPVLYFYHGFGDTGLSGITQGRIP
QIMDNLLAEGKIKPMLVVVPDTETDIKQAVAENFPPVERRKDFYPLNAAAADKELMEDII
PLISNRFTVRKDAAGRALAGLSQGGYQALVSGMSHLDSFGWLATFSGVTTTTVPNERVSR
QLAQPQAINQQLRNFTLVVGEKDGVTGKDIAGLKSQLEQQGVKFSYTSYPGLGHEMDV
WRPAYAEFVQR

Number of amino acids: 311

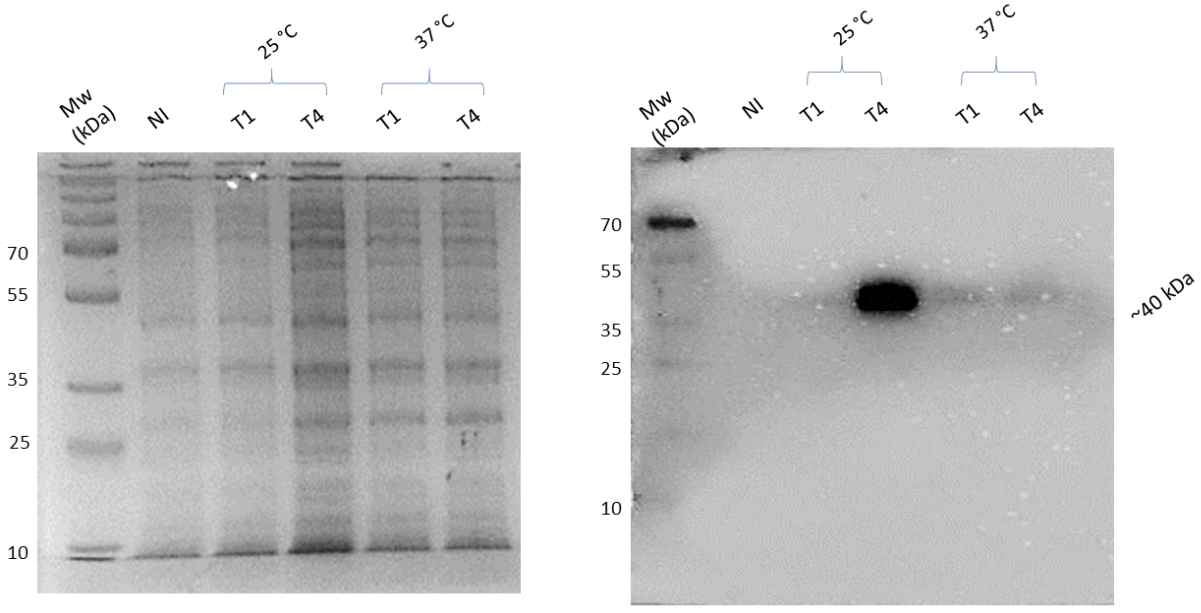
Molecular weight: 34548.04 Da

Theoretical pI: 5.80

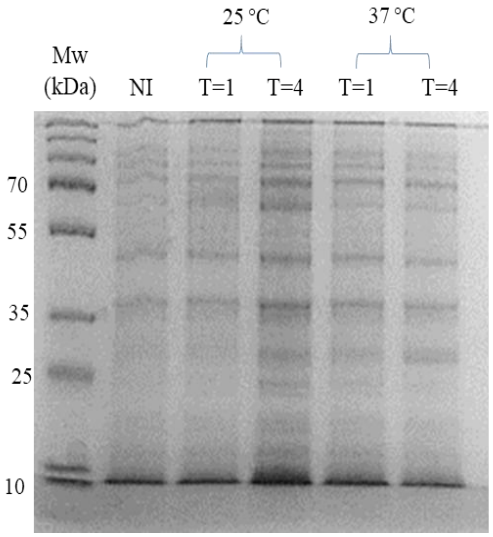
7.3 Amplified genes from the slug gut metagenomic DNA for future work

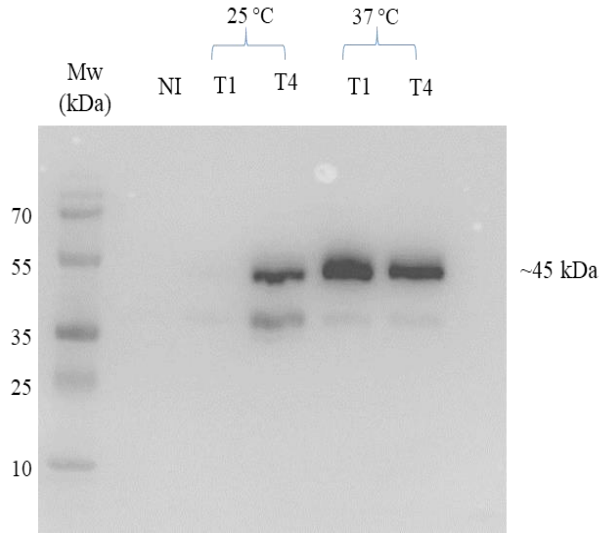


7.4 SDS-PAGE and western blot of expressed protein from Gene_Id_18328 (putative endo 1,4-beta xylanase)

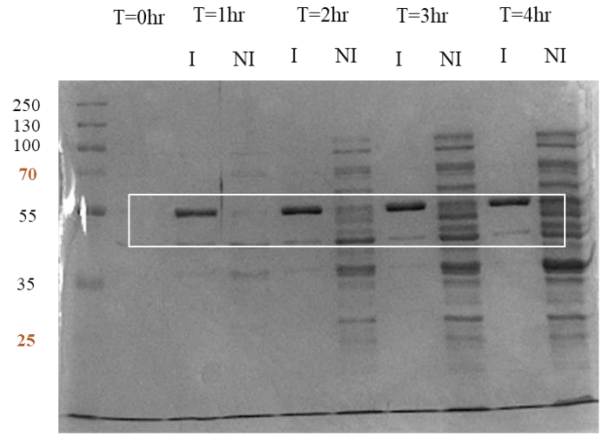


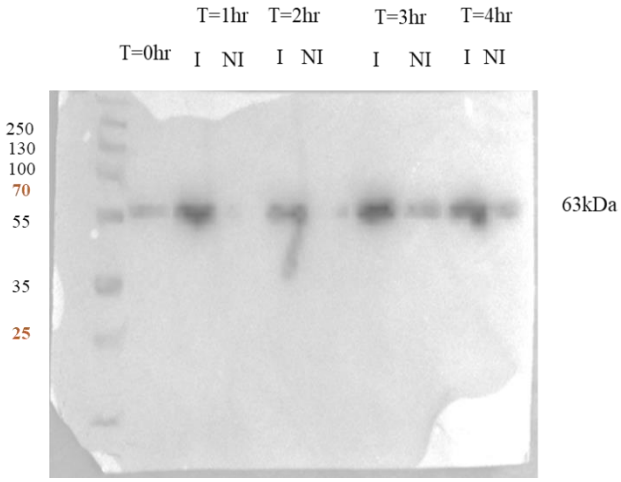
7.5 SDS-PAGE and western blot of expressed protein from Gene_Id_1443 (putative lignin peroxidase)





7.6 SDS-PAGE and western blot of expressed protein from Gene_Id_9459 (putative beta glucosidase)





7.7 Recipes for buffer preparations

7.7.1 Buffers used in a -L – rhamnosidase study

Table 7.1 cell lysis buffer

Reagent	**Needed**	
HEPES	20 mM	pH 7.4
NaCl	0.6M	
DTT	1 mM	
Glycerol	10% (optional)	
Triton-X-100	1%	
EDTA free protease inhibitor cocktail (Sigma-Merck, UK)	1 tablet in 10ml of lysis buffer	

To each 20g of cell pellet, 10ml of cell lysis buffer supplemented with 1 X EDTA free protease inhibitor cocktail (cOmplete™, EDTA-free Protease Inhibitor Cocktail, SKU 4693132001, Merck, UK)

Table 7.2 Equilibration/binding buffer

Reagent	**Needed**	
HEPES	20 mM	pH 7.4
NaCl	0.6M	
DTT	1 mM	
Imidazole	5 mM	

Table 7.3 Wash buffer

Reagent	**Needed**	
HEPES	20 mM	pH 7.4
NaCl	0.6M	
DTT	1 mM	
Imidazole	10 mM	

Table 7.4 Elution buffer

Reagent	**Needed**	

HEPES	20 mM	pH 7.4
NaCl	0.6M	
DTT	1 mM	
Imidazole	300mM	

7.7.1.1 Dialysis buffer

For every 1 ml of eluted protein fraction, 100 ml of dialysis buffer is required.

Table 7.5 Dialysis buffer

Reagent	**Needed**	
HEPES	20 mM	pH 6.5
NaCl	100 mM	
DTT	1 mM	
Glycerol	10%	

7.7.1.2 Enzyme activity reaction buffer

Table 7.6 Enzyme reaction buffer

Reagent	**Needed**	
HEPES	20 mM	pH 6.5
NaCl	50mM	

Table 7.7 Rhamnopyranoside substrate preparation

Reagent	**Needed**
A-L-rhamnopyranoside	50 mM in 50 % ethanol

7.7.2 Buffers used in esterase activity testing

Table 7.8 Lysis buffer

Reagent	**Needed**	
Potassium phosphate	50 mM	
NaCl	400 mM	
KCl	100 mM	
Glycerol	10%	pH 7.8
Triton-X-100	0.5%	
Imidazole	10 mM	
PMSF	1mM	
DNase I	2KU	
Lysozyme	50mg/ml	

Initially a 1 M stock solution of KH_2PO_4 and K_2HPO_4 was prepared. Into a volumetric flask, 90 ml of deionized water, 0.3 ml KH_2PO_4, 4.7 ml K_2HPO_4, 2.3 g NaCl, 0.75 g KCl, 10 ml glycerol, 0.5 ml Triton X-100 and 68 mg imidazole were added. The contents of the volumetric flask were thoroughly mixed by stirring and pH was adjusted to 7.8 with 4N HCl. The total volume was finally adjusted to 100 ml and the buffer was stored at +4°C.

