Fibre Degrading Enzymes from *Butyrivibrio proteoclasticus*

A thesis submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy

in

Biological Sciences

at

The University of Waikato

by

Marisa Till

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2011

Abstract

The rumen harbours a large and diverse microbial population that is responsible for the breakdown of plant material into smaller compounds, which can then be utilised by the animal.

Butyrivibrio proteoclasticus is an anaerobic, Gram-positive bacterium originally isolated from the rumen of New Zealand cows. The entire genome of *B. proteoclasticus* has been sequenced. This revealed that a large proportion of the genome is devoted to polysaccharide degradation and reassembly[1]. Prior to the start of the research described in this thesis, 44 of the genes from the *B. proteoclasticus* genome annotated as being involved in fibre degradation had been cloned and expression of many has been tested. Two of these enzymes were expressed, purified and had had their 3D structures determined. Further characterisation of these two enzymes is presented here, with site directed mutagenesis used to probe the proposed mechanism for each. The results support the proposed catalytic mechanisms of both enzymes. Kinetic parameters were measured with model substrates for both enzymes and this allowed comparison with similar enzymes produced by other organisms.

A further nine fibre degrading enzymes from *B. proteoclasticus* were chosen for structural and/or functional investigation.

Preliminary structural investigation gave crystallisation conditions for one enzyme, Xsa43E. The structure of Xsa43E is presented along with functional analysis of the enzyme. Four residues were found to be important for catalysis, three previously identified catalytic acidic residues common to all GH43 enzymes and a fourth residue, a histidine that is important for pK_a modulation of the catalytic acid. The large central ion present in the structure of Xsa43E is identified as calcium and shown to be important both for activity and structural stability.

Acknowledgements

I would like to thank my primary supervisor Associate Professor Vic Arcus. Thank you for the amazing amount of support and guidance you have given me. Also thank you for sharing your knowledge and expertise with me as well as your fantastic approach to science. I would also like to thank my other supervisors Dr Graeme Atttwood, Dr Shaun Lott and Dr Lyndsay Main for their expertise and their willingness to share it with me; your help throughout this research has been greatly appreciated.

I would like to thank all the members of the C.2.10 Proteins and Microbes Lab, past and present. You have all been a significant influence and help to me over the course of my research. Our little chats and your support through the rollercoaster that is a PhD have helped keep me sane and motivated.

I would like to thank the Tertiary Education Commission for the Bright Futures Scholarship I received as well as the Waihi Agricultural Education Trust and Agresearch for their help in funding this research.

I would like to thank Tony for the unlimited support and encouragement he has given me. Thank you Mum and Dad for your unwavering belief in me. Also to the rest of my family and friends who have put up with me, supported me and provided me with welcome distractions, thank you all.

Table of Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

AXE - Acetyl xylan esterase

Azo-CMC - Azo-CMCellulose

Azo-WAX - Azo-wheat arabinoxylan

CAZy - Carbohydrate Active EnZymes

CBD - Carbohydrate Binding Domains

CBM - Carbohydrate binding modules

CEs - Carbohydrate esterases

CP - Cell pellet

dNTP - Deoxyribonucleotide triphosphate

EDTA - Ethylenediaminetetraacetic Acid

FDE - Fibre Degrading Enzymes

FLPC - Fast Performance Liquid Chromatography

FOM - Figure of merit

GAX - Glucoroarabinoxylans

GCMS - Gas Chromatography Mass Spectrometry

GHs - Glycoside hydrolases

GTs - Glycosyltrasferases

ICP-MS - Inductively Coupled Plasma Mass Spectrometry

IMAC – Immobilised Metal Affinity Chromatography

MAD - Multiple wavelength Anomalous Dispersion

MALDI-TOF MS – Matrix Assisted Laser Desorption Ionisation Time of Flight Mass Spectrometry

MIR – Multiple Isomorphous Replacement

M_r – Molecular mass

ms-TMS - *N*-methyl-*N*(trimethylsilyl)-trifluoroacetic acid

NMR - Nuclear Magnetic Resonance

ORF - Open Reading Frame

PAGE - Poly Acrylamide Gel Electrophoresis

PCR- Polymerase Chain Reaction

PD - Nickel Pull Down

PDB - Protein Data Bank

PLs - Polysaccharide Lyases

p-NPA - *p*-nitrophenyl acetate

p-NPB - *p*-nitrophenyl butyrate

p-NP-D-X - *p*-nitrophenyl-ß-D-xylopyranose

p-NP-L-A - *p*-nitrophenyl-α-L-arabinofuranose

RMSD – Root Mean Square Deviation

rRNA - Ribosomal Ribonucleic Acid

SAD - Single wavelength Anomalous Dispersion

SDS - Sodium Dodecyl Sulfate

SEC - Size Exclusion Chromatography

SIR – Single Isomorphous Replacement

SN - Supernatant

SSRL - Stanford Synchrotron Radiation Laboratory

TLC - Thin Layer Chromatography

tris - tris(hydroxymethyl)aminomethane

WC - Whole Cell

Chapter One: Introduction

1.1 Plant cell walls

A plant cell wall is a physical barrier between the cell contents and its external environment. It protects and supports the cell, giving it shape[2, 3]. Between adjacent cells is an area called the middle lamella; this contains a sticky substance rich in pectin, a structural heteropolysaccharide, which essentially glues the cells together[2]. The cell wall is a complex matrix containing structurally independent but interacting polysaccharide networks, it contains up to 90% total polysaccharide in the form of cellulose, hemicelluose, pectin and lignin forming the main structural framework[4-6], as well as aromatic substances and proteins[3, 4]. The polysaccharide composition of a plant cell wall varies between species and within species[4]; certain cell types require specific cell wall components for specialisation[5]. In general the cell wall is made up of cellulose microfibrils and hemicelluloses that are amorphous polymers of different sugars. The hemicellulose polymers hydrogen bond to the surface of the cellulose microfibrils and crosslink them[6]. These structures are embedded in a gel like matrix of pectic substances and proteins[3, 7], and the hemicelluloses crosslink covalently to an outer layer of lignin via ester linkages[8]. The crosslinks add structural integrity to the plant cell wall and are also believed to limit cell growth and impede biodegradation[8].

Cellulose is a linear glucan made up of β-1,4 linked glucose units, also called elementary fibrils (Figure 1-1). It is estimated thirty six of these elementary fibrils associate together in a regular crystalline environment to form microfiobrils[9-11]. The microfibrils are attached to each other by hemicelluloses and other polymers such as pectin[11].

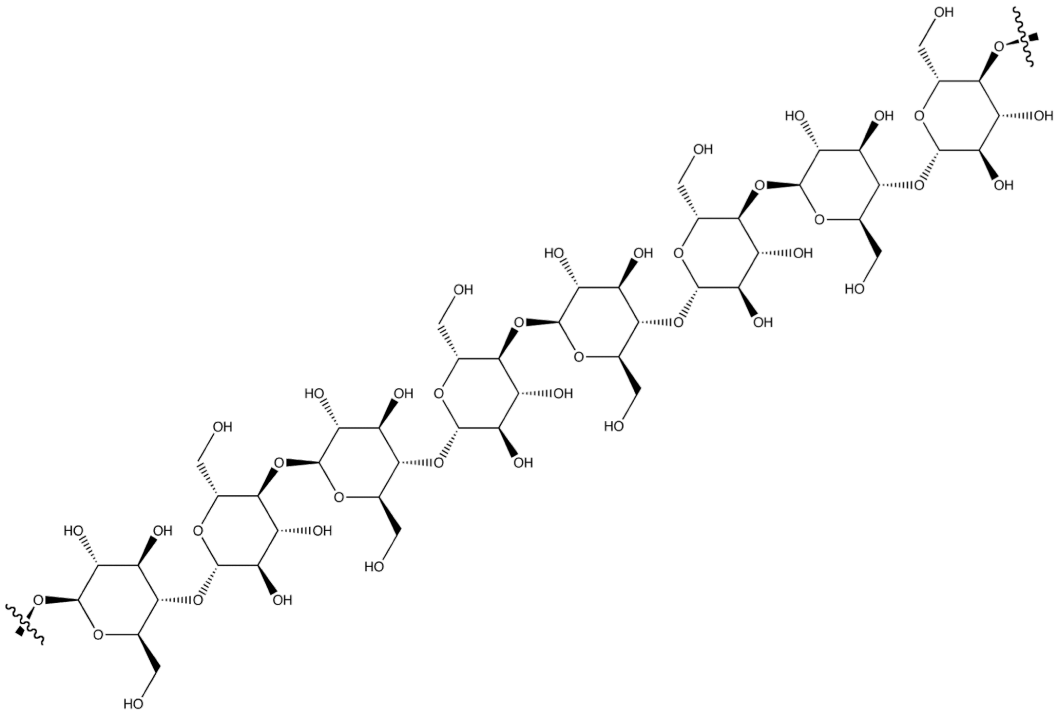


Figure 1-1: Structure of a Cellulose elementary fibril. Adapted from Onda *etal* (2008)[12].

Hemicelluloses are heteroglycans[13], meaning they contain a variety of sugar units and have more structural diversity than cellulose. The term hemicellulose was first proposed by Schulze in 1891 to describe the fractions isolated or extracted from plant materials with dilute alkali[14]. Hemicellulose covers a range of polysaccharides, the main ones being xylans, mannans, galactan and arabinan. A typical xylan structure is shown in Figure 1-2.

The classification of the hemicellulose depends on the major sugar moiety present[14]. Xylan is the most abundant of the hemicelluloses[15], it is a polymer of ß-1,4-linked xylose residues with a variety of side groups including L-arabinofuranose, D-galactose, acetyl, feruloyl and *p*-coumaric moities[8]. Where there is branching of the xylan backbone this prevents the formation of hydrogen bonds with the cellulose microfibers[8]. Xylans can be further classified depending on the other sugars incorporated into the structure, the most common of these being homoxylan,

arabinoxylan and glucuronoarabinoxylan[15].

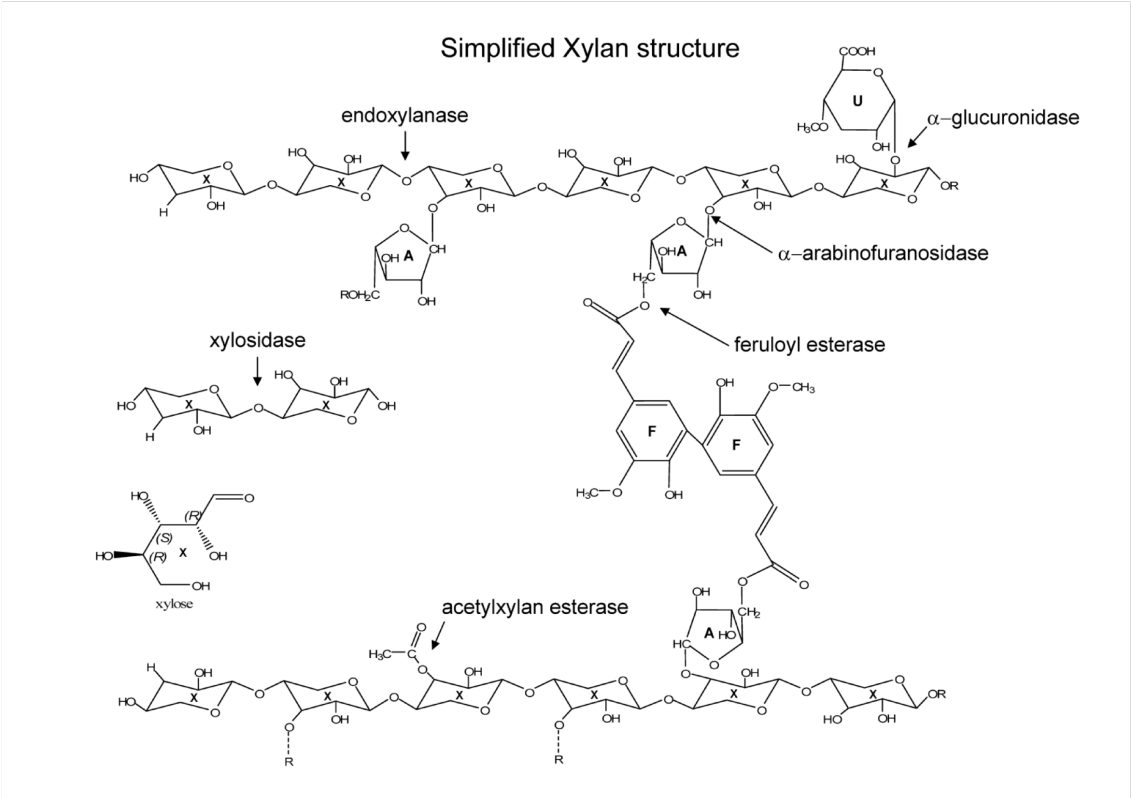


Figure 1-2: Simplified xylan structure with sites of enzyme activity[16]. A= arabinose, X= xylose, F= Ferullic acid, U= glucoronic acid.

In the *Poaceae* family of plants, which includes staple food grains and forage grasses, the main hemicellulose present in the cell wall is glucuronoarabinoxylan (GAX). GAXs consist of a linear ß-(1,4)-linked xylopyranose backbone to which α-L-arabinofuranosyl branches are attached via α-(1,3)- and/or α-(1,2)-linkages and α-D-glucuronic acid units are attached via α-(1,2)- links. The sugar units can also have other side groups attached including acetyl, methoxyl and feruloyl groups[5,7,11]. Plant cell walls have evolved to resist breakdown from microbial and mechanical forces[6]. The presence of these side chains poses a significant hurdle to enzymatic degradation and also increases the strength and rigidity of the cell wall[17]. Feruloyl

groups are especially important to structural integrity. They are attached to many of the arabinosyl units of the GAX and can undergo condensation with other feruloyl groups creating diferulates that crosslink the GAX to form a network, as well as forming ester linkages to lignin[5, 18]. The branching side groups of the hemicellulose present in a cell wall depends on the species, the tissue type and on the age of the tissue[13].

1.2 The Rumen

1.2.1 Physiology of the Rumen

Ruminants are cloven-hoofed mammals from the order *Artiodactyla* that includes cattle, sheep and deer. They obtain their food by browsing or grazing, meaning their diet consists of mainly plant material[19]. New Zealand ruminants graze a fresh forage diet that is high in protein and low in soluble carbohydrates[20]. Experimental observations have shown that under field conditions a ruminant spends equal amounts of time grazing, ruminating and resting[19]. Ruminating is the process of regurgitating and re-masticating the rumen contents. Ruminants regularly perform tasks such as eructation, belching to release gas from the digestive tract, and rumination; they also produce large amounts of salivary secretions. These features distinguish them from non-ruminants and provides favourable conditions for rumen microbiota[19].

The ruminant stomach has four compartments, the reticulum, the rumen, the omasum and the abomasum (Figure 1-3). The reticulum and the rumen are the first two stomachs and together they form a large fermentation vat called the reticulo-rumen, commonly referred to simply as the rumen[16].

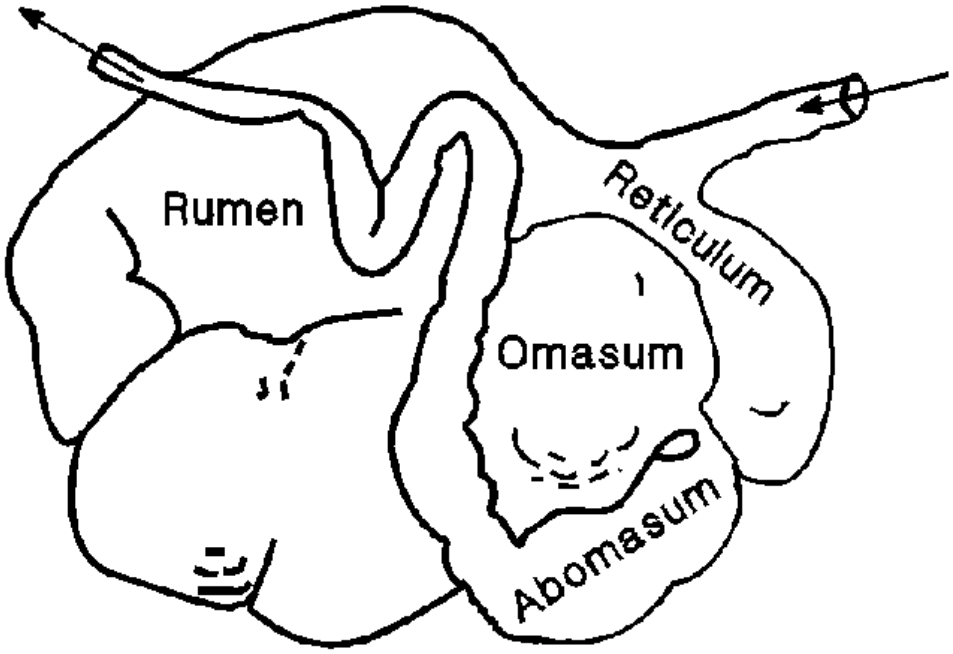


Figure 1-3: A schematic representation of the ruminant stomach showing the four compartments[21].

The entire stomach is sterile when the animal is born. The rumen has a poorly developed epithelial layer and is small in relation to the abomasum. As the young feeds the milk bypasses the rumen by the oesophageal groove[22]. Microbial colonisation of the stomach starts almost immediately with the first bacteria colonising the abomasum and the rumen via backflow. These tend to be lactic acid fermenters such as species of *Lactobacilli* and *Streptococci*. Once roughage begins to enter the rumen, the mechanical stimulus and presence of volatile fatty acids leads to the rumen developing into what is seen in an adult animal. The initial colonising microorganisms decrease to become a negligible part of the adult flora as proteolytic, celluloytic and other types of bacteria, fungi and protozoa begin to colonise the rumen and become a very complex community. Colonising species are acquired from other nearby animals, especially from the mothers mouth, from faeces and from forage material that other adults have licked or drooled on[22].

The rumen maintains a temperature of 40 ± 1 ºC by a combination of the hosts body temperature and the heat produced by fermentation[19, 22]. The pH is held within the

range of 4.5-7.0 and the rumen fluid has an osmotic pressure similar to that of blood[19]. The lower the pH the faster volatile fatty acids are absorbed and the acidity is countered by sodium bicarbonate and phosphate in the saliva. It has been reported that a cow can produce anywhere between 40 L to 190 L of saliva per day[22]. The saliva is largely a bicarbonate-phosphate mixture with a pH about 8.2, this is important for neutralisation, saliva is also an important source of water[19, 22]. The saliva contains many minerals required by the microbial population including sodium, potassium, calcium, magnesium and chloride[22].

The rumen has a large and diverse community of microorganisms that rapidly colonise and digest feed particles[23]. There is a symbiotic relationship between the ruminant and microorganisms in the rumen[23]; the microorganisms depend on the ruminant for the intake of food, mixing and propulsion, secretion of saliva, and removal and supply of substances through the rumen wall[19], and the ruminant relies on the microorganisms to breakdown forage material and ferment it to simpler compounds, including volatile fatty acids, that can be used by the animal for energy and growth[16]. One of the most important animal adaptations is the separation of the stomach into four compartments and the position of the rumen prior to the abomasum within the ruminant gastrointestinal tract. This means that the microbial community have first access to the ingested forage material[16].

Mixing of the rumen contents is very important for the digestion of plant material. It allows fresh ingesta to be inoculated with the resident micrgoorganisms, distributes the saliva throughout the rumen, increases absorption by continually replenishing the fermentation acids at the surface of the rumen epithelium and counteracts the flotation of solids during fermentation[19].

1.2.2 Fermentation within the Rumen

In the rumen, microorganisms adhere to the insoluble biomass of plant material[24] and through a complex process ferment it into volatile fatty acids that are then absorbed

by the host and used for energy; microbial cells provide the majority of the protein to the host[22]. The downfalls of this are that the volatile fatty acids then have to be re-synthesised to glucose and fat by the ruminant, and the ruminant cannot utilise the nucleic acids and some of the peptides from the microbial cells[22].

The most common volatile fatty acids are acetic, propionic and butyric acids. They have been reported in the rumen at a ratio of 7:2:1[22]. Acetic acid is produced by more rumen microorganisms than any other fermentation product. Propionate and butyrate are each formed by roughly a quarter of all ruminant microorganisms[19].

As well as these small volatile fatty acids, other compounds are also produced by microbial fermentation in the rumen. Significant amounts of ammonia are produced which is absorbed into the bloodstream and converted into urea in the liver. This is then either excreted or returned to the rumen where it is utilised by the microbial population[22].

The fermentation of sugars results in the production of a large amount of carbon dioxide, hydrogen and methane; these gases are either used in metabolism by other microorganisms or belched out by the animal. A cow can produce up to 1500 L of gas a day[22] and this gas can cause problems and even death when not efficiently removed.

The ruminant also relies on the microbial population for many vitamins and other processes. B-vitamins needed by the host are synthesised by the rumen microbes[22]. Sulfate present in feedstuff is converted into sulfide, excess levels of which have detrimental effects on the host. Species of bacteria from the *Desulfotomaculum* genus are known to be sulfate reducers in the rumen[22, 25].

1.2.3 The Rumen Ecosystem

Ruminants, like most herbivorous animals, rely on symbiotic micro-organisms contained in their digestive tracts to digest plant material[22]. Microorganisms adhered to solid particles have been shown to account for 70-80% of microbial biomass; the

other 20-30% consists of microorganisms in the rumen liquid[26]. The turnover of plant material in the rumen is much more efficient than in other environments such as soil[27] but it is still a relatively slow process so host organisms have adapted by slowing the passage of food in an enlarged area of the digestive tract. This allows a large microbial population to develop[22]. In ruminants this area is the fore-stomach, in non-ruminants it is generally areas of the intestines.

Investigations into the rumen ecosystem has increased our understanding of other anaerobic ecosystems due to the rumen being more strictly anaerobic than most systems[28]. The discovery that these microorganisms are strictly anaerobic and have specific nutritional requirements has allowed the culture of some species[28]. *Prevotella* spp. and *Butyrivibrio* spp. are the most commonly cultured rumen bacteria and it is frequently found that many divergent strains coexist within a single host [27].

The microbial ecosystem within the rumen is poorly understood[27]. Evidence for this includes the inability to culture many of the rumen microbes and the inability to inoculate the rumen with new bacteria; shortly after introduction the new species fall below detectable levels[23]. It is thought the established microbial community are so well adapted to their environment they can out compete organisms that could, in theory, thrive in this environment[22]. Significant effort has been devoted to studying rumen bacteria, less effort has been placed on studying the rumen protozoa and fungi, the latter were only first recognised as being present in the rumen in the late 1970s[23, 27]. The rumen microorganisms are well adapted to the physical conditions of the rumen. The environment has a continual supply of ingesta provided by the host and contains moderate concentrations of fermentation products which the host removes as well as other waste products, therefore the system remains relatively constant allowing a dense and diverse population to become established[19]. This population includes: bacteria, up to 10^{11} viable cells per mL comprising 200 species in over 50 genera; Ciliate protozoa, concentrations ranging from 10^{4}-10^{6} per mL made up of over 25 genera; Anaerobic rumen fungi, zoospore population densities of 10^{3}-10^{5} per

mL divided into 5 genera; Bacteriophage, at levels of 10^7-10^9 particles per mL[29].

The rumen microbial ecosystem is a stable and finely balanced system, however it is not fixed and will change when presented with a new diet to adapt to the different feedstock[30]. If the change is too sudden or drastic in amount or type of feedstuff, particularly if the shift is toward some of the modern intensive fattening diets, the balance can be upset with potentially fatal consequences to the animal[22].

1.2.3.1 Rumen Bacteria

There is an immense diversity in the types of rumen bacteria[19]. Most rumen bacteria are cocci and short rods of a variety of sizes but generally between 0.4-1.0 μm in diameter and 1-3 μm in length[19]. They have adapted to live in an environment that has an optimal pH of between 6 and 6.9 and an optimal temperature of 40 °C. They can also tolerate high levels of organic acids[30].

Cellulolytic bacteria of the rumen metabolise cellulose as their major feedstuff. There are many cellulolytic bacteria in the rumen but the most widespread is *Fibrobacter succinogenes*[29] (formerly *Bacteroides succinogenes*[31]), a Gram-negative rod that produces principally succinic acid with some acetic and formic acids[22].The *Ruminococci* species include *Ruminococci albus* and *Ruminococci flavefaciens* which are Gram-negative or Gram-variable small cocci that ferment cellulose forming hydrogen, formate and acetate, *R. albus* which also produces ethanol and lactate and *R. flavefaciens* also produces succinate. There are other isolates present that are believed to belong to this genus[22]. Other cellulolytic bacteria found in the rumen include *Butyrivibrio fibrisolvens*, a small Gram-negative vibrio that ferments a wide range of sugars forming butyric acid[22]; and *Eubacterium cellulosolvens* (formerly *Cillobacterium cellulosolvens*) which is a Gram-positive rod that can form hydrogen, formate, butyrate, lactate and valerate[22]. Many of these cellulolytic bacteria also hydrolyse xylans. Members of the *Butyrivibrio* genus are probably the most widely distributed hemicellulose fermenter, but the *Ruminococci*, some *Bacteroides* and

strains of *Eubacterium ruminantium* are also important in the degradation of hemicellulose[22]. Pectin is another polysaccharide present in plant material that can be hydrolysed and fermented by rumen bacteria; those known to utilise pectin include *Butyrivibrio fibrisolvens*, *Bacteroides ruminicola*, *Bacteroides succinogenes* and *Lachnospira multiparus.*

Energy dense cereal based feed supplements are being incorporated into many domesticated ruminant diets. This increases the availability of feedstuff for amylolytic bacteria. *Streptococcus bovis* is an amylolytic bacterium almost always present in the rumen in small amounts, it can ferment a range of sugars so it is not reliant on the presence of starch[22]. In the presence of high starch diets it could play an important role in releasing sugars for other non-amylolytic bacteria. *Bacteroides* species are probably the most important starch fermenting bacteria, especially for animals on intense cereal rations. *Bacteroides amylophilus* is a Gram-negative pleomorphic rod that can assume very odd shapes. It produces formic, acetic and succinic acids from starch[22].

Not all rumen bacteria can hydrolyse polysaccharides and thus must rely on either the small amounts of monomeric sugars present in plants or on the products of polysaccharide hydrolysis not utilised by the hydrolytic bacteria. *Lactobacilli* do not occur in large numbers in the adult rumen but can become predominant if the overeating of starchy foods occurs. A *Borrelia* sp. that is not cellulolytic, but can ferment cellobiose, has been seen to move through agar to grow near colonies of cellulolytic bacteria presumably to utilise their polysaccharide hydrolysis products[22]. *Eubacterium ruminantium,* a Gram-variable rod, probably lives mostly by sugar fermentation[22].

Bacteria can also utilise proteins, amino acids and fats present in the rumen contents. Many rumen bacteria including *Bacteroides*, *Selenomonas*, *Butyrivibrio*, *Succinivibrio*, and *Borrelia* perform proteolysis. *Butyrivibrio proteoclasticus* (formerly *Clostridium proteoclasticum*) was originally isolated as a proteolytic

bacteria[32]. Some species of bacteria can deaminate amino acids, but *Bacteroides ruminicola* appears to be the most active deaminating bacterium in the rumen. All these bacteria need carbohydrates as well, as they cannot grow solely on amino acids or protein[22]. Some bacteria use lipids as their feed source, *Anaerovibrio lipolytica* is a Gram-negative vibrio originally isolated as a long-chain glyceride fermenter. It ferments the glycerol portion of a glyceride producing acetic, propionic and succinic acids. *Selenomonads* seem particularly active in fermenting the glycerol produced by lipolysis[19].

Two of the important methanogenic bacteria in the rumen are the Gram-positive coccobacillus *Methanobacterium ruminantium* and *Methanobacterium mobilis*, a Gram-negative flagellated coccoid rod. Both require a very highly reduced media for isolation and utilise only hydrogen plus carbon dioxide or formate[22].

The great majority of the rumen bacteria are obligately anaerobic and require carbon dioxide for growth, but there are some species present that are capable of living with oxygen levels up to the equivalent of atmospheric levels[19]. Bacteria closely related to those in the rumen have been found in the gastrointestinal tracts of many other mammals including rabbits, horses, porcupine, and humans[19].

1.2.4 Role of *Butyrivibrio proteoclasticus* within the Rumen

Butyrivibrio proteoclasticus and other *Butyrivibrio* species have been found in the rumen of cattle, sheep and deer, as well as in other mammalian gastrointestinal tract; as their name indicates they are thought to be predominant producers of butyrate in the rumen[16, 33, 34].

The fermentation end products of *B. proteoclasticus* have been shown to include formate, butyrate, acetate, propionate, succinate and hydrogen[34]. The ability of *B. proteoclasticus* to break down protein and to utilise a range of carbohydrates places it among the proteolytic-saccharoclastic rumen bacteria which ferment the carbohydrates released from the breakdown of plant fibre (cellulose and

hemicellulose) and obtain their nitrogen by breaking down plant proteins[34]. *B. proteoclasticus* is likely to be involved in the primary hydrolysis of feed protein to peptides and amino acids but it is not involved in the fermentation of peptides based on the evidence that it does not produce ammonia as a result of fermentation and is unable to grow on peptone-yeast extract alone[34].

1.2.5 The Genus*: Butyrivibrio*

The *Butyrivibrio* genus (phylum *Firmicutes*, family *Lachnospiraceae*[35]) was first proposed as a genus by Bryant and Small (1956) with the type species being *Butyrivibrio fibrisolvens,* an isolate from the rumen of cows[36]. Their description of the proposed genus was as follows: an anaerobic, non-spore forming, monotrichous, Gram-negative curved rod that ferments glucose with the production of large quantities of butyric acid[36]. They noted it had similarities to other genera including *Vibrio*, *Desulfovibrio*, *Cellvibrio* and *Cellfalcicula* but differed from these genera by the large quantities of butyric acid it produced. In this research they investigated 48 strains they had previously isolated[37] and described a type species for the genus. Due to a lot of variation between the strains the type species was based on a strain among those with the most representative physiological characteristics. The organism, called *Butyrivibrio fibrisolvens*, is a Gram-negative, motile, slightly curved rod, 0.4-0.6 μm wide and 2-5 μm long with blunt tapered ends. Arrangements of cells include single, pairs and long chains. Flagellation is monotrichous and polar[36].

In 1976 Moore *et al* isolated and characterised *Butyrivibrio crossotus* from human faecal matter[38]. With the grouping of this organism into the *Butyrivibrio* genus an alteration to the genus description was necessary. *B. crossotus* isolates were seen to be lophotrichous, having multiple flagella at the same spot, but due to very few other differences from *B. fibrisolvens* a new genus was not justified[38].

The physical characteristics of *Butyrivibrio* were also challenged when Cheng *et al* (1977) showed that although *Butyrivibrio* sp. stained Gram-negative they in fact had

a Gram-positive morphological type[39].

Butyrivibrio sp. have been shown to be phylogenetically diverse[23]. They have been isolated from the gastrointestinal tracts of many animals[40] and although there are few described species in the genus, it has been reported that many isolates of the genus are genetically unrelated[40, 41]. Willems *et al* (1996) compared 16S rRNA, from 40 isolates of *Butyrivibrio* sp. which resulted in the grouping of these isolates in three remotely related groups[40]. Forster *et al* (1996) also reported that DNA/DNA relatedness investigations showed the genus consisted of genetically unrelated strains but that all strains were related to bacteria that grouped in the *Clostridium* cluster XIVa[41].

In 1996, vanGylswyk *et al* reported isolating a bacterium from the rumen of a cow. The description of the organism closely resembled *B. fibrisolvens* but with important differences, namely that the species produced butyrate as an end product of fermentation but was not xylanolytic, amylolytic or proteolytic, and that it also differed in cell morphology and contained no C18 fatty acids[42]. A new but closely related genus *Pseudobutyrivibrio* was suggested and the type species was called *Pseudobutyrivibrio ruminis*[42]. Also in 1996, Attwood *et al* isolated the novel bacterium *Clostridium proteoclasticum* from the rumen of cattle and noted that it was a *Butyrivibrio*-like organism[34]. In 2008 it was reclassified as *Butyrivibrio proteoclasticus*[33]. In 2003, Kopecny *et al* isolated two novel species of bacteria from the rumen fluid of cow and sheep; *Butyrivibrio hungatei* and *Pseudobutyrivibrio xylanivorans*[43]. 16S rRNA analysis of *B. hungatei* showed that the strain was located in *Clostridium* cluster XIVa and was closely related to *B. fibrisolvens* and *B. proteoclasticus* (then *C. proteoclasticum*)[43]. Analysis of the 16S rRNA gene from *Pseudobutyrivibrio xylanivorans* showed that the related isolates were located in *Clostridium* cluster XIVa and it was closely related to *Pseudobutyrivibrio ruminis* and *Butyrivibrio crossotus*[43].

Butyrivibrio species play an important part in the rumen ecosystem[19, 22] and are

efficient utilisers of xylan[23], producing butyrate as their major fermentation product. It has been suggested that the number of *Butyrivibrio* in the rumen is high enough to account for total butyrate production although other organisms have been seen to produce butyrate[19]. *Butyrivibrio* are also able to degrade starch[44] and protein[22] and they have been shown to be involved in the bio-hydrogenation of C-18 fatty acids[45, 46]. *Butyrivibrio fibrisolvens* has been shown to produce bacteriocins, small peptides that are inhibitory to the growth of other bacteria including other *Butyrivibrio* isolates[47-49]. *Butyrivibrio* species have been isolated from animals of many species and geographic locations[36, 40]. They are among the most commonly cultured rumen bacteria[27].

1.2.6 Description of *Butyrivibrio proteoclasticus*

Butyrivibrio proteoclasticus is an anaerobic, straight to slightly curved rod shaped Gram-positive bacterium with a single subpolar flagellum[34]. The result of staining experiments showed that younger colonies stain Gram-positive and older colonies stain Gram-negative, but based on cell envelope morphology it is classified as a Gram-positive bacterium[34, 50]. The cells occur in short chains with tapered ends and do not form spores; experimental data has indicated they have a generation time of 1.44 hours[34]. This species has extremely high proteinase activity that is predominantly serine-type activity[20]. Electron microscopy images of *B. proteoclasticus* are shown in Figure 1-4.

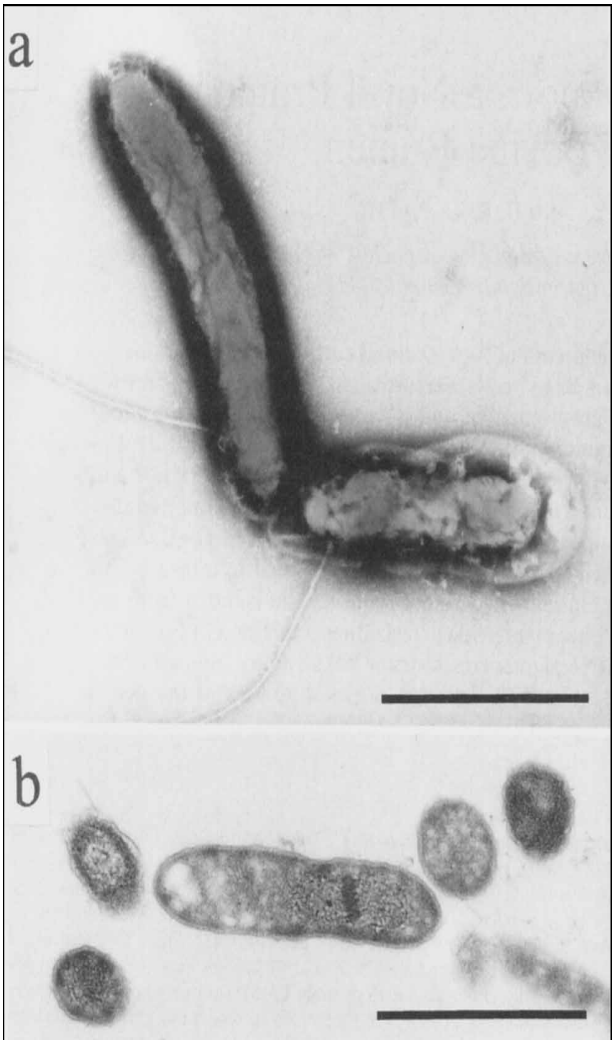


Figure 1-4: Electron micrographs of negatively stained strain B316^T cells, showing the flagellar arrangement (a), and thin-sectioned strain B316^T cells, showing the Gram-positive cell wall structure (b). Bars = 1.0 μm[34].

1.2.7 Isolation and Characterisation of *Butyrivibrio proteoclasticus*

The *B. proteoclasticus* type strain, termed strain B316^T, was originally isolated from the rumen of New Zealand cows on a forage diet containing mainly ryegrass and clover during a study of proteolytic rumen bacteria by Attwood *et al* in 1995[32, 34]. It has since been identified in rumen samples from animals on a range of diets, at levels between 2.01 x10^6 and 3.12 x10^7 cells per mL. This small range indicates that the population is relatively stable independent of dietary content[20]. Initially, this strain

was identified as *Butyrivibrio*-like but named *Clostridium proteoclasticum*. The description of the *Clostridia* genus is bacteria that are Gram-positive, obligately anaerobic, endospore-forming rods which are not able to carry out dissimilatory sulfate reduction[34]. However the *Clostridium* subcluster XIVa is both phenotypically and phylogenetically diverse, and many of its member species do not conform to the *Clostridia* genus discription[34]. In contrast, the description of a *Butyrivibrio* genus is typically curved, strictly anaerobic, non-spore forming, rod-shaped cells that are generally motile by means of one or more polar to subpolar flagella. They possess an atypical Gram-positive ultrastructure, although they stain Gram-negative, and have a fermentative metabolism with butyrate as the main product[33].

The initial classification of the bacterium as *C. proteoclasticum* was based on phenotypic characterisation and phylogenetic analysis of the 16S rRNA gene from strain B316^T, and indicated that is was most closely related to a member of *Clostridium* cluster XIVa, *viz*., *Clostridium aminophilum*, an amino acid fermenting organism isolated from the rumen[34] (similarity value = 92.2%). This resulted in the strain B316^T being tentatively assigned to a novel species of the genus *Clostridium, C. proteoclasticum,* although it was noted at the time that this classification was likely to change upon a review of *Clostridium* subcluster XIVa[34].

Initially, strain B316^T had a reported DNA G+C content of 28%[34], this is typical of a *Clostridia* whereas *Butyrivibrio* species tend to have a much higher G+C content between 38-42%[33]. However, later investigation found that the DNA G+C content originally reported for B316^T was wrong and a new value of 39.33% was determined after the entire genome had been sequenced. The new value is in the range of that characteristic of the genus *Butyrivibrio*[16, 33]. Phylogenetic analyses were also repeated to include gene sequences encoding 16S rRNA from a wider range of taxa from the *Clostridium* subcluster XIVa, including sequences from the genera *Butyrivibrio* and *Pseudobutyrivibrio* (Figure 1-5). The results showed that sequences from *C. proteoclasticus* cluster closely with sequences from *Butyrivibrio* and that it is most closely related to *Butyrivibrio hungatei,* although importantly *C. proteoclasticus* can

be distinguished from this species by its ability to form stearic acid from linoleic acid, a feature that has not been observed in any other strain of the genera *Butyrivibrio* or *Pseudobutyrivibrio*[33].

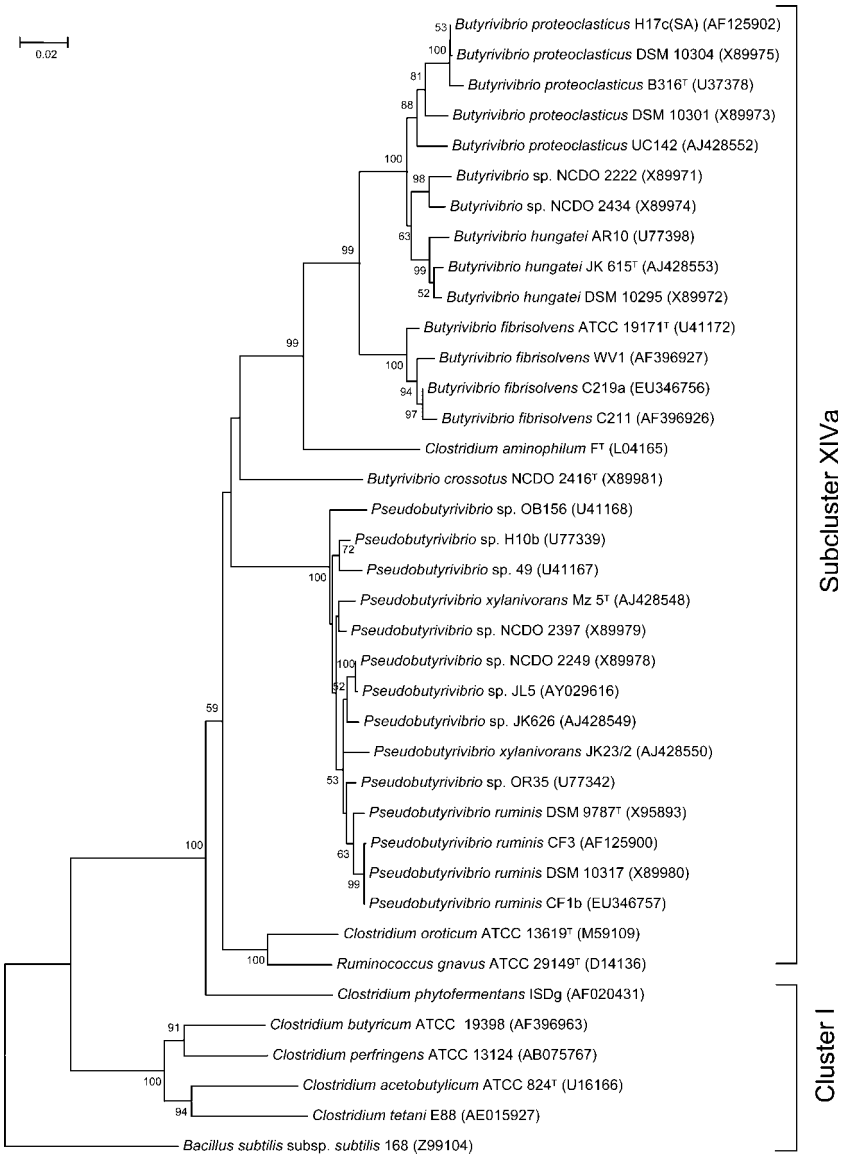


Figure 1-5: 16S rRNA gene phylogeny using near full-length sequences from clostridial cluster I and subcluster XIVa strains containing *B. proteoclasticus* taken from Moon *et al*[33].

Phylogenetic, genetic, morphological and physiological data indicated that

C. proteoclasticum should be reclassified as *B. proteoclasticus*, and that it was a different species within that genus based on both physiological and genetic analyses[33].

1.2.8 The Genome of *Butyrivibrio proteoclasticus*

The Attwood group at Agresearch has sequenced the entire genome of *B. proteoclasticus.* They discovered that the genome of *B. proteoclasticus* has an interesting architecture in that it contains a main chromosome and three smaller replicons. The main chromosome is just over 3.5 Mb in size; the replicons consist of a chromid of 302 kb and two mega-plasmids of 360 kb and 186 kb. The main chromosome encodes around 80% of the genes in the genome, the small chromid was labelled as such because it encodes two rRNA operons and contains unique copies of several genes from the minimal gene set[16]. Only about 15% of the genes on the two megaplasmids generated matches in BLAST searches with genes held in the public database. The majority of genes however generate no matches, thus their functions are currently unknown[16]. Many of the genes from *B. proteoclasticus* have low sequence identity with known genes in public databases and because of this have been annotated with 'putative' functions[51].

Of the two megaplasmids the smaller one, called pCY186, is readily lost from *B. proteoclasticus* cells on prolonged subculture in the lab without lethal consequence, however the larger one, pCY360, is maintained[16].

The sequencing of the *B. proteoclasticus* genome revealed that around 20% is devoted to the breakdown and reassembly of polysaccharides[1], this includes glycoside hydrolases (GHs), glycosyltrasferases and the genes associated in loci association with polysaccharide utilisation and exopolysaccharide production[1]. A large number of genes have been annotated as involved in fibre degradation. The genome contains 113 GHs, 121 glycosyltrasferases [1] as well as carbohydrate esterases (CEs), polysaccharide lyases (PLs) and carbohydrate binding domains (CBDs),

although some of the glycoside hydrolase genes contain a CBD as well[16]. All these genes are found on either the main chromosome or the smaller chromosome[16]. *B. proteoclasticus* can utilise a range of polysaccharides, including pectin, xylan and starch, these insoluble substrates require the bacterium to adhere to the surface of the substrate and secrete enzymes to initialise degradation. Many bacteria incorporate catalytic domains with scaffoldin, cohesion and dockerin components to create a large extracellular multi-enzyme complex called a cellulosome[52]. The cellulosome binds to the insoluble substrate as well as to the bacterium cell surface[52]. *B. proteoclasticus* lacks any scaffoldin or cohesion domains[1] and therefore must adhere to insoluble substrates in an alternative manner. There are nine multi-domain proteins produced by *B. proteoclasticus* that initiate this extracellular breakdown (Xyn10B, Amy13A, Lic16A, Xsa43J, Agn53A, Pme8B, Est12B , Pel1A and Bpr_I0264) and these are all secreted and contain cell-wall binding domains[1]. It is estimated one third of the polysaccharide degrading enzymes produced by *B. proteoclasticus* are secreted[1]. The attachment of these secreted enzymes to the cell wall of *B. proteoclasticus* allows for efficient uptake of oligosaccharides as they are released from the polysaccharide substrate[1]. The remaining two thirds of the enzymes involved in fibre degradation are intracellular. These enzymes belong to a range of glycoside hydrolases families and have a diverse array of annotated functions. This indicates that there are a broad range of oligosaccharides transported into the cell that can be metabolised by *B. proteoclasticus*[1]. Most of the genes encoding these intracellular polysaccharide degrading enzymes are clustered in polysaccharide utilization loci, which also contain genes for transporters, transcriptional regulators and environmental sensors[1]. This allows the bacterium to detect the type of oligosaccharides, and the sugar linkages within that, and produce the appropriate type of degrading enzymes and transporters to utilise it[1].

1.3 Enzymatic Catalysis

1.3.1 General Enzyme Catalysis and Kinetics

A catalyst is an atom or molecule that increases the rate of a reaction without itself being consumed. A catalyst does not affect the equilibrium of the reaction; it reduces the transition state energy allowing the equilibrium to be reached faster. A protein that interacts with a ligand resulting in a covalent change to that ligand is called an enzyme and the ligand a substrate. Enzymes catalyse the change in the substrate but are themselves unaltered[53]. An enzyme is a protein that is folded in such a way that allows it to become a catalyst. This fold usually has a binding pocket with amino acid residues that interact with the substrate in order to orientate it correctly, and catalytic residues that facilitate the change required to produce the product. The catalytic residues can work by either covalently bonding to the substrate resulting in an intermediate that is then converted to the product, leaving the enzyme in its original state or, via electrostatic effects where the catalytic residues interact with the substrate resulting in a strained environment; this lowers the amount of energy required for the substrate to reach the transition state and thus makes the change from substrate to product easier.

The kinetic parameters of the rate at which an enzyme will catalyse a reaction are very useful for comparing the efficiencies of different enzymes. Enzyme kinetics is typically steady state kinetics, where the production and degradation of enzyme-substrate intermediates is balanced resulting in a steady concentration of enzyme-substrate intermediates[54]. It is found experimentally that the initial rate (v) of the reaction increases with the concentration of substrate ([S]); this relationship is initially linear until significantly high [S] where v increases more slowly and at a sufficiently high [S], v tends toward a limiting value that is termed V_{max}; v also depends on the initial concentration of the enzyme, $[E]_0$. This relationship is presented in Figure 1-6. The catalysis from substrate to product by an enzyme is represented in Equation 1. The Michaelis-Menten equation (Equation 2) expresses

this relationship quantitatively, and is the basic equation of enzyme kinetics[52].

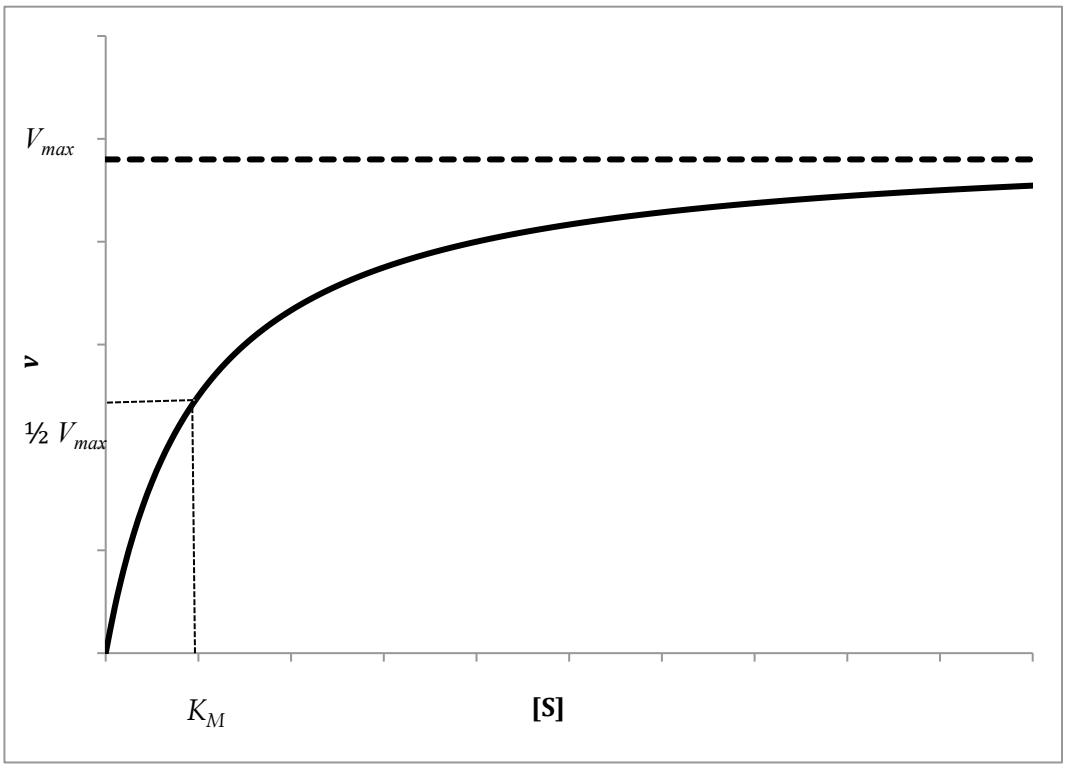


Figure 1-6: Reaction rate plotted against substrate concentration for a reaction obeying Michaelis-Menten kinetics.

The important parameters are K_M - the Michaelis-Menten equilibrium constant (Equation 3); and k_{cat} - the catalytic constant, which is a first order rate constant for the reaction of substrate to product[54].

Equation 1: Diagram of substrate binding and enzyme catalysis to form product.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} [ES] \xrightarrow{k_2} E + P$$

Equation 2: The Michaelis-Menten Equation**.**

$$v = \frac{[E]_0 . [S] k_{cat}}{K_M + [S]}$$

Equation 3: the Michaelis-Menton constant

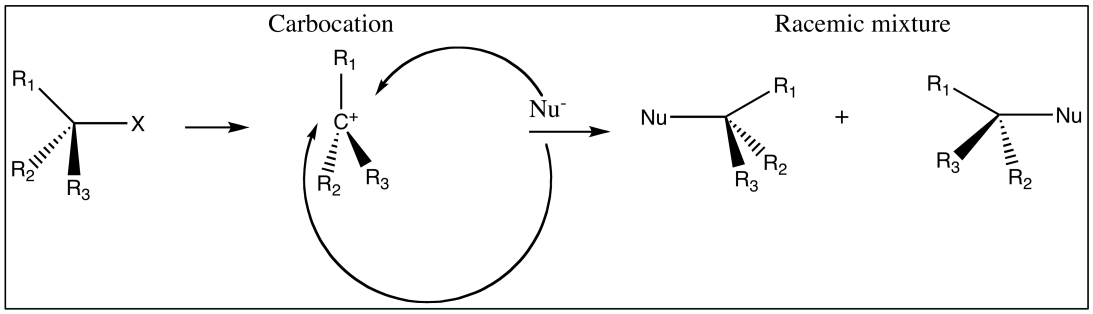
$$K_M = \frac{k_2 + k_{-1}}{k_1}$$

1.3.2 Enzymatic degradation of Carbohydrate Polymers

There are four main classes of enzymes considered active on polysaccharides. They are Glycoside hydrolases (GHs), Glycosyltransferases (GTs), Polysaccharide lyases (PLs) and Carbohydrate esterases (CEs). Many carbohydrate active enzymes also have additional domains that are involved in binding of carbohydrates.

Glycoside hydrolases hydrolyse the glycosidic bond between two sugar units and are crucial in the degradation of polysaccharides for energy utilisation and also within a plant cell wall for remodelling. The mechanism of hydrolysis is general acid catalysis[23] and involves a catalytic acid residue and a catalytic base. It can result in either the inversion or retention of the configuration at the anomeric carbon depending on the relative spatial location of the two catalytic residues[55]. The mechanism that results in retention of the anomeric carbon is a double displacement (Figure 1-7a), S_N1-like mechanism while the mechanism resulting in inversion (Figure 1-7b) is a single displacement S_N2-like mechanism[23, 56]. A true S_N1 mechanism results in a racemic mix however within an enzyme, access to the carbon cation may be restricted to one side resulting in retention.

(a) S_N1 Mechanism



(b) S_N2 Mechanism

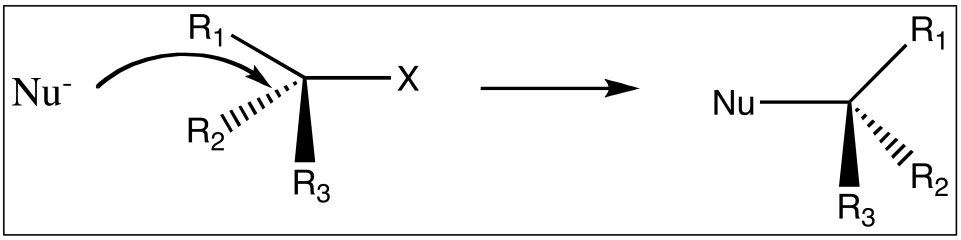


Figure 1-7: Representations of (a) S_N1 and (b) S_N2 mechanisms.

Glycosyltransferases catalyse the transfer of an activated monosaccharide donor onto an acceptor molecule creating a glycosidic bond[57]. These are involved in a range of processes, as a high number of all proteins are glycosylated. They can also glycosylate various other metabolites and are responsible for polysaccharide and cell wall formation[57]. Glycosyltransferases can have varying mechanisms that again result in either the retention or inversion of the configuration of the anomeric carbon. In inverting glycosyltransferase mechanisms a catalytic side chain (the base) deprotonates the incoming nucleophile (the acceptor). This results in displacement of the activated phosphate leaving group[56]. In retaining glycosyltransferase mechanisms the typical mechanism is a double-displacement reaction involving a covalently bound glycosyl-enzyme intermediate species, an oxocarbenium ion. This prevents nucleophilic attack from the opposite face of the reaction centre and only allows

retention of anomeric configuration in the product[56].

Polysaccharide lyases cleave the glycosidic bond of uronic acid containing polysaccharides via a β-elimination mechanism resulting in the formation of a double bond at the newly formed non-reducing end. This type of glycosidic bond cleavage is non-hydrolytic[55].

Carbohydrate esterases hydrolyse ester linkages within polysaccharides, between sugar units and side groups such as acetyl, methyl, feruloyl and *p*-coumaric groups. The carbohydrate esterases have a range of mechanisms but the most common is analogous to the mechanism of lipase and serine proteases that involves a Ser-His-Asp catalytic triad[55].

Another class of enzyme that is intimately involved with polysaccharide degradation is Carbohydrate Binding Domains (CBDs), also referred to as carbohydrate binding modules (CBMs). These domains are frequently seen as part of a carbohydrate active enzyme. When a CBD binds to its substrate it causes distortion in the packing and conformation of the polymer, which in turn assists substrate degredation[58]. Carbohydrate binding domains increase the rate of catalysis by increasing the local concentration of the catalytic domain at the surface of the carbohydrate substrate[59].

All these enzymes are grouped into families in a database called CAZy: **C**arbohydrate **A**ctive En**Zy**mes (http://www.cazy.org/)[55]. CAZy has been online since 1998 and families are created based on experimental characterisation of enzymes. Further enzymes are added to a family based on amino acid sequence similarity and structural information where available, therefore these classifications correlate better with mechanism and protein fold than substrate specificity[23]. In the CAZy database there are 122 separate glycoside hydrolase families grouped into 14 clans, 92 glycosyltransferase families, 22 PL families, 16 CE families and 61 carbohydrate binding domain families. These are continually being reassessed and upgraded as further information enters the public domain[55].

Enzymatic hydrolysis of polysaccharides requires a suite of specific enzymes to hydrolyse the polymer into monomer sugars. Cellulose needs to be broken down to cellodextrins in order for the organic carbon to become available for microbial growth[9, 60]. Cellulose can be effectively hydrolysed by three different enzymes that work synergistically; endoglucanases attack low crystallinity regions of the cellulose microfibril and create free ends by cleaving the cellulose chain, exoglucanases cleave cellobiose from the free ends of the chains and ß-glucosidases cleave the cellobiose into glucose units[11, 61]. The complete degradation of hemicellulose requires a much larger array of enzymes. Some of the most important glycoside hydrolase enzymes involved are: exo- and endo-1,4-ß-D-xylanases, that hydrolyse xylooligosaccharides; ß-mannosidases which cleave ß-mannose units; α-glucuronisidases that cleave α-glucuronic acid groups; α-L-arabinofuranosidases that cleave arabinose side groups from xylose backbones, arabinanases which cleave arabinose units from arabinose oligosaccharide chains and α-galactosidases that cleave α-galactose units. There are also a range of CE enzymes necessary for the complete degradation of hemicelluloses. Acetyl xylan esterases remove acetate groups attached to the sugar residues. These groups are proposed to be a significant impediment to other enzymes accessing the sugar backbone of hemicelluloses. In addition, feruloyl esterases, cleave ferulic acid side groups from the sugar units, and diferulic acid crosslinks between hemicellulose and lignin[7, 23, 61].

