Structural and functional studies of a Lytic Polysaccharide Monooxygenase

Rebecca Charlotte Gregory

Doctor of Philosophy

University of York

Chemistry

September 2016

Abstract

The enzymatic degradation of polysaccharides is a major goal of the biotechnology industry, most notably for both first (starch) and second (cellulose/chitin) generation biofuels. The classical model for enzymatic polysaccharide breakdown involves the action of endo- and exo-acting glycoside hydrolases which act in synergy; endo-acting enzymes generating free chain-ends for the processive exo-glycosidases. In 2010 a new enzyme class was discovered which overturned the classical hydrolase model. These enzymes, now known as Lytic Polysaccharide Monooxygenases (LPMOs), are understood to be copper-dependent oxygenase enzymes that cause chain cleavage within polysaccharides, facilitating their degradation by classical hydrolases. This research showcases the findings for an LPMO of the "Auxiliary Activity" class AA10 from the bacterium *Bacillus amyloliquefaciens* (*Ba*AA10). Methods of gene expression and protein production are described, including the incorporation of a SUMO-tag which resulted in increased yields of *Ba*AA10. The nature of the copper binding was assessed using Isothermal Titration Calorimetry and Differential Scanning Fluorimetry, which demonstrated the extremely tight binding of copper to the enzyme and its increased stability when copper is bound. Additionally, the detection of lactone and aldonic acid products of oxidative degradation, in MALDI-TOF Mass Spectrometry results, determined for the first time that *Ba*AA10 was active on both α and β forms of chitin. Finally, a Cu(II) structure of *Ba*AA10 was able to be obtained following the use of a spiral data collection technique as a method of preventing photoreduction of Cu(II) to Cu(I) in the X-ray beam. This structure can now provide us with details regarding the copper active site and, together with potential spectroscopic studies in the future, can help us to determine the mechanism of action of these LPMO enzymes, so that we can better exploit their use within the biofuel industry.

List of Contents

List of Figures

List of Tables

Acknowledgements

There are many people I would like to thank for their assistance over the course of my PhD. Firstly I would like to thank my supervisors, Gideon Davies and Paul Walton, for providing me with such a 'cutting edge' project, and for their patience and unlimited pastoral care over the past four years. I would also like to thank David Smith, Jennifer Potts and Alison Parkin for their help and suggestions during my Thesis Advisory Panel meetings.

The members of the Davies and Walton groups, both past and present, have been incredibly helpful, and I would particularly like to thank Luisa Ciano and Esther Johnston for their assistance with EPR experiments and analyses. In addition, special thanks must go to Glyn Hemsworth for guiding me through my PhD in every aspect, and being my rock when I needed help. I am truly grateful for his patience and the generosity of his time.

Various specific technical thanks go to Andrew Leech in the York Technology Facility for his help with SEC-MALLS, amongst other things; Adam Dowle for training and assistance on the MALDI-TOF machine; as well as many members of the YSBL who have helped me in the laboratory throughout my time here. I would also like to thank Sally Lewis for showing me the 'lab ropes' during my first few months and for her and Simon Grist's smooth operation of the YSBL wet-lab over the years. In addition, I would like to thank Tim Kirk for computer assistance and Catherine Jardine for advice and administrative care. Special thanks must go to Johan Turkenburg and Sam Hart for the many hours they spend collecting crystal data both remotely and at the synchrotron – I will never forget the snowy night spent collecting my crystal data at Diamond with Johan and Glyn!

On a personal note, I would like to acknowledge a few people who have supported me so much during my PhD, especially during the last year, and made me believe that I could do it. Sophie, Sherry, Imogen, Claire and Dan – thank you, your words and encouragement have been very beneficial and have not gone unnoticed. I would also like to thank my family (Mum, Dad, Rachael and Chasz) for their unconditional love and support always.

Finally, I wish to dedicate this thesis to a special grandad I still miss, and to the little niece I can't wait to meet soon.

Author's Declaration

All research carried out and presented in this thesis was completed by the author with the exception of collaborative work, which is stated within the main body of the text, or is detailed below. The author also declares that the writing of the thesis is all their own work. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Work carried out or overseen by others:

- The cloning of the *Ba*AA10 protein with a pelB leader sequence was carried out by Dr. Glyn Hemsworth, who performed the original crystallisation trials of the protein, leading to its structure being solved prior to the commencement of my PhD.

- SEC-MALLS experiments were performed and analysed by Dr. Andrew Leech in the York Technology Facility.

- Differential Scanning Calorimetry (DSC) was carried out at the Astbury Centre for Structural Molecular Biology in Leeds, under the supervision of Dr. Iain Manfield.

- Analysis and simulations of the EPR spectra for the *Ba*AA10 enzyme were carried out by Professor Paul Walton.

1. Introduction

1.1 The biotechnology era

Ever since the industrial revolution at the beginning of the 19th century, society has been progressing technologically at an ever increasing rate. For example, a journey by stagecoach from York to London in 1750 would have taken more than three days;[1] but now this journey can be made in less than five hours by car. Whilst this is beneficial to our way of life, increasing business, trade and tourism levels, an increase in population worldwide has subsequently led to an increased demand for energy. This increase in demand has resulted in non-renewable energy sources, such as oil, being in danger of running out for further generations, which is becoming a major issue within our current industrial climate. In addition, burning these non-renewable fossil fuels contributes to the release of carbon dioxide into the atmosphere, one of the main causes of global warming. Therefore, there is significant interest in developing renewable energy sources for the future. One area of research into alternative energy sources, which has gathered attention and subsequent momentum in recent years, has been the biofuel and biotechnology industry.[2]

Biofuels, in the context of this thesis ethanol produced from the fermentation of plant polysaccharides, can either be obtained by first generation or second generation sources[3] – first generation biofuels being produced directly from starch containing crops such as wheat, corn, and sugarcane (here the substrate is a disaccharide) grown specifically for this purpose, and second generation biofuels being produced from the waste accumulated from crops being grown for regular food production purposes and even specific energy crops.[4] In order to prevent any potential food shortages due to land and food sources being used up for first generation biofuel crop production,[5] second generation biofuel processes are becoming increasingly more favoured. Whether making first or second generation biofuels, the aim of this process is to release individual sugars contained within biomass, which can then be converted to bioethanol by fermentation or chemical means.

The processes currently used to convert biomass into bioethanol (Figure 1.1.1) and other useful chemicals, such as bio-based plastics and fertilisers, may consist of either physical degradation methods (e.g. steam explosion, acid hydrolysis), chemical methods (e.g. enzymatic degradation and conversion of products), or a combination of both.[6] Substantial developments in research into the biofuel industry from various different disciplines and industrial collaborations have led to an increase in awareness of and attention to bio-renewables within society. As this new

renewable age progresses, it could be suggested that we are entering into a new industrial revolution, one driven by biotechnology.

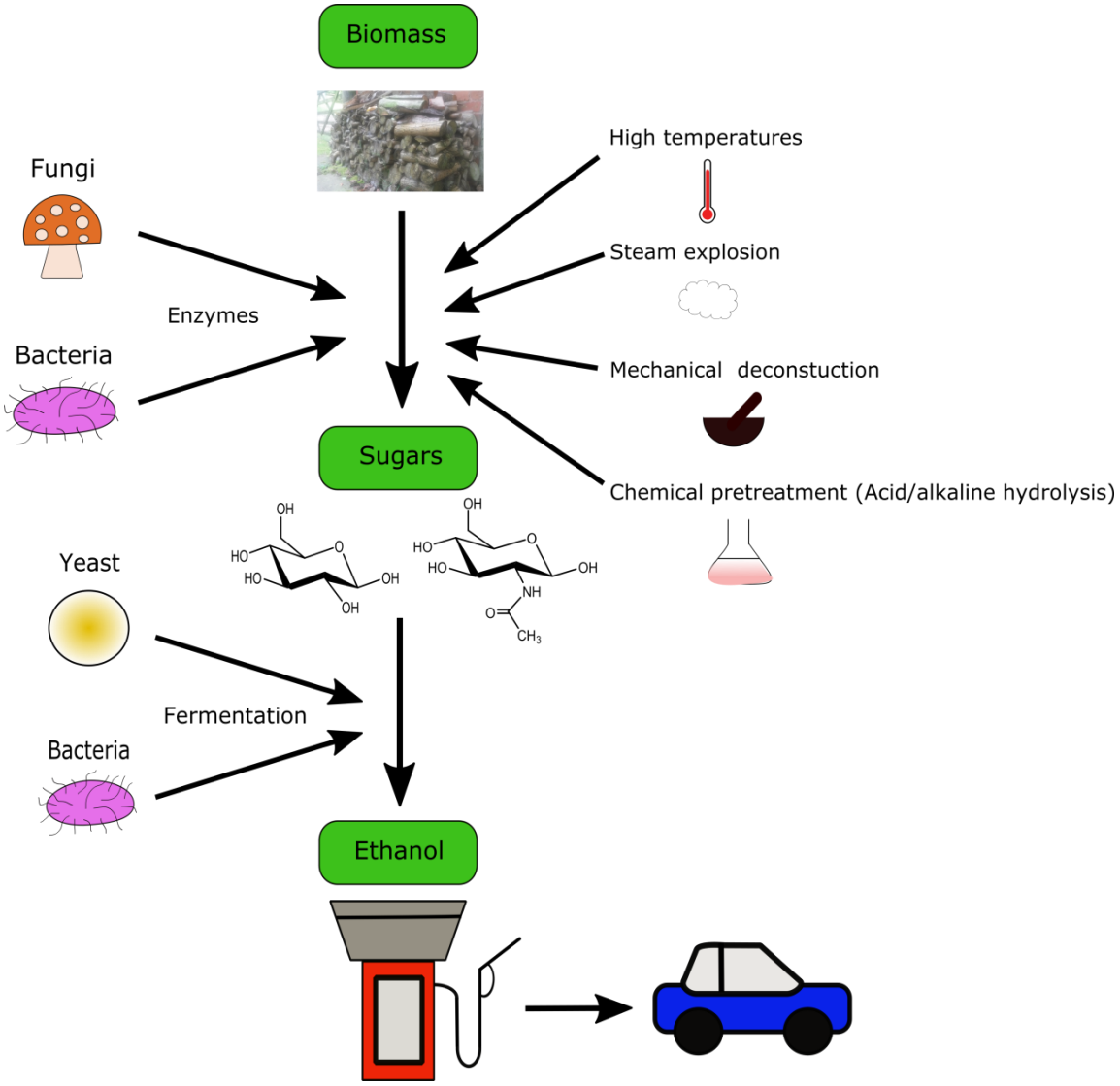

Figure 1.1.1 A flow diagram of the possible processes involved in the treatment and pre-treatment of biomass materials to their respective sugars in the production of bioethanol for fuel.

1.2 The Composition of Biomass for Second Generation Biofuel Production

In the recent drive towards second generation biofuels, understanding biomass, and developing the most effective ways of producing bioethanol products from it, has become an integral part of this type of biotechnology. Biomass typically consists of living or recently deceased plants or animals, hence rejecting the fossil fuel sources of coal, oil and other fossilised remains. One of the most commonly known types of biomass arises from the waste matter produced following crop harvesting and cultivation. The constant supply of waste materials resulting from crops such as wheat and corn, as well as wood chip sources and forest debris, means that deriving second generation biofuels from this type of biomass is a much more sustainable form of energy production. This biomass is known as lignocellulosic biomass as it is rich in cellulose and lignin, the materials that form the cell wall of plants (Figure 1.2.1).[7] Lignocellulose is a complex matrix consisting of diverse polymers including cellulose, "hemicelluloses" and lignin. Both cellulose and hemicellulose are complex polysaccharides, with cellulose consisting of chains of β (1,4)-linked glucose molecules arranged in a crystalline fashion, whilst hemicellulose contains various polysaccharides, formed from arabinose, galactose, glucose, xylose and mannose, with their composition varying depending on the plant source. Lignin, rather than being a polysaccharide, is comprised of three phenolic compounds – p-coumaryl alcohol, coniferyl alcohol and sinapyl alchohol – cross-linked together to form a highly complex and recalcitrant structure.[8] In addition to cellulose, hemicellulose and lignin, other polysaccharides, such as pectin, contribute to the overall make-up of different plant cell walls. In a related structure, the cell wall of fungi and the shells of insects and crustaceans, the β–1,4 linked polysaccharide of *N*-acetyl D-glucosamine, commonly known as chitin, is also found.[9] The plant cell wall (and these related structures) therefore represent a highly heterogeneous substrate which is very difficult to break down, a goal that must be achieved in order to release the sugars contained within it and to derive energy from these materials.[10]

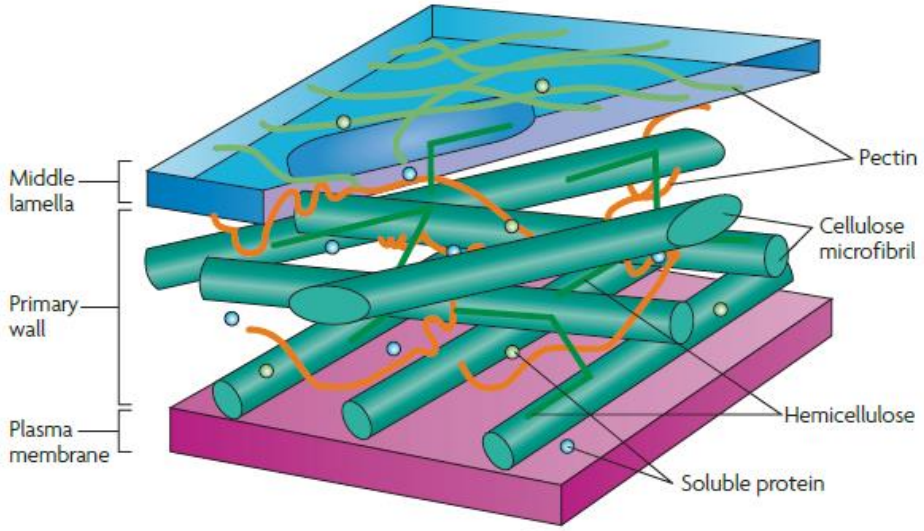

Figure 1.2.1 Representation of the plant cell wall, demonstrating the many components which contribute to the makeup of the primary cell wall housed between the plasma membrane and the middle lamella. The image shows how the different polysaccharide components form a complicated matrix, one which is extremely tough to break down. (Image taken from Sticklen, 2008.[7])

1.3 Cellulose: a tough substrate for bioethanol production

In order to derive biofuels from polysaccharides contained within the cell wall, the individual sugar units contained within these substrates need to be liberated. Long chain polysaccharides are the most abundant components of the plant cell wall and are made up of varying units of glucose and glucose derivatives. They contain large numbers of carbon atoms bonded together by strong glycosidic bonds that hold the monosaccharides together. The glycosidic bonds within polysaccharides such as cellulose have been proven to require a greater enthalpy of activation (~ 30 kcal/mol) for their spontaneous hydrolysis than other biological polymers, such as DNA.[11] Polysaccharides can exist either as linear chains or as highly branched structures. Well known branched polysaccharides include starch and glycogen. These can be classed as storage polysaccharides, as they are able to store energy for various organisms. The storage polysaccharides are generally easily broken down into their monosaccharide moieties, as the energy required for the organism must be easily accessible. In the case of starch its numerous accessible chain ends contribute to its ease of degradation, indeed the "sweetness" of starchy foods. The structural polysaccharides such as those that form the plant cell wall, on the other hand, are much more difficult to break down.

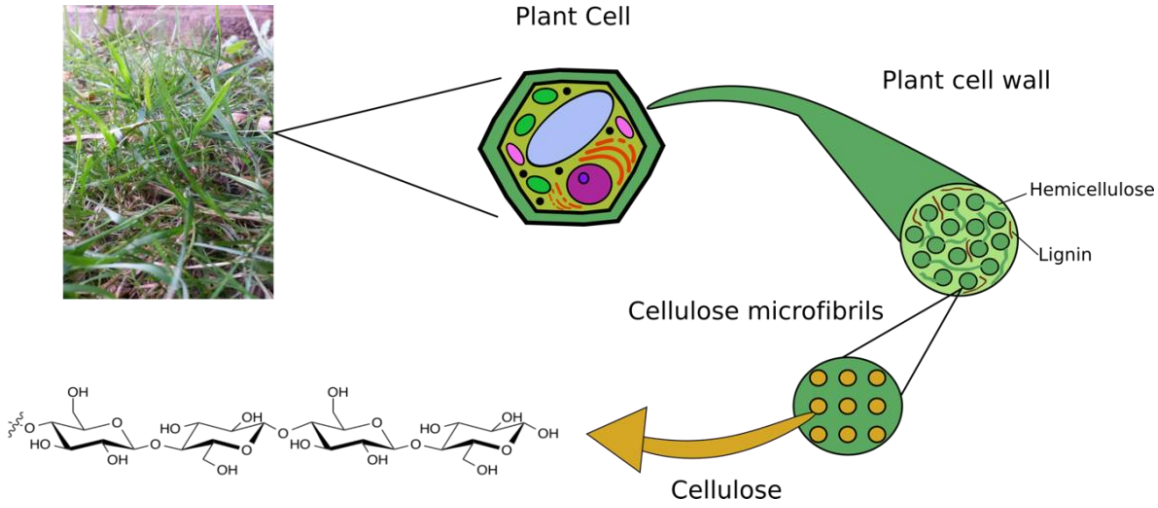

Figure 1.3.1 A schematic diagram to show the arrangement of cellulosic microfibrils, which are found within the plant cell wall, along with hemicellulose and lignin to form lignocellulose.

Cellulose is a structural polysaccharide, and is the most abundant polysaccharide in the plant cell wall. It is formed from linear chains of β-1,4-linked glucose units and individual chains can reach tens of thousands of sugars in length providing tensile strength and support to the cell. Cellulose is the most abundant biopolymer on earth,[12] and it is estimated that around 80 x 10^{12} kg is produced each year, resulting in it fast becoming a readily available source for biofuel production.[13] As shown in Figure 1.3.1, the biopolymer is found tightly packed together in microfibrils, with individual chains packed tightly and held together by hydrogen bonds. Within the cellulosic microfibrils, cellulose is found in both amorphous and crystalline forms, with regions of varying crystallinity. In industry, cellulose is mainly involved in the production of paper and cardboard,[12] as well as in various textiles, mainly cotton, and as a precursor to the synthetic fibre, Rayon.[14] Studies into the hydrolysis of individual glycosidic bonds within cellulose were carried out by Wolfenden *et al.*,[11] who determined the half-life of cellulose at 25 °C to be 4.7 million years, which endorses the tough nature of the polysaccharide.

1.4 Chitin: another tough to degrade polysaccharide

Another polysaccharide that shares many features with cellulose is chitin (Figure 1.4.2). Chitin is made up of monomeric units of *N*-acetyl D-glucosamine to form a long chain polymer covalently linked by β-1,4 glycosidic bonds. It can, therefore, be considered a "derivative" of the more abundant polysaccharide cellulose, with the 2-hydroxyl group being replaced with an N-acetyl group. This provides chitin with great strength, due to the hydrogen bonds between the layers within the chitin structure. For this reason, it is still possible to find polysaccharides such as chitin in fossilised samples of insects even after millions of years. One example is that of fossilised insects from the Oligocene era in Enspel, Germany, containing preserved chitin after 24.7 million years.[15]

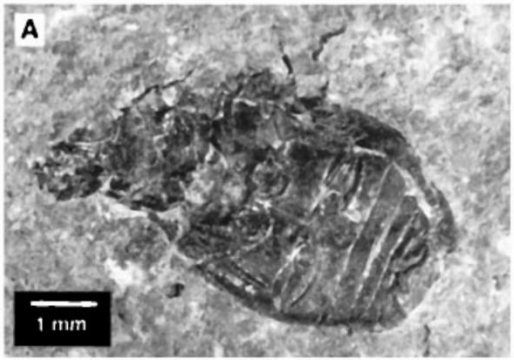

Figure 1.4.1 Photomicrograph of a beetle from the Oligocene era in Enspel, Germany, which was found to contain preserved chitin. (Taken from Stankiewicz *et al.*[15])

Chitin is mainly found as a crystalline structure, and can be present as α-chitin, β-chitin or γ-chitin.[16] The polymer molecules within α-chitin are arranged antiparallel to one another, whereas in β-chitin they are arranged in a parallel manner. γ-chitin consists of both antiparallel and parallel forms of the chitin molecules. β-chitin is thought to be more reactive than the other forms, due to fewer intermolecular hydrogen bonds present between the polymer layers.[16] β-chitin is found naturally in squid pens, cuttlefish and tubeworms, but is not as abundant in nature as α-chitin, which can be found in crabs, lobsters, shrimp and insects,[16] as well as in the cell walls of fungi.[9] It is estimated that 10^{11} tons of chitin is produced annually, most of which is found in the oceans.[15] The subsequent accumulation of shellfish waste within the food industry therefore results in large quantities of difficult to degrade waste being thrown away and simply unused. By harnessing the energy generated from degrading the waste into useful products, the methods of degradation of chitin can be applied to cellulose degradation from

lignocellulosic biomass, hence contributing to the biofuel industry.[17, 18] In addition to its biofuel potential, chitin has other applications in various different industries. For example, it is used in the food industry as a thickener and a stabiliser,[19] as well as being used for surgical threads in medicine.[20] It can easily be converted to the product chitosan, as well as used as a biopesticide in the agricultural industry.[21] Another application for chitin is within scientific research, as a method of purifying proteins using a chitin binding tag attached to the protein of choice, and using a column of chitin beads to which the protein will be able to bind.[22]

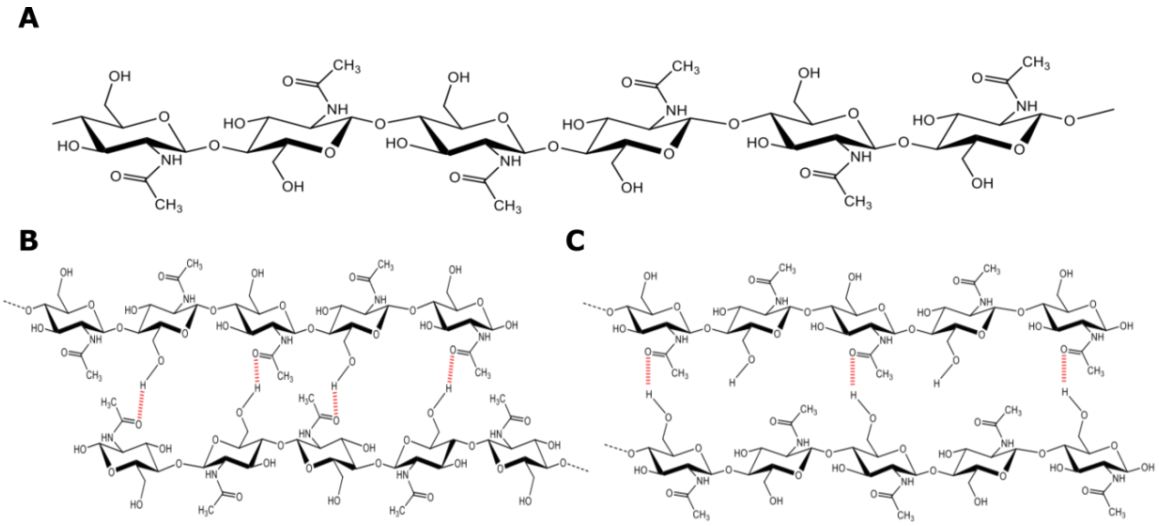

Figure 1.4.2 (A) The general structure of one chain of the polysaccharide chitin, consisting of monosaccharide units of *N*-acetyl glucosamine. (B) The alpha (α) form of chitin, where chains of the polysaccharide are arranged in an anti-parallel fashion and hydrogen bonds are formed between the anti-parallel chains (shown in red). (C) The beta (β) form of chitin, where the chains are arranged in a parallel fashion and forms fewer hydrogen bonds (again, shown in red) than the alpha form.

1.5 Classical degradation of biomass

The study of how microbes are able to degrade and derive energy from biomass began in earnest in the 1950s. Reese *et al.* (1950) were the first to propose that in nature the degradation of recalcitrant cellulose occurs via a two-step process: 1) the cellulose is converted into long chains of anhydroglucose (oligosaccharides) arranged in a linear fashion; 2) the β-1,4 glycosidic bonds linking the anhydroglucose together are hydrolysed, resulting in individual glucose units.[23] This two-step process invoked the "C1-Cx" notion involving not only hydrolytic enzymes in the Cx part of the process, but also an undefined C1 component which

disrupted the overall structure of the cellulose substrate, enabling it to be more accessible to degradation by the hydrolytic enzymes.[23] Many other studies into hydrolytic enzymes present in fungi and bacteria were carried out in order to understand how the organisms are able to degrade polysaccharides such as cellulose and chitin,[17, 18, 24, 25] and the products obtained upon hydrolysis.[26]

Traditionally, degradation of polysaccharides, such as chitin, occurs by the synergistic action of enzymes, of which there are three different types: *endo*-1,4-β-chitinases, processive *exo*-chitinases and β-hexosaminidases.[27] *Endo*-enzymes create chain ends by cleaving internal glycosidic bonds within the polysaccharide molecule. These non-reducing chain ends can then be hydrolysed by the *exo*-acting enzymes, which degrade the polysaccharide processively to create disaccharides, such as chitobiose.[27] The final step in this classical degradation of polysaccharides to soluble monosaccharides is the conversion of these disaccharides into *N*-acetylglucosamine by hexosaminidases. Similar schemes exist for cellulose, with *endo* and *exo* glucanases, termed cellobiohydrolases, and β-glucosidase to produce glucose as a precursor for ethanol.[27] One of the main problems with this degradation method, however, is the low level of synergy between the *endo*- and *exo*-enzymes, rendering degradation of many polysaccharides, such as cellulose, mannan and chitin to be slow and commercially ineffective.[27]

Dating back to the Reese paper,[23] it had long been considered that other enzymes may contribute to the breakdown of these materials. For example, it is known that white rot fungi possess enzymes which are able to degrade lignin and cellulose found in wood and, in the case of *Lentinula edodes*, are able to degrade lignin by direct oxidation.[28] In addition, the discovery of cellobiose dehydrogenase as an oxidative enzyme which enhanced the degradation of cellobiose first introduced the notion that oxidative mechanisms may be involved in the enzymatic degradation of biomass substrates into bioethanol.[29, 30] Cellobiose dehydrogenase consists of a haem-binding cytochrome domain linked to a FAD (flavin adenine dinucleotide) dehydrogenase enzyme, which uses electron transfer methods to oxidise cellobiose at the C1 position to produce cellobiono-1,5-lactone, a precursor to cellobionic acid (aldonic acid).[29, 31] Studies confirmed that the inclusion of this oxidative cellobiose dehydrogenase enzyme improved the degradation of cellulose when used as a pre-treatment, or in combination with typical cellulases.[32, 33] By investigating how organisms, such as fungi and bacteria, are able to utilise various enzymes at their disposal in the degradation of recalcitrant polysaccharides for their benefit, it is possible to exploit the evolutionary machinery of nature in our quest for more effective and efficient methods of biomass degradation.

1.6 Lytic Polysaccharide Monooxygenases: a new age in biofuel production

The notion that oxidative methods were involved in the concert of enzymes used by natural organisms to degrade polysaccharide substrates was first introduced as early as 1974, when Eriksson *et al.* found that the degradation of cellulose by the *Sporotrichum pulverulentum* fungus rapidly consumed oxygen from the atmosphere, indicating that an oxidation reaction was taking place.[30] In addition to the cellobiose:quinone oxidoreductase enzyme, reported previously in Westermark *et al.*,[29] a further unknown enzyme was isolated from the fungus which was able to produce twice the degradation of cellulose when present in the cocktail of endo and exo-glucanase enzymes, but only in the presence of oxygen. Although this paper was largely ignored at the time, the introduction of an oxidative step into the enzymatic degradation of biomass substrates has now led to a reinterpretation of the 'two-step' process put forward by Reese *et al.*[23]

A recently characterised class of enzymes has been found to play a vital role amongst the cocktail of enzymes that achieve the degradation of recalcitrant polysaccharides in nature.[34, 35] These enzymes are now known as the lytic polysaccharide monooxygenases (LPMOs).[35-38] The excitement surrounding LPMOs emerged in 2010/11 when they were discovered to be able to increase the degradation of polysaccharide substrates by a factor of 2 when used in concert with classical cellulases, as shown by Harris *et al.* in their study on the *Trichoderma reesei* enzyme.[39] At approximately the same time Vaaje-Kolstad *et al.* showed that a protein from *Serratia marcescens*, initially thought to be a carbohydrate binding domain CBM33,[40, 41] was able to boost the activity of chitinases and showed for the first time that these enzymes were oxidative enzymes, using a mechanism that was distinct from the classical glycoside hydrolases.[34] There are now many reports of the "boosting" effect that LPMOs have on the action of glycoside hydrolases on polysaccharide substrates.[34, 36, 42, 43] These include a study by Nakagawa *et al.*,[44] which showed that an AA10 from *Streptomyces griseus* was able to increase the efficiency of chitinase enzymes by 30- and 20-fold, on both α and β forms of chitin, respectively. The utility of LPMOs in polysaccharide processing is therefore now clear, and there is subsequently a worldwide drive to ensure that these enzymes are used optimally to allow the production of second generation biofuels.

1.7 CAZy Classification of LPMOs

The CAZy classification system was established by Henrissat *et al.*,[45] and delivered online (http://www.cazy.org/) in 1998, as a way of collectively grouping Carbohydrate-Active Enzymes (CAZymes) on the basis of amino-acid sequence similarities.[46] Originally used for

classifying glycoside hydrolases,[45, 47-50] the database was expanded to include glycosyltransferases[51, 52] so that enzymes that both cleave and create carbohydrates could be grouped according to their amino acid sequence similarities. By classifying the enzymes in this way, it is possible to infer the structure (where a homolog is known), reaction stereochemistry, and to some extent if sub families have been defined, potential substrate specificity. This is a very useful tool and so the CAZy database has since been expanded to include polysaccharide lyases and carbohydrate esterases.[53] In addition, the decision was made to include non-catalytic proteins as associated modules in the database. Carbohydrate binding modules are found within carbohydrate active enzymes such as glycoside hydrolases, and so despite them not having a direct enzymatic effect on carbohydrate substrates, the distinct modules are able to bind to the substrate in question enabling the catalytic enzymes greater access to carry out their specific function.[54] Following the discovery of LPMOs, and the renewed importance of oxidative enzymes in carbohydrate processing, a new class of carbohydrate active enzymes was added to CAZy in 2013 – the auxiliarly activity (AA) class.[55] This class includes both the lytic polysaccharide monooxygenase enzymes, as well as lignin degrading enzymes, which both provide supporting roles in the degradation of lignocellulose. LPMOs, which were originally designated as GH61[35] and CBM33[36] in CAZy, were therefore reclassified into families AA9 and AA10 respectively. Recently there has been an additional expansion of the AA classification with the discovery of the AA11[56] and AA13[57] families of LPMOs, as chitin and starch active LPMOs respectively. The LPMOs therefore encompass four families in the CAZy database. By studying both the structures and activities of the different classes, useful insights into how the LPMO enzymes work in nature, including their mechanism of action, can be achieved for industry to take advantage of in its quest for more efficient biomass degradation, reviewed in Dimarogona *et al.*,[58] Harris *et al.*,[59] Span *et al.*,[60] Beeson *et al.*[61] and Hemsworth *et al.*[62]

1.8 The Activity of Lytic Polysaccharide Monooxygenases

Prior to 2010/11 considerable work had been carried out on the proteins now known as LPMOs, but their activity and function had not yet been determined.[35, 40, 41] Originally the AA10, CBP21 (‘chitin-binding protein 21’), from *Serratia marcescens*, was identified as a carbohydrate binding module (CBM) that simply disrupted chitin physically without any enzymatic action.[40] In a parallel discovery, AA9 enzymes (then GH61) were shown to aid degradation in conjunction with classical hydrolases, on cellulose.[35] However, the initial non-enzymatic definition of these enzymes turned out to be incorrect, as the CBP21 was subsequently shown to have oxidative activity on chitin.[34] Analysis of the CBP21 degradation

products from its action on chitin, using Mass Spectrometry showed aldonic acids were the main degradation products.[34] Furthermore, analysis using ^{18}O-labelling revealed the oxidised ends contained one atom of oxygen from O_2, which coupled with the requirement for a reducing agent (typically ascorbate), showed that the enzymes work by an oxidative mechanism, rather than hydrolytically. In fact, research into these AA10 enzymes showed that they were metalloenzymes which are able to break down chitin into individual, mostly even-numbered, oligosaccharides (i.e. with degrees of polymerisation = 2, 4, 6 etc.).[34] It was originally suggested that this pattern of even numbered oxidation occurs because the structural repeating unit of chitin is the disaccharide chitobiose with each second C-H bond similarly disposed, shown in Figure 1.8.1.

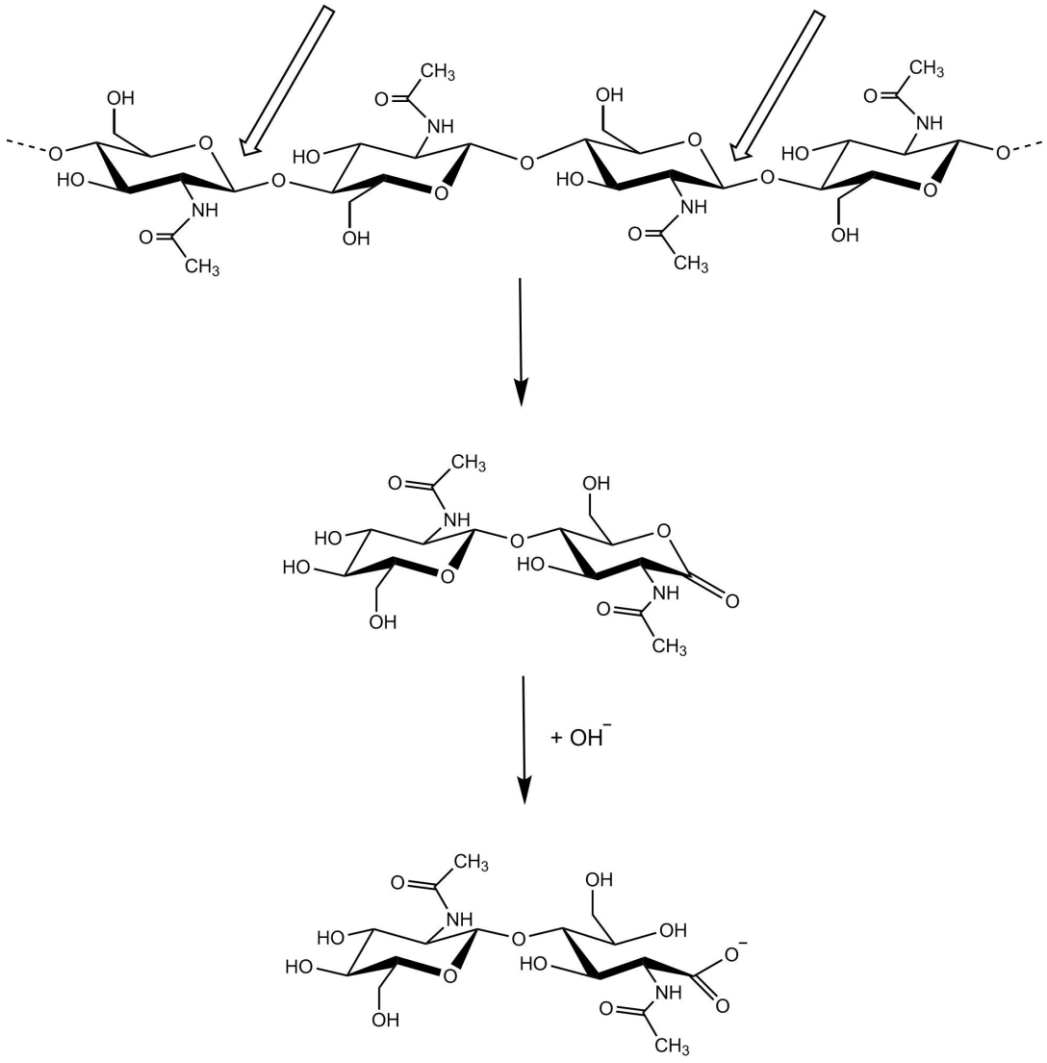

Figure 1.8.1 Schematic reaction mechanism of the breakdown of chitin to a lactone, before hydrolysis to an aldonic acid. Attack at the C1 positions on the chitin is demonstrated by the block arrows.

Following on from the initial demonstration of oxidative activity on chitin by CBP21,[34] activity was then demonstrated on cellulose for AA9 enzymes[63, 64] and other members of the family AA10 (full details of which can be seen in Table 1.8.1.). In addition to C1 oxidation releasing aldonic acid products, other regioselectivities have also emerged for some LPMOs, with reports of AA9 and AA10 enzymes that can also attack the C4 position, and some early reports of C6 specificity, though the validity of this observation has been cast into doubt more recently.[65-67] Activity by the only AA11 enzyme to date from *Aspergillus oryzae* has been studied, indicating that it also possesses activity on chitin, via C1 oxidation releasing aldonic acid products.[56] In contrast, the AA13 from *Aspergillus nidulans*, instead of degrading insoluble substrates, such as cellulose and chitin, has been found to have activity on recalcitrant 'retrograded' starch,[57] further broadening the range of substrates attacked by lytic polysaccharide monoxygenases.