Table 7.9 Esterase Purification buffer

Reagent	**Needed**	
Sodium phosphate	50 mM	
NaCl	300 mM	
Glycerol	10%	
Imidazole	10 - 300 mM	pH 7.5

Binding buffer = 10 mM imidazole, wash buffer 1 = 20 mM imidazole, wash buffer 2 = 50 mM imidazole, Elution buffer 1 = 80 mM imidazole, Elution buffer 2 = 150 mM imidazole and Elution buffer 3 = 300mM imidazole.

Table 7.10 dialysis buffer

Reagent	Needed	
Sodium phosphate	50 mM	pH 7.5
NaCl	10 mM	

Table 7.11 Esterase enzyme reaction buffer

Reagent	Needed	
Sodium phosphate	50 mM	pH 7.5

7.7.3 SDS-PAGE running buffer (10X)

Prepare the following.

Table 7.12 SDS-PAGE gel running buffer (10X) concentrated

Reagent	Amount (g)
Tris base	30.3
Glycine	144.4
SDS	10

After preparing, dissolve in 1 L of Milli-Q-filtered H_2O and store at 4°C.

7.7.4 SDS-PAGE Sample Loading Buffer (2X)

Prepare 10 ml of the following.

Table 7.13 SDS-PAGE gel sample loading buffer (2X) concentrated

Reagent	Needed	
Tris-HCl	62.5 mM	pH 6.8,
glycerol	25%	
SDS	2%	
Bromophenol Blue	0.01%	

This is a modified SDS-PAGE loading buffer (Laemmli,1970) with increase in the concentration of glycerol.

7.7.5 Western blot transfer buffer (10X)

To prepare 1L of a 10X western transfer buffer (without methanol), transfer to a bottle.

Table 7.14 Western blot transfer buffer (10X) concentrated

Reagent	Concentration
Tris base	250 mM
Glycine	192 M

To dilute to a working concentration of 1X, 100 ml of the 10 X was dissolved with 700 ml of distilled water and toped up to 1 L with 200 ml of ethanol (20% v/v).

7.7.6 Phosphate buffered saline (1 X PBS)

From the manufacturer we obtained 10 tables (Oxoid™ Phosphate Buffered Saline Tablets, ThermoFisher Scientific, UK) which were dissolved in 1L of deionised water and autoclaved at 115°C for 10 minutes to achieve 1X PBS.

7.7.7 Blocking solution for western blotting 5%

1g of semi-skimmed milk dry powder was dissolved in 20 ml of 1X PBS supplemented with 0.1% tween-20 (Sigma, UK)

7.7.8 Antibody solutions for western blotting

1g of semi-skimmed milk dry powder was dissolved in 20 ml of 1X PBS supplemented with 0.1% tween-20 (Sigma, UK) and a Monoclonal Anti-6X His tag antibody produced in mouse (Merck, UK) at 1:1000 dilution.

Secondary antibody solution: 1g of semi-skimmed milk dry powder was dissolved in 20 ml of 1X PBS supplemented with 0.1% tween-20 (Sigma, UK) and Goat Anti-Mouse IgG Antibody, HRP conjugate (Merck, UK) at 1:10000 dilution.

7.7.9 Western blotting protein detection

The detection of proteins after immobilisation onto PVDF membranes was achieved by preparing a working solution following instructions detailed in SuperSignal West Dura (ThermoFisher Scientific, UK)

Chapter 8. References

Abbott, D. W. and Boraston, A. B. (2008) 'Structural Biology of Pectin Degradation by Enterobacteriaceae', *Microbiology and Molecular Biology Reviews*. doi: 10.1128/MMBR.00038-07.

Aguiar, R. and Wink, M. (2005) 'How do slugs cope with toxic alkaloids?', *CHEMOECOLOGY* . Springer, 15(3), pp. 167–177. doi: 10.1007/S00049-005-0309-5.

Ahmad, R. and Sardar, M. (2015) 'Enzyme Immobilization: An Overview on Nanoparticles as Immobilization Matrix', *Biochemistry & Analytical Biochemistry*, 4(2), pp. 2–8. doi: 10.4172/2161-1009.1000178.

Akin, D. E. *et al.* (1995) 'Alterations in structure, chemistry, and biodegradability of grass lignocellulose treated with the white rot fungi Ceriporiopsis subvermispora and Cyathus stercoreus.', *Applied and environmental microbiology*. American Society for Microbiology, 61(4), pp. 1591–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7747973 (Accessed: 20 August 2018).

Alessi, A. M. *et al.* (2018) 'Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies', *Biotechnology for Biofuels*. BioMed Central, 11(1), p. 166. doi: 10.1186/s13068-018-1164-2.

Amado, N. G. *et al.* (2009) 'Isoquercitrin isolated from Hyptis fasciculata reduces glioblastoma cell proliferation and changes β-catenin cellular localization', *Anti-Cancer Drugs*. doi: 10.1097/CAD.0b013e32832d1149.

America's Energy Future Panel on Alternative Liquid Transportation Fuels *et al.* (2009) 'Liquid Transportation Fuels from Coal and Biomass', in *Energy*. Washington: National Academies Press, pp. 1–6.

Andrews, S. (2018) *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 10 May

2022).

Arsalan, A. and Younus, H. (2018) 'Enzymes and nanoparticles: Modulation of enzymatic activity via nanoparticles', *International Journal of Biological Macromolecules*. Elsevier, 118, pp. 1833–1847. doi: 10.1016/J.IJBIOMAC.2018.07.030.

Aspeborg, H. *et al.* (2012) 'Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5).', *BMC evolutionary biology*, 12, p. 186. doi: 10.1186/1471-2148-12-186.

Aust, S. D. (1995) 'Mechanisms of degradation by white rot fungi', in *Environmental Health Perspectives*. doi: 10.1289/ehp.95103s459.

Azzam, A. M. (1989) 'Pretreatment of cane bagasse with alkaline hydrogen peroxide for enzymatic hydrolysis of cellulose and ethanol fermentation', *Journal of Environmental Science and Health, Part B*, 24(4), pp. 421–433. doi: 10.1080/03601238909372658.

Babayan, A. *et al.* (2017) 'Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells', *Oncotarget*. Impact Journals, LLC, 8(34), p. 56066. doi: 10.18632/ONCOTARGET.10701.

Bailes, J. *et al.* (2012) 'Effect of Gold Nanoparticle Conjugation on the Activity and Stability of Functional Proteins', in *Nanoparticles in Biology and Medicine*. NJ: Humana Press, pp. 89–99. doi: 10.1007/978-1-61779-953-2_7.

Baudrexl, M. *et al.* (2019) 'Biochemical characterisation of four rhamnosidases from thermophilic bacteria of the genera Thermotoga, Caldicellulosiruptor and Thermoclostridium', *Scientific reports*. Sci Rep, 9(1). doi: 10.1038/S41598-019-52251-0.

Béguin, P. and Aubert, J.-P. (1994) 'The biological degradation of cellulose', *FEMS Microbiology Reviews*, 13, pp. 25–58. doi: 10.1111/j.1574-6976.1994.tb00033.x.

Bertino-Grimaldi, D. *et al.* (2013) 'Bacterial community composition shifts in the gut of Periplaneta americana fed on different lignocellulosic materials', *SpringerPlus*. doi:

10.1186/2193-1801-2-609.

Biddanda, B. A. (1988) *Microbial aggregation and degradation of phytoplankton-derived detritus in seawater. 11. Microbial metabolism*, *Marine Ecology - Progress series*. Available at: https://www.int-res.com/articles/meps/42/m042p089.pdf (Accessed: 27 July 2019).

Biely, P. and Mackenzie, C. R. (1986) 'Cooperativity of esterases and xylanases in the enzymatic degradation of acetyl xylan', *Bio/Technology*, 4(8), pp. 731–733. doi: 10.1038/NBT0886-731.

Biomass Biorefinery Netweork (2017) *The ten green chemicals which can create growth, jobs and trade for the UK*. Available at: https://www.bbnet-nibb.co.uk/wp-content/uploads/2019/08/UKBioChem10_Highligts.pdf (Accessed: 3 August 2021).

Blanchette, R. A. (1991) 'Delignification by Wood-Decay Fungi', *Annual Review of Phytopathology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 29(1), pp. 381–403. doi: 10.1146/annurev.py.29.090191.002121.

Blum, D. L. *et al.* (1999) 'Characterization of an acetyl xylan esterase from the anaerobic fungus Orpinomyces sp. strain PC-2', *Applied and Environmental Microbiology*. American Society for Microbiology, 65(9), pp. 3990–3995. doi: 10.1128/AEM.65.9.3990-3995.1999.

Boerjan, W., Ralph, J. and Baucher, M. (2003) 'Lignin Biosynthesis', *Annual Review of Plant Biology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA, 54(1), pp. 519–546. doi: 10.1146/annurev.arplant.54.031902.134938.

Boisvert, S. *et al.* (2012) 'Ray Meta: Scalable de novo metagenome assembly and profiling', *Genome Biology*. BioMed Central Ltd., 13(12), pp. 1–13. doi: 10.1186/GB-2012-13-12-R122/TABLES/2.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*. Oxford University Press, 30(15), p. 2114. doi: 10.1093/BIOINFORMATICS/BTU170.

Bonzom, C. *et al.* (2018) 'Feruloyl esterase immobilization in mesoporous silica particles and characterization in hydrolysis and transesterification', *BMC Biochemistry*. BioMed Central, 19(1). doi: 10.1186/S12858-018-0091-Y.

Borges, L. M. S., Merckelbach, L. M. and Cragg, S. M. (2014) 'Biogeography of wood-boring crustaceans (Isopoda: Limnoriidae) established in European coastal waters.', *PloS one*. Public Library of Science, 9(10), p. e109593. doi: 10.1371/journal.pone.0109593.

Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Analytical Biochemistry*. Academic Press, 72(1–2), pp. 248–254. doi: 10.1016/0003-2697(76)90527-3.

Brady, D. and Jordaan, J. (2009) 'Advances in enzyme immobilisation', *Biotechnology letters*. Biotechnol Lett, 31(11), pp. 1639–1650. doi: 10.1007/S10529-009-0076-4.

Brena, B., González-Pombo, P. and Batista-Viera, F. (2013) 'Immobilization of Enzymes: A Literature Survey', *Methods in molecular biology (Clifton, N.J.)*, 1051, pp. 15–31. doi: 10.1007/978-1-62703-550-7_2.