It had generally been considered that animals can only breakdown cellulose and hemicellulose due to symbiotic microorganisms found in their gut; however examples of carbohydrate active enzymes from animal sources have been reported in the literature. In 1998 Wantanabe *et al*[62] reported isolating an endogenous cellulase gene in the termite *Reticulitermes speratus*. *R. speratus* had previously been thought to digest cellulose exclusively through a symbiotic protist[62]. Endogenous cellulase genes have since been found in many invertebrates such as insects, nematodes and mollusks[63, 64]. The endo-1,4-β-xylanase gene of animal origin belonging to glycoside hydrolase family 10 was initially reported from the freshwater snail *Ampullaria*

crossean. It was described as a "multifunctional cellulase"[65], and since then xylanase enzymes have been reported from gastropods[66], nematodes[67] and mollusks[63].

1.4 Potential of Fibre Degrading Enzymes in Industrial Applications

People are using enzymes to aid in a multitude of manufacturing processes. In ancient times enzymes have been used in processes such as production of wine, cheese, and bread[14], the applications in today's world are much broader. The potential applications for fibre degrading enzymes are many, and range across a myriad of industries including the food, animal feed, textile, fuel, pharmaceutical and chemical industries as well as the pulp and paper industry and waste management[68].

In the food industry fibre degrading enzymes are used in a range of applications including extraction and clarification of fruit juices, coffee and oils. Enzymatic removal of cell walls facilitates the release of flavours, polysaccharides and proteins; this results in easier fermentation in products such as soybeans. Enzymes are also used for improving the rehydration of dried vegetables and the production of oligosaccharides and soluble sugars from cellulosic wastes[14, 68]. In the wine and beer industry enzymes are used to hydrolyse oligosaccharides and help in the filtration of beer and increase the aroma in wine[68]. In the baking industry xylanases are added into wheat flour to improve dough handling and the quality of the baked products[14].

In the animal feed industry, enzymes are used as supplements given directly to animals to aid digestion. They are also used in pre-treating lignocellulosic material, dehulling of cereal grains and treating silage to improve the digestibility and thus energy available to the animal[68]. Xylanases are used as food additives in the poultry industry; this improves the feed conversion efficiency in chicks and results in increased weight gain[14].

Enzymes are used in the textile industry to alter and enhance the colour and softness of materials, especially cotton[68]. They are also used to degum fibre sources such as

flax, hemp and jute[14]. Xylanases are used in the pulp and paper industry for the pre-treatment of pulp and paper; resulting in brighter paper products and reduces the use of harsh chemicals, such as chlorine, in paper pulp bleaching[14, 69].

Enzymes are being utilised to provide renewable sources of feed-stock molecules for the chemical and pharmaceutical industries[69]. Methods are available for converting polysaccharides into acetate, propionate, lactate or succinate, via monomer sugars.

Almost half of the global carbon fixed annually by photosynthesis is incorporated into the plant cell wall[23]. Plant material is the most renewable carbon source on earth. Unlocking these fermentable sugars efficiently is a difficult task[6], however, once achieved it potentially has significant economic and environmental benefits in many different industries including the production of biofuels.

Currently biofuels are of great interest as a sustainable alternative to ever-decreasing reserves of fossil fuels[70]. Both the United States of America and the European union have expressed the desire for a significant proportion of their transport fuels to be biofuels by the year 2030[71]. Currently starch from corn grain and simple sugars from sugar cane and beets are being used as a starting material for biofuel[71]. This has severe impacts globally as these products are important feedstuffs for people and animals; thus it is causing an increase in global food costs. The increasing desire for biofuels makes it important to shift the raw materials away from feedstuffs and toward using agricultural and horticultural by-products, such as wheat and barley straw from the beer brewing industry. The shift in starting material is essential for the sustainability of the industry, it also has benefits in offsetting the environmental concerns associated with disposing of these byproducts[72]. Significant development is required to create effective processing protocols for these alternative starting materials.

A particular focus area for biofuel production is the utilisation of xylans and heteroxylans as well as cellulose[72]. One of the prerequisites of ethanol production

from lignocellulosic material is the efficient generation of fermentable monomeric sugars from the carbohydrate polymers present in the biomass[72]. Enzymatic hydrolysis of cellulose and hemicellulose is deemed the limiting step in biofuel production[73], because currently available enzyme preparations are expensive and inefficient. Significant research needs to be directed into this field to make the production of biofuel sustainable and cost effective[71]. There are two main ways these improvements will happen – by exploring the diversity of fibre degrading enzymes to find new, more effective alternatives, and by applying protein engineering principles in attempts to improve these enzymes[71]. The latter of these is still heavily dependent on the first, as we need to be able to understand how the structure of an enzyme affects its function in order to engineer enzymes with the appropriate properties.

The idea of using cocktails containing a range of enzymes with different actions in order to fully degrade hemicellulose material is becoming increasingly popular. These cocktails can be used to degrade polysaccharide material into monosaccharides for use as starting materials for other industries. This relies on the understanding of a large range of fibre degrading enzymes, their specificity, optimal conditions and any synergy that exists between the compoments of the cocktail[74].

1.5 Three Dimensional Protein Structures

A protein is a string of amino acids that folds into a three-dimensional (3D) structure; in enzymes this allows the catalytic site and the substrate-binding pocket to be formed. The functional properties of a protein depends on its 3D structure[75]. The elucidation of this structure allows the investigator insight into which residues are important in substrate recognition, binding and catalysis within the protein.

There are two main methods used for determining 3D structures of proteins: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy[75]. X-ray crystallography is the most common technique and the one used in this project.

X-ray crystallography initially requires a protein crystal that diffracts X-rays. The molecules within the crystal are packed in a highly ordered fashion. Each equivalent molecule interacts with the neighbouring molecules in an identical way. The 3D array of molecules within the crystal makes up a lattice which is used to define the unit cell of the crystal[76]. The unit cell is a 3D block with the edges defined as the three lattice translations, the unit cell is defined as the smallest possible volume or the smallest repeating unit. The crystal is made of unit cells stacked in three dimensions.

X-rays are electromagnetic radiation at short wavelengths and can be monochromatic (one specific wavelength) or polychromatic (a broad window that allows X-ray radiation across a range of wavelengths)[75]. Monochromatic X-rays with a wavelength between 0.5 – 1.6 Å are used for protein crystallography as they penetrate into protein crystals and are scattered strongly[76]. X-rays are frequently generated either by a home source or at a Synchrotron. A home source produces an X-ray beam by focussing electrons on a metal target in a vacuum. By having a potential difference of 10s of kV, each electron in the resulting beam has enough energy to produce an X-ray photon[77]. The wavelength of the resulting X-rays is characteristic of the metal anode. In contrast, synchrotron radiation is produced by a stream of electrons travelling around a circular track at a velocity of ~99% of the speed of light; as they circle they emit radiation in a tangential direction[76]. The synchrotron produces a continuous X-ray spectrum; this allows the full range from 0.5-1.6 Å to be available to the user. Photons from the X-ray beam enter the crystal and are diffracted by the electrons of the atoms inside the crystal.

X-rays scatter in two ways, coherently and incoherently. Only coherent scattering (diffraction) is useful for crystallography. This is when a photon is absorbed by an electron which then vibrates and emits a photon of the same energy and wavelength in a random direction. The photons of the diffracted beam exiting the crystal need to be in phase in order for the reflection to be observed. A useful conceptual construct is to envisage the photons being diffracted by different planes within the crystal called

Bragg's planes (Figure 1-8).

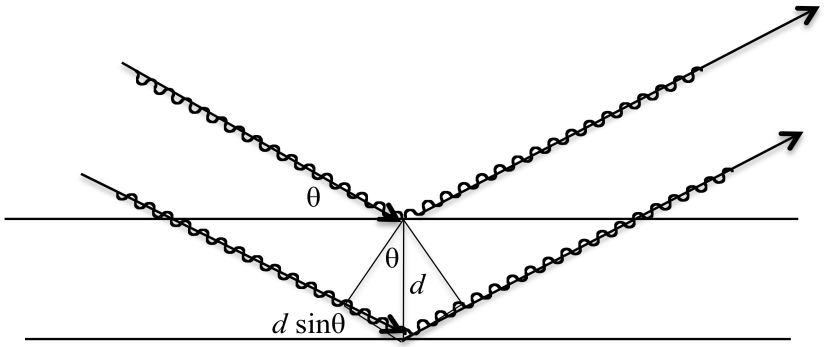


Figure 1-8: schematic of diffraction off Braggs planes.

A diffraction spot will be observed when the distance between the planes (*d*) and the angle of the photon entering the crystal (θ) are related to the wavelength (λ) as described by Bragg's Law (Equation 4).

Equation 4: Bragg's Law

$$n\lambda = 2d\sin\theta$$

A diffraction pattern is made up of many individual spots from the many diffracted photons that exit the crystal. The diffraction pattern is recorded either on image plates or more commonly by electronic detectors[53]. The crystal needs to be rotated and diffraction patterns collected at many different angles in order to determine the reflection in 3D space. Each spot is defined by three properties: the amplitude, the position (*h/k/l*) and the phase. The first two are easily determined; the spot position (*h/k/l*) can be measured and the amplitude is proportional to the square root of the intensity of the reflection[76]. The phase cannot be determined from the diffraction pattern directly. Because the wavelengths of X-rays are so short the detector is insensitive to the phase of the incoming beam.

The diffraction pattern is related to the electron density in the crystal by a

mathematical function[78]. If the electron density is considered as a sum of repeating mathematical functions then the diffraction pattern is the Fourier transform of those functions[77]. Fourier transform of a repetitive scattering object (protein molecule) represents a set of orders of diffraction, each of which is described by an amplitude and a phase[76]. The Fourier transform of the structure factors is called the electron density equation (Equation 5). Where ρ is the electron density, x/y/z represent the spatial coordinates, V is the volume of the unit cell, $h/k/l$ represent the reciprocal space indicies, $|F_{h/k/l}|$ is the amplitude of the structure factor and $\varphi_{h/k/l}$ is the phase of the structure factor[79].

Equation 5: the electron density equation[79].

$$\rho_{x/y/z} = \frac{1}{V} \sum |F_{h/k/l}| \left[\cos(\varphi_{h/k/l} - 2\pi(h/k/l) \cdot (x/y/z)) + i \sin(\varphi_{h/k/l} - 2\pi(h/k/l) \cdot (x/y/z)) \right]$$

Computation of the electron density can be achieved by an inverse Fourier transform of the diffraction pattern[78]. For the inverse Fourier transform it is necessary to know both the amplitude and phase of the diffraction waves. The amplitude can be measured however there is no way to measure the phase directly. This results in what is commonly referred to as the 'Phase Problem'.

There are three strategies to overcome the Phase problem: (1) using a model to infer the phases (molecular replacement), or (2) perturbing the diffraction pattern without perturbing the structure; this can be done in two ways, introduce heavy atoms into the crystal or (3) into the protein prior to crystallisation [77].

Molecular replacement uses a closely related structure as a search fragment[80]. This uses phase information from a 3D structure that has already been determined. There are fewer protein folds than sequences and the similarity of two structures correlates well with sequence identity. Ideally molecular replacement can be used if the sequence identity between the model structure and the target protein is above 30%, however it is achievable with lower identities. For molecular replacement the model

structure has to be placed in the right orientation and the right place within the unit cell. The orientation is defined by three rotation angles and the placement in the unit cell relies on translation variables. Traditionally computational techniques are used to find the orientation via firstly the 'rotation function' that uses the intramolecular vectors of the Patterson map, that rely solely on orientation[77]. Once orientation is determined the 'translation function' can establish the position[77, 81]. Least-squares and maximum likelihood are two approaches used for finding a solution with molecular replacement[82]. Maximum likelihood is based on the idea that the best model on the evidence of the data is the one that explains what has been observed with the highest probability[82, 83]. Maximum likelihood rotation targets, using a programme called BEAST, have been shown to be more sensitive to the correct orientation than traditional targets[83]. There are two current types of translation functions - brute force approaches that evaluate every sampled translation point in real space or approaches that generate a Fourier grid calculated by a fast Fourier transform and generate values for all points on the Fourier grid in real space simultaneously[83]. The latter is a lot faster.

PHASER is a programme for phasing protein crystal structures; it can use both molecular replacement and experimental phasing methods however in this work it was only used for molecular replacement. The algorithms that run PHASER are based on maximum likelihood probability theory and multivariate statistics[82]. PHASER uses maximum likelihood phasing algorithms for both the rotation and translation functions in molecular replacement.

Isomorphous replacement is based on the comparison of diffraction patterns from multiple crystals one being the native enzyme and the other(s) being the same enzyme containing one or more strong scattering centres (heavy atoms). It is essential that all crystals have the same unit cell dimensions and symmetry i.e are isomorphous.

Incorporation of a heavy atom changes the scattered intensities. Heavy atoms contribute disproportionately to the overall intensity because an atom contributes to

the scattering intensity in proportion to the square of the number of electrons it contains. Isomorphous replacement relies on measuring the diffraction data from a native crystal and one with heavy atoms incorporated at fixed positions. The heavy atom(s) will be mostly responsible for any difference in the scattering intensity and this can then be used to generate a Patterson map. This map will be relatively simple as it is based on a few heavy atoms. By knowing where the heavy atoms are located in the crystal their contribution to the structure factors can be established. The structure factors are related to the phase and amplitude of the scattered wave. Ideally the structure factors of the derivative protein (F_{PH}) are equal to the structure factors of the native protein (F_P) plus that of the heavy atoms (F_H). If these are represented by vectors, the length of which corresponds to the amplitude and the direction corresponds to the phase, this equation can be represented by a triangle (Figure 1-9) with the length and orientation of one side (F_H) known and the length on the other two sides known, this allows the direction of the those two sides to be calculated, resulting in two possibilities for the phases of F_P. By incorporating data from a crystal containing a heavy atom at a different location within the crystal the combination of all these data results in only one phase value being consistent with all observations[77].

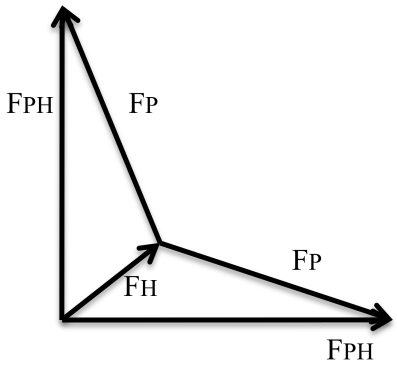


Figure 1-9: Structure factors represented by vectors for the heavy atoms (F_H) the derivatised protein (F_{PH}) and the native protein (F_P) show there are two possible phase values for F_P[77].

Isomorphic replacement relies on three assumptions (1) the positions of the heavy

atoms within the crystal are well defined, (2) the measurement of the amplitude of the structure factors is accurate and (3) the structures of the native and derivatised crystals are isomorphic. This method is strongly affected by experimental errors brought about by failure to satisfy each of these assumptions.

Anomalous dispersion uses a similar principle to isomorphous replacement. X-rays stimulate an oscillation in the electrons of atoms. If the frequency of oscillation is different from the natural frequency of oscillation of the electrons they will oscillate with the same phase. But if it is similar to the natural frequency it can bring the atom to an excited state and the electronic vibration gets out of step, instead of re-radiating in phase with the incident beam the radiated energy has a different phase. The intensity of the coherent scattering is reduced because a proportion of the energy is absorbed to bring about the transition. This is referred to as anomalous scattering[77]. This results in a small shift in both amplitude and phase causing a break down of Friedel's Laws. Friedels law states that in normal diffraction, the intensities of diffraction of orders h and $-h$ are identical. F_h and F_{-h} have the same amplitude, but the sign of the phase angles of the scattered waves are opposite[76]; h and $-h$ are Friedel's pairs. This small shift results in the Friedel's pairs having different amplitudes. The contribution of the heavy atoms to the structure factors can be established if there is a model of their location within the crystal. Phase determination is based on the violation of Friedel's law seen at wavelengths close to an absorption edge. Small changes in the wavelength around the absorption edge of the heavy metal result in large changes in the intensity in the diffraction pattern[78]. At wavelengths very near the energy of an electron transition, anomalous scattering can become a significant fraction of the total scattering[76]. The intensity differences can be sufficient to determine phase information[75, 80]. By collecting multiple datasets at different wavelength near the absorption edge of an element (the heavy atom) phase information analogous to that in multiple isomorphous replacement (MIR) can be obtained[77].

The increasing ease with which synchrotron radiation is accessed has meant

anomalous dispersion experiments are becoming increasingly popular[78]. The use of synchrotron radiation allows the wavelength to be finely tuned so that it is close to an absorption edge of the target atom.

Once diffraction images have been collected there are a range of programs available for processing the data and solving the structure. The programmes used for this project were MOSFLM[84], CCP4[85] and Phenix[86].

The integration of spots recorded on an area detector involves the diffraction intensities and position of the spots being recorded. MOSFLM[84] is the programme used for indexing as well as integrating spots. A program called SCALA[87] from the CCP4 suite is used for data reduction. This is done by using the redundancy of multiple measurements of symmetry-related reflections to put all observations on a common scale[87]. Analysis of the resulting data set allows information about the quality of the data to be gathered. It will give information about the overall quality of the data and areas of poor data that can then be excluded. Once the data is scaled the quality can be assessed and the resolution limit set at an appropriate place. The R_{merge} or R_{sym} indicates the quality of the data, as it is not biased by the improvement in merged intensities[87].

Phenix is an alternative to CCP4 that has been developed as a comprehensive crystallographic software system with an emphasis on automation[86]. The programmes AutoSol and AutoBuild from Phenix were used.

AutoSol carries out automated structure solution based on multiple anomalous dispersion (MAD), single anomalous dispersion (SAD), single isomorphous replacement (SIR) and MIR data sets, or a combination of them. After experimental phasing AutoSol generates the highest quality solution[88]. AutoSol uses a range of programs from the Phenix suite. It uses phenix.xtriage to analyse the data for twinning, non-crystallographic symmetry, unexpected strong or weak reflections or other factors that may complicate the structure determination process. The heavy

atom substructure is found using a hybrid substructure search called HySS. The phases are calculated using PHASER[82] for SAD data, or SOLVE for MAD, MIR and SIR data. Statistical density modification is performed using RESOLVE. Once density modification maps have been generated the model is built into the map and the map-model correlation is used to indicate the best solution[88].

Due to the lack of atomic resolution in X-ray crystallography experiments the data may not allow a sufficiently accurate structure of the protein[89]. The accuracy of the structure must be monitored. To avoid introducing bias into the analysis of the structure accuracy a set of reflections are excluded from the data and used to calculate an R-factor R_{free} (Equation 6) this is an unbiased estimate of the improvement of the model.

Equation 6: Calculation for R_{free}

$$R_{free} = \frac{\sum_{testset} \left\| F_{obs} \right| - \left| F_{calc} \right\|}{\sum_{testset} \left| F_{obs} \right|}$$

When a structure is sufficiently refined it is validated by tools such as SFCHECK[89] or PROCHECK[90]. These programmes collate a number of objective criteria for measuring the quality of the X-ray data and assessing the agreement between the model and those data[89]. Data can be evaluated in their entirety, or alternatively specific residues or areas can be analysed individually. The final structure can also be validated post refinement by the Ramachandran plot. This is an independent evaluation of structure quality. It shows the distribution of (ϕ,ψ) conformation angles from the peptide chain and groups them into favourable, allowed and unallowed regions[76]. It illustrates how the ϕ and ψ angles cluster and reveals outliers[91]. The clustering is based on angles accurately determined from small molecules and allows a few standard deviations freedom. The bond lengths in structures tend to be tightly restrained during refinement and this can lead to distortion in the bond angles[92]. The Ramachandran plot treats residues proline and glycine differently as they have

geometric properties different to that of the other amino acids.

Structural visualisation was done using two programmes COOT[93] and PyMOL[94]. COOT is primarily used as a map-fitting program whereas PyMOL[94] is used as a molecular display program. COOT[93] is a molecular graphics programme used to visualise the X-ray crystallographic data[93]. COOT was built around two main libraries mmdb – a library for handling the macromolecular coordinates and Clipper a library for crystallographic objects and their computation[93]. COOT is used to manually improve the model to better fit the electron density. Residues can be added, deleted and mutated; individual coordinates can be moved and ligands can be imported and placed into the structure. Manual building is monitored by refinement to insure successful improvement of the structure.

1.5.1 Structural Genomics

Structural genomics aims to determine the 3D structures from all genes of a representative group of proteins in an organism[95]. The increasing popularity of structural genomics has led to the development of robotics for increased throughput crystallisation trials. This allows small volumes of pure concentrated protein to be laid down in many different crystallisation conditions leading to detection of suitable conditions being more efficient. The development of complicated, automated software for processing the data from the X-ray diffraction patterns and generating structural information also aids the efficiency of structural genomics. Large scale structural genomics centres rely on these automated systems to get maximum throughput[95].

1.6 Aims of the Project

The rumen acts like a fermentation vat and plays host to a diverse microbial community. These microorganisms breakdown the plant material into volatile fatty acids and other compounds that can then be utilised by the host. *Butyrivibrio*

proteoclasticus is present in significant numbers within the rumen and has an important role being one of the butyrate producing species. A significant proportion of the *B. proteoclasticus* genome is dedicated to genes annotated as being involved in fibre degradation[1]. This consists of a suite of genes encoding for fibre degrading enzymes with a variety of functions necessary for the complete degradation of hemicellulose. By investigating the structure and functional capabilities of these enzymes a detailed understanding of how they work together can be established. A lot of research has been done into similar fibre degrading enzymes, however there remains a lot that is not known. Along with identifying the function of these enzymes, the relationship between their actions and any potential synergy between them will give insight into why this bacterium invests so much energy in maintaining such a diverse suite of enzymes.

Prior to the start of this project, 44 of the genes from the *B. proteoclasticus* genome annotated as being involved in fibre degradation had been cloned and expression of many of them had been attempted. Two of these enzymes had had their 3D structures determined.

The first aim for this PhD research is to investigate the activity profile of the two enzymes whose 3D structures had previously been determined. One was a feruloyl esterase, Est1E, and the other was an acetyl xylan esterase, Est2A. As part of this investigation kinetic studies of these enzymes with model substrates were performed as well as site directed mutagenesis of active site residues in order to gain a more complete understanding of the catalytic mechanism and substrate binding interactions.

The second aim was to gain a better understanding of a set of fibre degrading enzymes produced by *B. proteoclasticus* by subjecting them to crystallisation trials and preliminary functional investigation. The results from this research gave us a better understanding of the range of activities of a selection of these enzymes and has led to the determination of the 3D structure, substrate specificity and mechanism of a

further fibre degradation enzyme from *B. proteoclasticus*, Xsa43E. Based on Blast searches of the sequence Xsa43E was annotated as being a xylosidase/arabinofuranosidase belonging to CAZy family GH43.

The elucidation of the catalytic mechanism of these enzymes, combined with the structural data, will allow a better understanding of their importance to *B. proteoclasticus* as well as their role within the rumen as a whole. It will also provide insight into the relationship between protein structure and specific catalytic function; this is important information for the development of engineered proteins to better fulfil our industrial requirements. It would also allow us to compare the efficiencies of these enzymes with related enzymes from other sources.

Chapter Two: Materials and Methods

2.1 DNA-related Materials and Protocols

2.1.1 General DNA Materials and Methods

2.1.1.1 Primers

Primers for gene amplification and nested PCR were designed and ordered (Invitrogen, USA) previously. Primers for site directed mutagenesis were designed by manual inspection of the gene sequence and incorporation of the alternative codon. All primers were ordered from Invitrogen. All stocks were made up to a concentration of 100 pmol and stored at –20 ºC. Working stocks were made up to 10 pmol with ultrapure water and stored at –20 ºC.

2.1.1.2 Agarose Gels and Electrophoresis

Agarose gels were prepared by dissolving 1.5% (w/v) agarose in 1x TAE buffer (see Appendix I) and 1x Cybersafe Dye (Invitrogen, USA), these were poured into casts and allowed to set for a minimum of 30 minutes before being used. DNA samples were prepared by mixing the DNA with 6x DNA loading dye (see Appendix I) in a 5:1 ratio prior to being loaded into the wells. Gels were run in 1xTAE running buffer for 40 minutes at 100 watts. A 1kb plus ladder (Invitrogen, USA) was also loaded to allow estimation of DNA fragment size. Gels were visualised using a Safe Imager (Invitrogen, USA).

2.1.1.3 Gel Purification of DNA

Once a gel was run and visualised, if necessary, the gel piece with the appropriate sized band was excised from the gel with a sterilised scalpel. DNA was purified from gel pieces using the Qaigen PCR Purification Kit, following the manufacturers instructions.

2.1.1.4 Quantification of DNA

DNA was quantified using a Nanodrop ND-1000 spectrophotometer.

2.1.1.5 Antibiotics

Antibiotic stocks were made up to concentrations as in Table 2.1 and stored in 1 mL aliquots at –20 ºC. Antibiotics were diluted 1:1000 in media.

Table 2-1: Antibiotic Concentrations

Antibiotic	Concentration
Ampicillin	100 $mg.mL^{-1}$ ($MQ\text{-}H_2O$)
Kanamycin	50 $mg.mL^{-1}$ ($MQ\text{-}H_2O$)

2.1.1.6 Glycerol Stocks

Glycerol stocks were made by mixing cells grown over night in LB media (500 µL) with glycerol (500 µL) and flash freezing them in liquid nitrogen then stored at -80 ºC.

2.1.1.7 Sequence Alignments

ClustalW[96] (www.ebi.ac.uk/clustalw/*)* was used to generate multiple sequence alignments.

2.1.1.8 Primer Design

Primers were designed using Vector NTI® (Invitrogen, USA) or Geneious Pro (version 4.8.3, Biomatters Ltd, New Zealand).

2.1.2 Transformation of DNA

2.1.2.1 Preparation of Electrocompetent Cells

An LB-agar plate with no antibiotic was streaked with cells from a glycerol stock of the desired *E. coli* strain (BL21 DE3, DH5α or DL41) stored at -80 °C. The plate was incubated at 37 °C overnight. A colony from the plate and suspended in 5 mL LB media. This LB culture was incubated overnight at 37 °C (200 rpm). The culture was then used to seed 500 mL of LB media. The 500 mL LB culture was grown at 37 °C (180 rpm) until OD_{600} of between 0.5-0.7 was reached. The culture was then chilled on ice for 20 minutes, transferred to pre-cooled centrifuge bottles and centrifuged at 4600 rpm for 15 minutees at 4 °C. The supernatant was discarded and the cells were resuspended in ice cold 10% glycerol (500 mL). The cells were centrifuged at 4600 rpm for 15 minutes at 4 °C. The supernatant was discarded and the cells resuspended in 250 mL of ice cold glycerol. The cells were again centrifuged at 4600 rpm for 15 minutes at 4 °C. The supernatant was discarded and the cells were resuspended in 20mL ice cold 10% glycerol. The cells were centrifuged at 4600 rpm for 15 minutes at 4 °C. The supernatant was discarded and the cells resuspended in 1-2 mL of ice cold 10% glycerol. The cells were then divided into 50 µL aliquots and flash frozen in liquid nitrogen and stored at -80 °C.

2.1.2.2 Electroporation

2 µL of plasmid DNA, either from miniprep, of Gateway BP or LR reaction mixtures, was added to 50 µL of electrocompetent cells on ice. The mixture was placed between the electrodes of a pre-chilled 0.2 cm electroporation cuvette (Bio-Rad Laboratories, USA) and electroporated with a Bio-Rad Gene Pulser™ (Bio-Rad Laboratories) with the settings 2.5 kV, 25 µF, 200 Ω resistance. The electroporated cells were immediately transferred into 1 mL SOC medium and incubated at 37 °C for 30-60 minutes before being plated on LB-agar plates with the appropriate antibiotic. The plates were incubated at 37 °C overnight.

2.1.3 Polymerase Chain Reaction (PCR) Amplification of Target Sequences

To generate material for site directed mutagenesis the open reading frame (ORF) of the gene of interest was amplified from the pDEST17 vector using PCR. A deoxyribonucleotide triphosphate (dNTP) stock was made up to give a final concentration of 10mM for each dNTP. The polymerase enzyme used in all PCR reactions was Platinum *pfx* (Invitrogen, USA).

A typical reaction (50 µL) was made up of the following:

Pfx buffer	7.5 µL
$MgSO_4$ (50 mM)	1 µL
dNTPs (10 mM each)	1.5 µL
Forward Primer (100 µM)	0.5 µL
Reverse Primer (100 µM)	0.5 µL
Template DNA (1 $ng.\mu L^{-1}$)	1 µL
Pfx Polymerase	0.75 µL
MQ water	37.75 µL

Buffer and $MgSO_4$ were supplied by Invitrogen, USA. Reactions of different volumes were based on the same ratios.

The following reaction conditions were used in an Eppendorf PCR machine:

94 °C for 135 sec,
94 °C for 45 sec
55 °C for 45 sec, } 30 cycles
68 °C for 45 sec
68 °C for 300 sec
10 °C hold

2.1.4 PCR for Site Directed Mutagenesis

Site directed mutagenesis was preformed by incorporating custom oligonucleotides

containing the desired mutation into the gene. This was accomplished by generating two sections of the ORF, from the start to the mutation site and from the mutation site to the end. These two sections were generated using the reaction mixture and PCR reaction protocol as set out in section 2.1.2. The two sections of the ORF where then used in another PCR reaction resulting in generation of the whole ORF. The reaction mixture was the same as that set out in section 2.1.2 and the template used was equi-molar amounts of the two sections of the ORF. The PCR reaction protocol was the same as that in section 2.1.2.

AttB sites were incorporated into the ORF by the method outlined in Moreland *et al* (2005)[95].

The mutant genes were incorporated into the same vectors as the original ORF (pDEST17) via a Gateway donor vector (pDONR221) as outlined below. All mutant genes were sequenced to confirm the insertion of the mutation.

2.1.4.1 Gateway Cloning BP reaction

Approximately 15 femtomoles each of PCR product and pDONR221 vector DNA were mixed with 1 µL 5x BP clonase enzyme (Invitrogen, USA), TE buffer pH 8.0 was added to give a final volume of 5 µL. The reaction was incubated at room temperature for 3-18 hours. The reaction was stopped by the addition of 0.5 µL Proteinase K enzyme (Invitrogen, USA) and incubated at 37 °C for 10 minutes. 1 µL of this reaction was transformed into DH5α cells as outlined in section 2.1.2.2, plated onto LB plates with Kanamycin and incubated at 37 °C overnight.

2.1.4.2 Gateway cloning LR reaction

Approximately 75 ng each of entry clone DNA and destination vector DNA (pDEST17) were mixed with 1µL of 5x LR clonase enzyme (Invitrogen, USA), TE buffer pH 8.0 was added to give a final volume of 5 µL. The reaction was incubated at room temperature for 3-18 hours. The reaction was stopped by the addition of

0.5 µL Proteinase K enzyme (Invitrogen, USA) and incubated at 37 °C for 10 minutes. 1 µL of this reaction was transformed into DH5α cells as outlined in section 2.1.2.2, plated onto LB plates with Ampicillin and incubated at 37 °C overnight.

2.1.5 Analysis of Transformed Colonies by PCR

Colonies grown from electroporated cells were checked for the presence of the insert in the plasmid by colony PCR. 1 colony was resuspended in 10 µL of ultrapure water. 1 µL of this suspension was used as template in the recipe set out in section 2.1.2. The reaction conditions for the PCR reaction were the same as in section 2.1.2.

The products were visualised by agarose gel electrophoresis as described in section 2.1.1.2

2.1.6 Purification of Plasmid DNA (Mini preps)

Plasmid DNA extraction was performed on an overnight culture grown in 3 mL of LB with the appropriate antibiotic at 37 ºC, using the Qiagen Miniprep kit and following the manufacturers instructions.

2.1.7 DNA Sequencing

DNA sequencing was performed by the Waikato DNA Sequencing Facility at University of Waikato, School of Science and Technology, Hamilton.

2.2 Protein Expression and Purification Materials and Protocols

2.2.1 SDS-Poly Acrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE gels consisted of a 4% acrylamide stacking layer on top of either a 10% or 12% acrylamide resolving layer. SDS-PAGE gel recipes can be found in

Appendix I.

Samples were mixed with 4x SDS loading dye (Appendix I) in a ratio of 3:1 and heated at 96 ºC for 5 minutes. The gels were run in SDS-running buffer (Appendix I) at 15 mA or 70 V until the dye front passed the stacking layer, then at 20-25 mA or 160 V until the dye front reached the bottom of the gel. Each gel also had a lane in which the Precision plus protein ladder (Biorad) was run.

Protein was visualised on the gel by a 4-step Fairbanks staining protocol (see Appendix I).

2.2.2 Small Scale Expression Tests

Small scale expression tests were performed specific to the target protein as outlined in the methods for the appropriate chapters.

2.2.2.1 Nickel Pull Down Experiments

20 µL of Ni Sepharose High Performance beads (GE Healthcare, USA) were transferred to a 1.5 mL eppendorf tube, 1 mL of phosphate buffer was added, the beads left to settle then removed. The phosphate buffer wash step was repeated three times. The supernatant from the expression culture was applied to the washed beads and incubated at room temperature for 30 minutes. The tube was shaken gently every five minutes. The supernatant was removed and the phosphate wash step repeated three times. On removal of the phosphate buffer, 20 µL of 4x SDS loading dye was added to the beads and they were incubated at 90 °C for five minutes. The samples were then load onto an SDS-PAGE gel.

2.2.3 Large Scale Expression

2.2.3.1 Expression of protein using IPTG induction

A colony was taken from an LB-Agar plate with the appropriate antibiotic that had

been streaked out previously from glycerol stocks, and added to 5 mL of LB media with the appropriate antibiotic. The culture was grown overnight and used to seed a larger culture (either 500 mL or 1 L) with the appropriate antibiotic in a 2 L baffled conical flask. The large LB culture was grown at 37 °C until an OD_{600} between 0.4 and 0.6. IPTG was added to the culture to give a final concentration of 1 mM, then it was transferred to 28 °C for between 16-20 hours.

2.2.3.2 Expression of Seleno-Methionine Protein

Seleno-methionine protein was grown by transforming the relevant plasmid into DL41(DE3) *E. coli* cells, and plating them onto LB agar plates with the appropriate antibiotic and incubating them overnight at 37 ºC. A colony was transferred from an LB-agar plate with the appropriate antibiotic to 5 mL of PA-0.5G (See Appendix I) with the appropriate antibiotic. This culture was then transferred to 400 mL of PASM-5052 autoinduction media (See Appendix I) and incubated at 28 ºC for 24 hours.

2.2.3.3 Cell Harvesting for large-Scale Expression Cultures

Cells from large scale expression cultures were harvested by centrifugation, cultures were transferred into 750 mL centrifuge bottles and centrifuged at 4600 rpm for 45 minutes. The supernatant was discarded and the cell pellet was resuspended in 20-30 mL of lysis buffer (50 mM Tris/HCl, pH 8, 150 mM NaCl), transferred to a 50 mL Falcon tube and stored at –20 °C until required.

2.2.3.4 Lysis of large-Scale Expression Cultures

One Complete mini EDTA-free (Roche) tablet was added to the cell suspension. The cells were lysed by sonication using a Misonix somicator (USA) at level 7. 15 second bursts were followed by 15 second breaks for a total time of 2 minutes or until sufficient lysis.

The lysed suspension was returned to a 50 mL falcon tube and centrifuged at 13000xg for 30 minutes. The cell pellet was discarded and the supernatant was filtered to 0.2 μm using a Minisart filter.

2.2.4 Fast Performance Liquid Chromatography (FPLC)

FPLC was used to purify proteins from cell lysate. FPLC was performed using either an ÄKTA Basic™ or ÄKTA Prime™ FPLC system. The appropriate column for purification was connected to the FPLC system to be run. Elution of protein was monitored by absorbance at 280 nm.

2.2.5 Immobilised Metal Affinity Chromatography (IMAC)

2.2.5.1 Preparation of IMAC column

IMAC was performed using either a 5 mL HiTrap™ Chelating HP column (GE Healthcare) or a 5 mL HiTrap™ FF column (GE Healthcare). Prior to sample loading the columns were loaded with Ni^{2+} ions using 5 mL 100 mM $NiCl_2$ and then washed with 10 mL lysis buffer. After use the columns were stripped of Ni^{2+} using 100 mM EDTA (5 mL), then flushed with 10 mL MQ water and 5 mL 20% ethanol for storage.

2.2.5.2 Purification of His-tagged proteins by IMAC

The filtered supernatant from lysed expression cultures (section 2.2.3.4) was loaded onto an IMAC column pre-equilibrated with lysis buffer at a rate of up to 5 $mL.min^{-1}$. The column was then washed with 10 mL lysis buffer at a flow rate of up to 5 $mL.min^{-1}$, followed by 20 mL of a mix lysis buffer and elution buffer (50 mM Tris/HCl, pH 8, 150 mM NaCl, 500 mM Imidazole) at a ratio of 49:1. The bound proteins were then eluted by running a gradient from 2% to 100% elution buffer over 98 mL at a rate of 1 $mL.min^{-1}$. Elution of protein was monitored by absorbance at 280 nm. The fractions containing protein were analysed by SDS-PAGE (section 2.2.1).

2.2.6 Size Exclusion Chromatography (SEC)

SEC was used to further purify proteins after IMAC purification. SEC was performed using a HiLoad™ 16/60 Superdex™ 200pg (S200 16/60) column. The protein sample was concentrated down to <5 mL and filtered with a 0.2 μm filter. The protein sample was loaded onto the column pre-equilibrated into running buffer (50 mM Tris/HCl, pH 8, 150 mM NaCl) and eluted with 1 column volume (124 mL). SE columns were run at 0.5 $mL.min^{-1}$ with 2 ml fractions collected, fraction collection started after 0.25 column volumes. Elution of protein from the column was monitored by absorbance at 280 nm. The fractions containing the desired proteins were analysed by SDS-PAGE (section 2.2.1). Calibration of the S200 16/60 was carried out following manufacturers instructions. The resulting calibration curve with equation can be seen in Figure 2-1.

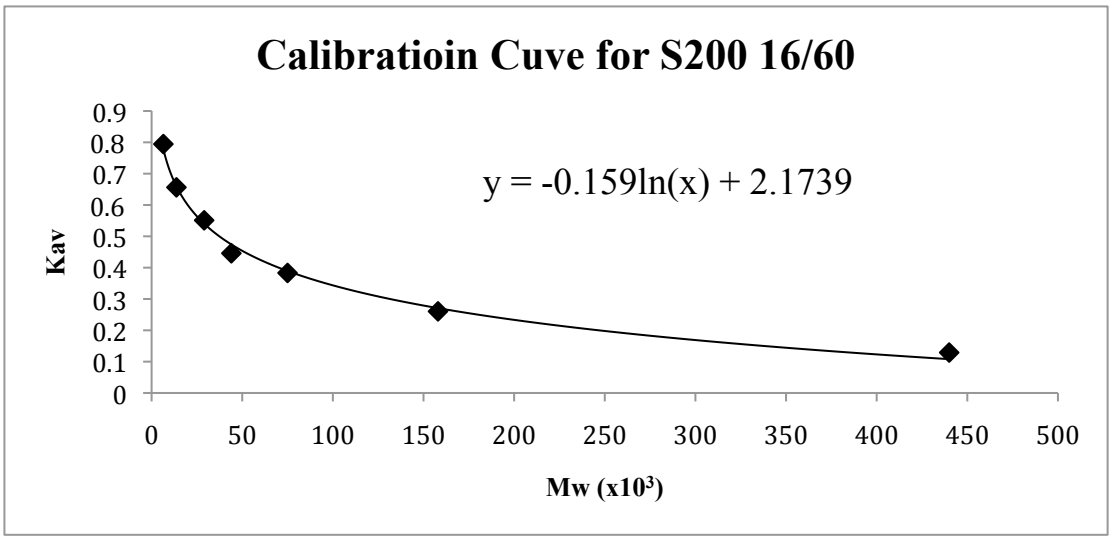


Figure 2-1: Calibration curve of S200 16/60. Calibration done using both the low and high molecular weight kits provided by GE and following manufacturers instructions.

2.2.7 Concentration of Purified Protein

Protein samples were concentrated using 20 mL, 2 mL and 600 μL Vivaspin concentrators with molecular weight limits of 5, 10 or 30 kDa. The concentrators

were prepared by adding 50% glycerol and spun at the maximum speed of the concentrator for 5 minutes in a Hereus Multifuge 3 S-R centrifuge. Then the glycerol was removed and the protein sample was added to the concentrator and spun at 3500 pm (20 mL), 4600 rpm (2 mL) or 10,000 rpm (600 µl) at 4 ºC until the desired concentration or volume was reached.

2.2.8 Protein Concentration Determination

Protein concentration was determined by a Nanodrop ND-1000 Spectrophotometer which gives a reading of concentration in $mg.mL^{-1}$ based on 1 Abs unit = 1 $mg.mL^{-1}$. These values were then corrected using the formula $A=\varepsilon.c.l$ where A= Absorbance at 280 nm, ε= the theoretical extinction coefficient of the protein, c= concentration, and l=pathlength. A theoretical extinction coefficient for each protein was calculated based on the amino acid composition by submitting the sequence to the online tool Portparam[97].

2.3 General Protein Crystallisation Materials and Protocols

2.3.1 Initial Crystallisation Trials

A liquid handling robotics system was used to prepare the crystallisation trials in an 18 °C humidity controlled environment. Crystallisation trials used the sitting-drop vapour diffusion method. The initial crystallisation trials contained 480 different conditions based on 5 in house screens in 96-well Intelli-plates (Hampton Research, USA). A Multiprobe® II HTEx robot (PerkinElmer, USA) was used to dispense 75 µL of each condition into wells of an Intelli-plate. A honey bee sitting drop 8-channel liquid handling robot (Cartesian™ Dispensing Systems, USA) set up sitting drops containing 100 nL of protein with 100 nL mother liquor taken from the well solution. All crystallisation trials were laid down at the School of Biological Sciences at the University of Auckland.

2.3.2 Crystal Screens and Crystal Optimisation

When positive results were seen in the initial crystallisation trials fine screening was carried out to optimise the conditions for crystallisation. The fine screens were typically carried out using hanging drop vapour diffusion in 24-well VDX^{TM} Plates (Hampton Research, USA). The top of each well was lined with grease and 500 µL of mother liquor was transferred into the well. 2 µL of mother liquor from the well was mixed with 2 µL of protein on a 22 mm square unsiliconised glass cover. The cover slip was then inverted and placed on top off the pre-greased well. Variations in the fine screens were using the sitting drop method in Micro-Bridges (Hampton Research, USA), or using siliconised glass cover slides (Hampton Research, USA) or by reducing the drop size to 1 µL + 1 µL.

2.3.3 Cocrystallisation Experiments

Crystals were grown in mother liquor containing a set concentration of a substrate or other compound of interest. Crystallisation was performed as set out in section 2.3.2 but protein was mixed directly with mother liquor containing the desired compound on the cover slip.

2.4 Crystal Preparation and X-ray Data Collection Methods

2.4.1 X-Ray Data Collection

2.4.1.1 Home Source Data Collection

Home-source data was collected at the School of Biological Sciences, The University of Auckland using CuKα radiation (λ=1.5418Å). The X-rays were generated by a Micro-Max^{TM}-007HF generator (Rigaku, Japan) and images were collected using a Mar345 detector. The crystals were kept at 110 K by a stream of cooled nitrogen gas. The home source was used to screen crystals prior to sending to the synchrotron, not to collect full data sets from crystals.

2.4.1.2 Synchrotron Data Collection

Multiple wavelength anomalous dispersion (MAD) and Single wavelength anomalous dispersion (SAD) data were collected using the rotation method from seleno-methionine substituted protein crystals at Beamline 11-1 of the Stanford Synchrotron Radiation Laboratory (SSRL), Stanford University, USA equipped with a MarUSA MarMosaic -325 CCD detector. The beamline was operated over the Internet using the SSRL's remote experiment set up. A preliminary X-ray fluorescence scan was used to determine the best wavelengths to use for data collection. Data for native protein crystals were collected on beamline MX1, equipped with a ADSC Quantum 210r detector, at the Australian Synchrotron.

2.4.2 Cryo-protectant Testing

Cryo-protectants consisted of mother liquor containing increasing amounts of glycerol. Cryo-protectants were tested by mounting a drop on a nylon loop fixed in a copper base and flash freezing it in liquid nitrogen before mounting it on the home source X-ray beam (section 2.4.1.1). Cryo-protectants were shot for 1 minute and the resulting image was analysed for presence of ice-rings. The cryo-protectant chosen was the one that contained the least amount of glycerol but gave no ice ring.

2.4.3 Mounting and Flash-Freezing Crystals

Crystals were removed from the drop using a nylon monofilament cryo-loop (Hampton Research, USA) fixed in a cooper cryo-cap with a magnetic base (Hampton Research, USA) and transferred through a number of drops containing increasing amounts of glycerol until reaching the sufficient level for cryo-protection.

The crystal remained in the each drop for between 10 seconds and 2 minutes. The crystal was then removed with a cryo-loop and flash frozen by immersion in liquid nitrogen.

2.5 General Data Processing and Structure Solution Methods

2.5.1 General Data Processing and Collection Strategy

2.5.1.1 Integration of Data

The programme MOSFLM[84] was used to integrate data sets. In MOSFLM[84] the automated spot finder was used to find spots, spots were added manually if necessary. The diffraction patterns were indexed using the autoindexing function in MOSFLM[84] then the mosaicity of the crystal was estimated. Prior to integration the cell parameters were refined using two sets of four consecutive images 90° apart. Integration was done in MOSFLM[84] in one batch unless it was necessary to process the data in smaller batches.

2.5.1.2 Scaling and Merging of Reflections

Scaling and merging of integrated reflections was done using SCALA[87] from the CCP4[85] suite of programmes via the CCP4i interface. In all SCALA[87] runs 5% of reflections were marked as the R_{free} set to ensure validation during structural refinement. For data sets with anomalous differences the data were processed with the anomalous flag set to keep the Bijvoet pairs (I+ and I-) separate during merging. MAD datasets were scaled separately.

2.5.1.3 Matthews Coefficient

The Matthews Coefficient was determined using the CCP4 programme MATTHEWS_COEF[85].

2.5.2 Structure Solutions Using Molecular Replacement

Structural solutions using molecular replacement methods was performed using PHASER[82] from the CCP4 suite of programmes.

2.5.3 Structure Solutions Using Anomalous Dispersion

Experimental Phase information was obtained using the anomalous dispersion of the element selenium, from crystals grown from seleno-methionine protein.

2.5.3.1 Structure Solution Using Multiple Anomalous Dispersion (MAD)

Structural solutions using MAD data was performed using the Autosol wizard[88] in the Phenix[86] programme.

2.5.3.2 Structure Solution Using Single Anomalous Dispersion (SAD)

Structural solutions using SAD data was performed using the Autosol wizard[88] from the Phenix programme[86].

2.5.4 Model Building and Refinement

2.5.4.1 Automated Model Building

Automated model building was done using the PHENIX AutoBuild wizard[98], which uses RESOLVE[99] for iterative model building and phenix.refine for refinement of the model.

2.5.4.2 Manual Building Using Coot

Manual model building was carried out in Coot[93]. The model was built into SigmaA-weighted $2|F_o|-|F_C|$ electron density maps contoured at 1.0 σ. Model building was also helped by SigmaA-weighted $|F_o|-|F_C|$ maps contoured to ±3.0 σ.

2.5.4.3 Model Refinement

Model refinement was normally performed using Refmac5 in the CCP4 suite of programs[85]. During initial manual building phenix.refine was used for model refinement to incorporate one round of simulated annealing with a starting

temperature of 5000 K, reducing in steps of 100 K down to a final temperature of 300 K. Model refinement was performed regularly during manual building to ensure manual alterations were true. When refinement was close to finished, waters were added using Refmac5. The added waters were examined in COOT[93] and manually checked for appropriate symmetrical electron density and hydrogen bonding interactions with the protein.

2.5.5 Model Validity

Validity of the final structure was checked Using PROCHECK[90] or SFCHECK[89] from the CCP4 suite.

2.5.6 Structure Analysis

Structure models were visualised using COOT[93] or PyMOL[94]. Structure diagrams presented in this thesis were created using PyMOL[94].

2.6 General Activity Data Collection Methods

Activity assays were done according to protocols specific to the substrate and will be outlined in the appropriate chapters.

2.6.1 Spectrophotometric Assays

Kinetic data for *p*-nitrophenol releasing substrates was collected using either a Heλios γ UV spectrophotometer (Thermospectronics, USA) in a 1 cm cuvette with a total volume of 1 mL; the reaction was followed in half-second intervals for between 15 seconds and 1 minute. Alternatively kinetic reactions were monitored in a BMG *FLUOstar Optima* F plate reader measuring absorbance in a 96 well plate with a total volume of 200 μL, readings were taken every minute for each well for between 30 and 80 minutes. Substrates, buffers and wavelengths used were specific to the reaction and are described in the appropriate chapters.

2.6.2 Thin Layer Chromatography (TLC)

Assays involving the degradation of sugar oligosaccharides were visualised using TLC. Up to 2 µL was spotted onto the base of a TLC plate with appropriate standards, the plate was run in a solvent system of chloroform: acetic acid: water (6:7:1). The plate was run until the solvent was within 1 cm of the top, the plate was then removed and allowed to dry. The plate was visualised by spraying with a solution of 5% sulphuric acid in ethanol and incubated at 110 °C for a minimum of 10 minutes.

2.7 Methods Specific to Chapter Three

2.7.1 Activity Assays

2.7.1.1 Kinetic Assay with p-Nitrophenyl acetate

p-Nitrophenyl acetate assay was done using 100 mM *p*-nitrophenyl acetate as a substrate made up in 20% ethanol, 100 mM Sodium Phosphate buffer, pH 7.2. Varying amounts of substrate were added to a cuvette and made up to 1 mL with buffer prewarmed to 37 °C. Enzyme was added to initiate the reaction. The reaction was monitored at 400 nm using a UV spectrophotometer as outlined in section 2.6.1. Measurements were taken at half-second intervals for 30 seconds. Each concentration of substrate was done in triplicate as well as a blank containing no enzyme. Rates were corrected for non-enzymatic degradation.

2.7.1.2 Kinetic assay with p-Nitrophenyl butyrate

p-Nitrophenyl butyrate assay was done using 100 mM *p*-nitrophenyl butyrate as a substrate made up in 20% ethanol and 100 mM Sodium Phosphate buffer, pH 7.2. The assays were preformed as outlined in Section 2.7.1.1.

2.7.1.3 Activity determination with other model substrates

p-Nitrophenyl laurate and *p*-Nitrophenyl palmitate were also used as substrates. Both were made up in 50% ethanol due to solubility issues. Enzyme was added to a sample of 10 mM substrate in 100 mM Sodium Phosphate buffer, pH 7.2, and an initial reading at 400 nm was taken, a reading of a second sample with no enzyme, the blank, was also taken. The samples were then incubated at 37 °C for 15 minutes, then a second reading at 400 nm was taken. Activity was indicated by an increase in absorbance above that of the blank.

2.7.2 Site Directed Mutagenesis of Est1E and Est2A

Site directed mutagenesis was preformed on the Est1E or Est2A gene through incorporation of custom oligonucleotides containing the desired mutation. The primers were designed to mutate Est1E Ser-105 to Ala, Est2A Ser-142 to Ala and Est2A His-351 to Ala, these are shown in Table 2-2.

Table 2-2: Primers for the mutation of Est1E and Est2A gene.

Est1E S105A Fwd	ATCTACATGGCAGGACACGCGCAGGGCGGACTCTCTGTA
Est1E S105A Rev	TACAGAGAGTCCGCCCTGCGCGTGTCCTGCCATGTAGAT
Est2A S142A Fwd	GAATTTATCGGAGACGCCATAACATCAGGAGAA
Est2A S142A Rev	TTCTCCTGATGTTATGGCGTCTCCGATAAATTC
Est2A H351A Fwd	TTCGGATCTCATATGGCCCCAGGACCCAAATCA
Est2A H351A Rev	TGATTTGGGTCCTGGGGCCATATGAGATCCGAA

The primers were designed using Vector NTI® and ordered from Invitrogen (USA). The custom oligonucleatide was incorporated by PCR as outlined in section 2.1.4.

The mutant genes were incorporated into the same vectors as the original as described in section 2.1.4, they were then sequenced to confirm the insertion of the mutation.

2.7.2.1 Expression and Purification ofEst1E and Est2A mutants

Mutant enzymes were expressed as outlined in section 2.2.3 and purified in two steps

as outlined in 2.2.5 and 2.2.6 respectively.

2.7.2.2 Activity Assay of Est1E and Est2A mutants

Activity of Est1E and Est2A were established using *p*-nitrophenyl acetate. 10 µL of enzyme (1 $mg.mL^{-1}$) was added to a total of 1 mL of 100mM sodium phosphate buffer containing a final concentration of 15 mM substrate. Blanks were run as well with lysis buffer in place of enzyme to account for non-enzymatic decay of substrate. Absorbance of samples were monitored continuously for five minutes using a Heλios γ UV spectrophotometer (Thermospectronics, USA) at 400 nM.

2.7.2.3 Crystallisation trials of Est1E S105A

Crystallisation trials of Est1E S105A were performed as outlined in section 2.3.1 with purified protein at 25 $mg.mL^{-1}$. Fine screens were done as described in section 2.3.2 around the crystallisation conditions of the wild type protein, which was 17-22% PEG 3350, 0.3-0.5 M NaH_2PO_4. The mother liquor also contained ferulic arabinose.

2.7.2.4 Data collection and processing for Est1E S105A

Data was collected from a single crystal using synchrotron radiation at the Australian synchrotron. Images were collected with 1° oscillations and a one second exposure time. A total of 180 images were collected. The data was integrated using MOSFLM[84] (section 2.5.1.1) into spacegroup $P22_12_1$. The data was then scaled and merged using SCALA[87] (section 2.5.1.2), and molecular replacement was done using PHASER[82] (section 2.5.2) with the Est1E chain A as a model, PHASER[82] was asked to find two copies of the model. Manual model building refinement and structural validation was done as outlined in section 2.5.4, using COOT[93], Refmac5 and PROCHECK.

2.7.2.5 Crystallisation of Est2A S142A and Est1D H351A

Est2A S142A and Est2A H351A were laid down in fine screens around the condition that gave crystals of the native enzyme (8% (w/v) PEG 4000, 0.1 M sodium acetate pH 4.6). The mother liquor also contained 1% (w/v) arabinotetraacetate (Sigma).

2.7.2.6 Data collection and processing for Est2A H351A

Data was collected from a single crystal using synchrotron radiation at the Australian synchrotron. Each image had a 1° oscillation and a 1 second exposure time. A total of 180 images were collected. The data was integrated using MOSFLM[84] (section 2.5.1.1) into space group P_1. The data was then scaled and merged using SCALA[87] (section 2.5.1.2), and molecular replacement was done using PHASER[82] (section 2.5.2) with the Est2A structure as a model. Manual model building was done in COOT[93], refinement was done using Refmac5 and the final solution was validated using PROCHECK (section 2.5.4).

2.8 Methods Specific to Chapter Four

2.8.1 Gas Chromatography Mass Spectrometry (GCMS) analysis of Enzymatic release of sugars from Natural Substrates

2.8.1.1 Natural Substrate Assays

Substrates (Hemicellulose, Cellulose and Ryegrass from Grasslands AgResearch and Arabinoxylan from Megazymes International Ltd, Ireland) were made up in 100 mM phosphate buffer pH 7.2 to a final concentration of 1 $g.L^{-1}$. 900 µL of substrate was mixed with 100 µL of enzyme and the solution incubated at 40 °C for 3 hours. Due to the insolubility of the majority of the substrate tubes were inverted regularly to aid mixing. Samples were then centrifuged (10,000 xg, 5 minutes) and stored at -20 °C until required.

2.8.1.2 Assay Derivatisation for GCMS Analysis

200 µL aliquots of sample were taken, 10 µL of Inisitol (25 mM) was added as an internal standard. The samples were then dried under vacuum. In a fumehood, 80 µL of methoxyamine HCl solution (0.5 g in 125 mL pyridine) was added to each sample, the sample was heated gently for three minutes. 80 µL of *N*-methyl-*N*(trimethylsilyl)-trifluoroacetic acid (ms-TMS) was added to each sample then it was again heated gently for three minutes. Samples were then ready for GCMS analysis.

2.8.1.3 GCMS Analysis of Samples

Samples were analysed by GCMS at the Mass Spectrometry suite Grasslands Research Centre in Palmerston North.

2.8.2 Azo-CMCellulose (Azo-CMC) Assay

Azo-CMCellulose assays were performed using the substrate Azo-CM-Cellulose (catalogue number S-ACMCL, Megazymes International Ltd, Ireland). Assays were performed as outlined in the manufacturers instructions.

2.8.3 Azo-wheat arabinoxylan (Azo-WAX) Assay

Azo-WAX assays were performed using the Xylanase (Azo-Wax) kit (catalogue number K-AZOWAX, Megazymes International Ltd, Ireland). Assays were performed as outlined in the manufacturers instructions.