A study by Agger *et al.*[68] showed that LPMOs were also able to degrade polysaccharides other than cellulose and chitin, as a fungal AA9 enzyme from *Neurospora crassa* was found to degrade hemicellulose, in particular xyloglucan. This opened up further opportunities to discover LPMO enzymes able to degrade other materials in the plant cell wall, or indeed other polysaccharides. For example, more recently, activity on cellulose, xyloglucan and even glucomannan has been found for an AA9 enzyme from *Gloeophyllum trabeum*.[69] With the discovery of LPMOs which are now able to break down more soluble substrates, such as xyloglucan,[68-70] starch[57] and soluble oligosaccharides[71] (cellotriose to cellohexaose), the lack of kinetics assays performed by LPMO enzymes due to their insoluble substrates can be addressed. Frandsen *et al.* were able to perform kinetics on an AA9 from *Lentinus similis* using a fluorescence-labelled cellotetraose upon which *Ls*AA9 had shown activity.[71] Such studies are able to provide a platform for further kinetics studies on LPMOs in the future.

Table 1.8.1 A table of AA10 enzymes in the literature to date, with their determined substrate, confirmation of the solved structure and the method/s used, as well as the key references for each enzyme.

Bacterial organism	Activity	Structure	References
Bacillus amyloliquefaciens	Chitin	Yes (X-ray diffraction and EPR)	Hemsworth *et al.*[72]
Bacillus licheniformis	Chitin	Cu-site only (EPR)	Forsberg *et al.*[73]
Bacillus thuringiensis	Chitin	No	Zhang *et al.*[74]
Cellvibrio japonicus	Cellulose,[75] Chitin[76]	Yes (X-ray diffraction)	Gardner *et al.*,[75] Forsberg *et al.*[76]
Enterococcus faecalis	Chitin	Yes (X-ray diffraction)	Vaaje-Kolstad *et al.*,[77] Gudmundsson *et al.*[78]
Hahella chejuensis	Cellulose	Yes (X-ray diffraction)	Ghatge *et al.*[79]
Jonesia denitrificans	Chitin	Yes (X-ray diffraction)	Mekasha *et al.*[80]
Lactococcus lactis	Chitin	No	Vaaje-Kolstad *et al.*[81]
Serratia marcescens	Chitin	Yes (X-ray diffraction[41] and NMR[82])	Suzuki *et al.*,[83] Vaaje-Kolstad *et al.*,[34, 40, 41] Aachmann *et al.*,[82] Nakagawa *et al.*[84]
Streptomyces coelicolor	Cellulose	Yes (X-ray diffraction and EPR[73])	Forsberg *et al.*[36, 73, 85]
Streptomyces griseus	Chitin	Yes (X-ray diffraction)	Nakagawa *et al.*[44]
Streptomyces lividans	N.D.	Yes (X-ray diffraction)	Chaplin *et al.*[86]
Thermobifida fusca	Cellulose	Yes (X-ray diffraction)	Forsberg *et al.*,[36, 73, 85] Moser *et al.*[87]
Vibrio cholerae	Chitin	No	Loose *et al.*[88]

1.9 Lytic Polysaccharide Monooxygenase Structures

The first crystal structure of an LPMO was that of the AA10, CBP21 from *Serratia* marcescens[41] in 2005 (Figure 1.9.1). However, it wasn't until 2010 that this enzyme was shown to exhibit an active site consisting of an N-terminal histidine plus a further histidine residue which bind a metal ion.[34] Similar topology and structure had been shown for the AA9 (then GH61) enzymes by Karkehabadi *et al.*[35] and Harris *et al*.,[39] where the active site lies on an extended flat surface on the protein 40 Å x 30 Å in size, enabling interactions with the solid substrate of cellulose or chitin to occur (see Figure 1.9.1). These AA9 structures were proven to be structurally homologous to the AA10 structure of *Serratia* marcescens,[41] confirming the two classes (GH61 and CBM33, at the time) to be structurally, and hence functionally, similar. Despite knowing the 3D structures of these enzymes, there was still confusion over the metal ion binding to the N-terminus of the proteins in the middle of the flat surface. It was not until structural and Isothermal Titration Calorimetry (ITC) studies by Quinlan *et al.* in 2011 that the identity of the metal was found to be copper.[38]

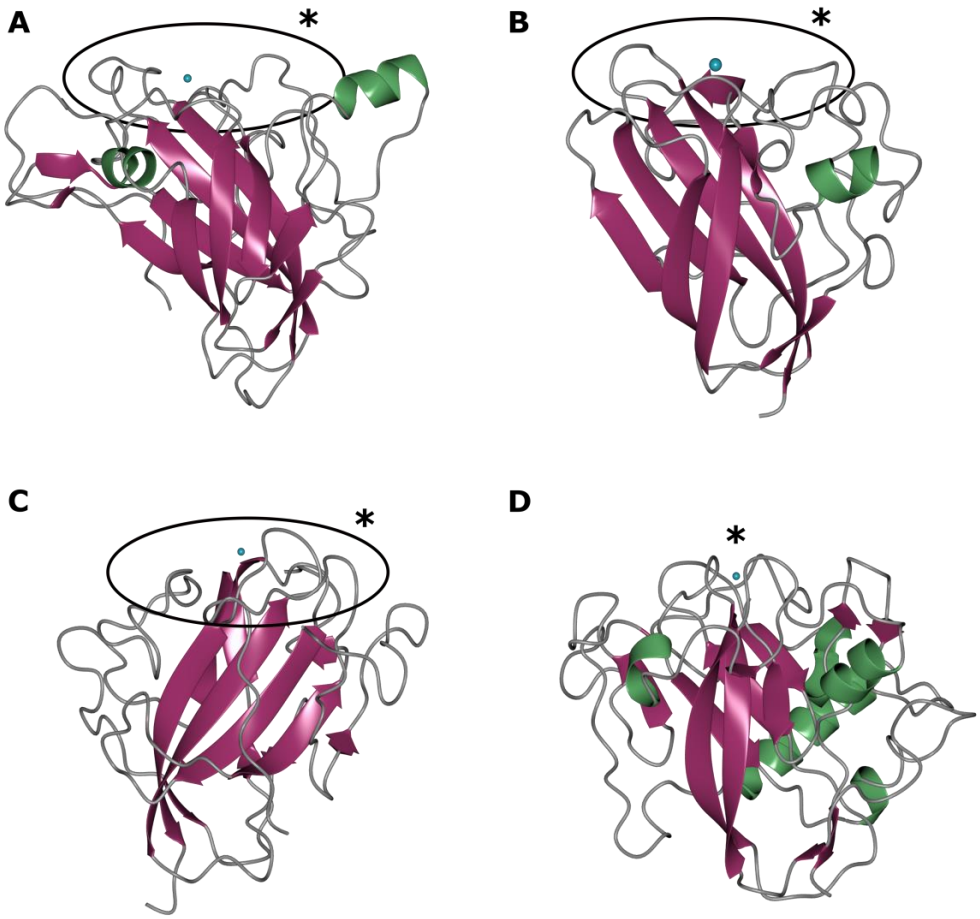

Figure 1.9.1 Structures of the Lytic Polysaccharide Monooxygenase enzymes of each Auxiliary Activity class showing their homologous structures and N-terminal metal binding sites: (A) the AA9 structure from *Hypocrea jecorina* solved by Karkehabadi *et al.*;[35] (B) AA10 structure from *Serratia marcescens*, which was the first LPMO enzyme to be solved;[40] (C) the more recent structure of the AA11 from *Aspergillus* oryzae;[56] (D) the AA13 structure also from the *Aspergillus oryzae* fungus.[57] The extended flat surface where the copper binding active site is found in the AA9, AA10 and AA11 structures is shown by the asterisked circle; and the groove in the AA13 structure is denoted by an asterisk.

In 2014, the AA11 class of LPMOs was discovered by a method known as 'module walking', whereby unknown domains linked to known LPMO classes, such as the AA9 class, are investigated to determine if they appear alone in other fungi sources. If they are found not linked to other modules, then it is likely they are a new class of enzyme, such as a new LPMO class. The only structure of an AA11 known to date is that from the *Aspergillus oryzae* fungus (shown as C in Figure 1.9.1).[56] Furthermore, in 2015, the AA13 class of LPMOs was identified using a similar method to the discovery of the AA11 enzymes. The structure of the AA13

enzyme from *Aspergillus oryzae*,[57] shown in Figure 1.9.1, was the first LPMO to display a groove on its surface. This is suggested to be a modification required in order to interact with its substrate, starch, which has a helical structure as opposed to the linear structures of cellulose and chitin. More recently, a unique structure of an AA9 LPMO from *Lentinus similis* with its oligosaccharide substrate bound to the enzyme has been obtained by Frandsen *et al.*[71] This structure provides great insight into where the substrate binds on the surface of the LPMO, showing the proximity of the C4-H bond to the ligands bound to the copper in the active site of the enzyme. The close proximity of this bond corroborates the fact that this enzyme carries out C4 oxidation (see Section 1.12).

1.10 The curious case of the 'histidine-brace' – LPMO active site architecture

3D structures of LPMOs provided information on the overall form of this class of enzyme, but the initial structures provided little clue as to their mechanism of action.[35, 41] Crystal structures (Figure 1.9.1) show that the copper active site lies in the middle of an extended flat face (~ 30 x 40 Å) showing few contours, other than in a few isolated cases.[56, 57] The copper binding site is somewhat unusual in that the metal is bound by two histidine sidechains together with the N-terminal amino group. The unusual nature of the metal binding site hindered correct identification of the metal ion which took some time to be defined, as Na, Mg, Ca, Ni, Mn and Zn were all suggested or inferred from various PDB files.[34, 35, 39, 77] With the realisation that LPMOs were oxidative enzymes, the redox active metal, copper, was finally confirmed as the redox cofactor used by LPMOs in a highly cited publication from Quinlan *et al.* in 2011.[38] Furthermore, Quinlan *et al.*[38] coined the term 'histidine brace' to describe the T-shaped N_3 arrangement of the histidine residues that coordinate the copper ion in the active site of these enzymes (Figure 1.10.1). This term is now in wide usage, not only for LPMOs but also for particulate methane monooxygenase (pMMO) enzymes, where the same structural motif appears, albeit coordinated to two copper ions. EXAFS studies on these enzymes show that two copper ions are involved in the metal binding site, resulting in a dimeric copper complex, but the discovery of LPMOs has now cast into doubt whether pMMOs truly bind two copper ions in their active site.[89-91]

In addition to identifying copper as the natural LPMO cofactor, Quinlan *et al.* were also the first to note the presence of the unusual τN-methylation of the N-terminal histidine, which had been present in previous structures but had not been recognised. This N-methylation is now routinely seen in fungal LPMOs purified from their natural sources or fungal expression systems,[41, 92] but not in cases where expression was carried out in *E. coli*[56] or *Pichia pastoris*,[37, 65] which appear to lack the biochemical machinery required to perform this modification. The

role of this methylated histidine is still not fully known, but possibilities include involvement in the mechanistic action in some way, or simply for redox control purposes. Indeed it has been proposed by Frandsen *et al.* that this N-methylated histidine in AA9 enzymes is involved in stabilising a lone pair to π^* electrostatic interaction[93] that occurs between an imidazole and the oxygen within a glycosyl ring.[71]

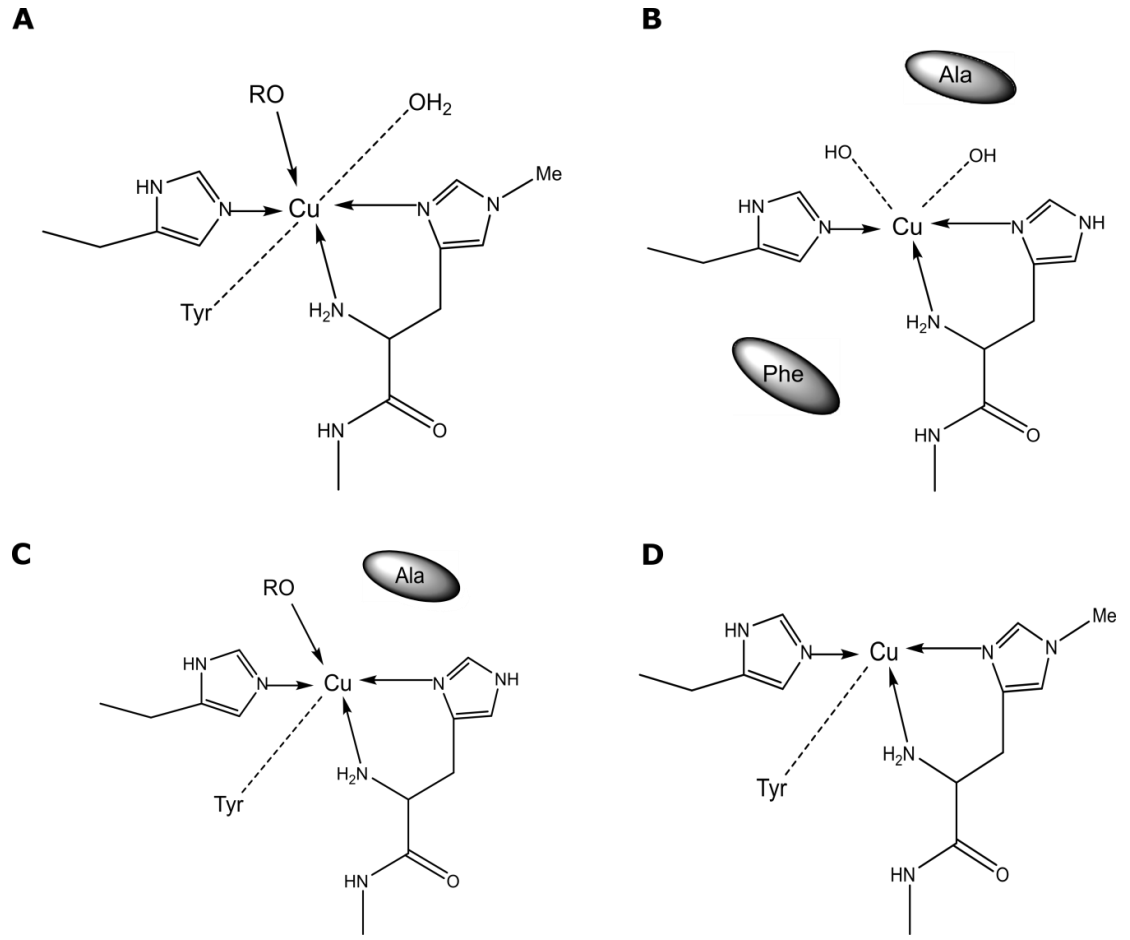

Figure 1.10.1 Copper active sites of the four Lytic Polysaccharide Monooxygenases: AA9 (A); AA10 (B); AA11 (C); AA13 (D), in each case showing the conserved 'histidine brace' coordinating to the copper amongst the different classes. The schematic figures show coordinating water molecules, or surrounding amino acids which provide a potential effect on the geometry of the copper active site.

AA10 LPMOs, which are largely of bacterial origin, appear to lack the N-methylation observed in fungal enzymes, but remain active nonetheless. Aside from the conserved histidine brace residues, all of the AA10 structures, along with sequence-alignments,[72] show further conserved residues surrounding the active site of the enzymes. In most cases phenylalanine is present, which replaces the common tyrosine found in the AA9 enzymes (Figure 1.10.1). In addition, a more highly conserved alanine residue is present, potentially blocking the axial coordination site of the copper ion due to steric hindrance and affecting its ability to accept ligands in this position.[72] In the case of AA11 LPMOs, a single crystal structure of this family has been determined. The structure is of an LPMO from *Aspergillus oryzae*, *Ao*AA11.[56] The structure shows that the AA11 class has sequence and crystal structure features which are similar to both AA9 and AA10 LPMOs. For example, the presence of the three N atoms coordinating to the

copper metal centre in the 'histidine brace', along with the overall shape where a substrate would bind to the enzyme's flat surface. Interestingly in the case of *Ao*AA11, the environment around the copper metal at the active site is somewhat intermediate between an AA10 and AA9 enzyme, as the phenylalanine residue typically found in AA10s has been replaced by a tyrosine, making it similar to the AA9 enzymes in this case, which could affect the steric hindrance around the copper site, and hence its mechanism of action. In the final known LPMO class, AA13, again a single crystal structure is known which also shows strong similarities to the AA9 and AA11 classes in that the familiar 'histidine brace' is present in the active site (Figure 1.10.1), as well as exhibiting the methylated N-terminal histidine residue most commonly found in fungal AA9 enzymes. It is notable in this regard that EXAFS studies on the AA13 class performed before the structural determination had suggested a potentially different coordination geometry at the copper ion involving a third histidine side chain, which was then shown not to be the case in the subsequent structural determination.[94]

One of the difficulties for structural biologists in studying the LPMO active site has been the propensity for the copper ion to undergo photoreduction from Cu(II) to Cu(I) in the X-ray beam during data collection. Indeed, this may have contributed to the confusion and difficulties in assigning the correct metal cofactor used by LPMOs during the early days of LPMO analysis. Photoreduction of copper is common in X-ray structural analysis and has been noted in various publications.[72, 78, 92] Photoreduction occurs due to damage caused by the X-rays, which are able to excite and transfer electrons to the protein, creating radicals. There are many examples of photoreduction occurring from synchrotron X-ray radiation damage for various metalloenzymes, including X-ray induced photoreduction of Fe(III) to Fe(II) in ferric myoglobin nitrate,[95] noted by changes in the UV-vis spectrum for the protein. A further example is the photoreduction at cryogenic temperatures of manganese and iron in the di-metal sites of ribonucleotide reductase.[96] In the context of LPMOs, the early structures are likely to have contained a mixture of Cu(II) and Cu(I) oxidation states, with most having the copper ion in the reduced state. The first explicit study into this phenomenon was carried out on an AA10 from *Bacillus amyloliquefaciens* and formed part of the wider study reported in this thesis (see Chapter 5). The first use of reduced-intensity X-ray collection strategies to obtain Cu(II) structures of LPMOs was carried out in 2014 on AA10 LPMOs of an *Enterococcus faecalis* bacterium. These structures were published by *Gudmundsson et al.*,[78] who investigated the photo-reduction of Cu(II) to Cu(I) by X-rays, and showed that it is possible to obtain structures with Cu(II) using helical or spiral data collection methods, as it enables the least amount of damage from the X-ray beam.

Given the difficulties in the interpretation of structural data due to photoreduction of the metal centre, a very powerful and useful tool used to ascertain the LPMO metal centre geometry of the Cu(II) state has been Electron Paramagnetic Resonance (EPR) Spectroscopy.

1.11 LPMO Electron Paramagnetic Resonance Spectroscopy Analyses

Electron Paramagnetic Resonance (EPR) Spectroscopy is a powerful tool for the analysis of redox metal centres in solution. The technique investigates the magnetic environment of unpaired electron(s), which is determined in large part by the relative energies and magnetic couplings of the molecular orbitals associated with the unpaired electron(s), which in turn is determined by structural arrangement of atoms in the vicinity of the unpaired electrons. Accordingly, the magnetic environment of the electron(s) as determined by EPR can be used to elucidate atomic structure. In terms of LPMOs this is a powerful tool, as the Cu(II) form of the enzyme contains a single unpaired electron (S = ½) at the copper active site, such that information can be directly obtained regarding the active site from a simple EPR experiment. Moreover, since there is only a single unpaired electron, the subsequent analysis of spectra is not complicated by significant electron-electron interactions (so-called fine coupling) and it is possible to obtain anisotropic EPR spectra of Cu(II)-LPMOs which contain three-dimensional information about the electronic environment. This information is somewhat complex and abstract, where a full analysis requires in-depth orbital determination. For the purposes of copper bioinorganic chemistry, however, this analysis is made significantly easier by the use of so-called Peisach-Blumberg plots (PB plots).[97] These plots use the property that the values of the *g* and *A* tensors, which are used to define the electronic environment, are anti-correlated. The *g* tensor is a measure of the strength of the magnetic field, known as the orbital angular momentum, created when an electron orbits a nucleus. It can be measured in the x, y and z directions, providing us with three separate *g* values (g_x, g_y and g_z). The *A* tensor represents the coupling the electron has with the nucleus, so will be affected by the strength of the magnetic field created by the electron. In other words, as a *g* value goes up then its corresponding *A* value falls. A plot of *g* values against *A* values therefore yields a linear correlation. In the context of Cu(II) species, it is normal to plot g_z against A_z to yield a PB plot, an example of which is shown in Figure 1.1.1 for a copper-dependent quercetin 2, 3-dioxygenase enzyme from *Aspergillus japonicus*.[98] The position of the values on the plot then allows the user to gain an approximate idea not only of the copper coordination geometry but also the types of ligating atoms.

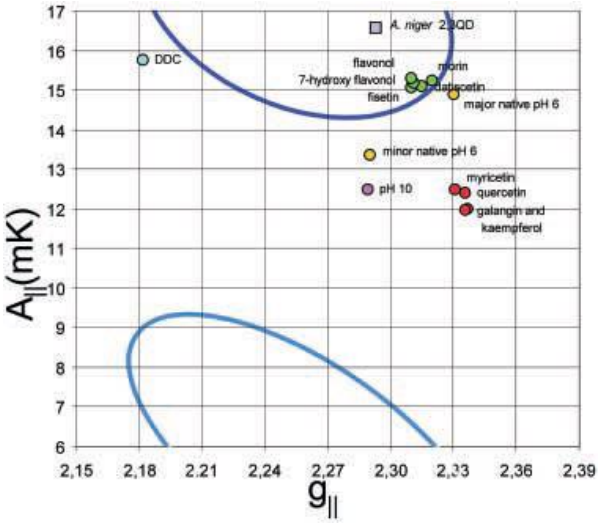

Figure 1.11.1 An example of a Peisach-Blumberg plot displaying the likely areas where type 2 copper sites in proteins may be found (dark blue/top circle), and type 1 copper proteins (light blue/bottom circle). (The plot is of $g_{//}$ and $A_{//}$ values for a copper-dependent quercetin 2, 3-dioxygenase from *Aspergillus japonicus* as well as for various small molecule Cu(II) model complexes that possess different flavonols as ligands taken from their EPR spectra.[98])

Figure 1.11.2 shows the spectra produced for an enzyme from each of the four LPMO classes (AA9, AA10, AA11 and AA13), and provides insights into the coordination and geometry around the metal sites. The first EPR analysis of the copper binding site in an AA9 was performed by Quinlan *et al*, which showed that *Ta*AA9 possessed a type 2 copper coordination geometry at the active site.[38] Subsequent analyses of fungal LPMOs from three CAZy classes - AA9, AA11 and AA13, showed that they all produce very similar copper signals, where the singly-occupied molecular orbital (SOMO) contains mainly $d(x^2\text{-}y^2)$ character, also placing them within the typical type 2 classification of copper models according to Peisach and Blumberg.[97] Type 2 copper centres are often square planar in geometry and the copper is coordinated by N and O atoms from residues such as His, Asp or Tyr. Unlike the type 1 copper sites, no sulfur atom coordinates to the copper ion, resulting in these proteins having weak absorption in UV-vis spectra at 700 nm, and no distinctive blue colour.

NMR experiments have shown binding of both Cu(II) and Cu(I) to AA10s, indicating that Cu(I) binds to the enzyme more strongly than Cu(II).[82] Due to the paramagnetic nature of the Cu(II) ion, however, any in-depth analysis of the coordination geometry around the copper active site could not be obtained in this way. EPR spectroscopy on some members of the AA10 family has shown reduced A_z and g_z values compared to EPR spectra of AA9 enzymes. This means that the $d(x^2\text{-}y^2)$ SOMO contains some $d(z^2)$ character which can only arise from a

structural distortion away from the canonical distorted axial geometry seen in type 2 enzymes. Indeed the EPR parameters of AA10 enzymes place them between the normal type 1 and type 2 classifications of copper coordinations according to Peisach *et al*.[97] Subsequent structural studies did indeed show that AA10 enzymes exhibit compressed or distorted trigonal-bipyramidal coordination geometries (Figure 1.11.2B).[72] To further complicate matters, this arrangement at the AA10 copper centre is not absolutely conserved across the family and is mainly found in chitin active AA10s. Interestingly, cellulose active AA10s display EPR spectra much more similar to AA9s and fall more closely into the class 2 classification of copper centres.[73] The reason for these differences is currently unclear, but delineating the effects of these differences is going to be crucial to allow further insight into the reaction mechanism used by these enzymes.

A further difference in the EPR spectra of LPMOs is found in the spectrum for the AA13 enzyme, which possesses a significant amount of resolved superhyperfine coupling in the perpendicular region (Figure 1.11.2D). It is still not known completely why this occurs for this enzyme, however it has been speculated that due to the enzyme's active site being in a groove as opposed to on a flat surface, the enzyme is structurally more ordered, allowing for a higher degree of coupling to the nitrogen atoms coordinated to the copper ion to occur.[57]

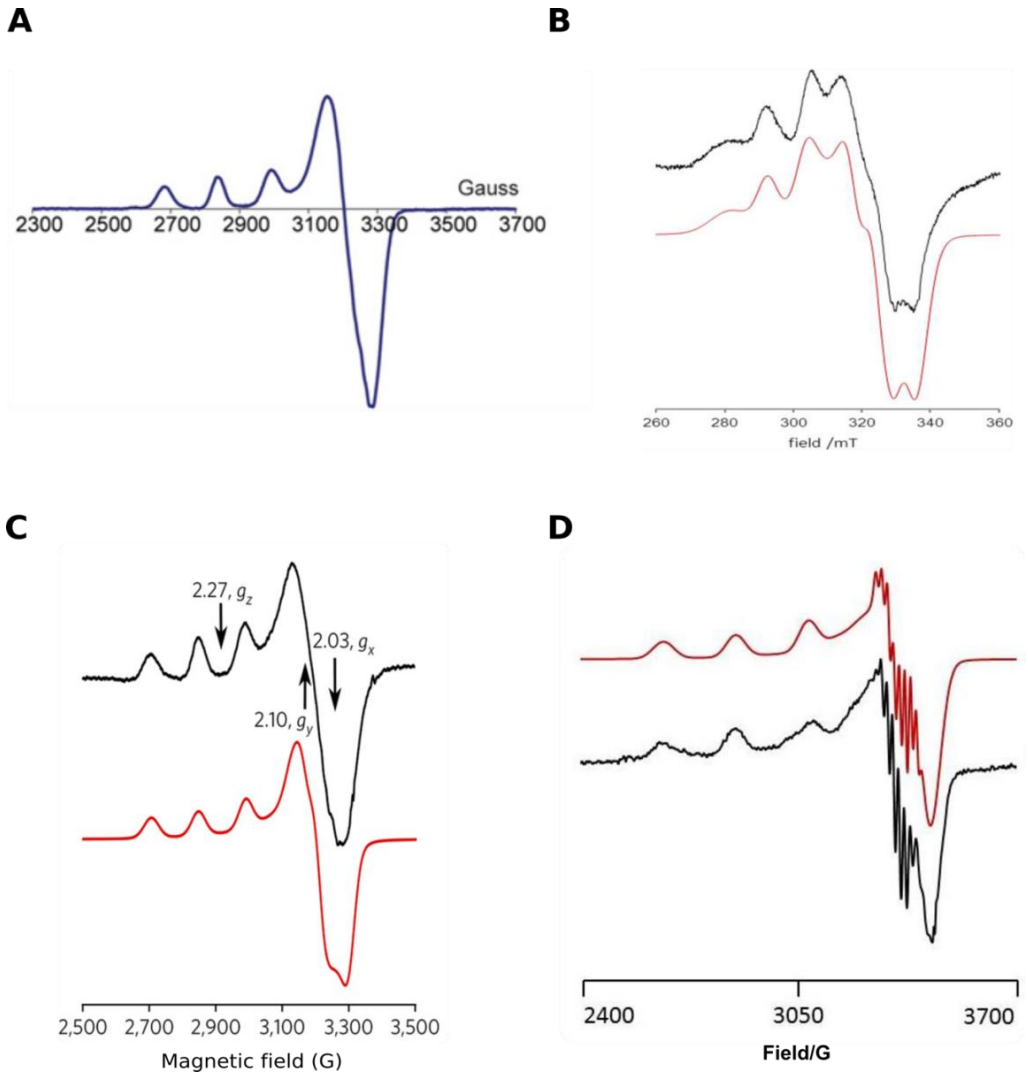

Figure 1.11.2 Published EPR spectra for each of the different classes of Lytic Poysaccharide Monooxygenase enzymes. (A) *Ta*AA9 EPR spectrum using X-band EPR at 140 K, pH 5 with ~ 15 % v/v glycerol.[38] (B) X-band EPR spectrum of *Ba*AA10 (black) along with the simulated spectrum (red), recorded at 155 K, pH 5 with ~ 15 % v/v glycerol.[72] (C) *Ao*AA11 EPR spectrum using X-band EPR at 150 K, pH 5 with 10 % v/v glycerol, again with the simulated spectrum in red.[56] (D) X-band EPR spectrum of *An*AA13 taken at 155 K, pH 5 with 10 % v/v glycerol, along with the simulated spectrum again in red.[57]

Following the determination of the structure of the *Ls*AA9 bound to its oligosaccharide substrate by Frandsen *et al.*,[71] the structure of the enzyme together with substrate was also investigated further using EPR. The spectra obtained contained well-resolved supehyperfine coupling, similar to that seen by *An*AA13 in Figure 1.11.2D. The induction of this superhyperfine coupling was attributed to the addition of another nitrogen magnetically coupled to the copper, suggesting that when the substrate is bound to the enzyme, the NH_2 of

the N terminus of the enzyme contributes to the magnetic environment of the copper ion, raising the energy of the SOMO, hence stabilising the enzyme's Cu(II) oxidation state.[71]

1.12 Proposed mechanism of action of Lytic Polysaccharide Monooxygenases

By gaining a thorough understanding of the active site architectures and coordination geometries of the copper centre in LPMOs, it is hoped that the mechanism of reaction for how these enzymes can catalyse such powerful chemistry with a single copper centre can be discerned. The current state of our knowledge regarding the general LPMO enzymatic reaction pathway has been well documented in papers by Phillips *et al.*,[63] Beeson *et al.*,[64] and in reviews such as Span *et al.* (2015)[60] and Hemsworth *et al.*[99] The current favoured general mechanism is that the oxidative reaction by the enzymes takes place via two separate one-electron reductions, the first occurring at the Cu(II) ion, which is reduced to Cu(I) by an electron transfer from a corresponding enzyme known as cellobiose dehydrogenase (CDH). Cellobiose dehydrogenase consists of an N-terminal heme domain which shuttles an electron, generated by the flavin-dependent dehydrogenase (FAD) domain of the CDH, to the LPMO.[31, 63] Once the copper ion within the LPMO is thus reduced it is able to bind oxygen, forming a copper-superoxide intermediate. This intermediate is able to attack either the C1 or C4 of the carbohydrate substrate, extracting the hydrogen from the C-H bond, generating a radical on the remaining carbon of the substrate and a copper-hydroperoxide species. A further electron transfer from the CDH is able to cleave the O-O bond within the copper bound species to form a copper oxyl radical, which can couple with the substrate radical to form a hydroxyl (OH) at the carbon atom. The addition of this oxygen causes destabilisation of the corresponding glycosidic bond, which results in an elimination reaction producing a lactone on the sugar, at the same position (C1 or C4) the attack occurred (see Figure 1.12.1).[63, 64] The choice to attack the C1 or C4 position of a polysaccharide substrate depends entirely on the LPMO, leading to the suggestion that amino acid residues surrounding the copper centre play an important part in the mechanism of enzymatic action. From sequence analyses it has been proposed that for the fungal AA9 enzymes, they may be classified into Type 1 and Type 2 PMOs (Type 1 oxidising at the C1 carbon and Type 2 oxidising at the C4 position).[63, 64] As attack by AA10 enzymes only seem to be performed at the C1 position so far, they have not been separated into different types. Indeed, further research into the products of degradation of various LPMO enzymes has concentrated on whether products are as a result of C1 or C4 oxidation,[100] often using isotopically labelled oxygen (O^{18}) to determine the origin of certain oxygen atoms within the products.[64] In addition, research into cellobiose dehydrogenase has continued, to determine how the electron transfer reaction occurs within nature and its synergistic activity with its

associated LPMO.[101] However, it is known from genomic information that some LPMO enzymes, AA10s in particular, do not possess cellobiose dehydrogenase enzymes, resulting in the question of where they acquire their electrons for reduction. One recent publication by Garajova *et al.* has found that certain flavoenzymes, glucose dehydrogenase and aryl-alcohol quinone oxidoreductases, were also efficient at providing a source of electrons for AA9 LPMOs.[102] Therefore, this enables the potential for further enzymes found within bacteria or fungi to be electron-donors for the LPMO enzymes.

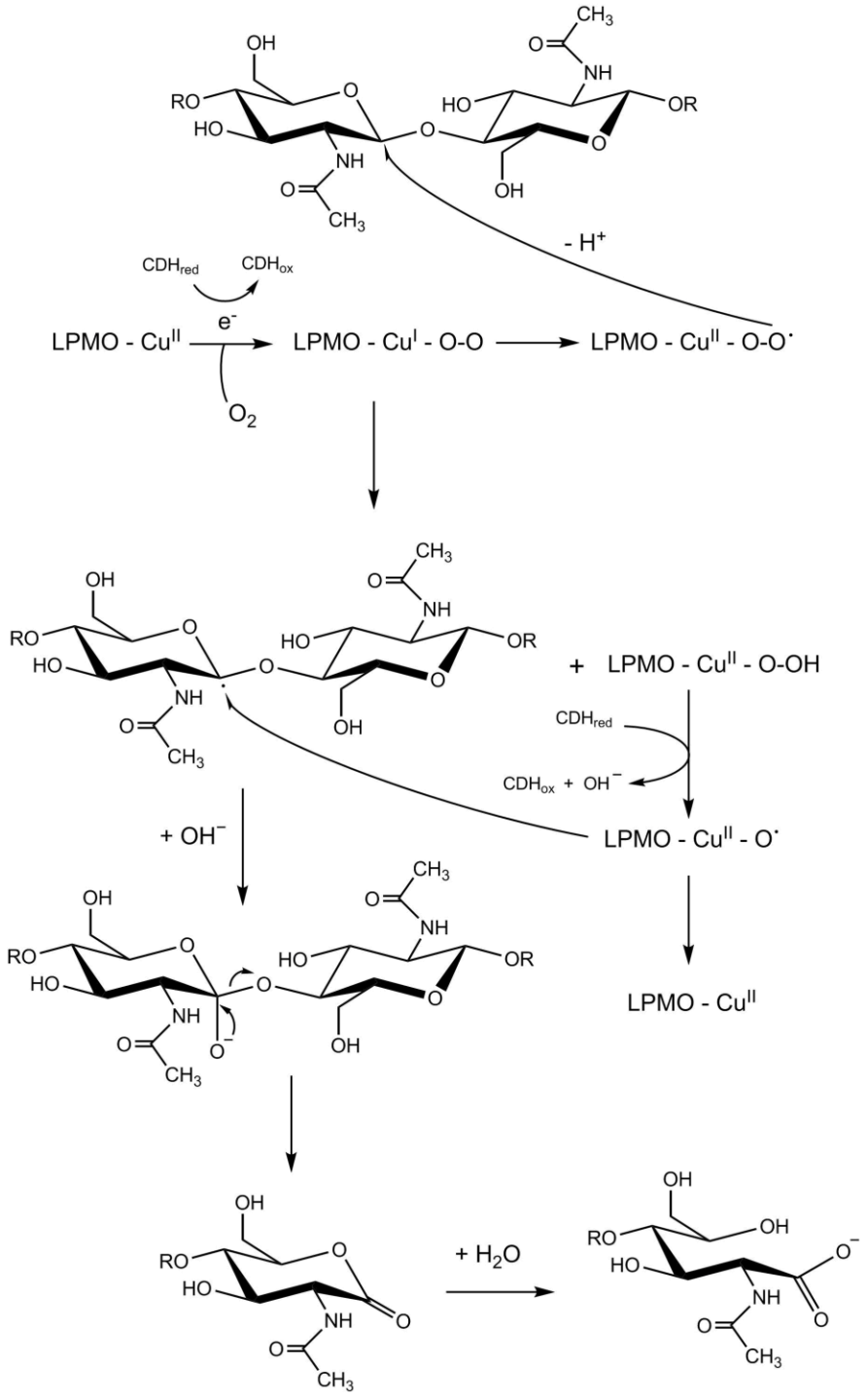

Figure 1.12.1 Detailed reaction mechanism of the proposed method of action of LPMOs, in this case on chitin with the attack occurring at the C1 position (R represents the rest of the chitin chain). The pathway shows the removal of a hydrogen atom, followed by production of radical species which can react to form the unstable oxygen species at the C1 position leading to cleavage of the glycosidic bond, generating the lactone product. This lactone can then be hydrolysed to an aldonic acid.