Breznak, J. A. and Brune, A. (1994) 'Role of Microorganisms in the Digestion of Lignocellulose by Termites', *Annual Review of Entomology*, 39(1), pp. 453–487. doi: 10.1146/annurev.en.39.010194.002321.

Brodelius, P. and Mosbach, K. (1987) '[14] Overview', *Methods in Enzymology*. Academic Press, 135, pp. 173–175. doi: 10.1016/0076-6879(87)35075-X.

Brulc, J. M. *et al.* (2009) 'Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(6), pp. 1948–53. doi: 10.1073/pnas.0806191105.

Brune, A. (2014) 'Symbiotic digestion of lignocellulose in termite guts', *Nature Reviews*

Microbiology. Nature Publishing Group, 12(3), pp. 168–180. doi: 10.1038/nrmicro3182.

Caldini, C. *et al.* (1994) 'Kinetic and immobilization studies on fungal glycosidases for aroma enhancement in wine', *Enzyme and Microbial Technology*. doi: 10.1016/0141-0229(94)90168-6.

Cardoso, A. M. *et al.* (2012) 'Metagenomic Analysis of the Microbiota from the Crop of an Invasive Snail Reveals a Rich Reservoir of Novel Genes', *PLoS ONE*. Edited by O. Terenius. Public Library of Science, 7(11), p. e48505. doi: 10.1371/journal.pone.0048505.

Celtic Renewables (2021) *Biochemical Production Process -*. Available at: https://www.celtic-renewables.com/process/ (Accessed: 19 October 2021).

Chapman, J., Ismail, A. and Dinu, C. (2018) 'Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks', *Catalysts*, 8(6), p. 238. doi: 10.3390/catal8060238.

Chaturvedi, V. and Verma, P. (2013) 'An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products.', *3 Biotech*. Springer, 3(5), pp. 415–431. doi: 10.1007/s13205-013-0167-8.

Chen, P. T. *et al.* (2020) 'Isolation, Expression and Characterization of the Thermophilic Recombinant Esterase from Geobacillus thermodenitrificans PS01', *Applied biochemistry and biotechnology*. Appl Biochem Biotechnol, 191(1), pp. 112–124. doi: 10.1007/S12010-020-03225-W.

Chen, S. *et al.* (2016) 'Therapeutic Effects of Quercetin on Inflammation, Obesity, and Type 2 Diabetes', *Mediators of Inflammation*. Hindawi Limited, 2016. doi: 10.1155/2016/9340637.

Cherry, J. R. and Fidantsef, A. L. (2003) 'Directed evolution of industrial enzymes: an update', *Current Opinion in Biotechnology*. Elsevier Current Trends, 14(4), pp. 438–443. doi: 10.1016/S0958-1669(03)00099-5.

Chum, H. L. *et al.* (1988) 'Organosolv pretreatment for enzymatic hydrolysis of poplars: I. Enzyme hydrolysis of cellulosic residues', *Biotechnology and Bioengineering*, 31(7), pp. 643–649. doi: 10.1002/bit.260310703.

Collins, J. and Hohn, B. (1978) 'Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 75(9), p. 4242. doi: 10.1073/PNAS.75.9.4242.

Colpa, D. I., Fraaije, M. W. and Van Bloois, E. (2014) 'DyP-type peroxidases: A promising and versatile class of enzymes', *Journal of Industrial Microbiology and Biotechnology*. doi: 10.1007/s10295-013-1371-6.

da Costa, R. R. *et al.* (2018) 'Enzyme Activities at Different Stages of Plant Biomass Decomposition in Three Species of Fungus-Growing Termites.', *Applied and environmental microbiology*. American Society for Microbiology, 84(5), p. AEM.01815-17. doi: 10.1128/AEM.01815-17.

Cui, Z. *et al.* (2007) 'Crystal Structure of Glycoside Hydrolase Family 78 α-L-Rhamnosidase from Bacillus sp. GL1', *Journal of Molecular Biology*. Academic Press, 374(2), pp. 384–398. doi: 10.1016/J.JMB.2007.09.003.

Das, S. and Vikalo, H. (2013) 'Base calling for high-throughput short-read sequencing: Dynamic programming solutions', *BMC Bioinformatics*. BioMed Central, 14(1), pp. 1–10. doi: 10.1186/1471-2105-14-129/TABLES/3.

Davis, R. H. and Perkins, D. D. (2002) 'Neurospora: A model of model microbes', *Nature Reviews Genetics*. doi: 10.1038/nrg797.

Deepika and Maurya, P. K. (2022) 'Health Benefits of Quercetin in Age-Related Diseases', *Molecules*. Multidisciplinary Digital Publishing Institute (MDPI), 27(8). doi: 10.3390/MOLECULES27082498.

Dekker, R. (1994) 'Enzymes in food and beverage processing (Part 1)', *Food Australia*,

46, p. 136. doi: 10.13140/2.1.4858.9760.

Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms', *Biotechnology Advances*. Elsevier, 27(3), pp. 297–306. doi: 10.1016/J.BIOTECHADV.2009.01.008.

DiCosimo, R. *et al.* (2013) 'Industrial use of immobilized enzymes', *Chemical Society Reviews*, 42(15), p. 6437. doi: 10.1039/c3cs35506c.

Dodd, D. and Cann, I. K. O. (2009) 'Enzymatic deconstruction of xylan for biofuel production.', *Global change biology. Bioenergy*. NIH Public Access, 1(1), pp. 2–17. doi: 10.1111/j.1757-1707.2009.01004.x.

Doi, R. H. (2008) 'Cellulases of Mesophilic Microorganisms', *Annals of the New York Academy of Sciences*, 1125(1), pp. 267–279. doi: 10.1196/annals.1419.002.

Duff, S. J. B. and Murray, W. D. (1996) 'Bioconversion of forest products industry waste cellulosics to fuel ethanol: A review', *Bioresource Technology*. Elsevier, 55(1), pp. 1–33. doi: 10.1016/0960-8524(95)00122-0.

DuPont, M. S. *et al.* (2004) 'Absorption of kaempferol from endive, a source of kaempferol-3-glucuronide, in humans', *European journal of clinical nutrition*. Eur J Clin Nutr, 58(6), pp. 947–954. doi: 10.1038/SJ.EJCN.1601916.

E4tech (2017) *UK Top Bio-based Chemicals Opportunities E4tech (UK) Ltd for LBNet*. Available at: www.e4tech.com (Accessed: 25 July 2019).

Ebringerová, A., Hromádková, Z. and Heinze, T. (2005) 'Hemicellulose', in *Polysaccharides I*. Berlin/Heidelberg: Springer-Verlag, pp. 1–67. doi: 10.1007/b136816.

El-Shanshoury, A. R. *et al.* (1994) 'The enumeration and characterization of bacteria and fungi associated with marine wood-boring isopods, and the ability of these microorganisms to digest cellulose and wood', *Marine Biology*. Springer-Verlag, 119(3), pp. 321–326. doi: 10.1007/BF00347528.

Eman Talukder, M. *et al.* (2017) 'Effect of various types of enzymatic treatment on Textile Materials and Optimize the process', *International Journal of Scientific & Engineering Research*, 8(2). Available at: http://www.ijser.org.

Esteghlalian, A. *et al.* (1997) 'Modeling and optimization of the dilute-sulfuric-acid pretreatment of corn stover, poplar and switchgrass', *Bioresource Technology*. Elsevier, 59(2–3), pp. 129–136. doi: 10.1016/S0960-8524(97)81606-9.

European Commission (2013) *Pulp and paper industry - European Commission*, *Pulp and paper industry*. Available at: https://ec.europa.eu/growth/sectors/raw-materials/industries/forest-based/pulp-paper_en (Accessed: 24 July 2018).

Evans, W. A. L. and Jones, E. G. (1962) 'A note on the proteinase activity in the alimentary tract of the slug Arion ater L.', *Comparative Biochemistry and Physiology*. Pergamon, 5(3), pp. 223–225. doi: 10.1016/0010-406X(62)90108-1.

Fan, L., Gharpuray, M. M. and Lee, Y.-H. (1987) 'Acid Hydrolysis of Cellulose', in *Biotechnology monographs*, pp. 121–148. doi: 10.1007/978-3-642-72575-3_4.

Fan, L. T., Lee, Y.-H. and Gharpuray, M. M. (1982) 'The nature of lignocellulosics and their pretreatments for enzymatic hydrolysis', *Advances in Biochemical Engineering*. Springer, Berlin, Heidelberg, 23, pp. 157–187. doi: 10.1007/3540116982_4.

Faulds, C. B. and Williamson, G. (1991) 'The purification and characterization of 4-hydroxy-3-methoxycinnamic (ferulic) acid esterase from Streptomyces olivochromogenes', *Journal of general microbiology*. J Gen Microbiol, 137(10), pp. 2339–2345. doi: 10.1099/00221287-137-10-2339.

Faulds, C. B. and Williamson, G. (1994) 'Purification and characterization of a ferulic acid esterase (FAE-III) from Aspergillus niger: Specificity for the phenolic moiety and binding to microcrystalline cellulose', *Microbiology*. Microbiology Society, 140(4), pp. 779–787. doi: 10.1099/00221287-140-4-779/CITE/REFWORKS.

Fernandes, S. and Murray, P. (2010) 'Metabolic engineering for improved microbial

pentose fermentation.', *Bioengineered bugs*. Taylor & Francis, 1(6), pp. 424–8. doi: 10.4161/bbug.1.6.12724.

Fernandez-Lafuente, R. (2009) 'Stabilization of multimeric enzymes: Strategies to prevent subunit dissociation', *Enzyme and Microbial Technology*, 45(6–7), pp. 405–418. doi: 10.1016/J.ENZMICTEC.2009.08.009.

Ferrer-Miralles, N. and Villaverde, A. (2013) 'Bacterial cell factories for recombinant protein production; expanding the catalogue', *Microbial Cell Factories*. BioMed Central, 12(1), p. 113. doi: 10.1186/1475-2859-12-113.

Flint, H. J. *et al.* (2012) 'Microbial degradation of complex carbohydrates in the gut.', *Gut microbes*. Taylor & Francis, 3(4), pp. 289–306. doi: 10.4161/gmic.19897.