2.8.4 Expression trials of Xyn10A

20 mL cultures of *E. coli* containing the Xyn10A plasmid were grown to an OD_{600} of 0.4 before being treated with IPTG at a final concentration of 1 mM. Four cultures were grown for 4 hours at 28 °C and another four at 37 °C. Each sample was spun down with centrifugation for 30 minutes at 4,600 rpm. One cell pellet grown at each

temperature was resuspended in 1 mL of 100 mM Tris/HCl buffers of pH 6, 7, 8 or 9. The cells were then lysed and spun down with centrifugation at 15,000 xg for 25 minutes. Each sample was analysed with a nickel pull down experiment as outlined in section 2.2.2.1. 25 µL samples were taken of the whole cell (WC), the cell pellet (CP), the supernatant (SN) and the nickel pull down (PD) and visualised on a 10% SDS-PAGE gel (2.2.1).

2.8.5 Construction of Xyn10A-16 pDEST17 Vector

The pDEST17 vector containing the Xyn10A gene was used as a template. A new primer was designed for the start of the gene removing the first 16 amino acids. This was used with the original reverse primer were used to amplify the target sequence. This was then incorporated back into a pDEST17 vector as outlined in section 2.1.4.

2.8.6 Expression Trials of Xyn10A-16

Expression trials for Xyn10A-16 were performed as set out in section 2.8.1. With the exception that the initial 20 mL culture contained *E. coli* containing the Xyn10A-16 plasmid.

2.8.7 Expression trials of Mxy10-43A

20 mL cultures of *E. coli* containing the Mxy10-43A plasmid were grown to an OD_{600} of 0.4 before being treated with IPTG at a final concentration of 1 mM. Cultures were grown for 4 hours at 37 °C. Each sample was spun down with centrifugation for 30 minutes at 4,600 rpm. A sample of the media (LB) was taken from each to check for secreted protein. The cells from each sample were resuspended in 1 mL of a different 100 mM Tric/HCl buffer varying in pH from 6 to 9 then lysed and spun down with centrifugation at 15,000 xg for 25 minutes. Each sample was analysed with a nickel pull down experiment as outlined in section 2.2.2.1. 25 µL samples were taken of the whole cell (WC), the cell pellet (CP), the supernatant (SN) and the nickel pull down (PD) and visualised on a 10% SDS-PAGE

gel (2.2.1).

2.8.8 Small Scale Expression Testing of Cel5C

The expression of soluble Cel5C was tested by growing cells at 28 °C and 37 °C to identify which temperature gave the best expression of soluble protein. Two 20 mL cultures were grown for 1 hour at 37 °C before the addition of IPTG (final concentration 1 mM) then transferred to the appropriate temperature. The cells were then grown for a further 3 hours at that temperature before being analysed with a nickel pull down experiment as outlined in section 2.2.2.1. 25 µL samples were taken of the whole cell (WC), the cell pellet (CP), the supernatant (SN) and the nickel pull down (PD) and visualised on a 12% SDS-PAGE gel (2.2.1).

2.9 Methods Specific to Chapter Five

2.9.1 Structural Characterisation of Xsa43E

2.9.1.1 Seleno-methionine incorporated Crystals

Seleno-methionine incorporated protein was expressed outlined in section 2.2.3.2. The protein was purified as set out in sections 2.2.5 and 2.2.6 except all buffers contained 2 mM β-mercaptoethanol. Seleno-methionine incorporation was confirmed by sending the protein for MALDI-TOF MS analysis (Chemistry Department, University of Waikato).

2.9.1.2 X-Ray Data Collection

Native Crystals

Native Crystals were sent to the Stanford Synchrotron Radiation Laboratory (SSRL) at Stanford University USA for data collection. One data set was collected from one crystal. A total of 360 images were collected at 1° oscillations with the wavelength set at 0.9795 Å.

Seleno-methionine Crystals

13 Se-met crystals were sent to Dr. Graeme Card at the SSRL. Data were collected from the best crystal, three datasets for multiple wavelength anomalous dispersion (MAD) and two datasets for single wavelength anomalous dispersion (SAD). An initial X-ray fluorescence scan indicated that the best wavelengths to collect data for MAD were 0.9783 nm (Peak), 0.9184 nm (Remote) and 0.9791 nm (Inflection). A total of 180° were collected at 1° oscillations. To maintain the integrity of the anomalous signal against radiation damage, the data for each wavelength was collected using the inverse beam method in batches of 20°. For the SAD datasets the wavelength used was 0.9795 nm. Two datasets were collected; the first dataset was collected to maximise the reflections in the lower resolution areas and the second was collected to maximise the reflections at the higher resolution limits.

2.9.1.3 Data processing

Data was integrated using MOSFLM[84] into the space group $P2_12_12_1$, then scaled and merged using SCALA[87] from the CCP4 suite as described in section 2.5.1.1 and 2.5.1.2 respectively. Each MAD dataset was integrated, scaled and merged individually. The two SAD datasets were integrated separately; data between 56-1.7 Å were taken from the low resolution data set and merged with data from between 1.7-1.3 Å taken from the high resolution data set. The two datasets were then combined prior to being run through SCALA[87].

2.9.1.4 Phase determination and improvement

MAD

The structure of Xsa43E was determined by MAD methods from Se-Met substituted Xsa43E using the program Phenix AutoSol at a resolution of 1.8 Å. Density modification, phase extension and initial automatic model building was carried out

using Phenix AutoSol to a final resolution of 1.8 Å.

SAD

A high-resolution structure was determined by SAD methods from the Se-met substituted Xsa43E. The images were scaled using the CCP4 programme SCALA[87]. The structure was then solved using Phenix AutoSol to a resolution of 1.3 Å, followed by AutoBuild as this data had the highest resolution therefore would generate the most accurate structure.

2.9.1.5 Refinement and Further model building

The output model from Phenix AutoBuild was put through an initial round of refinement in the programme Refmac5 from CCP4. The scaled mtz file from the SAD data collection was used for refinement. The output from the initial round of refinement was used as a starting point for manual building. Manual model building was done in COOT[93] and progress was checked with restrained refinement in Refmac5 from the CCP4 interface as set out in section 2.5.4.3. Electron density maps used for model building were SigmaA weighted $2|F_O|-|F_C|$ (contoured at 1.0 sigma) and $|F_O|-|F_C|$ (contoured at ± 3.0 sigma). Model building required inserting alternative conformations for some residues.

2.9.1.6 Model Completeness and Quality

When the model was sufficiently refined waters were added in Refmac5. The waters were manually checked to insure they were in sensible positions and were justified by sufficient electron density. Model completeness and quality were evaluated by running PROCHECK[90] in CCP4.

2.9.1.7 Crystallisation and Xray Data Collection for Xsa43E Mutants

Xsa43E H258A

Xsa43E protein was laid down at varying concentrations in fine screens based around the conditions 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2 that gave crystals of Xsa43E. Four crystals were sent to the Australian synchrotron for data collection. Data was collected from one crystal at a wavelength of 0.954 nm. A total of 180° were collected with images having a 1° oscillation and a 1 second exposure time. The data was processed in the same way as for the native enzyme (section 2.9.1). Images were integrated using MOSFLM[84] and then scaled and merged in SCALA[87] from CCP4. The Xsa43E structure was used as a model in PHASER[82] in CCP4 to find a molecular replacement solution. Manual model building in COOT[93] was used to change the mutated residue and any other necessary alterations. Validation was done with PROCHECK[90] from the CCP4 suite.

Xsa43E H258Q

Xsa43E H258Q protein was laid down at varying concentrations in fine screens based around the conditions 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2 that gave crystals of Xsa43E. Three crystals were sent to the Australian synchrotron for data collection. None of the crystals gave diffraction.

2.9.2 Functional Characterisation of Xsa43E

2.9.2.1 Substrate Specificity Assays

The Activity of Xsa43E was tested on the following range of substrates – arabinobiose, arabinohexaose, xylose oligosaccharides with a degree of polymerisation from 2 to 6, hemicellulose, ryegrass and arabinoxylan. All were from Megazymes International Limited other than the ryegrass and hemicellulose, which were from AgResearch Grasslands. Xsa43E was incubated with the substrate for a set amount of time at 37 °C in 20 mM sodium phosphate buffer pH 7.2. The mixture was

analysed by TLC (section 2.6.2). Specific activity for release of arabinose from arabinobiose and arabinoxylan was done using the lactose/galactose kit (Megazymes International Ltd) following the manufacturers instructions.

2.9.2.2 Model Substrate Assays

Kinetic data was obtained using 50 mM *p*-nitrophenyl-α-L-arabinofuranoside or 40 mM *p*-nitrophenyl-ß-D-xylopyranoside as substrates. Varying amounts of substrate were used with 1 µL of enzyme in a total volume of 200 µl made up with 20 mM sodium phosphate buffer pH 7.2. All amounts of substrate were done in triplicate and with a blank containing lysis buffer (20 mM Tris/HCl, 150 mM NaCl) instead of enzyme. The reactions were initiated by addition of enzyme and the absorbance monitored at one minute intervals at 405 nm for 30 minutes in a BMG *FLUOstar Optima* F plate reader (section 2.6.1). The data was corrected for non-enzymatic decay of the substrate and processed using Prism5 GraphPad software. Absorbance readings were converted to concentration of *p*-nitrophenol by a standard curve created specifically for this assay.

2.9.2.3 pH profile

pH profiles were perfomed with 20 µL of 50 mM *p*-nitrophenyl-α-L-arabinofuranoside, 175 µL of the appropriate buffer and 5 µL of enzyme. Each pH was tested in triplicate with a blank that contained lysis buffer instead of enzyme. The pH buffers used were sodium acetate - pH 4, 4.6 and 5.2, sodium phosphate – pH 5.7, 6.2, 6.8, 7.3,7.8 and 8.0 and glycine buffer - pH 8.8, 9.9 and 10.3. The reaction was initiated by addition of the enzyme and the change in absorbance at 405 nm recorded as in section 2.9.2.2. Data was processed using Microsoft Excel.

2.9.2.4 Temperature Profile

Temperature profiles were carried out using 50 µL of the substrate 50 mM

p-nitrophenyl-α-L-arabinofuranoside and 5 µL of enzyme in a total volume of 1 mL made up with 20 mM sodium phosphate buffer pH 7.2. Each temperature measurement was performed in triplicate with a blank that contained lysis buffer instead of enzyme. The buffer and substrate were allowed to equilibrate to temperature before addition of the enzyme. Initial readings were taken and the reaction mixture incubated at the appropriate temperature for 30 minutes before the final reading was taken. Data were processed as in section 2.9.2.1.

2.9.2.5 Assay for the Effect of Metal ions

Metal ion were added to Xsa43E at 1 mM and 10 mM (final concentration) and Xsa43E pretreated with EDTA with 1mM (final concentration) of various metal salts were added to the enzyme. The metal salts used were $CaCl_2$, $CoCl_2$, $CuCl_2$, $FeSO_4$, $MgCl_2$, $MnCl_2$, $NiCl_2$, $Sr(NO_3)_2$ and $ZnCl_2$. Aliquots of enzyme from the same purification were treated with one of the salt solutions and a control sample was treated with milli Q water. Enzyme used for the EDTA/metal ion test was pretreaed with 10 mM EDTA for four hours before being dialysed into the original buffer to remove the EDTA. The enzyme was left for 24 hour before the activity was tested on the model substrate *p*-nitrophenyl-α-L-arabinofuranoside. The assay was performed in a total volume of 200 µL containing 20 mM sodium phosphate buffer pH 7.2 and a final concentration of 10 mM substrate. The assay was initiated by addition of 1 µL of enzyme preparation (0.70 $mg.mL^{-1}$) and each enzyme treatment was tested in triplicate with the appropriate blank containing the same conditions without Xsa43E. Absorbance was monitored at 405 nm and the data were processed as in section 2.9.2.1.

2.9.2.6 Assay for the Effect of EDTA

EDTA was added to the enzyme in solution to give final concentrations ranging from 10 mM to 50 mM. The enzyme was then tested for activity either immediately or at varying time points over a 24 hour period. Activity was tested against the model

substrate *p*-nitrophenyl-α-L-arabinofuranoside at a final concentration of 5 mM in 20 mM sodium phosphate buffer pH 7.2. The reaction was initiated by addition of 5 µL of enzyme (2.5 $mg.mL^{-1}$). Activity tests always included both a positive control containing Xsa43E with no EDTA and a negative control containing EDTA but no enzyme.

2.9.3 Site Directed Mutagenesis

Site directed mutagenesis was performed on the Xsa43E gene through incorporation of custom oligonucleotides containing the desired mutation. This was achieved as outlined in section 2.1.4. The custom oligonecleotides used are detailed in Table 2-3.

The mutant genes were incorporated into the same vectors as the wild type gene as described in section 2.1.4.

Table 2-3: Custom oligonucleotides used to incorporate specific mutations into the Xsa43E gene.

D24A Fwd	5'−AAGGATATTTACACAGCAGCGCCTGCACCTATGGTGTAT−3'
D24A Rev	5'−ATACACCATAGGTGCAGGCGCTGCTGTGTAAATATCCTT-3'
D141A Fwd	5'−GGAGACTGGAACGATATTGCGCCTACAGTGTTTATAGAT-3'
D141A Rev	5'−ATCTATAAACACTGTAGGCGCAATATCGTTCCAGTCTCC-3'
E202A Fwd	5'−GGCACCAGCTATGGCGAAGCGCCGTGGTTTTATAAACGC-3'
E202A Rev	5'−GCGTTTATAAAACCACGGCGCTTCGCCATAGCTGGTGCC-3'
H258A Fwd	5'−GGCGGAGTATTTACAAACGCACCGGGAATTGCAGATTTT-3'
H258A Rev	5'−AAAATCTGCAATTCCCGGTGCGTTTGTAAATACTCCGCC-3'
H258Q Fwd	5'−GGCGGAGTATTTACAAACCAGCCGGGAATTGCAGATTTT-3'
H258Q Rev	5'−AAAATCTGCAATTCCCGGCTGGTTTGTAAATACTCCGCC-3'

2.9.3.1 Initial Activity Assays for Xsa43E Mutants

To test activity of the mutant enzymes 10 µL of enzyme was added to 1 mL of sodium phosphate buffered pH 7.2 containing a final concentration of 5 mM *p*-nitrophenyl-α-L-arabinofuranoside The samples were incubated at 37 °C for the duration of the experiment. The absorbance was measured at time points 0, 1, 4 and 24 hours at 400 nm. Positive and negative controls were also measured; the positive

control contained the Xsa43E enzyme and the negative contained lysis buffer instead of enzyme. All enzymes, including wildtype, were used at the same concentration.

2.9.3.2 pH Profiles for Xsa43E mutants.

pH profiles for Xsa43E H258A, Xsa43E H258Q and Xsa43E D141A were performed as for the wild type Xsa43E outlined in section 2.9.2.3.

Chapter Three – Characterisation of the Function and Catalytic Mechanism of Est1E and Est2A.

3.1 Introduction

3.1.1 The Importance of Ferulic Acid in Plant Cell Wall Structure

Plant cell walls are made up of three interconnecting networks: cellulose microfibrils, hemicellulose polysaccharides and lignin. They also contain aromatic substances and proteins[3]. The major hemicellulose in grasses and cereals is glucoroarabinoxylans (GAX) that have various side groups including arabinose units and acetyl groups. GAX also has phenolic acids covalently bound to the arabinose side chains with as well as deterring microbial attack; they have been shown to interfere with the attachment of ruminal bacteria to the plant cell wall impeding digestion. The most common phenolic acid linked to GAX is ferulic acid[100].

Ferulic acid is ubiquitous in complex plant cell walls and can form cross linking dimers and trimers between xylan chains[101]. Crosslinking dimers can be between polysaccharide chains or between a polysaccharide chain and lignin. The amount of ferulic acid dimers increases when the cell stops growing[102]. This increases the hydrophobicity of the cell wall and increases its mechanical rigidity[103]. The dimers are formed by peroxidase-mediated oxidation of trans-ferulic acid[100]. The rate of polysaccharide degradation by rumen microorganisms has been shown to decrease with increasing levels of ferulic acid dimers[101].

3.1.2 Feruloyl Esterases

Microorganisms require multiple esterases to remove all phenolic acid side groups from polysaccharide chains[102]. These esterases, including feruloyl esterases, have been shown to be produced by a wide range of microorganisms[17, 101]. Feruloyl esterases are classified as E.C. 3.1.1.73 and belong to the CE1 family of carbohydrate esterases in the CAZy database[55, 104]. These enzymes hydrolyse the ester bonds

releasing ferulic acid[101] from arabinose sidechains in hemicellulose (Figure 3-1).

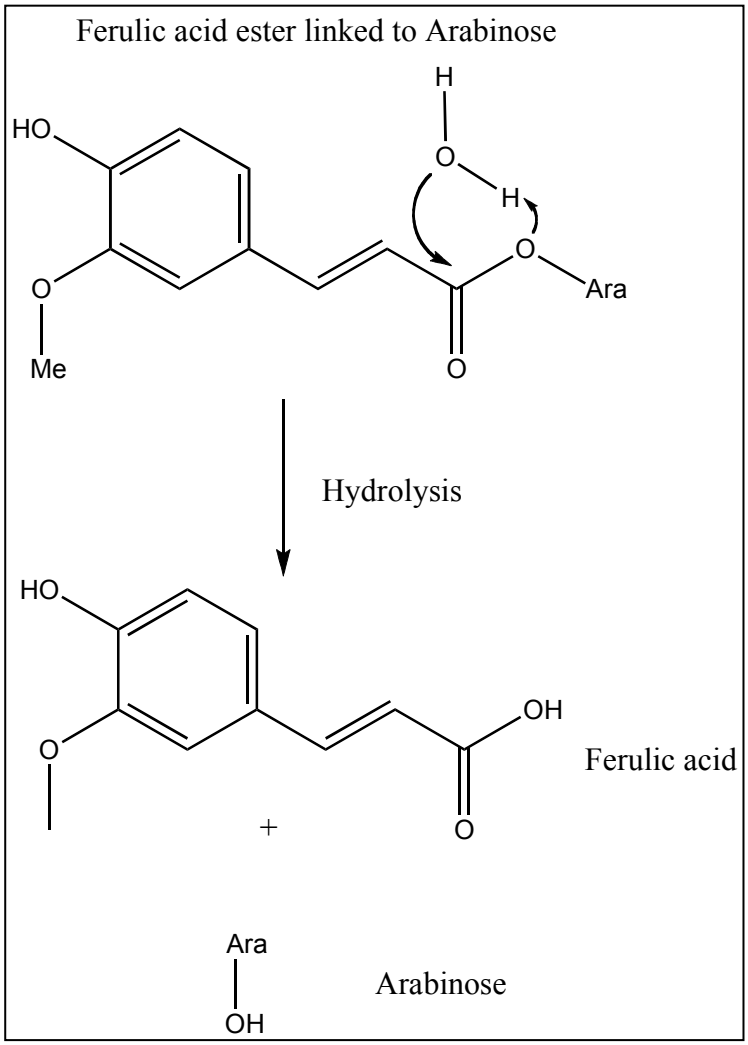


Figure 3-1: Hydrolysis reaction to cleave ferulic acid from arabinose.

Feruloyl esterases have been classified into four types, A, B, C and D, based on their protein sequence and substrate utilisation[104-106]. The three dimensional structures of feruloyl esterases from *Clostridium thermocellum*[107] and *Aspergillus niger*[108] have revealed they share a common α/ß hydrolase fold and a common catalytic triad consisting of Ser-His-Asp as is common to serine proteases, lipases and other esterases; the mechanism by which this facilitates the hydrolysis is shown in Figure 3-2. They also all share an oxyanion hole hypothesised to stabilise the tetrahedral

intermediate during the reaction[106]. Feruloyl esterases have now been recognised as a common part of hemicellulolytic enzyme systems and are becoming of interest for biotechnological applications[109].

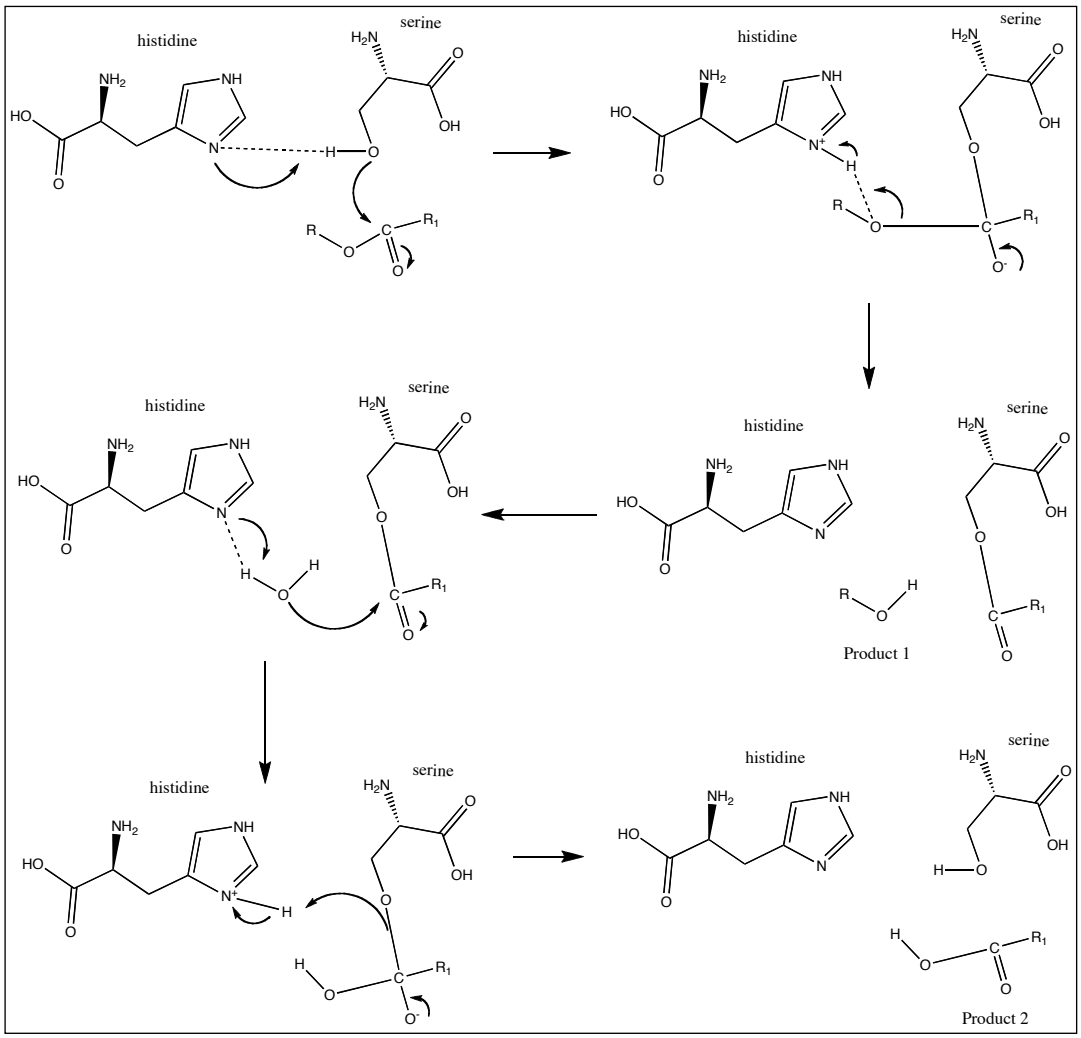


Figure 3-2: Mechanism for the hydrolysis of substrate by a serine esterase

3.1.3 Structure of Est1E

Est1E is a promiscuous feruloyl esterase from the rumen bacterium *Butyrivibrio proteoclasticus*[110]. The gene encoding Est1E, open reading frame (ORF) Bpr_I2870, is located on the main chromosome of *B. proteoclasticus*. Est1E was originally

annotated as a cinnamoyl esterase.

The structure of Est1E was solved both in the apo-form and with ferulic acid bound in the active site[110] (Figure 3-3). The structure consists of a canonical α/ß hydrolase fold with an insertion between the 6th and 7th strands of the central ß-sheet. This insertion creates a lid domain that covers the active site; there are no structural homologues for this lid domain in the PDB. The lid forms a flexible ß-sheet structure around a hydrophobic core. Est1E has four molecules in the asymmetric unit in the crystal, consisting of two dimers. The dimer interface is formed between four α-helicies – 2, 3, 4 and 5[110].

The active site contains three catalytic residues that are highly conserved throughout the α/ß hydrolase fold despite the large variation in function. The catalytic triad consists of a nucleophile, Ser-105, an acidic residue, Asp-197, and a histidine, His-225. Another common feature of the feruloyl esterases also seen in the Est1E structure is the oxyanion hole. In the apo-structure there is a phosphate ion located in the active site[110].

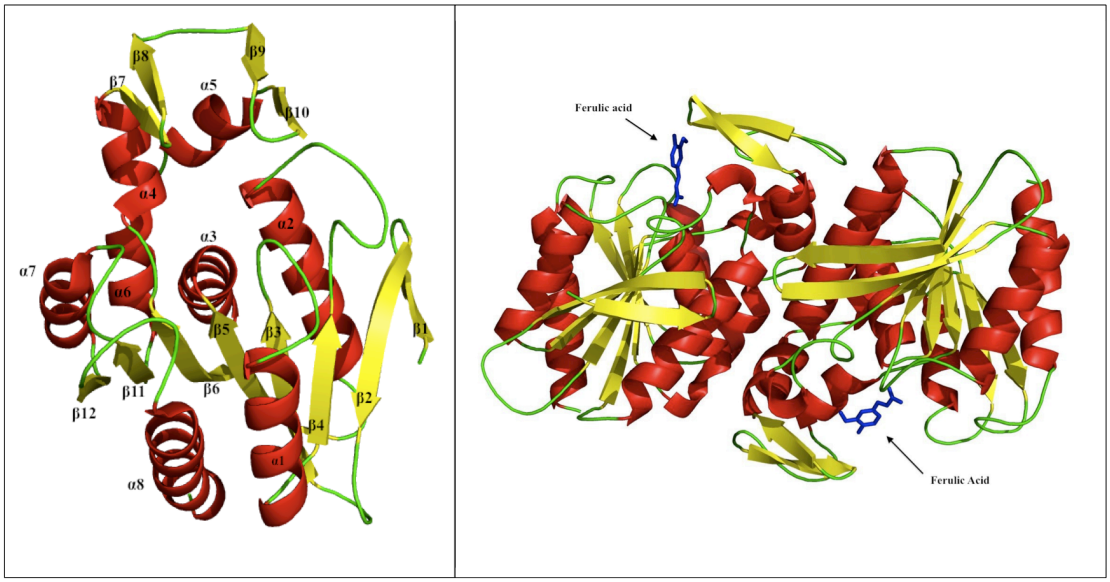


Figure 3-3: Structure of Est1E dimer with ferulic acid in blue bound in the active site adapted from Goldstone *et al* (2010)[110].

The structure of Est1E with ferulic acid shows that the substrate adopts different modes of binding in each of the two units of the dimer. The carboxylate tail is located next to Ser-105 in both monomers with one of the oxygen atoms from the acid group being in the oxyanion hole. In monomer A the aromatic head of the substrate is adjacent to Trp-160 in the hydrophobic pocket formed by the lid domain interacting with the main α/ß fold. In monomer B the aromatic head causes two leucine residues to shift and the methyl group occupies the hydrophobic pocket where Trp-160 is seen in monomer A (Figure 3-4)[110].

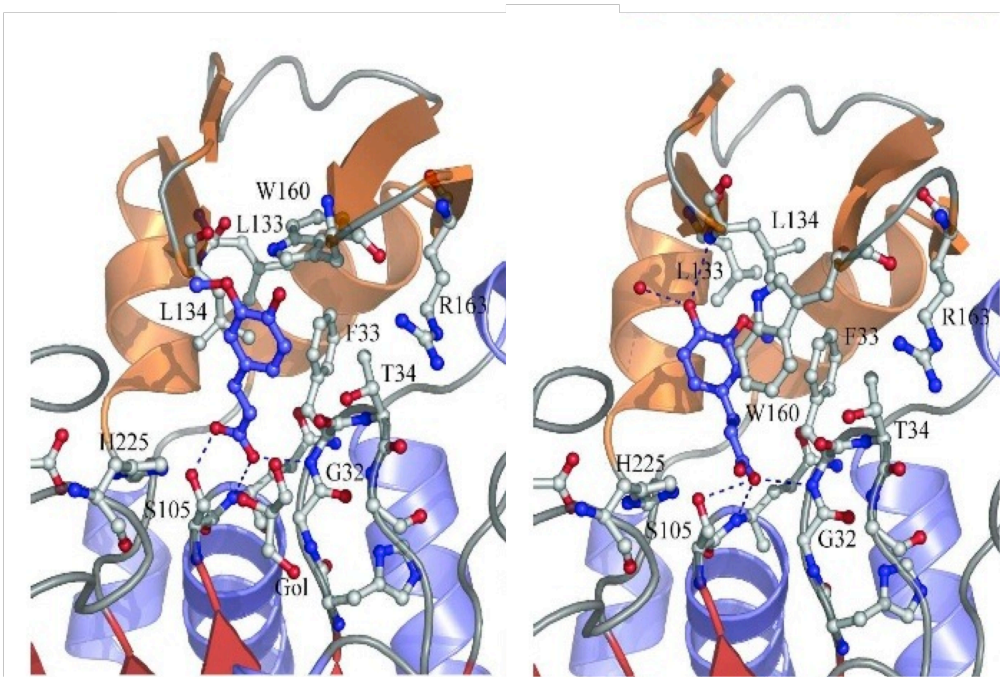


Figure 3-4: The two alternate conformations of ferulic acid bound in each of the two monomers in the Est1E dimer[110].

3.1.4 Functional Characterisation of Est1E

The activity of Est1E with natural substrates has also been previously investigated. Est1E releases a range of phenolic acids from various substrates (Table 3-1). Est1E was tested on a range of phenolic acid ethyl esters: ethyl ferulate, ethyl cinnamate and ethyl 3-coumarincarboxylate. It was able to cleave the ester linkage in all three of these but was most efficient against ethyl 3-coumarincarboxylate. Est1E was also tested on the natural substrates hemicellulose and birchwood xylan and the release of ferulic acid, trans-cinnamic acid, *p*-coumaric acid and 3-coumarincarboxylic acid was monitored. Est1E released ferulic acid, trans-cinnamic acid and *p*-coumaric acid from hemicellulose and trans-cinnamic acid, *p*-coumaric acid and 3-coumarincarboxylic acid from birchwood xylan[110]. The amount of the individual components in the substrates was not quantified.

Table 3-1: Activity of Est1E on natural substrates as analysed with GCMS. Errors are standard deviation. Taken from Goldstone *et al* (2010)[110].

Substrate	Product measured	Specific Activity ($\mu mol.mg^{-1}.min^{-1}$)
Ethyl ferulate	Ferulic acid	37 ± 1
Ethyl cinnamate	*Trans*-Cinnamic acid	2.6 ± 0.2
Ethyl 3-coumarincarboxylate	3-coumarincarboxylic acid	199 ± 10
Hemicellulose (ryegrass)	Ferulic acid	6.9 ± 0.4
	p-coumaric acid	4.1 ± 0.7
	Trans-Cinnamic acid	4.0 ± 0.1
Birchwood Xylan	Ferulic acid	0
	p-coumaric acid	2.9 ± 1
	Trans-Cinnamic acid	7.9 ± 0.2
	3-coumarincarboxylic acid	7.1 ± 3

3.1.5 Acetyl Xylan Esterases

Acetyl groups occur on about 70% of the xylose units in GAX[111] at the C-2 and/or C-3 positions. These groups are linked by ester bonds and alter the solubility of the polysaccharide.

The degradation of GAX requires a suite of enzymes. Xylanases are required to cleave the xylan backbone, but before this can be achieved the branching side groups need to be removed in order to allow better enzymatic access to the backbone. These sidegroups include phenolic acids, arabinose units and acetyl groups. Acetyl xylan

esterases (AXE) are enzymes that cleave these acetyl side groups in xylan. They are the most diverse type of carbohydrate esterases (CEs) as they can be found in eight of the sixteen CE families in the CAZy database[55]. AXEs are produced by a wide range of microorganisms including *Thermobifida*[112], *Neocallimastix*[113], *Fibrobacter*[114], *Aspergillus*[115] *Termitomyces*[116], *Streptomyces*[117], *Pseudomonas*[118] and *Bacillus*[119, 120]. They have also been identified from plant sources such as Malted finger millet *Elueusine coracana*[121].

Mechanistic investigation of AXEs has been limited although it is believed the majority of these enzymes share the mechanism of 'serine' esterases that have a Ser-His-Asp catalytic triad[122].

3.1.6 Structure of Est2A

Est2A is an acetyl xylan esterase produced by the rumen bacterium *B. proteoclasticus.* The structure of Est2A has been solved using SAD methods and there are structures for both seleno-methionine incorporated protein and native protein[123]. The two structures are essentially identical. The final model of native Est2A consists of eight monomers in the asymmetric unit in the form of two tetramers stacked perpendicular to each other in an edge to face arrangement. Each tetramer forms a 'donut' like shape with a large central hole. The structure of Est2A can be seen in Figure 3-5 as (a) a monomer, (b) a dimer, (c) a tetramer and (d) the two tetramers as seen in the asymmetric unit.

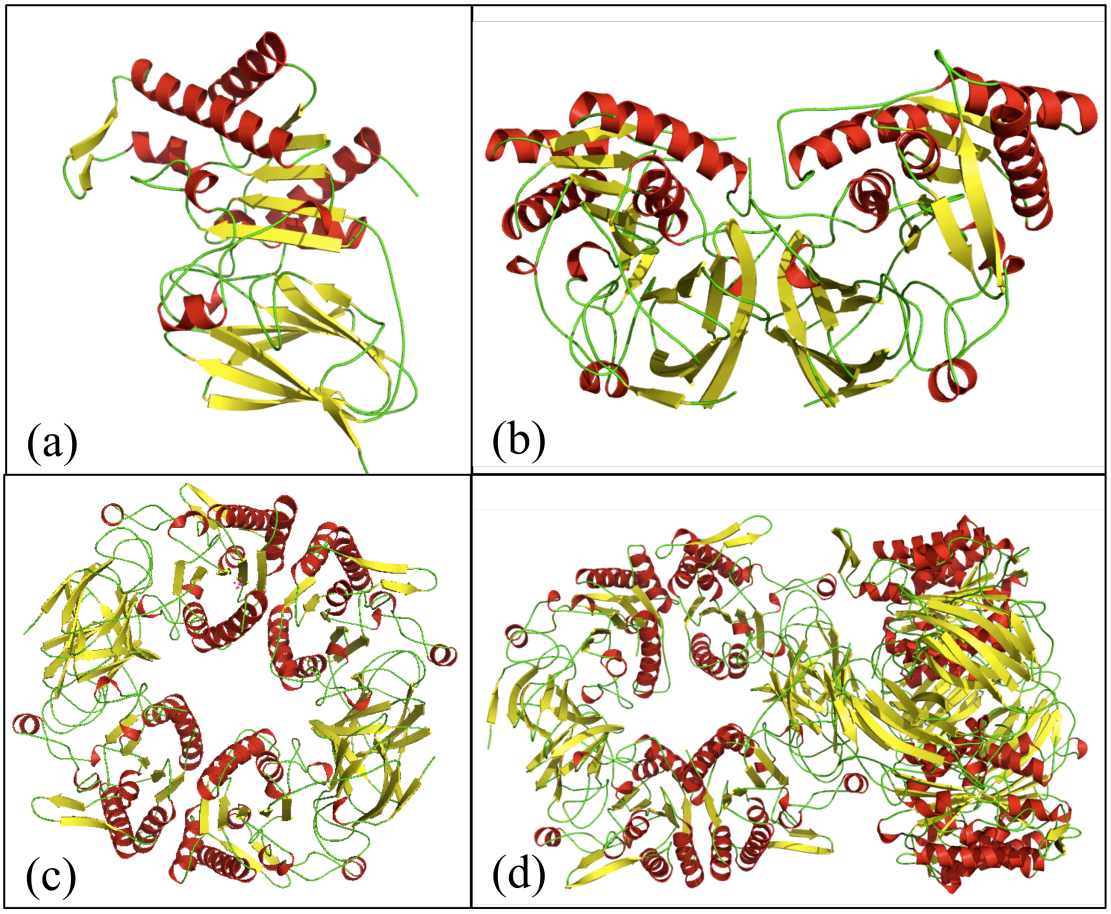


Figure 3-5: Structure of Est2A seen as (a) a monomer, (b) a dimer, (c) a tetramer and (d) the two tetramers as seen in the assymetric unit stacking in an edge to face manner[123].

Each Est2A monomer is made up of two domains: a catalytic C-terminal SGNH-hydrolase domain and an N-terminal jelly-roll domain. The monomers are arranged in the tetramer as two dimers with the C-terminal domains creating the interface between the two monomers and the N-terminal domains forming the second interface between the dimers, thus completing the tetramer ring structure.

The N-terminal jelly-roll domain consists of residues Val-3 through to Pro-130. The structure is formed by two opposing β-sheets, the first containing four strands - β1, β9, β4 and β7, and the second containing five strands - β2, β3, β8, β5 and β6[123].

A search of the PDB database revealed that this domain is most similar to

carbohydrate binding domains. Structural based sequence alignment indicated several residues as potentially important for polysaccharide binding[123]. The loop joining β-strand β2 and β3 appears to be in the vicinity of the binding groove, with Phe-33 and Phe-34 located in a position to interact with the polysaccharide chain. The side chains of Asn-108, Lys-25 and Lys-26 are also located in positions that would allow them to interact with the substrate. Est2A has a loop between α5 and α7 including α6 that appears to be blocking potential interactions with substrate in the binding groove. The loop is large and appears it would be easily displaced on binding of substrate[123] because there is space in the crystal structure that would accommodate the loop if it were to shift out of the binding site; also there is only a small portion if the loop within 5 Å of the rest of the protein indicating it is not held firmly in that position through intermolecular forces (Figure 3-6).

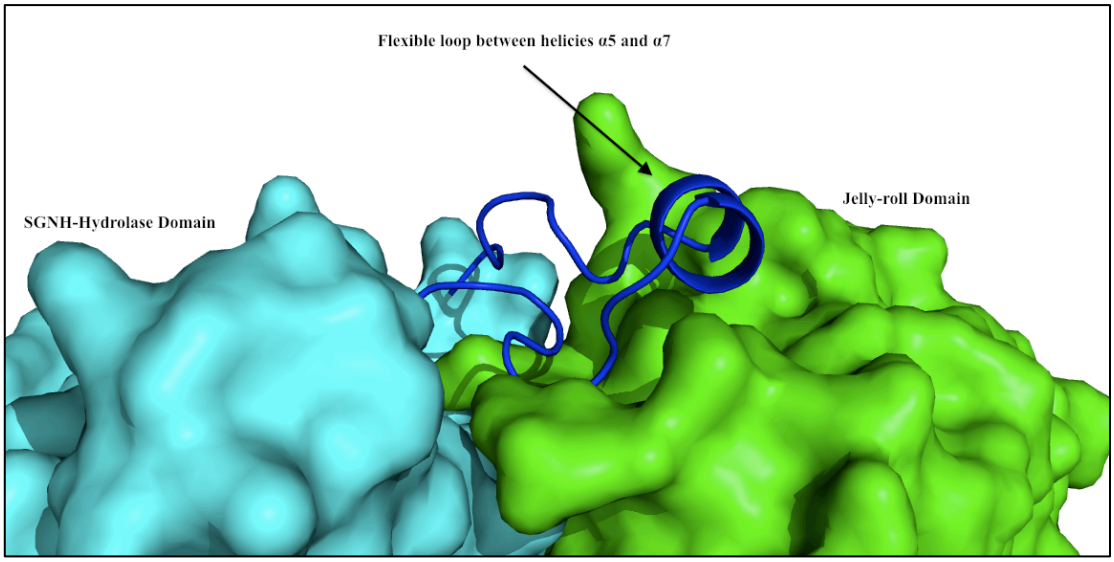


Figure 3-6: Surface of Est2A chain A showing the SGNH domain in aqua, the jelly roll domain in green and the flexible loop between helicies α5 and α7 shown in blue.

The C-terminal SGNH-hydrolase domain is made up of residues Val-131 through to Phe-375 and has a three layered α/β stack structure. The central ß sheet has five β-strands, β11, β10, β12, β15 and β16, arranged in a parallel manner. There are nine α helices packing against either side of the sheet. Helices α4 and α11 are packed against the central sheet in the centre of the ring while helices α6, α5, α9 and α10 are located on the opposite side of the sheet. The central ß-sheet contains the proposed catalytic dyad, Ser-142 and His-351, located at the C-terminal end[123]. The catalytic dyad was identified by comparison with *Streptomyces scabies* esterase (PDB accession code 1ESC). These two proteins have 20% sequence identity, an alignment of the Cα trace using DaliLite gave an RMSD of 3.0 Å[124].

The structure of Est2A also contains an acetate molecule. It is located in the active site next to Ser-142. One of the acetate oxygen atoms is located in the oxyanion hole formed by the backbone nitrogens of Ser-142, the adjacent Gly-188 at the N-terminus of helix α2 and the sidechain nitrogen of Asn-245. The oxyanion hole is present to stabilise the negatively charged oxygen of the tetrahedral reaction intermediate formed by the nucleophilic attack of the serine oxygen on the carbonyl carbon of the

acetyl group. The other acetate oxygen atom is positioned within hydrogen bonding distance of the Nε2 atom of His-351. The methyl group is directed away from the active site into a region containing hydrophobic residues including Ile-143 and Thr-244. The active site of Est2A is located at the base of a groove lined by Trp-158 and Trp-195 that are parallel to each other on either side above the active site. This groove runs the length of the SGNH fold above the active site forming an ideal binding cleft for the polysaccharide substrate (Figure 3-7)[123].

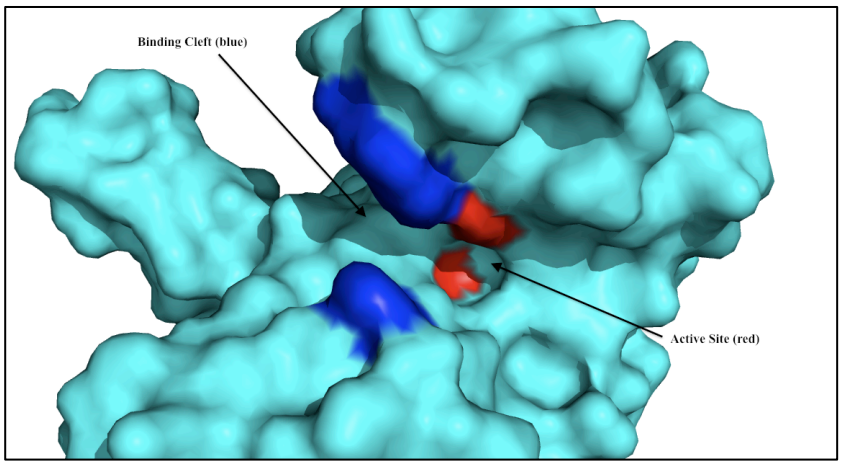


Figure 3-7: Surface of Est2A showing the active site in red and the binding cleft in blue.

3.1.7 Work Presented In This Chapter

To compliment the previous structural and functional investigation of Est1E I have established kinetic parameters with model substrates. I have also constructed a mutant, Est1E S105A, in order to test the hypothesis of the Ser-Asp-His catalytic triad and potentially gain insight into the substrate binding mechanism by co-crystallising the enzyme with substrate. The mutant was subject to structural and functional investigation. Kinetic parameters were also established for Est2A with model substrates. Two mutants of Est2A were made to test the hypothesis of a catalytic dyad and to gain insight into substrate binding by co-crystallisation with substrate as well.

3.2 Est1E Results and Discussion

3.2.1 Expression and Purification of Est1E

Est1E was expressed and the cells lysed as outlined in sections 2.2.3.1 and 2.2.3.3 respectively. Est1E was purified using immobilised metal affinity chromatography (IMAC) (section 2.2.5). The elution of protein was followed by monitoring the absorbance at 280 nm and visualised by running samples on a 12% SDS-PAGE gel (section 2.2.1). Est1E elutes from an IMAC Ni^{2+} column at ~150 mM imidazole (Figure 3-8).

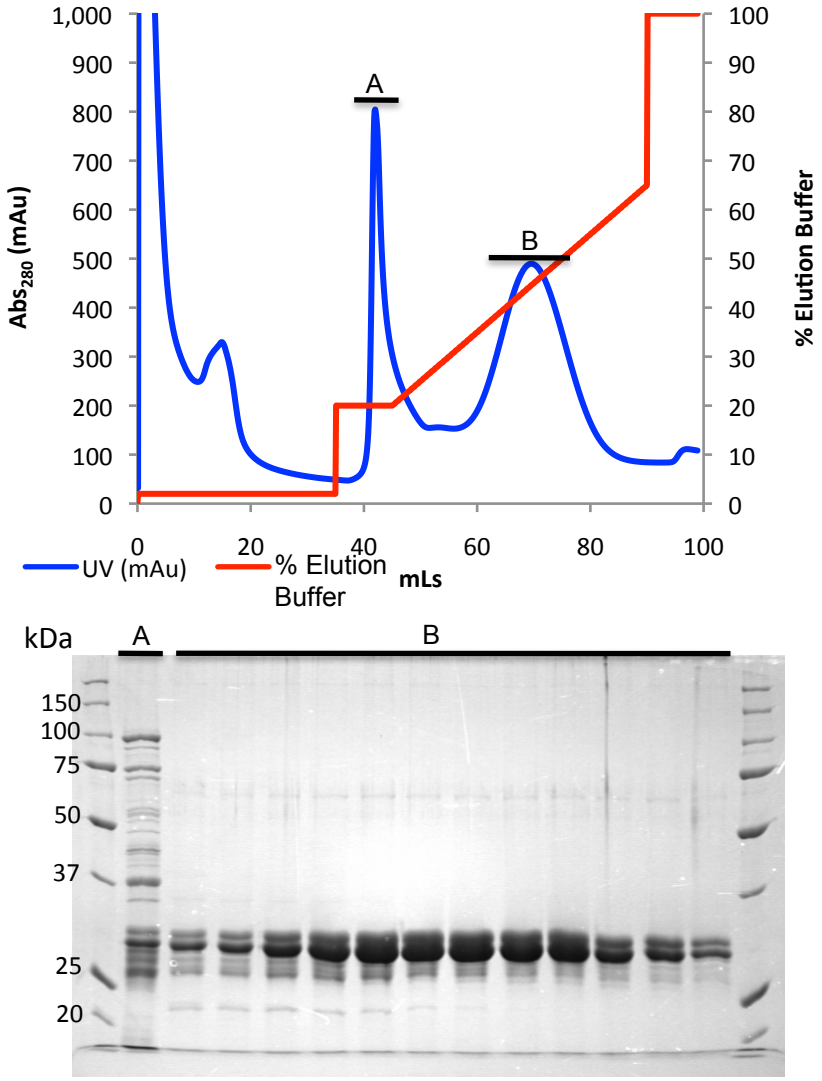


Figure 3-8: Est1E trace depicts absorbance profile with 12% SDS-PAGE gel of corresponding fractions.

Fractions containg Est1E from peak B were pooled and concentrated to less than five mL (section 2.2.7) in preparation for running through an S200 16/60 size exclusion column (method in section 2.2.6). Est1E eluted as a single peak from the S200 16/60 column at around 80 mL (Figure 3-9), this corresponds to a M_r of 62 kDa indicating Est1E is a dimer in solution.

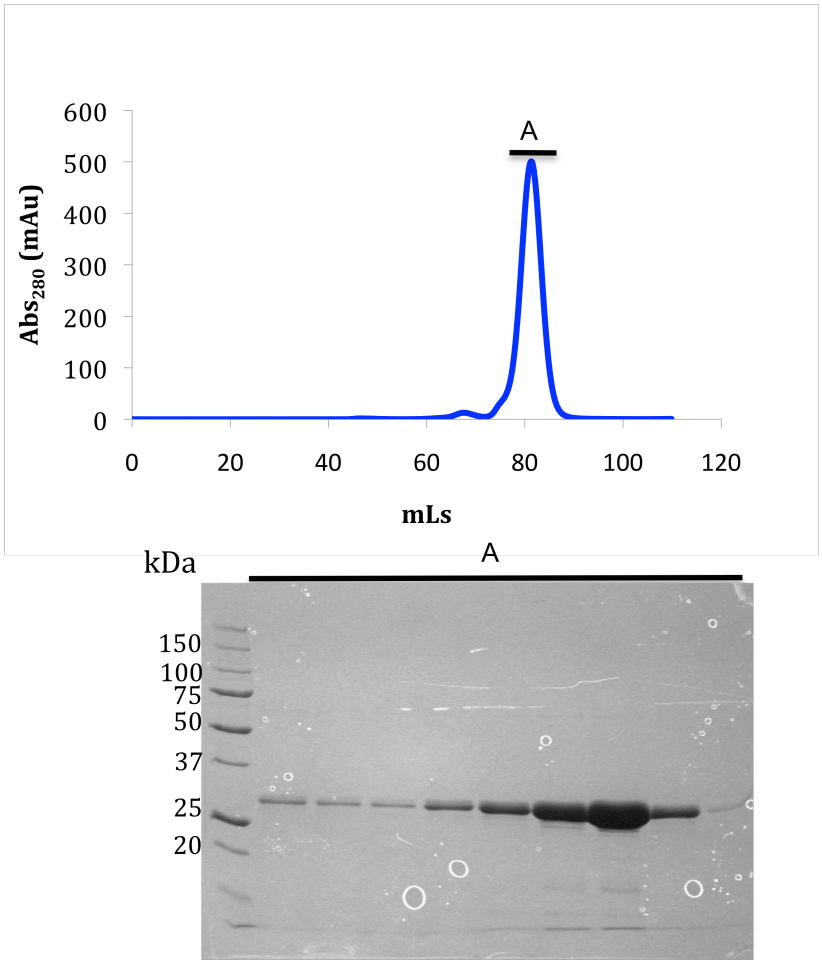


Figure 3-9: Chromatogram of Est1E eluting from S200 16/60 column with corresponding fractions run on a 12% SDS-PAGE gel.

3.2.2 Kinetic parameters for Est1E with Model Substrates

The activity of Est1E against two model substrates, *p*-nitrophenyl acetate (*p*-NPA) and *p*-nitrophenyl butyrate (*p*-NPB) (Figure 3-10), was determined by monitoring the absorbance of the reaction solution at 400 nm. These model substrates mimic the structures seen in the natural substrates shown in Figure 1-2.

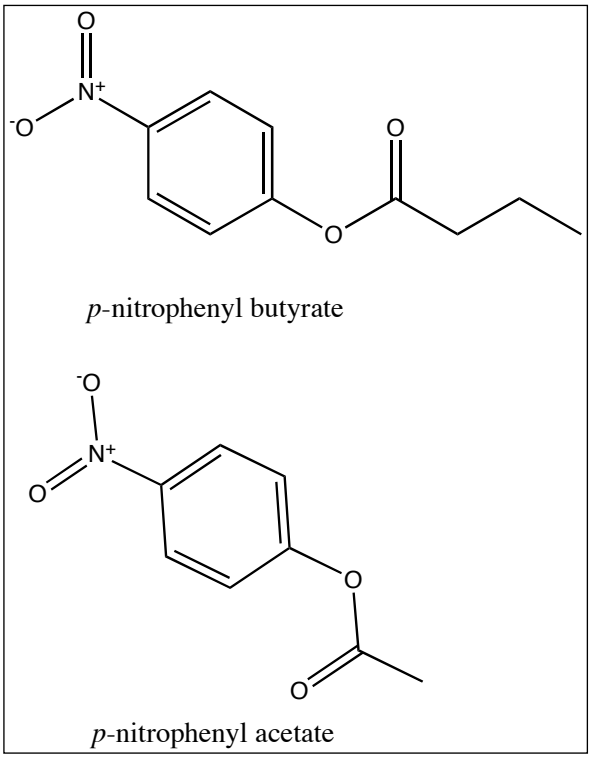


Figure 3-10: Structures of model substrates *p*-nitrophenyl butyrate and *p*-nitrophenyl acetate

Est1E readily degrades *p*-NPA. The activity of Est1E on *p*-NPA across a range of substrate concentrations follows Michaelis-Menten kinetics (Figure 3-11).

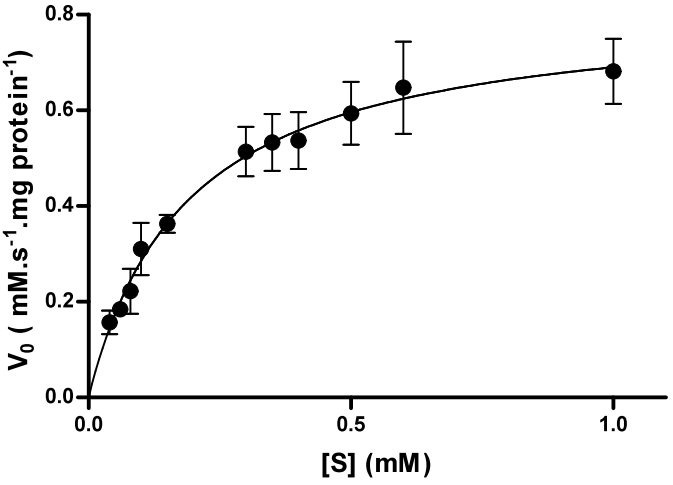


Figure 3-11: Rate of degradation of *p*-nitrophenyl acetate by Est1E at varying substrate concentrations with standard errors.

Although Est1E is not an acetyl esterase its activity on *p*-NPA allows the rate of ester cleavage to be followed in real time. The ability of Est1E to cleave this substrate indicates that the active site pocket is flexible and can accommodate a range of substrates.

p-Nitrophenyl-butyrate (*p*-NPB) is a second model substrate that was used to determine kinetic parameters for Est1E. *p*-NPB has a four carbon group attached to the *p*-nitrophenol ring. The rate profile with increasing substrate concentration (Figure 3-12) indicates it also follows Michaelis-Menten kinetics.

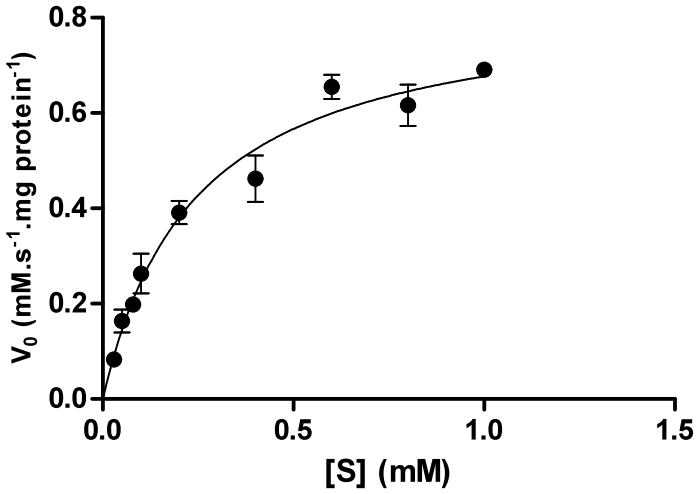


Figure 3-12: Rate of degradation of *p*-nitrophenyl butyrate by Est1E at varying substrate concentrations with standard errors.

The data for both substrates were analysed using Prism GraphPad to determine the kinetic parameters V_{max}, K_M and k_{cat} (Table 3-2).

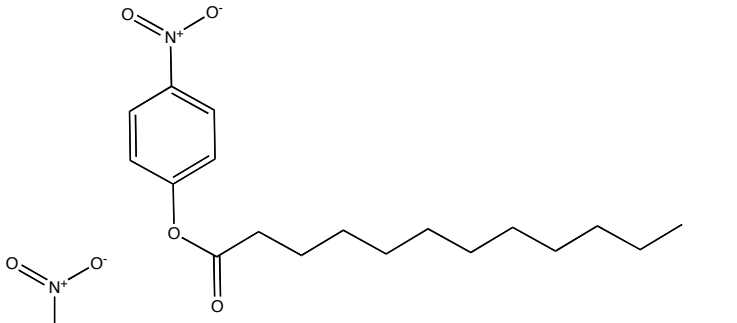
Table 3-2: Kinetic parameters for Est1E with Model substrates. Errors given are standard errors.

Est1E	V_{max} ($x10^{-4}$ $mM.s^{-1}$)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($x\ 10^2\ s^{-1}.mM^{-1}$)
p-Nitrophenyl Acetate	5.7 ± 0.3	0.19 ± 0.02	23 ± 1	1.2 ± 0.2
p-Nitrophenyl Butyrate	5.9 ± 0.3	0.24 ± 0.04	24 ± 1	1.0 ± 0.2

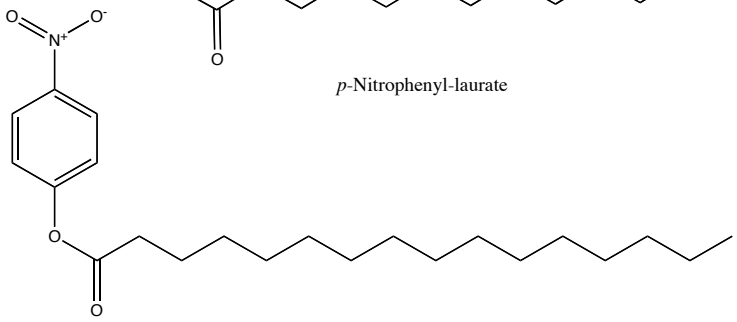
The kinetic parameters in Table 3-2 indicate that Est1E can efficiently hydrolyse both substrates, accommodating both the C4 and the C2 groups in the active site. The K_M and k_{cat} values for the two substrates are not significantly different; this indicates that neither substrate is interacting with the active site more effectively than the other.

3.2.3 Est1E activity on other model substrates

Est1E was tested for activity against two larger *p*-nitrophyl substrates, *p*-nitrophyl laurate (C-12) and *p*-nitrophyl palmitate (C-16) (Figure 3-13). Both substrates have unbranched, saturated long chain carbon tails. Est1E showed activity against both *p*-nitrophyl laurate and *p*-nitrophyl palmitate. Kinetic parameters were not determined due to both substrates having very poor solubility.



p-Nitrophenyl-laurate



p-Nitrophenyl-palmitate

Figure 3-13: Chemical structures of *p*-nitrophenyl laurate and *p*-nitrophenyl palmitate.