Despite the generalisation of how LPMOs work, the full extent of the mechanism of action, and the copper-oxygen species generated throughout, still remains a major enigma amongst the research into these enzymes. Significant evidence for the actual mechanism of action of at least one of the LPMO enzymes (AA9 class) was published by Kjaergarrd *et al.*,[103] who used a variety of methods to probe the Cu-AA9 reaction coordinate. These methods included EXAFS and XANES, as well as using density functional theories, which calculated the most thermodynamically favoured oxygen species to be an end-on Cu-AA9-superoxide complex involving a one electron reduction of O_2. Further insights into copper-oxygen species at enzymatic active sites has also recently been updated by Solomon *et al.*,[104] providing insights into how EPR spectra, as well as X-ray crystallographic studies, change upon coordination of ligands around the copper atom in an active site. Furthermore, Frandsen *et al.* were able to determine a likely method of action of the AA9 enzyme from *Lentinus similis* by successfully obtaining a structure bound to its more soluble oligosaccharide substrates. Using chloride ions bound in their structures in potential oxygen binding sites, they were able to suggest that any reactive oxygen species that is produced at the active site will only be generated if the polysaccharide substrate is bound to the surface of the enzyme, hence reacting in synergy with the substrate.[71]

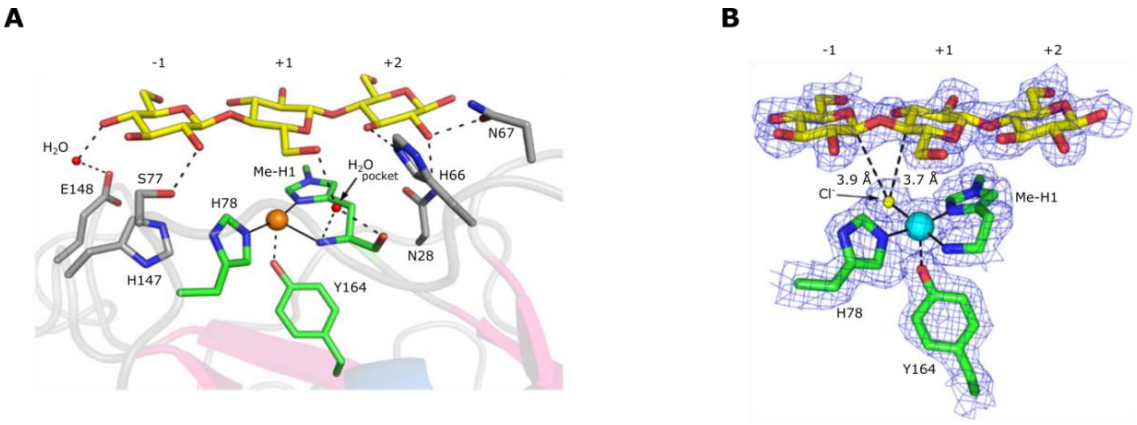

Figure 1.12.2 (A) Structure of the AA9 from *Lentinus similis* with cellotriose as its substrate bound to the active site, showing the various amino acid residue side chains involved in assisting with the binding on the surface of the enzyme. (In this structure, the copper has been photoreduced in the X-ray beam to Cu(I)). (B) 3D structure of the copper site in *Ls*AA9 portraying the electron density around the coordinating residues and the cellotriose substrate bound to the enzyme. In this structure the copper is found in its Cu(II) state and a chloride ion is bound in a cavity between the surface of the enzyme and the oligosaccharide substrate.[71]

Another integral part of the research into the LPMO enzymes is the efficiency of these enzymes and the synergy between them and their hydrolytic counterparts, as Morgenstern *et al.* explain in their review.[105] More knowledge of how efficient individual enzymes are and how they interact with the classical polysaccharide hydrolytic enzymes, will provide a better understanding of how to utilise these enzymes in an industrial setting in the future. Indeed, many studies are currently investigating the synergy between these LPMO enzymes and their classical hydrolase counterparts.[106] For example, a study of an AA10 from *Streptomyces griseus* has shown a 30-fold increase in α-chitin degradation, compared to a 20-fold increase in β-chitin degradation, showing how structural differences between the two types of chitin could make a difference on the binding of the AA10 for activity.[44]

The synergy between corresponding Carbohydrate Binding Modules (CBM), present among LPMOs, which act non-catalytically to bind to the substrate upon which the LPMO will act, is also an important area of research within this field. This is due to the modules providing an efficient method of allowing the LPMO to come in close contact with the substrate for it to bind and hence degrade. Many literature examples show LPMO activity on polysaccharide substrates regardless of whether a CBM is present or not, however, in some cases it does have an effect, as seen in Forsberg *et al.*,[85] where the removal of the CBM resulted in a reduction in the activity on phosphoric acid swollen cellulose (PASC). In addition, a recent study on CBM

modules involving deletion of the modules from the LPMO as well as replacements with CBMs from other organisms demonstrated the importance of the synergy between the two, and showed that the CBMs were both enzyme and substrate specific, inhibiting the action of the LPMO on the cellulose substrate.[107]

1.13 Summary and thesis aims

Second generation biofuels from plentiful biomass supplies have become increasingly important and interesting topics of various research in the past few years. The food vs. fuel debate is still ongoing, resulting in developments in these second generation biofuel production methods becoming key to the future of biofuels. Classical degradation methods involving the synergistic action of different enzymes relies on various enzymes working together in order to degrade the biomass substrate. With the discovery of Lytic Polysaccharide Monooxygenase enzymes, this process becomes more efficient, potentially resulting in the conversion of biomass into bioethanol being a more economically viable source of energy in the future.

The Lytic Polysaccharide Monooxygenases are copper-dependent fungal and bacterial enzymes that use an oxidative mechanism to cleave the glycosidic bonds within the polysaccharide substrates. The sequence similarities between enzymes show that the N-terminal histidine is very important in the mechanistic action of the enzyme. This histidine is involved in the 'histidine-brace' where the copper metal ion binds to three nitrogen atoms from two histidine residues and the N-terminus of the protein. Once reduced, the copper forms a reactive oxygen species with molecular oxygen, which is able to successfully attack the carbon atoms within the polysaccharide substrates, resulting in chain breaks to which classical enzymes can work more efficiently at degrading in a cocktail of enzymes.

My research focus has been on the bacterial AA10 class of LPMOs, in particular the AA10 from *Bacillus amyloliquefaciens* (*Ba*AA10), whose structure was solved prior to the start of my PhD.[72] However, the structure contained a reduced copper bound in the active site, resulting in the challenge of obtaining a copper(II) structure. Further questions to answer regarding the *Ba*AA10 enzyme at the beginning of my research involved confirmation of the binding of copper to the metal binding site and obtaining an accurate binding constant for the protein. In addition, the activity of *Ba*AA10 was to be researched, requiring both substrate determination and possible methods of action to be acquired.

Consequently, the subsequent chapter will address the production of the AA10 from *Bacillus amyloliquefaciens* from cloning to purification, and the steps taken to make the protein

production method more efficient, increasing the yield of the protein. Following this, the binding of copper in the metal site of the enzyme using Isothermal Titration Calorimetry (ITC) will be studied in Chapter 3, along with how various metals including copper affect the stability of the protein, hence confirming that the enzyme is copper-dependent. The fourth chapter will contain results of activity studies carried out with *Ba*AA10 and various polysaccharide substrates, including cellulose and chitin, with analysis of products by MALDI-TOF Mass Spectrometry. Finally, crystallisations of *Ba*AA10 and structural studies using X-ray crystallography will be presented in Chapter 5, before a conclusion chapter summarising the outcome of my PhD research as well as potential further experiments and research which could be carried out on the bacterial *Ba*AA10 and these Lytic Polysaccharide Monooxygenases in general. This work is published as part of Hemsworth *et al.*,[72] as well as more recently in Dalton Transactions (Gregory *et al*.[108]), both of which can be found in Appendix 2. In addition, thermostability research carried out on the glycosylated and de-glycosylated forms of a GH10 enzyme from *Malbranchea pulchella* using Differential Scanning Fluorimetry (DSF), which contributed to work published by Ribeiro *et al.*,[109] is also included in the Appendix 2.

2. Obtaining a Lytic Polysaccharide Monooxygenase from *Bacillus amyloliquefaciens*: the story of the cloning, expression and purification processes.

2.1 Abstract

The lytic polysaccharide monooxygense enzymes have proven themselves to be critical in the degradation of recalcitrant polysaccharides in both fungi and bacteria. In order to study their structure and function, efficient methods of production and purification are required to enable further investigations to take place. The bacterial LPMO from *Bacillus amyloliquefaciens*, *Ba*AA10, is the enzyme of choice in this research, and various methods have been used to obtain the pure protein. Initially, a periplasmic expression system was used and many results, including the structure published in 2013, were able to be produced in this way. However, for further analysis of this enzyme, a more efficient method of protein production was required, as yields following the original technique were reasonably low. This chapter, therefore, gives a detailed account of the development of a new method of expression and purification, involving utilising a SUMO tag within the protein construct for expression, cleaving the tag using SUMO protease, and final purification by gel filtration chromatography. This new process allowed a much greater yield of *Ba*AA10 to be produced, which could then subsequently be used for a variety of different techniques and analysis, including thermodynamics and stability, activity assays, crystallography and EPR spectroscopy.

2.2 Introduction

As discussed more fully in Chapter 1, the bacterial lytic polysaccharide monooxygenases are important enzymes in regards to their contribution to the breakdown of polysaccharides for biofuel production. Research has been published in the literature on AA10 enzymes from various bacteria including *Serratia marcescens*, which was the first to show catalytic activity on a polysaccharide,[34, 110] as well as *Streptomyces coelicolor*,[85] *Cellvibrio* japonicus[76] and *Enterococcus faecalis*.[78] All structures published on these enzymes have the common histidine-brace motif first identified in the AA9 enzyme from *Thermoascus aurantiacus*,[38] and all of the enzymes are known to be copper-dependent monooxygenases.

In this work, an AA10 from the bacterium *Bacillus amyloliquefaciens* was used to investigate the copper active site of the AA10 enzymes. *Bacillus amyloliquefaciens* can be translated as 'a bacteria which makes a liquefying amylase', and is a gram positive bacterium that was first discovered in soil in 1943 by Fukumoto.[111] The α-amylase it produces is used in starch hydrolysis, and the bacteria is also the source of the restriction enzyme BamH1. Being from the *Bacillus* genus, it is very similar to *Bacillus subtilis* in that it is capable of producing endospores which allow the bacteria to survive for longer periods of time. *Bacillus amyloliquefaciens* is non-pathogenic and its optimal temperature is between 30 and 40 degrees Celsius.

The AA10 enzyme from *Bacillus amyloliquefaciens* was originally identified as a 'chitin-binding protein', ChbB,[112] and was thought to act in a similar way to carbohydrate binding module proteins (CBMs), by simply binding to the substrate on which the other relevant glycoside hydrolases would act. According to the CAZy database the *Ba*AA10 protein is not directly linked to a CBM module, perhaps leading to the misidentification as a chitin-binding moiety itself.

Before any investigation into the potential substrate of this *Ba*AA10 protein could be carried out, however, an efficient and reliable method of production needed to be established, so that further experiments and analysis could take place. The most common and efficient purification method in protein production is the use of affinity tags, such as an N-terminal His-tag, which can then be removed to obtain the desired protein. However, the method used to cleave His-tags, which uses enzymes such as 3C protease, can disrupt the structure of the protein in the region where the affinity tag is cleaved. In *Ba*AA10 and other LPMO enzymes, the N-terminus forms part of the copper active site required for the role of the enzyme. Therefore, a purification protocol which enables the preservation of the N-terminal histidine residue as well as the N-terminus of the protein is required. Consequently, expressing the protein in the

periplasm of the cell using a pel-B leader sequence to direct the protein to the periplasm has enabled *Ba*AA10 to be purified with its N-terminus intact.[72] Another factor which led to this preferred method of purification, especially in the early stages of the LPMO research, was that some structures of LPMOs had been published in the PDB database with Ni as the metal bound in the active site. During these early stages of uncertainty as to the correct assignment of the metal ion in the active site, it was preferred not to include a nickel affinity tag, in order to prevent any excess nickel from binding. As a result, preparation of enzymes such as *Ba*AA10 using a periplasmic expression system became one of the methods of choice. A further advantage to periplasmic expression is that this system allows the formation of disulphide bonds within proteins, as a result of the oxidising environment of the periplasm. This oxidising environment allows enzymes such as disulphide oxidoreductase and disulphide isomerase to catalyse the formation and isomerisation of disulphide bonds.

A major disadvantage to producing proteins in the periplasm, however, is that expression yields are often significantly lower than with expression in the cytoplasm. This is because not all protein that is expressed is secreted into the periplasm. Some protein can also be found in the growth media, the cytoplasm and the cytoplasmic membrane. As a result, a method to overcome these low yields is required in order to produce enough protein for subsequent analysis and experimentation. The addition of a SUMO-tag to the N-terminus of the protein and expression in the cytoplasm can attempt to combat low yields, and the protein can be easily purified by a nickel affinity column with the use of the His-tag found within the SUMO-tag. The **S**mall **U**biquitin-like **Mo**difier (**SUMO**) is a protein which can be reversibly attached to any protein to increase its stability or solubility. In this case, the ability for it to be removed by SUMO-protease and leave the N-terminus of the protein intact is very beneficial.[113]

In this chapter, the expression of *Ba*AA10 using both periplasmic expression and SUMO-tag expression systems are discussed. The cloning of the *Ba*AA10 sequence with the SUMO-tag attached is also included, and the subsequent purifications using each method are given. The relative yields for both preparation methods of *Ba*AA10 are provided, which show the benefit of developing the expression and purification method used to obtain increased production of *Ba*AA10.

2.3 Materials and methods

2.3.1 Preparation of *Ba*AA10 using a periplasmic expression system

A pET-11a vector harbouring the codon-optimised *Ba*AA10 Optical Reading Frame (ORF) (Appendix 1) had previously been cloned and transformed into BL21 (DE3) *E.coli* cells by PDRA Glyn Hemsworth. The cells were cultured in 30 ml LB overnight at 37 °C, shaking at 180 rpm. 5 ml aliquots of these cells were then used to inoculate six 500 ml LB cultures. These were incubated at 37 °C, with aeration at 180 rpm until an OD_{600} of 0.4 was reached. At this point the temperature was reduced to 16 °C, and the cells were allowed to continue growing until the OD_{600} was 0.6-0.8. Production of *Ba*AA10 in the cells was then induced by adding IPTG to a concentration of 1 mM, and cultures were left overnight before being harvested by centrifugation at 10 000 g for 30 minutes. The cells were re-suspended in 60 ml buffer (50 mM Tris pH 8.0; 200 mM NaCl; 20 % w/v sucrose) and left on ice. After 30 minutes, the cells were centrifuged at 11 000 g for 10 minutes. The supernatant was kept and the pellets were re-suspended in two volumes of ice cold milli-Q water and kept on ice for a further 30 minutes. Cell debris was removed by centrifugation at 11 000 g for 10 minutes again to obtain the osmotic shock solution. This was combined with the sucrose supernatant solution and diluted to five times its volume (400 ml) with buffer A (50 mM Tris pH 8.0; 50 mM NaCl). The protein solution was loaded onto a 5 ml HiTrap Q-column using a peristaltic pump, collecting the flow-through which contains *Ba*AA10. The protein in the flow-through was then precipitated by the addition of ammonium sulfate to a final concentration of 3.5 M, and left stirring overnight at 4 °C. The protein was obtained by centrifugation of the solution at 38 000 g for 30 minutes, discarding the supernatant and resuspending the pellets in 20 ml gel-filtration (GF) buffer (20 mM NaOAc pH 5.0; 250 mM NaCl). This was then filtered using a 0.45 µm pore size and concentrated to less than 10 ml for loading onto a HiLoad 26/60 Superdex 75 column equilibrated in the GF buffer. The fractions corresponding to the pure protein were then combined and concentrated using a 10 kDa cut-off concentrator. The A_{280} was measured to determine the concentration of the protein for subsequent experiments, using 2.266 M^{-1} cm^{-1} as a molar extinction coefficient and 19821 Da as its molecular weight. Electrospray Ionisation - Time of Flight (ESI-TOF) Mass Spectrometry was carried out to confirm the mass of the *Ba*AA10.

2.3.2 Incorporation of a SUMO-tag via infusion cloning, for increased expression

In order to express *Ba*AA10 in the cytoplasm for increased yield of protein, a SUMO tag was cloned directly in front of the codon-optimised sequence for the *Ba*AA10 open reading frame (Appendix 1). This was cloned into the Champion pET-SUMO vector (Invitrogen), using In-fusion cloning (Clontech). The primers used in the PCR reaction were gaacagattggtggtCATGGCTACATCAAGGAACCGG and tacctaagcttgtctTTATTTCGTCAGATTCACGTCGATGAC, where lower case letters are the sequence for the overhangs required for inserting the sequence into the vector and upper case represents the complementary bases to the protein sequence. pET-SUMO(*Ba*AA10) was then transformed into BL21 (DE3) *E. coli* cells for overexpression.

2.3.3 Overexpression of *Ba*AA10-SUMO and purification of *Ba*AA10

Cells were cultured in LB media, together with kanamycin (30 μg/ml), at 37 °C before the temperature was reduced to 16 °C once an OD_{600} of 0.6 had been reached. Expression of *Ba*AA10 was induced when the OD_{600} reached 0.8, by the addition of IPTG to a final concentration of 1 mM. The cells were incubated at 16 °C overnight, with aeration at 180 rpm, before being harvested by centrifugation at 10 000 *g* for 30 minutes. The cell pellets were resuspended in 5x volumes of Buffer A (50 mM Tris pH 8; 200 mM NaCl; 30 mM Imidazole). Cells were subsequently lysed by sonication and the cell debris removed by centrifugation for 20 minutes in a Sorvall SS-34 rotor at 38 000 *g*. The resulting supernatant was then loaded onto a His-trap crude 5 ml Ni column (GE Healthcare) equilibrated in Buffer A. Following a 5 CV Buffer A wash of the column, a gradient from 0 to 100 % Buffer B (50 mM Tris pH 8; 200 mM NaCl; 300 mM Imidazole) was applied over 20 CVs. The fractions containing *Ba*AA10-SUMO were combined, concentrated and then diluted 10-fold with Buffer A to reduce the imidazole concentration to obtain a final concentration of protein of ~ 1 mg/ml. DTT was then added as a reducing agent to a concentration of 5 mM, along with 10 μg of SUMO protease per mg of *Ba*AA10-SUMO, in order to cleave the SUMO tag from the protein. The solution was then left shaking at 10 rpm at 20 °C overnight, before being passed through a 5 ml Ni (crude) column equilibrated in buffer A. The flow-through containing *Ba*AA10 was then collected and treated with 1 mM EDTA, to remove any metals present in the sample, before being concentrated to < 2 ml. The protein was then loaded onto a 16/60 superdex 75 (GE Healthcare) gel filtration column, which had been equilibrated with GF buffer (20 mM NaOAc pH 5; 250 mM NaCl). After a void volume of 40 ml, 1.6 ml fractions were collected and the peak fractions containing *Ba*AA10 were combined and concentrated to the required concentrations for subsequent experiments.

2.4 Results

2.4.1 Preparation of *Ba*AA10 using a periplasmic expression system

The gene coding for the *Ba*AA10 protein was successfully expressed in *E. coli* and the protein consequently purified following its secretion into the periplasm. The gel filtration trace shows the elution of *Ba*AA10 (Figure 2.4.1.1A). Fractions containing a high A_{280} response were run on an SDS-PAGE gel (Figure 2.4.1.1B), with the presence of protein bands at the expected molecular weight (19.8 kDa) being indicative of *Ba*AA10. Yields of approximately 3-4 mg per litre of culture were obtained using the periplasmic expression and purification methods.

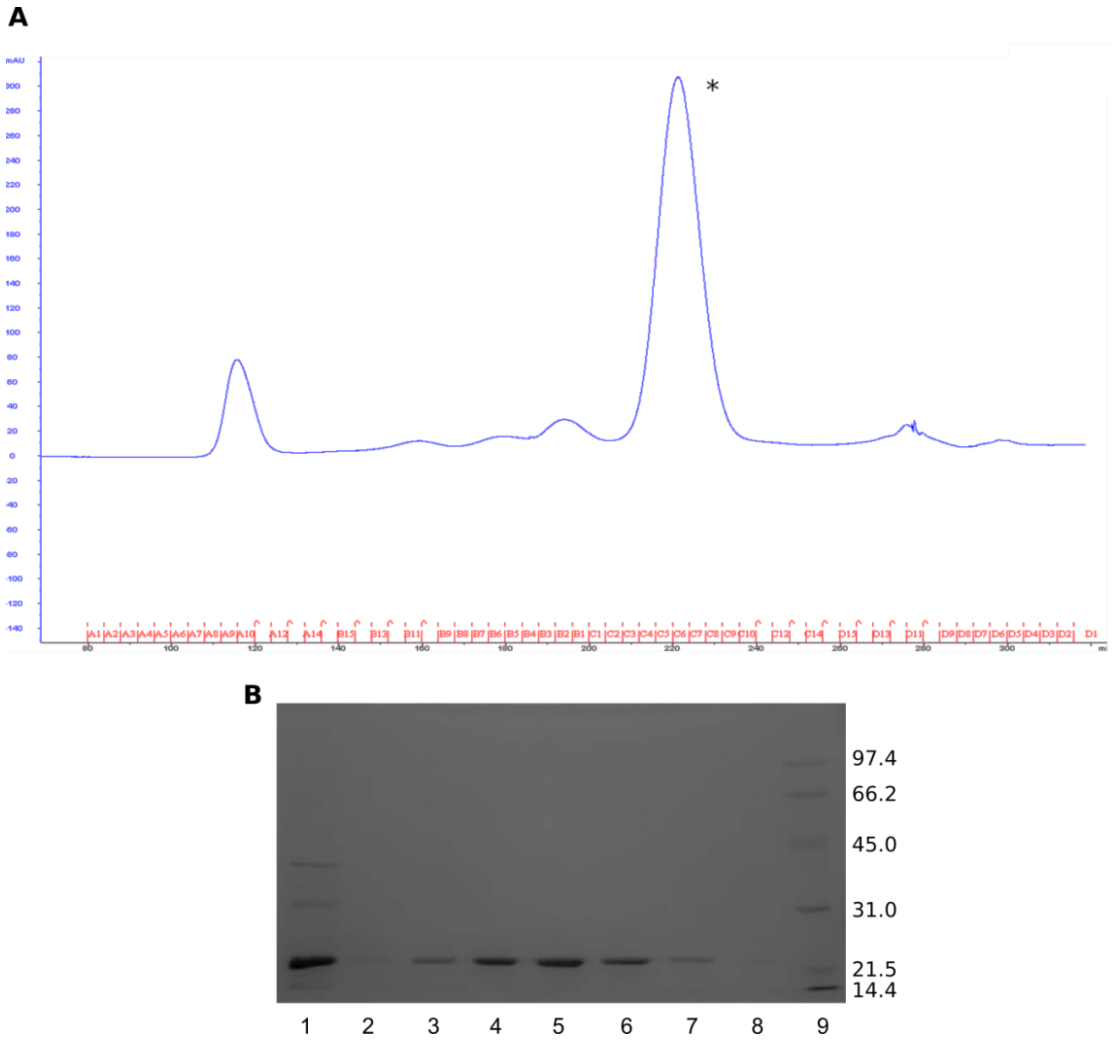

Figure 2.4.1.1 A) Gel filtration (26/60 S75) trace showing elution of the *Ba*AA10 protein after ~ 210 ml (indicated by the *). B) SDS-PAGE gel of purified *Ba*AA10 protein following gel filtration, where the protein loaded onto the column is in lane 1, lanes 2-8 are the fractions representing the peak indicated by the asterisk, and lane 9 is the low molecular weight marker (Biorad).

To prove further the purity of the *Ba*AA10 protein obtained, ESI-MS was carried out (Figure 2.4.1.2) which confirmed the presence of a single species at 19821.00 Da, corresponding to the molecular weight of *Ba*AA10 estimated using the online ExPASy ProtParam tool.[114]

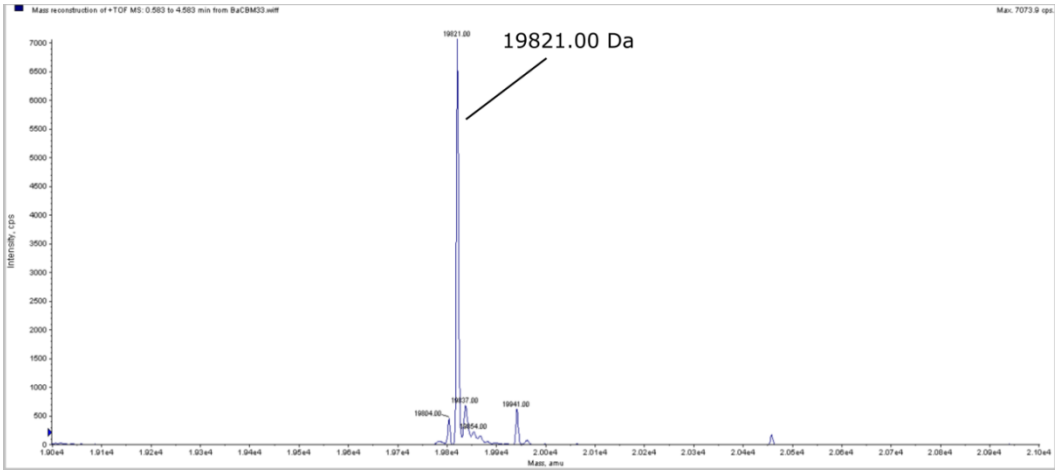

Figure 2.4.1.2 Electrospray Mass Spectrum (ESI-MS) of purified *Ba*AA10, with one main peak at 19821.00 Da, corresponding to the protein.

2.4.2 Infusion cloning of a SUMO-tag attached to the *Ba*AA10 protein sequence

The *Ba*AA10 gene with the SUMO tag attached was successfully cloned into the pET vector using In-Fusion cloning, resulting in a plasmid containing the sequence for *Ba*AA10 (Figure 2.4.2.1). Following transformation into competent Stellar cells and then colony PCR, successful expression into BL21 (DE3) *E. coli* cells was performed in order to allow expression of *Ba*AA10 on a larger scale.

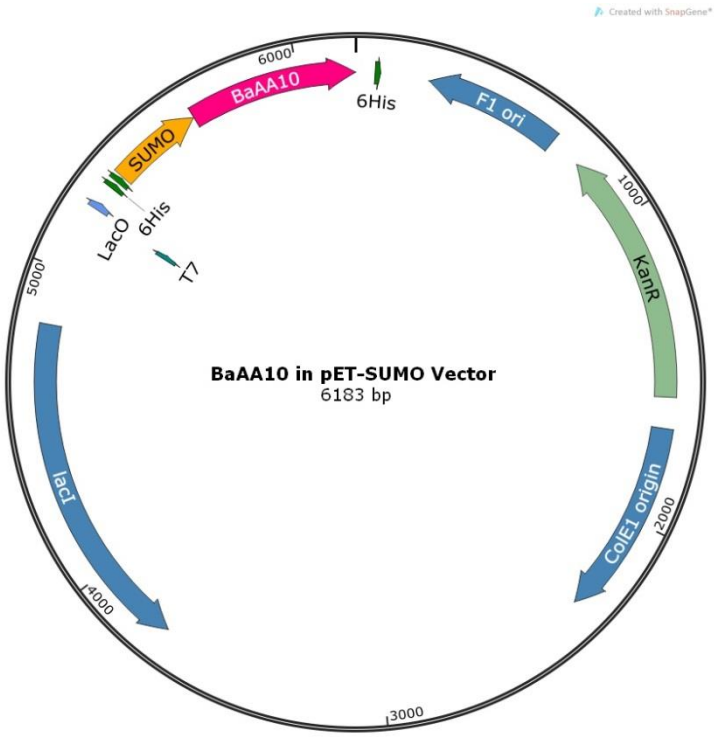

Figure 2.4.2.1 Representation of the pET vector harbouring the codon-optimised *Ba*AA10 sequence along with the SUMO tag situated at the N-terminus of *Ba*AA10. The figure shows the position of the His-tag within the SUMO tag for purification, as well as the kanamycin resistance gene for expression.

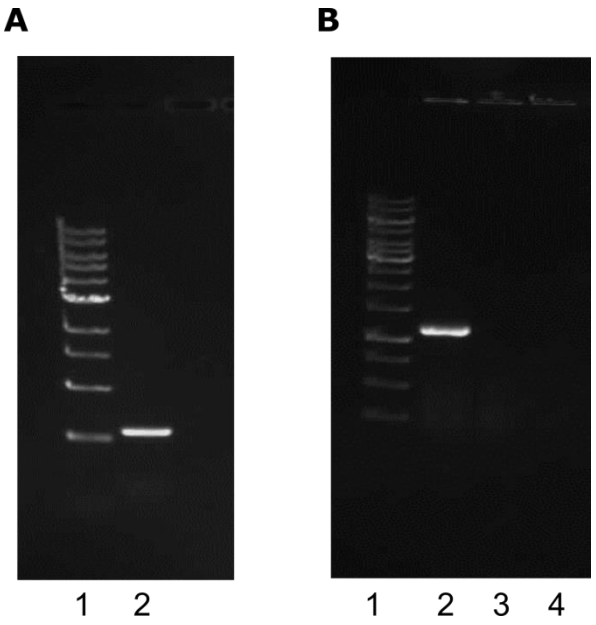

Figure 2.4.2.2 A) Agarose gel of the *Ba*AA10-SUMO insert that was used for In-Fusion cloning into the pET vector. Lane 1 contains the DNA marker and lane 2 is the amplified insert following PCR. B) An agarose gel following colony PCR to test three separate colonies for the pET vector containing the *Ba*AA10-SUMO insert after cloning by infusion. Lanes 2-4 each show a separate colony following colony PCR, indicating that the colony used for the sample in lane 2 is the only one that contains the vector. (Lane 1 is again the DNA marker.)

2.4.3 Overexpression of *Ba*AA10-SUMO and purification of *Ba*AA10

*Ba*AA10 with the SUMO tag attached was overexpressed and purified, using the SUMO tag's His-tag, by nickel affinity to obtain a yield of ~ 22 mg/L of cell culture. The trace and gel showing the pure *Ba*AA10-SUMO protein can be seen in Figure 2.4.3.1. Following SUMO-tag cleavage by SUMO protease and a second nickel affinity column to obtain the cleaved product, the *Ba*AA10 protein was further purified using gel filtration chromatography to obtain a final yield of ~ 9 mg/L of cell culture. This is a 3-fold increase in yield of pure *Ba*AA10 compared to previous yields obtained via the periplasmic expression system (which were ~ 3-4 mg/L of cell culture). The gel filtration trace and gel of the main peak showing the pure *Ba*AA10 protein can be seen in Figure 2.4.3.2, whilst a full gel of each stage of the purification process can be seen in Figure 2.4.3.3.

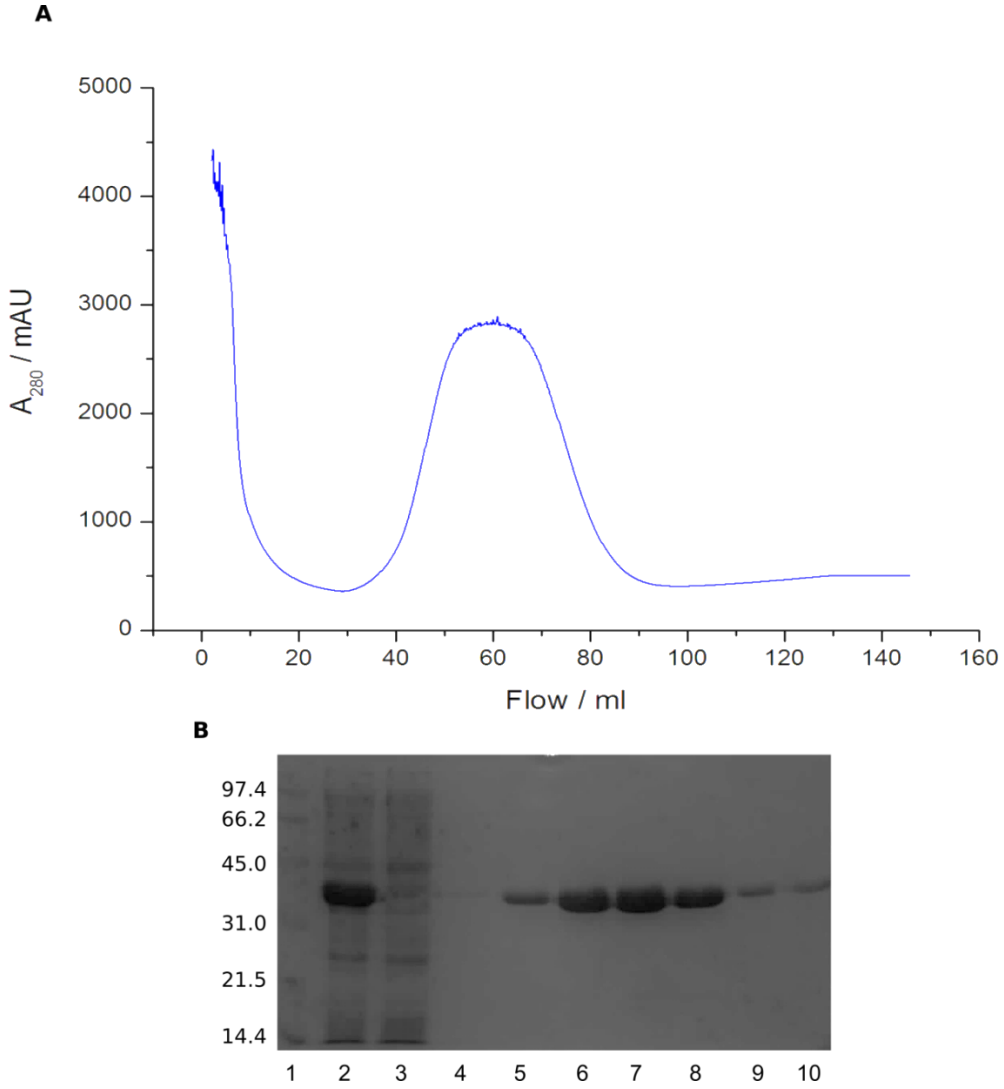

Figure 2.4.3.1 A) Chromatogram showing the elution of the *Ba*AA10-SUMO protein from a nickel column, measured by its UV absorption at 260 nm (A_{280}). B) SDS-PAGE gel of fractions of the peak from ~ 40 – 80 ml on the chromatogram in (A); alternate fractions were taken throughout the peak (lanes 4 – 10). Lane 1 is a low molecular weight marker (Biorad), lane 2 is a sample of the protein that was loaded onto the nickel column, and lane 3 is a sample of the flow-through solution from the nickel column.