Fontes, C. M. G. A. and Gilbert, H. J. (2010) 'Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates', *Annual Review of Biochemistry*, 79, p. 65681. doi: 10.1146/annurev-biochem-091208-085603.

Franco Cairo, J. P. L. *et al.* (2016) 'Expanding the Knowledge on Lignocellulolytic and Redox Enzymes of Worker and Soldier Castes from the Lower Termite Coptotermes gestroi', *Frontiers in Microbiology*, 7, p. 1518. doi: 10.3389/fmicb.2016.01518.

Franka Ganske and BMG LABTECH (2009) *Enzyme kinetic measurements on a Plate Reader | BMG LABTECH*. Available at: https://www.bmglabtech.com/enzyme-kinetic-measurements-performed-on-a-bmg-labtech-microplate-reader/ (Accessed: 20 December 2021).

Fujimoto, Z. *et al.* (2013) 'The structure of a Streptomyces avermitilis α-L-Rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement', *Journal of Biological Chemistry*. Elsevier, 288(17), pp. 12376–12385. doi: 10.1074/JBC.M113.460097/ATTACHMENT/45927239-ED93-4D42-80D3-2300C09D91FF/MMC1.PDF.

Gallardo, O., Diaz, P. and Pastor, F. I. J. (2003) 'Characterization of a Paenibacillus cell-associated xylanase with high activity on aryl-xylosides: a new subclass of family 10 xylanases', *Applied Microbiology and Biotechnology*, 61(3), pp. 226–233. doi: 10.1007/s00253-003-1239-1.

Ganesana, M. *et al.* (2011) 'Site-specific immobilization of a (His)6-tagged acetylcholinesterase on nickel nanoparticles for highly sensitive toxicity biosensors', *Biosensors and Bioelectronics*. Elsevier, 30(1), pp. 43–48. doi: 10.1016/J.BIOS.2011.08.024.

Gangwar, A. K., Tejo Prakash, N. and Prakash, R. (2016) 'An eco-friendly approach: Incorporating a xylanase stage at various places in ECF and chlorine-based bleaching of eucalyptus pulp', *BioResources*. North Carolina State University, 11(2), pp. 5381–5388. doi: 10.15376/BIORES.11.2.5381-5388.

Ge, L. *et al.* (2017) 'Enhancing the thermostability of α-L-rhamnosidase from Aspergillus terreus and the enzymatic conversion of rutin to isoquercitrin by adding sorbitol', *BMC Biotechnology*, 17(21), pp. 1–10. doi: 10.1186/s12896-017-0342-9.

Giavasis, I., Harvey, L. M. and McNeil, B. (2008) 'Gellan Gum', *Critical Reviews in Biotechnology*. Taylor & Francis, 20(3), pp. 177–211. doi: 10.1080/07388550008984169.

Gilbert, H. J. (2010) 'The biochemistry and structural biology of plant cell wall deconstruction.', *Plant physiology*, 153(2), pp. 444–455. doi: 10.1104/pp.110.156646.

Gilbert, J. and Hazlewood, P. (1993) *Review Article Bacterial cellulases and xylanases*, *Journal of General Microbiology*.

Goldemberg, J. (2007) 'Ethanol for a Sustainable Energy Future', *Science*, 315(5813), pp. 808–810. doi: 10.1126/science.1137013.

de Gonzalo, G. *et al.* (2016) 'Bacterial enzymes involved in lignin degradation', *Journal of Biotechnology*. doi: 10.1016/j.jbiotec.2016.08.011.

Goswami, K. and Sanan-Mishra, N. (2022) 'RNA-seq for revealing the function of the transcriptome', *Bioinformatics*. Academic Press, pp. 105–129. doi: 10.1016/B978-0-323-

89775-4.00002-X.

Govindarajan, R. *et al.* (2012) 'Microarray and its applications', *Journal of Pharmacy & Bioallied Sciences*. Wolters Kluwer -- Medknow Publications, 4(Suppl 2), p. S310. doi: 10.4103/0975-7406.100283.

Grand View Research (2022) *Flavonoids Market Size Worth $1.06 Billion By 2025*. Available at: https://www.grandviewresearch.com/press-release/global-flavonoids-market (Accessed: 13 May 2022).

Guerriero, G. *et al.* (2016) 'Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization', *Engineering in Life Sciences*. John Wiley & Sons, Ltd, 16(1), pp. 1–16. doi: 10.1002/elsc.201400196.

Guillotin, L. *et al.* (2019) 'Biochemical Characterization of the α- L -Rhamnosidase DtRha from Dictyoglomus thermophilum: Application to the Selective Derhamnosylation of Natural Flavonoids', *ACS Omega*. American Chemical Society, 4(1), pp. 1916–1922. doi: 10.1021/ACSOMEGA.8B03186/ASSET/IMAGES/LARGE/AO-2018-03186Y_0003.JPEG.

Gupta, M. N. *et al.* (2011) 'Nanomaterials as Matrices for Enzyme Immobilization', *Artificial Cells, Blood Substitutes, and Biotechnology*. Taylor & Francis, 39(2), pp. 98–109. doi: 10.3109/10731199.2010.516259.

Guzik, U., Hupert-Kocurek, K. and Wojcieszyńska, D. (2014) 'Immobilization as a Strategy for Improving Enzyme Properties-Application to Oxidoreductases', *Molecules*, 19(7), pp. 8995–9018. doi: 10.3390/molecules19078995.

Handelsman, J. (2004) 'Metagenomics: Application of Genomics to Uncultured Microorganisms', *Microbiology and Molecular Biology Reviews*, 68(4), pp. 669–685. doi: 10.1128/MMBR.68.4.669-685.2004.

Hansen, J. *et al.* (2010) 'Global surface temperature change', *Reviews of Geophysics*. Blackwell Publishing Ltd, 48(4). doi: 10.1029/2010RG000345.

Hazardous Substance Research Centers/South & Southwest Outreach Program (2003) *Environmental Impact of the Petroleum Industry Environmental Update #12*. Available at: http://www.epa.gov/ttn/chief/ap42/ch05/. (Accessed: 13 November 2021).

Hempel, F. *et al.* (2011) 'Algae as protein factories: Expression of a human antibody and the respective antigen in the diatom phaeodactylum tricornutum', *PLoS ONE*, 6(12), p. e28424. doi: 10.1371/journal.pone.0028424.

Hess, M. *et al.* (2011) 'Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen', *Science*. American Association for the Advancement of Science, 331(6016), pp. 463–467. doi: 10.1126/SCIENCE.1200387.

Ho, Y. T. *et al.* (2015) 'An instantaneous colorimetric protein assay based on spontaneous formation of a protein corona on gold nanoparticles', *Analyst*. Royal Society of Chemistry, 140(4), pp. 1026–1036. doi: 10.1039/c4an01819b.

Holck, J. *et al.* (2019) 'A carbohydrate-binding family 48 module enables feruloyl esterase action on polymeric arabinoxylan', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 294(46), pp. 17339–17353. doi: 10.1074/JBC.RA119.009523.

Hsiao, Y.-M. *et al.* (2008) 'Regulation of the pehA gene encoding the major polygalacturonase of Xanthomonas campestris by Clp and RpfF', *Microbiology*, 154(Pt3), pp. 705–713. doi: 10.1099/mic.0.2007/012930-0.

Hyatt, D. *et al.* (2010) 'Prodigal: Prokaryotic gene recognition and translation initiation site identification', *BMC Bioinformatics*. BioMed Central, 11(1), pp. 1–11. doi: 10.1186/1471-2105-11-119/TABLES/5.

International Energy Agency (2018) *The Future of Petrochemicals Towards more sustainable plastics and fertilisers Together Secure Sustainable*. Available at: www.iea.org/t&c/ (Accessed: 29 July 2021).

Izzo, V. *et al.* (2014) 'α-Rhamnosidase activity in the marine isolate Novosphingobium

sp. PP1Y and its use in the bioconversion of flavonoids', *Journal of Molecular Catalysis B: Enzymatic*, 105(2014), pp. 95–103. doi: 10.1016/j.molcatb.2014.04.002.

Jagmann, N. and Philipp, B. (2018) 'SpoT-mediated regulation and amino acid prototrophy are essential for pyocyanin production during parasitic growth of Pseudomonas aeruginosa in a co-culture model system with Aeromonas hydrophila', *Frontiers in Microbiology*. Frontiers Media S.A., 9(APR), p. 761. doi: 10.3389/FMICB.2018.00761/BIBTEX.

Janson, J.-C. (2011) *Protein purification : principles, high resolution methods, and applications*. John Wiley & Sons. Available at: https://www.wiley.com/en-us/Protein+Purification%3A+Principles%2C+High+Resolution+Methods%2C+and+Applications%2C+3rd+Edition-p-9780471746614 (Accessed: 26 June 2019).

Janusz, G. *et al.* (2017) 'Lignin degradation: Microorganisms, enzymes involved, genomes analysis and evolution', *FEMS Microbiology Reviews*. doi: 10.1093/femsre/fux049.

Jeon, J. H. *et al.* (2009) 'Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome', *Applied microbiology and biotechnology*. Appl Microbiol Biotechnol, 81(5), pp. 865–874. doi: 10.1007/S00253-008-1656-2.

Jones, D. T. and Woods, D. R. (1986) *Acetone-Butanol Fermentation Revisited*, *MICROBIOLOGICAL REVIEWS*.

Joynson, R. *et al.* (2017) 'Metagenomic Analysis of the Gut Microbiome of the Common Black Slug Arion ater in Search of Novel Lignocellulose Degrading Enzymes', *Frontiers in Microbiology*, 8, p. 2181. doi: 10.3389/fmicb.2017.02181.

Ju You, H., Jin Ahn, H. and Ji, G. E. (2010) 'Transformation of rutin to antiproliferative quercetin-3-glucoside by aspergillus niger', *Journal of Agricultural and Food Chemistry*. doi: 10.1021/jf102871g.

Jungschaffer, G., Schurz, J. and Taufratzhofer, E. (1983) 'Enzymatic Conversion Of

Lignocellulosic Materials To Sugars', *Journal of Wood Chemistry and Technology*. doi: 10.1080/02773818308085164.