3.2.4 Kinetic Activity of Est1E Compared with Other Feruloyl Esterases

The specific activity of Est1E on the model substrates ethyl ferulate (previously determined, Goldstone *et al*[110]), *p*-NPA and *p*-NPB are compared with the specific activities of other feruloyl esterases from various sources in Table 3-3.

Table 3-3: Specific activity of various feruloyl esterases on model substrates. est. = esterase.

Enzyme	**Organism**	**Substrate**	**Specific activity** ($\mu mol.min^{-1}.mg^{-1}$)
Ragi FAE	*Eleusine coracana*, Indaf-15[125]	Ethyl ferulate	3
AwfaeA	*Aspergillus awamori*[126]	Ethyl ferulate	1.85
Feruloyl est.	*Clostridium stercorarium*[127]	Ethyl ferulate	88
Est1E	*B. proteoclasticus*[110]	Ethyl ferulate	37
FAE-1	*Aspergillus niger*[128]	*p*-NPA	6.2
FAE-2	*Aspergillus niger*[128]	*p*-NPA	6.1
Feruloyl est.	*Aureobasidium pullulans*[129]	*p*-NPA	31.8
Est1E	*B. proteoclasticus*[110]	*p*-NPA	51
FAE-1	*Aspergillus niger*[128]	*p*-NPB	1.0
FAE-2	*Aspergillus niger*[128]	*p*-NPB	1.9
Feruloyl est.	*Aureobasidium pullulans*[129]	*p*-NPB	0.02
Est1E	*B. proteoclasticus*[110]	*p*-NPB	50.4

The feruloyl esterases from the plant *E. coracana*[125] and the fungus *A. awamori*[126] have very low specific activities with the model substrate ethyl ferulate. Est1E has a higher specific activity for ethyl ferulate[110] and the feruloyl esterase from *C. stercorarium*[127] has a specific activity that is higher again. This indicates that Est1E is efficient at cleaving ethyl ferulate compared with the feruloyl esterases reported in Table 3-3 with only the esterase from *C. stercorarium*[127] having a higher specific activity than Est1E.

The specific activities shown in Table 3-3 of Est1E on the model substrates *p*-PNA and *p*-PNB are 2 fold and 2000 fold higher than that of the feruloyl eseterase from the black yeast *A. pullulans*[129] respectively. There were no reported K_M or k_{cat} values for these substrates although it was noted that the enzyme was practically inactive with *p*-NPB. For the two feruloyl esterases FAE-1 and FAE-2 from *A. niger,* the specific

activity on *p*-NPA were roughly ten fold lower than Est1E for both. For the same two enzymes the specific activity for *p*-NPB[128] was around 50 fold lower that that of Est1E. These data (Table 3-3) indicate that Est1E from *B. proteoclasticus* has a much better specific activity on both model substrates than other feruloyl esterases reported in the literature.

The kinetic data for Est1E with these model substrates compliments the previous work on Est1E and allows the direct comparison of the specific activity of Est1E with feruloyl esterases from other sources. The mechanism by which Est1E catalyses these reactions has been hypothesised to be a Ser-Asp-His catalytic triad[110], to investigate this a mutation of Ser-105 to Ala was inserted into Est1E (Est1E S105A).

3.2.5 Expression and Purification of Est1E S105A

Est1E S105A was expressed and purified in the same way as Est1E (section 3.2.1). The IMAC chromatogram (Figure 3-14) looks very similar to that of Est1E (Figure 3-8). Est1E S105A elutes in significant quantities from the IMAC Ni^{2+} column with ~150 mM imidazole in the elution buffer.

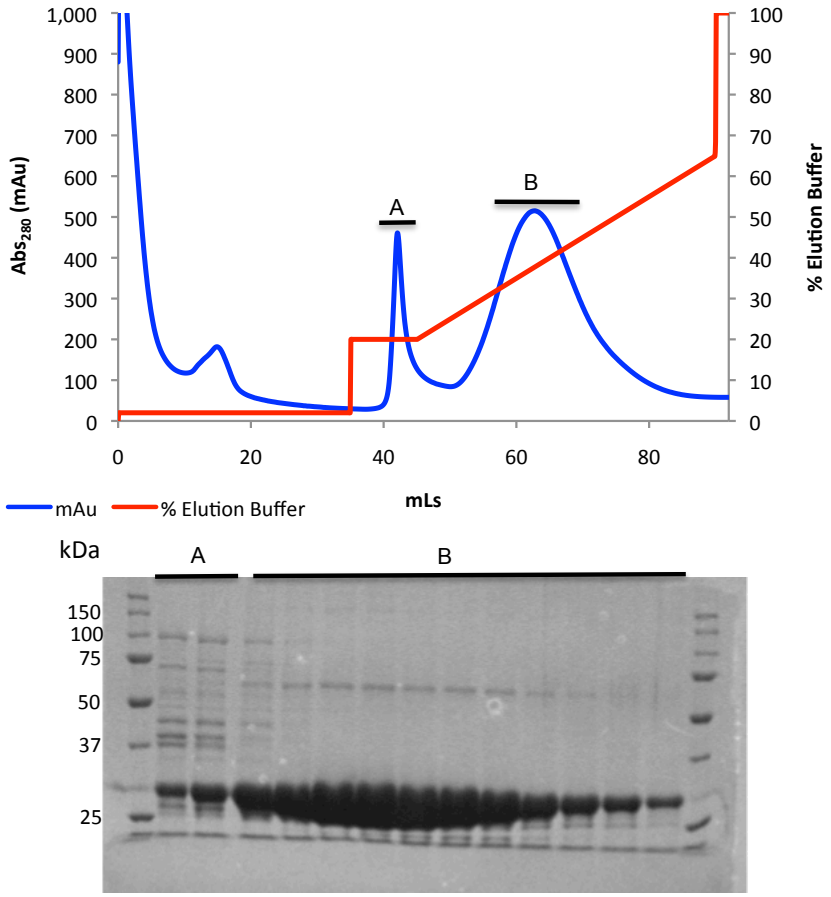


Figure 3-14: Est1E S105A purification by IMAC with 12% SDS-PAGE gel of corresponding fractions.

Fractions shown to contain Est1E S105A in peak B from Figure 3-14 were further purified by size exclusion chromatography using an S200 16/60 column. The chromatogram from the S200 16/60 column with the corresponding 12% SDS-PAGE gel is shown in Figure 3-15.

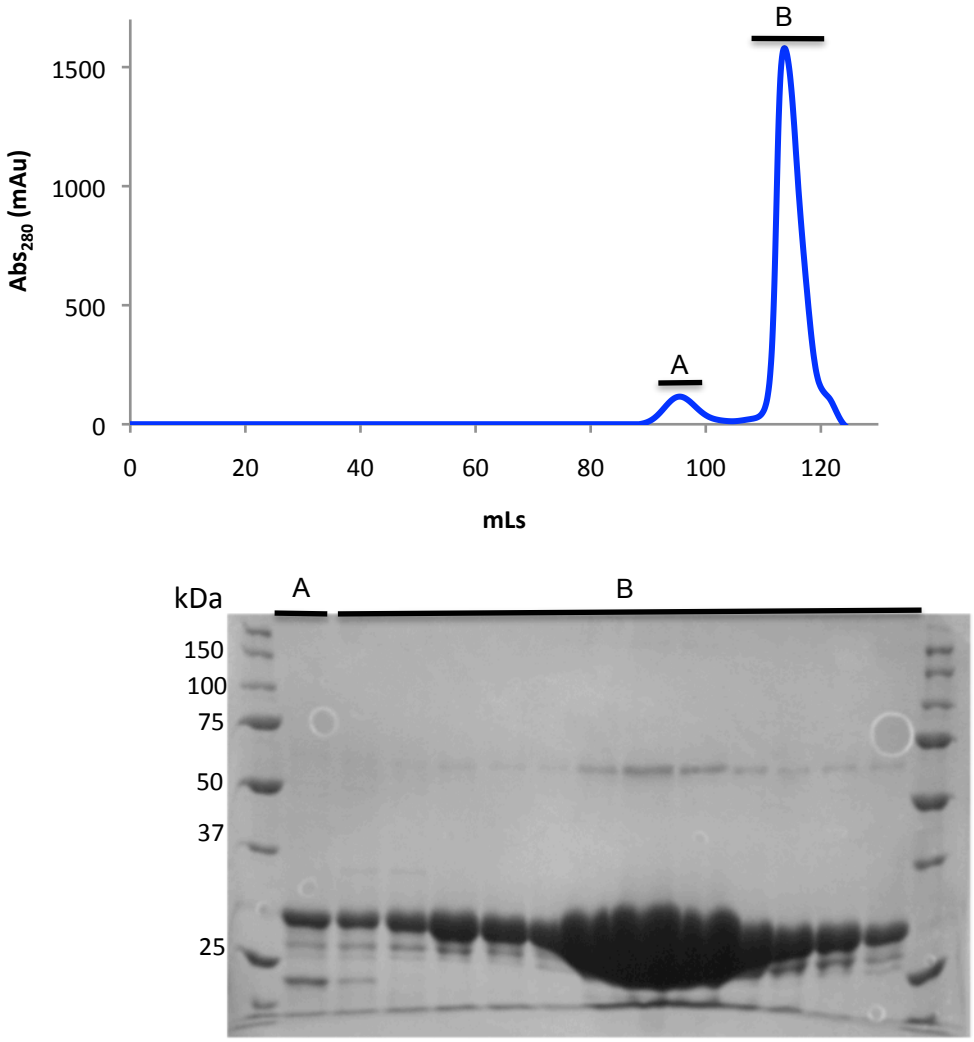


Figure 3-15: Est1E S105A S200 16/60 chromatogram with corresponding 12% SDS-PAGE gel.

Est1E elutes from the S200 16/60 size exclusion column at 80 mL (Figure 3-9), Est1E S105A elutes after a full column volume at the end of the column at 120 mL. This elution point corresponds to a M_r of 3.2 kDa, which is clearly not in agreement with the corresponding SDS-PAGE gel. This indicates that Est1E S105A is interacting with the column in some way. The conditions under which the column was run and the amount of protein are analogous to purification of Est1E therefore it is not due to different running conditions. Due to the different behaviour of Est1E S105A on the S200 16/60 size exclusion column, a native gel was run to

compare Est1E S105A with Est1E (Figure 3-16). Samples from both peaks A and B from Figure 3-15 were run on the gel as well as a sample of Est1E. Samples from both peaks of Est1E S105A appear the same on the native gel and both differ from the sample of Est1E in the distance into the gel they had run. The serine to alanine mutation was not expected to influence the isoelectric point but it is clear that it has done so. The apo-structure of native Est1E shows there is a phosphate ion in the active site[110], the mutation of the serine to alanine has the potential to prevent the phosphate binding in the active site. This could be responsible for the shift observed on the native gel; a phosphate ion has a three negative charge therefore Est1E S105A without the phosphate bound would result in it not running as far into the native gel as Est1E with a phosphate ion.

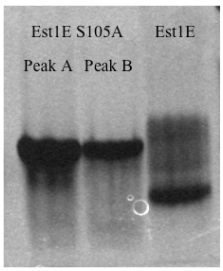


Figure 3-16: Native gel comparing Est1E with Est1E S105A.

3.2.6 Activity of Est1E S105A

Activity analysis of Est1E S105A using the model substrate *p*-NPA showed no degradation above that of the blank (Table 3-4) and indicates that Est1E S105A is inactive on these substrates. This supports the hypothesis that the catalytic mechanism of Est1E involves the highly conserved catalytic triad, Ser-Asp-His, seen commonly in enzymes with analogous α/ß hydrolase folds.

Table 3-4: Degradation of *p*-NPA with and without Est1E S105A present.

	Rate (x 10^{-5} $\Delta Abs_{400}.s^{-1}$)
Blank	5.0 ± 1.7
Est1E S105A	3.8 ± 0.9

3.2.7 Crystallisation and Structure of Est1E S105A

Purified Est1E S105A was placed in crystallisation trials (section 2.3.1). These trials gave no conditions in which crystals grew. Co-crystallisation screens were performed around the condition of the wildtype protein with ferulic arabinose included in the mother liquor. This screen gave small crystals within two weeks but these were poor quality and disintegrated on manipulation. After six months more conditions yielded better crystals. The best crystals of Est1E S105A grew in the condition 18% PEG 3350, 0.35 M NaH_2PO_4 with a protein concentration of 36.6 $mg.mL^{-1}$.

Two crystals of Est1E S105A cocrystallised with ferulic arabinose were sent to the Australian synchrotron and one dataset was collected to 2.5 Å resolution. The complete data statistics and refinement statistics can be seen in Table 3-5 and Table 3-6 respectively.

Table 3-5: Data collection statistics for Est1E S105A.

Est1E S105A	Native Data
Space group	$P22_12_1$
Wavelength	1.0012 Å
Cell parameters	
a	47.81 Å
b	59.14 Å
c	190.22 Å
α	90 °
β	90 °
γ	90 °
Resolution range (Å)	64.41-2.5 (2.64-2.5)
R_{merge}	0.228 (0.457)
No. of measured reflections	201756 (29939)
No. of Unique reflections	19463 (2771)
Mean I/σI	7.9 (4.4)
Completeness	100 (100)
Multiplicity	10.4 (10.8)

Table 3-6: Complete refinement statistics for Est1E S105A.

Refinement Statistics	Est1E S105A
R_{Work} (R_{free})	18.7% (24.3%)
Total No. Atoms	4077
No. of Protein Atoms	3950
Other Molecules/Ions	0
No. of Waters	131
RMS Deviation from Standard Geometry	
Bond Lengths (Å)	0.0173
Bond Angles (°)	1.7229
Average B-factors	
Protein	28.806
Water	29.047

The structure of Est1E is in spacegroup $P2_1$ for the native protein and in space group $C2$ with substrate bound. The structure of Est1E S105A is in space group $P22_12_1$. The asymmetric unit of the native structure has four molecules in it whereas Est1E S105A has two molecules in the asymmetric unit in accordance with the higher symmetry space group. The final *R-factor* and R_{free} were 18.7% and 24.3% respectively. Ramachandran analysis of the Est1E S105A structure showed 93.2% of residues were in favoured conformations and 5.4% were in allowed regions; 1.4% of the residues were outliers. The electron density map of the Est1E S105A does not indicate that there is any substrate bound in the active site. In the structure of Est1E without substrate there is a phosphate ion in the active site. This is not seen it the structure of Est1E S105A. This supports the reason for why it ran differently to Est1E on the native gel (Figure 3-16). In the structure of the native enzyme with ferulic acid in the active site, the substrate interacts with the Ser-105 residue. This absence of substrate or phosphate ion could indicate that the Ser-105 residue is important for substrate binding.

The RMSD for the Cα trace between A chains of both Est1E and Est1E S105A is 0.4 Å, and for B chains was 0.8 Å; it is important to note that the RMSD for the Cα trace from Est1E chain A and Est1E S105A chain B is 0.5 Å. When Est1E chain B is overlaid with Est1E S105A chain B there is a divergence of the peptide backbone between residues 156 and 164. This is part of the lid domain, which in the Est1E

dimer takes alternate conformations. These alternate conformations are not seen in Est1E S105A and this appears to be responsible for the higher symmetry space group as the lattice of molecules of Est1E and Est1E S105A overlay; the RMSD of the Cα traces for chain A and chain B of Est1E S105A is 0.4 Å. The cartoon representations of aligned Est1E and Est1E S105A structures (Figure 3-17) illustrates that they are essentially identical apart from the point mutation seen in the active site. Figure 3-18 illustrates the different orientation of Trp-160 in monomer B of Est1E S105A when overlaid with chain A from Est1E.

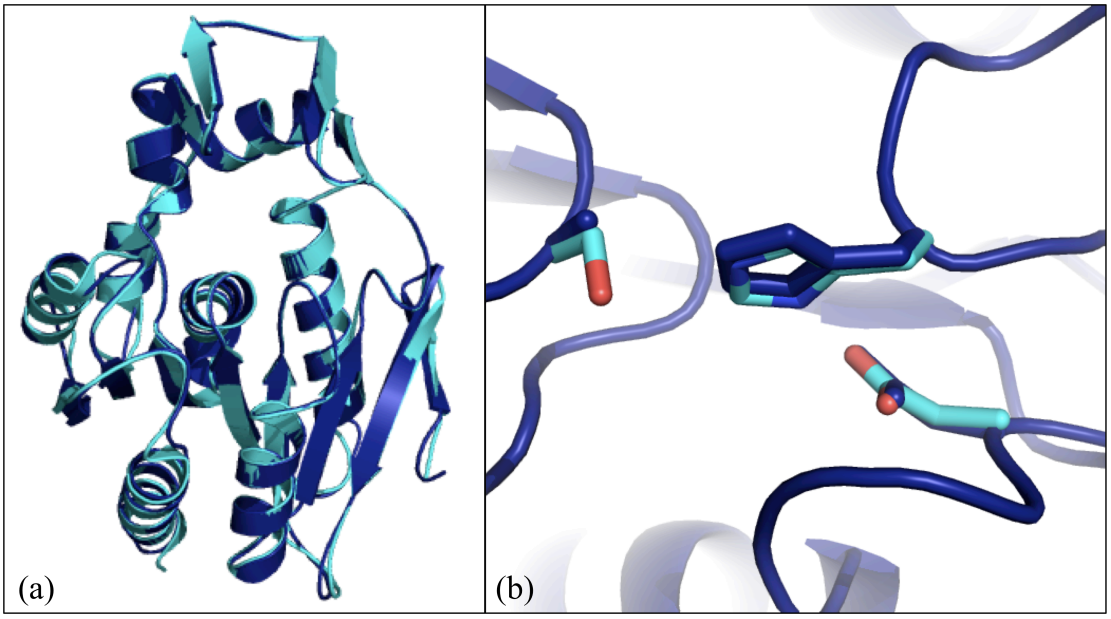


Figure 3-17: (a) Est1E S105A structure in blue overlaid with Est1E structure in aqua. (b) The active site of Est1E in aqua and Est1E S105A in blue.

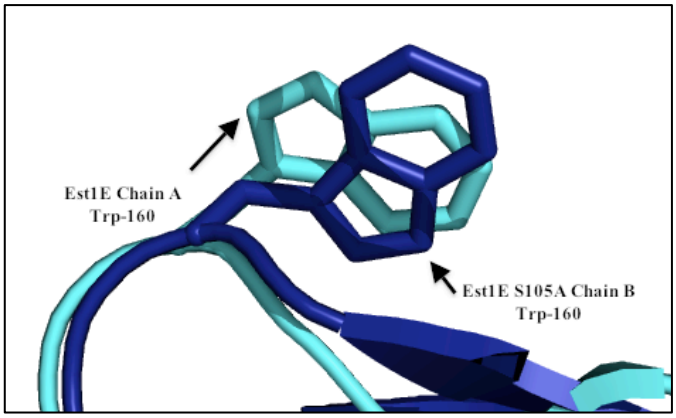


Figure 3-18: Difference seen in Trp-160 between Est1E chain A (aqua) and Est1E S105A chain B (blue).

3.3 Est2A Results and Discussion

3.3.1 Expression and Purification of Est2A

Est2A was expressed and purified as detailed by prior researchers and in the same manner as outlined in section 3.2.1. Est2A eluted from the IMAC column at ~200 mM imidazole (Figure 3-19).

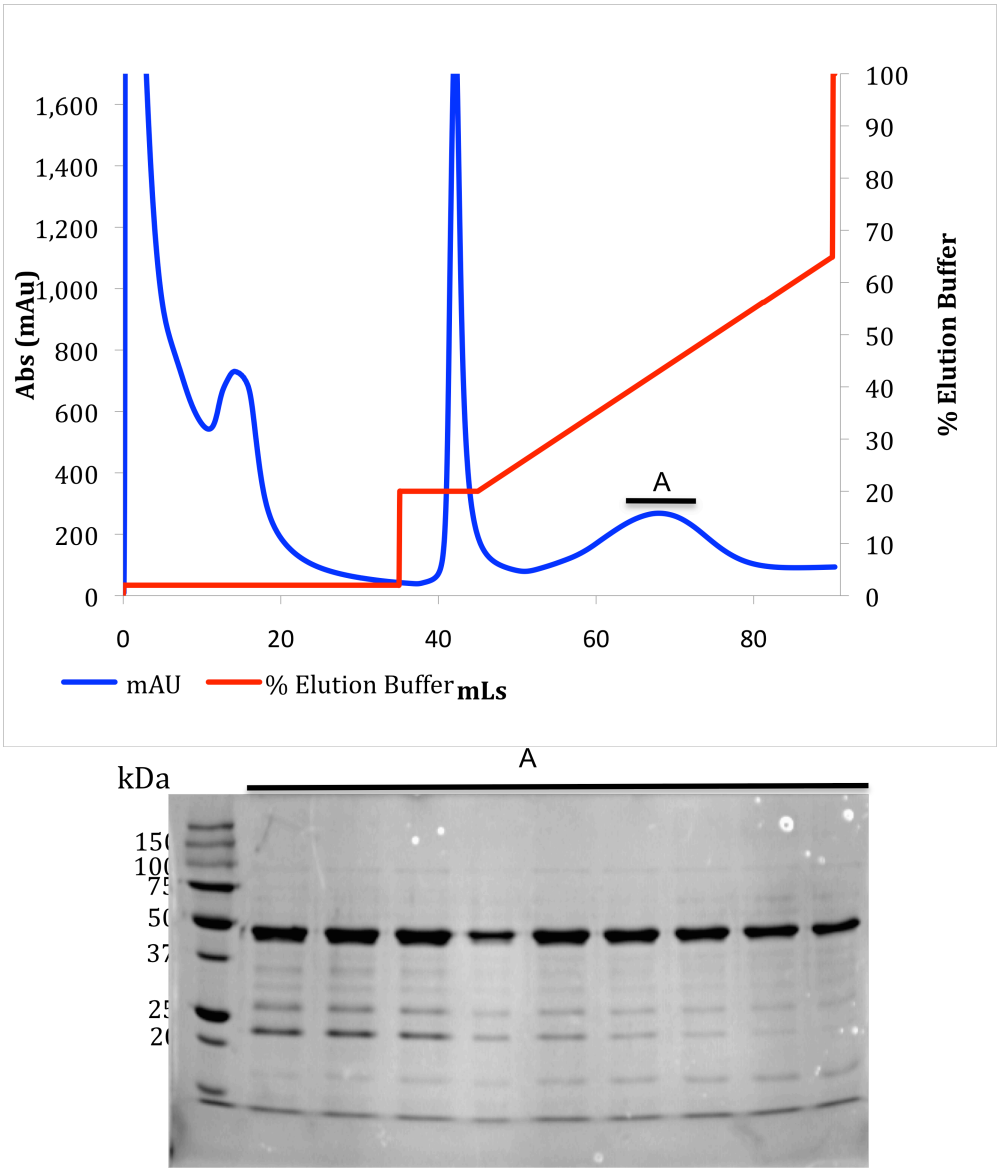


Figure 3-19: IMAC trace for Est2A including SDS-PAGE gel of appropriate fractions.

Previous purification of this enzyme indicated that peak A from Figure 3-19 contained Est2A hence only these fractions were analysed by SDS-PAGE.

Fractions from peak A in Figure 3-19 were pooled and further purified by size exclusion chromatography with an S200 16/60 gel filtration column. Est2A elutes in a single peak at ~65 mL (Figure 3-20), this corresponds to a M_r of 195 kDa,

indicating Est2A elutes as a tetramer.

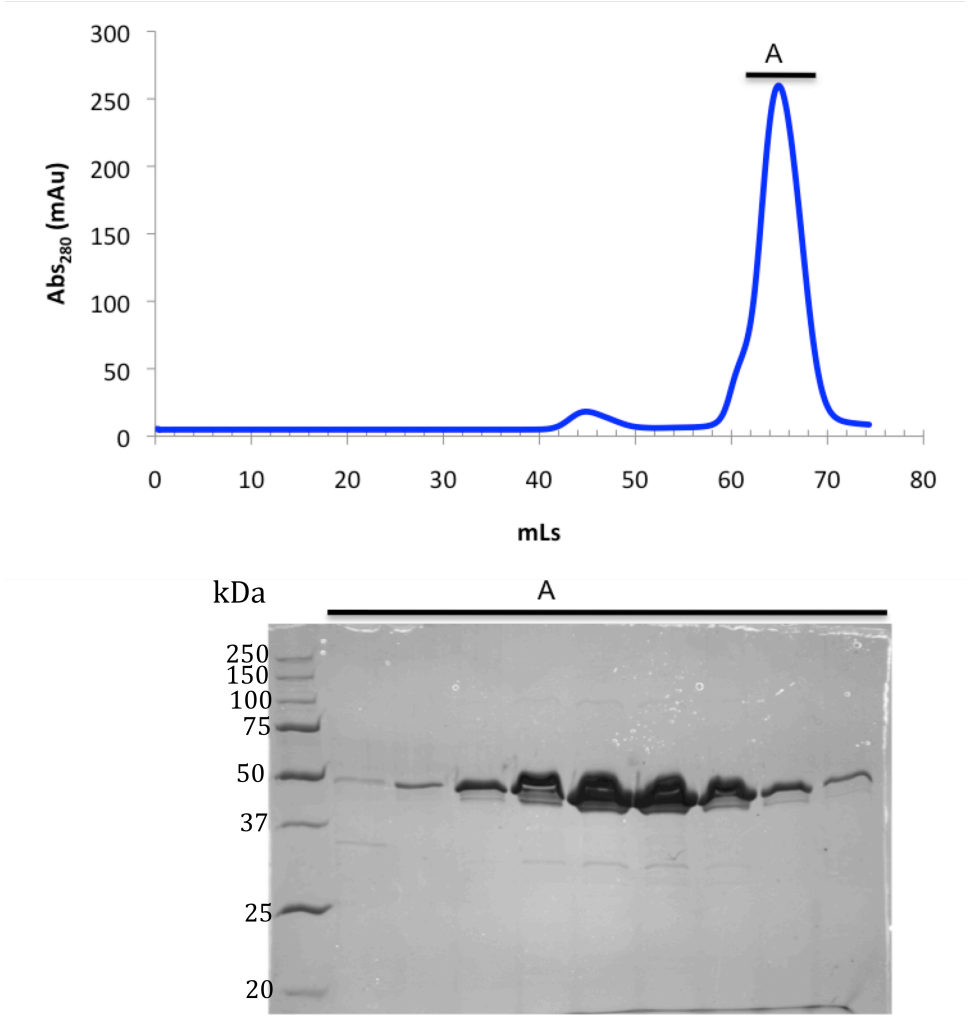


Figure 3-20: S200 16/60 chromatogram of Est2A with corresponding 12% SDS-PAGE gel. The protein samples have run in an abnormal fashion potentially due to the age of the gel.

3.3.2 Kinetic Parameters for Est2A with Model Substrates

The kinetic parameters of Est2A were determined with two model substrates, p-NPA and p-NPB. The results for Est2A with p-NPA and p-NPB are graphed in Figure 3-21 and Figure 3-22 respectively. The data indicates Est2A follows Michaelis-Menten kinetics with both substrate when analysed using Prism Graphpad software to give

the kinetic parameters V_{max}, K_M and k_{cat} (Table 3-7).

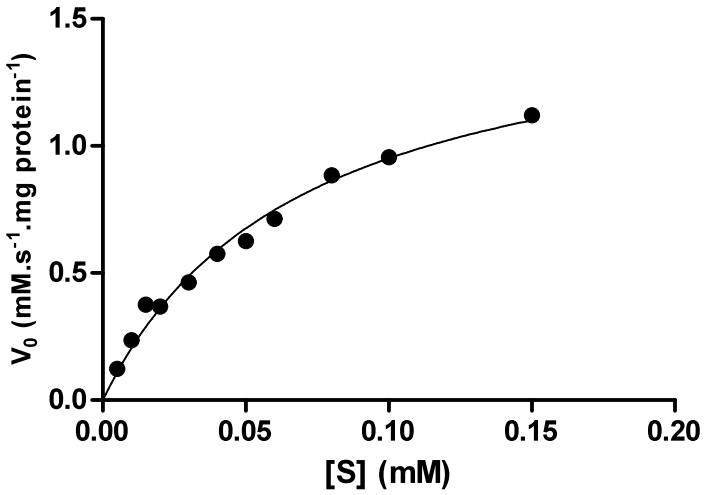


Figure 3-21: Rate of Est2A with *p*-nitrophenyl acetate at varying substrate concentrations. Errors are standard errors.

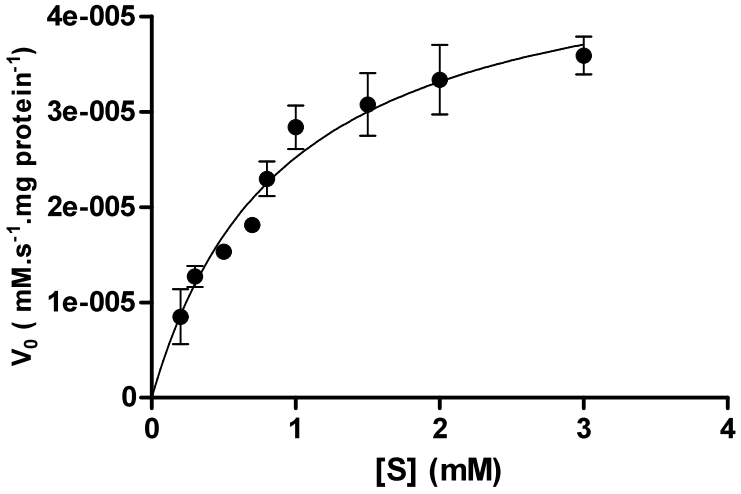


Figure 3-22: Rate of Est2A with *p*-nitrophenyl butyrate at varying concentrations. Errors are standard errors.

Comparison of the two graphs shows that at substrate concentrations 10 fold smaller the rate of degradation of *p*-NPA is two orders of magnitude faster than degradation

of *p*-NPB by Est2A.

Table 3-7: Kinetic parameters for Est2A with model substrates *p*-nitrophenyl acetate and *p*-nitrophenyl butyrate. Errors are standard errors.

Est2A	V_{max} (mM.s^{-1})	K_M (mM)	k_{cat} (s^{-1})	k_{cat} /K_M (s^{-1}.mM^{-1})
p-Nitrophenyl Acetate	3.5 ± 0.1 x10^{-3}	6.9 ± 0.6 x10^{-2}	139 ± 6	2.01 x 10^{3}
p-Nitrophenyl Butyrate	5.2 ± 0.4 x 10^{-5}	0.9 ± 0.2	2.1 ± 0.2	2.24

The kinetic parameters for Est2A with *p*-NPA and *p*-NPB reveal that the V_{max} for *p*-NPA is two orders of magnitude faster than for *p*-NPB. The affinity of Est2A for the *p*-NPA substrate is much better than that for *p*-NPB indicated by the K_M being over 10 fold smaller for *p*-NPA than *p*-NPB. The catalytic constant (k_{cat}) for Est2A with *p*-NPA is also significantly higher at 138.8 s^{-1} than for *p*-NPB at 2.1 s^{-1}. These results support the annotation of Est2A as an acetyl xylan esterase as all the kinetic parameters for the acetate substrate (*p*-NPA) are improved relative to those for the larger butyrate substrate (*p*-NPB). *p*-NPA is a closer representation of the natural substrate, acetyl side chains, than *p*-NPB. The four carbon chain of *p*-NPB attached to the *p*-nitrophenol ring causes a reduction of the enzymatic efficiency relative to the *p*-NPA. This indicates that the four carbon butyrate group is not accommodated by the active site as readily as the acetate group, which only has a two carbon chain.

3.3.3 Kinetic Activity of Est2A Compared with Other Acetyl Esterases

The specific activity of Est2A with the model substrate *p*-NPA is 194.4 μmol.min^{-1}.mg^{-1} and for *p*-NPB it is 2.91 μmol.min^{-1}.mg^{-1}. These data, along with the specific activity of a range of acetyl exterases, are presented in Table 3-8.

Table 3-8: Specific activity of a range of acetyl esterases from various sources with the model substrates *p*-nitrophenyl acetate (*p*-NPA) and *p*-nitrophenyl butyrate (*p*-NPB).

Enzyme	**Organism**	**Substrate**	**Specific activity** ($\mu mol.min^{-1}.mg^{-1}$)
YesT	*Bacillus subtilis*[119]	*p*-NPA	1580
AXE	*Bacillus pumilus*[130]	*p*-NPA	32
BhRGAE	*Bacillus halodurans* C-125[120]	*p*-NPA	15.43
acetyl esterase	*Streptomyces sp*. PC22[131]	*p*-NPA	9.8
ragi acetic acid esterase	*Eleusine coracana*, Indaf-15[121]	*p*-NPA	5.39
AxeA	*Streptomyces lividans*[132]	*p*-NPA	0
ReVvAXE1	*Volvariella volvacea*[133]	*p*-NPA	0
Est2A	*B. proteoclasticus*	*p*-NPA	194.4
Est2A	*B. proteoclasticus*	*p*-NPB	2.91
EstA3	derived from a drinking water metagenome[134]	*p*-NPB	513.6
EstCE1	derived from a soil metagenome[134]	*p*-NPB	31.1

The data indicate that Est2A has a significantly higher catalytic efficiency with *p*-NPA than all but one of the enzymes presented in Table 3-8. The specific activity of Est2A compared with the other esterase enzymes shows it to be relatively inefficient on *p*-NPB. It is clear that among the acetyl xylan esterases there is a large variation in activity, however it is important to note that not all reaction conditions were identical so comparisons between enzymes may be used as an indication only.

The structure of Est2A indicated that the catalytic mechanism relies on a catalytic dyad consisting of Ser-142 and His-351. Two individual amino acid substitutions were made in the Est2A gene; Ser-142 to Ala and His-351 to Ala, to test this proposed mechanism.

3.3.4 Expression and Purification of Est2A S142A and Est2A H351A

Both Est2A mutants expressed and purified in a similar manner. The purification of Est2A H351A is presented below as a representation of both mutants. The mutants eluted from the column with ~150 mM imidazole (Figure 3-23) and this was similar to the wildtype enzyme (Figure 3-19).

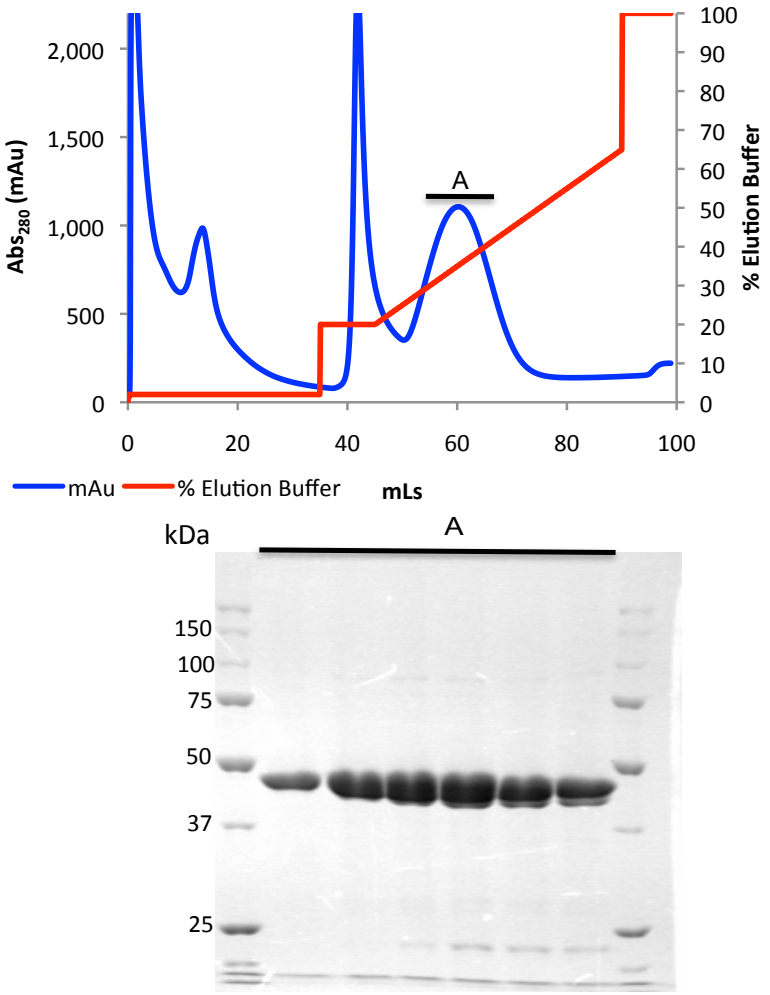


Figure 3-23: IMAC chromatogram of Est2A H351A with corresponding 12% SDS-PAGE gel.

The protein from peak A in Figure 3-23 was run through an S200 16/60 size exclusion column, the resulting chromatogram and corresponding 12% SDS-PAGE gel show the Est2A mutants elute from the column at ~65 mL (Figure 4-14). This is equivalent to the position the native enzyme elutes and correlates with a M_r of 179 kDa, indicating that the mutants are also tetramers in solution.

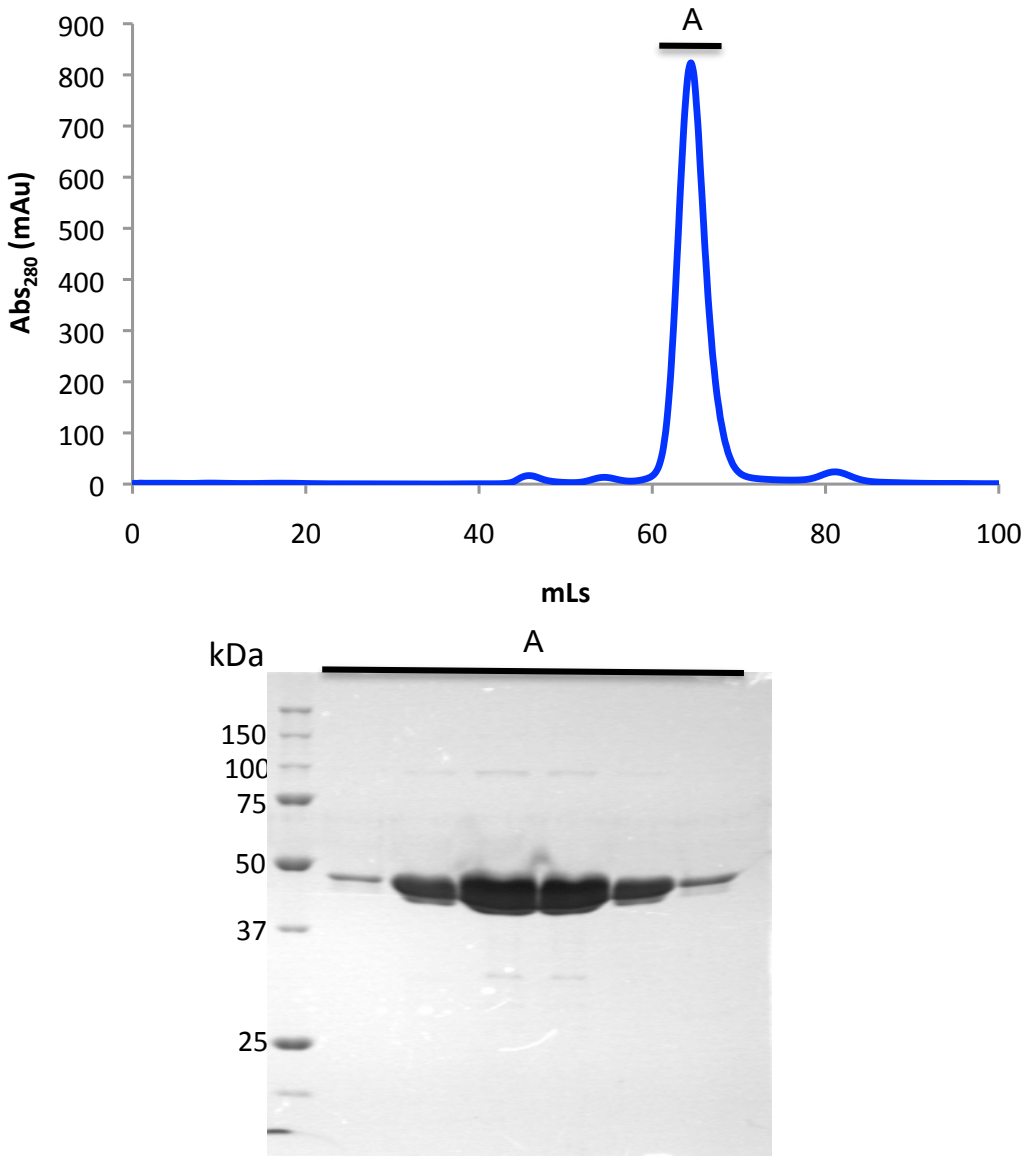


Figure 3-24: S200 16/60 chromatogram of Est2A H351A with 12% SDS-PAGE gel.

Fractions from peak A in Figure 4-14 were pooled and used for activity assays and crystallisation trials. Both mutants of Est2A purified in significant quantities and gave no indication of being unstable.

3.3.5 Activity of Est2A S142A and Est2A H351A

Both Est2A S142A and Est2A H351A were tested against the model substrate

p-NPA; both showed no activity above the blank indicating that both mutants are catalytically inactive (Table 3-9). This supports the proposed mechanism involving His-351 and Ser-142 as the catalytic dyad.

Table 3-9: Degradation of *p*-NPA with Est2A S142A, Est2A H351A and with no enzyme.

	Rate (x 10^{-5} $\Delta Abs_{400}.s^{-1}$)
Blank	7.5 ± 0.3
Est2A S142A	6.5 ± 0.5
Est2A H351A	7.3 ± 0.4

3.3.6 Crystallisation of Est2A H351A

Crystals of the Est2A mutants were laid down in fine screens around the condition that gave optimal crystals of Est2A (8% (w/v) PEG 4000, 0.1 M sodium acetate pH 4.6). Optimal crystals of Est2A H351A (Figure 3-25) were grown in 16% (w/v) PEG 4000, 0.1 M sodium acetate pH 4.6 with 1% w/w arabinotetraacetate as a substrate.

Figure 3-25: Est2A H351A crystals.

One dataset to 2.0 Å resolution was collected from a single crystal of Est2A H351A at the Australian synchrotron. The structure was solved using molecular replacement with native Est2A as a model. The final *R-factor* and R_{free} were 16.9% and 22.4%

respectively. Ramachandran analysis of the Est2A H351A structure indicated 97.3% of residues were in favoured conformations and 2.7% were in allowed regions. Full data collection and refinement statistics are given in Table 3-10 and Table 3-11 respectively. The RMSD between C-α for Est2A and Est2A H351A is 0.2 Å. Schematic representation of the Est2A H351 structure can be seen in Figure 3-26.

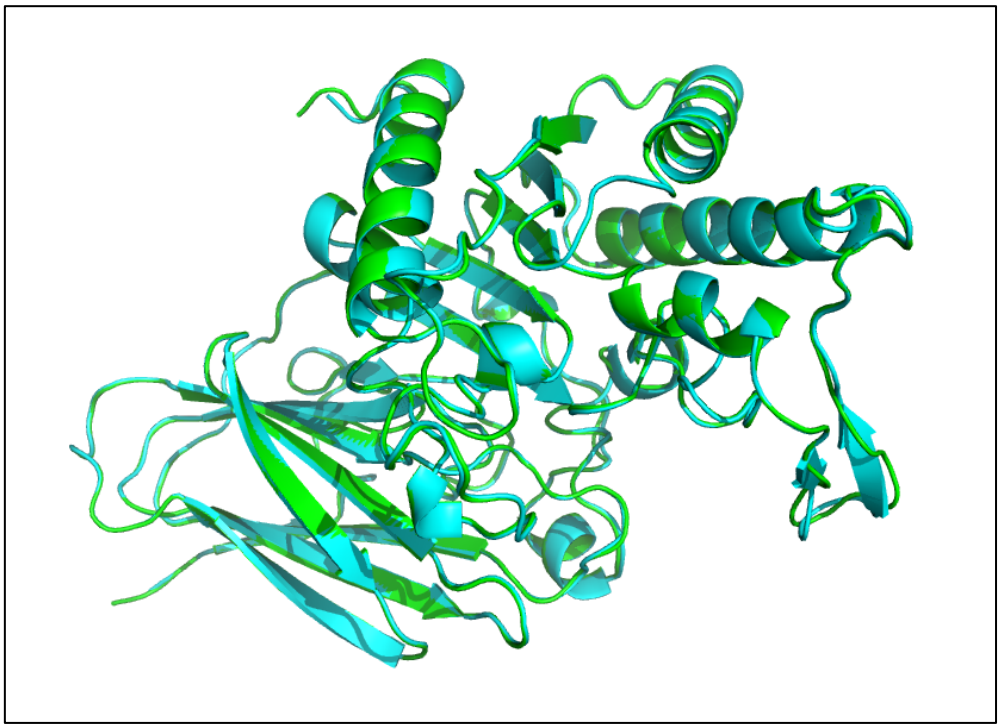


Figure 3-26: Overlaid structure of Est2A (green) and Est2A H351A (aqua) illustrates they are essentially identical.

Investigation of the electron density showed areas in the active site of five of the eight monomers that indicated potential presence of a ligand. The density was consistent with a molecule of PEG as shown in Figure 3-27. The active site of Est2A H351A also contained an acetate ion analogous to that in the structure of Est2A.

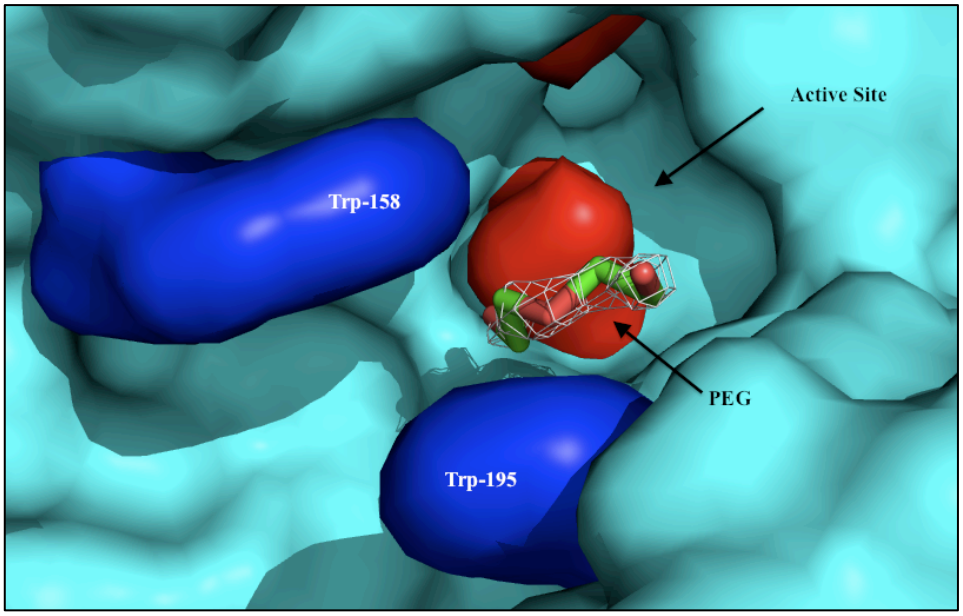


Figure 3-27: Active site of Est2A H351A occupied by a molecule of PEG. Electron Density is 2|Fo|-|Fc|map contoured at 1.0 σ.

By overlaying Est2A H351A with the structure of *Ct*CE2 from *Clostridium thermocellum* with cellohexaose (PDB accession code 2WAO) substrate binding interactions can be predicted. In *Ct*CE2 the residues involved in substrate binding and positioning are Trp-309, Tyr-184 and Trp-265[135]. The Est2A H351A structure has Trp-195 in the same position as Tyr-184, a methionine is in the corresponding position to Trp-309. There is no analogous residue at the position of Trp-265. Trp-158 in Est2A H351A is positioned on the opposite side of the substrate as Trp-195 and both are in positions that would allow them to interact with the substrate.

Table 3-10: Full data collection statistics for Est2A H351A.

Est2A H351A	Native Data
Space group	*P1*
Wavelength	0.9537 Å
Cell parameters	
a	90.11 Å
b	95.79 Å
c	98.79 Å
α	89.96°
β	99.74°
γ	92.50°
Resolution range (Å)	88.73-2.0 (2.11-2.0)
R_{merge}	0.161 (0.351)
No. of measured reflections	829350 (118180)
No. of Unique reflections	215871 (31205)
Mean I/σI	6.0 (3.1)
Completeness	98.2 (97.3)
Multiplicity	3.8 (3.8)

Table 3-11: Complete refinement statistics for Est2A H351A.

Refinement Statistics	Est2A H351A
RWork (*Rfree*)	16.9% (22.4%)
Total No. Atoms	25490
No. of Protein Atoms	23571
Other Molecules/Ions	35
No. of Waters	1760
RMS Deviation from Standard Geometry	
Bond Lengths (Å)	0.0243
Bond Angles (°)	1.9969
Average B-factors	
Protein	23.669
Water	29.541

3.4 Conclusion

The removal of side chains from hemicellulose increases the access of enzymes to the backbone sugars. This in turn increases the digestibility of the hemicellulose into single sugar units that can then be utilised as a fuel source by the animal

digesting it or used in industrial applications as a feedstock for other possesses.

Est1E is a feruloyl esterase that utilises a catalytic triad common to enzymes with α/ß hydrolase fold. Prior work showed Est1E can release a range of phenolic acids from hemicellulose as shown in assays with natural substrates presented in Table 3-1. I have shown that Est1E can work on a range of *p*-nitrophenol substituted model substrates with side chains varying from two carbons to sixteen carbons. The kinetic parameters for the two smaller substrates, *p*-NPA and *p*-NPB, are similar indicating that neither of the two substrates is preferred over the other.

The activity and structure of the mutant Est1E S105A is also presented here. The data show that this mutant is catalytically inactive which supports the proposed catalytic triad. There is a structure of Est1E with ferulic acid, a product of cleavage, bound in the active site. The structure of the mutant co-crystallised with ferulic arabinose did not appear to have substrate incorporated into the active site or the phosphate ion as seen in the Est1E apo- structure, this indicates that Ser-105 may also act in binding the substrate.

The activity of the acetyl xylan esterase Est2A is also presented. The kinetic parameters of Est2A with the two model substrates *p*-NPA and *p*-NPB clearly indicate that it cleaves the ester link of the acetate more readily than the butyrate substrate, this supports its annotation as a specific acetyl esterase. The proposed mechanism of Est2A includes a catalytic dyad; individual mutations of the two residues involved His-351 and Ser-142 to alanine both yielded inactive enzymes, supporting this mechanism. The structure of one mutant, Est2A H351A, is also presented here. Unfortunately with an acetate ion and a molecule of PEG bound in the active site instead of the substrate, the substrate binding interactions could not be indentified.

For *B. proteoclasticus* to completely degrade hemicellulose it requires a complex cocktail of enzymes; as well as esterases such as Est1E and Est2A it requires

enzymes to cleave the bonds between the sugar unit of the hemicellulose backbone and branching side groups. To gain further insight into the fibre degrading enzymes produced by *B. proteoclasticus* a preliminary investigation of nine enzymes was conducted. The aim was to express, purify and characterise a more diverse range of these enzymes. The results of this preliminary investigation are presented in chapter four.

Chapter Four: Preliminary Structural and Functional Investigation of Fibre Degrading Enzymes from *Butyrivibrio proteoclasticus*.

4.1 Introduction

B. proteoclasticus is a bacterium originally isolated from the rumen of New Zealand cattle[34]. Its genome contains a broad collection of genes encoding enzymes involved in the degradation and reassembly of polysaccharides. The enzymes involved in polysaccharide degradation are of interest for two reasons: (1) An increased understanding of their mechanism will allow us to gain insight into the workings of the highly complex rumen ecosystem and (2) to allow the potential application of these enzymes for the degradation of plant polysaccharides to simple sugars for use as feed-stocks for other processes such as the production of bio-fuels.

4.1.1 Previous work on Enzymes from *Butyrivibrio proteoclasticus*

Prior to the start of this project, Dr Dave Goldstone (University of Auckland) had PCR amplified 44 of the genes from *B. proteoclasticus* annotated as being involved in fibre degradation and cloned these into expression vectors. The genes were cloned from *B. proteoclasticus* B316 genomic DNA into pDEST17 Gateway expression vectors. All the enzymes tested in this chapter are from the suite of 44 previously cloned genes.

Microarray analysis of *B. proteoclasticus* grown on xylan revealed a suite of genes that were upregulated compared with growth on xylose (Attwood, personal comm.). *B. proteoclasticus* invests considerable energy in producing increased quantities of these enzymes when it is utilising xylan as a food source, presumably because they are important in the degradation of xylan and this makes them ideal targets as potential components of a hemicellulose degrading cocktail.

4.1.2 Work Presented in this Chapter

B. proteoclasticus is capable of growing on xylan, a hemicellulose composed of a backbone of xylose units with various side groups, hence it is necessary for the bacterium to produce a wide range of fibre degrading enzymes. In order to gather useful information about how this is achieved it is important to take a systematic approach to the enzymes targeted for investigation. In the complete degradation of hemicellulose the branching side chains need to be removed then the backbone degraded into single sugars. Est1E and Est2A (chapter three) are capable of removing phenolic acid and acetyl side groups respectively. Once these are removed arabinose side chains need to be cleaved by arabinofuranosidase enzymes; this increases enzymatic access to the xylan backbone. Cleavage of the backbone requires both endo- and exo- acting xylanase enzymes. The approach taken for this work involved investigation of enzymes with a range of annotated functions necessary for complete polysaccharide degradation, which, once characterised, would allow the creation of a cocktail that can facilitate this. Work presented in this chapter is aimed at preliminary characterisation of a representative group of arabinofuranosidases, xylanases and glucanases that are produced by *B. proteoclasticus*.

4.2 Cleaving arabinose side-chains in Xylan

Xylan chains have branching sidegroups, such as arabinose, prohibit the degradation of the xylan backbone. Removal of these side groups increases the biodegradability of the overall xylan; therefore an enzyme cocktail that can completely degrade xylan to its constituent sugars requires an arabinofuranosidase.

4.2.1 Selection of Arabinofurnaosidases for Investigation

The genome of *B. proteoclasticus* contains ten genes encoding enzymes that contain a GH43 family domain. This domain can have a range of activities typically involved in cleavage of arabinose and xylose units from various hemicelluloses; their

annotations indicate they have arabinofuranosidase/xylosidase activity. This seems to be a large number of enzymes that all have similar functions therefore a representative selection of these were chosen for investigation.

Four enzymes that contain GH43 domains were chosen for investigation, Xsa43A, Xsa43B, Xsa43C and Xsa43E. Sequence alignments of the GH43 domains of these enzymes can be seen in Figure 4-1.

```
                    1        10        20        30        40        50        60
                    |        |         |         |         |         |         |
Xsa43B_GH43Domain   -------------EDFYRN-PIRSGFFPDPSICRVGDDYYMANSS----------FIFFP
Xsa43C_GH43Domain   ----------------------------PD-------EGYYVVRT----------------
   Xsa43E_protein   -----MTETMTEEKTFHKNPIIKDIYTADPAPMVYGDTLYLYTTHDE-------DELVND
Xsa43A_GH43Domain   EIFQDISVQKSAKPITDHNPVMTQRYGADPYALVYNGRVYLYMTGDKPSYNADGSVKENT

Xsa43B_GH43Domain   CIPISHSKDLIHWEIIGHAITDPNWAGLDELDG--GRGYWAPDISYYNGKFYIAATYRLN
Xsa43C_GH43Domain   --------------------TDP-------------LGKWEDPVCILEGKGLIDSCPFWD
   Xsa43E_protein   FYTMNDWRCFSTKDMVNWTDHGAIFSLDDIGWA--DARAWAPQAVERNGKFYLYCPVHKR
Xsa43A_GH43Domain   YGNINTICVISSDDLVNWTDHGTVYAAGSKGAASWGPNSWAPAAAYKNIDGKDKFFLYFA

Xsa43B_GH43Domain   DTGPVYRRQIVVSSDKPQGPYSEPVFIDEDG---------IDPGLFNDDDGRRYMLLNRG
Xsa43C_GH43Domain   DDGKAY---------------------------------VIHGYAKSRIGFKSILG---
   Xsa43E_protein   NGGMAIAVGISDSPTGPFKDLGYPLVDEGDWN-------DIDPTVFIDDDGQAYLYFGNP
Xsa43A_GH43Domain   NNGNGIGVLSSDSPEGPFVDEIGGPLISRETPNCAEVTWLFDPAVLMDDGGSAYIYVGGG

Xsa43B_GH43Domain   ARIFELDKT---------GTRKISDATLLYYGSNK----RAP----EGPHIYKKDGYYYL
Xsa43C_GH43Domain   --IFEMSSD---------GLKAISEDHFIFDGNDP----KNPAVTIEGPKVYKRDGYYYI
   Xsa43E_protein   ELRYVLLNENMITYDKEVGIVKVPMTEEAFAKGSH----DTGTSYGEGPWFYKRNDLYYM
Xsa43A_GH43Domain   VPSPDKASNPGTARVCKLGEDMMSLDGDPIAIENVDYLFEDSGINKIGDTYYYSYCSNFN

Xsa43B_GH43Domain   LQAEGGTGPGHRVTVARSRELMGVYEPCPYNPIMRQDDPDALIQRCGHGDLVETQDGRWY
Xsa43C_GH43Domain   WAPAGGVKYGYQVVLRSR----DIHGPYEIKEVMHTG--NTVINGPHQGGLVDTVNGDEW
   Xsa43E_protein   VYAAFGVGKQNEHLAYSTSD--SPTGPWKYGGVLMTEEGGVFTNHPG----IADFKGHSY
Xsa43A_GH43Domain   VTPEATAELGFESGEIVTMKSDSPMGPFERCNAVLKNPERFFGLGGNNHHCMFEFNGEYY

Xsa43B_GH43Domain   MVYLCGRKIGDG-YSILGRETALDPVTWTADGWPIVNGLRGPSCMQIKPFPEKD
Xsa43C_GH43Domain   FIHFQDR-------GLYGRICHLQPVHWEN-GWPVVGVNADDNGVGEPVY----
   Xsa43E_protein   LFYHTGDLPGG---SLFHRSVCVAEFTYNE--DGTINPIPKCDGVEKIN-----
Xsa43A_GH43Domain   ITYHSRILEEKLGVNGGYRSTNVDKLVINDNNEPSKSTGTRAGVPQLKPFD---
```

Figure 4-1: Sequence alignment using ClustalW of four GH43 domains chosen for investigation.