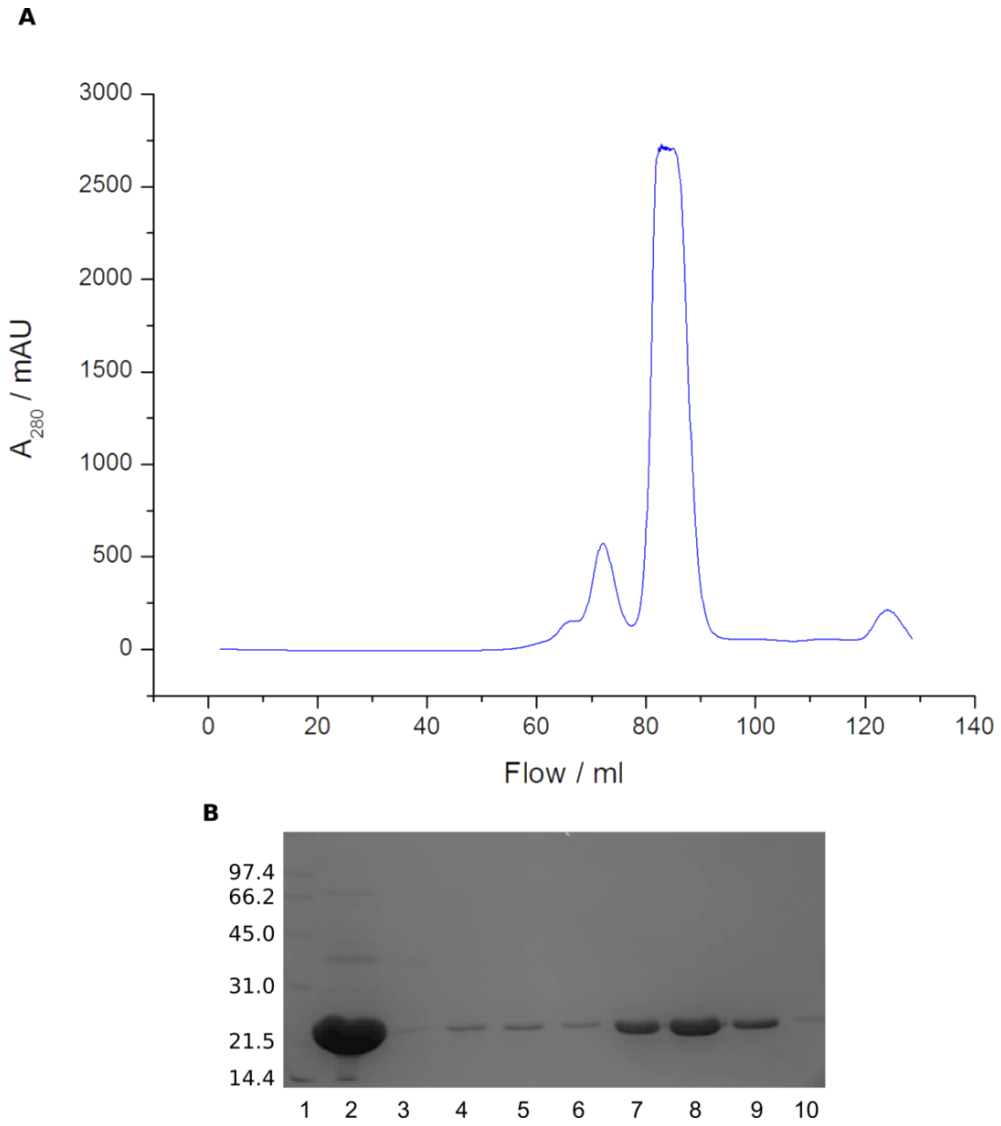

Figure 2.4.3.2 (A) Chromatogram showing the elution of *Ba*AA10 from a 16/60 S75 gel filtration column, measured by its UV absorption at 280 nm (A_{280}). B) SDS-PAGE gel of the fractions following the gel filtration, corresponding to the main peak at ~80 – 90 ml (lanes 3 – 10). From the marker in lane 1, we can see that this protein is in fact *Ba*AA10, which has been purified to remove all traces of other proteins seen in the load sample in lane 2.

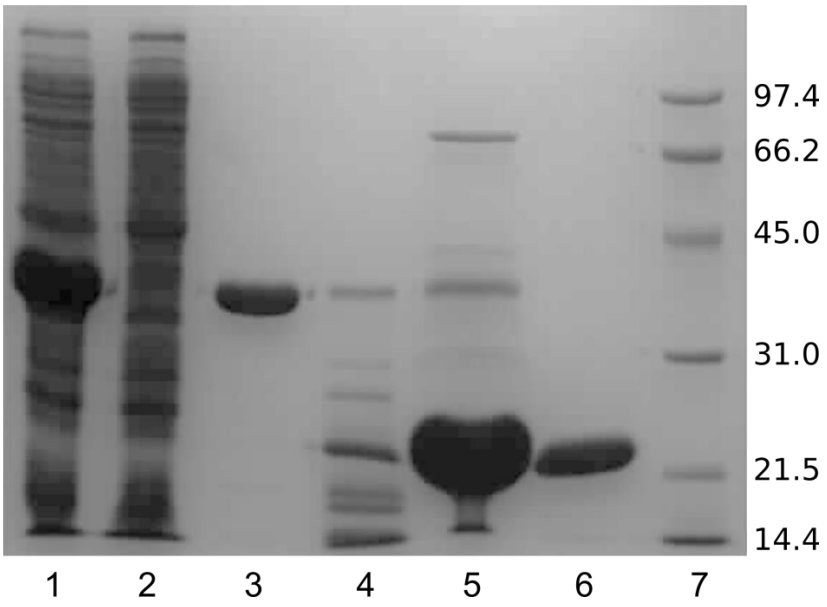

Figure 2.4.3.3 SDS-PAGE gel showing the purification process of *Ba*AA10. Lane 1 – Expression of *Ba*AA10-SUMO, prior to loading onto the nickel affinity column. Lane 2 – Flow-through of nickel affinity column. Lane 3 – Sample of *Ba*AA10-SUMO following purification by nickel affinity column. Lane 4 – A sample of the cleaved *Ba*AA10, following treatment with SUMO protease, prior to loading onto the second nickel affinity column for isolation of *Ba*AA10. Lane 5 – Concentrated *Ba*AA10 protein loaded onto the gel filtration column for purification. Lane 6 – Pure *Ba*AA10 following purification by gel filtration. Lane 7 – A low molecular weight marker (Biorad) for comparison of the molecular weights of the various proteins present at the different stages of purification.

2.5 Discussion and Conclusion

The lytic polysaccharide monooxygenase from the *Bacillus amyloliquefaciens* bacterium is part of the AA10 family, and has been successfully expressed in *E. coli* and purified for structural and functional analysis. At the beginning of this work, the method of protein production involved secretion into the periplasm for subsequent purification (described in this chapter). This method of purification allowed the LPMO to be produced, with its N-terminus intact, in reasonable quantities for further analysis. However, for efficient protein production, a yield of ~ 3-4 mg per litre of cell culture was deemed to be insufficient for further analysis, especially for chemical methods such as EPR, which often require more material than biophysical techniques. Therefore, a procedure was developed which should allow production of *Ba*AA10 at greater yields, whilst still producing the protein with its N-terminus intact, so that the copper-active site situated at the N-terminus of the protein could be synthesised correctly. This new procedure involved re-cloning the *Ba*AA10 sequence with a SUMO-tag attached into a new vector to be used for expression in *E. coli.*

Following successful cloning using the In-Fusion Cloning method, the pET vector was transformed into Stellar competent cells and colonies tested for the vector, before being transformed into BL21 (DE3) *E. coli* cells for expression tests and scale-up. The purification of *Ba*AA10 with the SUMO tag attached was now simpler than using periplasmic expression and purification methods, due to the presence of the His-tag attached to the SUMO-tag, enabling the protein to be purified by nickel-affinity and gel filtration methods. Using these techniques, a yield of ~9 mg per litre of cell culture was obtained, which is a significant increase on the previous method, which used periplasmic expression and purification techniques. Therefore, the *Ba*AA10 protein could now be produced more efficiently and at a greater yield following each preparation. This allowed more experiments to be carried out with fresh protein, and contributed to more results of various analysis experiments to determine the structure and function of *Ba*AA10, which are described in the subsequent chapters.

3. Copper binding to *Ba*AA10 and its impact on protein stability[1]

3.1 Abstract

During the discovery of the lytic polysaccharide monooxygenase enzymes and their role within the degradation of polysaccharides for biofuel production, the identity of the metal ion found in the site of what is now known as the 'histidine brace' became an important question. Many structures solved before 2011, and the seminal 2010 *Science* paper reporting the oxygenase activity suggested metals such as nickel, calcium, magnesium, sodium, zinc, manganese as the metal in question. In 2011 Quinlan *et al.* in *PNAS* reported that copper was the correct metal involved in the redox activity of LPMOs (in this case the AA9/GH61 enzyme family). During my PhD we sought to perform similar studies identifying the metal-ion in AA10 (formally CBM33) enzymes. In this chapter I report the binding of Cu^{2+} to BaAA10 using isothermal titration calorimetry enabled by the high level expression described in the previous chapter. Ultimately, the tight binding was addressed through the development of a displacement ITC method using Zn^{2+} prior to Cu titration. These studies were complemented using differential scanning fluorimetry to look at the increased stability of the enzyme upon Cu binding. I show that copper binds to the *Ba*AA10 enzyme with a binding constant of 43 nM, increasing its stability by approximately 22 °C. Issues with the stoichiometry of copper ions bound to protein molecules in the ITC experiments are also discussed.

[1] Part of this work has been published in Hemsworth et al (2013) The Copper Active Site of CBM33 Polysaccharide Oxygenases *J. Am. Chem. Soc.*, 2013, **135**, 6069–6077, a copy of which is attached in Appendix 2.

3.2 Introduction

In the first structure of an LPMO, then believed to be a chitin binding protein (CBP21) published in the academic literature by Vaaje-Kolstad *et al.*,[41] a metal site on the protein was identified and Na^{+} was modelled into the structure (PDB Code: 2BEM). At that point it was not known that these enzymes were oxygenases and there was no reason to believe that the metal played a role in the activity of the protein at the time. When, subsequently, the first reports were published by the same group in 2010 which showed that bacterial CBM33 proteins (now known as LPMOs from family AA10) were in fact oxygenases capable of polysaccharide degradation in an oxidative catalytic fashion,[34] and that these enzymes were structurally homologous to the GH61/AA9 enzymes published previously,[35, 39] the identity of the metal ion became much more relevant, albeit much more confusing. Indeed in their Science paper, Vaaje-Kolstad *et al.* published the same AA10 from *Serratia marcescens* having activity on chitin with both Mg^{2+} and Zn^{2+} bound.[34] The different AA9 structures which had previously been obtained by Karkehabadi *et al.*[35] in 2008 and Harris *et al.*[39] in 2010 also contained different metal ions in the active site coordinating to the two His residues, including Ni^{2+}, Mg^{2+} and Zn^{2+}. As Karkehabadi *et al.* had used nickel to aid the crystallisation of the GH61 from *Hypocrea jecorina*, they modelled and refined this mononuclear metal as nickel, but the authors did speculate as to whether other metals such as calcium might also be able to bind to the protein.[35] At the time, it was obviously not thought that these metal ions had redox activity within the enzyme; however it was known that the removal of such ions by EDTA resulted in loss of catalytic activity, as shown in Harris *et al.*,[39] where they state that the presence of various divalent metal ions, such as Mg^{2+} and Zn^{2+}, increase the hydrolytic activity of the GH61 enzyme. It was additionally suggested that, in addition to a putative catalytic role, one possible role of the metal ion was to assist in the stability of the enzyme. Therefore, the work in this chapter will cover metal-ion binding and stability.

In 2011 Quinlan *et al.*[38] were the first to correctly assign copper as the metal ion bound in the active site of these metalloproteins. Of primary importance, instead of using a metal-ion "protocol", used by others,[34] in which the enzyme was incubated with *n* mM EDTA and then different metals added at concentrations > *n* mM metal, Quinlan and colleagues demetalated and purified the enzyme completely free from metals using a method of column chromatography involving a nickel stationary phase,[115] before performing the readdition/activity assays. In this way they showed that the enzymes were in fact solely Cu^{2+} dependent oxygenases, work subsequently supported by Phillips *et al.*[63] in their studies on the *Neurospora crassa* AA9, whereby they found a greater increase in activity of cellobiose dehydrogenase linked to the copper bound AA9 than with the zinc or apo enzyme. It therefore

appeared that the misleading observations in Vaaje-Kolstad *et al.*[34] stemmed from the co-incubation with EDTA and addition of different metals. Under these conditions Cu binding is so tight that whatever metal was added in excess it simply displaced sufficient Cu from the EDTA and allowed Cu rebinding to the enzyme, hence all metals appeared to activate activity. Furthermore, Quinlan *et al.*[38] used electron paramagnetic resonance (EPR) spectroscopy to help determine the coordination of the copper ion in what they termed the 'histidine brace' at the N-terminus of the AA9 LPMO from *Thermoascus aurantiacus*. Having determined the identity of the metal ion upon which these enzymes depend, the extent of copper binding and how it affects the proteins' stability were useful questions to be answered in order to learn more about these enzymes, and the role of copper within them.

All of the above work, in the Marletta and Davies/Walton groups, was performed on the AA9 (formally GH61) family of fungal enzymes. The metal-ion in AA10 enzymes remained to be established. This chapter will therefore address the thermodynamics of copper binding to the AA10 enzymes, specifically the AA10 protein from *Bacillus amyloliquefaciens*. Binding using isothermal titration calorimetry (ITC) is reported, as well as the stability gained by the copper within the protein using differential scanning fluorimetry (DSF). Some of these results were published towards the start of my research in *Journal of the American Chemical* Society[72] (see Bibliography/Appendix). Following on from this published data, certain questions arose regarding the stoichiometry of copper ion to protein, due to the published ITC data containing *n* values which never equated to 1 (expected value if one copper atom binds to one protein molecule, as stated in the paper and seen in structural information). These questions are important, as it could be possible for the protein to be a dimer requiring two protein atoms per copper ion, or the protein could contain more than one copper ion such as in the metalloprotein copper methane monooxygenase, which harbours a dimeric copper centre amongst a similar 'histidine brace' site.[90, 91] However, this was thought not to be the case for the LPMOs, as structural information from EPR and crystallographic data suggested otherwise.[38] Consequently, other biophysical techniques were undertaken, such as SEC-MALLS and differential scanning calorimetry (DSC), in order to provide further evidence for a 1:1 stoichiometry of copper within the lytic polysaccharide monooxygenases. These methods and their results are explained in detail in this chapter. The role of the metal on activity of the *Ba*AA10 enzyme on its polysaccharide substrate will be addressed in Chapter 4.

3.3 Materials and methods

3.3.1 Investigation of copper binding to *Ba*AA10 using Isothermal Titration Calorimetry (ITC)

Copper binding to *Ba*AA10 was measured using a VP-ITC (MicroCal). Experiments were carried out at both pH 5 (20 mM NaOAc; 250 mM NaCl) and pH 7 (20 mM Bis-Tris; 250 mM NaCl), with $CuCl_2$ used as the source of Cu^{2+} in each case. The concentration of *Ba*AA10 in the cell was between 30 to 60 µM, with the concentration of $CuCl_2$ in the syringe 10 times greater than that of the protein in the cell. An initial injection of 2 µl was titrated into the cell, before subsequent injections of 10 µl at 5 minute and 8 minute intervals for pH 5 and 7 respectively. The initial injection was discarded in the analysis of the titration data, which was fit using the Origin 7 software package (MicroCal). The protein in these cases was loaded into the cell using a plastic syringe to prevent any copper from the metal syringe binding to the protein prior to copper titration. In an attempt to obtain the minimum amount of contact with a metal-based cell cavity during the titration experiment, a run was carried out on an ITC_{200} (MicroCal), which has a smaller cell cavity. This experiment was carried out at pH 5 (20 mM NaOAc; 250 mM NaCl) with *Ba*AA10 at a concentration of 36 µM and a concentration of $CuCl_2$ of 360 µM. An initial injection of 1 µl was titrated into the cell, before 18 subsequent injections of 2 µl at 2 minute intervals. All experiments were carried out at 25 °C.

3.3.2 Displacement ITC using zinc, to measure the binding constant of copper to *Ba*AA10

In order to obtain an accurate Cu^{2+} binding constant of *Ba*AA10 at pH 5, displacement ITC[116-118] was carried out using Zn^{2+} as the competing metal. The experiments were performed using an ITC_{200} calorimeter (GE Healthcare). First, a simple Cu^{2+} binding experiment was performed at 298 K and pH 5, with the *Ba*AA10 concentration at 50 µM inside the cell and 500 µM $CuCl_2$ in the syringe. A Zn^{2+} binding experiment was then carried out under the same conditions, but with the *Ba*AA10 enzyme in the cell at a concentration of 200 µM, and 2 mM $ZnCl_2$ in the syringe. A greater concentration of enzyme was used for the Zn titration, as we knew from preliminary experiments[72] that Zn has much weaker binding, so greater concentrations were used to obtain substantial measurements to use in the displacement calculations. Fitting the data using the Origin 7 (MicroCal) software, the measurements generated were $N = 0.63$, $K = 1.23 \times 10^5$ M^{-1}, $\Delta H = 5424$ cal/mol and $\Delta S = 41.5$ cal/mol/K. Finally, the displacement experiment was carried out with 50 µM *Ba*AA10 and 2 mM $ZnCl_2$ present in the cell, and titrating in $CuCl_2$ at a

concentration of 500 µM. For each experiment, 38 injections of 1 µL were carried out at 2 min intervals, and the buffer used for all solutions was 20 mM NaOAc pH 5; 250 mM NaCl.

3.3.3 Resolving stoichiometry issues using SEC-MALLS

Size Exclusion Chromatography – Multi-Angle Laser Light Scattering (SEC-MALLS)[119, 120] was used to confirm the homogeneity, and oligomeric state +/- Cu^{2+} of the LPMO protein *Ba*AA10. Three samples were run on a S75 10/30 column: BSA (Bovine Serum Albumin) as a control protein, *Ba*AA10 with 10-fold concentration of EDTA added, and *Ba*AA10 with stoichiometric amounts of $CuCl_2$ added. 120 µl of each sample was used, and the concentration of *Ba*AA10 was 2 mg/ml. The size exclusion column was equilibrated with fresh buffer (20 mM NaOAc pH 5; 250 mM NaCl), which had been filtered through a 0.22 µm filter. The column was linked to a **Wyatt Dawn HELEOS-II** 18-angle light scattering detector and a **Wyatt Optilab rEX** refractive index monitor to determine the products obtained following chromatography. Analysis of the products was conducted using the Astra software on the Wyatt systems.

3.3.4 Investigating the effects of copper on the stability of *Ba*AA10

Thermostability studies[121] were carried out on an Agilent MX3000P qPCR machine. Copper ($CuCl_2$) and zinc ($ZnCl_2$) were added in stoichiometric amounts with the enzyme to measure any changes in protein stability with each of the metals present. A sample with EDTA in stoichiometric amounts was also used as a control. 15 µl of enzyme was combined with 15 µl of SYPRO orange (1000-fold diluted from stock) in qPCR tubes. The final concentration of *Ba*AA10 was 1.35 mg/ml (65 µM), and the buffer used was 20 mM sodium acetate pH 5, 250 mM NaCl. The fluorescence was measured whilst the temperature was increased from 25 to 96 °C, at 1 °C steps every 30 s. 517 and 585 nm were used as the excitation and emission wavelengths, respectively. The resulting melting temperatures (T_m) were calculated by fitting a sigmoidal curve to the data using the MTSA program for MATLAB.[122]

3.3.5 Resolving copper-binding issues using Differential Scanning Calorimetry (DSC)

Differential Scanning Calorimetry[123] experiments were performed on a VP-Capillary DSC machine at the Astbury Centre for Structural Molecular Biology at the University of Leeds.

Three samples were run at pH 5 (20 mM NaOAc; 250 mM NaCl buffer). These included 'apo'-*Ba*AA10 (as purified in Section 2.3.3, but used after 6 days in storage tubes with possible contact with copper), Cu-*Ba*AA10 (copper loaded with $CuCl_2$ in a stoichiometric amount), and EDTA-*Ba*AA10 (EDTA added at 10x the protein concentration, approx. 0.66 mM). For each of these samples, the concentration of the protein was 66 µM. A further experiment was carried out at pH 6 (20 mM MES pH 6; 250 mM NaCl buffer), at a concentration of 1.0 mg/ml in an attempt to prevent aggregation after unfolding. The sample run at pH 6 had been pre-treated with 10 mM EDTA before being run down a S75 16/60 size-exclusion column. All samples were degassed twice, with stirring, for 8 minutes to prevent any bubbles developing during the heating process. Fresh buffer was also degassed and used in the reference cell for each protein sample. Buffer-buffer runs were carried out before and after each set of experiments to enable background correction. Between protein runs, the sample cell was cleaned with 3 x H_2O, 2 x buffer, followed by 1 x degassed buffer to ensure complete removal of any aggregated protein from the cell. For each run the initial pressure inside the cell was recorded, as an initial pressure of ~ 29 Pa or above should indicate that there are no air bubbles present in the sample. A temperature cycle of 10 ° to 90 °C was set up running at 90 °C / hour, allowing for emptying, cleaning and refilling the cell to occur once it had cooled to 25 °C.

3.4 Results

3.4.1 Investigation of copper binding to *Ba*AA10 using Isothermal Titration Calorimetry (ITC)

Studies to determine the binding constant of copper to *Ba*AA10 were carried out by titrating the protein with $CuCl_2$ using a VP-ITC calorimeter. The result at pH 5 (Figure 3.4.1.1A) showed that Cu^{2+} binds so strongly to the *Ba*AA10 that it is not possible to calculate an accurate binding constant (K_D calculated to be 5.09 nM ± 3.41 $\times 10^{-5}$ nM). Figure 3.4.1.1 compares this ITC result to the experiment where a higher pH of 7 was used. The result obtained (3.4.1.1B) shows weaker binding to copper at higher pH. The weaker copper binding to *Ba*AA10 at the higher pHs could be due to the fact that there are more hydroxyl groups present in solution which compete with the enzyme to bind the copper. A representation of how pH can affect the copper site in the protein, and how this in turn affects its ability to bind copper, is shown in Figure 3.4.1.2. At pHs below that of the pKa of histidine (pH 6), the active site is likely to be protonated, so one would expect weaker binding at lower pHs. As seen from these ITC experiments, however, this is not the case, likely because hydroxyl ions present at higher pHs compete to bind with the copper, hence resulting in weaker copper binding at these higher pHs, such as pH 7.

The more sigmoidal curve at pH 7 (Figure 3.4.1.1B) calculates a binding constant of 80 nM with a stoichiometry, *n*, of 0.80. The stoichiometry calculated from Figure 3.4.1.1A was also similar, at $n = 0.75$. The low *n* values in these experiments (i.e. $n < 1$) are most likely due to the *Ba*AA10 protein readily picking up copper during the purification or ITC process, so it may already have copper bound to it. In an attempt to obtain a more accurate binding constant of copper to the protein, an ITC_{200} was used which has a smaller cell cavity, holding just 380 µl compared to the 1.8 ml of the VP-ITC machine. The result for this experiment carried out at pH 5 can be seen as C in Figure 3.4.1.1. Once again copper binding was extremely strong at this pH, resulting in an estimated K_D of 4.19 $\times 10^{-3}$ nM. However, the stoichiometry value for this experiment ($n = 0.86$) was slightly larger than that measured at the same pH (pH 5) using the VP-ITC, suggesting that the use of this smaller sample size may have been partially successful in limiting the amount of copper from the machine that the protein came into contact with.

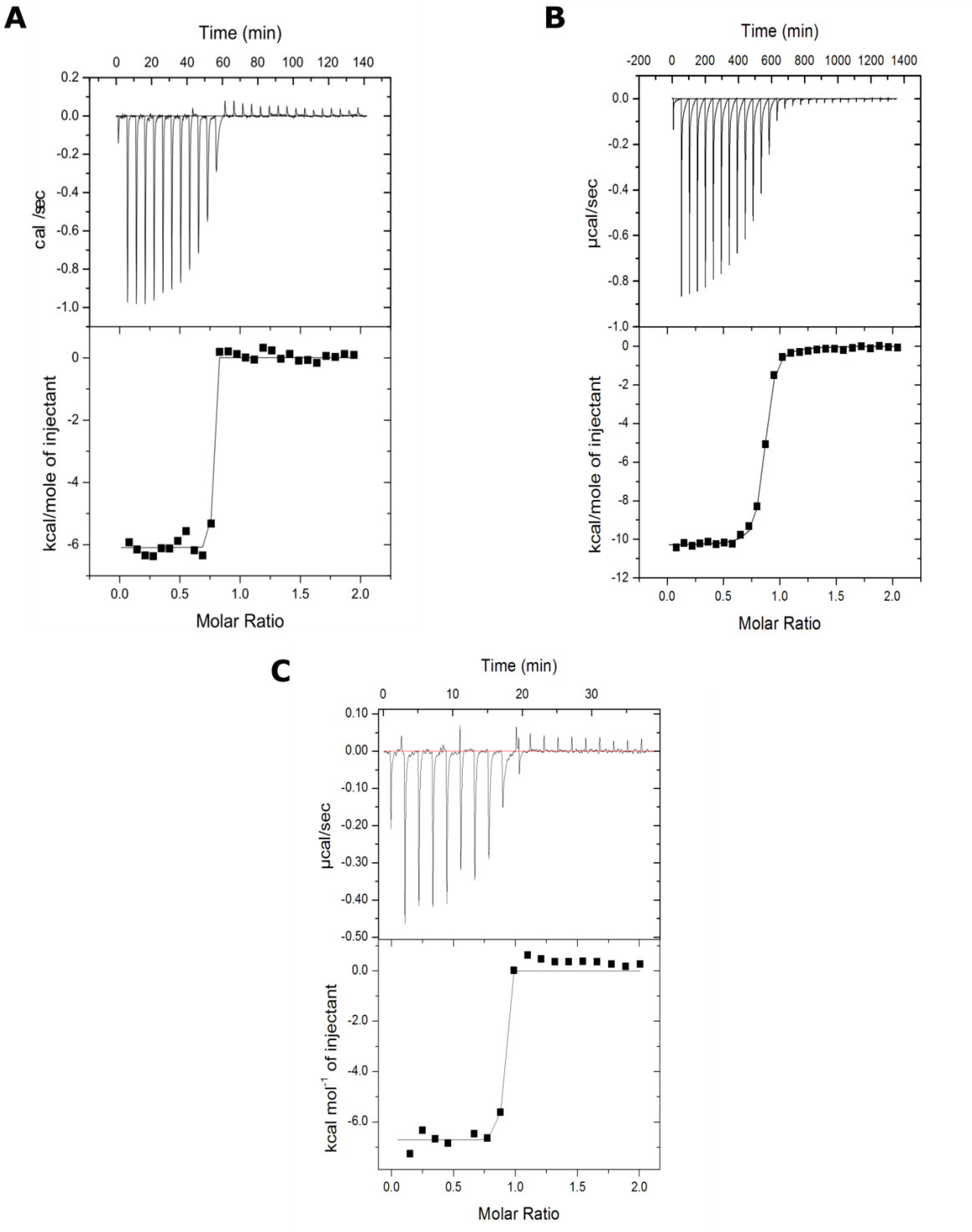

Figure 3.4.1.1 Isothermal titration plots of copper binding to *Ba*AA10 at pH 5 (A) and pH 7 (B) where both experiments were carried out on a VP-ITC machine, as well as the plot of copper binding at pH 5 (C) carried out on an ITC_{200} machine, which has a smaller cell cavity.

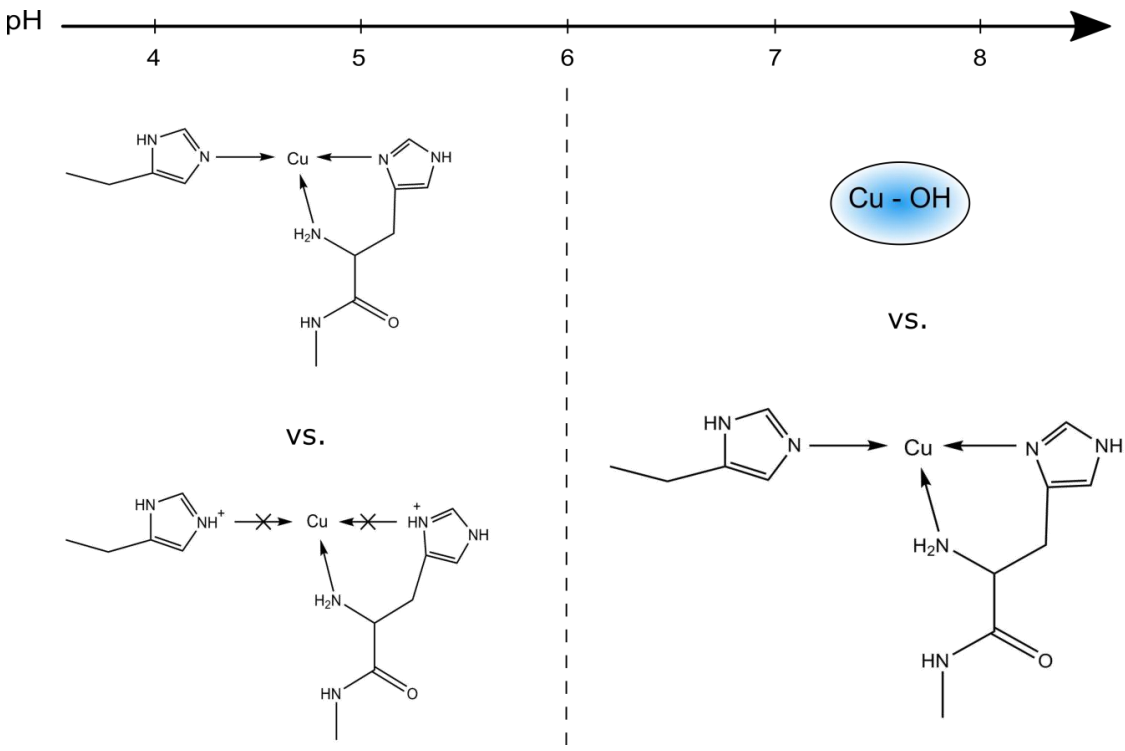

Figure 3.4.1.2 Representation of how the active site around the copper changes in the protein according to the pH and its environment. At low pHs (below pH 6, which is the pKa of histidine), the imidazoles within the copper site are likely to become protonated, theoretically preventing copper from binding as efficiently. However, at higher pHs (i.e. above pH 6), there are more hydroxide molecules in the protein environment competing for copper binding. Therefore, copper binding is actually, albeit counter intuitively, better at a lower pH such as pH 5, than at a higher pH of pH 7.

3.4.2 Displacement ITC using zinc, to measure the binding constant of copper to *Ba*AA10

Results of Cu^{2+} binding to *Ba*AA10 at pH 5.0 indicated extremely tight binding of the metal to the protein's active site, too tight to be measured accurately by normal ITC methods as there are insufficient experimental data to define the slope accurately, shown in Section 3.4.1 of this Chapter. Consequently, displacement ITC using a competing metal was used to accurately determine a binding constant. In this case Zn^{2+} was used, as it is known to be a more weakly binding metal to *Ba*AA10.[72] The binding constant (K_d) for Zn^{2+} at pH 5.0 was determined to be 8.13 µM. The competitive binding titration, displacing the Zn^{2+} with Cu^{2+}, produced results that could more accurately determine the K_d for Cu^{2+} binding of 43 nM ± 2 nM at pH 5.0, the titration curve of which can be seen as Figure 3.4.2.1. The χ^2 value of 8.78 for the displacement titration shows the reliability of the binding curve obtained.

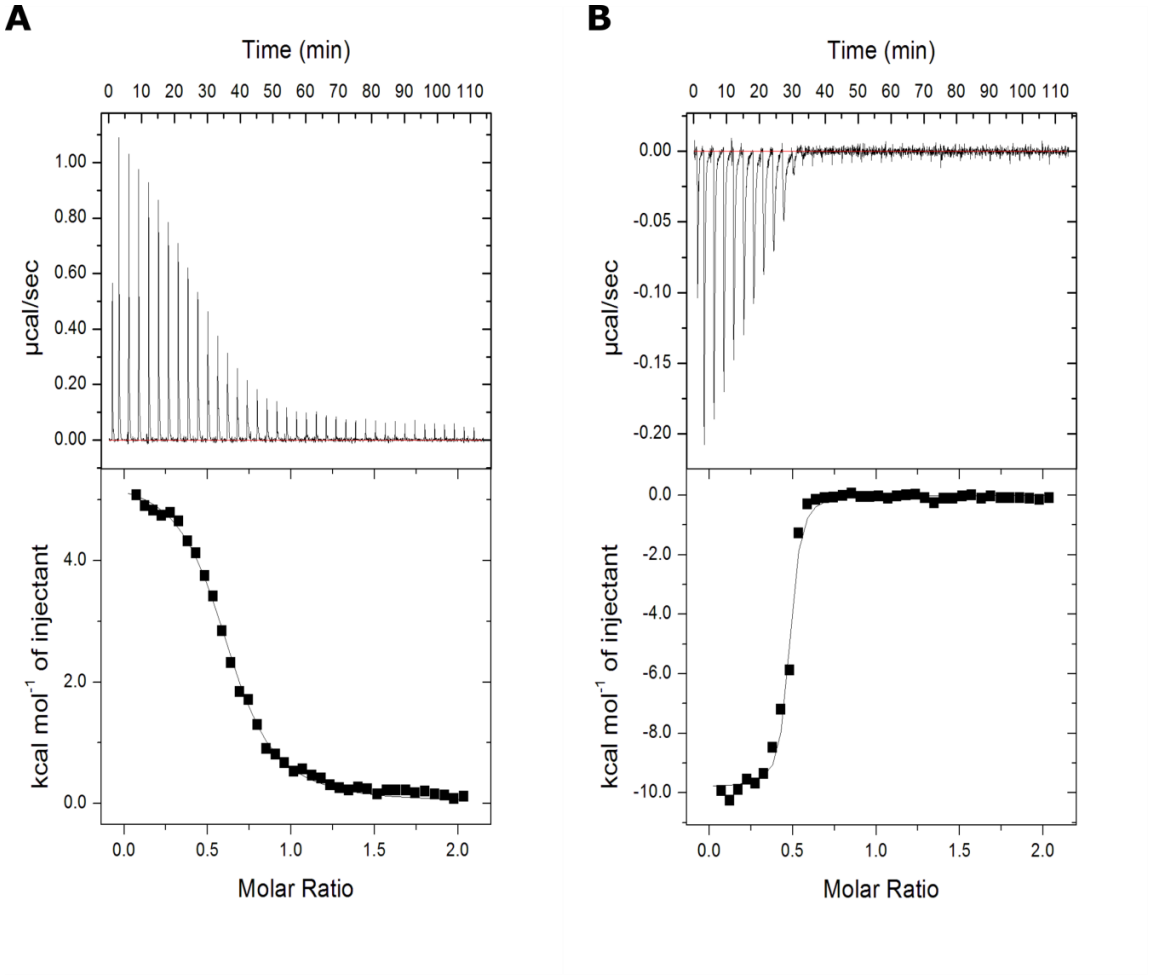

Figure 3.4.2.1 Isothermal titration calorimetry curves showing binding of Cu^{2+} to *Ba*AA10 by a displacement reaction with Zn^{2+}. (A) is the binding curve of Zn^{2+} to *Ba*AA10, and (B) is the displacement reaction, titrating Cu^{2+} into *Ba*AA10 with Zn^{2+} bound.

3.4.3 Resolving stoichiometry issues using SEC-MALLS

SEC-MALLS was run to establish the homogeneity, and oligomeric state, of the *Ba*AA10 protein with and without Cu^{2+}, as isothermal titration calorimetry (ITC) always produced a stoichiometry value of $n < 1$ for a monomer binding to one copper atom. We wanted to rule out the possibility that two molecules of *Ba*AA10 bound to a single Cu^{2+} ion (which would result in $n = 0.5$). ITC results have always given n values of 0.4 – 0.8. Therefore, SEC-MALLS was used to ensure that the protein is in fact a monomer and not a dimer, in both copper free and copper-loaded states. The results showed a single peak in both copper-free and copper-loaded samples with molecular weights (Mw) of 1.977 x 10^4 g/mol and 1.991 x 10^4 g/mol, respectively. These molecular masses agree with the proposed 1.9821 x 10^4 g/mol weight of the protein, allowing for errors. The plot of molecular mass against time produced from the light

scattering is shown in Figure 3.4.3.1, and shows the overlapping traces of Cu-*Ba*AA10 and EDTA-*Ba*AA10. In addition the polydispersity values for both of these sample peaks were both 1.001, showing that only one species is present in each of the peaks.

Therefore, SEC-MALLS showed that the *Ba*AA10 protein was a monomer in both copper-free and copper-loaded forms, and that the protein does not form a dimer when copper is added. There is a slight difference in the masses of the two samples, with the copper-loaded protein having a marginally larger molecular weight, and so eluting slightly earlier from the column. This could be due to a small change in configuration of the protein when copper is bound in the active site, perhaps changing the shape of the molecule. Possible reasons for sub-stoichiometric Cu titrations are discussed in 3.4.5.