Kambhampati, S. and Eggleton, P. (2000) 'Taxonomy and Phylogeny of Termites', in *Termites: Evolution, Sociality, Symbioses, Ecology*. Dordrecht: Springer Netherlands, pp. 1–23. doi: 10.1007/978-94-017-3223-9_1.

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 of the United States of America*. National Academy of Sciences, 110(25), pp. 10189–10194. doi: 10.1073/pnas.1301502110.

Khoshnevisan, K. *et al.* (2011) 'Immobilization of cellulase enzyme on superparamagnetic nanoparticles and determination of its activity and stability', *Chemical Engineering Journal*. Elsevier, 171(2), pp. 669–673. doi: 10.1016/J.CEJ.2011.04.039.

King, A. J. *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 of the United States of America*. National Academy of Sciences, 107(12), pp. 5345–5350. doi: 10.1073/pnas.0914228107.

Kmezik, C. *et al.* (2021) 'A polysaccharide utilization locus from the gut bacterium dysgonomonas mossii encodes functionally distinct carbohydrate esterases', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology Inc., 296. doi: 10.1016/J.JBC.2021.100500.

Krajewska, B. (2004) 'Application of chitin- and chitosan-based materials for enzyme immobilizations: a review', *Enzyme and Microbial Technology*. Elsevier, 35(2–3), pp. 126–139. doi: 10.1016/J.ENZMICTEC.2003.12.013.

Krpetić, Ž. *et al.* (2012) 'Importance of Nanoparticle Size in Colorimetric and SERS-Based Multimodal Trace Detection of Ni(II) Ions with Functional Gold Nanoparticles',

Small. John Wiley & Sons, Ltd, 8(5), pp. 707–714. doi: 10.1002/smll.201101980.

Krpetić, Ž. *et al.* (2013) 'High-resolution sizing of monolayer-protected gold clusters by differential centrifugal sedimentation', *ACS Nano*, 7(10), pp. 8881–8890. doi: 10.1021/nn403350v.

Kuhad, R. C. *et al.* (2011) 'Microbial Cellulases and Their Industrial Applications Ramesh', *Enzyme Research*, 2011, pp. 1–10. doi: 10.4061/2011/280696.

Kujawska, A. *et al.* (2015) 'ABE fermentation products recovery methods—A review', *Renewable and Sustainable Energy Reviews*. Pergamon, 48, pp. 648–661. doi: 10.1016/J.RSER.2015.04.028.

Kumla, J. *et al.* (2020) 'Cultivation of mushrooms and their lignocellulolytic enzyme production through the utilization of agro-industrial waste', *Molecules*. MDPI AG, 25(12). doi: 10.3390/MOLECULES25122811.

Kwon, K. S. *et al.* (1994) 'Detection of β-glucosidase activity in polyacrylamide gels with esculin as substrate', *Applied and Environmental Microbiology*. American Society for Microbiology, 60(12), pp. 4584–4586. doi: 10.1128/AEM.60.12.4584-4586.1994.

Laemmli, U. K. (1970) 'Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4', *Nature*. Nature Publishing Group, 227(5259), pp. 680–685. doi: 10.1038/227680a0.

Lam, K. N. *et al.* (2015) 'Current and future resources for functional metagenomics', *Frontiers in Microbiology*. Frontiers Media S.A., 6(OCT), p. 1196. doi: 10.3389/FMICB.2015.01196/BIBTEX.

Langmead, B. *et al.* (2019) 'Scaling read aligners to hundreds of threads on general-purpose processors', *Bioinformatics*. Oxford University Press, 35(3), pp. 421–432. doi: 10.1093/BIOINFORMATICS/BTY648.

Levasseur, A. *et al.* (2013) 'Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes', *Biotechnology for Biofuels*. doi: 10.1186/1754-

6834-6-41.

Levin, A. (2019) *Clean Energy Opportunities And Dirty Energy Challenges*. Available at: https://www.census.gov/newsroom/press-kits/2018/pop-estimates-national-state.html. (Accessed: 20 September 2021).

Li, D. *et al.* (2007) 'Immobilization of glucose oxidase onto gold nanoparticles with enhanced thermostability', *Biochemical and Biophysical Research Communications*. Academic Press, 355(2), pp. 488–493. doi: 10.1016/J.BBRC.2007.01.183.

Li, D. *et al.* (2015) 'MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph', *Bioinformatics (Oxford, England)*. Bioinformatics, 31(10), pp. 1674–1676. doi: 10.1093/BIOINFORMATICS/BTV033.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2011) 'A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data', *Bioinformatics*, 27(21), pp. 2987–2993. doi: 10.1093/bioinformatics/btr509.

Li, H. (2012) 'Exploring single-sample snp and indel calling with whole-genome de novo assembly', *Bioinformatics*, 28(14), pp. 1838–1844. doi: 10.1093/bioinformatics/bts280.

Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler transform', *Bioinformatics*, 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi: 10.1093/bioinformatics/btp698.

Li, L. L. *et al.* (2009) 'Bioprospecting metagenomes: Glycosyl hydrolases for converting biomass', *Biotechnology for Biofuels*. BioMed Central, 2(1), pp. 1–11. doi: 10.1186/1754-

6834-2-10/TABLES/3.

Lockridge, O. and Quinn, D. M. (2010) 'Esterases', *Comprehensive Toxicology, Second Edition*. Elsevier Inc., 4, pp. 243–273. doi: 10.1016/B978-0-08-046884-6.00414-0.

Loix, C. *et al.* (2017) 'Reciprocal Interactions between Cadmium-Induced Cell Wall Responses and Oxidative Stress in Plants', *Frontiers in Plant Science*. Frontiers, 8, p. 1867. doi: 10.3389/fpls.2017.01867.

Loman, N. J. *et al.* (2013) 'A Culture-Independent Sequence-Based Metagenomics Approach to the Investigation of an Outbreak of Shiga-Toxigenic Escherichia coli O104:H4', *Journal of the American Medical Association*. American Medical Association, 309(14), pp. 1502–1510. doi: 10.1001/JAMA.2013.3231.

Lombard, V. *et al.* (2014) *The carbohydrate-active enzymes database (CAZy) in 2013*, *Nucleic Acids Research*. doi: 10.1093/nar/gkt1178.

Lozinsky, V. I. (2018) 'Cryostructuring of Polymeric Systems. 50. Cryogels and Cryotropic Gel-Formation: Terms and Definitions', *Gels*. Multidisciplinary Digital Publishing Institute (MDPI), 4(3). doi: 10.3390/GELS4030077.

Lu, J. *et al.* (2017) 'Bracken: Estimating species abundance in metagenomics data', *PeerJ Computer Science*. PeerJ Inc., 2017(1). doi: 10.7717/PEERJ-CS.104.

Maccaferri, S., Biagi, E. and Brigidi, P. (2011) 'Metagenomics: Key to Human Gut Microbiota', *Digestive Diseases*, 29(6), pp. 525–530. doi: 10.1159/000332966.

Maki, M., Leung, K. T. and Qin, W. (2009) 'The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass.', *International journal of biological sciences*. Ivyspring International Publisher, 5(5), pp. 500–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19680472 (Accessed: 13 August 2018).

Mansfield, J. *et al.* (2012) 'Top 10 plant pathogenic bacteria in molecular plant pathology', *Molecular Plant Pathology*. John Wiley & Sons, Ltd, 13(6), pp. 614–629. doi: 10.1111/J.1364-3703.2012.00804.X.

Manzanares, P. *et al.* (2000) 'Purification and characterization of an α-L-rhamnosidase from Aspergillus nidulans', *Letters in Applied Microbiology*, 31, pp. 198–202. doi: 10.1046/j.1365-2672.2000.00788.x.

Manzanares, P. *et al.* (2003) 'Construction of a Genetically Modified Wine Yeast Strain Expressing the Aspergillus aculeatus rhaA Gene, Encoding an α-L-Rhamnosidase of Enological Interest', *Applied and Environmental Microbiology*. American Society for Microbiology, 69(12), pp. 7558–7562. doi: 10.1128/AEM.69.12.7558-7562.2003/ASSET/01406EFA-1A10-42B7-87BA-53FB72A1777C/ASSETS/GRAPHIC/AM1230643002.JPEG.

Manzanares, P. *et al.* (2007) 'α-L-rhamnosidases: Old and new insights', in *Industrial Enzymes: Structure, Function and Applications*. doi: 10.1007/1-4020-5377-0_8.

Manzaneres, P. *et al.* (2001) 'Purification and Characterization of Two Different α-L-Rhamnosidases, RhaA and RhaB, from Aspergillus aculeatus', *Applied and Environmental Microbiology*, 67(5), pp. 2230–2234. doi: 10.1128/AEM.67.5.2230-2234.2001.

Margolin, A. L. and Navia, M. A. (2001) 'Protein Crystals as Novel Catalytic Materials - Margolin - 2001 - Angewandte Chemie International Edition - Wiley Online Library', *Angewandte Internation Edition Chemie*, 40(12), pp. 2204–2222. Available at: https://onlinelibrary.wiley.com/doi/10.1002/1521-3773(20010618)40:12%3C2204::AID-ANIE2204%3E3.0.CO;2-J (Accessed: 8 August 2021).

Margolles-Clark, E. *et al.* (1996) 'Acetyl xylan esterase from Trichoderma reesei contains an active-site serine residue and a cellulose-binding domain', *European journal of biochemistry*. Eur J Biochem, 237(3), pp. 553–560. doi: 10.1111/J.1432-1033.1996.0553P.X.

McMillan, J. D. (1992) *Processes for Pretreating Lignocellulosic Biomass: A Review*. Springfield. Available at: https://www.nrel.gov/docs/legosti/old/4978.pdf (Accessed: 20 August 2018).

McNamara, J. T., Morgan, J. L. W. and Zimmer, J. (2015) 'A Molecular Description of Cellulose Biosynthesis', *Annual Review of Biochemistry*, 84(1), pp. 895–921. doi: 10.1146/annurev-biochem-060614-033930.

Menzies, R. J. (1957) 'The Marine Borer Family Limnoriidae (Crustacea, Isopoda). Part I: Northern and Central America: Systematics, Distribution, and Ecology', *Bulletin of Marine Science*. University of Miami - Rosenstiel School of Marine and Atmospheric Science, 7(2), pp. 101–200. Available at: https://www.ingentaconnect.com/content/umrsmas/bullmar/1957/00000007/00000002/art00001 (Accessed: 20 August 2018).