The genes encoding Xsa43A and Xsa43E show upregulation in response to growth on xylan as opposed to xylose in microarray experiments, whereas the genes of Xsa43B and Xsa43C are not changed. Xsa43A is secreted by *B. proteoclasticus* whereas the other three are intracellular enzymes. Xsa43A also has a CBM6 domain, Xsa43B has

an ankyrin repeat, a GH43 domain and appears to have a third domain that shows no putative homologues in blast searches. Xsa43B also contains an additional domain that has no annotation. Xsa43E contains only the GH43 domain. These four enzymes are a good cross section of the ten GH43 enzymes produced by *B. proteoclasticus* and are summarised in Table 4-1.

Table 4-1: Summary of GH43 enzymes chosen for further investigation.

Enzyme	ORF	# Amino Acids	M_r (Da)	EC Number	Microarray (fold up-regulated)
Xsa43A	Bpr_I0302	536	57542	3.2.1.55	21.7
Xsa43B	Bpr_I1584	643	73406	3.2.1.37	0
Xsa43C	Bpr_I1585	571	64379	3.2.1.37	0
Xsa43E	Bpr_I2319	313	35279	3.2.1.37	9.1

4.2.2 Expression and Purification of Potential Arabinofuranosidases

4.2.2.1 Expression and Purification of Xsa43A

Xsa43A was expressed and purified as outlined in section 3.2.1. The IMAC purification trace and corresponding SDS-PAGE gel shows Xsa43A eluted from the IMAC column with ~150 mM imidazole (Figure 4-2).

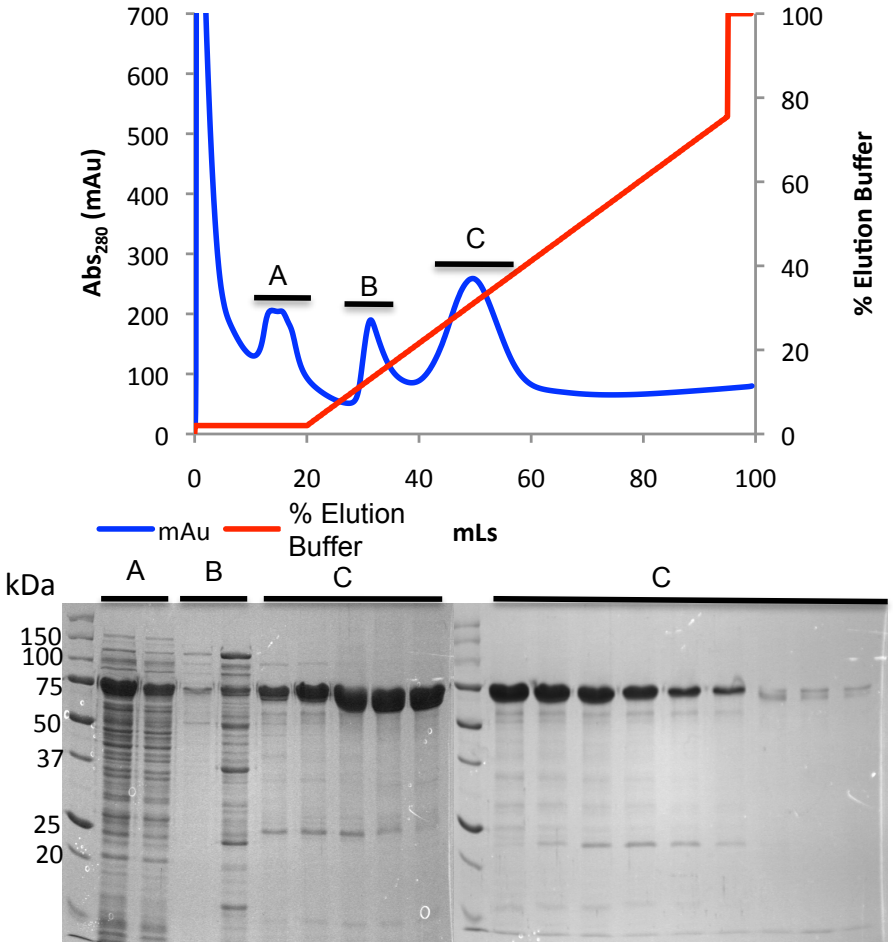


Figure 4-2: IMAC chromatogram for Xsa43A IMAC purification with 12% SDS-PAGE gel.

Fractions shown to contain Xsa43A from peak C in Figure 4-2 were purified further using an S200 16/60 size exclusion column. The chromatogram shows the majority of Xsa43A eluting from the column around 78 mL (Figure 4-3) with a small amount eluting around 40 mL. The elution at 78 mLs indicates the M_r is 68 kDa which is close to the calculated molecular mass of Xsa43A. Protein in Peak A appears to be eluting in the void volume of the column and thus is most likely an aggregate.

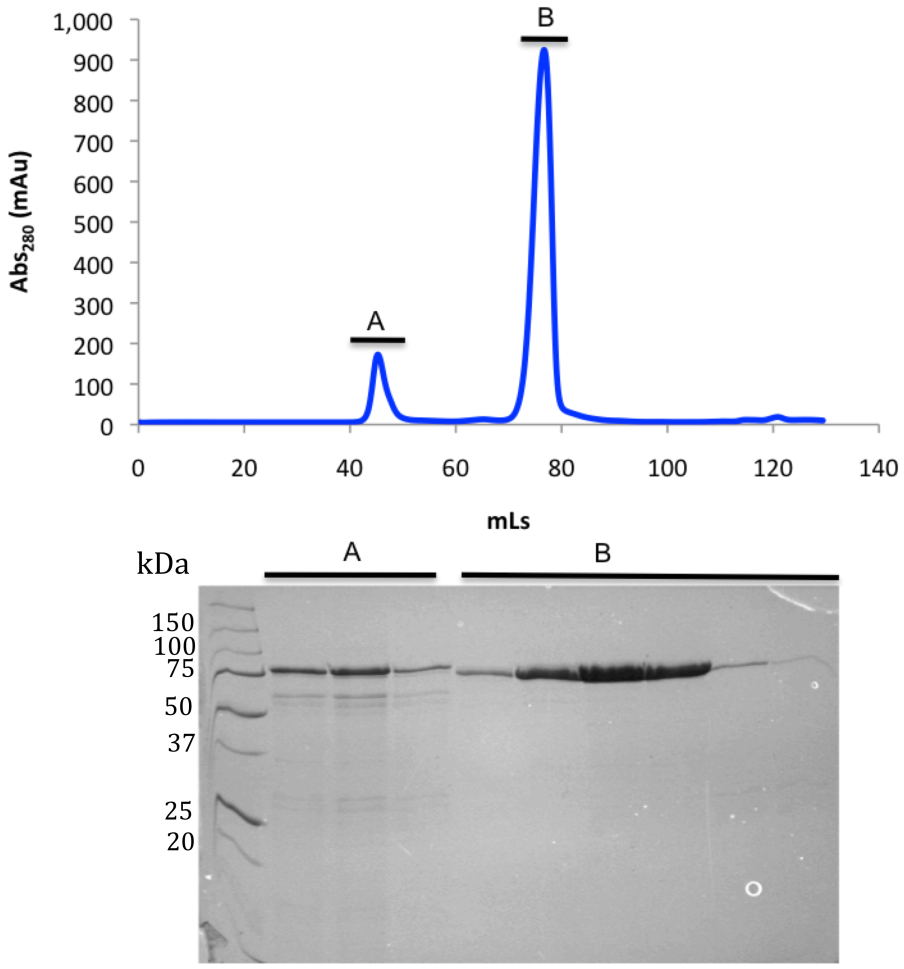


Figure 4-3: S200 16/60 chromatogram for Xsa43A with 12% SDS-PAGE gel.

4.2.2.2 Expression and Purification of Xsa43B

Xsa43B was expressed and purified as outlined in section 3.2.1. Xsa43B elutes from the IMAC column at ~150 mM imidazole (Figure 4-4).

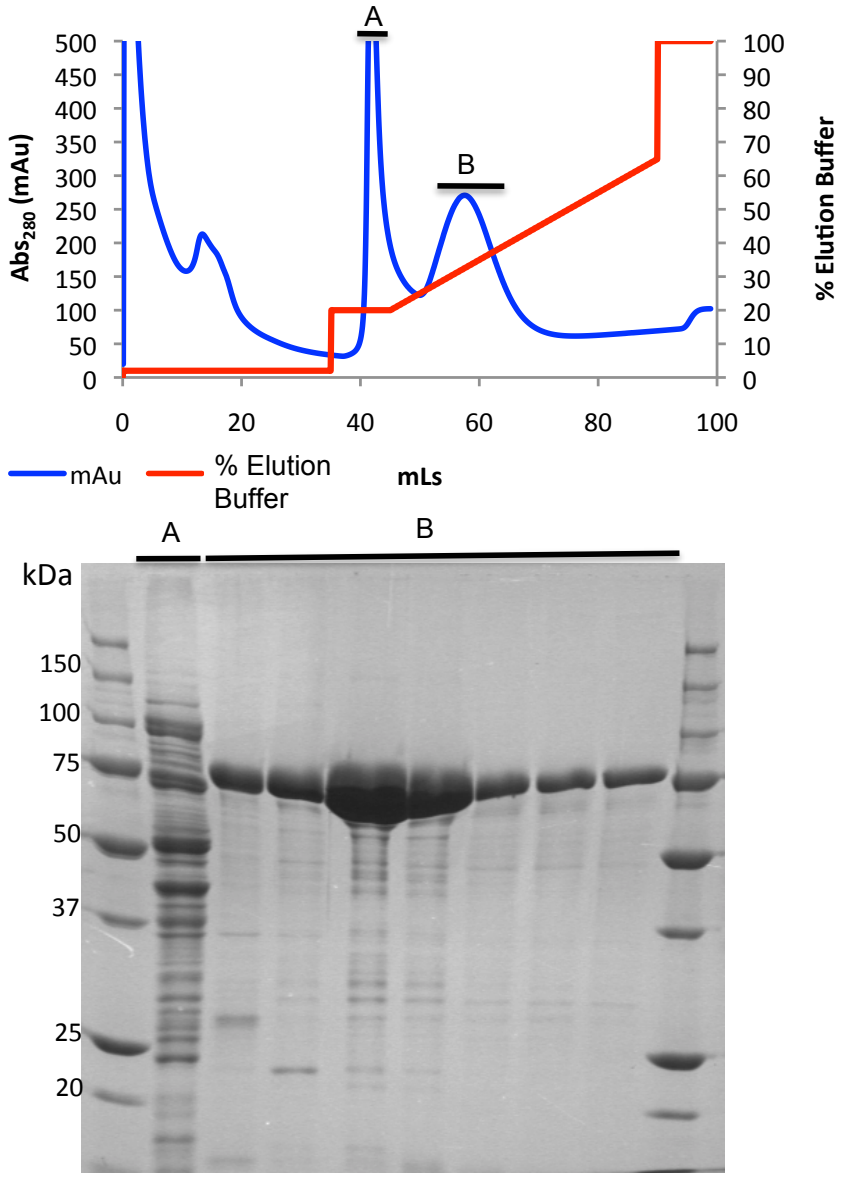


Figure 4-4: Xsa43B IMAC purification with corresponding 12% SDS-PAGE gel.

Fractions shown to contain Xsa43B from peak B in Figure 4-4 were purified through an S200 16/60 size exclusion column. The resulting chromatogram and 12% SDS-PAGE gel show the majority of Xsa43B elutes in a single peak between 60-65 mL (Figure 4-5). This elution volume corresponds to a protein of ~192 kDa, equating to 2.6 times the M_r of Xsa43B, indicating it is most likely a dimer but due to

the multiple domains may have a larger than expected hydrodynamic radius.

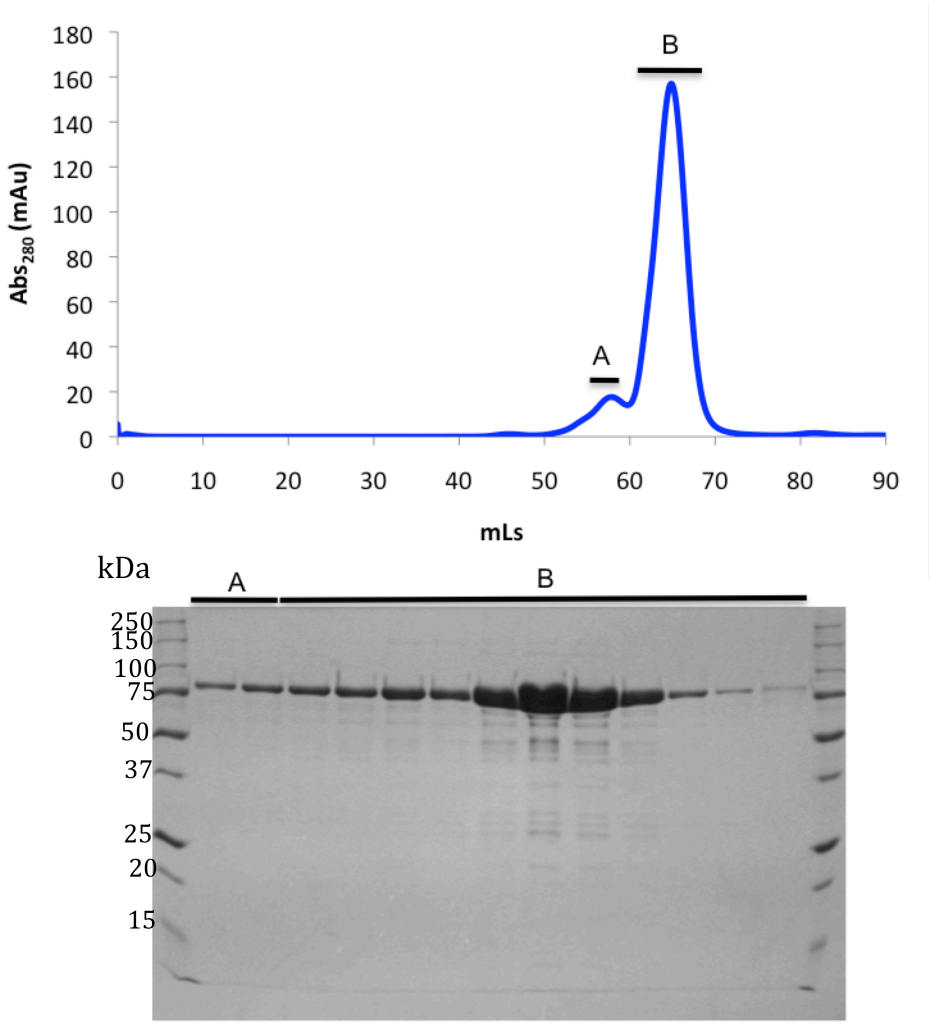


Figure 4-5: Size exclusion chromatogram of Xsa43B with the corresponding 12% SDS-PAGE gel.

4.2.2.3 Expression and Purification of Xsa43C

Xsa43C was expressed and purified as outlined above in section 4.2.2.4. The chromatogram and 12% SDS-PAGE gel for Xsa43C IMAC purification show Xsa43C elutes from the IMAC column with ~200 mM imidazole (Figure 4-6).

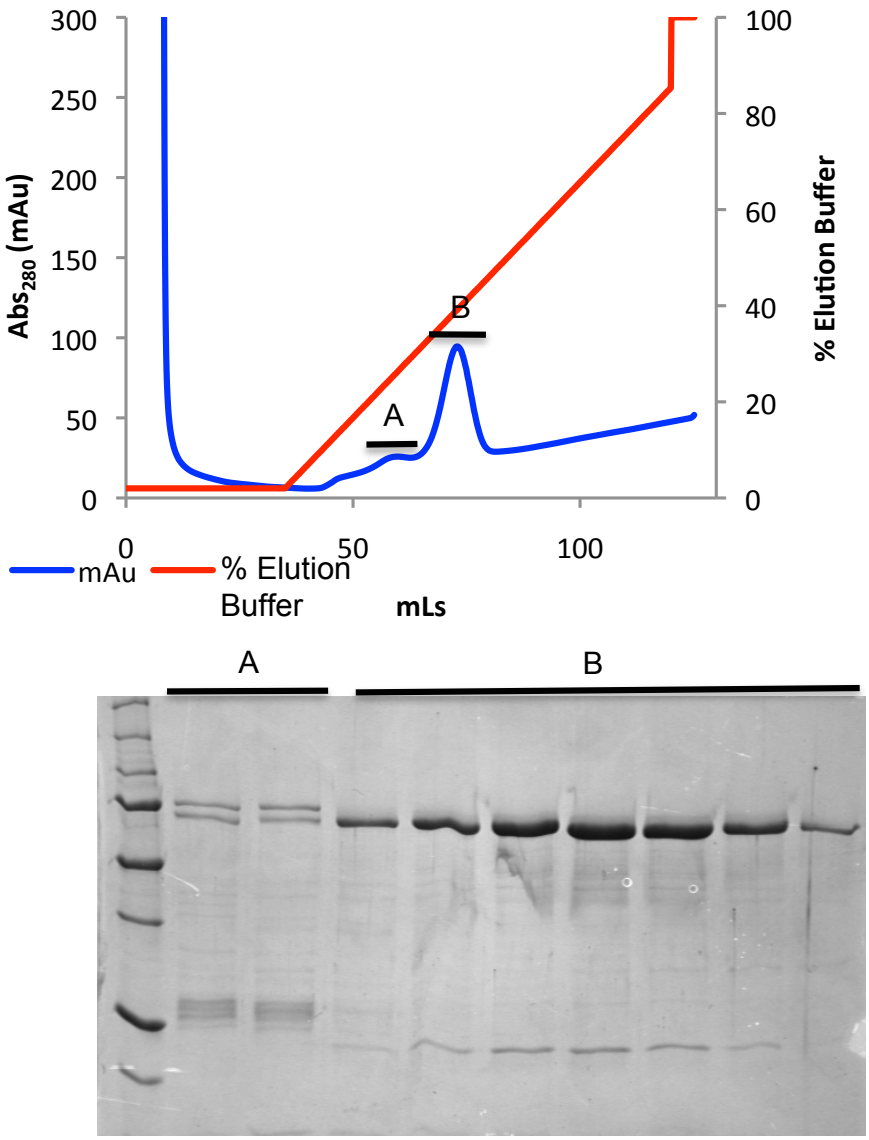


Figure 4-6: IMAC chromatogram of Xsa43C with corresponding 12% SDS-PAGE gel.

Fractions containing Xsa43C from peak B in Figure 4-6 were purified further using an S200 16/60 size exclusion column. The resulting chromatogram and the 12% SDS-PAGE gel show that a small amount of Xsa43E elutes around 65 mL but the majority of the protein elutes from the column at ~75 mL (Figure 4-7). The elution point corresponds to a M_r of 81 kDa, indicating Xsa43C elutes as a monomer.

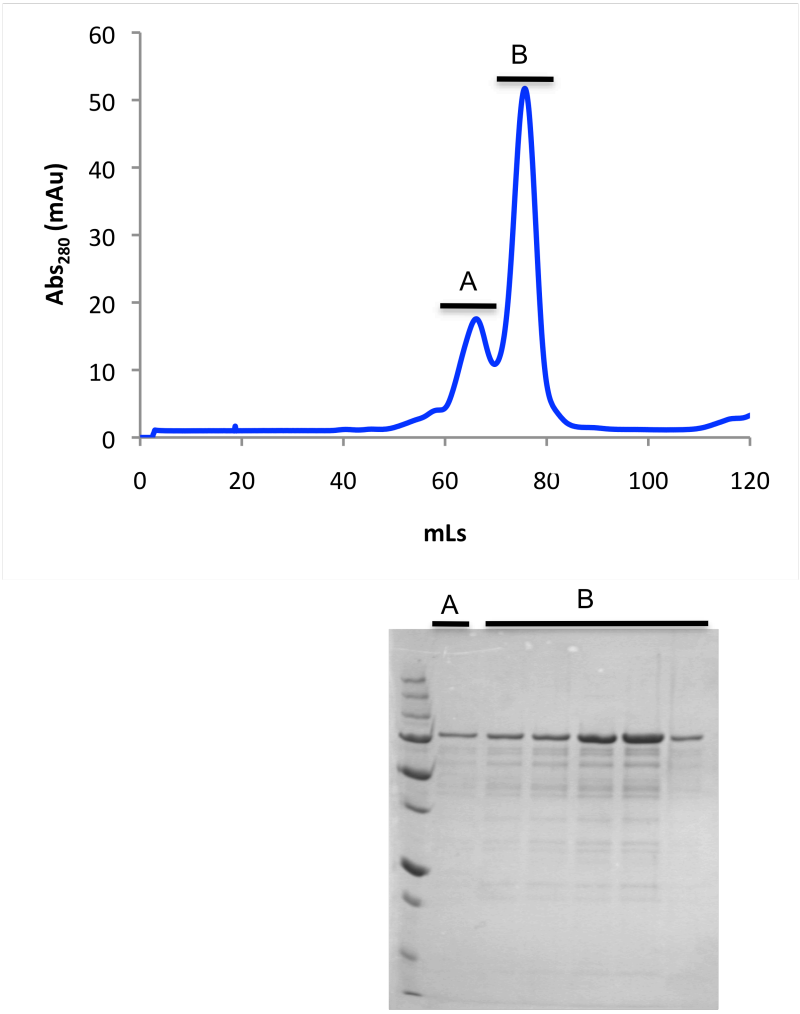


Figure 4-7: S200 16/60 chromatogram of Xsa43C with corresponding 12% SDS-PAGE gel.

4.2.2.4 Expression and Purification of Xsa43E

Xsa43E was expressed and the protein was purified as outlined in section 3.2.1. The IMAC elution profile showed that Xsa43E elutes from the column with ~150 mM imidazole (Figure 4-8).

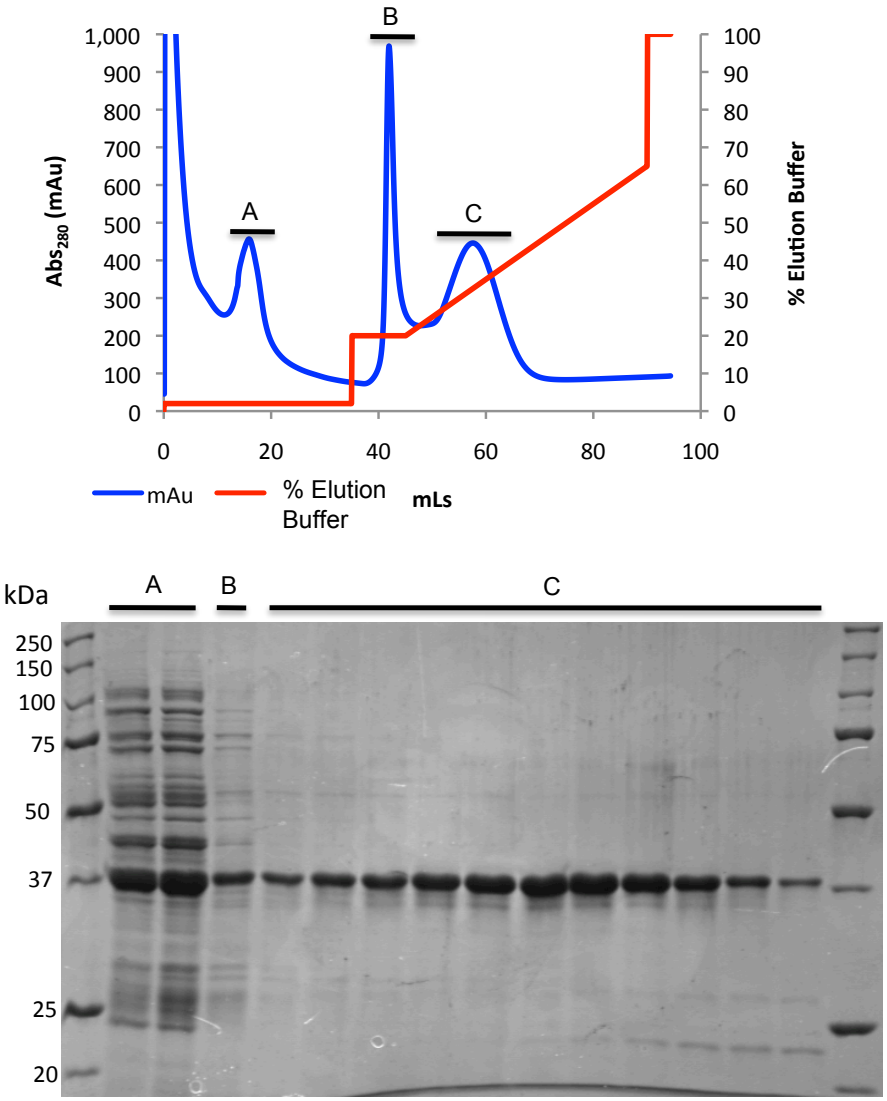


Figure 4-8: Xsa43E IMAC chromatogram with corresponding 12% SDS-PAGE gel of fractions.

There is a significant amount of Xsa43E remaining in the wash fractions potentially due to overloading of the column. Fractions of the peak C from Figure 4-8 were purified further by size exclusion on an S200 16/60 column (Figure 4-9).

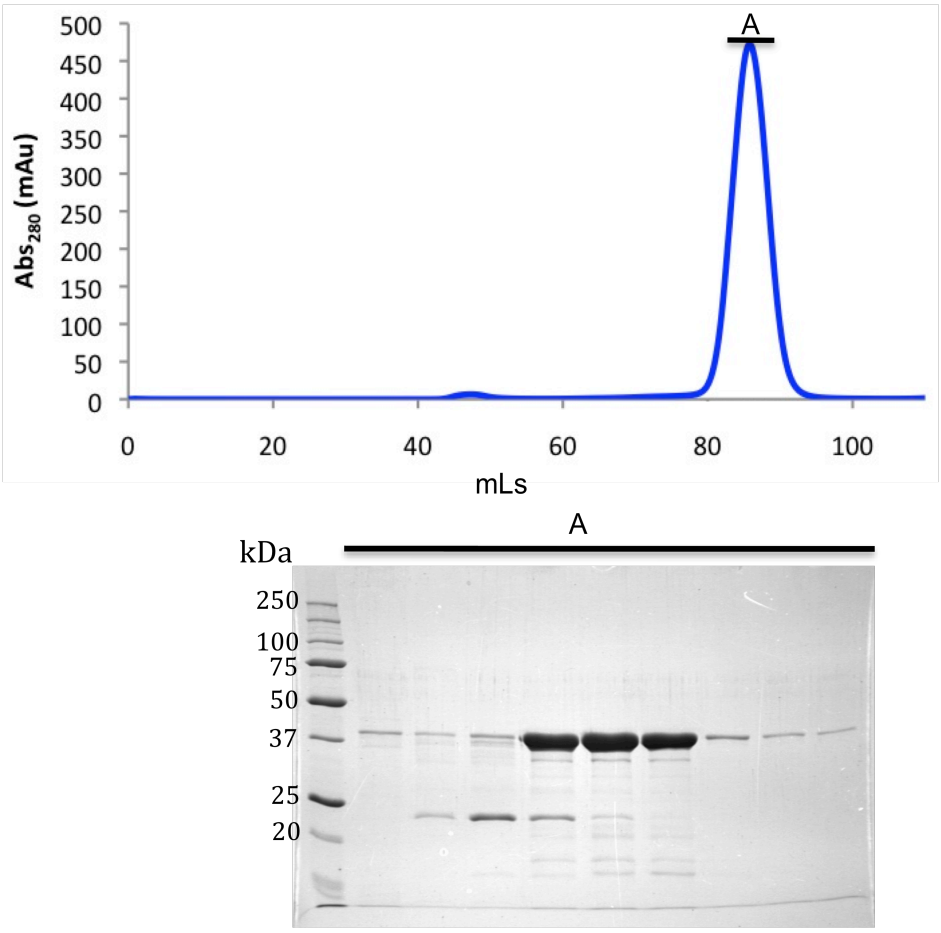


Figure 4-9: Size exclusion profile of Xsa43E protein with the 12% SDS-PAGE gel.

Xsa43E elutes in a single peak around 85 mL (Figure 4-9). This equates to a M_r of 36 kDa, which is consistent with Xsa43E being a monomer in solution.

4.2.3 Crystallisation

4.2.3.1 Crystallisation attempts

Attempts were made to crystallise three of these enzymes, Xsa43A, Xsa43B and Xsa43E. These attempts included crystallisation trials and fine screens as set out in section 2.3. Xsa43E grew the best crystals in a precipitant solution of 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2 (Figure 4-10). A fine screen was

performed around this condition.

Figure 4-10: Best crystals of Xsa43E from initial fine screen.

4.2.3.2 Crystal Optimisation

Fine screens of Xsa43E were performed around the conditions 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2; using the hanging drop method as outlined in section 2.8.1. Protein was laid down in crystallisation drops at a concentration of 60.6 mg.mL^{-1}. Crystals grew in all but one of the conditions after ten days. The best crystals were grown in 18% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2. These crystals scaled and integrated in the space group $P2_12_12_1$ with the unit cell dimensions a=58.3, b=76.4, c=84.1 Å, $\alpha=\beta=\gamma=90$.

Full elucidation of the structure and catalytic mechanism of Xsa43E is presented in Chapter Five.

4.2.4 Functional Characterisation of Selected GH43 Enzymes

Preliminary functional analysis of the four GH43 enzymes were performed by incubating each protein with four natural substrates: ryegrass, cellulose, hemicellulose (all prepared by Grasslands AgResearch) and arabinoxylan (Megazymes International Ltd). The reaction mixtures were analysed for the release of sugars using gas chromatography mass spectrometry (GCMS) as outlined in section 2.8.1. The comparison of the GCMS traces with external standards (Figure

4-11) allowed the identification of sugars being released but the quantity was below the range of the standards therefore the sugars could not be quantified. The retention time of the sugars can be seen in Appendix II.

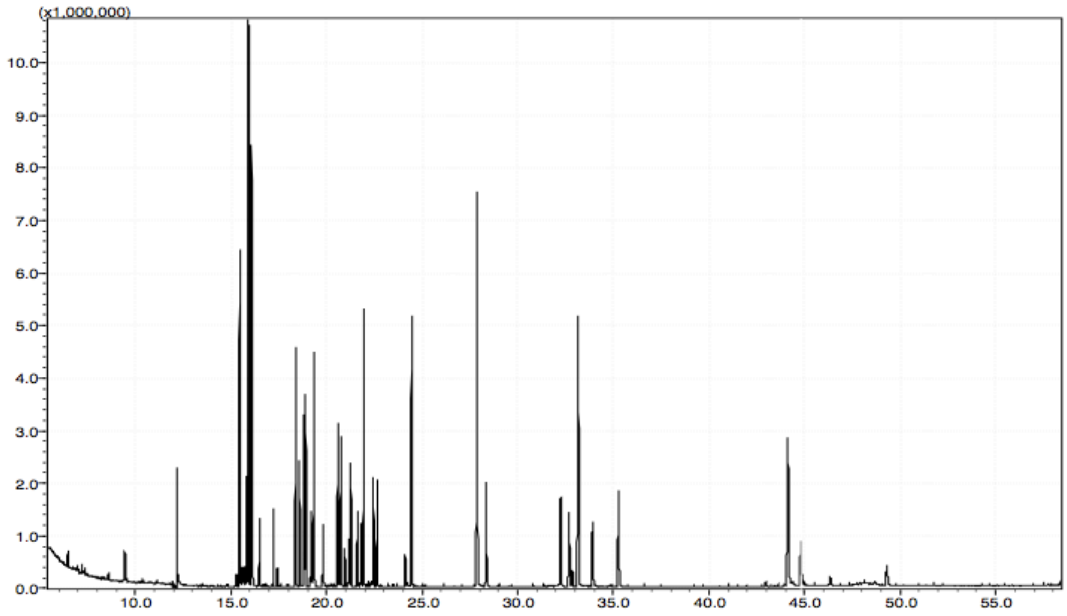


Figure 4-11: GCMS trace of external standards.

The sugars released by these enzymes are outlined in Table 4-2. Arabinose and xylose elute in overlapping peaks therefore if a peak was seen at that point the enzyme was said to release arabinose/xylose as the results cannot distinguish between the two. The GCMS trace of Xsa43E with arabinoxylan is shown in Figure 4-12 indicating the release of arabinose or xylose.

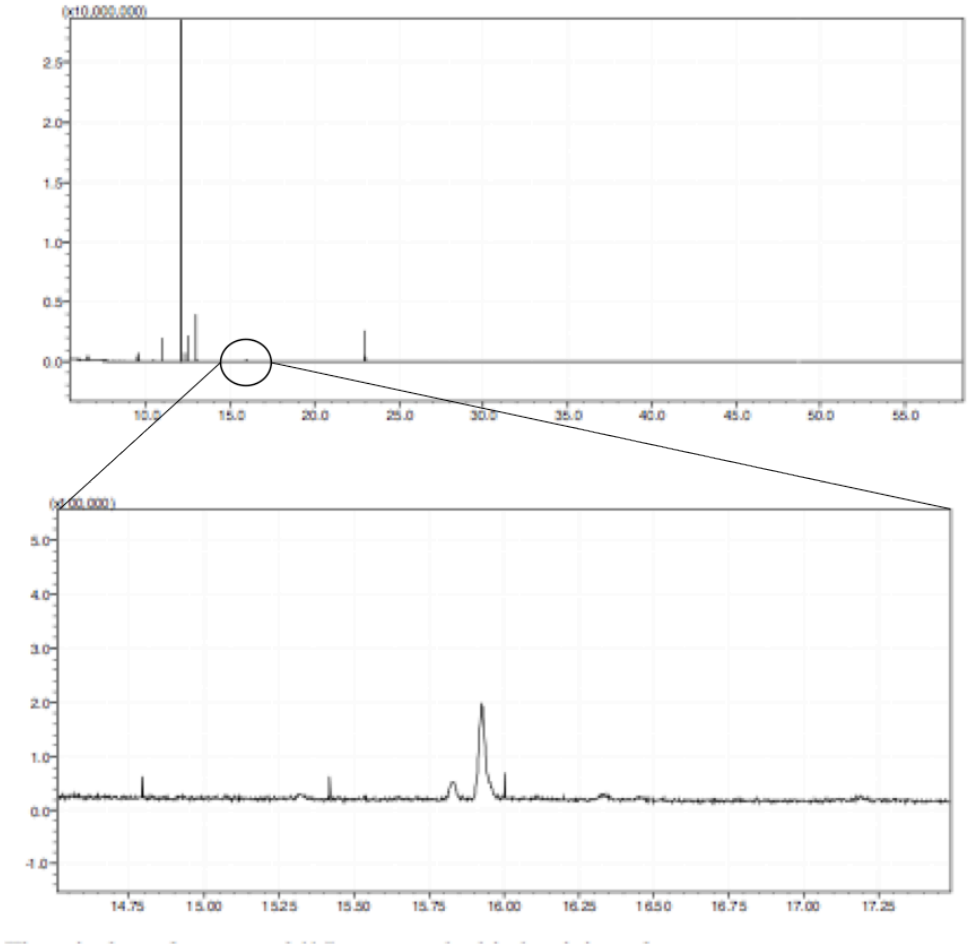


Figure 4-12: GCMS trace of sugars released from arabinoxylan by Xsa43E with a zoom in on the arabinose/xylose peak.

Table 4-2: Sugars released from natural substrates by GH43 enzymes.

Enzyme	Sugar(s) released from Substrates
Xsa43A	Glucose
Xsa43B	none
Xsa43C	Arabinose/xylose, Glucose
Xsa43E	Arabinose/xylose

All these enzymes are annotated as being a xylosidase or arabinofuranosidase and therefore the release of arabinose or xylose from the natural substrate would be expected. This was seen for Xsa43E and Xsa43C. The other two xylosidases or arabinofuranosidases, Xsa43A and Xsa43B, did not release arabinose or

xylose Xsa43C also released glucose, as did Xsa43A.

The four enzymes were also tested on two model substrates: Azo-CM-cellulose (Azo-CMC), which tests for endo-1,4-ß-D-glucanase activity (section 2.8.2) and Azo-wheat arabinoxylan (Azo-WAX), which tests for endo-1,4-ß-D-xylanase activity (section 2.8.3). None of the enzymes showed any activity on these model substrates.

Xsa43C and Xsa43E appear to be the best choices for further investigation as they released arabinose or xylose from the natural substrates. These enzymes will require further investigation to establish their suitability.

4.3 Cleavage of the Xylan backbone

The backbone of xylan is composed of ß-1,4-linked xylopyranose sugars. Once the side groups have been cleaved enzymatic degradation of this backbone becomes significantly more efficient. Endo- and exo-1,4-xylanases are responsible for the cleavage of the backbone into smaller chains.

4.3.1 Selection of Xylanase Enzymes

B. proteoclasticus produces six enzymes annotated as having a GH10 xylanase domain. Three of these enzymes were chosen for further investigation, Xyn10A, Xyn10D and Mxy10-43A. All three have a GH10 family domain, sequence alignments of the GH10 domains can be seen in Figure 4-13.

```
                        1        10        20        30        40        50        60
                        |        |         |         |         |         |         |
                Xyn10A  MYKQGRSLVAVMAASVMLTLTACGGTTDHQSAAVSEAESISATSDNSLNSTSTEESNKDV
   Xyn10D_GH10Domain    ------------------------------------------------MNLKTAYEP----
Mxy10-43A_GH10Domain    ---------------------------------------GNQGQVVEANLRNTYGAQ---

                Xyn10A  TDASSESSSSEQEGDEDNVPALRDCVEQKMGCRIGCAITGKEPWDIPLWNLVTTHFNAVT
   Xyn10D_GH10Domain    -------------------------------YFKMGAAISRWNLHTPAHTKLLTAQFNSFT
Mxy10-43A_GH10Domain    -------------------------------YGYSGTCINLYQLRDAGQLNFLKTHYNSIT

                Xyn10A  LGNELKPDSLFGYSNGKCPGTQEAELNGET------IMVPVLNFDNPEKILNKFVKWNEV
   Xyn10D_GH10Domain    CENDMKP---MYYLDKDANKKDPEKYN----------LSPALTFEN-------AIPYLEF
Mxy10-43A_GH10Domain    LENEMKPDALLGGSARLISVNDAKSRGYYIPDGYSEQYVPQINFNT-------VDEVMKI

                Xyn10A  HPDREIKVRGHVLVWHSQTPEWFFHQDYDKTKPYVSKEEMDKRQEWYIREVLTHFTGEDS
   Xyn10D_GH10Domain    AKDNKIAMRGHTLVWHNQTPKWFFCERYNENFPMADRETILARLESYIHGVLDFVQTN--
Mxy10-43A_GH10Domain    CYQNGLKMRAHTLVWHSQTPSWFFRNGYSGNGGFVNQSTMDKRLEMYVKTVMNHVYNS--

                Xyn10A  PYKDLFYGWDVVNEAVSDATGTYRTDNENPNEDLSNDTHGNNSSWWHVYQSEEFIINAFK
   Xyn10D_GH10Domain    -YPGIIYAWDVVNEIVDEG--------------------AFRKSIWTETVGEDFFIKAFE
Mxy10-43A_GH10Domain    QYGSVVYAWDVANEVIHANNS------------------GWEAVYGNNRTNATYVKKAFN

                Xyn10A  YANKYAP-----ADLELYYNDYNECMAKKREGIVALLKEVKEQEGEPGTGTRITAMGMQG
   Xyn10D_GH10Domain    FARKYAA-----PEVSLFYNDYETAQPWKRDYIIANVLKPLMDKG------LVDGMGMQS
Mxy10-43A_GH10Domain    YAYDCLEYFKLTDSVKLFYNDYNTYMEVNN--VIKLVNYINQGKK------VCAGVGMQS

                Xyn10A  HYGVDSPSYTEVETSAKAYGDVVGNIQITEWDLSSSDDYDGSPEAKEKEYEKMRKTYNLQ
   Xyn10D_GH10Domain    HLLMDHPDINEYRTALEMYGATGLQIHITELDMHNADPSEESMHALATRYQEFFQTY---
Mxy10-43A_GH10Domain    HLGTGFPSVDYYTTALRSFLNAGFEVQITEMDITNKGDSDMATYAYNLFKNIN-------

                Xyn10A  YYALQSVQNSGVNVTGITFWGTVDKYSWLQHRSTVGGGSNKEKAQCPLLFDDLYKPKPAF
   Xyn10D_GH10Domain    ---LDAKKSGKANITSVTFWNLLDENSWLSG--------FRRETSYPLVFKRKCEAKEAY
Mxy10-43A_GH10Domain    -----SLKKNGGKITSITWWGLSDQTTWISN-------------SAPLLFSRPGTPKQAY

                Xyn10A  WVFAETN------------
   Xyn10D_GH10Domain    YTVLKAAVSDDSIDKWVPD
Mxy10-43A_GH10Domain    NKV----------------
```

Figure 4-13: Sequence alignment in ClustalW of the three GH10 domains chosen for investigation.

Xyn10A is a secreted enzyme that contains a single GH10 domain, Xyn10D is an intracellular enzyme with a GH10 domain and a CE10 esterase domain; Mxy10-43A is a secreted enzyme identified as being important in the initiation of extracellular polysaccharide degradation by the bacterium[1], it has five domains, a GH10 xylanase domain, a GH43 domain and three carbohydrate binding domains. Previous analysis of these enzymes is shown in Table 4-3.

Table 4-3: Xylanase enzymes chosen for further investigation.

Enzyme	ORF	# Amino Acids	M_r (Da)	EC Number	Microarray (fold up-regulated)
Mxy10-43A	Bpr_I0737	1394	150747	3.2.1.8	0
Xyn10A	Bpr_I0304	476	53630	3.2.1.8	22.7
Xyn10D	Bpr_I1083	692	79691	3.2.1.8	5.8

4.3.2 Expression trials of Xylanses

4.3.2.1 Expression trials of Xyn10A

Previous attempts at expression and purification of Xyn10A using the standard conditions were unsuccessful. Expression trials were performed as detailed in section 2.8.4 across a range of growth temperature and lysis buffer/pH combinations. From each combination four stages were sampled, the whole cell (WC) after growth with IPTG, the cell pellet (CP), the supernatant (SN) after cell lysis and centrifugation, and the beads from the nickel pull down (PD). The fractions were visualised with 10% SDS-PAGE gels (Figure 4-14).

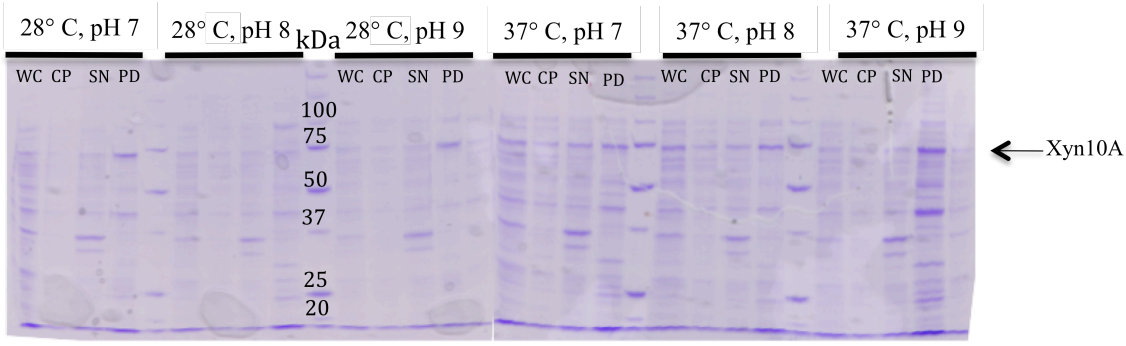


Figure 4-14: 10% SDS-PAGE gels of expression test for Xyn10A across a range of conditions. WC= whole cell, CP= cell pellet, SN = supernatant, PD = nickel pull down.

The results indicated that no conditions gave significant quantities of soluble protein in the nickel pull downs.

Analysis of the protein sequence using PrediSi (www.predisi.de) indicated that the first 16 amino acids of the protein were a signal sequence. It was hypothesised that by removing this from the expressed protein the amount of soluble protein expressed may increase. The signal sequence was removed and the shorter gene was cloned into a pDEST17 vector as outlined in section 2.8.5. The shorter gene was named Xyn10A-16.

4.3.2.2 Expression trials of Xyn10A-16

Expression trials were performed using the shorter protein Xyn10A-16 as outlined in section 2.8.6. The results indicated that expression of soluble Xyn10A-16 was significantly increased over that of Xyn10A across all temperature and pH combinations (Figure 4-15). Large bands were seen in the nickel pull down fractions of all the combinations tested at around 75 kDa, which is larger than the predicted molecular mass of 52.4 kDa. Sequencing of the vector showed that the correct sequence was in the expression vector and in frame.

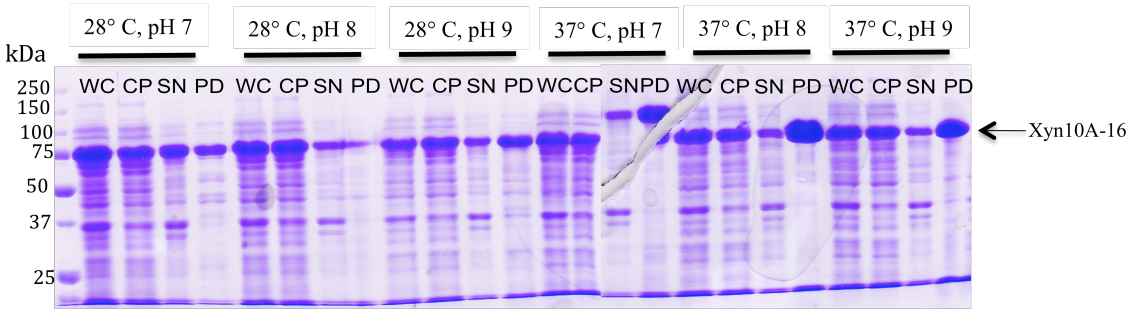


Figure 4-15: 10% SDS-PAGE gels of expression test for Xyn10A-16. WC= whole cell, CP= cell pellet, SN = supernatant, PD = nickel pull down.

Maximum expression of soluble protein was seen at 37 °C and lysis buffer at pH 8. This was chosen as the conditions to use for large scale expression.

4.3.2.3 Previous Analysis of Mxy10-43A

Mxy10-43A is a large multidomain protein encoded by the ORF Bpr_I0737 of

B. proteoclasticus. Mxy10-43A is a 1362 amino acid protein with a calculated molecular weight of 147028 Da. Annotation of this gene indicates the protein is made up of five domains: three carbohydrate binding modules (CBM) together with a GH10 and a GH43 catalytic domain[1]. These catalytic domains are annotated as having endo-1,4-ß-xylanase and xylosidase activity respectively. Mxy10-43A is secreted from *B. proteoclasticus*.

4.3.2.4 Expression trials of Mxy10-43A

Expression trials of Mxy10-43A were performed as detailed in section 2.8.7. Five samples from each culture and were visualised on 10% SDS-PAGE gels (Figure 4-16). A sample of the LB media (LB) after centrifugation of culture to check for extra-cellular secretion, whole cells prior to lysis (WC), insoluble cell pellet (CP) after lysis and centrifugation, supernatant after cell lysis and centrifugation (SN) and the nickel pull down fraction (PD).

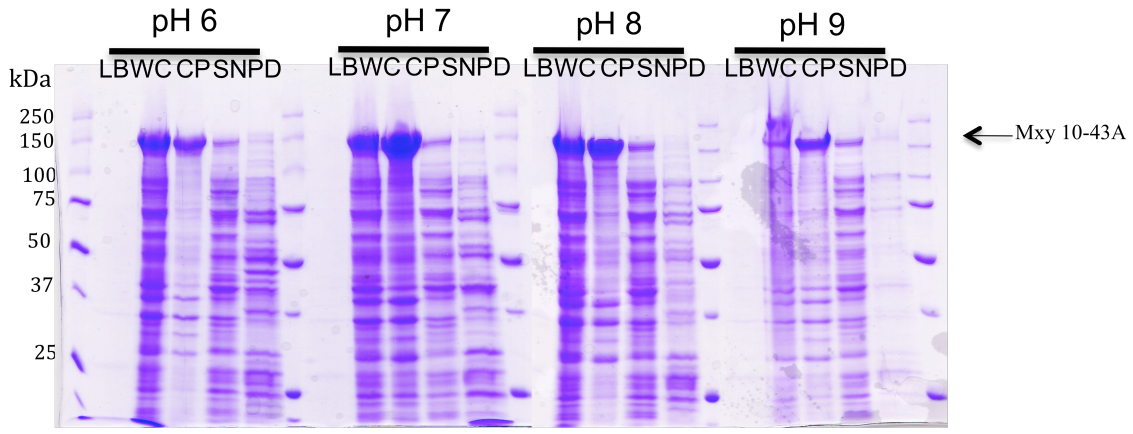


Figure 4-16: Expression trials of Mxy10-43A across a pH range. LB - LB media after centrifugation of culture, WC- whole cells prior to lysis, CP - insoluble cell pellet after lysis and centrifugation, SN- supernatant after cell lysis and centrifugation, PD – Nickel pull down fraction.

None of the conditions screened indicated successful expression of soluble protein; a large band of the appropriate size, representing over-expressed protein, is clearly seen

in the cell pellet fraction of each preparation. This indicates that Mxy10-43A is insoluble in all the conditions screened. Further attempts to express soluble protein should include removal of the signal sequence, if that fails expressing individual domains could be explored.

4.3.3 Expression and Purification of Xylanases

4.3.3.1 Expression and Purification of Xyn10A-16

Large scale expression and purification of Xyn10A-16 was performed as outlined in section 3.2.1 except that Xyn10A-16 was grown for 3 hours at 37 °C post induction with IPTG, instead of 16-20 hours at 28 °C, based on the results of the small scale expression tests. The IMAC trace and corresponding SDS-PAGE gel show Xyn10A-16 eluted from the IMAC column at ~150 mM imidazole (Figure 4-17).

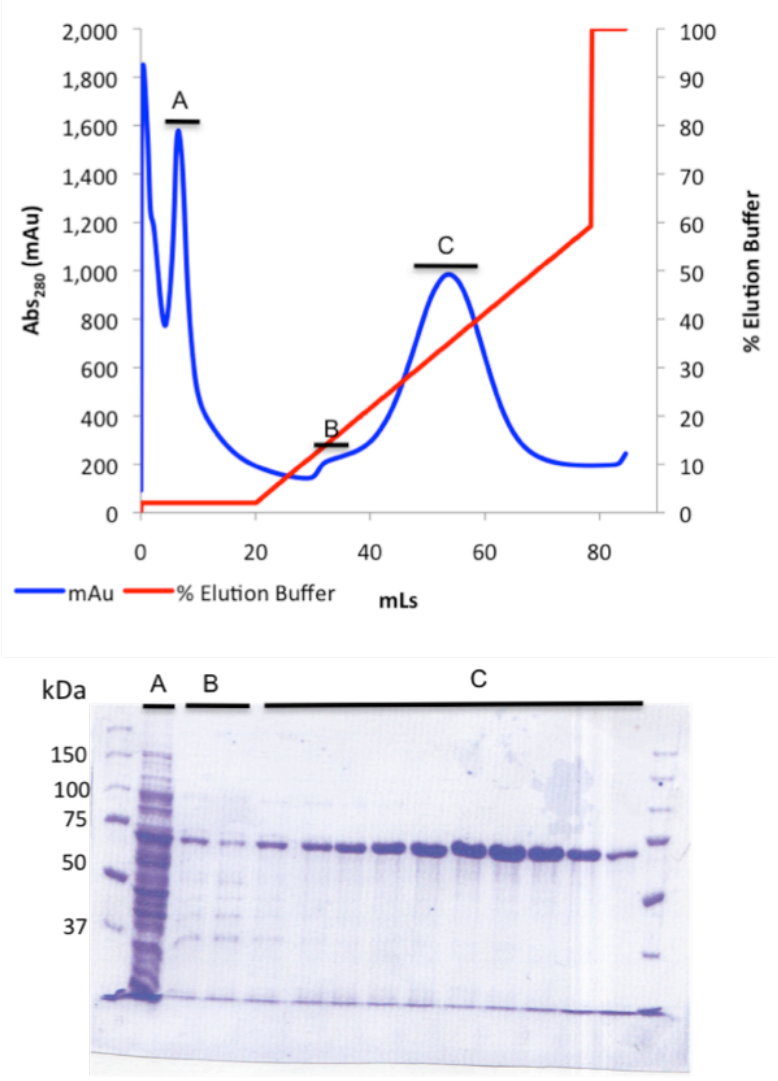


Figure 4-17: IMAC chromatogram for Xyn10A-16 with corresponding 10% SDS-PAGE gel

Xyn10A-16 in fractions from peak C from Figure 4-17 was purified further using an S200 16/60 column. The chromatogram and corresponding SDS-PAGE gel from this purification step show Xyn10A-16 elutes in a single peak at ~75 mL (Figure 4-18). This corresponds to an M_r of 75 kDa.

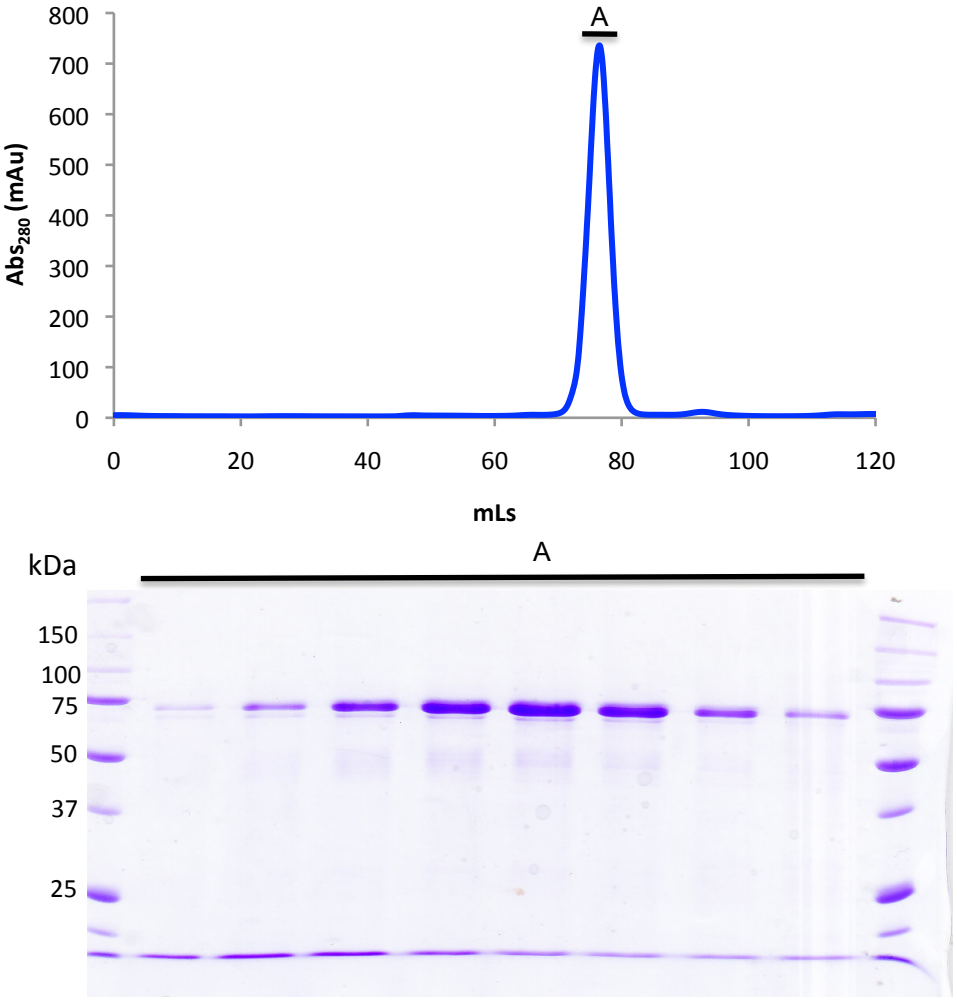


Figure 4-18: Size Exclusion Chromatogram for Xyn10A-16.

4.3.3.2 Expression and Purification of Xyn10D

Xyn10D was expressed and purified as outlined in section 3.2.1; Xyn10D eluted from the IMAC column with ~250 mM imidazole (Figure 4-19).