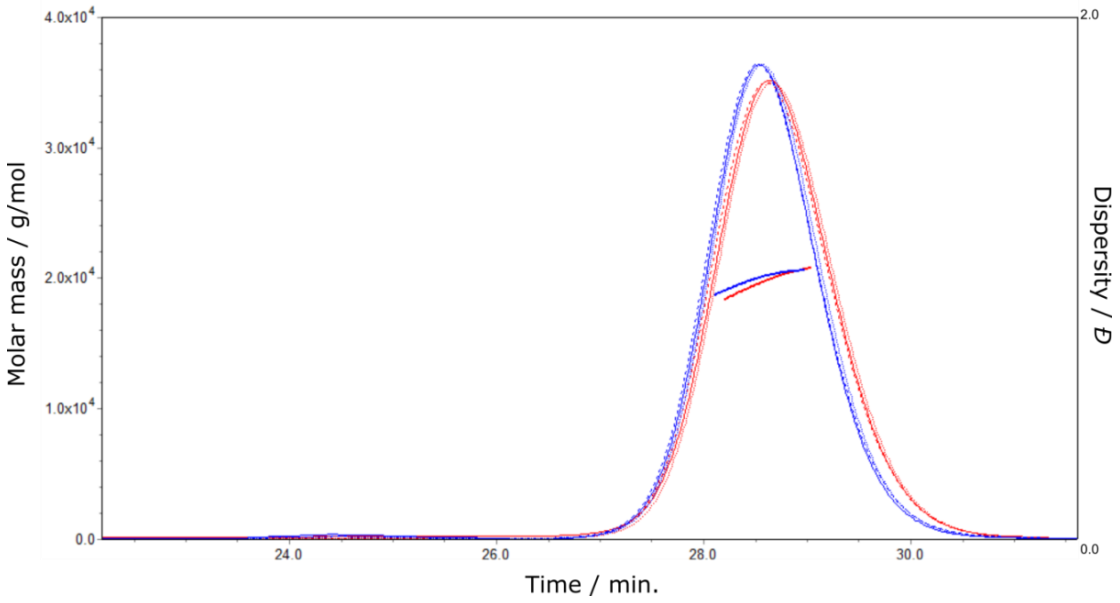

Figure 3.4.3.1 Size exclusion chromatogram of molecular mass vs. elution time for *Ba*AA10 with EDTA (red) and copper (blue). The dispersity of each sample is also shown (the 'straight' lines), which indicate that both samples are monomeric**.**

3.4.4 Investigating the effects of copper on the stability of *Ba*AA10

In order to determine the stability of the *Ba*AA10 protein in complex with a Cu(II) metal ion compared to copper-free and in complex with another metal (Zn(II) in this case), differential scanning fluorimetry (DSF)[121, 124] was used to obtain respective melting temperatures. SYPRO-orange dye was used as a source of fluorescence, which binds to the hydrophobic residues inside the protein during unfolding as the temperature increases. This provides a dissociation curve (Figure 3.4.4.1), from which corresponding melting temperatures for each sample, and hence an indication of the stability, can be obtained. The melting temperatures for each sample were calculated using a five-parameter equation written into the MTSA program[122] on MATLAB, and are shown in Table 3.4.4.1, along with their R^2 values of best fit. These R^2 values show a good fit of the dissociation curve to the data in all cases. In the sample containing copper, the fluorescence signal was significantly reduced, presumably due to quenching of the dye from the copper ions. However, a melting curve could still be obtained showing an increase in melting temperature from 43.6 °C for the *apo-Ba*AA10 to 46.4 °C and 65.2 °C with Zn^{2+} and Cu^{2+}, respectively, showing that there is a significant increase in the protein's stability when a metal ion is present. The data supports the calorimetric observation that the protein binds Cu^{2+} more strongly than other metals, with a considerable increase in melting temperature of 21.6 °C.

Table 3.4.4.1 Melting temperatures for *Ba*AA10 (labelled as 'apo'–*Ba*AA10 here, but was subjected to EDTA treatment during its purification process prior to the final gel filtration step (see Section 2.3.3)), as well as melting temperatures for the protein with EDTA, copper and zinc, together with their corresponding R^2 values.

Sample	T_m / °C	R^2
'apo'-*Ba*AA10	43.6	0.99997
*Ba*AA10 + EDTA	41.8	0.99994
*Ba*AA10 + Cu^{2+}	65.2	0.99466
*Ba*AA10 + Zn^{2+}	46.4	0.99973

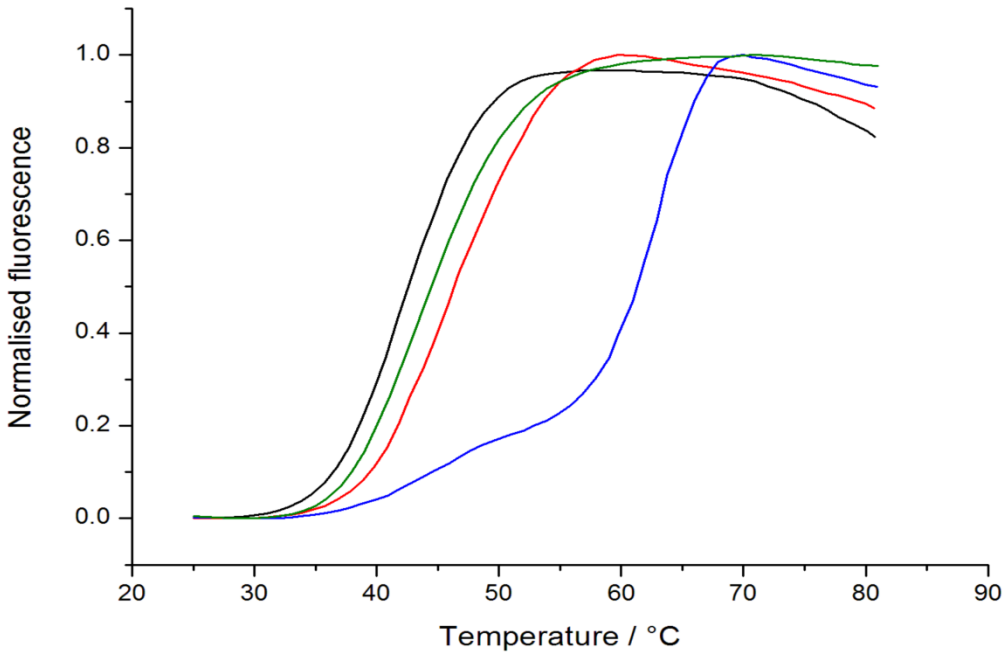

Figure 3.4.4.1 Dissociation curves to show the melting profiles of 'apo'-*Ba*AA10 (green), *Ba*AA10 with EDTA (black), *Ba*AA10 with zinc (red) and *Ba*AA10 with copper (blue). The original fluorescence for each has been normalised for comparison between the samples, and is plotted against the relative temperature.

3.4.5 Resolving copper-binding issues using Differential Scanning Calorimetry (DSC)

Given the sub-stoichiometric *n* values in ITC and the clean monomeric behaviour of *Ba*AA10, the inconsistency in binding stoichiometry was sought to be resolved. In another (commercial & unpublished with Novozymes, Gideon Davies *personal communication*) project Novozymes had shown that some AA9 enzymes could never be fully demetallated and that this was the cause of the sub-stoichiometric *n* values in ITC on those systems. With that impetus, differential scanning calorimetry (DSC) was carried out in order to see whether copper could be fully removed from the protein sample. Therefore, EDTA was added to the *Ba*AA10 protein as a chelating agent to remove copper and run alongside samples with copper added, as well as an 'apo' protein sample which had originally been purified with an EDTA step (as in Section 2.3.3) prior to being stored for 6 days when potential contact with copper could have occurred. The samples were initially run at pH 5, and the melting curves can be seen in Figure 3.4.5.1. To our surprise, both the 'apo' and copper-loaded samples gave almost identical unfolding curves with a melting temperature of 65.9 °C, whilst the protein sample with EDTA unfolded at a lower temperature of 49.9 °C. These melting temperature values are consistent with DSF (Differential Scanning Fluorimetry) results in the presence of Cu^{2+} reported in section 3.4.4 of this chapter.

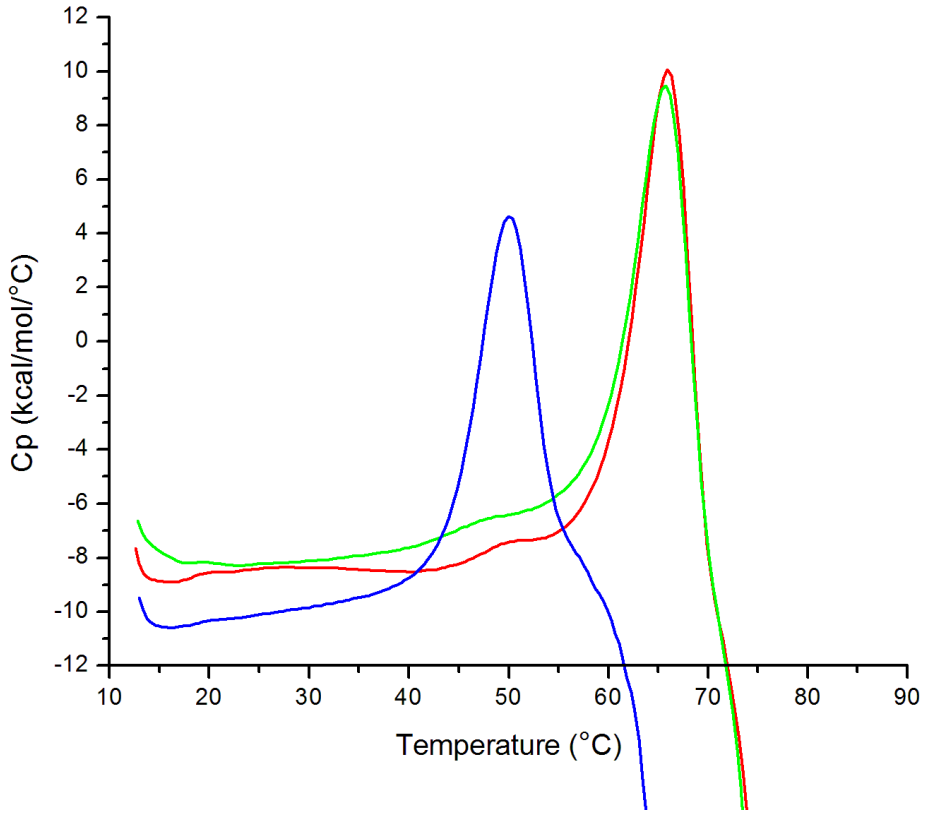

Figure 3.4.5.1 Differential Scanning Calorimetry (DSC) results for 'apo'-*Ba*AA10 (red), Cu-*Ba*AA10 (green) and EDTA-*Ba*AA10 (blue), where 'apo'-*Ba*AA10 refers to protein as it has been purified in Section 2.3.3 with an EDTA treatment step as part of the final stages of purification. All samples were run at pH 5, and the protein was at a concentration of 66 µM.

The crucial observation here, shown in panel A of Figure 3.4.5.2, is that the 'apo'-*Ba*AA10 is indeed partially copper-loaded, despite having been treated with EDTA during its original purification process, albeit roughly a week prior to carrying out this experiment. By integrating the peaks using Origin, the ratio of how much copper-free to copper-loaded protein can be estimated as 33 % and 67 %, respectively. By carrying out ITC on this same sample, we can then corroborate this ratio to that of the stoichiometry generated by the ITC experiment. The ITC result is shown as B in Figure 3.4.5.2, and this curve has an *n* value of only 0.379, giving a percentage ratio of 38 % to 62 %, supporting that of the DSC result. The non-unity stoichiometries frequently observed with some LPMOs therefore reflect partially loading of even protein purified with an EDTA treatment step prior to ITC.

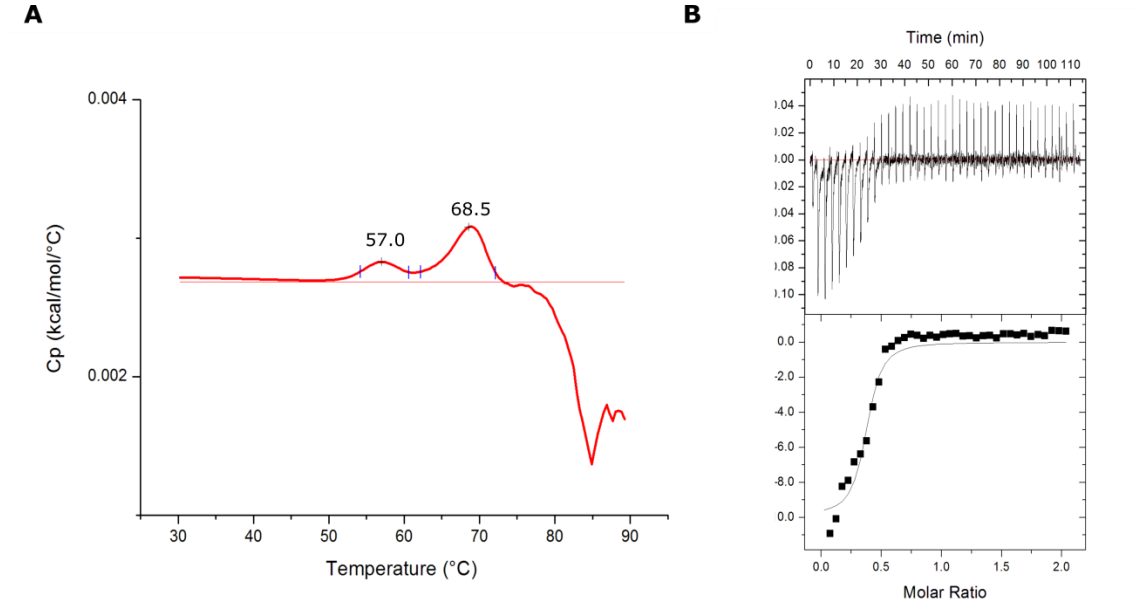

Figure 3.4.5.2 (A) Differential scanning calorimetry (DSC) plot of an 'apo'-*Ba*AA10 sample carried out at pH 6, which would have included an EDTA treatment step in the final stages of purification (see Section 2.3.3). Two peaks can be seen, which have been marked by the Origin software, and a baseline (straight red line) has been added so that the area under these peaks could be determined. (B) Isothermal titration calorimetry (ITC) plot of the same *Ba*AA10 sample used for the DSC in (A).

3.5 Discussion and Conclusion

Despite previous incorrect assignments of the metal ion involved at the active site of LPMOs[34, 35, 39] it is now clear, for both AA9[38, 63] and now AA10 (this work[108] and Hemsworth *et al.*[72]), that copper is the metal in question that binds at the N-terminus of the protein at the 'histidine-brace' site.[38] The thermodynamics of copper binding has been difficult to attain, due to the extremely strong binding nature of the copper to the protein, resulting in difficulties with removing all of the metal from the protein and in obtaining defined curves in ITC. Here I have shown that an accurate binding constant can be established at higher pHs (pH7) when hydroxyl ions compete for binding with copper, resulting in weaker copper binding to the protein. For accurate determination at lower pHs, displacement ITC can be used with zinc as a competing ion; the group has also used this approach on other LPMOs, such as for an AA11 from *Aspergillus oryzae*.[56] However, there are still issues with the stoichiometry of copper binding, with maximum *n* values of only 0.86 obtained and not 1.00, even using an ITC machine with a smaller cell cavity to limit the amount of metal the protein might come into contact with during the experiment setup and run. These issues have resulted in the use of other methods, such as SEC-MALLS to establish proof that the protein is indeed a monomer and only one copper ion binds to one protein molecule, and DSC to show that the *Ba*AA10 protein still contains some copper, even after being treated with EDTA beforehand in an attempt to remove all of the copper.

Research into the stability of the protein with copper present was investigated using DSF, a technique that can measure the protein's melting temperature. This showed that there is an extremely significant increase in *Ba*AA10 stability when copper is present, as a very large increase in melting temperature of 21.6 °C was obtained. These results show that copper binding is not only tight, but implicit in the stability of LPMO enzymes. With the establishment of Cu as the active centre metal, I was able to go on and study the activity of the enzyme. The specific activity of the *Ba*AA10 enzyme will be addressed in the subsequent chapter.

4. Determining the activity of *Ba*AA10, and the effect the substrate has on the stability of the protein

4.1 Abstract

Lytic polysaccharide monooxygenases have been proven to be effective in assisting in the degradation of various polysaccharides, in particular cellulose derived from plant cell walls and chitin from crustacean and insect shells. The activity of the AA10 enzyme from *Bacillus amyloliquefaciens* was yet to be determined; hence the investigation into the substrate targeted by *Ba*AA10 forms the basis of this chapter. Studies were carried out using various polysaccharide substrates including both α (anti-parallel chains) and β (parallel chains) forms of chitin – the most likely substrate based on sequence similarities with other AA10 enzymes, the genetic context of the enzyme in the host organism and the prior knowledge that the protein binds chitin. Following analysis of polysaccharide degradation products by Matrix Assisted Laser Desorption Ionisation – Time Of Flight (MALDI-TOF) Mass Spectrometry, chitin was confirmed as the substrate for *Ba*AA10 following the production of lactone and aldonic acid oxidation products. The presence of these products indicated that the LPMO carries out C1 oxidation of chitin, with aldonic acids of predominantly even numbered degrees of polymerisation obtained. Furthermore, Differential Scanning Fluorimetry (DSF) was used to show that the stability of the protein increases in the presence of this substrate. These findings will form the basis for further investigation into the chitin LPMO mechanism of action with knowledge of how the enzyme interacts with its substrate being the key to delineating the chemical events that take place at the active site. For example, activity results for *Ba*AA10 revealed the requirement for a crystalline substrate upon which the enzyme can perform C1 oxidation. However, it was shown that activity did not depend on the arrangement of the chitin chains within the crystalline structure, with oxidation occurring on both anti-parallel and parallel forms of chitin. Hopefully this knowledge will help us to assess the nature of the flat binding surface of *Ba*AA10, leading into a more in depth look at the structure of the enzyme, especially around the copper coordination site using X-ray crystallography, which will be described in the subsequent chapter.

4.2 Introduction

At the time of writing, it is well established that LPMOs are copper-dependent enzymes.[27, 38, 99] In order to define the molecular events that occur at the copper active centre and lead to oxygen activation and subsequent polysaccharide degradation, further knowledge is required on how these enzymes interact with their substrates. The recent publication of the structure of an AA9 LPMO with a substrate bound provides the first structural snapshots of the possible mechanism for activation of oxygen in the presence of substrate in AA9 enzymes.[71] Given the markedly different EPR spectra observed for chitin active AA10s,[72, 73, 85] similar knowledge of the interaction of chitin active AA10s with their substrates will be key to unravelling why chitin active AA10s display such dramatic differences in the copper coordination geometry. Research into substrate specificity for all AA10s has mainly focussed on sequence similarities and analysis of degradation products by High Performance Anion Exchange Chromatography (HPAEC), MALDI-TOF Mass Spectrometry or Size Exclusion Chromatography (SEC).[34, 36, 77]

Reports into the substrate specificities of AA9 and AA10 enzymes (then GH61 and CBM33), prior to my PhD research commencing, had revealed that all fungal AA9 enzymes were cellulose acting monooxygenases, and all but one of the AA10 enzymes were chitin degrading monooxygenases (Table 4.2.1).

Table 4.2.1 All known activities of LPMO enzymes, both AA9 and AA10 classes, up to 2011/12.

LPMO Class	Organism	Substrate	Reference
AA9 (GH61)	*Trichoderma reesi*	Cellulose	Harris *et al.*[39]
	Neurospora crassa	Cellulose	Phillips *et al.*,[63] Beeson *et al.*[64]
	Thermoascus aurantiacus	Cellulose	Quinlan *et al.*[38]
AA10 (CBM33)	*Serratia marcescens*	Chitin	Suzuki *et al.*,[83] Vaaje-Kolstad *et al.*[34, 41]
	Thermobifida fusca	Cellulose/Chitin	Moser *et al.*[87]
	Streptomyces coelicolor	Cellulose	Forsberg *et al.*[36]
	Enterococcus faecalis	Chitin	Vaaje-Kolstad *et al.*[77]

Consequently, oxidation at C1 has been detected for the AA10 enzymes,[34, 77] apart from the AA10s that degraded cellulose, which carried out both C1 and C4 oxidation processes.[36, 85] Analysis of degradation products of cellulose reveals reducing end lactone and aldonic acids from C1 oxidation,[34] whilst non-reducing end ketoaldose and gemdiol products are formed as a result of C4 oxidation (Figure 4.2.1),[66] as reviewed in Hemsworth *et al.*[62] In addition, most AA10 enzymes that show activity on chitin are able to show activity on both α and β forms of chitin,[34, 77] whereby α chitin denotes chains of *N*-acetyl glucosamine units running in an anti-parallel fashion, and those of β chitin running in a parallel fashion.[16] Regardless of the arrangement of the polysaccharide chains within chitin, a dominance of even-numbered products is often seen amongst the AA10 enzymes that have both chitin[34] and cellulose[36] as their substrate. This may provide us with evidence for the necessity of crystalline substrates for the AA10 enzymes which act on one side of the substrate.

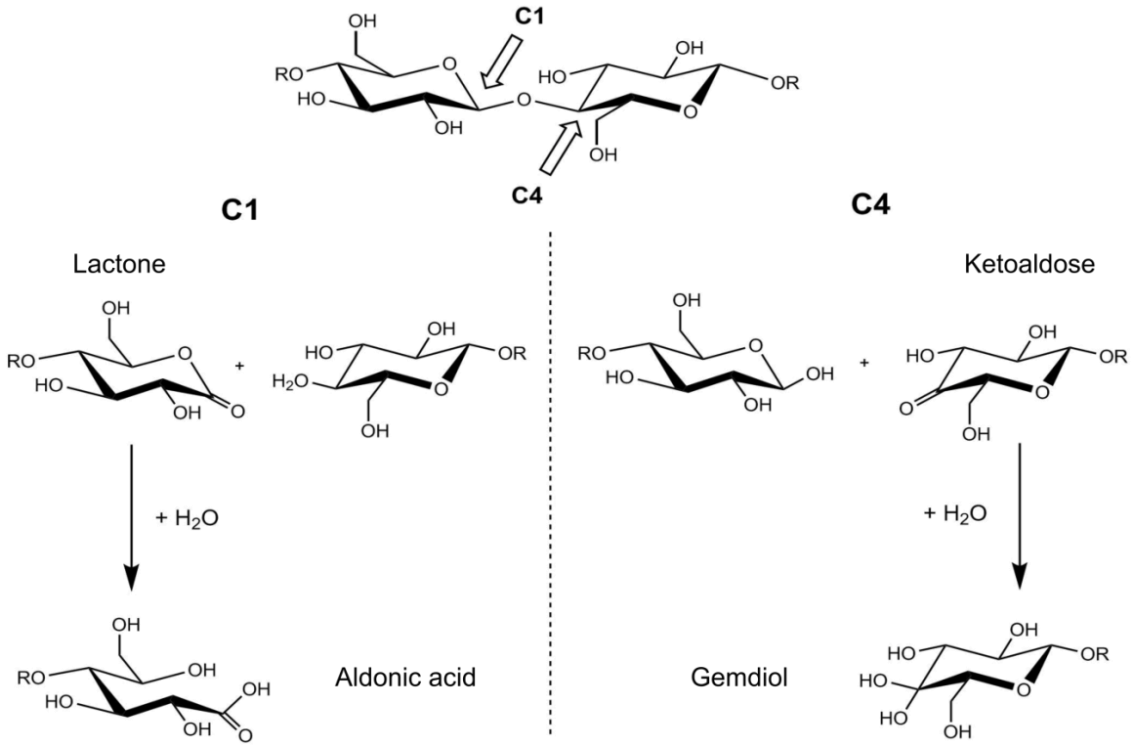

Figure 4.2.1 Reaction pathway for cellulose, demonstrating both the C1 and C4 oxidation pathways. The reaction for C1 oxidation generating a lactone product which may be hydrolysed to an aldonic acid product is shown on the left hand side of the figure, whilst the reaction for C4 oxidation producing a ketoaldose which can be hydrolysed to a gemdiol product is shown on the right hand side of the figure.

As the first AA10 activity was demonstrated on chitin,[34] and chitin binding had been observed some years before,[41] it was inferred that many AA10 enzymes would also be active on chitin due to their sequence similarities. Analysis of the AA10 sequence from *Bacillus amyloliquefaciens* in comparison to AA10 enzymes from other species shows a strong sequence similarity to other enzymes known to show activity on or binding to chitin (Figure 4.2.2), giving some indication as to the likely substrate for *Ba*AA10.

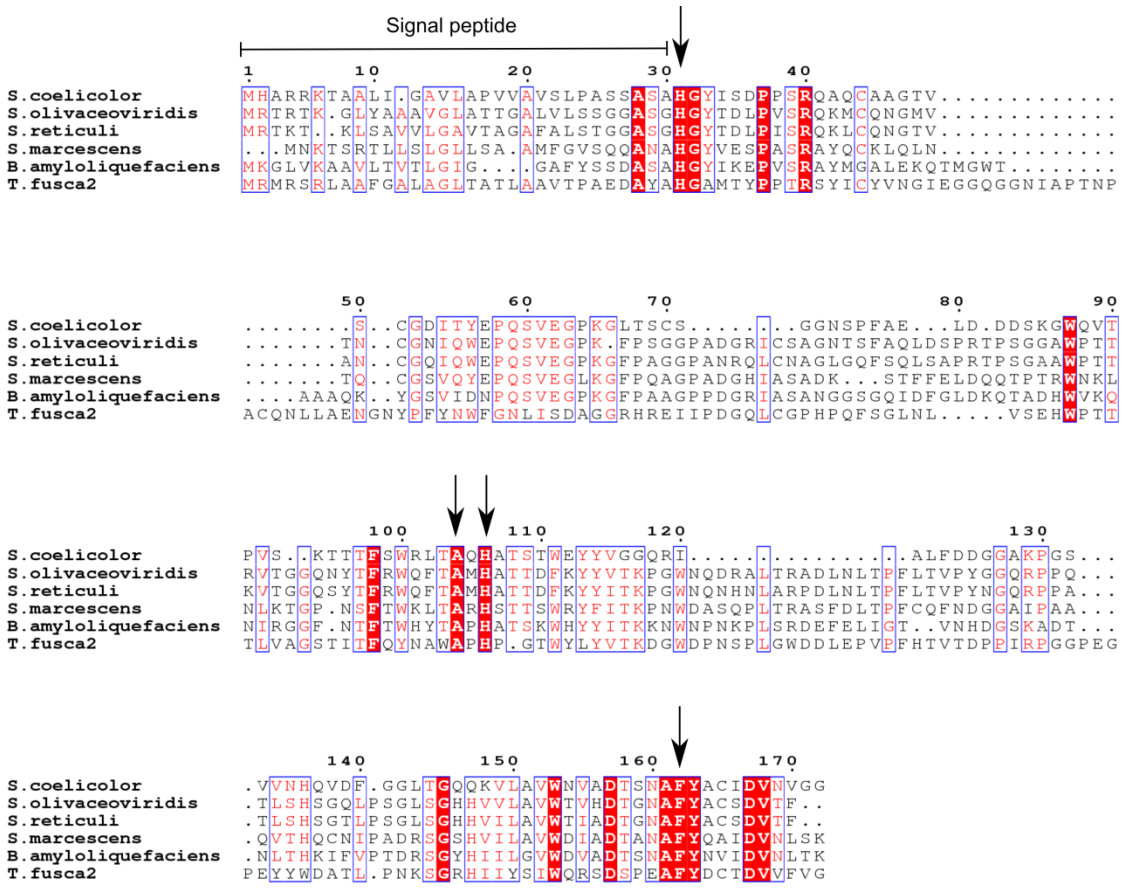

Figure 4.2.2 Sequence alignment of AA10 enzymes from various bacterial organisms compared to the sequence of the AA10 from *Bacillus amyloliquefaciens.* The alignment shows fully conserved amino acid residues (highlighted in red), as well as similarities (red text with blue outline) amongst AA10 proteins from *Streptomyces coelicolor* (NCBI: Q9Z9M5), *Streptomyces olivaceoviridis* (NCBI: Q54501), *Streptomyces reticuli* (NCBI: O87962), *Serratia marcescens* (NCBI: O83009), *Bacillus amyloliquefaciens* (NCBI: Q9F9Q5), and an AA10 protein from *Thermobifida fusca* (NCBI: Q47PB9). The signal peptide region for each of the sequences is labelled, as well as the highly conserved histidines, alanine and phenylalanine residues found within the copper active site of the AA10 LPMO enzymes, as seen in Figure 1.10.1. Alignment carried out with ClustalOmega[125, 126] and formatted using ESPript 3.0.[127]

In addition to the sequence similarities, a report published in 2001 by Chu *et al.*[112] showed by binding experiments using simple analysis of the products using SDS-PAGE, that the protein then known as ChbB from *Bacillus amyloliquefaciens* (the enzyme described in this thesis as *Ba*AA10) was able to bind both α and β forms of chitin. Consequently, sequence analysis and previous substrate binding knowledge indicate that *Ba*AA10 is likely to be a chitin active LPMO, but proof of this was lacking. In this chapter I will, therefore, describe a series of experiments analysing the substrate specificity of *Ba*AA10 by testing its activity on various polysaccharide substrates including α and β chitin from different sources. In addition, the stability of the protein when its substrate is present was investigated by measuring its melting temperature with the substrate using differential scanning fluorimetry (DSF).

4.3 Materials and methods

4.3.1 Activity experiments and analysis of degradation products by MALDI-TOF Mass Spectrometry

In order to determine whether *Ba*AA10 was active on chitin, MALDI-TOF Mass Spectrometry was used to analyse any degradation products resulting from enzyme activity. 1 ml reactions were set up consisting of 0.2 % w/v of the solid substrate under investigation, 1 mM ascorbic acid as reducing agent, 1 μM $CuCl_2$ and 1 μM *Ba*AA10. In order to try and remove any soluble fragments in the chitin samples, the substrates were washed with Milli-Q water and then with 10 mM ammonium acetate pH 5 buffer, before being filtered and dried under vacuum. Additionally, in an attempt to make the substrates more "amenable for catalysis", 60 mg of the chitin was suspended in 20 ml 0.2 M acetic acid. The suspension was then sonicated using a 3 mm probe five times for one minute with one minute intervals. It was then dialysed into 10 mM ammonium acetate buffer (pH 5). This process was an attempt to follow the method used to create 'chitin whiskers' produced by Vaaje-Kolstad *et al*.[34], based on a method by Fan *et al.*,[128] to endeavour to make the substrates more soluble for activity-based experiments, but was only carried out on shrimp and crab chitin. Reactions were also carried out using various other substrates to test for any additional activity, such as on cellulose (Avicel), starch, pectin, xylan, mannan, glucomannan, guar gum, locust bean gum and arabinoxylan. Control reactions were carried out without *Ba*AA10 for each new substrate used, as well as control reactions containing $CuCl_2$ and ascorbic acid to show any reactivity on the substrate is not a result of potential Fenton chemistry.[129] The reactions were carried out in 10 mM ammonium acetate pH 5, and were incubated for ~ 16 hours at 30 °C and rotated end-over-end to ensure the solid substrates were constantly being mixed and to guarantee suitable aeration for the reaction to proceed.

After incubation, the reactions were quenched with 1 μM EDTA before being centrifuged for 1.5 minutes at 13300 g. 1 μl of the supernatant was then spotted onto a MALDI Mass Spectrometry plate and mixed with 1 μl of a 10 mg/ml DHB (2,5-dihydroxybenzoic acid) matrix solution prepared in 50% acetonitrile, 0.1 % Trifluoroacetic acid. An ultraflex MALDI-TOF Mass Spectrometer was used to analyse the samples, using the Bruker flexControl software. CalMix1 was used to calibrate the spectrometer. Samples were run at 800 shots of 100 Hz laser frequency and a laser power of 40 mW. All spectra were processed and analysed using the Bruker flexAnalysis software

4.3.2 Investigating the effect of ligand on the stability of *Ba*AA10

Thermal shift analyses[121, 124] were carried out using Differential Scanning Fluorimetry (DSF) on an Agilent MX3000P qPCR machine to determine the melting temperatures of the *Ba*AA10 enzyme with and without squid chitin as a substrate. Cellulose (Avicel) and chitohexaose were also used as control substrates in the experiment. Copper ($CuCl_2$) was used in stoichiometric amounts with the enzyme, in order to detect any change in melting temperature upon the addition of chitin to Cu-*Ba*AA10. Solutions of 30 µl were made up in separate eppendorf tubes and 15 µl of each was taken, along with 15 µl of SYPRO orange dye (diluted 1000-fold from the stock solution) to make a final volume of 30 µl in the qPCR tubes. The buffer used was 20 mM sodium acetate pH 5 with 250 mM NaCl, and the final concentration of the protein was 1.35 mg/ml (~ 65 µM). The fluorescence was measured whilst the temperature was increased from 25 to 96 °C, at 1 °C steps every 30 s. 517 and 585 nm were used as the excitation and emission wavelengths, respectively. The resulting melting temperatures (T_m) were calculated by fitting a sigmoidal curve to the data using the MTSA program for MATLAB.[122]

4.4 Results

4.4.1 MALDI Mass Spectrometry results and analysis of degradation products

The activity of *Ba*AA10 was tested on a range of different polysaccharides, including α and β-chitin. Following incubation with the different substrates, MALDI-TOF Mass Spectrometry was carried out to determine the products of degradation. No degradation products were identified in the spectra produced from the cellulose (Avicel), starch, pectin, xylan, mannan, glucomannan, guar gum, locust bean gum or arabinoxylan reactions. The only positive results obtained from these reactions were those containing chitin, from both α (crab and shrimp) and β (squid pen) forms. Figure 4.4.1 shows the various spectra obtained for reactions with α chitin from crab, with and without enzyme (Figures 4.4.1.1C and 4.4.1.1A, respectively) as well as without ascorbic acid as a reducing agent (Figure 4.4.1.1B). The spectrum produced in Figure 4.4.1.1C shows oxidative degradation products, primarily aldonic acids obtained upon hydrolysis of the lactones formed following oxidation by *Ba*AA10. A schematic diagram of the breakdown of chitin and the pathway undertaken to obtain the various products obtained is shown in Figure 4.4.1.1D. The results show that degradation products of chitin by oxidation occurs only when both the AA10 enzyme and the reducing agent are present, as small base levels of chitooligosaccharides from non-oxidative degradation are observed in the control reactions, as they naturally occur in the substrate (Figure 4.4.1.1A and 4.4.1.1B).

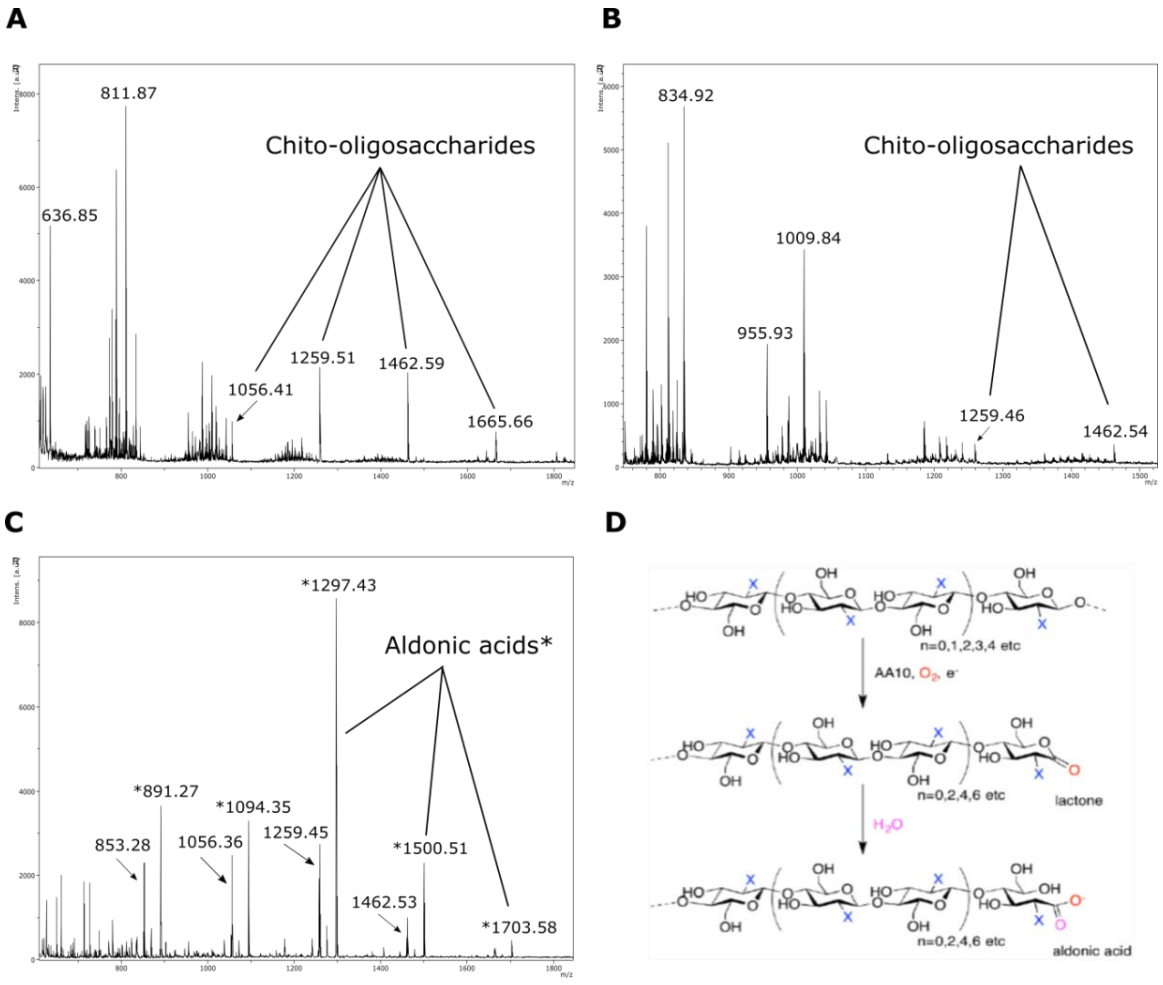

Figure 4.4.1.1 MALDI-TOF Mass Spectrometry results upon testing the activity of *Ba*AA10 with chitin from crab (α chitin). (A) shows the spectrum for chitin without *Ba*AA10, but with both copper and ascorbic acid to rule out any possibility of degradation by Fenton chemistry.[63, 129] (B) is the spectrum obtained for chitin with *Ba*AA10, but in the absence of ascorbic acid, and (C) is the spectrum obtained for chitin with *Ba*AA10 in the presence of ascorbic acid as a reducing agent. (D) is a schematic diagram of the degradation of chitin using an AA10, and the products that are obtained in the degradation process.