Mewis, K. *et al.* (2016) 'Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization.', *Applied and environmental microbiology*, 82(6), pp. 1686–1692. doi: 10.1128/AEM.03453-15.

Michon, F. *et al.* (1987) *Structure of the Complex Group-Specific Polysaccharide of Group B Streptococcust*. Available at: https://pubs.acs.org/sharingguidelines.

Middleton, E. (1998) 'Effect of plant flavonoids on immune and inflammatory cell function', *Advances in experimental medicine and biology*. Adv Exp Med Biol, 439, pp. 175–182. doi: 10.1007/978-1-4615-5335-9_13.

Mikheenko, A., Saveliev, V. and Gurevich, A. (2016) 'MetaQUAST: Evaluation of metagenome assemblies', *Bioinformatics*. Oxford University Press, 32(7), pp. 1088–1090. doi: 10.1093/BIOINFORMATICS/BTV697.

Millett, M. A., Baker, A. J. and Satter, L. D. (1976) 'Pretreatments to enhance chemical, enzymatic, and microbiological attack of cellulosic materials.', *Biotechnology and bioengineering symposium*, (6), pp. 125–153.

Mishra, A. and Malhotra, A. V. (2009) 'Tamarind xyloglucan: a polysaccharide with versatile application potential', *Journal of Materials Chemistry*. The Royal Society of Chemistry, 19(45), pp. 8528–8536. doi: 10.1039/B911150F.

Mohnen, D. (2008) *Pectin structure and biosynthesis*, *Current Opinion in Plant Biology*. Elsevier Current Trends. Available at: https://www.sciencedirect.com/science/article/pii/S1369526608000630?via%3Dihub#! (Accessed: 14 May 2019).

Moraïs, S. *et al.* (2010) 'Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate.', *mBio*. American Society for Microbiology, 1(5), pp. e00285-10. doi: 10.1128/mBio.00285-10.

Mordor Intelligence (2022) *Flavonoid Market: Industry Share, Size, Growth* . Available at: https://www.mordorintelligence.com/industry-reports/flavonoid-market (Accessed: 13 May 2022).

Mouafo Tamnou, E. B. *et al.* (2021) 'Biodegradation of polyethylene by the bacterium Pseudomonas aeruginosa in acidic aquatic microcosm and effect of the environmental temperature', *Environmental Challenges*. Elsevier, 3, p. 100056. doi: 10.1016/J.ENVC.2021.100056.

Mueller, M. *et al.* (2018) 'Rhamnosidase activity of selected probiotics and their ability to hydrolyse flavonoid rhamnoglucosides', *Bioprocess and Biosystems Engineering*. Springer, 41(2), pp. 221–228. doi: 10.1007/s00449-017-1860-5.

Mullis, K. B. (1990) 'The Unusual Origin of the Polymerase Chain Reaction', 262(4), pp. 56–65. doi: 10.2307/24996713.

Mutter, M. *et al.* (1994) 'Rhamnogalacturonan [alpha]-L-Rhamnopyranohydrolase (A Novel Enzyme Specific for the Terminal Nonreducing Rhamnosyl Unit in Rhamnogalacturonan Regions of Pectin)', *Plant Physiology*. Oxford Academic, 106(1), pp. 241–250. doi: 10.1104/PP.106.1.241.

Neelakanta, G. and Sultana, H. (2013) 'The Use of Metagenomic Approaches to Analyze changes in Microbial communities', *Microbiology Insights*, 6, pp. 37–48. doi: 10.4137/MBI.S10819.

Nelson, J. and Griffin, E. G. (1916) 'ADSORPTION OF INVERTASE', *Z. Physiol. Chem*, 72(2), p. 722. doi: 10.1021/ja02262a018.

Nemati, M. and Webb, C. (2011) 'Immobilized Cell Bioreactors', *Comprehensive Biotechnology*. Pergamon, 2, pp. 489–504. doi: 10.1016/B978-0-444-64046-8.00084-7.

Nevalainen, K. M. H., Te'o, V. S. J. and Bergquist, P. L. (2005) 'Heterologous protein expression in filamentous fungi.', *Trends in biotechnology*, 23(9), pp. 468–74. doi: 10.1016/j.tibtech.2005.06.002.

Ngara, T. R. and Zhang, H. (2018) 'Recent Advances in Function-based Metagenomic Screening', *Genomics, Proteomics & Bioinformatics*. Elsevier, 16(6), p. 405. doi: 10.1016/J.GPB.2018.01.002.

Nimni, M. E. *et al.* (1987) 'Chemically modified collagen: A natural biomaterial for tissue replacement', *Journal of Biomedical Materials Research*. John Wiley & Sons, Ltd, 21(6), pp. 741–771. doi: 10.1002/JBM.820210606.

Nurk, S. *et al.* (2013) 'Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer, Berlin, Heidelberg, 7821 LNBI, pp. 158–170. doi: 10.1007/978-3-642-37195-0_13.

Nurk, S. *et al.* (2017) 'MetaSPAdes: A new versatile metagenomic assembler', *Genome Research*. Cold Spring Harbor Laboratory Press, 27(5), pp. 824–834. doi: 10.1101/GR.213959.116/-/DC1.

O'Neill, E. C. *et al.* (2015) 'Crystal structure of a novel two domain GH78 family α-rhamnosidase from Klebsiella oxytoca with rhamnose bound', *Proteins*. Wiley-Blackwell, 83(9), p. 1742. doi: 10.1002/PROT.24807.

de Oliveira, V. F. *et al.* (2018) 'Short-chain esters enriched biofuel obtained from vegetable oil using molecular distillation', *The Canadian Journal of Chemical Engineering*. John Wiley & Sons, Ltd, 96(5), pp. 1071–1078. doi: 10.1002/CJCE.23044.

Ondov, B. D., Bergman, N. H. and Phillippy, A. M. (2011) 'Interactive metagenomic visualization in a Web browser', *BMC bioinformatics*. BMC Bioinformatics, 12. doi: 10.1186/1471-2105-12-385.

ONS (2017) *Chemicals Sector Report*, *UK Parliment*. Available at: https://www.ons.gov.uk/economy/grossdomesticproductgdp/datasets/ukgdpolowlevelaggregates.

Orejas, M., Ibáñez, E. and Ramón, D. (1999) 'The filamentous fungus Aspergillus nidulans produces an α-L-rhamnosidase of potential oenological interest', *Letters in Applied Microbiology*. John Wiley & Sons, Ltd, 28(5), pp. 383–388. doi: 10.1046/J.1365-2672.1999.00539.X.

PA, K., CB, F. and Williamson, G. (1996) 'Purification and characterization of a novel esterase induced by growth of Aspergillus niger on sugar-beet pulp', *Biotechnology and Applied Biochemistry*. John Wiley & Sons, Ltd, 23(3), pp. 255–262. doi: 10.1111/J.1470-8744.1996.TB00381.X.

Pace, N. R. *et al.* (1986) 'The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences', in. doi: 10.1007/978-1-4757-0611-6_1.

Panche, A. N., Diwan, A. D. and Chandra, S. R. (2016) 'Flavonoids: an overview', *Journal of Nutritional Science*. Cambridge University Press, 5, pp. 1–15. doi: 10.1017/JNS.2016.41.

Panda, T. and Gowrishankar, B. S. (2005) 'Production and applications of esterases', *Applied microbiology and biotechnology*. Appl Microbiol Biotechnol, 67(2), pp. 160–169. doi: 10.1007/S00253-004-1840-Y.

Pandey, B., Demchenko, A. V. and Stine, K. J. (2012) 'Nanoporous gold as a solid support for protein immobilization and development of an electrochemical immunoassay for prostate specific antigen and carcinoembryonic antigen', *Microchimica Acta*, 179(1–2), pp. 71–81. doi: 10.1007/s00604-012-0870-x.

Patra, J. K., Das, G. and Shin, H.-S. (2018) *Microbial biotechnology. Volume 2, Application in food and pharmacology*. Available at: https://books.google.co.uk/books?id=7TtMDwAAQBAJ&pg=PA30&lpg=PA30&dq=Voget+et+al.+2003;+Yun+et+al.+2004&source=bl&ots=f0IUNNEio5&sig=ACfU3U3mrE1TIxno0LSvMhRIChx_nuS1_A&hl=en&sa=X&ved=2ahUKEwij6JPa0dXjAhXSTcAKHfxWD9cQ6AEwCXoECAkQAQ#v=onepage&q=soil&f=false (Accessed: 27 July 2019).

Peng, Y. *et al.* (2012) 'IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth', *Bioinformatics*. Oxford Academic, 28(11), pp. 1420–1428. doi: 10.1093/BIOINFORMATICS/BTS174.

Peng, Y. *et al.* (2016) 'Accurately Determining Esterase Activity via the Isosbestic Point of p-Nitrophenol', *BioResources*. North Carolina State University, 11(4), pp. 10099–10111. doi: 10.15376/BIORES.11.4.10099-10111.

Petkova, G. A. *et al.* (2012) 'Gold and silver nanoparticles for biomolecule immobilization and enzymatic catalysis.', *Nanoscale research letters*. Springer, 7(1), p. 287. doi: 10.1186/1556-276X-7-287.

Pope, P. B. *et al.* (2010) 'Adaptation to herbivory by the Tammar wallaby includes bacterial and glycoside hydrolase profiles different from other herbivores.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(33), pp. 14793–8. doi: 10.1073/pnas.1005297107.

Poutanen, K. *et al.* (1990) 'Deacetylation of xylans by acetyl esterases of Trichoderma reesei', *Applied Microbiology and Biotechnology 1990 33:5*. Springer, 33(5), pp. 506–510. doi: 10.1007/BF00172542.

Purewal, S. S. and Sandhu, K. S. (2021) 'Debittering of citrus juice by different processing methods: A novel approach for food industry and agro-industrial sector', *Scientia Horticulturae*. Elsevier, 276, p. 109750. doi: 10.1016/J.SCIENTA.2020.109750.

Puri, M. *et al.* (2008) 'Biochemical Basis of Bitterness in Citrus Fruit Juices and Biotech

Approaches for Debittering', *Critical Reviews in Biotechnology*. Taylor & Francis, 16(2), pp. 145–155. doi: 10.3109/07388559609147419.