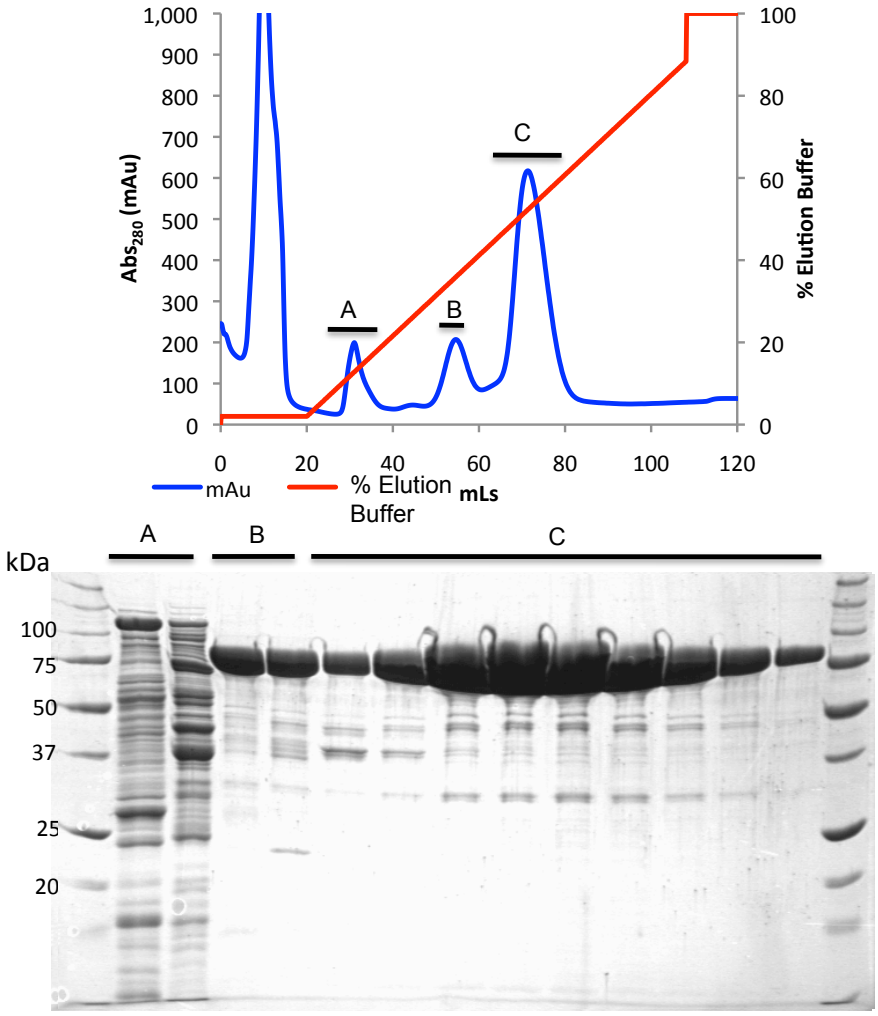


Figure 4-19: IMAC trace of Xyn10D with corresponding 12% SDS-PAGE gel.

Xyn10D contained in fractions from peak C in Figure 4-19 was concentrated and loaded on to a S200 16/60 size exclusion column. The resulting chromatogram from the S200 16/60 (Figure 4-20) indicates the majority of Xyn10D elutes as a single peak at ~65 mL. This equates to a M_r of 205 kDa, which is 2.6 times the calculated M_r for Xyn10D. Because Xyn10D is a multidomain protein this could elongate the shape of the folded protein making its hydrodynamic radius appear larger and causing it to elute differently.

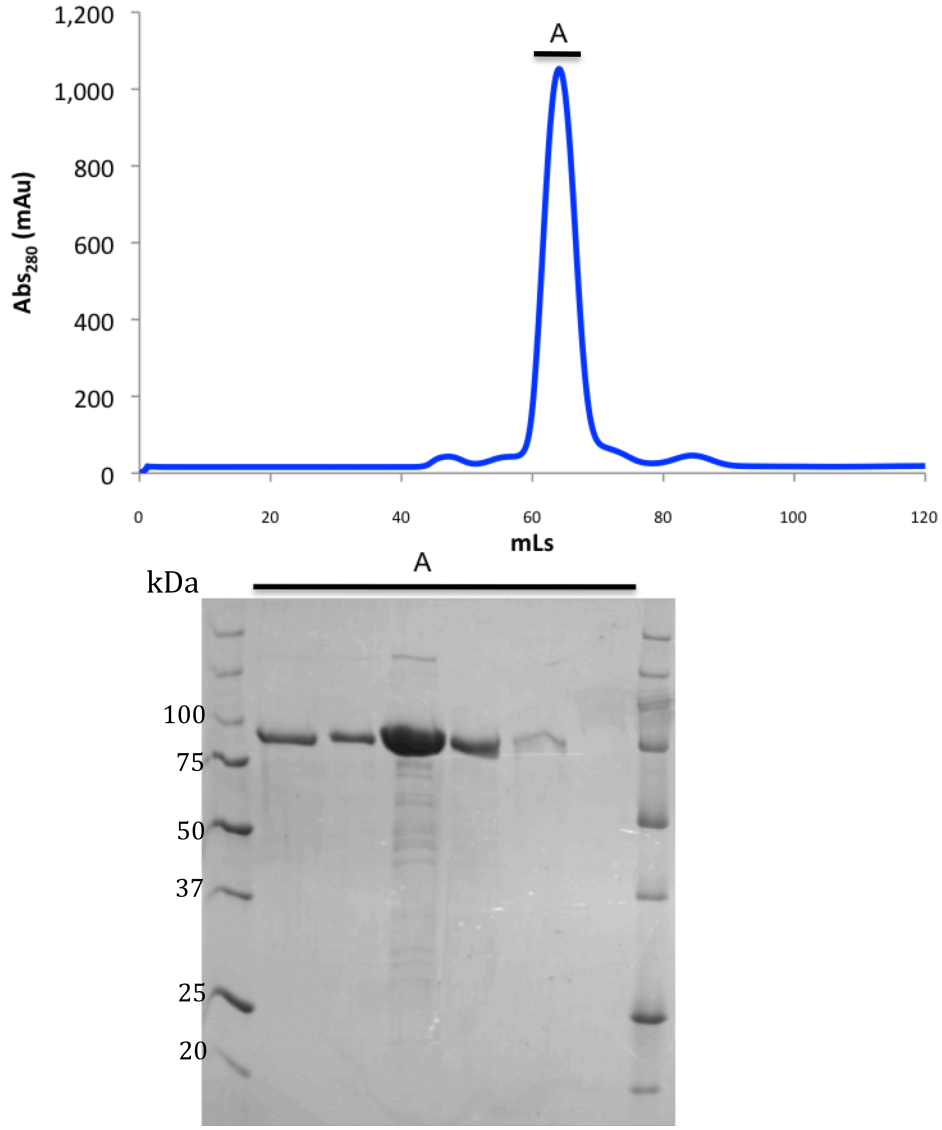


Figure 4-20: S200 16/60 chromatogram with corresponding 12 % SDS-PAGE gel for Xyn10D.

4.3.4 Crystallisation Attempts

Xyn10A-16 and Xyn10D were subject to crystallisation trials. Unfortunately no suitable crystals of either protein were produced.

4.3.5 Functional Analysis of Xyn10D

Xyn10D was tested for activity on the same four natural substrates as the four GH43

enzymes (section 4.2.4). The only sugar seen in detectable quantities released by Xyn10D was glucose. This result was not expected, as Xyn10D is a xylansae therefore if sugars were seen it was expected to be xylose. This could indicate that the Xyn10D has a more versatile functionality than just xylanase activity Further investigation into the activity of Xyn10D included testing it against the model substrates Azo-CMC and Azo-WAX. Xyn10D showed activity against Azo-WAX but it was inactive against the Azo-CMC substrate. The results show that Xyn10D does have endo-1,4-ß-D-xylanase activity but not endo-1,4-ß-D-glucanase, this supports its annotation. Further investigation would be required for full characterisation of this enzyme. However based on its activity on Azo-WAX and the GCMS results showing it to release glucose, indicating it could be a multifunctional enzyme, Xyn10D shows potential as an enzyme that could be included in a hemicellulose degrading cocktail as a xylanase.

4.4 Enzymes with Annotated Glucanase activity

B. proteoclasticus cannot grow on crystalline cellulose as a carbon source, however the genome contains genes annotated as having glucanase activity. Two enzymes Lic16B and Cel5C were chosen for characterisation in attempt to understand their importance to *B. proteoclasticus*. Lic16B is an intracellular enzyme that is annotated as being an endo-1,3(4)-ß-glucanase. The gene encoding Lic16B is upregulated 4.6 times when the bacterium is grown on xylan as opposed to xylose indicating it plays a role in xylan degradation. Cel5C is a secreted enzyme that contains two domains, one annotated as having endoglucanase activity and the other contains a carbohydrate binding domain and is also annotated as having 1,4-ß-cellobiosidase activity which is classified as exo-glucanase activity. Previous analysis of these enzymes is shown in Table 4-4.

Table 4-4: Glucanase enzymes chosen for further investigation.

Enzyme	ORF	# Amino Acids	M_r (Da)	EC Number	Microarray (fold up-regulated)
Lic16B	Bpr_I1522	252	29254	3.2.1.6	4.6
Cel5C	Bpr_I1710	547	61122	3.2.1.4	0

4.4.1 Expression and Purification of Glucanase enzymes

4.4.1.1 Small Scale Expression Testing of Cel5C

A small scale expression test of Cel5C was performed as outlined in section 2.8.8 and analysed on a 12% SDS-PAGE gel (Figure 4-21). A negative control was also included of the whole cell fraction of *E. coli* containing the plasmid that had not been treated with IPTG.

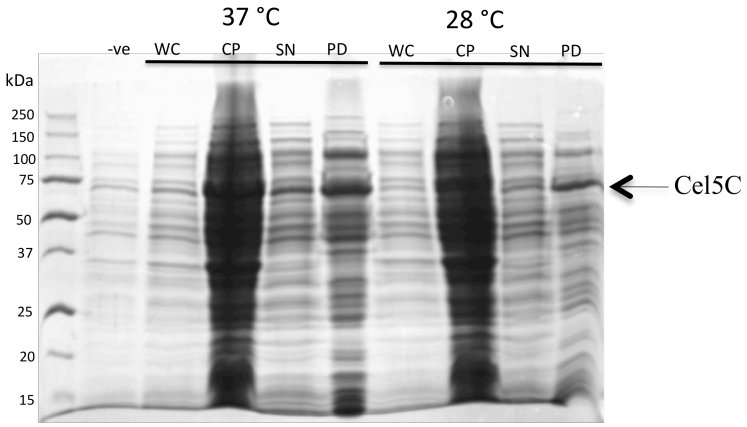


Figure 4-21: Small scale expression test of Cel5C at 37 °C and 28 °C. -ve = negative, WC = whole cell, CP = cell pellet, SN = supernatant and PD = nickel pull down.

The results of the small scale expression test indicated that more soluble protein was expressed at 37 °C than at 28 °C.

4.4.1.2 Large Scale Expression and Purification of Cel5C

Cel5C was expressed and purified as previously described in section 3.2.1 with the exception that post induction with IPTG the cultures were grown for 3 hours at 37 °C, instead of 16-20 hours at 28 °C, based on the results of the small scale expression tests. The absorbance trace indicates Cel5C elutes from the IMAC column with elution buffer containing 250 mM imidazole (Figure 4-22).

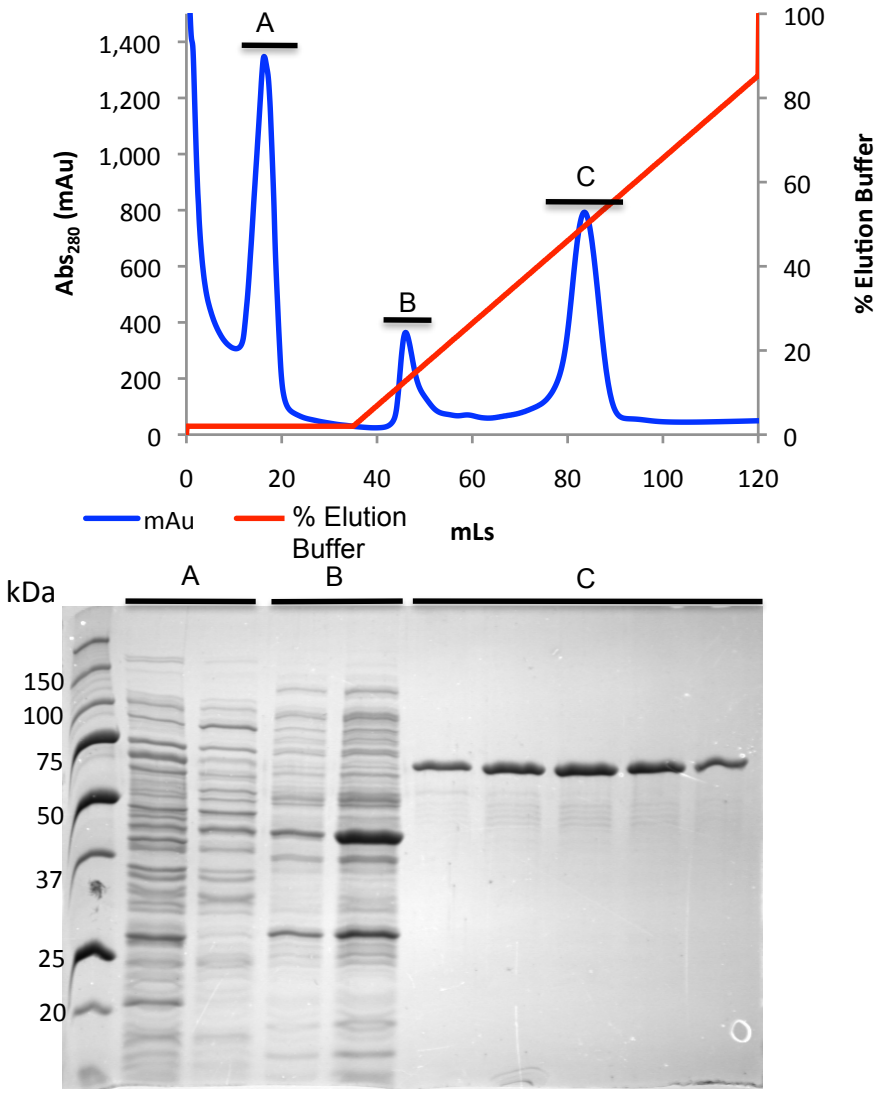


Figure 4-22: IMAC purification of Cel5C with SDS-PAGE gel.

Fractions containing Cel5C were pooled and further purified using an S200 16/60 size exclusion column. The resulting chromatogram and 12% SDS-PAGE gel shows Cel5C elutes around 90 mL (Figure 4-23). This equates to a M_r of 27 kDa, which is considerably smaller than the calculated M_r of Cel5C which is 61 kDa and does not agree with the M_r from the SDS-PAGE gel.

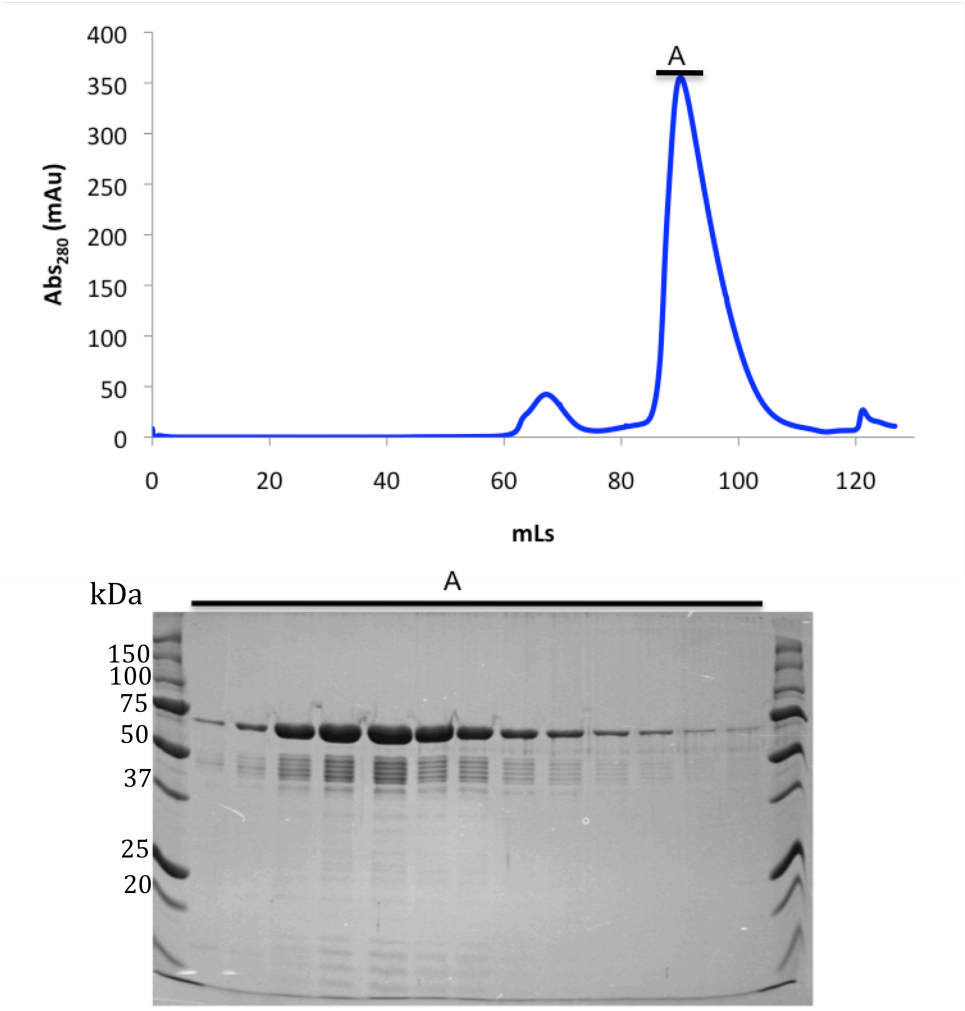


Figure 4-23: Size exclusion chromatogram for Cel5C with corresponding 12% SDS-PAGE gel.

There is an interesting banding pattern seen in the SDS-PAGE gel. The second domain is ~10kDa, which in the sequence is preceded by a 766 Da linker containing a seven amino acid (PDPQPVD) section that is repeated five times. The bands appear

to be the loss of the larger section then successive repeat sections.

4.4.2 Crystallisation Trials of Cel5C

Crystallisation trials of Cel5C were laid down at two concentrations (15 $mg.mL^{-1}$ and 35 $mg.mL^{-1}$). Both trials gave no positive results. The banding pattern in the SDS-PAGE gel from the S200 16/60 column that indicated Cel5C was degrading. The heterogeneous mix of species would not favour crystallisation. Future attempts to crystallise the protein should include expressing each domain separately and trying to crystallise them individually.

4.4.3 Functional Analysis of Glucanases

Both Cel5C and Lic16B were tested on the natural substrates mentioned in section 4.2.4 and the results are presented in Table 4-5.

Table 4-5: Sugars released from natural substrates by glucanase enzymes.

Enzyme	Sugar(s) released from natural substrates
Lic16B	Arabinose/xylose, Glucose
Cel5C	Glucose

The release of arabinose and/or xylose from the natural substrates by Lic16B indicates that its catalytic capabilities may be more than just endo-1,3(4)-ß-glucanase activity. This needs further investigation to confirm but that would help to explain why the gene was upregulated in response to growth on xylan.

Both enzymes were also tested for activity on the two model substrates Azo-CMC and Azo-WAX. Cel5C was active on both substrates indicating it has both endo-1,4-ß-D-glucanase and endo-1,4-ß-D-xylanase activity. This suggests that Cel5C is a multifunctional enzyme, although further investigation is needed to confirm this the potential of a multifunctional enzyme is exciting. It would allow a potential fibre degrading cocktail to require fewer enzymes resulting in lower production costs while

maintaining the range of activities. Lic16B showed no activity on either substrate.

4.5 Discussion

From the preliminary investigation into nine of the fibre degrading enzymes from *B. proteoclasticus*, I have determined all express and purify from *E. coli* (DE3) cells using Gateway pDEST17 expression vectors except for Mxy10-43A. Xyn10A expressed well after removal of the signal sequence. Large scale expression of the eight soluble enzymes in this chapter gave significant quantities of clean protein when purified by IMAC and size exclusion chromatography. The proteins have good stability with no indication of any proteins precipitating from solution during purification or storage.

The results from the preliminary functional investigation illustrate the importance of determining the function of an enzyme experimentally. The annotations are based on protein sequence similarity with enzymes previously characterised; they should only be used as a prediction of function. Genome annotations are limited and not always a good indication of the full functionality of the enzyme.

To obtain complete degradation of hemicellulose a cocktail containing a range of enzymes with varying functions is needed. Based on the results of these preliminary investigation enzymes with arabinofuranosidase activity and xylanase activity have been identified that could potentially add to this cocktail. Full functional investigation would be required to define the optimal conditions and substrate specificity of these enzymes before they were included in such a cocktail.

Xsa43E was the only protein laid down in crystallisation trials that has produced crystals in fine screens; a full functional and structural investigation along with the elucidation of the mechanism of action of this enzyme is presented in chapter five.

Chapter Five: The Structural and Functional Characterisation of Xsa43E.

5.1 Introduction

5.1.1 Arabinoxylan

Glucuronoarabinoxylan (GAX) is one of the most common forms of hemicellulose in the cell wall of forage grasses. It consists of a xylopyranose backbone with arabinose side groups usually branching from the 2-O or 3-O position of the xylose sugars. Arabinose groups can condense with phenolic acid esters. The addition of these side groups has two benefits to the cell wall; it increases the amount of hydrogen bonding with other strands of hemicellulose or lignin and restricts access of enzymes to the xylan backbone, impeding enzymatic degradation.

The complete degradation of hemicellulose requires a suite of different enzymes due to the complex nature of the polysaccharide structure.

5.1.2 GH43 Family Enzymes

Of the fibre degrading enzymes produced by *B. proteoclasticus*, ten have been annotated as belonging to the GH family 43, from the GH-F clan of the CAZy database[55]. Family GH43 contains 1287 genes, 1130 from bacterial origin, 143 from eukaryotes, seven from archaea and seven unclassified entries. GH43 family enzymes have a number of annotated functions including β-xylosidase (EC 3.2.1.37) that hydrolyse xylose from β-(1,4)-D-xylans, β-1,3-xylosidase (EC 3.2.1.-) that hydrolyse (1,3)-β-D-glycosidic linkages in (1,3)-β-D-xylans. There are also α-L-arabinofuranosidase (EC 3.2.1.55) that cleave α-L-arabinose units connected by (1,3)- and/or (1,5)-linkages in arabinans and arabinoxylans, arabinanase (EC 3.2.1.99) that are endo- acting and hydrolyse (1,5)-α-arabinose linkages in (1,5)-arabinans. The group also contains enzymes with xylanase activity (EC 3.2.1.8) these are endo-acting and hydrolyse (1,4)-β-D-xylosidic linkages in xylans, and enzymes with

galactan activity 1,3-β-galactosidase (EC 3.2.1.145) hydrolyse non-reducing β-D-galactose residues in (1,3)-β-D-galactopyranans[55]. The GH 43 family has 33 structures in the protein data bank (PDB) from 15 different proteins; all have a 5 bladed ß-propeller domain and some also contain additional domains. The first reported 5 bladed ß-propeller structure was of tachylectin[136]. Since then the domain has been seen in four different GH families: GH families 32 and 68 of the GH-J clan and GH families 43 and 62 of the GH-F clan[55]. All enzymes from the GH43 family have an inverting catalytic mechanism with three residues identified as being essential for catalysis; a general base that activates a water molecule which then acts as the nucleophile, a general acid that is the proton donor and a third acid residue that is hypothesised to modulate the pK_a of the general acid, as well as orientating both the substrate and general acid. This third residue is essential for activity[137].

5.1.3 Inverting Mechanism of GH43 Enzymes

Members of the GH43 family of enzymes have an inverting mechanism; the configuration of the anomeric carbon of the sugar substrate is inverted as opposed to being retained. This is achieved by having the catalytic acid and the catalytic base on opposite sides of the substrate in the enzyme's binding pocket. As the catalytic base activates a water molecule which attacks the anomeric carbon from one side, a catalytic acid donates a proton to the leaving group on the other side, breaking the glycosidic bond between the two sugar moieties and in the process inverting the configuration of the anomeric carbon (Figure 5-1).

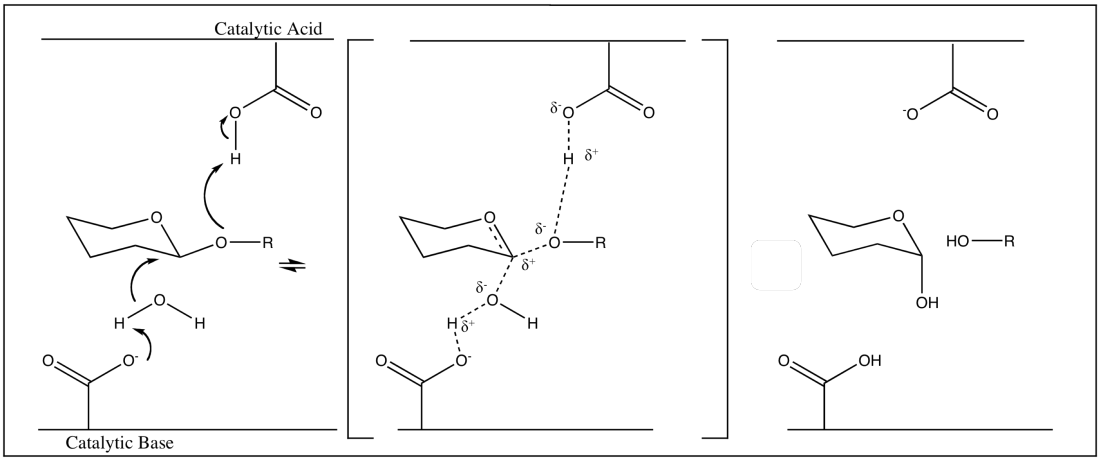


Figure 5-1: Inverting mechanism of GH43 enzymes.

5.2 Results and Discussion

5.2.1 Structural Determination

5.2.1.1 Seleno-methionine protein

Seleno-methionine protein was expressed using DH41 *E. coli* expression cells in defined media as set out in section 2.2.3.2. The protein was purified the same way as the native protein (sections 2.2.5 and 2.2.6) except all buffers contained 2 mM β-mercaptoethanol. The protein was analysed using MALDI-TOF MS which confirmed the incorporation of seleno-methionine (section 2.9.1.1). Fine screens were conducted with seleno-methionine incorporated protein at a concentration of 14 $mg.mL^{-1}$. The fine screen was focussed around the best result from the initial crystallisation trials and varying the pH more significantly ranging from 3.2 to 5.5.

5.2.1.2 Crystals

Seleno-methionine crystals were grown in 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 3.4. Native crystals of Xsa43E were grown in 18% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate buffer pH 4.2. Native crystals diffracted to a maximum resolution of 2.2 Å and seleno-methionine crystals diffracted to a

maximum resolution of 1.3 Å using synchrotron radiation. Both MAD and SAD datatsets were collected from a single seleno-methionine crystal. The MAD datasets were collected first, then SAD data was collected in two sets, a high and low resolution pass, in order to avoid saturating the low resolution reflections.

5.2.1.3 Crystallisation and X-ray Diffraction

These crystals scaled and integrated in the space group $P2_12_12_1$ with the unit cell dimensions a=58.3, b=76.4, c=84.1 Å, α=β=γ=90. Crystals of Se-Met Xsa43E diffracted to a maximum resolution of 1.3 Å and also belonged to the space group $P2_12_12_1$ with the cell dimensions a=59.1, b=76.5, c=82.5 Å, all angles equal 90°. The Matthews coefficient (2.67) predicts one Xsa43E molecule in the asymmetric unit. Autosol located nine sites, eight of which were selenium and one was calcium, with a figure of merit (FOM) of 0.50. Following density modification with RESOLVE the FOM improved to 0.73. Complete data collection statistics for all data sets are shown in Table 5-1, Table 5-2 and Table 5-3.

Table 5-1: Complete data collection statistics for the native dataset of Xsa43E.

	Native Data
Space group	$P2_12_12_1$
Wavelength	0.9795
Cell parameters	
a	58.32 Å
b	76.43 Å
c	84.05 Å
α	90°
β	90°
γ	90°
Resolution range (Å)	83.92-2.2 (2.32-2.20)
R_{merge}	11.0% (54.8%)
No. of Measured Reflections	286293 (41529)
No. of Unique Reflections	19728 (2831)
Mean I/σI	22.0 (5.1)
Completeness	100 (100)
Multiplicity	14.5 (14.7)

Table 5-2: Complete data collection statistics for the MAD dataset of Xsa43E.

	Peak	MAD Inflection	Remote
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Wavelength	0.9783	0.9791	0.9184
Cell parameters			
a	59.05 Å	59.04 Å	59.05 Å
b	76.48 Å	76.48 Å	76.49 Å
c	82.51 Å	82.49 Å	82.50 Å
α	90°	90°	90°
β	90°	90°	90°
γ	90°	90°	90°
Resolution range (Å)	59-1.8 (1.9-1.8)	59-1.8 (1.9-1.8)	59-1.8 (1.9-1.8)
R_{merge}	5.2% (15.6%)	5.3% (15.3%)	4.9% (13.9%)
No. of Measured Reflections	113889 (16374)	113710 (16253)	113908 (16310)
No. of Unique Reflections	24603 (3789)	24612 (3789)	24289 (3740)
Mean I/σI	19.3 (8.1)	19.1 (8.2)	20.3 (9.3)
Completeness	70.9 (75.4)	70.9 (75.5)	70.0 (74.4)
Anomalous Completeness	65.7 (70.0)	65.7 (70.0)	64.8 (69.0)
Multiplicity	4.6 (4.3)	4.6 (4.3)	4.7 (4.4)
Anomalous Multiplicity	2.6 (2.4)	2.6 (2.3)	2.6 (2.4)

Table 5-3: Complete data statistics for the SAD dataset of Xsa43E.

	SAD
Space group	$P2_12_12_1$
Wavelength	0.9795
Cell parameters	
a	59.13 Å
b	76.54 Å
c	82.56 Å
α	90°
β	90°
γ	90°
Resolution range (Å)	56.2-1.33 (1.4-1.33)
R_{merge}	7.7% (51.0%)
No. of Measured Reflections	463689 (44441)
No. of Unique Reflections	82628 (11406)
Mean I/σI	14.9 (2.3)
Completeness	95.9 (91.7)
Anomalous Completeness	88 (78.3)
Multiplicity	5.6 (3.9)
Anomalous Multiplicity	3.0 (2.2)

5.2.1.4 Phase determination and improvement

MAD

The structure of Xsa43E was determined by MAD methods from Se-Met substituted Xsa43E using the program Phenix AutoSol at a resolution of 1.8 Å. Density modification, phase extension and initial automatic model building was carried out using Phenix AutoSol to a final resolution of 1.8 Å.

SAD

A high-resolution structure was determined by SAD methods from the Se-met substituted Xsa43E. The images were scaled using the CCP4 programme SCALA[87]. The structure was then solved using Phenix AutoSol to a resolution of 1.33 Å, followed by AutoBuild, as this data had the highest resolution it was used for refinement to generate the most accurate structure.

5.2.1.5 Refinement and Further model building

The output model from Phenix AutoBuild was put through an initial round of refinement in the programme Refmac5 from CCP4. The scaled mtz file from the SAD data collection was used for refinement. The output from the initial round of refinement was used as a starting point for manual building. Manual model building was done in COOT[93] and progress was checked with restrained refinement in Refmac5 from the CCP4 interface as set out in section 2.5.4.3. Electron density maps used for model building were SigmaA weighted $2|F_O|-|F_C|$ (contoured at 1.0 sigma) and $|F_O|-|F_C|$ (contoured at ± 3.0 sigma). Model building required inserting alternative conformations for some residues.

5.2.1.6 Model Completeness and Quality

When the model was sufficiently refined waters were added in Refmac5. The waters

were manually checked to insure they were in sensible positions and were justified by sufficient electron density. Model completeness and quality were evaluated by running PROCHECK[90] in CCP4.

5.2.2 Structure of Xsa43E

One monomer consisting of one catalytic domain is present in the asymmetric unit of the crystal. Residues 9-313 are visible in the electron density. The final structure also contains 319 water molecules, a calcium ion and a molecule of tris(hydroxymethyl)aminomethane (tris). The highest resolution (1.33 Å) was obtained from a crystal of seleno-methionine derivitised protein and the final *R-factor* and R_{free} for these data are 16.9% and 18.6% respectively (complete refinement statistics can be seen in Table 5-4).

Table 5-4: Refinement statistics for the structure of Xsa43E.

Refinement Statistics	
R_{Work} (R_{free})	16.9% (18.6%)
Total No. Atoms	2827
No. of Protein Atoms	2499
Other Molecules/Ions	2 (1 tris, 1 Ca)
No. of Waters	319
RMS Deviation from Standard Geometry	
Bond Lengths (Å)	0.031
Bond Angles (°)	2.343
Average B-factors	
Protein	15.653
Water	26.678

The structure of Xsa43E has 96.4% of the residues within the most favoured region and 3.6% in allowed regions of the Ramachandran plot (Structure validation using ProCheck[90] in CCP4) (Figure 5-2).

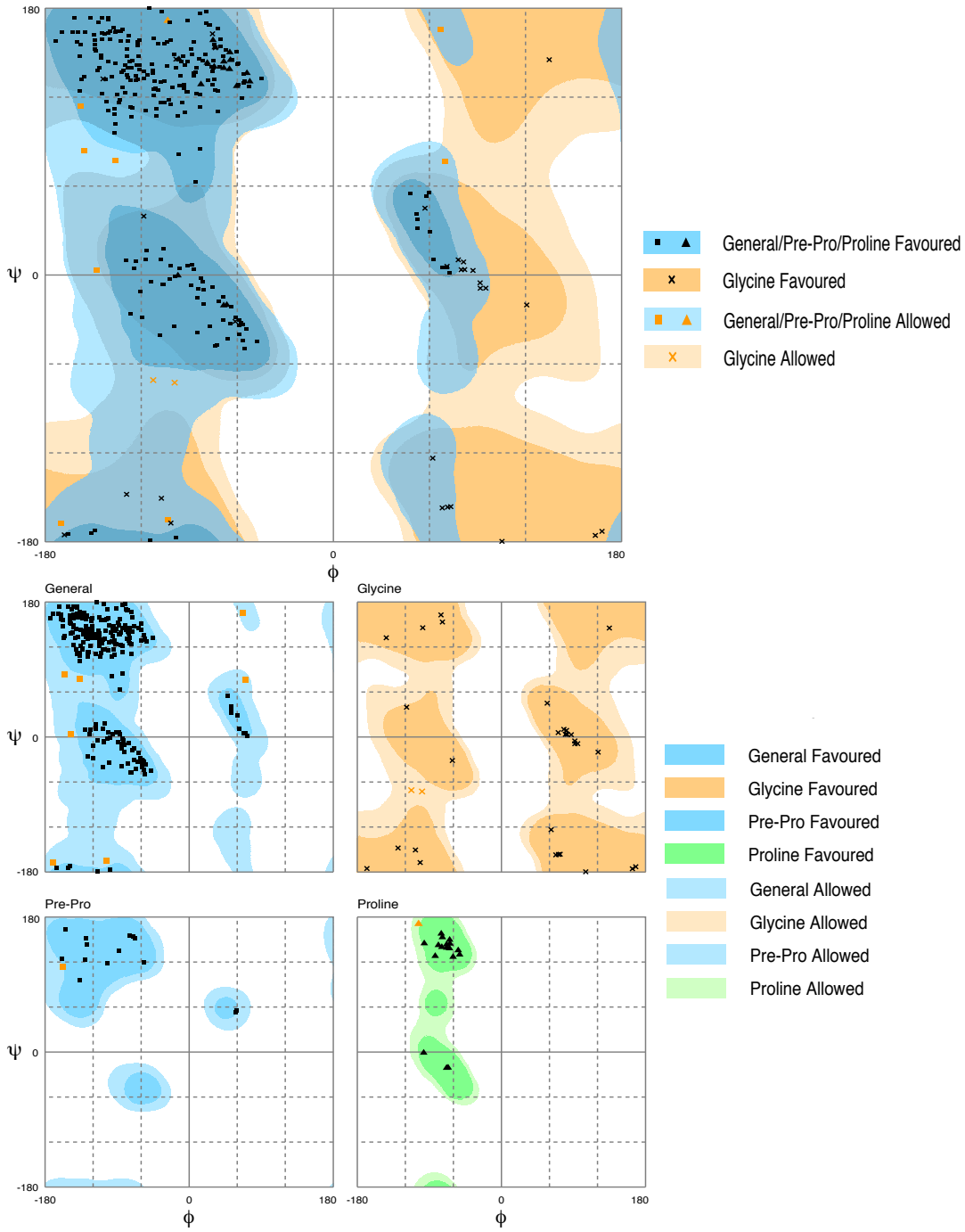


Figure 5-2: Ramachandran plot for Xsa43E produced using Procheck in CCP4.

Xsa43E has one domain with the overall structure consisting of a 5 bladed

β-propeller. This domain was first reported for tachylectin[136] but since has been seen in all GH43 structures as well as in GH families 32, 62 and 68[55]. The 5 blades of the propeller form a cylindrical shape and are organised around a central axial cavity. This central cavity is approximately 30 Å long and contains the active site. A cartoon representation of the structure is depicted in Figure 5-3.

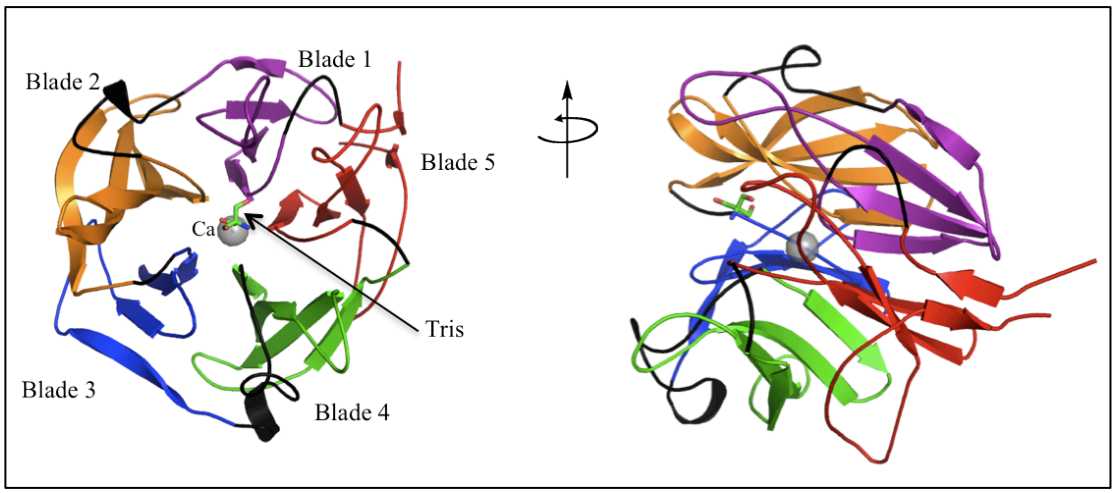


Figure 5-3: Structure of Xsa43E presented as a cartoon representation in two orientations with the central metal ion and molecule of tris bound in the active site.

Each blade of the propeller consists of four antiparallel β-strands connected by loops of varying sizes. The loops between the first and second strand of each blade are short hairpins; the loops between the third and fourth strands are also relatively short. The loops between the second and third and the fourth ß-strand of one blade and the first strand of the next blade are longer, and they form the substrate binding cavity and active site. The inner ß-strands of each blade start from the same relative position along the central axis, just after the active site. In the case of Xsa43E the strands are preceded by a proline residue. The four strands that make up each blade twist through approximately 90° from the inside to the outside of the structure. The N-terminal and C-terminal ends of the protein have short ß-sheets (coloured red in Figure 5-3); the C-terminal strand is the fourth strand of the fifth blade of the propeller and hydrogen bonds to the N-terminal end, which acts almost as a fifth strand in that blade in a

'molecular velcro' fashion effectively sealing the structure.

In the structure there are two short α-helices. In the loop between the fourth ß-strand of the first blade and the first ß-strand of the second blade the residues Lys-75 to Asp-77 form a short helix. The second helix is located from residues Glu-186 to Phe-189 and is located on the loop between the fourth ß-strand of the third blade and the first ß-strand of the fourth blade.

Electron density consistent with a molecule of tris is seen in the active site, this would have originated from the purification buffer. The tris molecule is interacting with Asp-24, Ala-87, Asp-141 and Arg-285.

The electron density map showed an area of strong electron density in the centre cavity of the structure indicating it was occupied by a large ion. Modelling of various metal ions into the site indicated it could be a calcium ion, a chloride ion or a manganese ion. The ion was positioned close to the active site and was coordinated to His-258 residue and six water molecules in a pentagonal bipyramid geometry.

5.2.2.1 The Identity of the Central Ion

The enzyme was dialysed into ultra pure water and sent for inductively coupled plasma mass spectrometry (ICP-MS) analysis (Chemistry Department, the University of Waikato). The results indicated that calcium was the only element present in significant quantities. Mg and Zn showed significant levels however neither satisfied the electron density when modelled into the structure and neither are likely to be in 7 co-ordinate geometry[138]. The other elements tested (Al, Ti, V, Cr, Fe, Mn, Co, Ni, and Cu) are in very low concentrations (Table 5-5).

Table 5-5: Results of Xsa43E from ICP-MS analysis. Numbers are parts per billion.

	Mg	Al	Ca	Ti	V	Cr	Fe	Mn	Co	Ni	Cu	Zn
Blank	1.1	1.2	6.3	0.9	0.0	1.1	4.9	0.9	1.0	0.9	1.0	2.5
Xsa43E	231.7	30.8	3900.9	16.0	2.5	16.3	0.0	4.1	0.3	72.1	49.4	997.4

The presence of calcium is further supported by the literature which shows that calcium is the most likely metal to be in seven co-ordinate geometry[138]. The structure of an 1,5-α-L-arabinanase from *Geobacillus stearothermophilus* to 1.06 Å resolution also identified the central atom as calcium[139]. Of the 15 GH43 family structures five identify calcium as the central ion, six have a water molecule in the corresponding position, two have a sodium ion there, one has a magnesium ion and one has nothing in that position.

5.2.2.2 The Active Site and Known Catalytic Residues

The three catalytic residues previously reported in GH43 family structures[138,140,141] were identified by superimposing the structure of the arabinanase Arb43A from *Cellvibrio japonicus* (PDB accession code 1GYD)[137] onto the structure of Xsa43E. In Xsa43E Asp-24 is the catalytic base, Glu-202 is the catalytic acid, and Asp-141 is the third catalytic residue believed to modulate the pK_a of the catalytic acid as well as orientating the catalytic acid and the substrate. These residues are located on the innermost strands of blades 1, 4 and 3 respectively. The OD2 of Asp-141 is 3.8 Å from the OE1 atom of Glu-202 and 4.5 Å from the OE2 atom of Glu-202. There is a distance of 6.1 Å between the two catalytic residues Asp-24 and Glu-202. The spatial arrangement of the active site is shown in Figure 5-4 with His-258 and the central calcium ion.

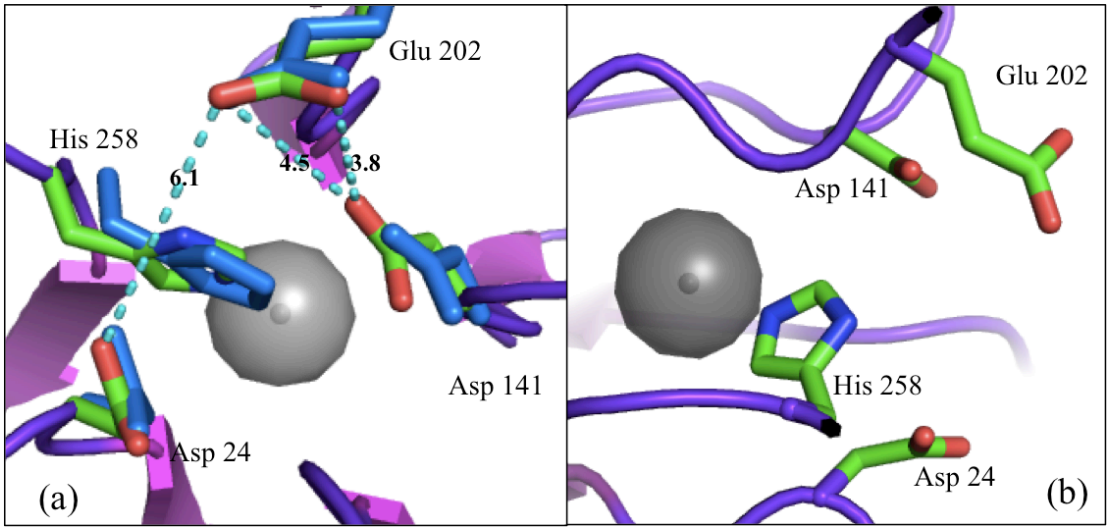


Figure 5-4: The active site of Xsa43E with the essential catalytic residues; (a) view from the active site entrance, with overlaid catalytic residues from Arb43A shown in blue (b) view perpendicular to the active site entrance. Calcium is coloured grey, carbon is green, nitrogen blue and oxygen red for Xsa43E.

5.2.2.3 The Calcium Ion Environment

The central calcium ion is in pentagonal bi-pyramid seven coordinate geometry. It is within hydrogen bonding distance of five equatorial water molecules. It is also coordinated to a water molecule in an axial position while the other axial position is occupied by the Nε2 in the imidazole ring of His-258. All bonding distances for the calcium are between 2.4 and 2.5 Å. These distances are consistent with other Ca^{2+} ligands in structures in the PDB. The calcium ion environment is depicted in Figure 5-5.

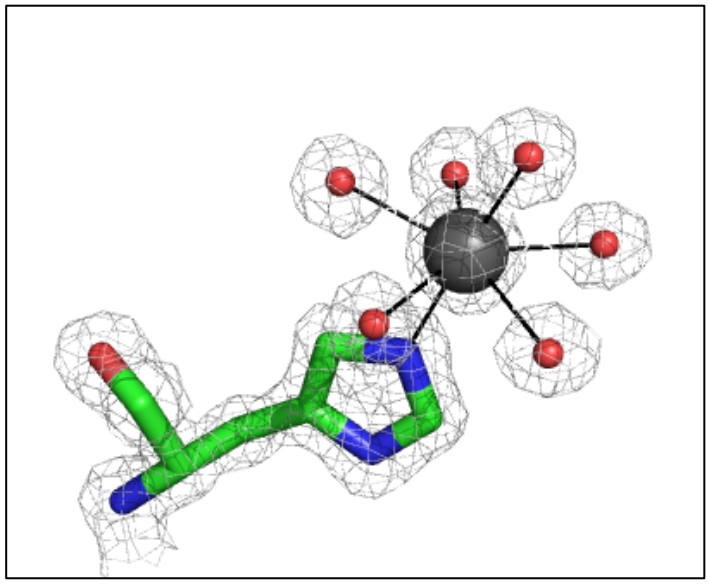


Figure 5-5: The calcium ion coordinated to seven ligands, six water molecules and Nε2 of His 258. The distances between the calcium ion and its ligands are between 2.4-2.5 Å. Calcium is coloured grey; carbon is green, nitrogen blue and oxygen red. Electron Density is 2 | FO | - | FC | map contoured at 1.0 σ.

Each of the five equatorial water molecules are positioned mid way between two of the blades of the propeller; this allows each water molecule to hydrogen bond to two inner most β-strands of consecutive blades of the structure as shown in Figure 5-6. The proline residues that sit at the N-termini of the innermost strands interact with these equatorial water molecules in an interesting way. The water molecules hydrogen bond to the oxygen atoms of either a proline residue or the preceding residue.

Other members of the GH43 family have conserved proline residues in at least three of the five positions seen in Xsa43E but no other structure has all five proline residues. In Xsa43E the proline residues are located where the loops stop and the inner most ß-strands of each blade start; they are also located near the catalytic residues. Pro-25 is the proline in the first blade positioned next to Asp-24, the general base, Pro-88 in the second blade participates in hydrogen bonding to an equatorial water, this is part of the conserved WAP element present in all GH43 enzymes,

Trp-86 and Ala-87 both have been shown to interact with the substrate[140]. Pro-142, located in the third blade, is adjacent to Asp-141 (Figure 5-6), which is essential for catalysis. Pro-204 is located two residues away from the general acid Glu-202; and Pro-259 is next to His-258. The residues analogous to His-258 in other GH43 enzymes have not been shown to be important in previous publications; however it is shown later in this chapter to be important for activity. Structural alignments show that Pro-25 and Pro-88 are conserved throughout the group, except in the 3NQH structure, which has neither. Pro-142 is partially conserved, being present in eight of the sixteen structures. Pro-204 is present in all but three structures and Pro-259 is only present in the structure of Xsa43E. Pro-259 is next to His-258, which is a ligand for the central calcium ion. Because the histidine is held in place by the calcium ion Pro-259 is not as important as the other proline residues for forming the active site. Proline residues are relatively inflexible compared to most amino acids; in Xsa43E they may help to position the catalytic residues in the correct spatial configuration and generally assist in formation of the active site pocket.

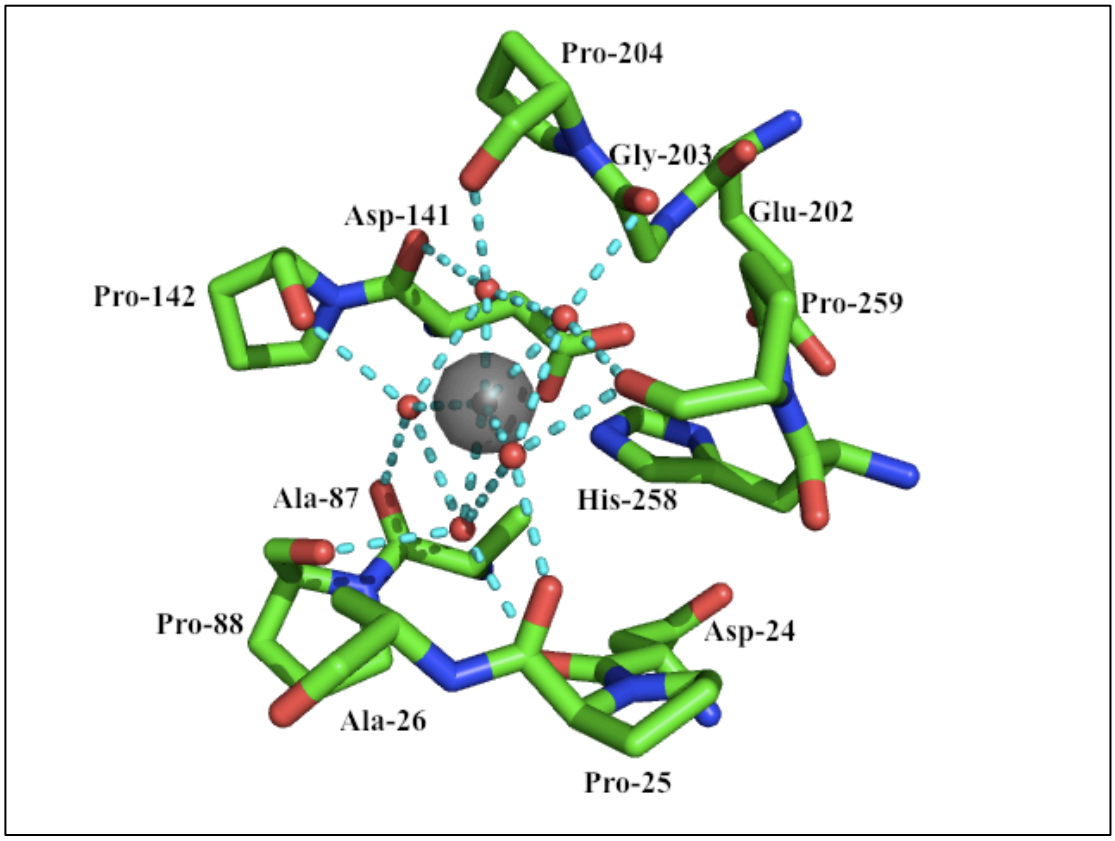


Figure 5-6: The calcium ion coordinated to five equitorial waters that each hydrogen bond to residues on the inner most ß-strands of two consecutive blades of the propeller. Calcium is coloured grey; carbon is green, nitrogen blue and oxygen red. The dashed lines show bonding interactions, all distance between are between 2.4 and 3.1 Å.

5.2.2.4 Structure Comparison with other GH43 Enzymes

The GH43 enzyme structures of all 15 proteins that are available in the PDB have a common 5 bladed ß-propeller domain. Eight of the structures also have a second domain as well. Of the 15 structures nine are monomers, three form dimers and four are tetramers in their crystal structures. All structures have four antiparallel ß-sheets per blade that rotate roughly through 90° from the centre to the outside of the structure. The sequence identity between all GH43 structures from the PDB is shown in Table 5-6, with the RMSDs shown in Table 5-7.

Table 5-6: Sequence identity between GH43 structures.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Xsa43E		**0.323**	**0.174**	**0.274**	**0.353**	**0.229**	**0.224**	**0.224**	**0.244**	**0.209**	**0.214**	**0.239**	**0.189**	**0.229**	**0.174**
2	3KST	**0.323**		**0.234**	**0.249**	**0.313**	**0.259**	**0.259**	**0.239**	**0.254**	**0.219**	**0.234**	**0.254**	**0.204**	**0.264**	**0.239**
3	3K1U	**0.174**	**0.234**		**0.189**	**0.199**	**0.204**	**0.194**	**0.199**	**0.179**	**0.184**	**0.174**	**0.194**	**0.557**	**0.219**	**0.134**
4	1GYD	**0.274**	**0.249**	**0.189**		**0.244**	**0.502**	**0.502**	**0.269**	**0.259**	**0.279**	**0.249**	**0.269**	**0.214**	**0.318**	**0.154**
5	3C7E	**0.353**	**0.313**	**0.199**	**0.244**		**0.254**	**0.254**	**0.234**	**0.239**	**0.219**	**0.229**	**0.264**	**0.204**	**0.219**	**0.204**
6	1WL7	**0.229**	**0.259**	**0.204**	**0.502**	**0.254**		**0.960**	**0.254**	**0.254**	**0.254**	**0.254**	**0.264**	**0.219**	**0.308**	**0.194**
7	3CU9	**0.224**	**0.259**	**0.194**	**0.502**	**0.254**	**0.960**		**0.249**	**0.254**	**0.249**	**0.249**	**0.264**	**0.209**	**0.313**	**0.194**
8	3C2U	**0.224**	**0.239**	**0.199**	**0.269**	**0.234**	**0.254**	**0.249**		**0.692**	**0.801**	**0.741**	**0.701**	**0.189**	**0.303**	**0.124**
9	1YRZ	**0.244**	**0.254**	**0.179**	**0.259**	**0.239**	**0.254**	**0.254**	**0.692**		**0.662**	**0.672**	**0.726**	**0.194**	**0.284**	**0.169**
10	1Y7	**0.209**	**0.219**	**0.184**	**0.279**	**0.219**	**0.254**	**0.249**	**0.801**	**0.662**		**0.771**	**0.726**	**0.184**	**0.303**	**0.139**
11	1YIF	**0.214**	**0.234**	**0.174**	**0.249**	**0.229**	**0.254**	**0.249**	**0.741**	**0.672**	**0.771**		**0.786**	**0.194**	**0.284**	**0.149**
12	2EXH	**0.239**	**0.254**	**0.194**	**0.269**	**0.264**	**0.264**	**0.264**	**0.701**	**0.726**	**0.726**	**0.786**		**0.194**	**0.284**	**0.149**
13	3AKF	**0.189**	**0.204**	**0.557**	**0.214**	**0.204**	**0.219**	**0.209**	**0.189**	**0.194**	**0.184**	**0.194**	**0.194**		**0.204**	**0.109**
14	3LV4	**0.229**	**0.264**	**0.219**	**0.318**	**0.219**	**0.308**	**0.313**	**0.303**	**0.284**	**0.303**	**0.284**	**0.284**	**0.204**		**0.169**
15	3NQH	**0.174**	**0.239**	**0.134**	**0.154**	**0.204**	**0.194**	**0.194**	**0.124**	**0.169**	**0.139**	**0.149**	**0.149**	**0.109**	**0.169**	

Table 5-7: RMSD for Structures of GH43 enzymes.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Xsa43E		**2.183**	**2.361**	**2.246**	**1.516**	**2.371**	**2.338**	**2.058**	**2.006**	**2.002**	**1.966**	**1.994**	**2.465**	**2.266**	**2.527**
2	3KST	**2.183**		**2.179**	**1.858**	**2.100**	**2.069**	**2.001**	**1.881**	**1.834**	**1.867**	**1.881**	**1.854**	**2.205**	**2.190**	**2.490**
3	3K1U	**2.361**	**2.179**		**2.128**	**2.306**	**2.350**	**2.234**	**2.095**	**2.119**	**2.090**	**2.066**	**2.094**	**1.074**	**2.388**	**2.442**
4	1GYD	**2.246**	**1.858**	**2.128**		**2.137**	**1.152**	**1.204**	**1.909**	**1.894**	**1.871**	**1.881**	**1.905**	**2.111**	**1.856**	**2.433**
5	3C7E	**1.516**	**2.100**	**2.306**	**2.137**		**2.340**	**2.272**	**1.936**	**1.910**	**1.898**	**1.876**	**1.901**	**2.413**	**2.211**	**2.296**
6	1WL7	**2.371**	**2.069**	**2.350**	**1.152**	**2.340**		**0.641**	**1.969**	**1.993**	**1.975**	**1.963**	**2.018**	**2.347**	**1.900**	**2.660**
7	3CU9	**2.338**	**2.001**	**2.234**	**1.204**	**2.272**	**0.641**		**1.893**	**1.913**	**1.897**	**1.886**	**1.945**	**2.225**	**1.784**	**2.544**
8	3C2U	**2.058**	**1.881**	**2.095**	**1.909**	**1.936**	**1.969**	**1.893**		**0.509**	**0.473**	**0.478**	**0.467**	**2.191**	**1.935**	**2.623**
9	1YRZ	**2.006**	**1.834**	**2.119**	**1.894**	**1.910**	**1.993**	**1.913**	**0.509**		**0.466**	**0.495**	**0.440**	**2.193**	**1.937**	**2.553**
10	1Y7	**2.002**	**1.867**	**2.090**	**1.871**	**1.898**	**1.975**	**1.897**	**0.473**	**0.466**		**0.308**	**0.408**	**2.158**	**1.860**	**2.549**
11	1YIF	**1.966**	**1.881**	**2.066**	**1.881**	**1.876**	**1.963**	**1.886**	**0.478**	**0.495**	**0.308**		**0.388**	**2.150**	**1.831**	**2.547**
12	2EXH	**1.994**	**1.854**	**2.094**	**1.905**	**1.901**	**2.018**	**1.945**	**0.467**	**0.440**	**0.408**	**0.388**		**2.179**	**1.947**	**2.575**
13	3AKF	**2.465**	**2.205**	**1.074**	**2.111**	**2.413**	**2.347**	**2.225**	**2.191**	**2.193**	**2.158**	**2.150**	**2.179**		**2.445**	**2.455**
14	3LV4	**2.266**	**2.190**	**2.388**	**1.856**	**2.211**	**1.900**	**1.784**	**1.935**	**1.937**	**1.860**	**1.831**	**1.947**	**2.445**		**2.572**
15	3NQH	**2.527**	**2.490**	**2.442**	**2.433**	**2.296**	**2.660**	**2.544**	**2.623**	**2.553**	**2.549**	**2.547**	**2.575**	**2.455**	**2.572**	

When the structures are overlaid the three catalytic residues, equivalent to Asp-24, Asp-141 and Glu-202 in Xsa43E, are conserved in the same positions in all but one of the structures, regardless of the catalytic function of the enzyme. The structure (PDB accession code 3NQH) from *Bacteroides thetaiotaomicron*, does not have the catalytic base equivalent to Asp-24. There is no publication of the structure, or related activity, to indicate whether it is still catalytically active. As well as the catalytic residues all structures of GH43 family enzymes, except 3NQH, have the equivalent of the Tyr-35 residue conserved, as well as the WAP region (residues 86-88 in Xsa43E). All but 3NQH and 1UV4 also have the equivalent of His-258 conserved. 3NQH has been excluded from further comparisons as it shows very little conservation with the other structures. All structures have either a His, Pro or Ala residue preceding the catalytic base (Asp-24). At the position of Thr-39 all structures have either a Thr or

Ser residue. The equivalent residues of Phe-97 and Tyr-98 in the other GH43 structures are all Phe, Tyr or Trp residues. Trp-124 is highly conserved as either a Trp (12 structures) Phe (two structures) or Tyr (one structure). The residue after the catalytic acid is either a Gly or Ala in 13 other structures. Leu-162 is either a Leu or Ile in 13 of the other structures. Other residues highly conserved throughout this group of enzymes include Asp-62 Trp-66, Gly-95, Lys-99, Asp-147, Gly-150, Pro-204 and Tyr-214 which are all present in 12 other structures, Tyr-213 is present in 11 other structures.