When *Ba*AA10 is present, as well as ascorbic acid as a reducing agent, chitin can be degraded into lactones before being converted to aldonic acids by the addition of water. Taking a six numbered product (i.e. the hexasaccharide, where degree of polymerisation, n, is 6), the molecular weights of the lactone product plus Na+ (1257 m/z) and aldonic acid products plus Na^+ (1275 m/z) and minus a hydrogen plus $2Na^+$ (1297 m/z) can be compared to that of the normal non-oxidised chitooligosaccharide with n = 6 (1259 m/z). These results clearly indicate oxidative activity on chitin has been carried out by *Ba*AA10.

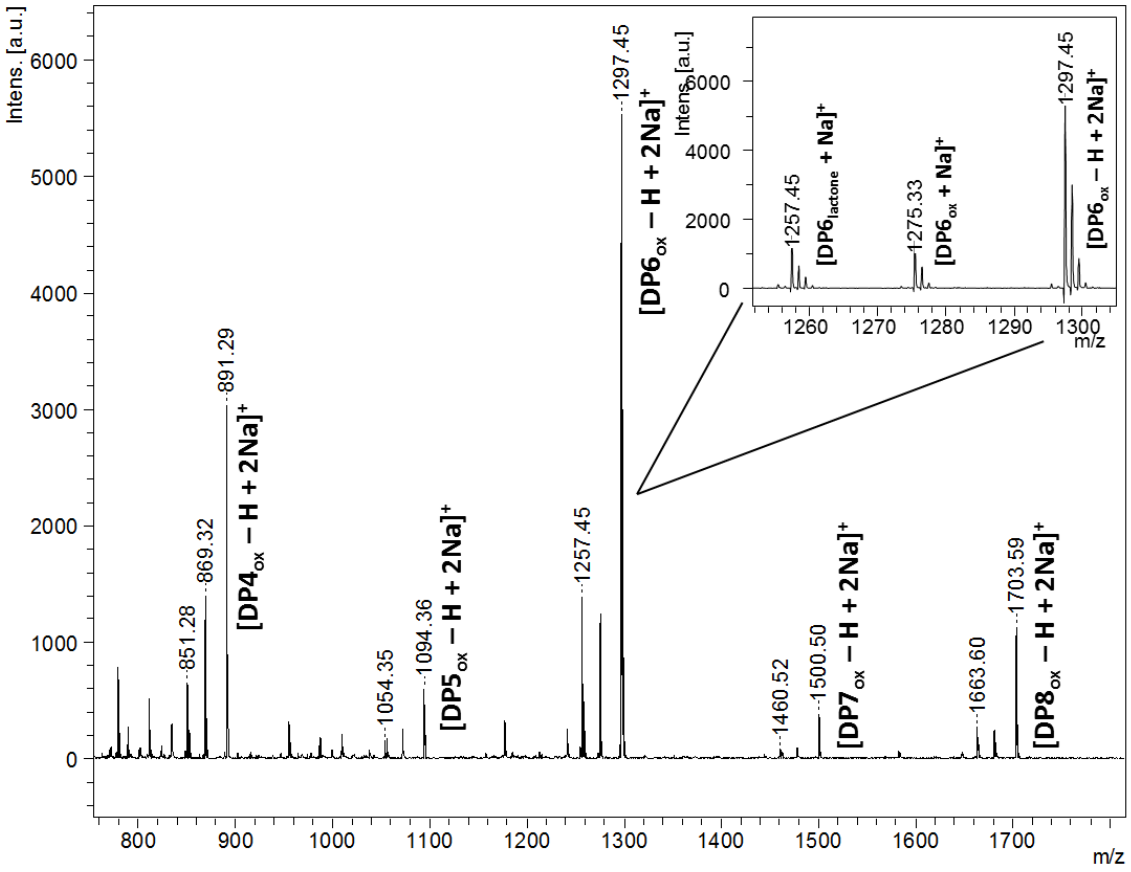

Figure 4.4.1.2 MALDI-TOF Mass Spectrum of the oxidative degradation products of chitin from squid pen (β chitin). Inset is a closer look at the oxidation products obtained with regards to the DP6 chitooligosaccharide (degree of polymerisation = 6).

The spectrum given in Figure 4.4.1.2 shows the analysis of degradation products of β chitin by MALDI-MS following incubation of *Ba*AA10 with squid pen chitin, the purest chitin substrate used (reflected simply in its white, as opposed to brown/yellow colour). Three main peaks can be seen for each degree of polymerisation, which most likely signify the lactone, the aldonic acid and the deprotonated aldonic acid, as a result of C1 oxidation (see schematic in Figure 4.4.1.1D). It must be noted that the peaks for the deprotonated aldonic acid are always stronger than the other two for each dp, possibly due to the amount of Na+ in the protein sample. The oxidative degradation products show a strong preference towards the even-numbered products i.e. dp = 4, 6, 8 over the odd-numbered products that are produced (dp = 5, 7). It is also worth noting the very large peak of the hexamer deprotonated aldonic acid at 1297.45 m/z, possibly displaying a preference by *Ba*AA10 to form the hexameric oxidation product, although one must remember that mass spectrometry is not a quantitative measurement. Furthermore, the lactone products obtained from oxidative degradation could also be assigned to C4 oxidation, further analysis of which cannot be carried out by MALDI-TOF Mass Spectrometry.

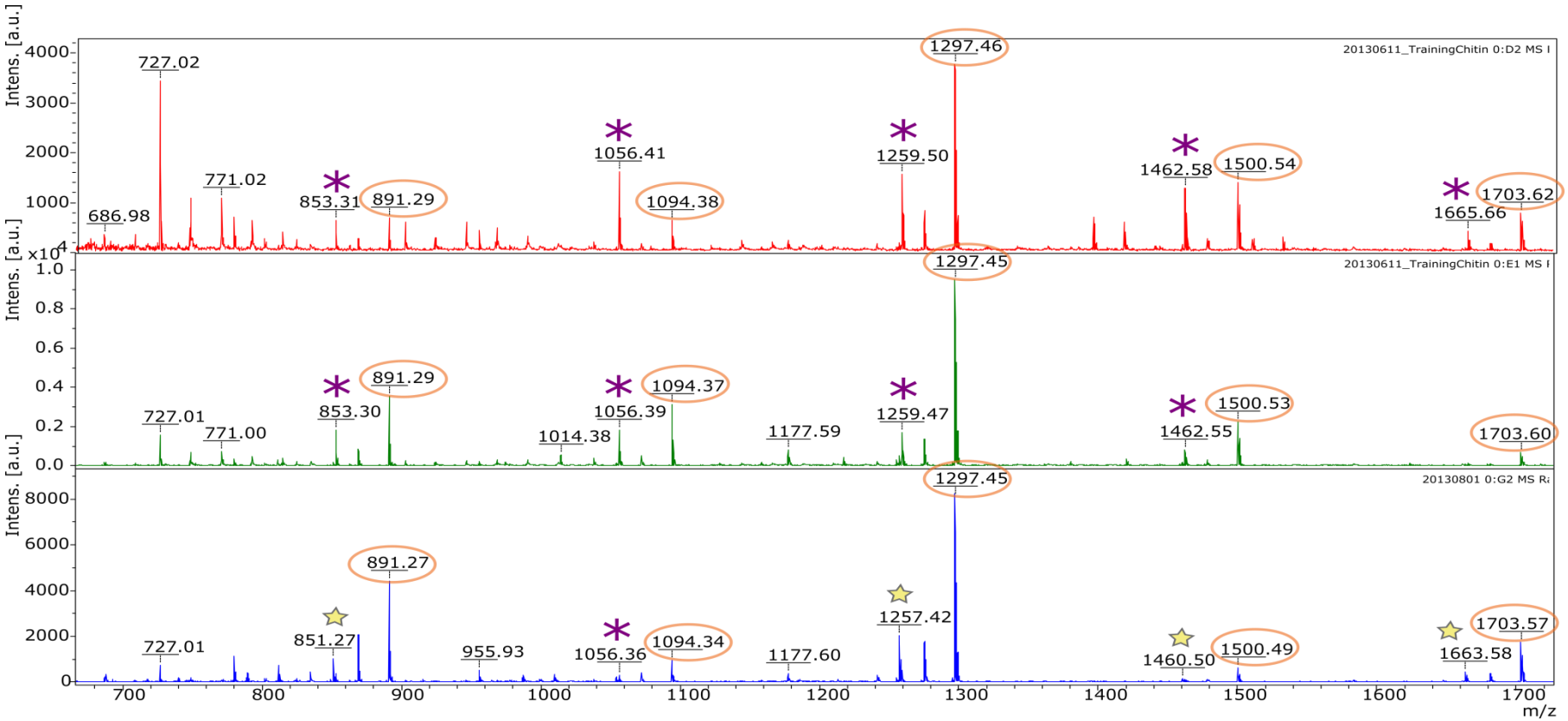

Figure 4.4.1.3 Comparison of the activity of *Ba*AA10 on chitin from different sources: the red spectrum is crab (α chitin); green is shrimp (α chitin); blue is squid pen (β chitin). All aldonic acids are circled in orange, whilst any chitooligosaccharides not produced from oxidative action are marked with purple asterisks. In the case of the products from the squid pen chitin sample, the presence of lactones (denoted by yellow stars) from oxidative action is more common to see in the annotated spectrum than the corresponding chitooligosaccharides. However, these peaks could also be assigned to products of C4 oxidation, which is impossible to define using this method of mass spectrometry.

From Figure 4.4.1.3, it is possible to see that the chitin from both crab and shrimp contained a significant proportion of chito-oligosaccharides, which make interpretation of any products obtained by oxidation of the chitin more challenging. Despite attempted cleaning and synthesis of chitin "whiskers" by sonication in acetic acid,[34, 128] some chito-oligosaccharide products still remained in the crab and shrimp chitin forms. However, it is interesting to note that the *Ba*AA10 enzyme appears to be active on crab, shrimp and squid forms of chitin, which means that it is active on both α-chitin and β-chitin in parallel and anti-parallel arrangements.

4.4.2 Differential Scanning Fluorimetry (DSF) of *Ba*AA10 with substrate

Screening for substrate activity using MALDI-TOF Mass Spectrometry can be a challenging and labour intensive process, and this can be made even more difficult due to the presence of soluble oligosaccharides in the substrate samples, as observed here for α chitin. As an alternative approach for identifying likely substrate preferences, the interaction of *Ba*AA10 with chitin was further investigated using Differential Scanning Fluorimetry (DSF). This technique was used to investigate protein stability in the presence of chitin, with increased protein stability indicative of an interaction with the substrate, which may provide an initial high throughput screen towards identifying LPMO substrates (Figure 4.4.2.1). The thermal stability of LPMOs in the presence of substrate has not been investigated previously; therefore, having established *Ba*AA10 as a chitin active enzyme, we used it as a test case here.

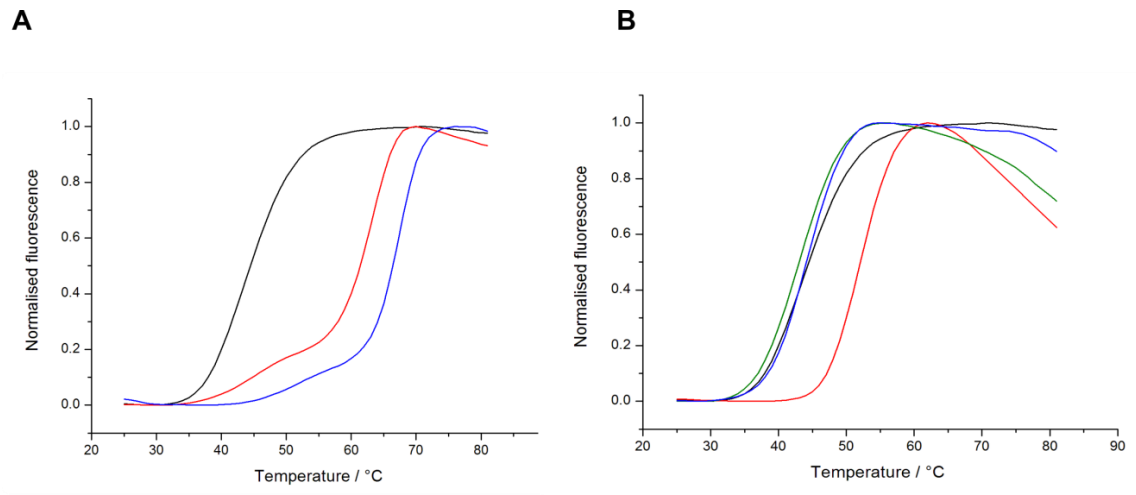

Figure 4.4.2.1 Dissociation curves for the unfolding of *Ba*AA10 ('apo'-*Ba*AA10 is black in both A and B). (A) shows the melting curves of *Ba*AA10 upon the addition of copper (red) as well as the addition of chitin to the copper bound protein (blue). (B) shows the melting curves of *Ba*AA10 upon the addition of chitin (red) compared to the additions of cellulose (green) and chitohexaose (blue).

The stability of *Ba*AA10 was measured, along with the stability upon the addition of copper and upon the addition of copper and chitin as a substrate. Dissociation curves were obtained (Figure 4.4.2.1), from which the melting temperatures (T_m) for each sample were calculated using the MTSA program on MATLAB.[122] The subsequent melting temperatures and their respective R^2 values of best fit are shown in Tables 4.4.2.1 and 4.4.2.2. These R^2 values show a good fit of the dissociation curve to the data in all cases. In samples that contained copper, the fluorescence signal was significantly reduced, presumably due to fluorescence quenching of the SYPRO-orange dye resulting from the presence of metal ions. However, melting curves could still be obtained showing an increase in melting temperature of *Ba*AA10 from 65.2 °C to 68.7 °C in the presence of chitin, which is a 3.5 °C increase in stability in the presence of substrate.

Table 4.4.2.1 Melting temperatures of 'apo'-*Ba*AA10, purified as described in Section 2.3.3, as well as of the protein with chitin compared to cellulose and chitohexaose. The corresponding R^2 values of best fit of the melting curves are also shown.

Sample	T_m / °C	R^2
'apo'-*Ba*AA10	43.6	0.99997
*Ba*AA10 + Chitin	51.9	0.99992
*Ba*AA10 + Cellulose	42.9	0.99989
*Ba*AA10 + Chitohexaose	44.2	0.99991

Table 4.4.2.2 Melting temperatures and corresponding R^2 values of *Ba*AA10 with copper, zinc or EDTA, either with or without chitin, to demonstrate the increase in stability when chitin is added to the protein regardless of the metal bound in the active site.

Sample	T_m / °C	R^2
*Ba*AA10 + Cu^{2+}	65.2	0.99466
*Ba*AA10 + Cu^{2+} + Chitin	68.7	0.99723
*Ba*AA10 + Zn^{2+}	46.4	0.99973
*Ba*AA10 + Zn^{2+} + Chitin	53.6	0.99984
*Ba*AA10 + EDTA	41.8	0.99994
*Ba*AA10 + EDTA + Chitin	50.6	0.99997

To see the increase in stability of the enzyme in more detail, the same experiment was carried out with zinc and also with EDTA. These experiments show that, even in the absence of the known redox metal cofactor Cu^{2+}, a significant stabilisation of the protein in the presence of substrate is still observed with the T_m increasing by 7.2 °C increase in the sample with chitin and Zn^{2+}, and an 8.8 °C increase in enzyme stability in samples containing EDTA (see Table 4.4.2.2). Furthermore, chitin in the absence of enzyme was tested to see if it interfered with the fluorescence readings. However, the dissociation curve obtained for this (data not shown), showed a flat baseline indicating that there was no fluorescence change detected in this sample, proving that any change in fluorescence could solely be credited to the enzyme unfolding. As further proof for the use of this method to probe substrate preference, controls using cellulose (Avicel) and chitohexaose as other potential substrates showed no significant change in melting temperature with these poly/oligosaccharides, indicating that chitin alone binds significantly to *Ba*AA10. This not only reaffirms the fact that chitin is the substrate for *Ba*AA10 but, notably, shows that a soluble chitin-derived hexasaccharide, chitohexaose, had no measurable effect on stability, showing that it is the crystalline polymer alone that is the enzyme's substrate.

4.5 Discussion and conclusion

Sequence analysis of *Ba*AA10 indicated that this LPMO was a likely chitin specific enzyme, according to similarities in sequence with other AA10 enzymes. Furthermore, a study by Chu *et al.*[112] had shown previously that this protein, then named ChbB, was able to bind both α and β forms of chitin. By testing the activity of *Ba*AA10 on chitin, as well as a range of other polysaccharide substrates, such as cellulose and starch, it was determined using MALDI-TOF MS that the enzyme only had activity on chitin. A selection of chitin from various sources was tested and *Ba*AA10 was found to be active on both α (from crab and shrimp) and β (from squid pen) forms of chitin, with the most definitive results obtained using β chitin from squid pen. The difference in the number of peaks seen in the mass spectra for the different forms of chitin might be due to the higher purity of squid pen chitin. An alternative explanation could be that *Ba*AA10 degrades squid pen chitin more efficiently because it is β chitin, which potentially is easier to degrade due to fewer interactions between the layers of *N*-acetyl glucosamine arranged in a parallel fashion[16] giving a lower activation energy,[130] but it is difficult to discern which is most likely at this time. The results obtained by MALDI-TOF MS provide us with the knowledge of how chitin is degraded by the AA10 LPMO enzymes, which undertake the oxidative pathway via lactones to aldonic acids.

In addition to establishing the substrate preference of *Ba*AA10, and demonstrating its LPMO activity for the first time, it was also found that the presence of this substrate, chitin, increases the stability of the protein. Whether in the presence of copper or in its absence, an increase in melting temperature of 3.5 °C and 8.8 °C respectively, was induced in the protein resulting from the addition of chitin. Significantly, this offers evidence that this methodology could be used to screen substrate preferences in new enzymes in a higher throughput way, but also suggests that only when the substrate is present does the enzyme become ready for oxidative activity, which requires the enzyme to be in a more stable state. From subsequent EPR analysis (addressed in Chapter 6), it can be seen that chitin induces superhyperfine coupling in spectra obtained, indicating that the substrate does stabilise the environment around the copper in some way. How the substrate affects the environment around the copper could perhaps be studied by looking at the crystallographic structure, in particular around the copper, in more detail. Therefore, subsequent structural data and analysis of *Ba*AA10 was completed and is addressed in the next chapter.

5. Crystallographic studies to establish the 3D structure of a Cu(II) form of *Ba*AA10

5.1 Abstract

The copper active site, and how the copper interacts with its environment, is an area of significant interest for research as it has the potential to provide important information regarding the oxidative pathway for the activity of the lytic polysaccharide monooxygenase enzymes. To this end, the 3D structures of the enzymes have become an integral part in the detailed investigation of the copper centre. Unfortunately, most 3D analyses of AA enzymes result in photoreduction of the Cu(II) to Cu(I) in the X-ray beam; this makes it difficult to compare X-ray structures with solution spectroscopy of the Cu(II) state. Our initial solution of the *Ba*AA10 enzyme indeed revealed copper in its reduced Cu(I) state, providing insufficient details as to the complete environment around the copper upon substrate binding and subsequent oxidative action. Therefore, with an aim to obtain a 3D structure of Cu(II)-*Ba*AA10, here I present how crystal optimisations and spiral data collection were undertaken in a bid to prevent photoreduction of the copper, leading to our first observation of a Cu(II) form of the enzyme. The structure of AA10 enzymes such as that from *Bacillus amyloliquefaciens* can now be compared to those of other 'Auxiliary Activity' enzymes (AA9, AA11 and AA13) in light of their spectroscopic observations. Having successfully achieved a structure with copper in its Cu(II) state, comparisons with data produced from EPR studies are made to further understand the nature of the copper centre in the AA10 enzymes.

5.2 Introduction

Solving the structures of newly-discovered enzymes, such as these lytic polysaccharide monooxygenases, is extremely important in helping to determine how their structure might relate to their activity and their potential mechanism of action. In addition, the structure of the protein around the copper active site can provide us with the environment in which any potential substrate will be exposed to prior to glycosidic bond cleavage and subsequent degradation of the polysaccharide. Prior to the start of this work, the 3D structures of various AA9 enzymes, such as those from *Hypocrea* jecorina[35] and *Thermoascus* aurantiacus,[38] were determined, providing an idea of the overall structure of the LPMO enzymes. Before this, however, the structure of an AA10 enzyme (then, however, thought to be a chitin-binding protein, known as CBP21) from the bacterium *Serratia marcescens* had been solved (Figure 5.2.1A)[41], albeit with an incorrect assignment of the metal at the active site. Despite not knowing of the potential enzymatic action of this protein structure, the existence of a structure for comparison with other LPMO enzymes, including future AA10 enzymes, has been of vital importance.

At the commencement of this research, the 3D structure of *Ba*AA10 had been solved using crystallographic methods, by PDRA Glyn Hemsworth.[72] The structure solution involved use of the molecular replacement method, using the only solved structure of an AA10 enzyme available at the time, that of *Serratia marcescens*.[41] A comparison of the two AA10 structures can be seen in Figure 5.2.1, which shows the similarities in overall structure of the enzymes, which contains an elongated globular unit composed of β-sheets linked to a smaller protrusion of mainly α-helices and flexible loops. An initial similarity search carried out by Vaaje-Kolstad *et al.* upon determining the 'CBP21' structure found that the CBM33 protein had a structure similar to that of the fibronectin type III (FnIII)[131] chitin binding domain of the chitinase enzyme from *Serratia marcescens* (ChiA).[132] Interestingly, this domain is common in carbohydrate active enzymes found in bacteria, providing some insight into how substrates such as chitin might be able to bind to these LPMO enzymes.

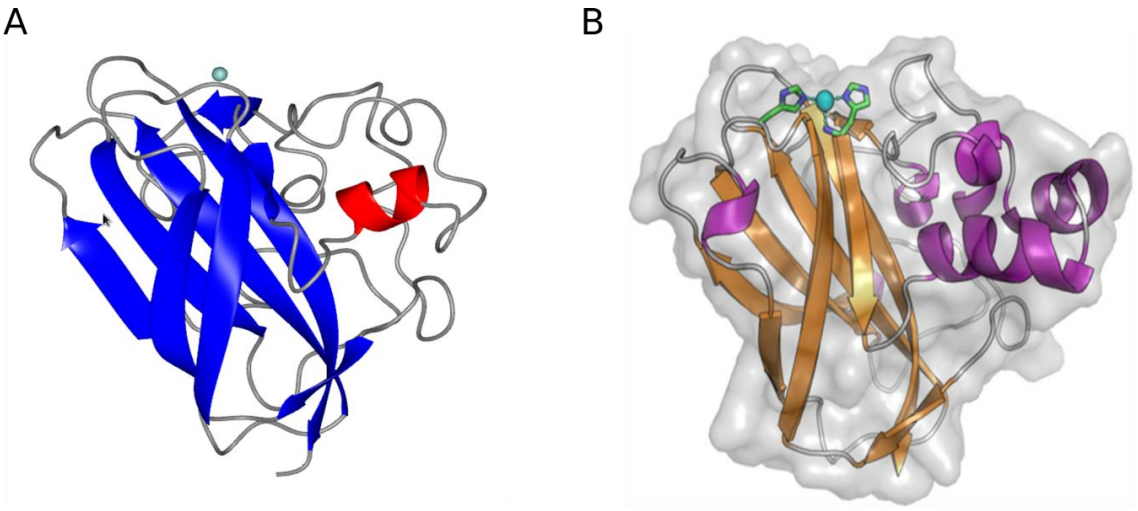

Figure 5.2.1 The crystal structures of A) 'CBP21' from *Serratia* marcescens[41] (PDB Code: 2BEM); B) copper-bound *Ba*AA10, solved in 2012,[72] showing the copper in the active site, which was reduced to Cu(II) in the X-ray beam.

Structures of the *apo*, Cu-bound and Cu-bound/ascorbate soaked *Ba*AA10 were solved by Dr. Glyn Hemsworth, allowing analysis of the copper active site to be carried out. As a result, it was determined that the Cu(II) used in co-crystallisations of the protein was being photoreduced in the x-ray beam to Cu(I). This was confirmed by Cu K-edge fluorescence spectroscopy (XANES studies), as the Cu-bound crystal produced the same pattern and behaviour as that of the chemically reduced sample by ascorbate in solution, identified by a 'pre-edge' feature in the XANES at *ca.* 8983 eV.[72] Similar photoreduction of the copper in the active site of other lytic polysaccharide monooxygenases, such as that of the AA10 from *Enterococcus faecalis*, has also been reported.[78]

One way of preventing photoreduction of the Cu(II) from taking place is to reduce the amount of X-ray exposure to the crystal whilst collecting data. In order to reduce the exposure time of the crystal to the X-ray beam, a spiral data collection technique can be carried out, similar to that performed by Gudmundsson *et al.* in their redox studies of *Ef*AA10.[78] Spiral data collection involves performing a "line scan" across the length of the crystal, whilst simultaneously rotating the crystal by a certain degree after each pulse of the X-ray beam (Figure 5.2.2). In this manner, each exposure is to a previously unexposed part of the crystal. In order to achieve successful spiral data collection, long crystals wide enough to perform a line scan across the crystal in various rotations are required. Previous crystallisations of *Ba*AA10 proved to be unsuccessful in achieving these types of crystals, with showers of small needle-like crystals forming in most crystallisation trays. This indicates that the protein may be

forming crystals too quickly to be able to obtain the larger crystals required for successful data collection. In an attempt to obtain better crystals, seeding using seed beads can be used to aid the process of nucleation by providing a template for molecules to act upon to form a successful crystal lattice.[133]

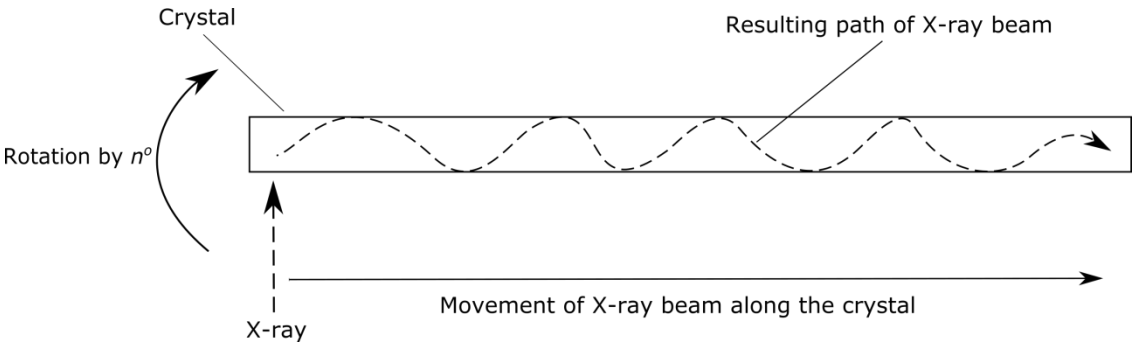

Figure 5.2.2 Schematic diagram to demonstrate spiral data collection, where *n* is the number of degrees the crystal is rotated during data collection.

By attempting to determine a Cu(II) structure of the *Ba*AA10 enzyme, a clearer idea of the geometry and environment around the copper in the active site of the enzyme, one which can be correlated with EPR observations, can be established. The crystallisation methods described in this chapter enabled large enough crystals to be obtained for successful spiral data collection, culminating in a Cu(II)-*Ba*AA10 structure, which is analysed with regards to its Cu(I) counterpart and EPR spectra obtained of the enzyme.

5.3 Materials and methods

5.3.1 Crystallisation and optimisation of *Ba*AA10 crystals for X-ray diffraction

Copper chloride ($CuCl_2$) was added to *Ba*AA10 in a stoichiometric amount, and the protein was then buffer-exchanged into 20 mM NaOAc pH 5 to reduce the amount of salt present, before concentrating to ~ 5 mg/ml for crystallisation. Initial crystallisation trials were set up using 0.15 µl of protein and 0.15 µl of reservoir solution with the PACT (Qiagen) screen on a TTPLabtech Mosquito crystallisation robot. The protein crystallised in almost all conditions of the PACT screen, however only showers of microcrystals were obtained. Larger crystals were found in condition D1 of the PACT screen: 0.1 M MMT pH 4.0; 25 % PEG-1500. Therefore, optimisations were set up based around this condition in 48 well sitting-drop trays with differing concentrations of PEG-1500 (20 %, 22 %, 24 %, 26 %, 28 % and 30 %).

In order to obtain longer rod-shaped crystals for spiral data collection, further optimisation of the crystals obtained previously was necessary. Consequently, a seed stock of the *Ba*AA10 protein was made by crushing the thin crystals found in one of the drops of the sitting drop tray (0.1 M MMT pH4; 26 % PEG 1500) and mixing them with seed beads (Hampton Research) in the mother liquor. This was vortexed, before being used to set up a MRC 2-drop 96-well plate on a Douglas Oryx crystallisation robot. Conditions A7 – A11 and B7 – B11 of the PACT screen were used in the reservoir wells (A7 – A11: 0.1 M Acetate pH 5, 20 % PEG 6000, 0.2 M NaCl, LiCl, NH_4Cl, $MgCl_2$ or $CaCl_2$; B7 – B11: 0.1 MES pH 6, 20 % PEG 6000, 0.2 M NaCl, LiCl, NH_4Cl, $MgCl_2$ or $CaCl_2$), after an initial PACT screen showed that these conditions resulted in the largest crystals following use of the seeding technique. Drops of 0.25 µl reservoir solution, 0.30 µl protein and 0.05 µl seed bead solution were then dispensed. The concentration of *Ba*AA10 used for this tray was 4.7 mg/ml, and $CuCl_2$ had been added to the protein in a stoichiometric amount.

5.3.2 X-ray diffraction of *Ba*AA10 using spiral data collection techniques

The longest Cu-*Ba*AA10 protein crystals obtained were placed in cryo-protectant for 30 seconds before being cry-cooled in liquid nitrogen. Diffraction data for the crystals was collected on the I04-1 beamline at the Diamond Light Source (DLS). A cryo-protectant of mother liquor (0.1 M NaOAc pH 5, 20 % PEG-6000, 0.2 M $CaCl_2$) together with 20 % ethylene glycol was used. A spiral data collection technique, inspired by the work of Gudmundsson *et al.*[78], was used for data collection. This involved running a line scan along the rod-shaped crystal, whilst the crystal was rotated by 180°, ensuring that a new point on the crystal was

always exposed to the X-ray beam during data collection. In order to further reduce the chances of photoreduction of the Cu occurring in the X-ray beam, a 5 μm beam was used, and the X-ray beam was attenuated to 5 % transmission.

Data was processed using XDS,[134] followed by the Aimless reduction pipeline on the CCP4i2 software.[135] The previous *Ba*AA10 PDB structure (2YOX)[72] was used to determine the phases and solve the structures. All water molecules and copper ions were removed from the original structure used for refinement to prevent any model bias. Refinement of the structures was carried out using Refmac[136] and model building completed with Coot[137], both through CCP4i2.[135] The final structure was deposited for publication in the Protein Data Bank (PDB) with code 5IJU.

5.4 Results

5.4.1 Crystallisations and optimised crystals of *Ba*AA10

The structure of *Ba*AA10 was solved prior to the commencement of my PhD research. Therefore, it was known that the AA10 protein from *Bacillus amyloliquefaciens* could be crystallised in almost all conditions of the PACT screen. The structure was solved, for example, from crystals formed in condition D1 of the PACT screen (0.1 M MMT pH 4, 25 % PEG-1500). Further PACT screen crystallisations of *Ba*AA10 confirmed that the enzyme crystallises in almost every condition, however it can be seen from Figure 5.4.1.1A that *Ba*AA10, despite readily forming crystals, crystallises into showers of lots of small, thin needles, which are difficult to fish successfully for data collection. A sitting-drop optimisation tray, set up around the conditions found in D1 of the PACT screen (0.1 M MMT pH 4, 20-30 % PEG-1500), produced *Ba*AA10 crystals which were thicker in nature, albeit still very small and needle-like in size (Figure 5.4.1.1B).

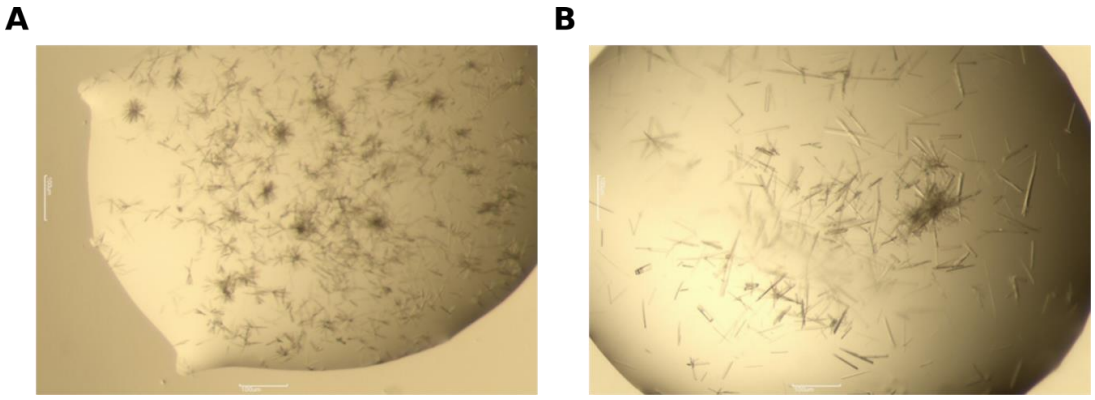

Figure 5.4.1.1 *Ba*AA10 crystals at 5.2 mg/ml before (A) and after (B) optimisation. (A) is condition D1 of the PACT screen: 0.1 M MMT pH 4.0; 25 % PEG-1500. (B) is the optimised condition: 0.1 M MMT pH4; 26 % PEG 1500.

A number of crystals from these sitting-drop optimisation trays were fished for data collection attempts using the microfocus I24 beam at the Diamond Light Source synchrotron. This beamline has a smaller beam size, which was necessary to prevent damage by X-ray photoreduction. However, initial processing during data collection indicated that the copper was still being reduced from Cu(II) to Cu(I) in the X-ray beam, as well as producing poor resolution and diffraction of the crystals. Consequently, a spiral data collection technique, similar to that used by Gudmundsson *et al.*,[78] was used, which would further reduce the risk of

Cu photoreduction due to the method involving the X-rays hitting a different part of the crystal every time. However, as this type of data collection involves a line scan, the best results for spiral data collection require as long and thick rod-shaped crystals as possible, resulting in the need for further optimisations of the *Ba*AA10 crystals to take place.

One successful attempt at producing such crystals was through the seeding technique using seed beads (Hampton research) and setting up trays on a Douglas Oryx Crystallisation Robot. An initial PACT screen using the seed beads was used to obtain larger crystals, so further trays using the most successful conditions (A7 – A11: 0.1 M Acetate pH 5, 20 % PEG 6000, 0.2 M NaCl, LiCl, NH_4Cl, $MgCl_2$ or $CaCl_2$; B7 – B11: 0.1 MES pH 6, 20 % PEG 6000, 0.2 M NaCl, LiCl, NH_4Cl, $MgCl_2$ or $CaCl_2$) were set up to obtain a large number of crystals for data collection attempts. These crystallisation conditions were very effective in obtaining rod-shaped crystals (Figure 5.4.1.2) both long and wide enough for spiral data collection using the IO4-1 beam at the Diamond Light Source synchrotron.

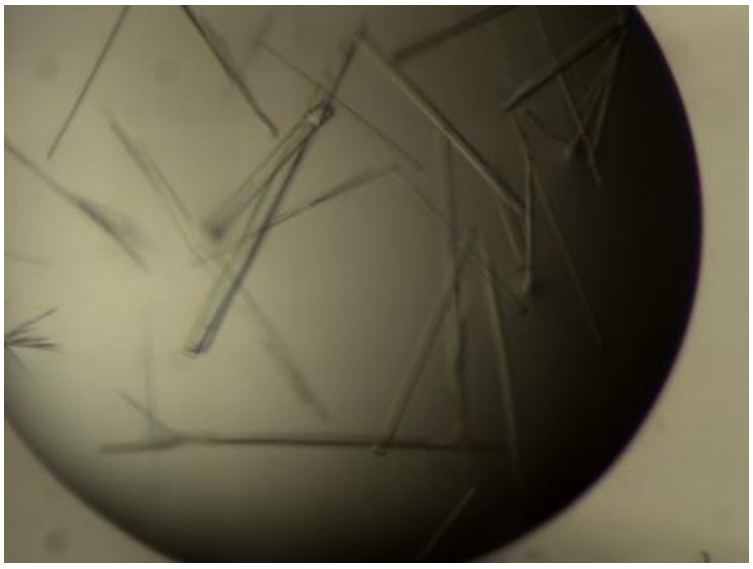

Figure 5.4.1.2 Long rod-shaped crystals of *Ba*AA10 following crystallisation by seeding methods. These crystals were suitable for fishing and X-ray data collection at the synchrotron, where the concentration of protein was 4.7 mg/ml and the reservoir condition was 0.1 M NaOAc pH 5, 20 % PEG-6000, 0.2 M $CaCl_2$.