Puri, M. and Banerjee, U. C. (2000) 'Production, purification, and characterization of the debittering enzyme naringinase.', *Biotechnology advances*, 18(3), pp. 207–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14538108 (Accessed: 5 August 2019).

Puri, M., Kaur, H. and Kennedy, J. F. (2005) 'Covalent immobilization of naringinase for the transformation of a flavonoid', *Journal of Chemical Technology & Biotechnology*. John Wiley & Sons, Ltd, 80(10), pp. 1160–1165. doi: 10.1002/JCTB.1303.

Qin, J. *et al.* (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*. doi: 10.1038/nature08821.

Ragauskas, A. J. *et al.* (2014) 'Lignin Valorization: Improving Lignin Processing in the Biorefinery', *Science*, 344(6185), pp. 1246843–1246843. doi: 10.1126/science.1246843.

Raveendran, S. *et al.* (2018) 'Applications of microbial enzymes in food industry', *Food Technology and Biotechnology*, 56(1), pp. 16–30. doi: 10.17113/ftb.56.01.18.5491.

Razeq, F. M. *et al.* (2018) 'A novel acetyl xylan esterase enabling complete deacetylation of substituted xylans', *Biotechnology for Biofuels*. BioMed Central Ltd., 11(1), pp. 1–12. doi: 10.1186/S13068-018-1074-3/FIGURES/8.

Rees, H. C. *et al.* (2003) 'Detecting cellulase and esterase enzyme activities encoded by novel genes present in environmental DNA libraries', *Extremophiles*. doi: 10.1007/s00792-003-0339-2.

Ridley, B. L., O'Neill, M. A. and Mohnen, D. (2001) 'Pectins: structure, biosynthesis, and oligogalacturonide-related signaling', *Phytochemistry*. Pergamon, 57(6), pp. 929–967. doi: 10.1016/S0031-9422(01)00113-3.

Rinaldi, R. *et al.* (2016) 'Paving the Way for Lignin Valorisation: Recent Advances in Bioengineering, Biorefining and Catalysis', *Angewandte Chemie International Edition*.

Wiley-Blackwell, 55(29), pp. 8164–8215. doi: 10.1002/anie.201510351.

Rodrigues, R. C. *et al.* (2013) 'Modifying enzyme activity and selectivity by immobilization.', *Chemical Society reviews*, 42(15), pp. 6290–307. doi: 10.1039/c2cs35231a.

Rodriguez, M. E. *et al.* (2010) 'Characterization of α-rhamnosidase activity from a Patagonian Pichia guilliermondii wine strain', *Journal of Applied Microbiology*. John Wiley & Sons, Ltd, 109(6), pp. 2206–2213. doi: 10.1111/J.1365-2672.2010.04854.X.

Roma-Rodrigues, C. *et al.* (2020) 'Gene therapy in cancer treatment: Why go nano?', *Pharmaceutics*. MDPI AG, 12(3). doi: 10.3390/PHARMACEUTICS12030233.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia coli: advances and challenges.', *Frontiers in microbiology*. Frontiers Media SA, 5(172), pp. 1–17. doi: 10.3389/fmicb.2014.00172.

Roy, A., Yang, J. and Zhang, Y. (2012) 'COFACTOR: an accurate comparative algorithm for structure-based protein function annotation', *Nucleic Acids Research*, 40, pp. 471–477. doi: 10.1093/nar/gks372.

Ruprecht, C. *et al.* (2019) 'High level production of flavonoid rhamnosides by metagenome-derived Glycosyltransferase C in Escherichia coli utilizing dextrins of starch as a single carbon source', *Metabolic engineering*. Metab Eng, 55, pp. 212–219. doi: 10.1016/J.YMBEN.2019.07.002.

Saha, B. C. (2003) 'Hemicellulose bioconversion', *Journal of industrial microbiology & biotechnology*. J Ind Microbiol Biotechnol, 30(5), pp. 279–291. doi: 10.1007/S10295-003-0049-X.

Saida, F. *et al.* (2006) 'Expression of highly toxic genes in E. coli: special strategies and genetic tools', *Current protein & peptide science*. Curr Protein Pept Sci, 7(1), pp. 47–56. doi: 10.2174/138920306775474095.

Salzberg, S. L. *et al.* (2012) 'GAGE: A critical evaluation of genome assemblies and

assembly algorithms', *Genome research*. Genome Res, 22(3), pp. 557–567. doi: 10.1101/GR.131383.111.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *DNA sequencing with chain-terminating inhibitors (DNA polymerase/nucleotide sequences/bacteriophage 4X174)*.

Sauer, M. (2016) 'Industrial production of acetone and butanol by fermentation-100 years later.', *FEMS microbiology letters*. Oxford University Press, 363(13). doi: 10.1093/femsle/fnw134.

Schmidt, T. M., DeLong, E. F. and Pace, N. R. (1991) 'Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing', *Journal of Bacteriology*, 173(14), pp. 4371–4378. doi: 10.1128/jb.173.14.4371-4378.1991.

Schmieg, B. *et al.* (2019) 'Advantages of hydrogel-based 3D-printed enzyme reactors and their limitations for biocatalysis', *Frontiers in Bioengineering and Biotechnology*. Frontiers Media S.A., 6(211), pp. 1–12. doi: 10.3389/FBIOE.2018.00211/BIBTEX.

Schmitt, J., Hess, H. and Stunnenberg, H. G. (1993) 'Affinity purification of histidine-tagged proteins.', *Molecular biology reports*, 18(3), pp. 223–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8114690 (Accessed: 1 August 2019).

Schoch, C. L. *et al.* (2020) 'NCBI Taxonomy: A comprehensive update on curation, resources and tools', *Database*. Oxford University Press. doi: 10.1093/DATABASE/BAAA062.

Schoenfeld, T. *et al.* (2008) 'Assembly of viral metagenomes from yellowstone hot springs', *Applied and environmental microbiology*. Appl Environ Microbiol, 74(13), pp. 4164–4174. doi: 10.1128/AEM.02598-07.

Seemann, T. (2014) 'Prokka: Rapid prokaryotic genome annotation', *Bioinformatics*. Oxford University Press, 30(14), pp. 2068–2069. doi: 10.1093/bioinformatics/btu153.

Sheldon, R. A. (2010) 'Cross-Linked Enzyme Aggregates as Industrial Biocatalysts', *Organic Process Research and Development*, 15(1), pp. 213–223. doi: 10.1021/op100289f.

Sheldon, R. A. (2011) 'Characteristic features and biotechnological applications of cross-linked enzyme aggregates (CLEAs)', *Applied microbiology and biotechnology*. Appl Microbiol Biotechnol, 92(3), pp. 467–477. doi: 10.1007/S00253-011-3554-2.

Sheldon, R. A. and van Pelt, S. (2013) 'Enzyme immobilisation in biocatalysis: Why, what and how', *Chemical Society Reviews*, 42(15), pp. 6223–6225. doi: 10.1039/c3cs60075k.

Sigoillot, J. C. *et al.* (2012) 'Fungal Strategies for Lignin Degradation', *Advances in Botanical Research*. doi: 10.1016/B978-0-12-416023-1.00008-2.

Singh Arora, D. and Kumar Sharma, R. (2010) 'Ligninolytic fungal laccases and their biotechnological applications', *Applied Biochemistry and Biotechnology*, 160(6), pp. 1760–1788. doi: 10.1007/s12010-009-8676-y.

Singh, S. M. and Panda, A. K. (2005) 'Solubilization and refolding of bacterial inclusion body proteins', *Journal of bioscience and bioengineering*. J Biosci Bioeng, 99(4), pp. 303–310. doi: 10.1263/JBB.99.303.

Sivaloganathan, D. M. and Brynildsen, M. P. (2021) 'Pseudomonas aeruginosa prioritizes detoxification of hydrogen peroxide over nitric oxide', *BMC Research Notes*. BioMed Central Ltd, 14(1), pp. 1–6. doi: 10.1186/S13104-021-05534-7/FIGURES/1.

Slaytor, M. (1992) 'Cellulose digestion in termites and cockroaches: What role do symbionts play?', *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*. Pergamon, 103(4), pp. 775–784. doi: 10.1016/0305-0491(92)90194-V.

Solomon, E. I., Sundaram, U. M. and Machonkin, T. E. (1996) 'Multicopper Oxidases and Oxygenases', *Chemical Reviews*. doi: 10.1021/cr950046o.

Stackebrandt, E. and Hippe, H. (1986) 'Transfer of Bacteroides amylophilus to a new genus Ruminobacter gen. nov., nom. rev. as Ruminobacter amylophilus comb. nov.', *Systematic and Applied Microbiology*. Urban & Fischer, 8(3), pp. 204–207. doi: 10.1016/S0723-2020(86)80078-9.

Su, X. *et al.* (2012) 'Heterologous Gene Expression in Filamentous Fungi', *Advances in Applied Microbiology*, 81, pp. 1–61. doi: 10.1016/B978-0-12-394382-8.00001-0.

Sukul, P., Lupilov, N. and Leichert, L. I. (2018) 'Characterization of ML-005, a novel metaproteomics-derived esterase', *Frontiers in Microbiology*. Frontiers Media S.A., 9(AUG). doi: 10.3389/FMICB.2018.01925.

Sun, J. *et al.* (2012) 'Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa', *Eukaryotic Cell*. doi: 10.1128/EC.05327-11.

Sun, Y. and Cheng, J. (2002) 'Hydrolysis of lignocellulosic materials for ethanol production: a review', *Bioresource Technology*. Elsevier, 83(1), pp. 1–11. doi: 10.1016/S0960-8524(01)00212-7.

Sundberg, M. and Poutanen, K. (1991) 'Purification and properties of two acetylxylan esterases of Trichoderma reesei — VTT's Research Information Portal', *JournalBiotechnology and Applied Biochemistry*, 13(1), pp. 1–11. Available at: https://cris.vtt.fi/en/publications/purification-and-properties-of-two-acetylxylan-esterases-of-trich (Accessed: 19 May 2022).

Suurnäkki, A. *et al.* (1997) 'Hemicellulases in the bleaching of chemical pulps', *Advances in biochemical engineering/biotechnology*. Springer, Berlin, Heidelberg, 57, pp. 261–287. doi: 10.1007/BFB0102077.