5.2.2.5 Substrate Binding

In GH43 enzymes the active site lies in a groove that is shown to accommodate multiple sugar moieties. The substrate binding groove has defined subsites which accommodate these sugars. The subsites are labelled based on their proximity to the active site. Subsites -1 and +1 are the sites either side of the glycosidic bond being cleaved. In Xsa43E subsite -1 accommodates the arabinose side chain while subsite +1 accommodates the xylose sugar that the arabinose branches from.

Within the GH43 family of enzymes there are xylosidases, xylanases, arabinanases and arabinofuranosidases. Xsa43E shows the most similarity to the α-L-arabinofuranosidase from *Bacillus subtilis Bs*AXH-m2,3[141] (PDB accession code 3C7E), with a sequence similarity of 32%. The Cα trace of Xsa43E overlaid on *Bs*AXH-m2,3 has an RMSD of 2.0 Å. A stereo view of Xsa43E with *Bs*AXH-m2,3 is shown in

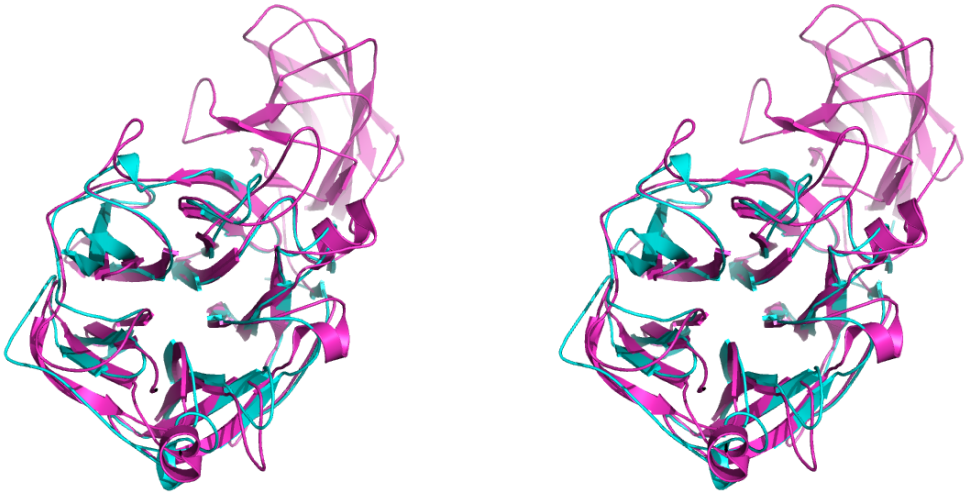


Figure 5-7: Wall-eyed stereo view of Xsa43E (cyan) overlaid with *Bs*AXH-m2,3 (pink).

The surface topography of Xsa43E has a pocket with the active site at its entrance and an overlying groove when viewed from the surface. This topography allows for a xylan backbone to be bound in the groove with room in the pocket for an arabinose sidechain. Based on other GH43 family enzyme substrate interactions it is possible to predict the residues that interact with the arabinose side chain of the substrate. The residues of importance can be seen in relation to xylobiose modelled from aligning the Xsa43E structure with that of *Bs*AXH-m2,3 in Figure 5-8.

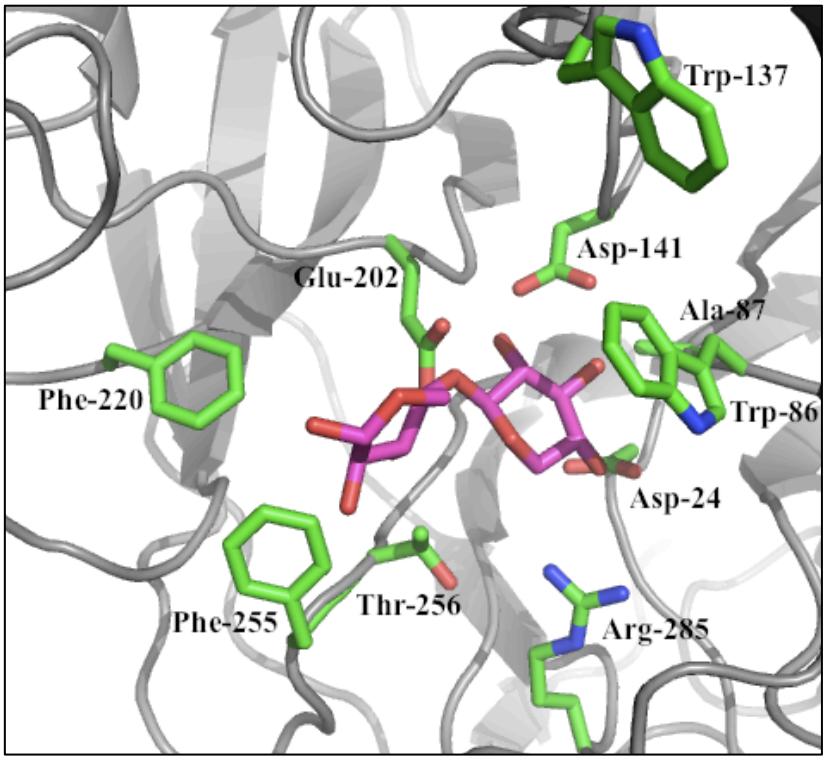


Figure 5-8: Xsa43E with xylobiose (pink) modelled in showing all residues believed to interact with the substrate. Xsa43E resiudues are coloured green for carbon, nitrogen blue and oxygen red.

The hydroxyl groups in L-arabinofuranose and D-xylopyranose have similar spatial arrangements, therefore by overlaying the structure of the *Geobacillus stearothermophilus* T-6 ß-xylosidase XynB3[140] (PDB accession code 2EXH, sequence identity with Xsa43E 22%) with the structure of Xsa43E some of the hydrogen bonding interactions with the substrate can be predicted (Cα trace between Xsa43E and XynB3 has RMSD of 2.6 Å). Asp-141 can form hydrogen bonds to the 2-O of the arabinose with both oxygen atoms in the side chain; it can also form a hydrogen bond from Oδ1 to the 3-O of the arabinose. The general base, Asp-24, can hydrogen bond to the 3-O via Oδ2. Arg-285 is retained in GH43 ß-xylosidases and α-L-arabinofuranosidases; it is not seen in the GH43 family α-L-arabinanases. Arg-285 has been shown in XynB3 to interact with the 4-O group of the xylose in the -1 subsite. Arabinose does not have an analogous 4-O group. In the *Bs*AXH-m2,3 structure the corresponding Arg-321 forms hydrogen bonds with the xylose in the +1

subsite. Trp-86 and Ala-87 form part of the WAP conserved region seen in all GH43 enzymes and have been shown to interact with the sugar in the -1 subsite; in the case of Xsa43E this would be the arabinose sidechain. Trp-86 forms hydrophobic stacking interactions with the sugar and the backbone N of Ala-87 has been shown to hydrogen bond to the 3-O. In Xsa43E these residues are conserved and the N of Ala-87 is seen interacting with the tris molecule in the structure therefore it is highly probable that these substrate interactions also occur in Xsa43E.

The structure of *Bs*AXH-m2,3 has been reported with partial substrate xylotetraose bound. To gain insight into the substrate binding mechanism of the xylan backbone in Xsa43E the two structures were overlaid. This revealed many of the same residues that interact with the xylan backbone in *Bs*AXH-m2,3 are also present in Xsa43E. Two residues in *Bs*AXH-m2,3 that form hydrophobic interactions with the xylan backbone are Phe-244 and Trp-160 and the corresponding residues in Xsa43E are Phe-220 and Trp-137. The side chain of Trp-137 in Xsa43E is offset approximately 90° from that of Trp-160 in *Bs*AXH-m2,3 but has the ability to move to take up this position on binding substrate. Also Phe-220 is in a slightly different position from Phe-244 of *Bs*AXH-m2,3 but it appears it is still able to interact with the substrate. Another residue, Phe-255, in Xsa43E appears to be situated in a position to allow hydrophobic stacking with the substrate as well. The xylose unit, in subsite +1, interacts with Asn-288 and the catalytic acid Glu-225 in the *Bs*AXH-m2,3 structure. The Asn-288 residue corresponds to Thr-256 in Xsa43E and the catalytic residue Glu-202 is in the same position as Glu-225 in *Bs*AXH-m2,3. Thr-256 from Xsa43E corresponds to the Thr-207 from the XynB3 and it has been reported for XynB3 that the backbone nitrogen of Thr-207 hydrogen bonds to the 3-O of the sugar in the +1 position[141]. This would be analogous to the xylose unit in which the arabinose group branches from in the substrate for Xsa43E. It is clear from comparisons with other GH43 enzymes that the substrate binding of the xylose and arabinose sugars either side of the glycosidic bond being cleaved are bound by hydrogen bonds, whereas the sugar units further from the active site are not held as firmly, instead the interactions

with the enzyme occur via hydrophobic interactions. This would allow oligosaccharides of varying chain length to be accommodated by the substrate binding groove.

A delineated structure of Xsa43E with the catalytic and proposed binding substrates indicated can be seen in Figure 5-9.

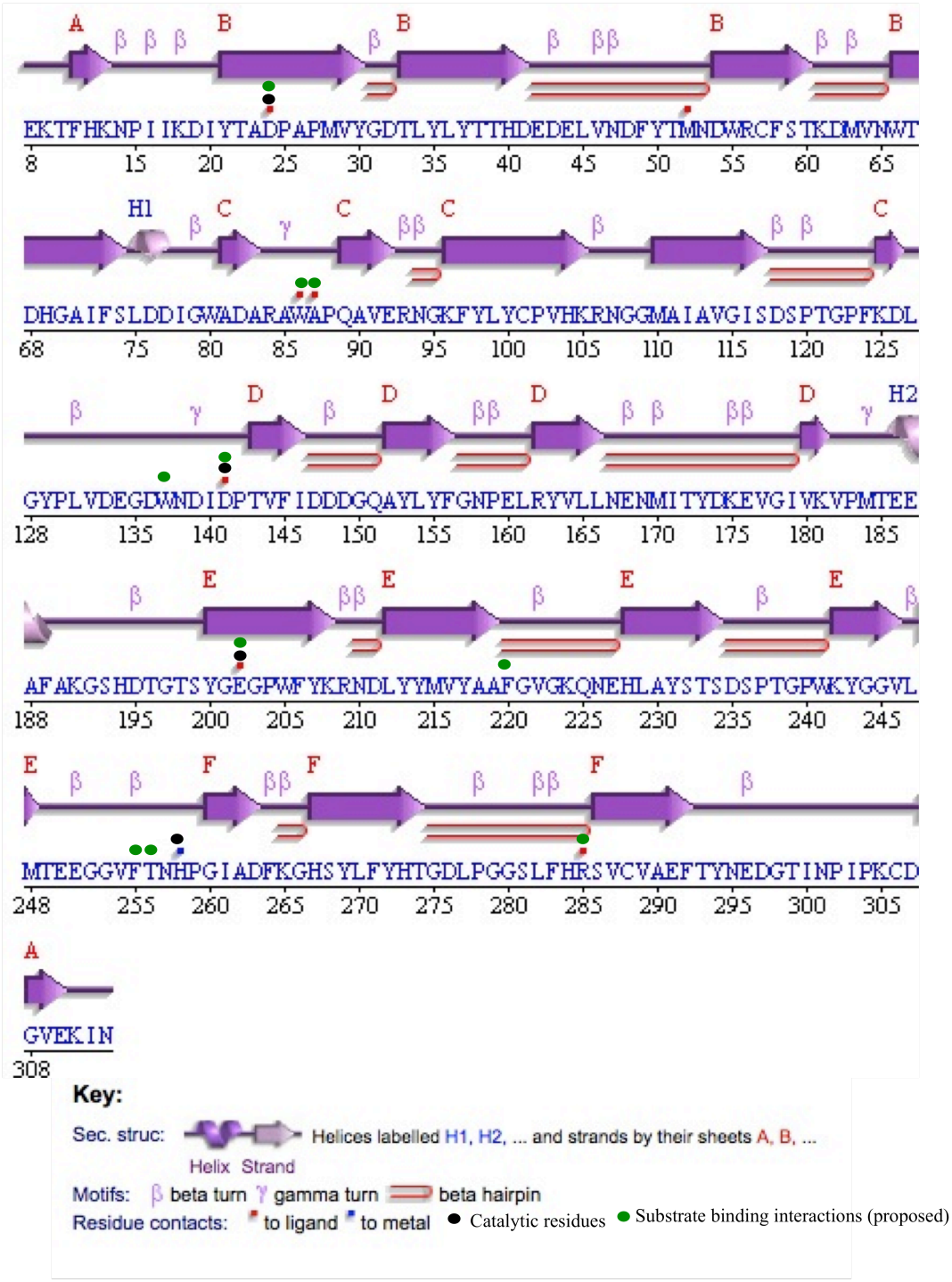


Figure 5-9: Delineated secondary structure of Xsa43E with key features

5.2.3 Elucidation of the Mechanism of Xsa43E

5.2.3.1 The GH43 Family proposed mechanism

The mechanism of GH43 family enzymes requires three catalytic residues, a general base, which activates a water molecule that then attacks the anomeric carbon of the substrate; a general acid, which donates a proton to the leaving group and a third residue that has been shown to be important for catalysis. This residue is suggested to have a range of functions including orientation of the general acid and substrate, stabilisation of the transition state and pK_a modulation of the general acid[139].

In the first structure of a GH43 family enzyme, Arb43A from *Cellvibrio japonicus* (PDB accession code 1GYD) a glutamic acid was identified as the catalytic acid (Glu-221), an aspartic acid (Asp-38) was proposed to be the catalytic base and a third residue, an aspartic acid (Asp-158) was suggested to function as a pK_a modulator as well as positioning the catalytic acid[137]. Individual substitution of these three residues in Arb43A inactivated the enzyme or considerably reduced the activity[142]. Based on the positions of the Asp-38 (6 Å from the anomeric carbon of the substrate) and Glu-221 (adjacent to the glycosidic oxygen atom at the site of bond cleavage) they were assigned as the catalytic base and catalytic acid respectively[142]. Sequence alignments of proteins from GH families 32, 43, 62 and 68, all of which contain the 5-bladed ß-propeller structure, revealed that the three catalytic acids mentioned above are conserved throughout[142]. While GH 43 enzymes have an inverting mechanism the other three families have a retaining mechanism. This difference results from the positioning of the catalytic acid and base. With the retaining mechanism these two residues are about 5 Å apart whereas in an inverting mechanism they are further apart[143]. The increased distance in the inverting mechanism is a function of the residues being on opposite sides of the substrate when it is in the active site.

The ß-xylosidase, XynB3, from *Geobacillus stearothermophilus* (PDB accession code 2EXI) identified the equivalent three acidic residues as being important for

catalysis; again the same role for each residue was given[140]. Asp-128 (equivalent to Asp-158 from Arb43A) was shown to be 4 Å from the general acid, it was also shown to interact with the substrate leading to it being assigned the role of positioning the general acid, modulating its pK_a, orientating the substrate and stabilising the transition state[140].

Although many more structures of GH43 family enzymes have been published identifying the three equivalent catalytic residues and showing all to be important for catalysis, there is a lack of experimental evidence to confirm the exact role of this third catalytic acid.

5.2.3.2 pK_a Modulation of the Catalytic Acid Glu-202.

Previous GH43 family enzyme publications nominate the equivalent of Asp-141 as a pK_a modulator of the catalytic acid[138,142-144]. The calcium ion at the centre of the domain has the potential to influence Glu-202 through an electron withdrawing pathway via His-258 which then hydrogen bonds to a water molecule which is also in hydrogen bonding distance of the carboxylic acid of of Glu-202 (Figure 5-10). This electron withdrawing pathway indicates the calcium ion and His-258 maybe the true pK_a modulators of Glu-202 allowing it to be protonated in the enzymatic environment at neutral pH. This represents a pK_a shift of ~2.7 pH units which is equivalent to ΔG for the interaction of 16.19 kJ.mol^{-1}.

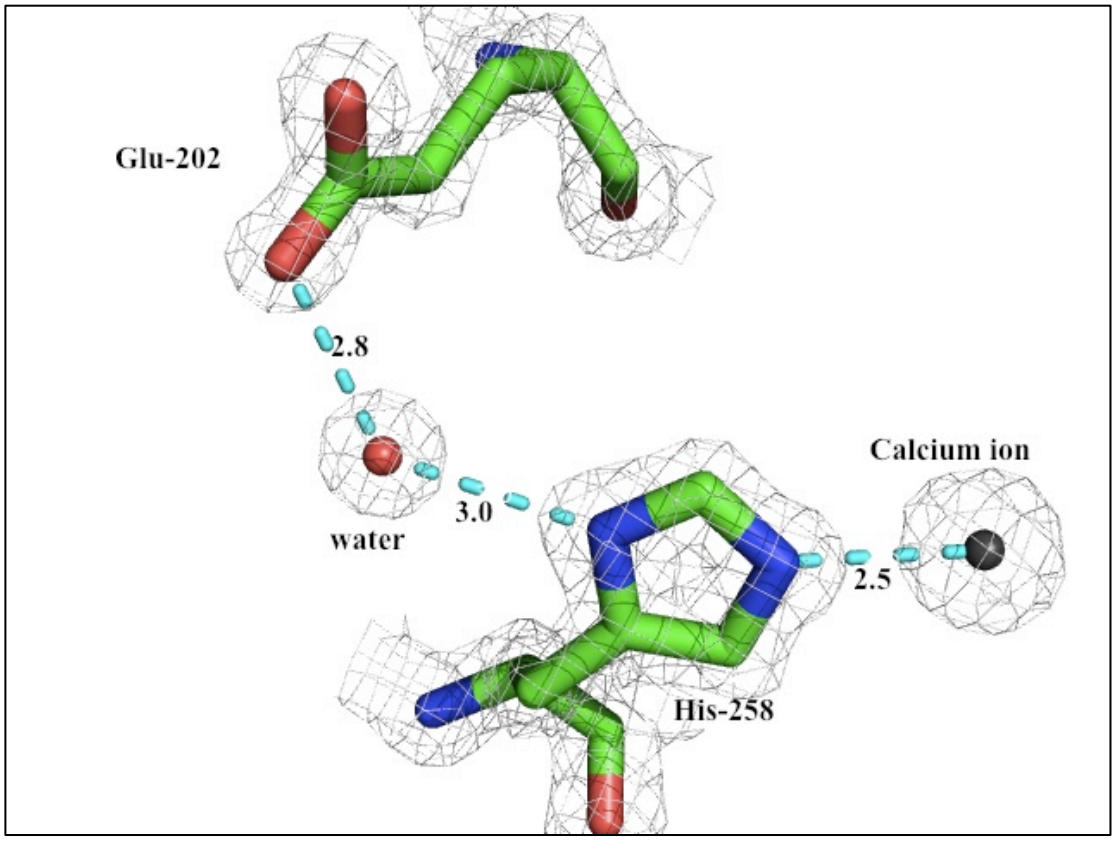


Figure 5-10: Electron withdrawal pathway from Glu-202 via His-258 by the Calcium ion. Electron Density is 2 | FO | - | FC | map contoured at 1.0 σ.

Of the 16 GH43 family structures all have the histidine residue conserved except for two, 1UV4 and 3NQH. 1UV4 is an arabinanase from *Bacillus subtilis* and has two water molecules corresponding to the positions of the two nitrogen atoms of the imidazole ring of the histidine. These water molecules are 2.7 Å apart, which implies they are within hydrogen bonding distance. Along the backbone of the chain there is a glycine in the equivalent position of the histidine; this structure is unique among the GH43 structures as it has 2 ions modelled as calcium in the centre of the structure[144], one is analogous to that seen in Xsa43E, the other is 6 Å from the first calcium on the opposite side to the active site and is interacting with three acidic residues Asp-77, Glu-135 and Asp-257. 3NQH does not have the central calcium ion or the histidine but has an arginine residue in a similar position to the histidine. There is no published

information on this structure but it appears to have a different mechanism of action as it is also missing the catalytic base.

Not all of the structures have identified the ion at the centre of the structure as being calcium, and only eight structures have an ion with seven co-ordinate geometry in the central position.

Of the 16 GH43 family enzymes that have structures available on the PDB database, eight have the water molecule conserved in the position between the analogous histidine and glutamic acid residues; the other eight structures do not. Of the eight structures that do have the water present, the enzymes have varying functions including xylosidases, xylanases and arabinofuranosidases, indicating this mechanism is not isolated to arabinofuranosidases. Of the eight structures that are missing the water, two are 1UV4 and 3NQH which also lack the histidine residue and the other six have a tyrosine with the OH from the side chain in the position of the water molecule, 2.6 Å from the catalytic acid. The presence of the tyrosine would allow the electron-withdrawing pathway to be maintained.

Xsa43E has a pathway by which the central calcium ion can influence the catalytic acid Glu-202, this pathway may also be present in other GH43 family enzymes.

5.2.3.3 Substrate Specificity of Xsa43E

Xsa43E was tested on a range of natural and synthetic substrates in order to understand the substrate specificity of the enzyme. Xsa43E was able to cleave arabinobiose with a specific activity of 0.002 μmol.min^{-1}.mg^{-1}; but showed no activity on arabinohexose. The inability to cleave the longer arabinose oligosaccharide indicates Xsa43E is not an arabinanase. It showed no ability to cleave ß-D-xylo-oligosaccharides with a degree of polymerisation between two and six. Xsa43E cleaved arabinose units from the commercially available substrate arabinoxylan with a specific activity of 0.010 μmol.min^{-1}.mg^{-1}. When Xsa43E was

tested for activity against arabinoxylan in conjunction with a xylanase from *Trichoderma virdie*, the amount of arabinose cleaved was significantly higher at 0.025 μmol.min^{-1}.mg^{-1}. Specific activity for all substrates tested is shown in Table 5-7. The active site pocket and the positioning of the catalytic residues indicated that Xsa43E would cleave arabinose from the xylose backbone of arabinoxylan. The results shown in Table 5-7 support this and also indicate that Xsa43E acts more efficiently on short arabinoxylan segments given its increased activity with the addition of a xylanase. The xylanase cleaves the xylan backbone into smaller segments, which may allow Xsa43E better access to cleavage sites. Xsa43E is a non-secreted enzyme and as *B. proteoclasticus* can transport short oligosaccharides into the cell[1], this may be why Xsa43E works more effectively on short arabinoxylan pieces.

Table 5-8: Specific Activity Investigation for Xsa43E with a range of substrates.

Enzyme	Substrate	Specific Activity (μmol.min^{-1}.mg^{-1})
Xsa43E	Arabinobiose	0.0020
	Arabinohexose	0
	Xylobiose	0
	Xylotriose	0
	Xylotetraose	0
	Xylopantose	0
	Xylohexose	0
	Hemicellulose	0
	Ryegrass	0
	Arabinoxylan	0.010
Xsa43E with Xylanase from *T. virdie*	Arabinoxylan	0.025

Xsa43E was also tested on a range of different glycosidic linkages in model substrates. Xsa43E can cleave the glycosidic bond in *p*-nitrophenyl-α-L-arabinofuranose (*p*-NP-L-A) and *p*-nitrophenyl-ß-D-xylopyranose (*p*-NP-D-X), but cannot cleave the glycosidic bond in *p*-nitrophenyl-α-D-xylanopyranose.

5.2.3.4 Specific Activity of Xsa43E on Model Substrates

In order to observe the rate of degradation of substrate by Xsa43E, two model substrates *p*-NP-L-A and *p*-NP-D-X were tested. The rate of degradation at various substrate concentrations is shown in Figure 5-11 for both substrates. Kinetic parameters V_{max}, K_M, k_{cat}, and k_{cat}/K_M were calculated for both substrates (Table 5-9) using Prism5 GraphPad.

p-nitrophenyl-α-L-arabinfuranose *p*-nitrophenyl-ß-D-xylopyranose

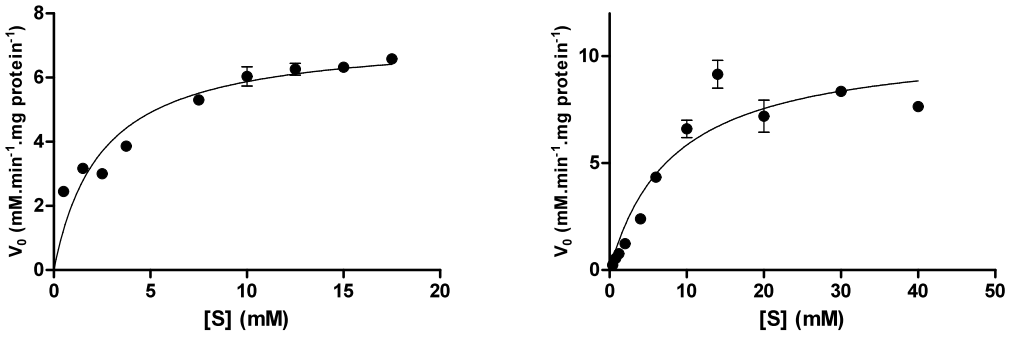


Figure 5-11: Rate of Xsa43E activity on model substrates at varying substrate concentrations. Errors are standard errors.

Table 5-9: Kinetic data for Xsa43E with model substrates *p*-nitrophenyl-α-L-arabinofurnaoside (*p*-NP-L-A) and *p*-nitrophenyl-β-D-xylopyranoside (*p*-NP-D-X) with standard errors.

Substrate	V_{max} (mM)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}.s^{-1}$)
PNP-L-A	0.022 ± 0.001	3.6 ± 0.5	0.70 ± 0.03	0.20 ± 0.04
PNP-D-X	0.077 ± 0.006	8 ± 2	1.25 ± 0.10	0.15 ± 0.05

It is important to note that the K_Ms for both substrates are high and this indicates these are not ideal. This is most likely due to the leaving group being a substituted phenol ring as opposed to a sugar moiety; however due to the ability to follow these reactions in real time they were used to establish rate data. It is not clear if the kinetic

profile of Xsa43E with *p*-NP-D-X (Figure 5-11) is showing product inhibition at the higher concentrations of substrate; therefore it was treated as if there was no product inhibition.

The results show that *p*-NP-L-A is the preferred substrate in terms of binding, indicated by the lower Michealis-Menten constant (K_M) which represents the dissociation constant of the substrate-enzyme complex[54]. The catalytic constant, k_{cat}, shows that for *p*-NP-L-A the rate of hydrolysis is slower than for *p*-NP-D-X. Although the catalytic efficiency (k_{cat}/K_M) values for Xsa43E with the two substrates indicate that *p*-NP-L-A is the preferred substrate the two values are not significantly different. It is important to note that D-xylopyranoside and L-arabinofuranoside are very similar in terms of the spatial positioning of the hydroxyl groups therefore it is reasonable that the activity of Xsa43E on both model substrates is similar.

5.2.3.5 pH and Temperature Range of Xsa43E

The activity of Xsa43E was tested across a pH range from 4.0 to 9.9. Xsa43E is active above pH 4.6 and below 9.9 (Figure 5-12). The pH optima is between 6.8 and 7.3; the pH of the rumen is maintained close to pH 7 but may drop lower due to volatile fatty acid production[19].

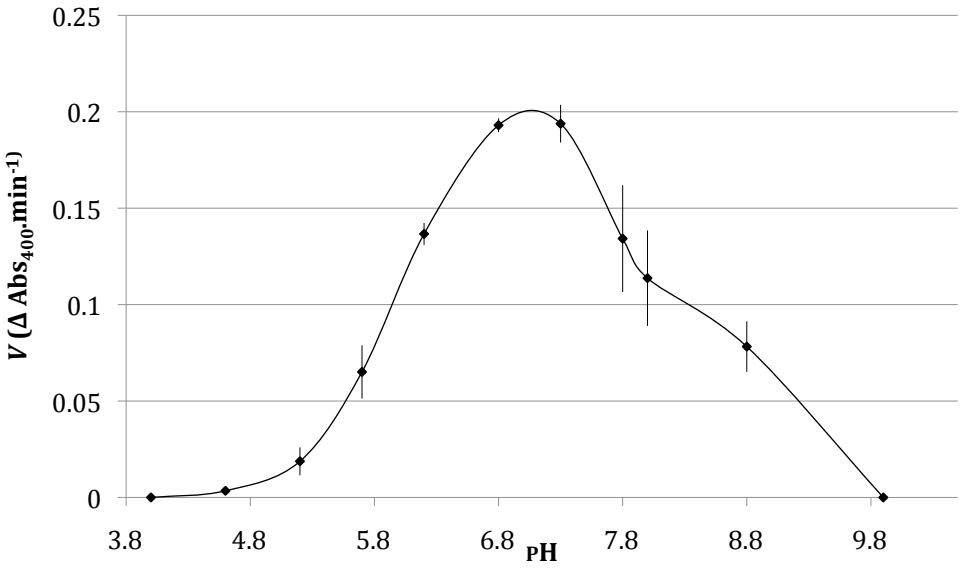


Figure 5-12: Rate of *p*-NP-L-A degradation by Xsa43E across a range of pHs. Corrected for non-enzymatic decay. Error bars represent standard deviations.

The activity of Xsa43E was also measured across a range of temperatures from 18 °C to 52 °C as shown Figure 5-13. It showed an increase in activity up to a maximum of 47 °C before falling dramatically. The rumen is maintained at 40 °C ± 1 °C[19] and Xsa43E shows maximum activity close to that temperature.

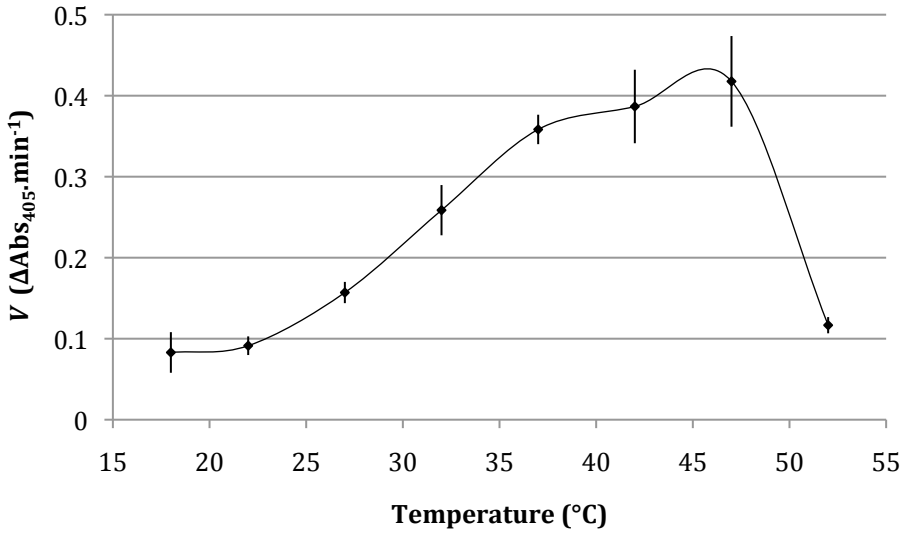


Figure 5-13: Temperature profile of Xsa43E. Error bars represent standard deviation.

5.2.3.6 Effects of Metal Ions on Xsa43E

At the centre of the Xsa43E structure is a calcium ion in pentagonal bipyramidal coordination. I have identified an electron withdrawing pathway that indicates the calcium may alter the pK_a of the catalytic acid. The five equatorial ligands of the calcium ion each interact with two consecutive blades of the ß-propeller structure and this appears to help organise the active site and potentially the entire domain. Xsa43E was tested to see if the addition of different metal ions would alter the activity. A range of divalent cations was added to Xsa43E at final concentrations of 1 mM and 10 mM and EDTA treated Xsa43E at a final concentration of 1 mM (Figure 5-14).

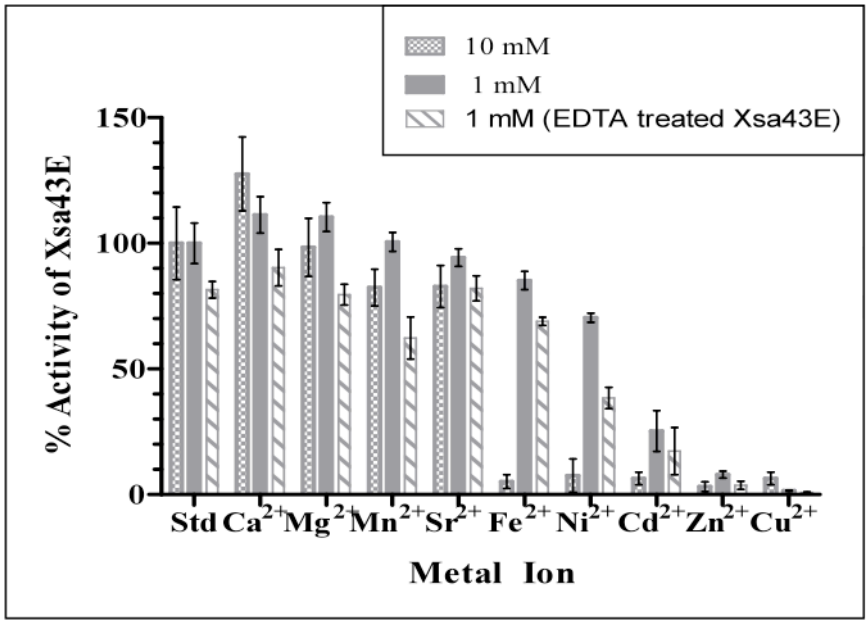


Figure 5-14: Effects of divalent cations on the activity of Xsa43E with *p*-NP-L-A. Activity is displayed as a percentage with standard errors, with 100% being the activity of the untreated enzyme (std).

Addition of Ca^{2+} at 10 mM increased the activity of the enzyme by 27% and this could be due to calcium displacing other metals that may be occupying the active site of some of the enzyme molecules, at 1 mM it increased the activity by 11% but this was not significant due to an overlap within errors. Mg^{2+}, Mn^{2+} and Sr^{2+} did not have significant effects on the activity of Xsa43E at either concentration. Addition of 10 mM divalent transition metal cations, other than Mn^{2+}, caused a dramatic decrease in the activity. The corresponding solutions contained precipitate and attempts to establish the corrected concentrations of soluble protein were unsuccessful. It was assumed that the majority of the protein had precipitated from solution. At 1 mM Fe^{2+} and 1 mM Ni^{2+} the activity decreased to 85% and 70% respectively; addition of 1 mM Cd^{2+} decreased the activity to 25% whereas 1 mM Zn^{2+} and 1 mM Cu^{2+} both reduced the activity to less than 10%. Treatment of Xsa43E with EDTA prior to the addition of 1 mM metal ions decreases the activity for all metal ions compared with the untreated enzyme; however the trend across the metal ions is similar to the untreated enzyme.

It is believed that the coordination of the central ion is important for the overall structure of Xsa43E and placement of the active site residues via the proline residues outlined in section 5.2.2.3, as well as activity by altering the pK_a of the catalytic acid. Calcium is an alkali earth metal from group II of the periodic table; unlike transition metals, group II elements have a general pattern of behaviour due to their valence electron configuration. Mg, Ca and Sr all belong to Group II, therefore it is assumed that if these metals displace the central calcium ion they can still maintain the environment in a similar way to the calcium. The transition metals, other than Mn^{2+}, reduced the activity of the enzyme; this could be due to their inability to maintain the pentagonal bi-pyramid coordination potentially resulting in the active site of the enzyme being distorted. Alternatively, they may cause the enzyme to unfold due to the five blades not being organised because of the altered geometry of the central ion. Interestingly the ß-xylosidase, XynB3, from *Geobacillus stearothermophilus* showed no decrease in activity with metal ions Fe^{3+}, Ca^{2+},Co^{2+}, Mg^{2+}, Ni^{2+}, Zn^{2+} and Cu^{2+} present in the reaction solution[143].

5.2.3.7 Effects of EDTA on Xsa43E

EDTA is a chelating agent that strongly binds to divalent metal ions in solution. Addition of EDTA decreases the activity of Xsa43E at concentrations above 10 mM. Figure 5-15 shows the effect of increasing EDTA concentrations on Xsa43E and Figure 5-16 shows the effect of 50 mM EDTA on Xsa43E over 24 hours.

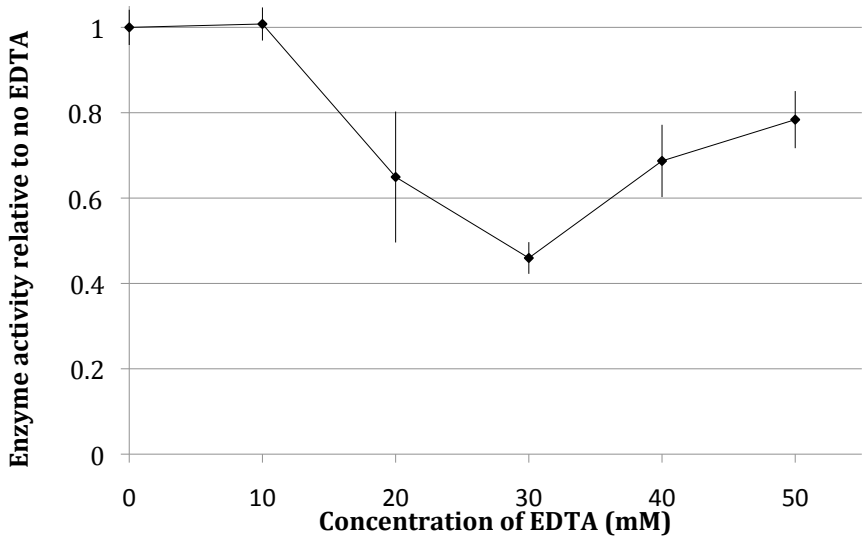


Figure 5-15: Relative activity of Xsa43E with increasing concentrations of EDTA. Activity is displayed as a decimal of the activity of Xsa43E with no EDTA. Error bars represent standard deviations.

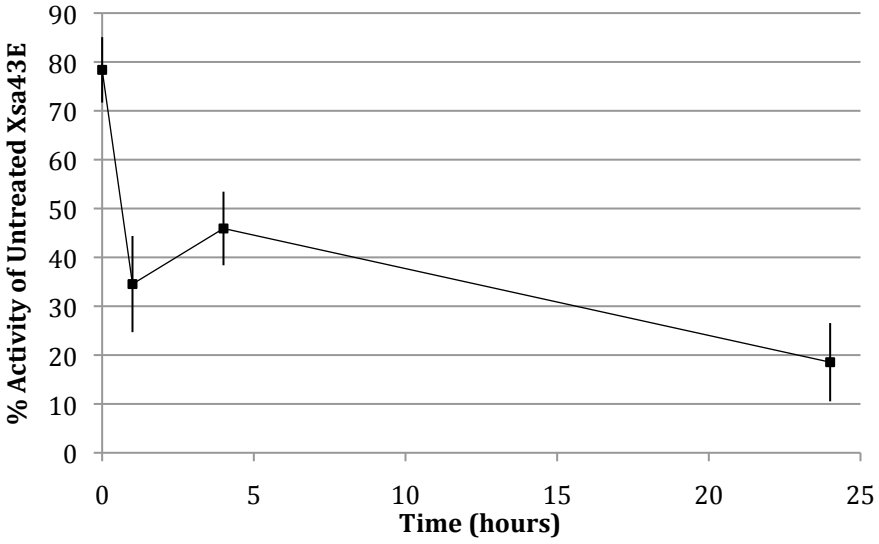


Figure 5-16: Time course showing relative activity of Xsa43E with 50 mM EDTA compared with untreated enzyme over 24 hours. Error bars represent standard deviations.

Although there is no clear trend in enzyme activity with EDTA concentration when the enzyme is tested immediately after addition of EDTA, there is evidence that when

the enzyme is incubated with EDTA over a period of time the activity is reduced. This supports the significance of the calcium ion to the enzymatic activity. By removing the calcium ion the electron withdrawing pathway is removed and the pK_a of the catalytic acid Glu-202 is no longer affected meaning at a pH of 7.2 it would be unlikely to be protonated and thus unable to act as the proton donor. The reduction in activity is believed to be due to the calcium being removed from the enzyme by the chelating EDTA and as this is in equilibrium, only some of the enzyme population has the calcium removed. The ß-xylosidase, XynB3, from *G. stearothermophilus* reported no drop in activity with 1 mM EDTA[143] and the arabinanase Abn2 from *B. subtilis* also showed no decrease in activity with 1 mM EDTA[145]. These concentrations were final concentrations in the reaction mixture whereas in the experiments with Xsa43E, the enzyme was incubated with higher concentrations before being diluted 40 fold in the reaction mixture.

5.2.4 Mechanistic Investigation through Site Directed Mutagenesis

The residues involved in the catalysis of Xsa43E are Asp-24, Asp-141, Glu-202 and potentially His-258. The first three are well known to be necessary for catalytic activity of the enzyme. Glu-202 is the catalytic acid that donates a proton to the leaving group. The His-258 residue is positioned between the calcium ion in the centre of the structure and the active site. In the structure of Xsa43E there is a clear electron withdrawing pathway between the calcium ion and the catalytic acid via His-258 and a water molecule (Figure 5-10). This, combined with the fact that Asp-141 is at least 3.8 Å away from Glu-202, led to the investigation of the role of all four residues in catalysis. Each residue was individually mutated to alanine, the mutated enzymes were expressed, purified and tested for activity under appropriate conditions. The mutants of the main catalytic residues, Asp-24 and Glu-202, were tested for activity under standard conditions. The mutants of the other residues, Asp-141 and His-258, were originally tested under standard condition then also tested

across a broad pH range in order to determine which was the true pK_a modulator.

5.2.4.1 Expression and Purification of Xsa43E Mutants

Expression and purification for all mutants was performed as for the wildtype enzyme. All mutants had similar purification traces and SDS-PAGE gels therefore purification of Xsa43E H258A is illustrated as a representation of all mutants (Figure 5-17). The chromatogram shows the Xsa43E mutants elute in a single peak at ~150 mM imidazole.

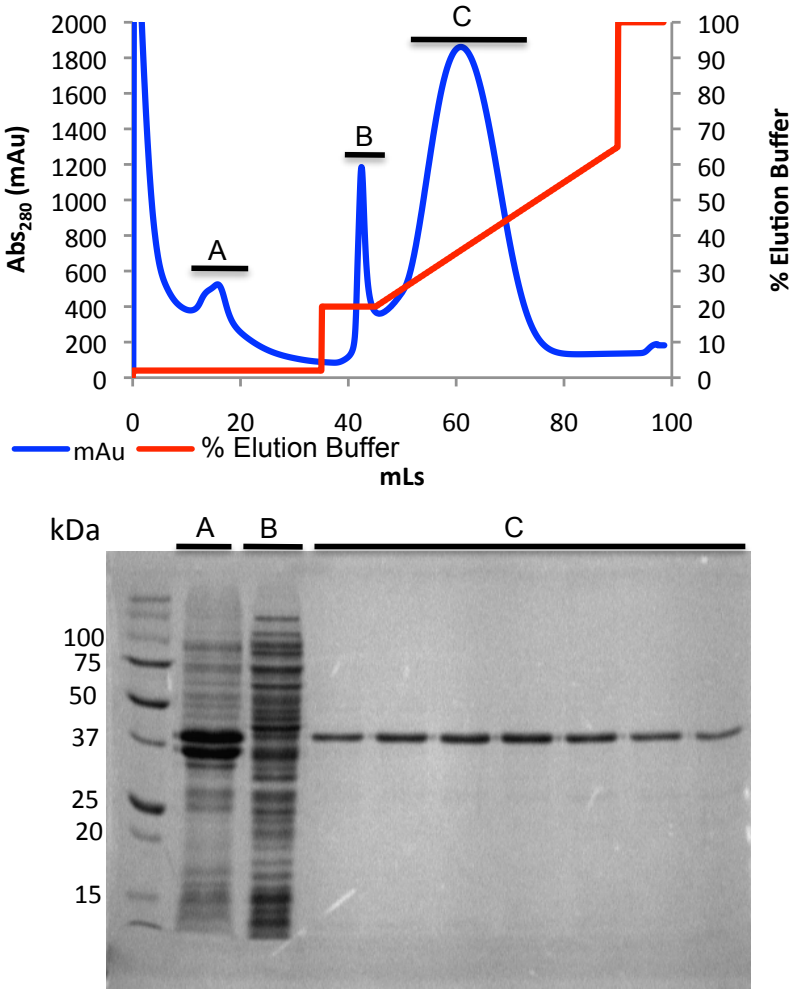


Figure 5-17: IMAC chromatogram and corresponding 12% SDS-PAGE gel for Xsa43E H258A.

Fractions from peak C of Figure 5-17 were further purified using an S200 16/60 column, the chromatogram shows the Xsa43E mutants elute in a single peak at ~85 mL (Figure 5-18), this is equivalent to a M_r of 38 kDa indicating the mutants are also monomers in solution.

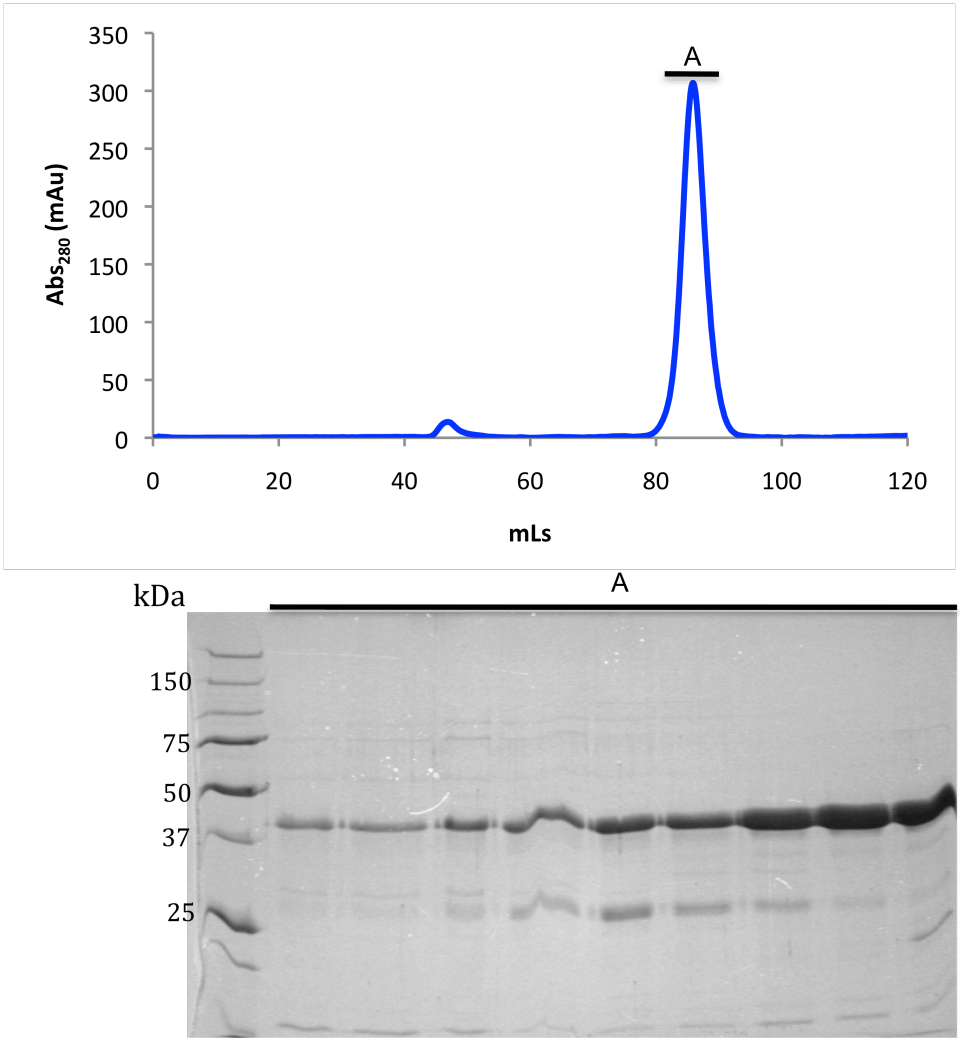


Figure 5-18: S200 16/60 purification chromatogram of Xsa43E H258A and corresponding 12% SDS-PAGE gel for Xsa43E H258A.

The mutants all express in substantial quantities and purify in a single peak in a similar position to the wildtype.

5.2.4.2 Activity of Mutant Enzymes

The four mutants, D24A, D141A, E202A and H258A, were tested for activity against the model substrate *p*-NP-L-A over a time course of 24 hours at pH 7.2. Values for the time points are reported in terms of ΔAbs_{400} since time zero hours. All of the mutations effectively remove all significant activity (Table 5-10).

Table 5-10: Activity of Xsa43E mutants with *p*-nitrophenyl-α-L-arabinofuranoside over 24 hour time course. Numbers given are change in absorbance from time zero at 400 nm corrected for non-enzymatic decay of the substrate.

	ΔAbs_{400}			Relative
	Time (Hours)			Activity
Sample	1	4	24	(24 hrs)
Xsa43E	0.010	0.101	0.408	100%
Xsa43E D24A	-0.002	0.001	-0.020	0%
Xsa43E D141A	-0.010	0.000	0.011	3%
Xsa43E E202A	-0.009	0.001	-0.003	0%
Xsa43E H258A	-0.005	0.003	-0.006	0%

The absorbance values of the assays were corrected for non-enzymatic decay of the substrate (i.e. the blank values) and also reported as a percentage of the wildtype. The only mutant with any activity above the blank was Xsa43E D141A at 3% of the wildtype enzyme's activity, this is not a considerable amount of activity therefore further investigation is required to establish the full catalytic capability of the Xsa43E D141A mutant. However these data support the hypothesised role of Asp-141 in the positioning of the substrate and the catalytic acid; without Asp-141 the efficiency of the enzyme is significantly reduced, almost to zero, but the catalytic mechanism can still function if the activity seen is real. Due to the low levels of activity, kinetic parameters such as K_M and k_{cat} could not be determined for Xsa43E D141A because the substrate was not soluble at high enough concentrations. The data

in Table 5-10 shows the importance of His-258 and supports the hypothesis that His-258 is the true pK_a modulator of the catalytic acid. This hypothesis was investigated further by determining the structure of Xsa43E H258A.

5.2.4.3 Structure of Xsa43E H258A

Crystals of Xsa43E H258A grew in the condition 20% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.2 with a protein concentration of 12 mg.mL^{-1}. A dataset of Xsa43E H258A was collected at the Australian synchrotron. Complete data statistics and refinement statistics can be seen in Table 5-11 and Table 5-12 respectively. The structure of Xsa43E H258A has 95.4% of the residues within the most favoured region and 4.6% in allowed regions of the Ramachandran plot (Structure validation using ProCheck[90] in CCP4).

Table 5-11: Complete data statistics for the native dataset of Xsa43E H258A.

Xsa43E H258A	
Space group	$P2_12_12_1$
Wavelength	0.9537 Å
Cell parameters	
a	58.74 Å
b	76.57 Å
c	87.86 Å
α	90°
β	90°
γ	90°
Resolution range (Å)	76.6-2.6 (2.74-2.60)
R merge	0.087 (0.317)
Number of measured reflections	83130 (12312)
Number of Unique reflections	12320 (1767)
Meal I/σI	17.0 (5.7)
Completeness	100 (100)
Multiplicity	6.7 (7.0)

Table 5-12: Refinement Statistics for Xsa43E H258A.

Refinement Statistics	
RWork (*Rfree*)	15.5% (21.3%)
Total No. Atoms	2675
No. of Protein Atoms	2485
Other Molecules/Ions	9 (1trs 1ca)
No. of Waters	320
RMS Deviation from Standard Geometry	
Bond Lengths (Å)	0.02
Bond Angles (°)	1.759
Average B-factors	
Protein	24.635
Water	29.844

The structure of Xsa43E H258A was the same as that for Xsa43E except that the histidine residue has been replaced with an alanine. The RMSD between the C-α trace for Xsa43E and Xsa43E H258A is 0.4 Å. The two structures of Xsa43E and Xsa43E H258A have minimal difference when overlaid (Figure 5-19 (a)) and the active sites are also essentially identical (Figure 5-19 (b)).

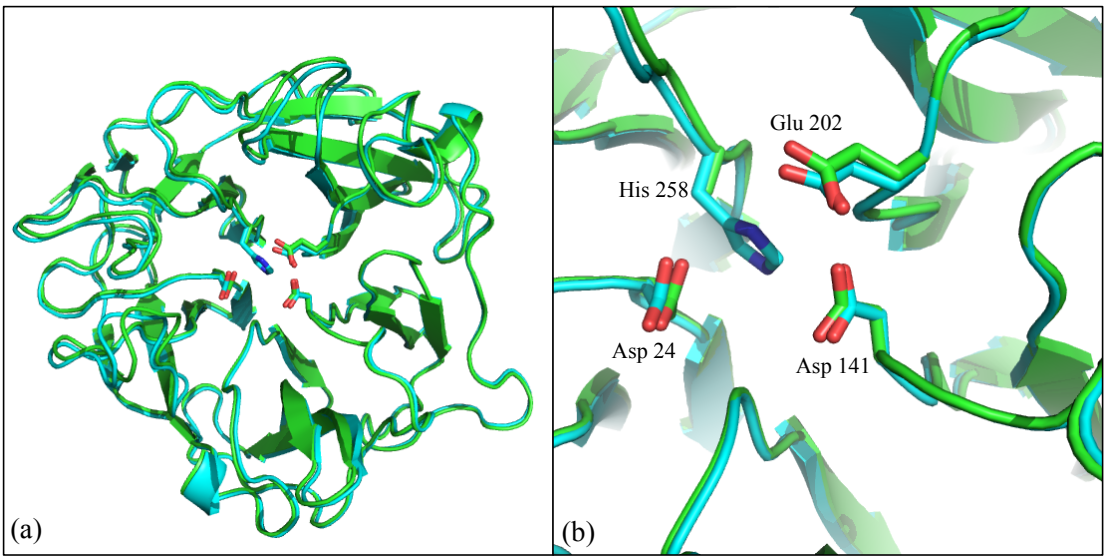


Figure 5-19: Overlay of Xsa43E and Xsa43E H258A. (a) shows the entire structures of Xsa43E (cyan) and Xsa43E H258A (green) overlaid and (b) shows the active site residues of the two structures.

The central calcium ion still has a seven co-ordinate geometry but with an extra water molecule replacing the histidine residue as a ligand (Figure 5-20). Attempts to cocrystallise Xsa43E with arabinobiose were unsuccessful; crystals were grown, and data collected and processed but there was no electron density present for the substrate. This could be due to the crystals being grown at pH 4.2; the low pH is likely to disrupt the binding of the substrate.

Future work on this mutant could include screening crystallisation conditions with a more neutral pH in attempts to get substrate bound in the crystals.