5.4.2 Data collection of *Ba*AA10 crystals by spiral data collection

Data was collected using a spiral collection technique, which involved carrying out a line scan from one end of the crystal to the other at the same time as rotating the crystal by a certain

degree. In this case, the crystals were rotated by 180 degrees, to enable enough data to be collected at the same time as minimising the collection of data from already exposed sections of the crystal. Spiral data collection proved difficult for these crystals, as it was hard to centre the crystals at each end. Out of the eleven Cu-*Ba*AA10 crystals sent to Diamond, eight data sets were collected, with two of these comprising poor quality data; the other three crystals could not be centred correctly for the run. With this difficulty in centring the crystals and the amount of time it took, the use of even larger and thicker crystals could make this process easier, as the beam would be more likely to hit the crystal when it is rotated by 90 degrees if it is thicker in size.

In addition to the difficulty encountered in collecting the data, the data collected were weak, as only 5 % transmission of the beam was used. This meant that the diffraction spots had poor signal to noise, and so the auto-processing carried out by xia2 and Mosflm at Diamond was not successful in most cases, as data was either processed in the wrong space-group, or data failed to process altogether. Therefore, manual processing using XDS[134] was used to provide the most likely space group and cell dimensions, and hence solve the structures using the aforementioned *Ba*AA10 structure (PDB: 2YOX).[72] As the data was collected at low intensity in an attempt to avoid photoreduction of the copper in the x-ray beam, careful processing of the data had to be undertaken. Therefore, based on the criterion that CC (1/2) must have a value greater than 0.5,[138] the data was processed at a resolution of 1.7 Å where CC (1/2) was equal to 0.54, which signified that the data present at this resolution are indeed meaningful and able to be processed and refined satisfactorily.

From the collected data, one data set was chosen to use for refinement. During refinement, water molecules and copper ions were not added to the structures until the final stages of refinement, so as not to bias the models. The details of refinement, including values for the various R factors, as well as for data processing can be found in Table 5.4.2.1. The crystal structure was composed of two molecules in the asymmetric unit, and crystals grew in the $P2_1$ space group. Both molecules in the asymmetric unit were highly similar to one another (a superposition of all the Cα atoms in the two molecules generates a RMSD of 0.16 Å), and as such subsequent structural description will be given on solely the A chain of the protein structure.

Table 5.4.2.1 Crystallographic data obtained for *Ba*AA10, including refinement statistics

Data Collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	34.8, 73.5, 75.5
α, β, γ (°)	90.0, 100.4, 90.0
Resolution range (Å)	37.11 – 1.70 (1.73 – 1.70)
Completeness (%)	99.8 (99.4)
Redundancy	3.3 (2.8)
$\langle I / \sigma (I) \rangle$	5.3 (1.3)
Half-set correlation CC (1/2)	0.985 (0.54)
$R_{p.i.m.}$	0.108 (0.573)
Overall *B* factor from Wilson plot ($Å^2$)	7.3
Refinement	
Resolution range (Å)	37.11 – 1.70 (1.74 – 1.70)
Completeness (%)	99.7 (99.3)
No. of reflections, working set / test set	38958 / 2062
Final R_{cryst} / R_{free}	0.223 / 0.264
No. of non-H atoms	
Protein	2782
Ion	3
Solvent (Ethylene glycol)	8
Water	282
Total	3075
R.m.s. deviations: Bonds(Å); Angles (°)	0.014; 1.64
Average B factors ($Å^2$)	
Protein	9
Ion	11
Solvent (Ethylene glycol)	22
Water	15
Ramachandran plot	
Most favoured (%)	97
Allowed (%)	3

5.4.3 Crystallographic structure and structural comparisons of *Ba*AA10 following X-ray data collection and processing

The overall structure of *Ba*AA10 complies with the general shape and nature of the LPMO enzymes (Figure 5.4.3.1), consisting of a fundamental β-sandwich domain adjacent to an α-helical region linked by flexible loops. The enzyme's structure contains an extended flat surface, where the copper ion is situated in the N-terminal active site, coordinated to three nitrogen atoms originating from the NH_2 terminus, one nitrogen from the N-terminal histidine residue (His28) and a further nitrogen from another histidine elsewhere in the molecule (His125). This coordination site, which involves the bridging of two histidine residues, has commonly become known as the 'histidine-brace', following the introduction of the term in 2011 by Quinlan *et al*.[38] The extended flat surface, which harbours the copper within the 'histidine brace', measures approximately 30 x 40 Å, and provides the enzyme with a suitable binding-site for long chain crystalline substrates, such as chitin, enabling them to more easily come into contact with the chemistry occurring at the copper ion. The structure of *Ba*AA10 in Figure 5.4.3.1 highlights the copper active site on the surface of the enzyme, with the copper ion in grey coordinated by three nitrogen atoms and two red water molecules.

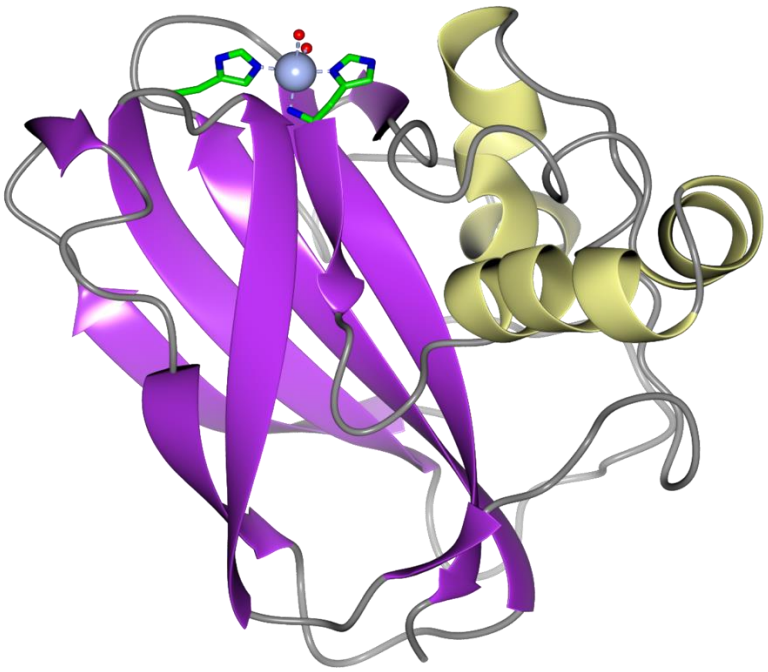

Figure 5.4.3.1 Tertiary structure of *Ba*AA10, showing the copper active site situated in the 'histidine-brace'[38] on the flat surface of the protein molecule, where its substrate is likely to bind. The presence of two water molecules around the copper ion (indicated as red spheres) show that the copper bound is Cu(II) and has not been reduced to Cu(I).

The surface of the enzyme provides an insight into the binding of potential substrates. Whilst the issue of substrate specificity amongst the LPMO enzymes remains a major question to be answered, it is generally thought that the amino acid residues found on the surface of the enzymes close to the copper active site allow binding of the substrate along the extended flat surface through hydrogen bonding, aromatic π-stacking interactions and Van der Waals' interactions. Indeed, Figure 5.4.3.2 of the surface of the enzyme looking down on the copper active site (indicated by the asterisked circle), shows that there is a large red area across the greatest width of the top surface where the copper active site is found. This red area indicates positively charged polar residues, capable of participating in H-bonding with a substrate molecule.

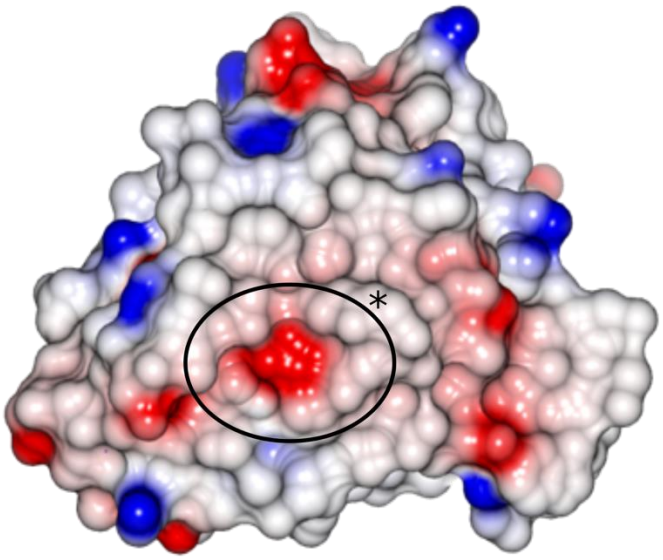

Figure 5.4.3.2 Crystallographic structure of *Ba*AA10 portraying the surface of the enzyme, looking down upon the copper active site (*), showing areas on the surface where positively charged (red) and negatively charged (blue) polar residues are found.

A detailed view of the electron density around the copper ion, which provides a clear visualisation at the spheres of electron density resulting from the presence of two water molecules, can be seen in Figure 5.4.3.3A. The structure and the coordination around the copper active site in particular is very similar to another AA10 enzyme from the *Enterococcus faecalis* bacterium, the crystal structure of which was published by Gudmundsson *et al.*[78] (PDB Code: 4ALC), who were also able to obtain the non-reduced structure by spiral data collection techniques. A comparison of the two structures can be seen in Figure 5.4.3.3B, where both structures clearly possess the two water molecules and have very similar positioning of the amino acid residues surrounding the copper ion.

These two figures clearly show the active site around the copper ion in more detail, in particular the two histidine residues forming the 'histidine brace' and coordination of copper with the N-terminus of the protein. The presence of both the alanine and phenylalanine side chains, which are situated in close proximity to the copper active site, demonstrates their respective positions in relation to the histidine residues and water molecules coordinating to the copper. These two residues may therefore have a steric effect on the geometry of the active site and may assist in orchestrating the method of oxygen binding and ultimate cleavage of the polysaccharide substrate. In addition, the electron density around the copper site shows that the

N-terminal histidine residue is not methylated in AA10 enzymes, in contrast to the 3D structures of various AA9 enzymes, when expressed in filamentous fungi hosts.[37, 38, 92, 139] However this could also be because the AA10 enzymes have been expressed in *E. coli*, which does not possess the correct machinery required for any post-translational methylation modifications.

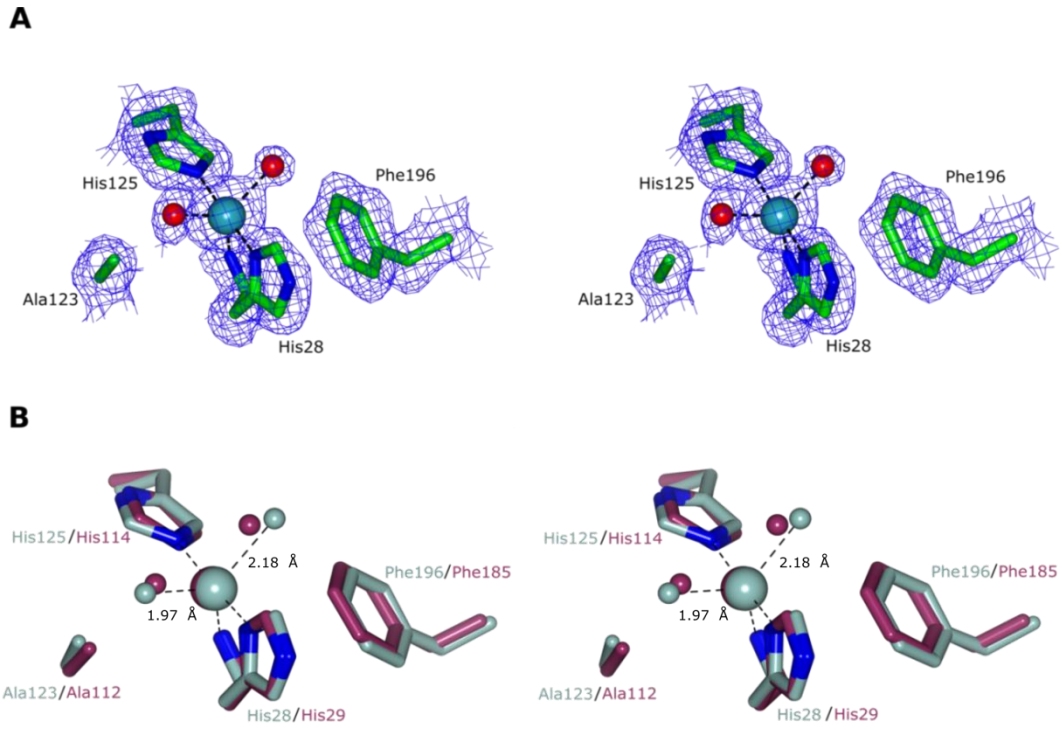

Figure 5.4.3.3 (A) Stereo image of the copper active site of *Ba*AA10 showing the two water molecules coordinated to the copper ion. The $2F_{obs}$-F_{calc} map is displayed at a contour level of 0.48 e/A^3 (1 σ). (B) Stereo image of the comparison between the Cu(II) active site of *Ba*AA10 (blue/grey) with that of the AA10 enzyme from *Enterococcus faecalis* (purple).[78]

The presence of the two water molecules has until now been undetected in attempts to obtain the 3D structure of the AA10 enzyme from *Bacillus amyloliquefaciens* in its Cu(II) state. Previous structures of *Ba*AA10[72] have not been able to show electron density for the two water molecules coordinating to the copper ion, indicating that with only three coordinating atoms present in the structure, the copper was in a Cu(I) state, with the copper having been photoreduced in the X-ray beam from Cu(II) to Cu(I). Indeed, this was proved to be the case following fluorescence XANES studies carried out at Diamond Light Source, which showed that the Cu K-edge results of *Ba*AA10 crystals possess the same features and demonstrate the

same behaviour as samples reduced chemically by ascorbate in solution.[72] Figure 5.4.3.4A shows a more detailed look at the coordination of the protein to the copper ion, and compares the Cu(I) structure from Hemsworth *et al.*[72] to the Cu(II) structure obtained during this research. The main differences are in the basic geometries around the copper ion, with the Cu(I)-*Ba*AA10 structure clearly demonstrating a T-shaped geometry, whilst the Cu(II)-*Ba*AA10 structure possesses a distorted trigonal bipyramidal geometry (schematically represented in Figure 5.4.3.4B). The bond angles between the water molecules and the N-terminal NH_2 of the protein demonstrate that neither of these water molecules form a direct *trans* relationship with this position. The O-Cu(II)-NH_2 bond angles are 131.9 ° and 128.4 ° for the two water molecules, when one would predict angles of 120 ° for both of these angles in a true trigonal bipyramidal geometry. The fact that these bonds are greater than 120 degrees provides evidence of the distortion within the Cu(II) geometry in the structure of AA10 enzymes.

In addition to the bond angles, there is further evidence of distortion in the copper geometry, in that the compressed bond lengths amongst the atoms that coordinate to the copper in the active site. The copper-ligand bond lengths include the axial positions of $N\varepsilon2_{His125}$-Cu= 2.05 Å, $N\delta1_{His28}$-Cu= 2.01 Å and equatorial nitrogen and water positions of NH_{His28}-Cu=2.24 Å, $O_{H2O\text{-}412}$-Cu= 1.97 Å and $O_{H2O\text{-}487}$-Cu= 2.18 Å. As in the AA10 from *Enterococcus faecalis,*[78] the two axial nitrogen distances increase slightly (0.07 Å) upon reduction from Cu(II) to Cu(I), whilst the distance to the equatorial nitrogen decreases. The O_{H20}-Cu bond lengths demonstrate that neither of the two water molecules in the Cu(II) structure are positioned directly *trans* to the equatorial nitrogen atom coordinated to the copper, providing further evidence of the distorted trigonal bipyramidal geometry.

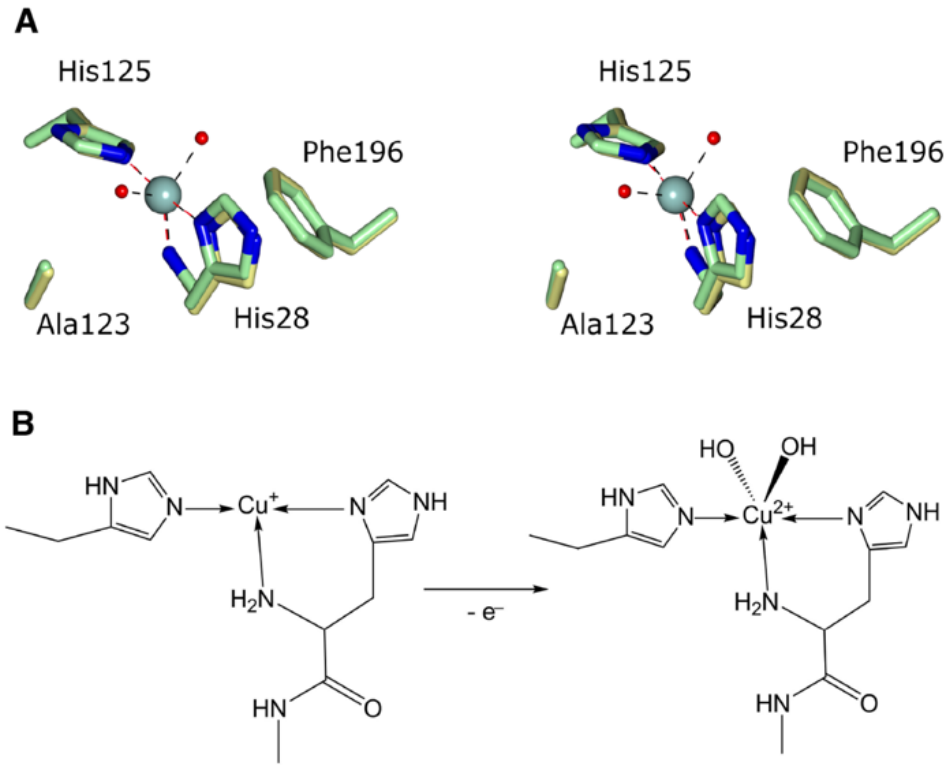

Figure 5.4.3.4 (A) A structural comparison of Cu(II)-*Ba*AA10 (yellow) with the structure of Cu(I)-*Ba*AA10 obtained originally (green),[72] portrayed in stereo. The water molecules observed in the Cu(II) structure are shown as red spheres. (B) Schematic representation of the T-shaped geometry observed in the reduced Cu^{+} species and the distorted trigonal bipyramidal geometry found in the Cu^{2+} species.

Crystallographic studies have been able to provide structural data regarding the specific details of bond angles and lengths around the copper ion within the active site of the *Ba*AA10 enzyme. This data corroborates with results obtained by Electron Paramagnetic Resonance (EPR) spectroscopy, which was performed as an alternative to crystallography. As copper in its Cu(II) state is paramagnetic, EPR can be undertaken in order to obtain spectra which can be interpreted to provide a picture of the geometry around the copper centre. EPR studies carried out on *Ba*AA10 have produced copper spectra previously unseen in copper EPR literature (Figure 5.4.3.5B). Compared to the copper spectrum obtained from the *Ta*AA9 enzyme (Figure 5.4.3.5A),[38] which displays a regular Type II copper spectrum, the *Ba*AA10 enzyme exhibits both axial and rhombic characteristics, indicating the highly distorted trigonal bipyramidal geometry that results from $d(z^2)$ orbital mixing in the $d(x^2\text{-}y^2)$ SOMO.[72] This characteristic spectrum for *Ba*AA10 has been maintained throughout our EPR studies of the enzyme, including after improvements in protein production were made using a SUMO tag expression and purification method (see Chapter 2). The proposed 'distorted trigonal

bipyramidal' geometry of the copper active site of *Ba*AA10 was confirmed by the crystallographic studies carried out in this chapter, following the attainment of a Cu(II) structure.

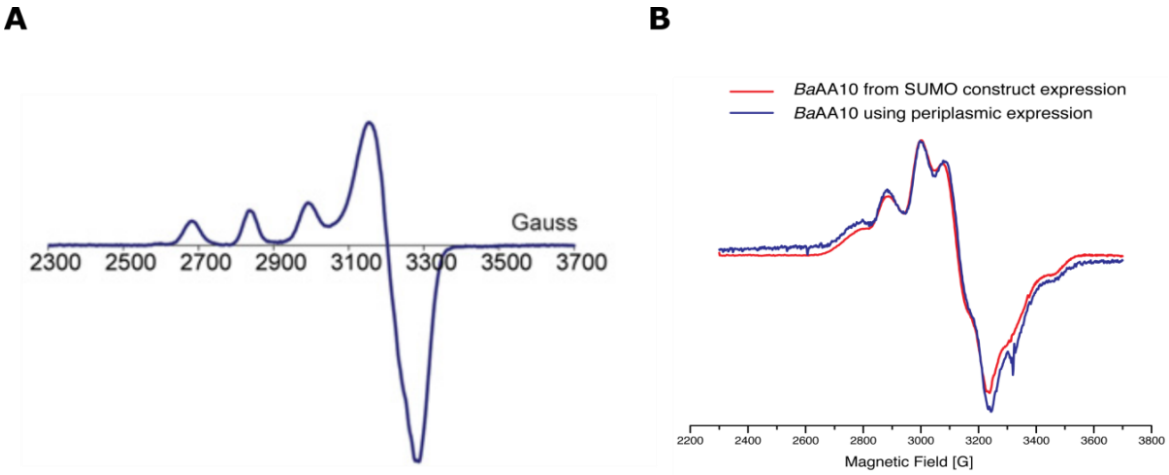

Figure 5.4.3.5 A) X-band EPR spectrum of *Ta*AA9, portraying a typical Cu(II) motif (9 GHz, 140 K).[38] B) X-band EPR spectrum of *Ba*AA10,[72] carried out prior to this study using protein isolated from a periplasmic expression protocol (9 GHz, 155 K, blue spectrum), as well as protein produced following the SUMO tagged production method (9 GHz, 150 K, red spectrum).

5.5 Discussion and conclusion

Structural studies of *Ba*AA10 carried out prior to the onset of this study involved the use of both crystallographic and EPR techniques. These studies were able to provide information regarding the overall structure of the enzyme by crystallography, as well as offer an idea about the geometry of the copper centre within the enzyme using EPR. *Ba*AA10 was found to be a globular protein composed of a core of β-sheets and an adjacent α-helical unit. As in all other LPMO structures, the enzyme possesses an extended flat surface, upon which chitin can bind as a substrate. This binding surface harnesses the copper centre, which was found to possess distorted trigonal bipyramidal geometry. The copper is coordinated to three nitrogen atoms, which originate from two histidine residues on the surface of the protein and from the N-terminus of the protein, as well as coordinating two water molecules.[72] Confirmation of the copper active site geometry and environment whilst the copper is in the Cu(II) state had, however, yet to be achieved. The aim of this research was thus to determine a structure of *Ba*AA10 with copper in its Cu(II) state.

Following crystal optimisations via seeding methods, *Ba*AA10 crystals were formed which were longer and thicker than the small needle-like crystal showers obtained in previous crystallisation attempts. These longer crystals were easier to handle for X-ray data collection preparation as well as being more suitable for data collection using a spiral data collection technique, which was used to prevent photoreduction of the copper in the X-ray beam. Consequently, data collection using these crystals by spiral data collection techniques was carried out and resulted in data which could be processed at 1.7 Å and subsequently refined to obtain a Cu(II)-*Ba*AA10 structure. Analysis of the structure and comparison with the Cu(I)-*Ba*AA10 structure was carried out, leading to the conclusion that the enzyme undergoes very few structural changes upon reduction of the copper for subsequent oxidative degradation of the substrate. From this, it can be inferred that the copper is readily available to be reduced whenever the enzyme requires it to assist in the substrate degradation process, and so forms a structure which would require as little stress or strain as possible to achieve this. How the AA10 enzymes achieve this is still unknown, as the exact oxidative mechanism of action has not yet been determined for this class of LPMO. However, the proposed mechanism of action of the AA9 enzymes has been published by Frandsen *et al.*[71], which suggests that the binding of oxygen and the polysaccharide substrate is synergistic, and that the production of any reactive oxygen species will only occur when the substrate is present at the active site. Further studies similar to those carried out by Frandsen *et al.*[71], that is obtaining a structure with a sugar substrate bound in order to discover exactly how the substrate binds along the surface of the AA10 enzymes, would be fundamental in determining the mechanistic pathway undertaken by

these enzymes. Determining crystal structures of an AA10-substrate complex will only be possible if an AA10 is discovered that degrades a more soluble substrate that can bind to an AA10 crystal, or a method of solubilising long chain polysaccharide substrates for co-crystallisations is uncovered. Alternatively, a structure with oxygen bound would provide more information on how AA10 enzymes bind oxygen for use in the redox degradation process. Indeed, crystal soaks with azide (NaN_3 at 100x the protein concentration) as a potential mimic of oxygen, were carried out and data collected again by spiral data collection, but these experiments did not yield any significant results. Therefore, improved methods in obtaining an oxygen or oxygen mimic (azide or cyanide) bound crystal structure should be researched further, as well as EPR studies of the copper centre when these small molecules are bound to establish the effect on the metal site's environment in the presence of oxygen. The combination of both methods (X-ray crystallography and EPR) would assist in the determination of the copper-oxygen species capable of performing the redox reactions in the cleavage of polysaccharide substrates. Consequently, this knowledge will be extremely useful in trying to establish the potential mechanism of action of the AA10 LPMO enzymes.

6. Conclusion and future perspectives

6.1 Summary of the field

The use of biofuels as a renewable energy source has become increasingly important in recent years, in particular the use of second generation biofuels. In 2013, the world's first industrial plant using enzymatic conversion methods to generate bioethanol from agricultural waste and crops was opened by Beta Renewables in Northern Italy.[140] This facility put into practise the academic and industrial research carried out on biomass degrading enzymes to successfully degrade long chain polysaccharides found within lignocellulosic material to produce cellulosic ethanol.

Improved biofuels processes increasingly demand new and enhanced enzyme solutions. Historically, the industrial degradation of plant cell wall polysaccharides has relied on the synergistic action of classical enzymes, such as *endo*- and *exo*-(processive) acting glucanases, which together are able to break down cellulose into cellobiose. β-glucosidases then convert the cellobiose into glucose, which can undergo subsequent fermentation to produce ethanol.[27, 141, 142] However, cellulose is a highly recalcitrant substrate linked by strong glycosidic bonds, the likes of which result in cellulose having an estimated half-life of 4.7 million years.[11] Therefore, the introduction of new enzymes, such as the lytic polysaccharide monooxygenases described in this work, provides a more efficient and hence cost effective method of enzymatic degradation of polysaccharide biomass.[59]

Lytic polysaccharide monooxygenases have thus become increasingly important within the development of new ways to degrade polysaccharides found within biomass for biofuel production. Although knowledge of these proteins had existed for a number of decades, they had incorrectly been reported as non-enzymatic carbohydrate binding modules (CBM33)[40, 112] and as glycoside hydrolases (GH61).[35, 143] Finally, in 2010/11, the enzymes were shown to be polysaccharide monooxygenases,[34] and subsequently mononuclear copper-dependent enzymes.[38] LPMOs perform a redox reaction to cleave the glycosidic bonds within polysaccharides. The introduction of chain breaks by LPMOs allows for further breakdown and conversion of the oligosaccharide products to glucose by the classical enzymes, before fermentation to bioethanol. Lytic polysaccharide monooxygenases are found in both fungi and bacteria, where the organisms degrade polysaccharide substrates as forms of energy. Consequently, by reproducing these enzymes in the laboratory and studying their structures and functions, it is possible to learn from nature and be able to reproduce our own energy in a similar way.

6.2 Conclusions of the work in this thesis

Currently, the lytic polysaccharide monooxygenases are classified into four 'Auxiliary Activity' classes (AA9, AA10, AA11 and AA13) according to the CAZy database.[46, 55] The research into these LPMO enzymes presented throughout this work has been on a bacterial AA10 enzyme; an AA10 from *Bacillus amyloliquefaciens* in particular. By investigating the enzyme's structure and activity, insights into this bacterial LPMO class have be achieved, providing us with vital information which can be applied and used towards the potential use of such enzymes within the 'cocktail of enzymes' used to produce biofuels from lignocellulosic biomass. Here, I have outlined the methods of protein production used to obtain the *Ba*AA10 enzyme, including both periplasmic secretion and intracellular production methods. In addition, cloning methods to include a SUMO-tag and subsequent purification methods are outlined, which managed to produce increased yields of the protein.

The stability of *Ba*AA10 was investigated to confirm copper as the metal ion in the active site of the enzyme. This was achieved by isothermal titration calorimetry (ITC) and a more accurate measurement for the binding constant K_D of copper to the protein was obtained using displacement ITC methods with zinc as the weaker-binding competitive metal. A more accurate K_D was required due to the extremely tight binding 'curves' produced from titrating copper into the protein in previous studies,[72] resulting in difficulties obtaining an accurate gradient for the curve and hence an accurate binding constant. Following on, from ITC studies, thermostability studies were carried out, which demonstrated a large increase in melting temperature for *Ba*AA10 when copper is bound to the protein, indicating the metal is able to stabilise the protein.

Activity studies carried out on *Ba*AA10 resulted in chitin being confirmed as the substrate for the enzyme, which had been expected due to prior research into the enzyme's sequence homologs, suggesting that it was a chitin binding protein (before LPMOs were discovered to be enzymatic).[112] The activity of the enzyme was investigated by overnight substrate reactions using ascorbate as a source of electrons, followed by analysis and elucidation of the products obtained by Matrix Assisted Laser Desorption Ionisation Time-Of-Flight (MALDI-TOF) Mass Spectrometry. This confirmed the presence of oxidative products, both aldonic acids and lactones, due to the action of the lytic polysaccharide monooxygenase at the C1 position along the chitin chain of both α and β forms of chitin. Notably, the enzyme requires a crystalline substrate (it is not active on soluble chito-oligosaccharides), but oddly it does not seem to discriminate greatly between parallel or antiparallel crystals. In addition, thermostability studies carried out with the addition of chitin resulted in an increase in melting temperature, indicating that the stability of the protein increases when its substrate is bound.

Finally, X-ray crystallography was used to further probe the structure of *Ba*AA10, which had been solved prior to my research by PDRA Glyn Hemsworth.[72] This original structure included copper which had been photoreduced in the X-ray beam from Cu(II) to Cu(I), making it difficult to correlate the Cu structure with spectroscopic observations of the Cu(II) form. By optimising crystals using various techniques, such as seeding, larger crystals of *Ba*AA10 were obtained, which could then be used for spiral data collection techniques at the synchrotron in an attempt to reduce the amount of radiation damage by the X-ray beam. In this way, a Cu(II) structure of *Ba*AA10 was obtained, clearly showing the distorted trigonal bipyramidal geometry around the copper ion, proposed by previous studies conducted by electron paramagnetic resonance (EPR) spectroscopy.[72] This structure is extremely beneficial towards understanding the mechanism of action of the AA10 enzymes, as the knowledge that the geometry at the active site changes very little between the Cu(II) and the reduced Cu(I) states can help us to understand how the enzymes might perform their redox reactions. Finally, a Cu(II) structure has been obtained that allows EPR to be correlated correctly with 3D structure and atomic positions. Such work can now allow major new directions for the future.

6.3 Challenges for the future

Perhaps the major point that needs to be addressed for all LPMOs is their actual mechanism, or likely mechanisms, of action. Our future approach will be to study the copper oxygen species produced throughout the mechanistic pathway, and ultimately to combine that with theory, through density functional theory (DFT) calculations, to determine the electronic states by quantum mechanical modelling. An initial start on that was provided by Kjaergaard *et al.* in their studies on an AA9 enzyme from *Thermoascus aurantiacus*, which showed that a copper-superoxide species was thermodynamically favourable in the active site of the enzyme.[103] In addition, Kim *et al.*[144] carried out quantum mechanical calculations on both the mechanism of action originally put forward by Phillips *et al.*,[63] involving a copper-superoxo species and a separate hydroperoxo species involved in hydrogen abstraction and substrate hydroxylation, respectively; as well as another mechanism consisting of a copper-oxyl species followed by hydroxyl transfer via an oxygen rebound mechanism. Their results showed that oxygen binds end on (η^1) to the copper, and that the copper(II)-oxyl reaction mechanism pathway is energetically more favourable due to the lower energy barriers calculated. However, thermodynamics for the electron transfer steps has not been determined, providing challenges in the proof of this proposed pathway. An alternative method in the quest to determine the oxygen species and mechanism of action of the AA10 enzymes would be to obtain an X-ray structure with oxygen species bound, or an oxygen mimic such as azide or cyanide. In parallel

studies to X-ray crystallography, spectroscopic methods using oxygen or its mimics could be carried out, for example in EPR studies, to further investigate the copper environment when oxygen is bound.

In the particular case of AA10, in addition to structures and spectra with oxygen or oxygen mimics bound, structures and spectra with the polysaccharide substrate of the AA10 would assist in developing the knowledge of how the substrate binds and how the copper-oxygen species is able to attack the polysaccharide at the C1 position. Preliminary EPR experiments using chitin as the substrate for *Ba*AA10 have been carried out as a "proof of principle" prior to writing this thesis (the results of which are shown in Figure 6.3.1).

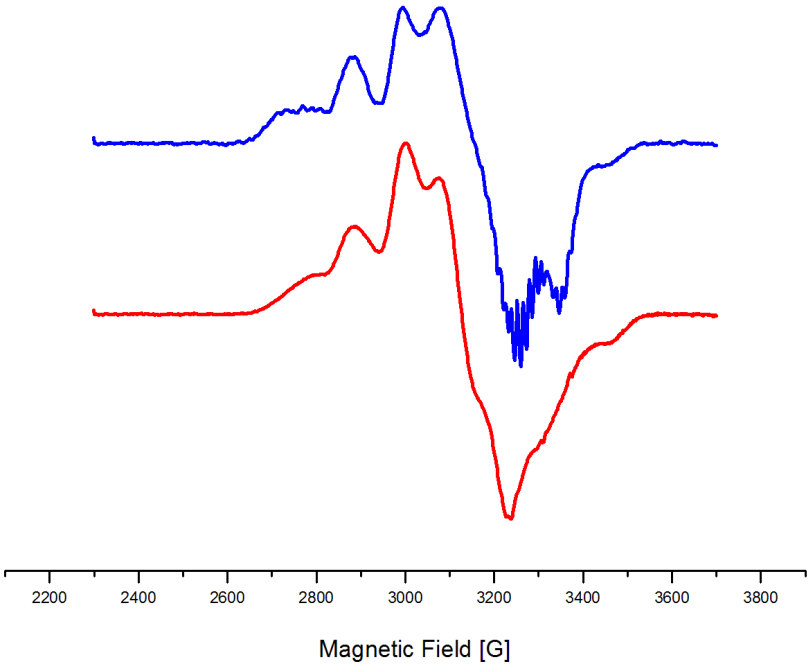

Figure 6.3.1 X-band EPR spectra of *Ba*AA10 (red/bottom spectrum) and *Ba*AA10 + squid pen chitin (blue/top spectrum). Both samples contained 10 % glycerol and were measured at pH 5 and 160 K.

One can see induction of "superhyperfine coupling" in the perpendicular region of the spectrum, similar to that recently reported upon the introduction of oligosaccharide substrates to *Ls*AA9 in the results published by Frandsen *et al.* (Figure 6.3.2B).[71] As this AA9 was shown to be active on soluble oligosaccharides, such as cellotriose and cellohexaose, a structure of the

substrate bound enzyme complex could be compared to the EPR spectra obtained when an oligosaccharide is present.