Terpe, K. (2006) 'Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems', *Applied Microbiology and Biotechnology*. Springer-Verlag, 72(2), pp. 211–222. doi: 10.1007/s00253-006-0465-8.

The Intergovernmental Panel on Climate Change (2021) *The Physical Science Basis Summary for Policymakers Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

The Lignocellulosic Biorefinery Network and Biotechnology and Biological Science Research Council Network in Industrial Biotechnology and Bioenergy (2019) *The ten green chemicals which can create growth, jobs and trade for the UK*.

The Royal Society and National Academy of Sciences (2020) *Climate Change: Evidence & Causes 2020*. Available at: https://www.nap.edu/catalog/18373] (Accessed: 18 September 2021).

ThermoFisher Scientific (2021) *Overview of Protein Expression Systems*. Available at: https://www.thermofisher.com/uk/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-protein-expression-systems.html (Accessed: 7 December 2021).

ThermoFisher Scientific (2022) *Gateway Recombination Cloning Technology*. Available at: https://www.thermofisher.com/uk/en/home/life-science/cloning/gateway-cloning/gateway-technology.html (Accessed: 21 May 2022).

Thomas, S. M., DiCosimo, R. and Nagarajan, V. (2002) 'Biocatalysis: Applications and potentials for the chemical industry', *Trends in Biotechnology*. doi: 10.1016/S0167-7799(02)01935-2.

Thomas, T., Gilbert, J. and Meyer, F. (2012) 'Metagenomics - a guide from sampling to data analysis.', *Microbial informatics and experimentation*. BioMed Central, 2(1), p. 3. doi: 10.1186/2042-5783-2-3.

Thrash, A., Hoffmann, F. and Perkins, A. (2020) 'Toward a more holistic method of genome assembly assessment', *BMC Bioinformatics*. BioMed Central, 21(Suppl 4). doi: 10.1186/S12859-020-3382-4.

Tokuda, G. and Watanabe, H. (2007) 'Hidden cellulases in termites: revision of an old hypothesis', *Biology Letters*. The Royal Society, 3(3), p. 336. doi: 10.1098/RSBL.2007.0073.

Topakas, E., Vafiadi, C. and Christakopoulos, P. (2007) 'Microbial production,

characterization and applications of feruloyl esterases', *Process Biochemistry*, 42(4), pp. 497–509. doi: 10.1016/J.PROCBIO.2007.01.007.

Turati, D. F. M. *et al.* (2019) 'Thermotolerant lipase from Penicillium sp. section Gracilenta CBMAI 1583: Effect of carbon sources on enzyme production, biochemical properties of crude and purified enzyme and substrate specificity', *Biocatalysis and Agricultural Biotechnology*. Elsevier, 17, pp. 15–24. doi: 10.1016/J.BCAB.2018.10.002.

Turkevich, J., Stevenson, P. C. and Hillier, J. (1951) 'A study of the nucleation and growth processes in the synthesis of colloidal gold', *Discussions of the Faraday Society*. doi: 10.1039/DF9511100055.

United Nations (2021) *What Is Climate Change?*, *Climate Action*. Available at: https://www.un.org/en/climatechange/what-is-climate-change# (Accessed: 18 September 2021).

Väisänen, T. *et al.* (2016) 'Utilization of agricultural and forest industry waste and residues in natural fiber-polymer composites: A review', *Waste Management*. doi: 10.1016/j.wasman.2016.04.037.

Vakulchuk, R., Overland, I. and Scholten, D. (2020) 'Renewable energy and geopolitics: A review', *Renewable and Sustainable Energy Reviews*. Pergamon, 122(109547), pp. 1–12. doi: 10.1016/J.RSER.2019.109547.

Valentová, K. *et al.* (2014) 'Isoquercitrin: Pharmacology, toxicology, and metabolism', *Food and Chemical Toxicology*. doi: 10.1016/j.fct.2014.03.018.

Walle, T. (2004) 'Absorption and metabolism of flavonoids', *Free Radical Biology and Medicine*. doi: 10.1016/j.freeradbiomed.2004.01.002.

Wang, W. *et al.* (2019) 'Injectable, Self-Healing Hydrogel with Tunable Optical, Mechanical, and Antimicrobial Properties'. doi: 10.1021/acs.chemmater.8b04803.

Wang, X. *et al.* (2022) 'Recent advances and application of whole genome amplification in molecular diagnosis and medicine', *MedComm*. Wiley-Blackwell, 3(1). doi:

10.1002/MCO2.116.

Warnecke, F. *et al.* (2007) 'Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite', *Nature*. Nature Publishing Group, 450(7169), pp. 560–565. doi: 10.1038/nature06269.

De Winter, K. *et al.* (2013) 'Chemoenzymatic synthesis of α-l-rhamnosides using recombinant α-l-rhamnosidase from Aspergillus terreus', *Bioresource Technology*. doi: 10.1016/j.biortech.2013.08.083.

Woese, C. R. and Fox, G. E. (1977) *Phylogenetic structure of the prokaryotic domain: The primary kingdoms (archaebacteria/eubacteria/urkaryote/16S ribosomal RNA/molecular phylogeny)*.

Wood, D. E., Lu, J. and Langmead, B. (2019) 'Improved metagenomic analysis with Kraken 2', *Genome Biology*. BioMed Central Ltd., 20(1). doi: 10.1186/S13059-019-1891-0.

Yadav, A. K., Chaudhari, A. B. and Kothari, R. M. (2011) 'Bioconversion of renewable resources into lactic acid: an industrial view', *Critical Reviews in Biotechnology*. Taylor & Francis, 31(1), pp. 1–19. doi: 10.3109/07388550903420970.

Yadav, S., Yadava, S. and Yadav, K. D. S. (2017) 'α-l-rhamnosidase selective for rutin to isoquercitrin transformation from Penicillium griseoroseum MTCC-9224', *Bioorganic Chemistry*. Academic Press, 70, pp. 222–228. doi: 10.1016/J.BIOORG.2017.01.002.

Yang, J., Roy, A. and Zhang, Y. (2013a) 'BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions', *Nucleic Acids Research*, 41, pp. 1096–1103. doi: 10.1093/nar/gks966.

Yang, J., Roy, A. and Zhang, Y. (2013b) 'BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions', *Nucleic Acids Research*. Oxford University Press, 41(Database issue), p. D1096. doi: 10.1093/NAR/GKS966.

Yang, J., Roy, A. and Zhang, Y. (2013c) 'Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment', *Bioinformatics*. Oxford Academic, 29(20), pp. 2588–2595. doi: 10.1093/BIOINFORMATICS/BTT447.

Yang, P. and Mahmood, T. (2012) 'Western blot: Technique, theory, and trouble shooting', *North American Journal of Medical Sciences*, 4(9), p. 429. doi: 10.4103/1947-2714.100998.

Young, N. M., Johnston, R. A. Z. and Richards, J. C. (1989) 'Purification of the α-l-rhamnosidase of Penicillium decumbens and characterisation of two glycopeptide components', *Carbohydrate Research*. Elsevier, 191(1), pp. 53–62. doi: 10.1016/0008-6215(89)85045-1.

Yun, J. and Ryu, S. (2005) 'Screening for novel enzymes from metagenome and SIGEX, as a way to improve it', *Microbial Cell Factories*. BioMed Central, 4(1), pp. 1–5. doi: 10.1186/1475-2859-4-8/FIGURES/1.

Zachary, A. and Colwell, R. R. (1979) 'Gut-associated microflora of Limnoria tripunctata in marine creosote-treated wood pilings', *Nature*. Nature Publishing Group, 282(5740), pp. 716–717. doi: 10.1038/282716a0.

Zerbino, D. R. and Birney, E. (2008) 'Velvet: algorithms for de novo short read assembly using de Bruijn graphs', *Genome research*. Genome Res, 18(5), pp. 821–829. doi: 10.1101/GR.074492.107.

Zhanfang, M. and Teng, D. (2009) 'Bioconjugates of Glucose Oxidase and Gold Nanorods Based on Electrostatic Interaction with Enhanced Thermostability', *Nanoscale Research Letters*, 4, pp. 1236–1240. doi: 10.1007/s11671-009-9385-8.

Zhang, C., Freddolino, Peter L and Zhang, Y. (2017) 'COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information', *Nucleic Acids Research*, 45(2), pp. 291–299. doi: 10.1093/nar/gkx366.

Zhang, C., Freddolino, Peter L. and Zhang, Y. (2017) 'COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information'. Nucleic Acids Res, 45. Available at: https://academic.oup.com/nar/article-abstract/45/W1/W291/3787871 (Accessed: 20 April 2022).

Zhang, L. *et al.* (2021) 'Advances in Metagenomics and Its Application in Environmental Microorganisms', *Frontiers in Microbiology*. Frontiers Media S.A., 12, p. 3847. doi: 10.3389/FMICB.2021.766364/BIBTEX.

Zhang, S. *et al.* (2014) 'Improved thermostability of esterase from Aspergillus fumigatus by site-directed mutagenesis', *Enzyme and microbial technology*. Enzyme Microb Technol, 64–65, pp. 11–16. doi: 10.1016/J.ENZMICTEC.2014.06.003.

Zhang, Y. and Skolnick, J. (2004) 'SPICKER: A Clustering Approach to Identify Near-Native Protein Folds', *J Comput Chem*, 25, pp. 865–871.

Zhou, M. *et al.* (2014) 'Bioelectrochemistry of Microbial Fuel Cells and their Potential Applications in Bioenergy', *Bioenergy Research: Advances and Applications*. Elsevier, pp. 131–152. doi: 10.1016/B978-0-444-59561-4.00009-7.

Zhu, W., Lomsadze, A. and Borodovsky, M. (2010) 'Ab initio gene identification in metagenomic sequences', *Nucleic Acids Research*. Oxford Academic, 38(12), pp. e132–e132. doi: 10.1093/NAR/GKQ275.

Zhu, Y. (2007) 'Immobilized Cell Fermentation for Production of Chemicals and Fuels', in *Bioprocessing for Value-Added Products from Renewable Resources*. Elsevier, pp. 373–396. doi: 10.1016/B978-044452114-9/50015-3.