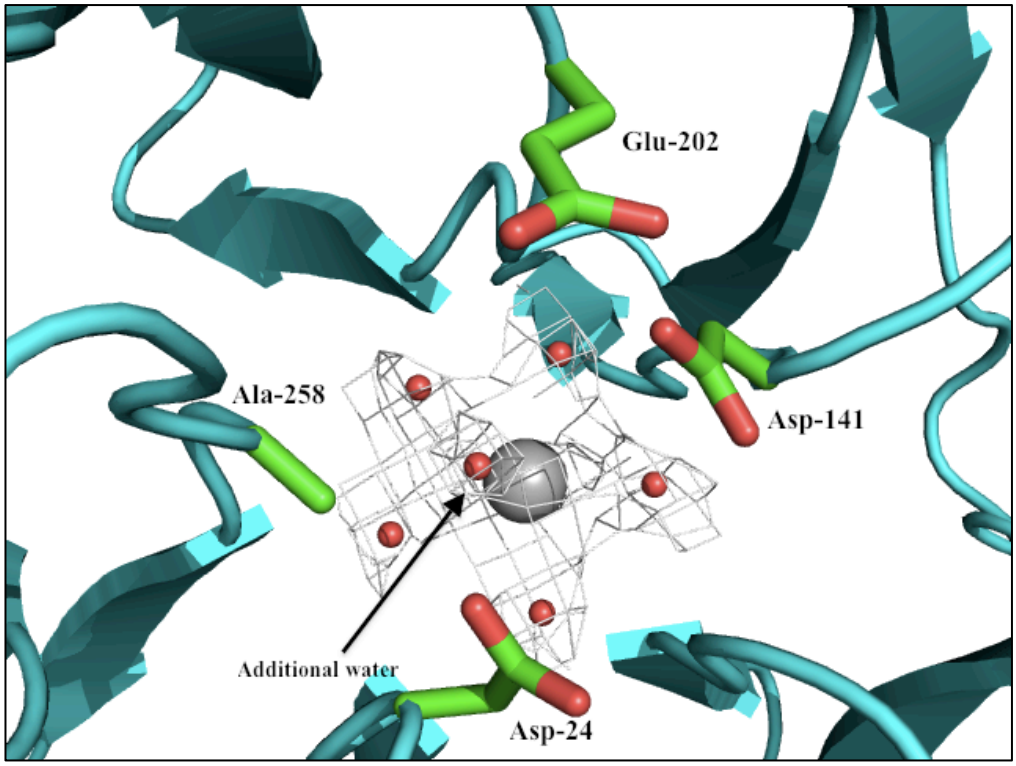


Figure 5-20: Active site of Xsa43E H258A with additional water ligand replacing the histidine previously co-ordinated to the central calcium ion. Electron Density is 2 | Fo | - | Fc | map contoured at 1.0 σ.

The activity of Xsa43E H258A demonstrated that the histidine is important for catalysis (section 5.2.4.2). The histidine was then replaced with a functional group that would allow the electron withdrawing pathway from the calcium ion to the

Glu-202 to be re-established. The residues that were most suitable were those with an amine functional group on the side chain; asparagine and glutamine. It was determined from the crystal structure of Xsa43E that glutamine would position the functional amine group in the position of the nitrogen atoms from the histidine ring better than an asparagine residue; therefore Xsa43E H258Q was cloned and characterised.

5.2.4.4 Xsa43E H258Q Crystals

The best crystals of Xsa43E H258Q grew in the condition 22% PEG 8000, 0.2 M NaCl, 0.1 M phosphate/citrate pH 4.6; protein was at a concentration of 4 $mg.mL^{-1}$. The crystals were tested (Australian Synchrotron) in multiple places along the crystal using the synchrotron beam. The three crystals gave no diffraction. These crystals were formed in the same way as the Xsa43E H258A crystals except at a lower concentration of protein.

5.2.4.5 Activity of Xsa43E H258Q

The activity of Xsa43E H258Q was established for the model substrate *p*-NP-L-A (Figure 5-21).

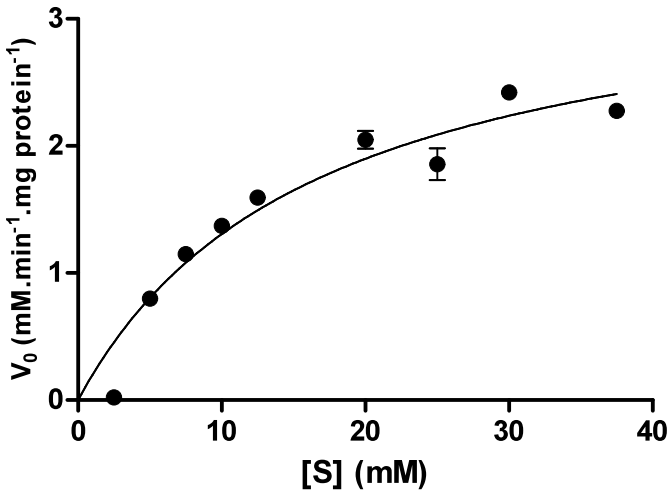


Figure 5-21: Rate of Xsa43E H258Q on *p*-nitrophenyl-α-L-arabinose, errors are standard errors.

The kinetic parameters are compared with Xsa43E (Table 5-13). The K_M increased by more than a factor of four and the k_{cat} decreased more than three fold. Although the H258Q mutant retains activity, the activity is significantly reduced and the overall efficiency of the enzyme, indicated by k_{cat}/K_M, is also reduced.

Table 5-13: Kinetic parameters of Xsa43E and Xsa43E H258Q with the model substrate *p*-nitrophenyl-α-L-arabinofuranoside.

Enzyme	Substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}.s^{-1}$)
Xsa43E	PNP-L-A	3.6 ± 0.5	0.70 ± 0.03	0.20 ± 0.04
Xsa43E H258Q	PNP-L-A	16.6 ± 3.2	0.20 ± 0.02	0.012 ± 0.003

The degradation of PNP-L-A by Xsa43E H258Q is very slow. The degradation of substrate is confirmed to be due to the presence of enzyme as blanks were run with every sample to account for any non-enzymatic decay of substrate. Also samples were run in triplicate and the results were consistent.

The ability of the H258Q mutant to maintain activity while the H258A mutant does not illustrates the importance of the calcium ion on the pK_a of the catalytic acid Glu-202. The imidazole group of a histidine and the amine group of a

glutamine are both polar allowing the calcium ion to elicit an effect on the glutamic acid residue and alter its pK_a; an alanine residue cannot facilitate this.

5.2.5 Comparison of Alternate pK_a Modulators Asp-141 and His-258

Xsa43E D141A retained a small amount of activity when tested in section 5.2.4.1. Further investigation into the activity profile of Xsa43E D141A, Xsa43E H258A and Xsa43E H258A was necessary to determine which residue (Asp-141 or His-258) is the p*Ka* modulator in the native enzyme. The activity of Xsa43E D141A, Xsa43E H258A and Xsa43E H258Q were tested against *p*-NP-L-A over a broad pH range (Figure 5-22).

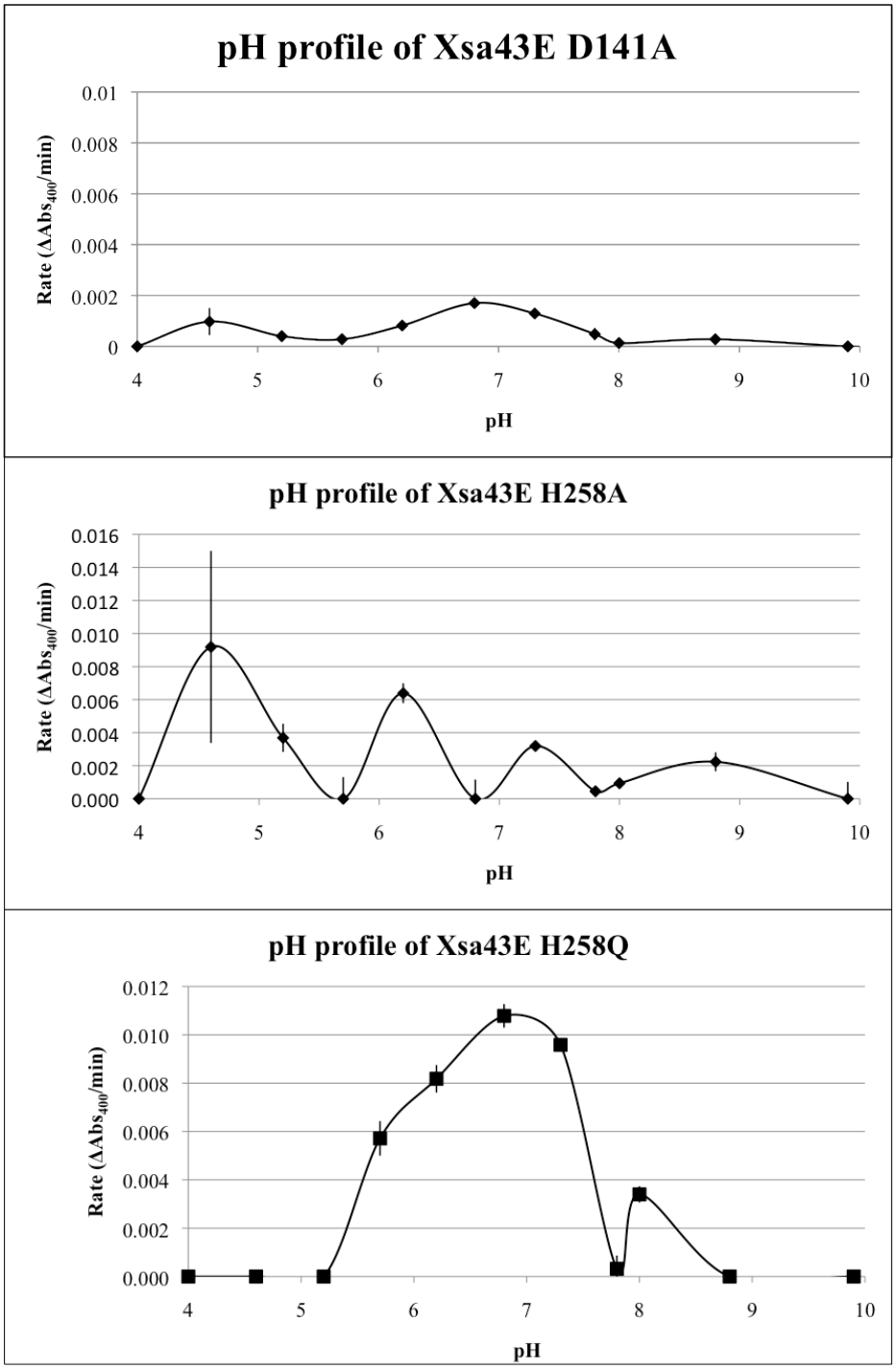


Figure 5-22: pH profile of Xsa43E D141A, Xsa43E H258A and Xsa43E H258Q. Errors are standard deviations.

Optimal activity of Xsa34E D141A was observed at pH 6.8 with some activity at pH 4.6, 6.2 and 7.3. Xsa43E H258A has activity equivalent to that of the

wildtype enzyme at pH 4.6; it is important to note there are significant errors for Xsa43E H258A at this pH that have been taken into account.

In the lower pH range the pK_a modulator is unnecessary for activity because the pH is similar to the unaffected pK_a, which for an isolated glutamic acid side chain is 4.1[146, 147], and for glutamic acid in a protein is around 4.5[148]. The pH profiles of Xsa43E H258A and Xsa43E D141A support the hypothesis that His-258 is better suited than Asp-141 to being the pK_a modulator of the catalytic acid. Xsa43E H258A maintains the equivalent activity of the native enzyme in the low pH range, whereas Xsa43E D141A retains very little activity across the entire pH range indicating it has an important role in positioning the substrate and the catalytic acid.

Xsa43E H258Q had a pH profile similar to that of the wild-type enzyme however the activity is about 20 fold less; at pH 5.2 and 4.6 the wild-type protein has very low activity, the loss of activity of the H258Q mutant at these pHs could be because of the level of activity being below observable levels.

The activity results of Xsa43E H258A and Xsa43E H258Q illustrate the importance of the electron withdrawing pathway between the calcium and the Glu-202 for activity. It also supports the role of His-258 in modulating the pK_a of the catalytic acid as opposed to Asp-141. A publication released in November, 2010 identifies the role of the central calcium ion and the corresponding histidine in another GH43 family enzyme as being important for activity of the enzyme, they also postulate that the equivalent histidine is the true pK_a modulator in GH43 family enzymes[149].

5.3 Conclusion

Xsa43E, the protein encoded by ORF Bpr_I2319 in *B. proteoclasticus,* was annotated as being either an arabinofuranosidase or a xylosidase belonging to the GH43 family of enzymes. I have characterised this enzyme both structurally and functionally, revealing it to be an α-L-arabinofuranosidase capable of cleaving arabinose units

from the xylan backbone of arabinoxylans. It has a 5 bladed ß-propeller structure that is common to all GH43 enzymes. Structural comparisons and site directed mutagenesis have revealed the importance of four residues (Asp-24, Asp-141, Glu-202 and His-258) in the catalytic mechanism. Along with the three previously identified residues, His-258 is necessary for the catalytic function of this enzyme, and potentially the family of enzymes due to its involvement in pK_a modulation of the catalytic acid.

The ion at the centre of the structure was identified as calcium and plays an important role in the function of the enzyme, as well as being integral to the structural arrangement of the active site.

Previous reports of GH43 enzyme structures have nominated the analogous Asp-141 as being the pK_a modulator of the catalytic acid[137]. The data presented here illustrate that in this enzyme, His-258 is a better candidate for that role, recent literature also supports this for another GH43 family enzyme[149]. This could also be applicable for all GH43 enzymes; using methods developed for Xsa43E this could be investigated for the other members of this family.

Chapter Six: Concluding Discussion

6.1 Background

The rumen is a complex fermentation vat that plays host to a variety of microorganisms. These microbes are responsible for the breakdown of plant material to volatile fatty acids and other compounds that can then be utilised by the host. *Butyrivibrio proteoclasticus* plays an important role within the rumen and a large proportion of its genome is dedicated to genes annotated as being involved in fibre degradation and reassembly[1]. This consists of a suite of genes encoding enzymes with a variety of functions necessary for the degradation of plant material.

Prior to this project 44 of the genes from *B. proteoclasticus* had been cloned into expression vectors. The 3D structures for two of these enzymes, Est1E, a feruloyl esterase, and Est2A an acetyl xylan esterase, had been determined.

The research involved in this project investigated eleven enzymes produced by *B. proteoclasticus;* two of which, Est1E and Est2A, had 3D structures and required further characterisation to compliment these. Crystallisation trials gave conditions for crystallisation of Xsa43E. This allowed the structural elucidation of Xsa43E; this was complimented with functional characterisation.

6.2 Characterisation of Est1E and Est2A

I investigated the kinetic parameters and proposed catalytic mechanisms for the two enzymes (Est1E and Est2A) that had previously elucidated 3D structures.

The kinetic parameters for Est1E with two model substrates *p*-nitrophenyl acetate and *p*-nitrophenyl butyrate were established. Comparisons of these kinetic data with other feruloyl esterases from the literature revealed Est1E is a promiscuous and efficient enzyme for cleaving ester linkages in a range of substrates. Est1E was also shown to be active against larger *p*-nitrophenyl substrates. Investigation of the catalytic

mechanism through site directed mutagenesis supports the hypothesis that Est1E utilises a catalytic triad consisting of Ser-105, Asp-197 and His-225. The structure of Est1E S105A showed very little difference when compared with the structure of the native enzyme.

The kinetic parameters for Est2A on the same two model substrates, *p*-nitrophenyl acetate, and *p*-nitrophenyl butyrate, were established. The data from these experiments supported the annotation of Est2A as an acetyl xylan esterase. This was illustrated by substantially improved kinetic parameters for the model substrate *p*-nitrophenyl acetate compared with *p*-nitrophenyl butyrate. Site directed mutagenesis was used to investigate the proposed catalytic dyad, consisting of Ser-142 and His-351. The results from this investigation supported the hypothesis that these residues made up the catalytic dyad. I have solved the crystal structure of Est2A H351A and shown there is practically no difference between the structures of the native enzyme and this mutant. Attempts to cocrystallise Est2A H351A mutant were unsuccessful meaning elucidation of the binding mechanism was not possible.

6.3 Preliminary Investigation of *Butyrivibrio proteoclasticus* Enzymes

Initial investigation into nine of the enzymes produced by *B. proteoclasticus* demonstrated that they were, in general, easy to express and purify with the exception of Mxy10-43A, a large multi domain protein. The gene for Xyn10A contained a 16 amino acid signal sequence, this was removed resulting in an increase in the levels of soluble protein expressed. Seven of the nine enzymes were screened to establish what sugars they release from natural substrates. The results indicated that the annotations of these enzymes were often inadequate for identifying their catalytic capabilities (Table 6-1)

Table 6-1: Preliminary investigation into seven FDE wih natural substrates showing the sugars the released and their annotations.

Enzyme	Annotation	Sugar released from Substrates
Xsa43A	xylosidase/arabinofuranosidase	Glucose
Xsa43B	xylosidase/arabinofuranosidase	none
Xsa43C	xylosidase/arabinofuranosidase	Arabinose/xylose, Glucose
Xsa43E	xylosidase/arabinofuranosidase	Arabinose/xylose
Lic16B	endo-1,3(4)-beta-glucanase	Arabinose/xylose, Glucose
Cel5C	endo-1,4-beta-glucanase	Glucose
Xyn10D	endo-1,4-ß-xylanase	Glucose

I have illustrated that annotations should be used as a guide to possible enzyme function but enzymes need to be characterised experimentally to truly elucidate their full enzymatic capacity.

6.4 Investigation of Xsa43E

Preliminary crystallisation experiments resulted in crystallisation conditions for the arabinofuranosidase Xsa43E. Xsa43E belongs to the GH43 family from the CAZy database. Structural investigation of this enzyme revealed that it contains one domain; a 5 bladed ß-propeller. Investigation of the structure revealed a metal ion at the centre of the structure, which I have determined is calcium. The position of the His-258 residue indicated that it is involved in the catalytic mechanism. By making four mutants; mutating the three previously identified catalytic acids and His-258 to alanine, I have shown that His-258 is important for catalysis. Functional investigation into the previously identified pK_a modulator Asp-141 and His-258 indicated that His-258 is more likely to be the pK_a modulator in Xsa43E. The histidine is necessary for the electron-withdrawing pathway from the calcium ion to the catalytic acid Glu-202; this causes a shift in the pK_a of Glu-202. Replacement of the histidine with a glutamine residue allows the pathway to be maintained and results in this mutant enzyme being active, however it is significantly less efficient than the wildtype. Kinetic parameters for Xsa43E with model substrates

p-nitrophenyl-α-L-arabinofuranose and *p*-nitrophenyl-ß-D-xylopyranose have also been established. The substrate specificity of Xsa43E was investigated revealing that it is an arabonofuranosidase able to cleave arabinose units from arabinoxylan, arabinobiose and degrade model substrates *p*-nitrophenyl-α-L-arabinofuranose and *p*-nitrophenyl-ß-D-xylopyranose but Xsa43E shows no activity of linear chains of ß-D-xylose or arabinohexaose.

6.5 Application potential

By characterising the function and exploring the potential of the enzymes produced by *B. proteoclasticus* it will allow us the potential to use them for alternative applications.

Characterisation of these enzymes will aid in creating an enzyme cocktail that can be used to completely degrade plant material into simple molecules (Figure 6-1). These molecules have a range of applications as starting materials for many industrial processes including renewable alternatives to fossil fuels.

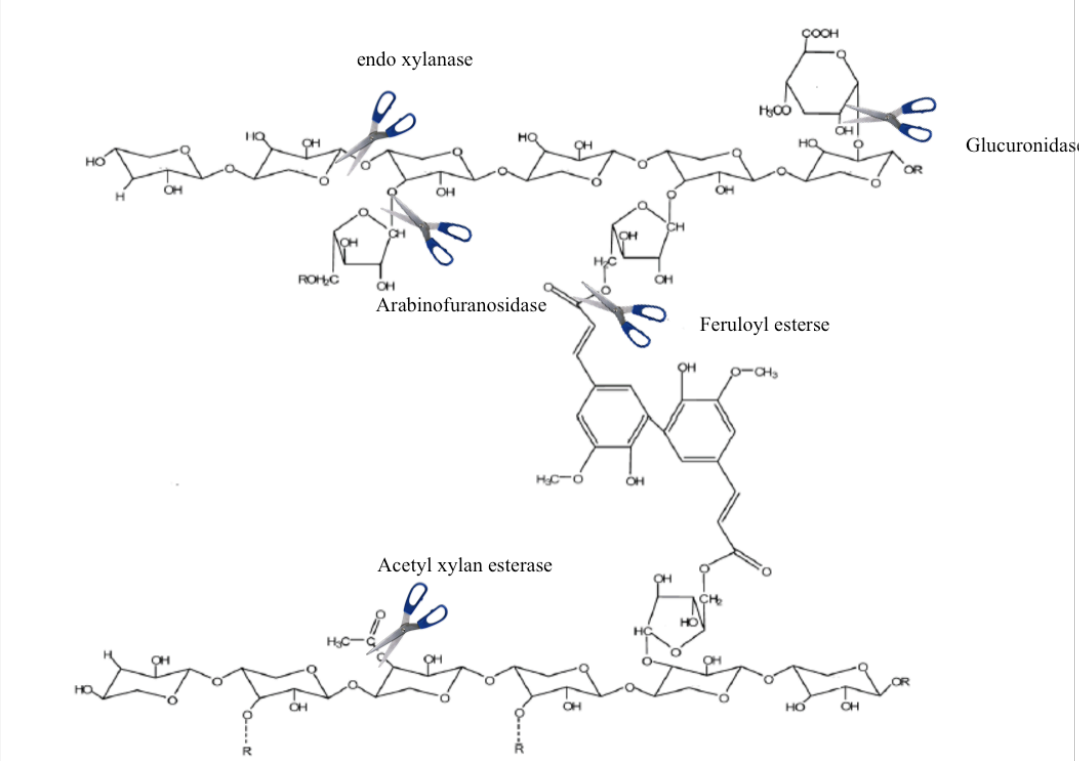


Figure 6-1: Simplified structure of xylan with enzymes indicating cleavage sites.

Research into functional characterisation of enzymes from the *B. proteoclasticus* genome has being conducted in conjunction with this project. This research focussed on the functional characterisation of potential candidates for a cocktail of enzymes that can completely degrade plant material. Est1E, Xsa43A, Xsa43D (an arabinofuranoside/xylosidase), Xsa43E and Cel5C have complimentary catalytic functions, combinations of these should lead to a larger proportion of the complex substrate being degraded. Preliminary assays of cocktails including combinations of these enzymes with various natural substrates have given varying results. The methods for these assays are still being developed.

6.6 Future Directions

The 44 genes from *B. proteoclasticus* that are cloned into expression vectors allow a base for the future investigation of these enzymes. In the future investigation into

enzymes deemed significant in the work by Kelly *et al* (2010)[1] will be a focus. There are nine large cell associated enzymes that were reported as making up the core of *B. proteoclasticus'* catalytic capacity to break down polysaccharides extracellularly. The structures and functional analysis of these enzymes will provide a significant challenge but will allow a better understanding of the dynamics of fibre degradation by *B. proteoclasticus*.

Establishing the catalytic capabilities of these enzymes will lead to the ability to create a cocktail of enzymes for the complete degradation of polysaccharides. This requires a significant amount of testing, using different combinations of enzymes. Another important avenue for future research is investigating the synergy between enzymes with different roles, preliminary work has been started on this with promising results[150]. Utilising this potential synergy and known function will allow an enzyme cocktail for the complete and efficient degradation of plant material.

Appendix I

TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0
LB-agar	1% bactotryptone, 0.5% yeast extract, 1% NaCl, 15 $g.L^{-1}$ agar pH 8.0
10 x TAE	400 mM Tris-acetate, 20 mM EDTA
1 x TAE	100 ml 10 x TAE + 900 ml H_2O
6 x DNA Loading Dye	0.05% bromophenol blue, 0.25% xylene cyanol, 30% glycerol
4 x SDS loading buffer	200 mM Tris-HCl pH 6.8, 8% SDS, 40% (v/v) glycerol, 0.4% bromophenol blue, 400 mM β-mercaptoethanol

SDS-PAGE Gel Recipes

Volume in mL	12 % Gel	10% Gel	Stacker
Milli Q Water	10.05	12.05	8.5
Resolving Buffer	7.5	7.5	-
Stacking Buffer	-	-	1.6
Acrylamide (30%)	12	10	2.125
10 % SDS	0.3	0.3	0.125
10% APS	0.15	0.15	0.063
Temed	0.015	0.015	0.0063

APS – Ammonium persulfate, Temed - N, N, N', N'-tetramethylethylenediamine

SDS running buffer	25m M Tris-HCl pH6.8, 0.1% SDS 190 mM glycine

Resolving Buffer	1.5 M Tris-HCl pH 8.8
Stacking Buffer	0.5 M Tris-HCl pH 6.8

Fairbanks Staining Protocol:

Cover gel in Fairbanks Staining Solution A, heat for 30 sec in microwave, allow to cool while gently shaking. Repeat with Fairbanks Staining Solution B. Repeat with Fairbanks Staining Solution C. Repeat with Fairbanks Staining Solution D.

Fairbanks Staining Solution A	0.05% coomassie blue R-250, 25% isopropanol, 10% acetic acid
Fairbanks Staining Solution B	0.005% coomassie blue R-250 10% isopropanol, 10% acetic acid
Fairbanks Staining Solution C	0.002% coomassie blue, 10% acetic acid
Fairbanks Staining Solution D	10% acetic acid
LB	1% bactotryptone, 0.5% yeast extract, 1% NaCl pH 8.0
PA-0.5G	50 mM Na_2HPO_4, 50 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% glucose, 0.1 x metals mix, 100 $\mu g.ml^{-1}$ each of 17 amino acids (no Cys, Tyr or Met). Individual components autoclaved to sterile filtered before adding to

	sterile H_2O.
PASM-5052	50 mM Na_2HPO_4, 25 mM KH_2PO_4, 25 mM $(NH_4)_2SO_4$, 1 mM $MgSO_4$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose, 1 x metals mix, 100nM vitamin B-12, 100 $\mu g.ml^{-1}$ each of 17 amino acids (no Cys, Tyr or Met), 10 $\mu g.ml^{-1}$ Methionine 125 $\mu g.ml^{-1}$ Seleno-methionine.
SOC	2% bactotryptone or bactopeptone, 0.55% yeast extract, 10 mM NaCl, 2.5 mM KCl 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.
1000 x metals mix	made from the sterile stock solutions of each component to give the following concentrations: 50 µM $FeCl_3$ in 0.12 M HCl (filter sterile), 20 µM $CaCl_2$, 10 µM $MnCl_2$, 10 µM $ZnSO_4$, 2 µM $CoCl_2$, 2 µM $CuCl_2$, 2 µM $NiCl_2$, 2 µM Na_2MoO_4, 2 µM Na_2SeO_3, 2 µM H_3BO_3

Appendix II

Retention Times of Sugars in GCMS

Sample	Retention time (min)
trans-Cinnamic acid TMS	15.46
Xylose TMS-oxime 1	15.8
Xylose TMS-Oxime 2	15.905
Arabinose TMS-Oxime1+2	15.92
Ribose TMS-oxime 1	16.078
Ribose TMS-oxime 2	16.115
Vanillic acid TMS	18.393
Sorbose TMS	18.59
Sorbose TMS-oxime 1	18.825
Sorbose TMS-oxime 2	18.913
Glucose TMS	19.2
Glucose TMS-oxime 1	19.345
Glucose TMS-oxime 2	19.825
Glucuronic acid TMS-oxime 1	20.6
Galacturonic acid TMS-oxime 1	20.753
Galacturonic acid TMS-oxime 1	20.753
Glucuronic acid TMS-oxime 2	20.96
Glucuronic acid TMS-oxime 2	20.96
Galacturonic acid TMS-oxime 2	21.163
Galacturonic acid TMS-oxime 2	21.163
glucose oxime 1	21.28
glucose oxime 2	21.28
p-Coumaric acid TMS	21.945
p-Coumaric acid TMS	21.945
Ferulic Acid TMS	24.445
Ferulic Acid TMS	24.445
Xylobiose TMS-oxime 1	27.87
Xylobiose TMS-oxime 1	27.87
Xylobiose TMS-oxime 2	28.358
Xylobiose TMS-oxime 2	28.358
Cellobiose TMS-oxime 1	33.17
Cellobiose TMS-oxime 1	33.17
Cellobiose TMS-oxime 2	33.918
Xylotriose TMS-oxime 1	44.125
Xylotriose TMS-oxime 2	44.805

References

1 Kelly WJ, Leahy SC, Altermann E, Yeoman CJ, Dunne JC, Kong ZH, Pacheco DM*, et al.* The Glycobiome of the Rumen Bacterium Butyrivibrio proteoclasticus B316(T) Highlights Adaptation to a Polysaccharide-Rich Environment. Plos One 2010, 5

2 Raven PH, Johnson GB, Lossos JB, Singer SR. Biology. 7th ed. 2005: McGraw-Hill

3 Carpita NC, McCann MC. The functions of cell wall polysaccharides in composition and architecture revealed through mutations. Plant and Soil 2002, 247: 71-80

4 Fernandez LEM, Obel N, Scheller HV, Roepstorff P. Characterization of plant oligosaccharides by matrix-assisted laser desorption/ionization and electrospray mass spectrometry. J Mass Spectrom 2003, 38: 427-437

5 Carpita N, McCann M, Griffing LR. The plant extracellular matrix: News from the cell's frontier. Plant Cell 1996, 8: 1451-1463

6 McCann MC, Carpita NC. Designing the deconstruction of plant cell walls. Curr Opin Plant Biol 2008, 11: 314-320

7 Koukiekolo R, Cho H-Y, Kosugi A, Inui M, Yukawa H, Doi RH. Degradation of Corn Fibre by *Clostridium cellulovorans* cellulases and Hemicellulases and contribution of Scaffolding protein CbpA. Applied and Environmental Microbiology 2005, 71: 3504-3511

8 de Vries RP, van Grieken C, vanKuyk PA, Wosten HAB. The value of genome sequences in the rapid identification of novel genes encoding specific plant cell wall degrading enzymes. Current Genomics 2005, 6: 157-187

9 Desvaux M. The cellulosome of *Clostridium cellulolyticum*. Enzyme and Microbial Technology 2005, 37: 373-385

10 Carpita NC. Structure and biogenesis of the cell walls of grasses. Annual Review of Plant Physiology and Plant Molecular Biology 1996, 47: 445-476

11 Taherzadeh MJ, Karimi K. ENZYME-BASED HYDROLYSIS PROCESSES FOR ETHANOL FROM LIGNOCELLULOSIC MATERIALS: A REVIEW. Bioresources 2007, 2: 707-738

12 Onda A, Ochi T, Yanagisawa K. Selective hydrolysis of cellulose into glucose

over solid acid catalysts. Green Chemistry 2008, 10: 1033-1037

13 Coughlan MP, Hazelwood GP. Hemicellulose and Hemicellulases. London: Portland Press Ltd 1993

14 Beg QK, Kapoor M, Mahajan L, Hoondal GS. Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol 2001, 56: 326-338

15 Fan ZM, Wagschal K, Lee CC, Kong Q, Shen KA, Maiti IB, Yuan L. The Construction and Characterization of Two Xylan-Degrading Chimeric Enzymes. Biotechnology and Bioengineering 2009, 102: 684-692

16 Attwood G, Altermann EH, Leahy SC, Moon CD, Cookson A, Yeomen C, Li D*, et al.* Hemicellulose degradation in the rumen anaerobe, *Butyrivibrio proteoclasticus.* 2008,

17 Mathew S, Abraham TE. Ferulic Acid: An Antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. Critical Reviews in Biotechnology 2004, 24: 59-83

18 Chapple C, Carpita N. Plant cell walls as targets for biotechnology. Curr Opin Plant Biol 1998, 1: 179-185

19 Hungate RE. The Rumen and Its Microbes. New York: Academic Press 1966

20 Reilly K, Attwood GT. Detection of Clostridium proteoclasticum and closely related strains in the rumen by competitive PCR. Applied and Environmental Microbiology 1998, 64: 907-913

21 Owensby CE, Cochran RM, Auen LM. Effects of Elevated Carbon Dioxide on Forage Quality of Ruminants. In: Koerner C, Bazzaz F eds., *Carbon Dioxide, Populations, and Communities*: Academic Press 1996: 363-371

22 Hobson PN. The Microflora of the Rumen. Durham, England: Meadowfield Press Ltd 1976

23 Krause DO, Denman SE, Mackie RI, Morrison M, Rae AL, Attwood GT, McSweeney CS. Opportunities to improve fiber degradation in the rumen: microbiology, ecology, and genomics. Fems Microbiol Rev 2003, 27: 663-693

24 Morrison M, Pope PB, Denman SE, McSweeney CS. Plant biomass degradation by gut microbiomes: more of the same or something new? Current

Opinion in Biotechnology 2009, 20: 358-363

25 Morvan B, RieuLesme F, Fonty G, Gouet P. In vitro interactions between rumen H-2-producing cellulolytic microorganisms and H-2-utilizing acetogenic and sulfate-reducing bacteria. Anaerobe 1996, 2: 175-180

26 Lee SS, Shin KJ, Kim WY, Ha JK, Han IK. The rumen ecosystem: As a fountain source of nobel enzymes - Review. Asian Australas J Anim Sci 1999, 12: 988-1001

27 Flint HJ. The rumen microbial ecosystem - some recent developments. Trends Microbiol 1997, 5: 483-488

28 Wallace RJ. Rumen Microbiology, Biotechnology and Ruminant Nutrition - the Application of Research Findings to a Complex Microbial Ecosystem. FEMS Microbiology Letters 1992, 100: 529-534

29 Hobson PN. The Rumen Microbial Ecosystem. London: Elsevier Applied Science 1989

30 Kamra P, Gokhale RS, Mohanty D. SEARCHGTr: a program for analysis of glycosyltransferases involved in glycosylation of secondary metabolites. Nucleic Acids Research 2005, 33: W220-W225

31 Montgomery L, Flesher B, Stahl D. TRANSFER OF BACTEROIDES-SUCCINOGENES (HUNGATE) TO FIBROBACTER GEN-NOV AS FIBROBACTER-SUCCINOGENES COMB NOV AND DESCRIPTION OF FIBROBACTER INTESTINALIS SP-NOV. Int J Syst Bacteriol 1988, 38: 430-435

32 Attwood GT, Reilly K. Identification of Proteolytic Rumen Bacteria Isolated from New-Zealand Cattle. Journal of Applied Bacteriology 1995, 79: 22-29

33 Moon CD, Pacheco DM, Kelly WJ, Leahy SC, Li D, Kopecny J, Attwood GT. Reclassification of Clostridium proteoclasticum as Butyrivibrio proteoclasticus comb. nov., a butyrateproducing ruminal bacterium. Int J Syst Evol Microbiol 2008, 58: 2041-2045

34 Attwood GT, Reilly K, Patel BKC. Clostridium proteoclasticum sp nov, a novel proteolytic bacterium from the bovine rumen. Int J Syst Bacteriol 1996, 46: 753-758

35 The UniProt C. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Research, 38: D142-148

36 Bryant MP, Small N. The Anaerobic Monotrichous Butyric Acid-Producing Curved Rod Shaped Bacteria of the Rumen. Journal of Bacteriology 1956, 72: 16-21

37 Bryant MP, Burkey LA. Cultural methods and some characteristics of some of the more numerous groups of bacteria inthebovinerumen. Journal of Dairy Science 1953, 36: 205- 217

38 Moore WEC, Johnson JL, Holdeman LV. Emendation of Bacteroidaceae and Butyrivibrio and Descriptions of Desulfomonas gen. nov. and Ten New Species in the Genera Desulfomonas, Butyrivibrio, Eubacterium, Clostridium, and Ruminococcus. Int J Syst Bacteriol 1976, 26: 238-252

39 Cheng KJ, Costerton JW. Ultrastructure of Butyrivibrio fibrisolvens: a gram-positive bacterium. Journal of Bacteriology 1977, 129: 1506-1512

40 Willems A, AmatMarco M, Collins MD. Phylogenetic analysis of Butyrivibrio strains reveals three distinct groups of species within the Clostridium subphylum of the gram-positive bacteria. Int J Syst Bacteriol 1996, 46: 195-199

41 Forster RJ, Teather RM, Gong J. 16S rDNA analysis of Butyrivibrio fibrisolvens: Phylogenetic position and relation to butyrate-producing anaerobic bacteria from the rumen of white-tailed deer. Letters in Applied Microbiology 1996, 23: 218-222

42 vanGylswyk NO, Hippe H, Rainey FA. Pseudobutyrivibrio ruminis gen nov, sp nov, a butyrate-producing bacterium from the rumen that closely resembles Butyrivibrio fibrisolvens in phenotype. Int J Syst Bacteriol 1996, 46: 559-563

43 Kopecny J, Zorec M, Mrazek J, Kobayashi Y, Marinsek-Logar R. Butyrivibrio hungatei sp nov and Pseudobutyrivibrio xylanivorans sp nov., butyrate-producing bacteria from the rumen. Int J Syst Evol Microbiol 2003, 53: 201-209

44 Cotta MA. Amylolytic Activity of Selected Species of Ruminal Bacteria. Applied and Environmental Microbiology 1988, 54: 772-776

45 Polan CE, Tove SB, McNeill JJ. Biohydrogenation of unsaturated fatty acids by rumen bacteria. Journal of Bacteriology 1964, 88: 1056–1064

46 Boeckaert C, Vlaeminck B, Fievez V, Maignien L, Dijkstra J, Boon N. Accumulation of trans C-18:1 Fatty Acids in the Rumen after Dietary Algal Supplementation Is Associated with Changes in the Butyrivibrio Community. Applied and Environmental Microbiology 2008, 74: 6923-6930

47 Kalmokoff ML, Teather RM. Isolation and characterization of a bacteriocin (butyrivibriocin AR10) from the ruminal anaerobe Butyrivibrio fibrisolvens

AR10: Evidence in support of the widespread occurrence of bacteriocin-like activity among ruminal isolates of B-fibrisolvens. Applied and Environmental Microbiology 1997, 63: 394-402

48 Kalmokoff ML, Cyr TD, Hefford MA, Whitford MF, Teather RM. Butyrivibriocin AR10, a new cyclic bacteriocin produced by the ruminal anaerobe Butyrivibrio fibrisolvens AR10: characterization of the gene and peptide. Canadian Journal of Microbiology 2003, 49: 763-773

49 Rychlik JL, Russell JB. Bacteriocin-like activity of Butyrivibrio fibrisolvens JL5 and its effect on other ruminal bacteria and ammonia production. Applied and Environmental Microbiology 2002, 68: 1040-1046

50 Wallace RJ, Chaudhary LC, McKain N, McEwan NR, Richardson AJ, Vercoe PE, Walker ND*, et al. Clostridium proteoclasticum*: a ruminal bacterium that forms stearic acid from linoleic acid. FEMS Microbiology Letters 2006, 265: 195-201

51 Villas-Boas SG, Moon CD, Noel S, Hussein H, Kelly WJ, Cao M, Lane GA*, et al.* Phenotypic characterization of transposon-inserted mutants of Clostridium proteoclasticum B316(T) using extracellular metabolomics. Journal of Biotechnology 2008, 134: 55-63

52 Demain AL, Newcomb M, Wu JHD. Cellulase, Clostridia, and Ethanol. Microbiology and molecular biology reviews 2005, 69: 124-154

53 Creighton TE. Proteins Structures and Molecular Properties. 2nd ed. New York: W. H. Freeman and Company 1993

54 Fersht A. Structure and Mechanism in Protein Science. New York: W. H. Freeman and Company 1999

55 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research 2009, 37: D233-D238

56 Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: Structures, functions, and mechanisms. Annual Review of Biochemistry 2008, 77: 521-555

57 Sado PE, Tessier D, Vasseur M, Elmorjani K, Guillon F, Saulnier L. Integrating genes and phenotype: a wheat-Arabidopsis-rice glycosyltransferase database for candidate gene analyses. Funct Integr Genomics 2009, 9: 43-58

58 Christiansen C, Abou Hachem M, Janecek S, Vikso-Nielsen A, Blennow A, Svensson B. The carbohydrate-binding module family 20-diversity, structure,

and function. Febs Journal 2009, 276: 5006-5029

59 Shoseyov O, Shani Z, Levy I. Carbohydrate binding modules: Biochemical properties and novel applications. Microbiology and molecular biology reviews 2006, 70: 283-+

60 Madkour M, Mayer F. Structural organization of the intact bacterial cellulosome as revealed by electron microscopy. Cell Biology International 2003, 27: 831-836

61 Jorgensen H, Kutter JP, Olsson L. Separation and quantification of cellulases and hemicellulases by capillary electrophoresis. Analytical Biochemistry 2003, 317: 85-93

62 Wantanabe H, Noda H, Tokuda G, Lo N. A cellulase gene of termite origin. Nature 1998, 394: 330-331

63 Sakamoto K, Toyohara H. Putative endogenous xylanase from brackish-water clam Corbicula japonica. Comp Biochem Physiol B-Biochem Mol Biol 2009, 154: 85-92

64 Byrne KA, Lehnert SA, Johnson SE, Moore SS. Isolation of a cDNA encoding a putative cellulase in the red claw crayfish Cherax quadricarinatus. Gene 1999, 239: 317-324

65 Wang J, Ding M, Li YH, Chen QX, Xu GJ, Zhao FK. Isolation of a multi-functional endogenous cellulase gene from mollusc, Ampullaria crossean. Acta Biochimica Et Biophysica Sinica 2003, 35: 941-946

66 Imjongjirak C, Amparyup P, Sittipraneed S. Cloning, genomic organization and expression of two glycosyl hydrolase family 10 (GHF10) genes from golden apple snail (Pomacea canaliculata). DNA Sequence 2008, 19: 224-236

67 Mitreva-Dautova M, Roze E, Overmars H, de Graaff L, Schots A, Helder J, Goverse A*, et al.* A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode Meloidogyne incognita. Molecular Plant-Microbe Interactions 2006, 19: 521-529

68 Bhat MK, Bhat S. Cellulose degrading enzymes and their potential industrial applications. Biotechnology Advances 1997, 15: 583-620

69 Uffen RL. Xylan degradation: a glimpse at microbial diversity. Journal of Industrial Microbiology & Biotechnology 1997, 19: 1-6

70 Zhang YHP, Himmel ME, Mielenz JR. Outlook for cellulase improvement: Screening and selection strategies. Biotechnology Advances 2006, 24: 452-481

71 Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science 2007, 315: 804-807

72 Meyer AS, Rosgaard L, Sorensen HR. The minimal enzyme cocktail concept for biomass processing. Journal of Cereal Science 2009, 50: 337-344

73 Ding SY, Xu Q, Crowley M, Zeng Y, Nimlos M, Lamed R, Bayer EA*, et al.* A biophysical perspective on the cellulosome: new opportunities for biomass conversion. Current Opinion in Biotechnology 2008, 19: 218-227

74 Eun JS, Beauchemin KA. Relationship between enzymic activities and in vitro degradation of alfalfa hay and corn silage. Anim Feed Sci Technol 2008, 145: 53-67

75 Branden C, Tooze J. Introduction to Protein Structure. 2nd ed. New York: Garland Publishing Ltd 1999

76 Blow D. Outline of Crystallography for Biologists. New York: Oxford University Press 2002

77 Read RJ, Leslie A, McCoy AJ, Evans G, Bloomer A, Evans P, Lesk A. Protein Crystallography Course, http://www-structmed.cimr.cam.ac.uk/links.html. ed. 2005

78 Read RJ. As MAD as can be. Structure 1996, 4: 11-14

79 McCoy AJ, Read RJ. Experimental phasing: best practice and pitfalls. Acta Crystallographica Section D-Biological Crystallography 2010, 66: 458-469

80 Uson I, Sheldrick GM. Advances in direct methods for protein crystallography. Current Opinion in Structural Biology 1999, 9: 643-648

81 McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. Acta Crystallographica Section D-Biological Crystallography 2005, 61: 458-464

82 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of Applied Crystallography 2007, 40: 658-674

83 Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. Acta Crystallographica Section D-Biological Crystallography 2004, 60: 432-438

84 Leslie AGW. The integration of macromolecular diffraction data. Acta Crystallographica Section D-Biological Crystallography 2006, 62: 48-57

85 Bailey S. THE CCP4 SUITE - PROGRAMS FOR PROTEIN CRYSTALLOGRAPHY. Acta Crystallographica Section D-Biological Crystallography 1994, 50: 760-763

86 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography 2010, 66: 213-221

87 Evans P. Scaling and assessment of data quality. Acta Crystallographica Section D-Biological Crystallography 2006, 62: 72-82

88 Terwilliger TC, Adams PD, Read RJ, McCoy AJ, Moriarty NW, Grosse-Kunstleve RW, Afonine PV, *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallographica Section D-Biological Crystallography 2009, 65: 582-601

89 Vaguine AA, Richelle J, Wodak SJ. SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. Acta Crystallographica Section D-Biological Crystallography 1999, 55: 191-205

90 Laskowski RA, Macarthur MW, Moss DS, Thornton JM. PROCHECK - A PROGRAM TO CHECK THE STEREOCHEMICAL QUALITY OF PROTEIN STRUCTURES. Journal of Applied Crystallography 1993, 26: 283-291

91 Kleywegt GJ, Jones TA. Phi/psi-chology: Ramachandran revisited. Structure 1996, 4: 1395-1400

92 Lovell SC, Davis IW, Adrendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, *et al.* Structure validation by C alpha geometry: phi,psi and C beta deviation. Proteins 2003, 50: 437-450

93 Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography 2004, 60: 2126-2132

94 DeLano WL. The PyMOL Molecular Graphics System. ed. Palo Alto, CA,

USA: DeLano Scientific 2002

95 Moreland N, Ashton R, Baker HM, Ivanovic I, Patterson S, Arcus VL, Baker EN*, et al.* A flexible and economical medium-throughput strategy for protein production and crystallization. Acta Crystallographica Section D-Biological Crystallography 2005, 61: 1378-1385

96 Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. Nucleic Acids Research 1994, 22: 4673-4680

97 Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM ed., *The Proteomics Protocols Handbook*: Humana Press 2005: 571-607

98 Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ*, et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallographica Section D-Biological Crystallography 2008, 64: 61-69

99 Terwilliger TC. SOLVE and RESOLVE: Automated structure solution and density modification. *Macromolecular Crystallography, Pt D* 2003: 22-37

100 Koseki T, Fushinobu S, Ardiansyah, Shirakawa H, Komai M. Occurrence, properties, and applications of feruloyl esterases. Appl Microbiol Biotechnol 2009, 84: 803-810

101 Yu P, McKinnon JJ, Christensen DA. Hydroxycinnamic acids and ferulic acid esterase in relation to biodegradation of complex plant cell walls. Canadian Journal of Animal Science 2005, 85: 255-267

102 Williamson G, Kroon PA, Faulds CB. Hairy plant polysaccharides: a close shave with microbial esterases. Microbiology-Sgm 1998, 144: 2011-2023

103 Kamisaka S, Takeda S, Takahashi K, Shibata K. DIFERULIC AND FERULIC ACID IN THE CELL-WALL OF AVENA COLEOPTILES - THEIR RELATIONSHIPS TO MECHANICAL-PROPERTIES OF THE CELL-WALL. Physiol Plant 1990, 78: 1-7

104 Crepin VF, Faulds CB, Connerton IF. Functional classification of the microbial feruloyl esterases. Appl Microbiol Biotechnol 2004, 63: 647-652

105 Koseki T, Takahashi K, Fushinobu S, Iefuji H, Iwano K, Hashizume K, Matsuzawa H. Mutational analysis of a feruloyl esterase from Aspergillus

awamori involved in substrate discrimination and pH dependence. Biochimica Et Biophysica Acta-General Subjects 2005, 1722: 200-208

106 Wong DWS. Feruloyl Esterase: A key enzyme in Biomass Degredation. Applied Biochemistry and Biotechnology 2005, 133: 87-112

107 Schubot FD, Kataeva IA, Blum DL, Chah AK, Ljungdahl LG, Rose JP, Wang B-C. Structural Basis and Substrate Specificity of the Feruloyl Esterase Domain of the Cellulosomal Xylanase Z from *Clostridium thermocellum*. Biochemistry 2001, 40: 12524-12532

108 McAuley KE, Svendsen A, Patkar SA, Wilson KS. Structure of a feruloyl esterase from Aspergillus niger. Acta Crystallographica Section D-Biological Crystallography 2004, 60: 878-887

109 Topakas E, Vafiadi C, Christakopoulos P. Microbial production, characterization and applications of feruloyl esterases. Process Biochemistry 2007, 42: 497-509

110 Goldstone DC, Villas-BÙas SG, Till M, Kelly WJ, Attwood GT, Arcus VL. Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium <I>Butyrivibrio proteoclasticus</I>. Proteins: Structure, Function, and Bioinformatics 2010, 78: 1457-1469

111 Hakulinen N, Tenkanen M, Rouvinen J. Three-dimensional structure of the catalytic core of acetylxylan esterase from Trichoderma reesei: Insights into the deacetylation mechanism. J Struct Biol 2000, 132: 180-190

112 Yang C-H, Liu W-H. Purification and properties of a acetylxylan esterase from *Thermobifida fusca*. Enzyme and Microbial Technology 2008, 42: 181-186

113 Cybinski DH, Layton I, Lowry JB, Dalrymple BP. An acetylxylan esterase and a xylanase expressed from genes cloned from the ruminal fungus Neocallimastix patriciarum act synergistically to degrade acetylated xylans. Appl Microbiol Biotechnol 1999, 52: 221-225

114 Kam DK, Jun HS, Ha JK, Inglis GD, Forsberg CW. Characteristics of adjacent family 6 acetylxylan esterases from Fibrobacter succinogenes and the interaction with the Xyn10E xylanase in hydrolysis of acetylated xylan. Canadian Journal of Microbiology 2005, 51: 821-832

115 Ghatora SK, Chadha BS, Saini HS, Bhat MK, Faulds CB. Diversity of plant cell wall esterases in thermophilic and thermotolerant fungi. Journal of Biotechnology 2006, 125: 434-445

116 Mukhopadhyay A, Hazra PP, Sengupta T, Ghosh AK, Sengupta S. Acetyl esterase production by *Termitomyces clypeatus*. Biotechnology Letters 1997, 19: 159-161

117 Puchart V, Gariepy M-C, Sharek F, Dupont C. Identification of catalytically important amino acid residues of *Streptomyces lividans* acetylxylan esterase A from carbohydrate esterase family 4. Biochemica et Biophysica Acta 2006, 1764: 263-274

118 Pesaresi A, Devescovi G, Lamba D, Venturi V, Degrassi G. Isolation, characterization, and heterologous expression of a carboxylesterase of Pseudomonas aeruginosa PAO1. Current Microbiology 2005, 50: 102-109

119 Martinez-Martinez I, Navarro-Fernandez J, Lozada-Ramirez JD, Garcia-Carmona F, Sanchez-Ferrer A. YesT: A new rhamnogalacturonan acetyl esterase from Bacillus subtilis. Proteins: Structure Function and Bioinformatics 2008, 71: 379-388

120 Navarro-Fernandez J, Martinez-Martinez I, Montoro-Garcia S, Garcia-Carmona F, Takami H, Sanchez-Ferrer A. Characterization of a new rhamnogalacturonan acetyl esterase from Bacillus halodurans C-125 with a new putative carbohydrate binding domain. Journal of Bacteriology 2008, 190: 1375-1382

121 Latha GM, Muralikrishna G. Purification and partial characterization of acetic acid esterase from malted finger millet (Eleusine coracana, Indaf-15). J Agric Food Chem 2007, 55: 895-902

122 Biely P, Mastihubova M, Puchart V. The vicinal hydroxyl group is prerequisite for metal activation of Clostridium thermocellum acetylxylan esterase. Biochimica Et Biophysica Acta-General Subjects 2007, 1770: 565-570

123 Goldstone D. Structure of Est2A. ed. 2007

124 Holm L, Park J. DaliLite workbench for protein structure comparison. Bioinformatics 2000, 16: 566-567

125 Latha GM, Srinivas P, Muralikrishnah G. Purification and characterization of ferulic acid esterase from malted finger millet (Eleusine coracana, indaf-15). J Agric Food Chem 2007, 55: 9704-9712

126 Fazary AE, Hamad HA, Lee JC, Koskei T, Lee CK, Ju YH. Expression of feruloyl esterase from Aspergillus awamori in Escherichia coli: Characterization and crystal studies of the recombinant enzyme. International Journal of Biological Macromolecules 2010, 46: 440-444

127 Donaghy JA, Bronnenmeier K, Soto-Kelly PF, McKay AM. Purification

and characterization of an extracellular feruloyl esterase from the thermophilic anaerobe Clostridium stercorarium. Journal of Applied Microbiology 2000, 88: 458-466

128 Hegde S, Muralikrishna G. Isolation and partial characterization of alkaline feruloyl esterases from Aspergillus niger CFR 1105 grown on wheat bran. World Journal of Microbiology & Biotechnology 2009, 25: 1963-1969

129 Rumbold K, Biely P, Mastihubova M, Gudelj M, Gubitz G, Robra KH, Prior BA. Purification and properties of a feruloyl esterase involved in lignocellulose degradation by Aureobasidium pullulans. Applied and Environmental Microbiology 2003, 69: 5622-5626

130 Degrassi G, Okeke BC, Bruschi CV, Venturi V. Purification and characterization of an acetyl xylan esterase from Bacillus pumilus. Applied and Environmental Microbiology 1998, 64: 789-792

131 Chungool W, Thongkam W, Raweesri P, Thamchaipenet A, Pinphanichakarn P. Production, purification, and characterization of acetyl esterase from Streptomyces sp PC22 and its action in cooperation with xylanolytic enzymes on xylan degradation. World Journal of Microbiology & Biotechnology 2008, 24: 549-556

132 Dupont C, Daigneault N, Shareck F, Morosoli R, Kluepfel D. Purification and characterization of an acetyl xylan esterase produced by Streptomyces lividans. Biochemical Journal 1996, 319: 881-886

133 Ding SJ, Cao J, Zhou R, Zheng F. Molecular cloning, and characterization of a modular acetyl xylan esterase from the edible straw mushroom Volvariella volvacea. FEMS Microbiology Letters 2007, 274: 304-310

134 Elend C, Schmeisser C, Leggewie C, Babiak P, Carballeira JD, Steele HL, Reymond JL*, et al.* Isolation and biochemical characterization of two novel metagenome-derived esterases. Applied and Environmental Microbiology 2006, 72: 3637-3645

135 Montanier C, Money VA, Pires VMR, Flint JE, Pinheiro BA, Goyal A, Prates JAM*, et al.* The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. Plos Biology 2009, 7: 687-697

136 Beisel HG, Kawabata S, Iwanaga S, Huber R, Bode W. Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab Tachypleus tridentatus. EMBO Journal 1999, 18: 2313-2322

137 Nurizzo D, Turkenburg JP, Charnock SJ, Roberts SM, Dodson EJ,

McKie VA, Taylor EJ*, et al. Cellvibrio japonicus* α-L arabinanase 43A has a novel five-blade β-propeller fold. Nature Structural Biology 2002, 9: 665-668

138 Dokmanic I, Sikic M, Tomic S. Metals in proteins: correlation between the metal-ion type, coordination number and the amino-acid residues involved in the coordination. Acta Crystallographica Section D-Biological Crystallography 2008, D64: 257-263

139 Alhassid A, Ben-David A, Tabachnikov O, Libster D, Naveh E, Zolotinitsky G, Shoham Y*, et al.* Crystal Structure of an inverting GH 43 1,5-α-L-arabinanase from *Geobacillus stearothermophillus* complexed with its substrate. Biochemical Journal 2009, 422: 73-82

140 Brux C, Ben-David A, Shallom-Shezifi D, Leon M, Niefind K, Shoham G, Shoham Y*, et al.* The structure of an inverting GH43 beta-xylosidase from Geobacillus stearothermophilus with its substrate reveals the role of the three catalytic residues. Journal of Molecular Biology 2006, 359: 97-109

141 Vandermarliere E, Bourgois TM, Winn MD, Campenhout SV, Volckaert G, Delcour JA, Strelkov SV*, et al.* Structural Analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. Biochemical Journal 2009, 418: 39-47

142 Pons T, Naumoff DG, Martinez-Fleited C, Hernandez L. Three Acidic Residues Are at the Active Site of a β-propeller Architecture in Glycoside Hydrolase Families 32, 43, 62 and 68. Proteins: Structure Function and Bioinformatics 2004, 54: 424-432

143 Shallom D, Leon M, Bravman T, Ben-David A, Zaide G, Belakhov V, Shoham G*, et al.* Biochemical characterization and identification of the catalytic residues of a family 43 beta-D-xylosidase from Geobacillus stearothermophilus T-6. Biochemistry 2005, 44: 387-397

144 Proctor MR, Taylor EJ, Nurizzo D, Turkenburg JP, Lloyd RM, Vardakou M, Davies GJ*, et al.* Tailored catalysis for plant cell-wall degradation: Redesigning the exo/endo preference of *Cellvibrio japonicus* arabinanase 43A. PNAS 2005, 102: 2697-2702

145 Inacio JM, de Sa-Nogueira I. Characterization of abn2 (yxiA), encoding a bacillus subtilis GH43 arabinanase, Abn2, an its role in arabino-polysaccharide degradation. Journal of Bacteriology 2008, 190: 4272-4280

146 Gunner MR, Mao JJ, Song YF, Kim J. Factors influencing the energetics of electron and proton transfers in proteins. What can be learned from calculations?

Biochimica Et Biophysica Acta-Bioenergetics 2006, 1757: 942-968

147 Oliveberg M, Arcus VL, Fersht AR. PK(A) VALUES OF CARBOXYL GROUPS IN THE NATIVE AND DENATURED STATES OF BARNASE - THE PK(A) VALUES OF THE DENATURED STATE ARE ON AVERAGE 0.4 UNITS LOWER THAN THOSE OF MODEL COMPOUNDS. Biochemistry 1995, 34: 9424-9433

148 Isom DG, Castaneda CA, Velu PD, Garcia-Moreno B. Charges in the hydrophobic interior of proteins. Proceedings of the National Academy of Sciences of the United States of America, 107: 16096-16100

149 de Sanctis D, Inacio JM, Lindley PF, de Sa-Nogueira I, Bento I. New evidence for the role of calcium in the glycosidase reaction of GH43 arabinanases. Febs Journal 2010, 277: 4562-4574

150 Haaning L. *B. proteoclasticus* enzyme analysis. 2010