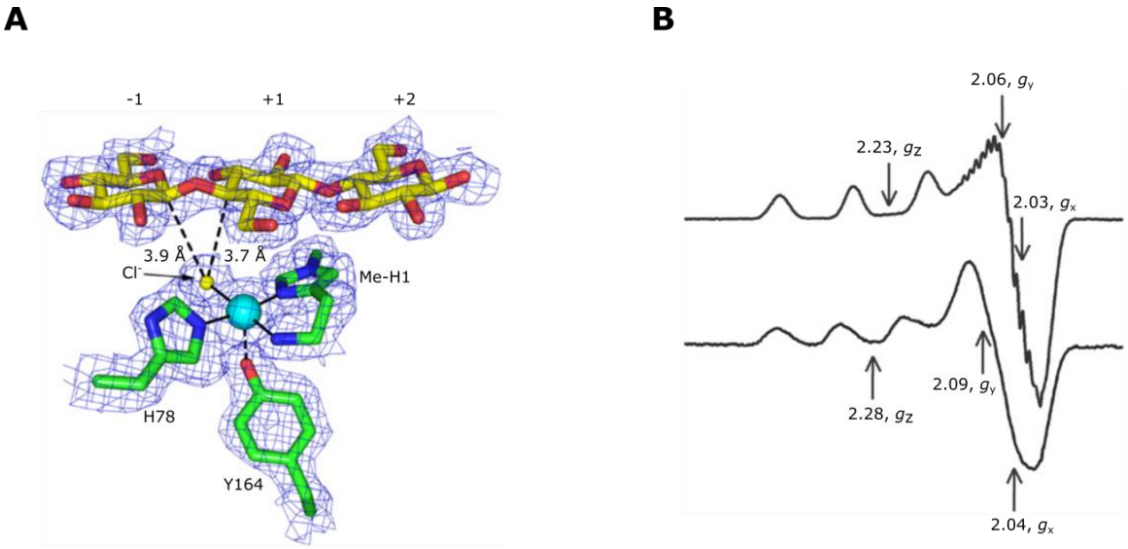

Figure 6.3.2 (A) 3D structure of the copper active site of *Ls*AA9 showing the electron density of the copper, its coordinating ligands and the cellotriose substrate bound on the surface of the enzyme. A chloride ion can be seen binding in a pocket between the substrate and enzyme. (B) X-band EPR spectra of *Ls*AA9 carried out in a chloride-rich with (top) and without (bottom) cellohexaose bound as a substrate.[71]

The EPR results for the *Ba*AA10 have not been analysed completely, for example the chloride effect has not been studied further, as in Frandsen *et al.*,[71] whereby the synergy between chloride (an oxygen surrogate) binding and substrate binding was investigated using isothermal titration calorimetry. These studies, as well as the increased affinity for chloride binding after oligosaccharide binding, suggested that an oxygen intermediate, such as Cu(III)-hydroxide or Cu(II)-oxyl reactive species, will only be produced upon substrate binding. In addition, further studies of the *Ls*AA9 EPR spectra using ^{1}H-HYSCORE spectroscopy were carried out, which enabled the allocation of magnetically coupled protons in the structure to take place, hence determining the addition of NH_2 coupling to the copper in the substrate bound complex.[71] Using similar methods, the coordination around the copper ion when chitin is bound to *Ba*AA10 could be resolved, providing us with insight into the possible mechanistic pathway of this class of LPMO.

The major, so far insurmountable for any crystalline-substrate-active LPMO, drawback in obtaining a substrate bound AA10 complex has been the insolubility of the polysaccharide

substrate. If an AA10 LPMO can be found which binds to a soluble saccharide, or a method of making the insoluble substrates more soluble for co-crystallisations with the AA10, then that would be very beneficial in leading to a substrate bound structure. Indeed, Frandsen *et al.*[71] were able to achieve a substrate bound structure of *Ls*AA9 (Figure 6.3.2A), as this enzyme binds to smaller, more soluble oligosaccharides. In this way, substrate development and co-crystallisations would provide us with a more in depth look at how the substrate binds to the flat surface of the AA10 enzymes, and provide insight into the method of action of these enzymes. In addition, the use of soluble substrates would allow for the development of facile kinetic experiments for LPMO enzymes, in the same way that energy-transfer substrates have been used for *Ls*AA9.[71] The lack in kinetics has been one of the main problems in analysing these enzymes, due to the lack of soluble substrates available to be studied. Indeed, some have even questioned the inclusion of 'lytic' in their title as a result of the absence of any actual data showing the enzyme does actually catalyse glycosidic bond cleavage.[61] Therefore, if the rates of reaction of the LPMOs can be determined through the use of more soluble substrates, then this would be a great step forward in our knowledge of these enzymes.

An additional mystery still to be solved within the research into lytic polysaccharide monooxygenases is the methylated N-terminal histidine residue often found within AA9 enzymes,[38, 63] as well as in other fungal LPMO enzymes such as in the AA13 class.[57] It has been proposed by Frandsen *et al.* that this methyl group on the N-terminal histidine in the *Ls*AA9 structure aids the stability of the lone pair to π^* electrostatic interaction[93] occurring between the ring oxygen of a glycosyl unit and the imidazole face of the N-terminal histidine residue.[71] As this interaction could potentially be applied to AA10 enzymes as well, it is therefore unclear whether or not the AA10 species requires the N-terminal histidine to be methylated to assist the mechanism of substrate binding and degradation. In order to prove the significance of the methylated histidine, comparisons would have to be made between methylated and non-methylated versions of the same LPMO from the same organism, by expressing the enzyme in a non-methylating host such as *Pichia pastoris*. It is known from previous studies that the methylated histidine may not be essential for activity, as non-methylated versions of AA9 enzymes remain active.[37, 65] However, structures with soluble oligosaccharides bound to enzymes with and without the methylated histidine could provide us with evidence on its purpose within the mechanism of substrate degradation.

In the long term, the lytic polysaccharide monooxygenases have great potential to be used in the biotechnology industry within a cocktail of enzymes for the degradation and conversion of second generation biomass materials into bioethanol. The inclusion of such enzymes can provide a more efficient method of polysaccharide deconstruction, as they increase the activity of classical degradation enzymes by their prior oxidative cleavage of the polysaccharide

substrate. An issue arises, however, regarding the synergy between these lytic polysaccharide monooxygenases and the more classical hydrolases. Research should be carried out towards obtaining the most efficient and cost effective mix of classical and LPMO enzyme used in the biofuel production processes. Furthermore, the LPMO enzymes are still generally new, with the AA13 class being discovered and characterised as recently 2015.[57] Therefore, it is very likely that further families of lytic polysaccharide monooxygenases will be discovered in the future, which should be analysed and compared to the four current classes. Indeed, the variety in LPMO enzymes found in nature in general poses an important question as to why so many different classes are required, especially within very similar fungal environments. For example, the fungus *Aspergillus oryzae* has been shown to possess both AA11[56] and AA13[57] LPMOs, and many fungi also contain a large number of LPMO enzymes. The question as to why so many are needed within nature and their uses still needs to be answered. Further unknowns exist within the scope of the lytic polysaccharide monooxygenases, such as the presence of an unusual arginine residue as opposed to the highly conserved N-terminal histidine residue.[37, 145] The purpose of the arginine terminated LPMOs has currently not been fully studied, demonstrating the range of research still to be carried out on these enzymes.

The role of carbohydrate binding modules to assist the degradation of polysaccharides by lytic polysaccharide monooxygenases also requires additional investigation. Most LPMO enzymes are found in nature with a non-catalytic carbohydrate binding module (CBM) attached to assist in the binding of the LPMO to its polysaccharide substrate. However, the extent of their role and synergy with specific enzymes has not been well studied to date. Studies have been carried out to determine the effect on activity following deletion of the CBM attached to the LPMO, including one by Forsberg *et al.*, who showed that the removal of the CBM from AA10 enzymes resulted in a 2-fold reduction in the oxidative activity of the LPMO on cellulose (PASC).[85] A recent study by Crouch *et al.*[107] involved the removal of CBMs from AA10 enzymes and the substitution of CBMs from other proteins. Results of this research showed that different CBMs were able to either increase the likelihood of polysaccharide degradation or inhibit the activity of the LPMO enzymes altogether, suggesting enzyme specificity and the need for determination of the correct CBM protein to work with each LPMO enzyme. Furthermore, substrate specificity was defined, enabling the possibility for tailoring different CBM proteins to LPMOs to obtain specific substrate activity. These studies show the importance of carbohydrate binding modules for efficient catalytic activity of lytic polysaccharide monooxygenases on polysaccharide substrates, as they assist the LPMO in coming into prolonged contact with its substrate, allowing for efficient oxidative action to occur.

In addition to regular carbohydrate binding modules, other modular enzymes could be used to promote the activity of the lytic polysaccharide monoxygenases, such as multi-enzyme complexes known as cellulosomes.[146-149] These multi-modular enzymes are found in anaerobic organisms and consist of a non-catalytic scaffoldin subunit which binds to insoluble substrates by a carbohydrate binding module (Figure 6.3.3). The scaffoldin also has cohesion modules which facilitate the binding of specific catalytic subunits via complementary binding modules known as dockerins.[150]

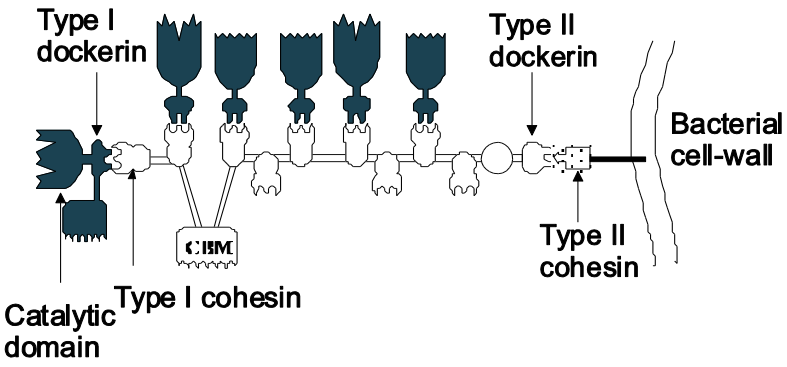

Figure 6.3.3 Representation of a multi-modular cellulosome complex, illustrating the proximity of the various catalytic and non-catalytic domains, which are able to bind and subsequently degrade insoluble polysaccharide substrates. (Image provided to Professor Gideon Davies by Professor Ed Bayer of the Weizmann Institute, Israel).

The proximity of these modules and synergy with the catalytic enzymes involved in the degradation of the insoluble substrate provide the cellulosomal system with great advantages regarding polysaccharide degradation. As lytic polysaccharide monooxygenases are not currently known to be found in anaerobic systems, they cannot benefit from the use of cellulosomes. However, the interlocking modular nature of the cellulosome complexes enables the incorporation of enzymes from other organisms, potentially including LPMOs. Indeed, Arfi *et al.*[151] were able to recombinantly incorporate LPMOs into "designer cellulosomes" by appending dockerin domains at their C-termini. These cellulosomal complexes, with LPMOs incorporated, showed a significant increase in the degradation of cellulose (2.6-fold enhancement) compared to in the absence of LPMOs, and furthermore their assembly into a complex showed additional benefit compared to using the same enzymes without complex assembly. Hopefully this research is just the beginning in the aim to make designer multi-modular oxidative systems using cellulosomes and tailored CBMs, creating an exciting direction for the LPMO enzymes both academically and for industrial applications in the future.

Appendix 1

Ba*AA10 Codon Optimised Gene Sequence:*

ATGAAATACCTGCTGCCGACGGCTGCTGCTGGTCTGCTGCTGCTGGCTGCCCAACC
GGCTATGGCGCATGGCTACATCAAGGAACCGGTCTCCCGTGCCTATATGGGCGCAC
TGGAAAAACAAACCATGGGTTGGACGGCGGCCGCACAGAAGTACGGCAGTGTTAT
TGATAATCCGCAATCCGTCGAAGGTCCGAAAGGTTTTCCGGCTGCAGGTCCGCCGG
ACGGTCGTATTGCCAGCGCAAACGGCGGTTCTGGCCAGATCGATTTCGGTCTGGAC
AAACAAACCGCCGATCATTGGGTTAAGCAGAATATTCGCGGCGGTTTTAACACCTT
CACGTGGCATTATACCGCTCCGCACGCGACGAGCAAATGGCACTATTACATCACCA
AAAAGAACTGGAATCCGAACAAGCCGCTGTCACGTGATGAATTTGAACTGATTGG
CACGGTGAATCATGATGGTTCGAAAGCGGACACCAACCTGACGCACAAGATCTTC
GTTCCGACCGATCGCAGCGGCTATCATATTATCCTGGGTGTGTGGGATGTTGCTGA
CACCTCTAATGCGTTTTACAACGTCATCGACGTGAATCTGACGAAATAA

Appendix 2

The peer-reviewed publications published during the course of this thesis work are included and listed below in order of publication date.

- Hemsworth G.R., Taylor E.J., Kim R.Q., Gregory R.C., Lewis S.J., Turkenburg J.P., Parkin A., Davies G.J., Walton P.H. (2013) The copper active site of CBM33 polysaccharide oxygenases. *J Am Chem Soc* 135(16):6069-6077.

- Ribeiro L.F.C., De Lucas R.C., Vitcosque G.L., Ribeiro L.F., Ward R.J., Rubio M.V., Damásio A.R.L., Squina F.M., Gregory R.C., Walton P.H., Jorge J.A., Prade R.A., Buckeridge M.S., Polizeli M.T.M..(2014) A novel thermostable xylanase GH10 from Malbranchea pulchella expressed in Aspergillus nidulans with potential applications in biotechnology. *Biotechnology for biofuels* 7(1):1-11.

- Gregory R.C., Hemsworth G.R., Turkenburg J.P., Hart S.J., Walton P.H., Davies G.J. (2016) Activity, stability and 3-D structure of the Cu(II) form of a chitin-active lytic polysaccharide monooxygenase from *Bacillus amyloliquefaciens*. *Dalton Transactions* (Published as an Accepted Manuscript).

List of abbreviations

µM	micro-Molar
AA	Auxiliary Activity
An	*Aspergillus nidulans*
Ao	*Aspergillus oryzae*
Asp	Aspartate
Ba	Bacillus amyloliquefaciens
CalMix1	Calibration Mix Solution 1
CAZy	Carbohydrate-active enzymes database
CBM	Carbohydrate binding module
CBP	Chitin-binding protein
CC	Correlation coefficient
CCP4	Collaborative Computational Project No. 4
CDH	Cellobiose dehydrogenase
CV	Column volume
DFT	Density functional theory
DHB	2, 5 – dihydroxybenzoic acid
DLS	Diamond Light Source
DNA	Deoxyribonucleic acid
DP	Degree of polymerisation
DSC	Differential Scanning Calorimetry
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid

Ef	Enterococcus faecalis
EPR	Electron paramagnetic resonance
ESI-TOF MS	Electrospray ionisation – time of flight mass spectrometry
ESPript	Easy sequencing in PostScript
eV	electronvolt
EXAFS	Extended X-ray absorption fine structure
ExPASy	Expert protein analysis system
FAD	Flavin adenine dinucleotide
FnIII	Fibronectin type III
GF	Gel filtration
GH	Glycoside hydrolase
GHz	Gigahertz
His	Histidine
HPAEC	High-performance anion-exchange chromatography
HYSCORE	Hyperfine sub-level correlation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K	Kelvin
K_D	Binding constant
kDa	kilo-Dalton
LB	Lysogeny broth
LPMO	Lytic polysaccharide monooxygenase
Ls	*Lentinus similis*
m/z	mass/charge

MALDI-TOF	Matrix-assisted laser desorption/ionisation – time of flight
MES	2-(*N*-morpholino)ethanesulfonic acid
mM	milli-Molar
MMT	DL-Malic acid, MES, Tris
NCBI	National Center for Biotechnology Information
nM	nano-Molar
NMR	Nuclear magnetic resonance
OD	Optical density
ORF	Optical reading frame
Pa	Pascal
PACT screen	pH, anion and cation testing screen
PASC	Phosphoric acid swollen cellulose
PB	Peisach-Blumberg
PCR	Polymerase chain reaction
PDB	Protein data bank
PDRA	Post-doctoral research associate
PEG	Polyethylene glycol
pMMO	particulate Methane Monooxygenase
qPCR	quantitative Polymerase Chain Reaction
RMSD	Root mean square deviation
rpm	revolutions per minute
S	Spin
SDS-PAGE	Sodium dodecyl sulphate - poly acrylamide gel electrophoresis

SEC	Size-exclusion chromatography
SEC-MALLS	Size-exclusion chromatography - Multi-angle laser light scattering
SOMO	Singly occupied molecular orbital
SUMO	Small ubiquitin-like modifier
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
Tyr	Tyrosine
UV-vis	Ultraviolet-visible spectroscopy
v/v	Volume per volume
XANES	X-ray absorption near edge structure
XDS	X-ray detector software

References

1. Daunton MJ (1995) *Progress and poverty: An economic and social history of Britain 1700-1850* (Oxford University Press).

2. Simpson-Holley M, Higson A, & Evans G (2007) Renewables: Bring on the biorefinery. *Chemical Engineer* (795):46-49.

3. Naik SN, Goud VV, Rout PK, & Dalai AK (2010) Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews* 14(2):578-597.

4. Zegada-Lizarazu W & Monti A (2011) Energy crops in rotation. A review. *Biomass and Bioenergy* 35(1):12-25.

5. Tenenbaum DJ (2008) Food vs. fuel: diversion of crops could cause more hunger. *Environ Health Perspect* 116(6):A254-257.

6. Kumar P, Barrett DM, Delwiche MJ, & Stroeve P (2009) Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. *Industrial & Engineering Chemistry Research* 48(8):3713-3729.

7. Sticklen MB (2008) Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nature reviews. Genetics* 9(6):433-443.

8. Rubin EM (2008) Genomics of cellulosic biofuels. *Nature* 454(7206):841-845.

9. Bowman SM & Free SJ (2006) The structure and synthesis of the fungal cell wall. *BioEssays : news and reviews in molecular, cellular and developmental biology* 28(8):799-808.

10. Sanderson K (2011) Lignocellulose: A chewy problem. *Nature* 474(7352):S12-14.

11. Wolfenden R, Lu X, & Young G (1998) Spontaneous Hydrolysis of Glycosides. *Journal of the American Chemical Society* 120(27):6814-6815.

12. Klemm D, Heublein B, Fink HP, & Bohn A (2005) Cellulose: fascinating biopolymer and sustainable raw material. *Angewandte Chemie* 44(22):3358-3393.

13. Gelfand I, *et al.* (2013) Sustainable bioenergy production from marginal lands in the US Midwest. *Nature* 493(7433):514-517.

14. Kauffman GB (1993) Rayon: The first semi-synthetic fiber product. *Journal of Chemical Education* 70(11):887.

15. Stankiewicz BA, Briggs DEG, Evershed RP, Flannery MB, & Wuttke M (1997) Preservation of Chitin in 25-Million-Year-Old Fossils. *Science* 276(5318):1541-1543.

16. Kurita K (2001) Controlled functionalization of the polysaccharide chitin. *Progress in Polymer Science* 26(9):1921-1971.

17. Cosio IG, Fisher RA, & Carroad PA (1982) Bioconversion of Shellfish Chitin Waste: Waste Pretreatment, Enzyme Production, Process Design, and Economic Analysis. *Journal of Food Science* 47(3):901-905.

18. Revah-Moiseev S & Carroad PA (1981) Conversion of the enzymatic hydrolysate of shellfish waste chitin to single-cell protein. *Biotechnology and bioengineering* 23(5):1067-1078.

19. Shahidi F, Arachchi JKV, & Jeon Y-J (1999) Food applications of chitin and chitosans. *Trends in Food Science & Technology* 10(2):37-51.

20. Shao K, *et al.* (2016) Fabrication and feasibility study of an absorbable diacetyl chitin surgical suture for wound healing. *Journal of Biomedical Materials Research Part B: Applied Biomaterials* 104(1):116-125.

21. Spiegel Y, Cohn E, & Chet I (1986) Use of chitin for controlling plant parasitic nematodes. *Plant and Soil* 95(1):87-95.

22. Arnau J, Lauritzen C, Petersen GE, & Pedersen J (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification* 48(1):1-13.

23. Reese ET, Siu RG, & Levinson HS (1950) The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *Journal of bacteriology* 59(4):485-497.

24. Carroad PA & Tom RA (1978) Bioconversion of shellfish chitin wastes: Process conception and selection of microorganisms. *Journal of Food Science* 43(4):1158-1161.

25. Tom RA & Carroad PA (1981) Effect of Reaction Conditions on Hydrolysis of Chitin by *Serratia marcescens* QMB 1466 chitinase. *Journal of Food Science* 46(2):646-647.

26. Levinson HS, Mandels GR, & Reese ET (1951) Products of enzymatic hydrolysis of cellulose and its derivatives. *Archives of biochemistry and biophysics* 31(3):351-365.

27. Horn SJ, Vaaje-Kolstad G, Westereng B, & Eijsink VG (2012) Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels* 5(1):45.

28. Vane CH, Drage TC, & Snape CE (2003) Biodegradation of Oak (Quercus alba) Wood during Growth of the Shiitake Mushroom (*Lentinula edodes*): A Molecular Approach. *Journal of Agricultural and Food Chemistry* 51(4):947-956.

29. Westermark U, Eriksson, K. E. (1974) Cellobiose:Quinone Oxidoreductase, A New Wood-degrading Enzyme From White-rot Fungi. *Acta Scandinavica B* 28:209-214.

30. Eriksson KE, Pettersson B, & Westermark U (1974) Oxidation: an important enzyme reaction in fungal degradation of cellulose. *FEBS letters* 49(2):282-285.

31. Tan T-C, *et al.* (2015) Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nature communications* 6.

32. Canam T, Town JR, Tsang A, McAllister TA, & Dumonceaux TJ (2011) Biological pretreatment with a cellobiose dehydrogenase-deficient strain of *Trametes versicolor* enhances the biofuel potential of canola straw. *Bioresource technology* 102(21):10020-10027.

33. Bao W & Renganathan V (1992) Cellobiose oxidase of *Phanerochaete chrysosporium* enhances crystalline cellulose degradation by cellulases. *FEBS letters* 302(1):77-80.

34. Vaaje-Kolstad G, *et al.* (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330(6001):219-222.

35. Karkehabadi S, *et al.* (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from *Hypocrea jecorina*, at 1.6 A resolution. *Journal of molecular biology* 383(1):144-154.

36. Forsberg Z, *et al.* (2011) Cleavage of cellulose by a CBM33 protein. *Protein Science* 20(9):1479-1483.

37. Wu M, *et al.* (2013) Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the Basidiomycota fungus *Phanerochaete chrysosporium*. *J. Biol. Chem.* 288(18):12828-12839.

38. Quinlan RJ, *et al.* (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci U S A* 108(37):15079-15084.

39. Harris PV, *et al.* (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49(15):3305-3316.

40. Vaaje-Kolstad G, Horn SJ, van Aalten DM, Synstad B, & Eijsink VG (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *J. Biol. Chem.* 280(31):28492-28497.

41. Vaaje-Kolstad G, Houston DR, Riemen AH, Eijsink VG, & van Aalten DM (2005) Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J. Biol. Chem.* 280(12):11313-11319.

42. Eibinger M, *et al.* (2014) Cellulose surface degradation by a lytic polysaccharide monooxygenase and its effect on cellulase hydrolytic efficiency. *J. Biol. Chem.* 289(52):35929-35938.

43. Vermaas JV, Crowley MF, Beckham GT, & Payne CM (2015) Effects of Lytic Polysaccharide Monooxygenase Oxidation on Cellulose Structure and Binding of Oxidized Cellulose Oligomers to Cellulases. *The Journal of Physical Chemistry B* 119(20):6129-6143.

44. Nakagawa YS, *et al.* (2015) A small lytic polysaccharide monooxygenase from *Streptomyces griseus* targeting alpha- and beta-chitin. *The FEBS journal* 282(6):1065-1079.

45. Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* 280 (Pt 2):309-316.

46. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42(Database issue):D490-495.

47. Henrissat B & Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *The Biochemical journal* 293 (Pt 3):781-788.

48. Henrissat B & Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. *The Biochemical journal* 316 (Pt 2):695-696.

49. Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* 3(9):853-859.

50. Henrissat B & Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Current opinion in structural biology* 7(5):637-644.

51. Campbell JA, Davies GJ, Bulone V, & Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *The Biochemical journal* 326 (Pt 3):929-939.

52. Coutinho PM, Deleury E, Davies GJ, & Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *Journal of molecular biology* 328(2):307-317.

53. Lombard V, *et al.* (2010) A hierarchical classification of polysaccharide lyases for glycogenomics. *The Biochemical journal* 432(3):437-444.

54. Boraston AB, Bolam DN, Gilbert HJ, & Davies GJ (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical journal* 382(Pt 3):769-781.

55. Levasseur A, Drula E, Lombard V, Coutinho PM, & Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* 6(1):41.

56. Hemsworth GR, Henrissat B, Davies GJ, & Walton PH (2014) Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature chemical biology* 10(2):122-126.

57. Lo Leggio L, *et al.* (2015) Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature communications* 6:5961.

58. Dimarogona M, Topakas E, & Christakopoulos P (2012) Cellulose degradation by oxidative enzymes. *Computational and Structural Biotechnology Journal* 2(3):1-8.

59. Harris PV, Xu F, Kreel NE, Kang C, & Fukuyama S (2014) New enzyme insights drive advances in commercial ethanol production. *Current opinion in chemical biology* 19:162-170.

60. Span EA & Marletta MA (2015) The framework of polysaccharide monooxygenase structure and chemistry. *Current opinion in structural biology* 35:93-99.

61. Beeson WT, Vu VV, Span EA, Phillips CM, & Marletta MA (2015) Cellulose Degradation by Polysaccharide Monooxygenases. *Annu Rev Biochem* 84:923-946.

62. Hemsworth GR, Johnston EM, Davies GJ, & Walton PH (2015) Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol* 33(12):747-761.

63. Phillips CM, Beeson WT, Cate JH, & Marletta MA (2011) Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS chemical biology* 6(12):1399-1406.

64. Beeson WT, Phillips CM, Cate JH, & Marletta MA (2012) Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. *J Am Chem Soc* 134(2):890-892.

65. Bey M, *et al.* (2013) Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from *Podospora anserina*. *Appl Environ Microbiol* 79(2):488-496.

66. Isaksen T, *et al.* (2014) A C4-oxidizing lytic polysaccharide monooxygenase cleaving both cellulose and cello-oligosaccharides. *J. Biol. Chem.* 289(5):2632-2642.

67. Bennati-Granier C, *et al.* (2015) Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. *Biotechnology for biofuels* 8:90.

68. Agger JW, *et al.* (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proc Natl Acad Sci U S A* 111(17):6287-6292.

69. Kojima Y, *et al.* (2016) Characterization of an LPMO from the brown-rot fungus *Gloeophyllum trabeum* with broad xyloglucan specificity, and its action on cellulose-xyloglucan complexes. *Appl Environ Microbiol*.

70. Nekiunaite L, *et al.* (2016) *Fg*LPMO9A from *Fusarium graminearum* cleaves xyloglucan independently of the backbone substitution pattern. *FEBS letters*.

71. Frandsen KEH, *et al.* (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nature chemical biology* 12(4):298-303.

72. Hemsworth GR, *et al.* (2013) The copper active site of CBM33 polysaccharide oxygenases. *J Am Chem Soc* 135(16):6069-6077.

73. Forsberg Z, *et al.* (2014) Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* 53(10):1647-1656.

74. Zhang H, Zhao Y, Cao H, Mou G, & Yin H (2015) Expression and characterization of a lytic polysaccharide monooxygenase from *Bacillus thuringiensis. Int J Biol Macromol* 79:72-75.

75. Gardner JG, *et al.* (2014) Systems biology defines the biological significance of redox-active proteins during cellulose degradation in an aerobic bacterium. *Molecular microbiology*.

76. Forsberg Z, *et al.* (2016) Structural and Functional Analysis of a Lytic Polysaccharide Monooxygenase Important for Efficient Utilization of Chitin in *Cellvibrio japonicus*. *J. Biol. Chem.*

77. Vaaje-Kolstad G, *et al.* (2012) Characterization of the Chitinolytic Machinery of *Enterococcus faecalis* V583 and High-Resolution Structure of Its Oxidative CBM33 Enzyme. *Journal of molecular biology* 416(2):239-254.

78. Gudmundsson M, *et al.* (2014) Structural and electronic snapshots during the transition from a Cu(II) to Cu(I) metal center of a lytic polysaccharide monooxygenase by X-ray photoreduction. *J. Biol. Chem.* 289(27):18782-18792.

79. Ghatge SS, *et al.* (2015) Multifunctional cellulolytic auxiliary activity protein *Hc*AA10-2 from *Hahella chejuensis* enhances enzymatic hydrolysis of crystalline cellulose. *Applied microbiology and biotechnology* 99(7):3041-3055.

80. Mekasha S, *et al.* (2016) Structural and functional characterization of a small chitin-active lytic polysaccharide monooxygenase domain of a multi-modular chitinase from *Jonesia denitrificans. FEBS letters* 590(1):34-42.

81. Vaaje-Kolstad G, Bunaes AC, Mathiesen G, & Eijsink VG (2009) The chitinolytic system of *Lactococcus lactis* ssp. lactis comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of alpha- and beta-chitin. *The FEBS journal* 276(8):2402-2415.

82. Aachmann FL, Sorlie M, Skjak-Braek G, Eijsink VG, & Vaaje-Kolstad G (2012) NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proc Natl Acad Sci U S A* 109(46):18779-18784.

83. Suzuki K, Suzuki M, Taiyoji M, Nikaidou N, & Watanabe T (1998) Chitin Binding Protein (CBP21) in the Culture Supernatant of *Serratia marcescens* 2170. *Bioscience, Biotechnology, and Biochemistry* 62(1):128-135.

84. Nakagawa YS, Eijsink VG, Totani K, & Vaaje-Kolstad G (2013) Conversion of α-chitin substrates with varying particle size and crystallinity reveals substrate preferences of the chitinases and lytic polysaccharide monooxygenase of *Serratia marcescens*. *J Agric Food Chem* 61(46):11061-11066.

85. Forsberg Z, *et al.* (2014) Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases. *Proc Natl Acad Sci U S A* 111(23):8446-8451.

86. Chaplin AK, *et al.* (2016) Heterogeneity in the histidine-brace copper coordination sphere in AA10 lytic polysaccharide monooxygenases. *J. Biol. Chem.*

87. Moser F, Irwin D, Chen S, & Wilson DB (2008) Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module proteins E7 and E8. *Biotechnology and bioengineering* 100(6):1066-1077.

88. Loose JS, Forsberg Z, Fraaije MW, Eijsink VG, & Vaaje-Kolstad G (2014) A rapid quantitative activity assay shows that the *Vibrio cholerae* colonization factor GbpA is an active lytic polysaccharide monooxygenase. *FEBS letters* 588(18):3435-3440.

89. Culpepper MA, Cutsail GE, 3rd, Hoffman BM, & Rosenzweig AC (2012) Evidence for oxygen binding at the active site of particulate methane monooxygenase. *J Am Chem Soc* 134(18):7640-7643.

90. Balasubramanian R & Rosenzweig AC (2007) Structural and mechanistic insights into methane oxidation by particulate methane monooxygenase. *Accounts of chemical research* 40(7):573-580.

91. Smith SM, *et al.* (2011) Crystal structure and characterization of particulate methane monooxygenase from *Methylocystis* species strain M. *Biochemistry* 50(47):10231-10240.

92. Li X, Beeson WTt, Phillips CM, Marletta MA, & Cate JH (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* 20(6):1051-1061.

93. Pavlakos I, *et al.* (2015) Noncovalent Lone Pair(No-pi!)-Heteroarene Interactions: The Janus-Faced Hydroxy Group. *Angewandte Chemie* 54(28):8169-8174.

94. Vu VV, Beeson WT, Span EA, Farquhar ER, & Marletta MA (2014) A family of starch-active polysaccharide monooxygenases. *Proc Natl Acad Sci U S A* 111(38):13822-13827.

95. Yi J, Orville AM, Skinner JM, Skinner MJ, & Richter-Addo GB (2010) Synchrotron X-ray-Induced Photoreduction of Ferric Myoglobin Nitrite Crystals Gives the Ferrous Derivative with Retention of the O-Bonded Nitrite Ligand. *Biochemistry* 49(29):5969-5971.

96. Sigfridsson KGV, *et al.* (2013) Rapid X-ray Photoreduction of Dimetal-Oxygen Cofactors in Ribonucleotide Reductase. *Journal of Biological Chemistry* 288(14):9648-9661.

97. Peisach J & Blumberg WE (1974) Structural implications derived from the analysis of electron paramagnetic resonance spectra of natural and artificial copper proteins. *Archives of biochemistry and biophysics* 165(2):691-708.

98. Kooter IM, *et al.* (2002) EPR characterization of the mononuclear Cu-containing *Aspergillus japonicus* quercetin 2,3-dioxygenase reveals dramatic changes upon anaerobic binding of substrates. *European journal of biochemistry / FEBS* 269(12):2971-2979.

99. Hemsworth GR, Davies GJ, & Walton PH (2013) Recent insights into copper-containing lytic polysaccharide mono-oxygenases. *Current opinion in structural biology* 23(5):660-668.

100. Westereng B, *et al.* (2016) Simultaneous analysis of C1 and C4 oxidized oligosaccharides, the products of lytic polysaccharide monooxygenases acting on cellulose. *Journal of chromatography. A* 1445:46-54.

101. Courtade G, *et al.* (2016) Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences* 113(21):5922-5927.

102. Garajova S, *et al.* (2016) Single-domain flavoenzymes trigger lytic polysaccharide monooxygenases for oxidative degradation of cellulose. *Scientific Reports* 6:28276.

103. Kjaergaard CH, *et al.* (2014) Spectroscopic and computational insight into the activation of O_2 by the mononuclear Cu center in polysaccharide monooxygenases. *Proc Natl Acad Sci U S A* 111(24):8797-8802.

104. Solomon EI, *et al.* (2014) Copper active sites in biology. *Chem. Rev.* 114(7):3659-3853.

105. Morgenstern I, Powlowski J, & Tsang A (2014) Fungal cellulose degradation by oxidative enzymes: from dysfunctional GH61 family to powerful lytic polysaccharide monooxygenase family. *Briefings in functional genomics* 13(6):471-481.

106. Hu J, Arantes V, Pribowo A, Gourlay K, & Saddler JN (2014) Substrate factors that influence the synergistic interaction of AA9 and cellulases during the enzymatic hydrolysis of biomass. *Energy & Environmental Science* 7(7):2308-2315.

107. Crouch LI, Labourel A, Walton PH, Davies GJ, & Gilbert HJ (2016) The Contribution of Non-catalytic Carbohydrate Binding Modules to the Activity of Lytic Polysaccharide Monooxygenases. *The Journal of biological chemistry* 291(14):7439-7449.

108. Gregory RC, *et al.* (2016) Activity, stability and 3-D structure of the Cu(II) form of a chitin-active lytic polysaccharide monooxygenase from *Bacillus amyloliquefaciens*. *Dalton Transactions* (45):16904-16912.

109. Ribeiro LF, *et al.* (2014) A novel thermostable xylanase GH10 from *Malbranchea pulchella* expressed in *Aspergillus nidulans* with potential applications in biotechnology. *Biotechnology for biofuels* 7(1):1-11.

110. Vaaje-Kolstad G, Horn SJ, Sorlie M, & Eijsink VG (2013) The chitinolytic machinery of *Serratia marcescens*--a model system for enzymatic degradation of recalcitrant polysaccharides. *The FEBS journal* 280(13):3028-3049.

111. Fukumoto J (1943) Studies on the production of bacterial amylase. I. Isolation of bacteria secreting potent amylases and their distribution. *Journal of the Agricultural Chemical Society of Japan (in Japanese)* 19:487-503.

112. Chu HH, Hoang V, Hofemeister J, & Schrempf H (2001) A *Bacillus amyloliquefaciens* ChbB protein binds beta- and alpha-chitin and has homologues in related strains. *Microbiology* 147(Pt 7):1793-1803.

113. Panavas T, Sanders C, & Butt TR (2009) SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. *Methods in molecular biology (Clifton, N.J.)* 497:303-317.

114. Gasteiger E. HC, Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. *John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press*:571-607

115. Carrer C, *et al.* (2006) Removing coordinated metal ions from proteins: a fast and mild method in aqueous solution. *Analytical and bioanalytical chemistry* 385(8):1409-1413.

116. Sigurskjold BW (2000) Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Analytical biochemistry* 277(2):260-266.

117. Velazquez-Campoy A, Ohtaka H, Nezami A, Muzammil S, & Freire E (2004) Isothermal titration calorimetry. *Current protocols in cell biology* Chapter 17:Unit 17 18.

118. Velazquez-Campoy A & Freire E (2006) Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat Protoc* 1(1):186-191.

119. Wen J, Arakawa T, & Philo JS (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. *Analytical biochemistry* 240(2):155-166.

120. Murphy RM (1997) Static and dynamic light scattering of biological macromolecules: what can we learn? *Current opinion in biotechnology* 8(1):25-30.

121. Lo M-C, *et al.* (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Analytical biochemistry* 332(1):153-159.

122. Schulz MN, Landstrom J, & Hubbard RE (2013) MTSA--a Matlab program to fit thermal shift data. *Analytical biochemistry* 433(1):43-47.

123. Brandts JF & Lin LN (1990) Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry* 29(29):6927-6940.

124. Niesen FH, Berglund H, & Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2(9):2212-2221.

125. Sievers F, *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7(1).

126. Goujon M, *et al.* (2010) A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Research* 38(suppl 2):W695-W699.

127. Robert X & Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research* 42(W1):W320-W324.

128. Fan Y, Saito T, & Isogai A (2008) Preparation of chitin nanofibers from squid pen beta-chitin by simple mechanical treatment under acid conditions. *Biomacromolecules* 9(7):1919-1923.

129. Pham AN, Xing G, Miller CJ, & Waite TD (2013) Fenton-like copper redox chemistry revisited: Hydrogen peroxide and superoxide mediation of copper-catalyzed oxidant production. *Journal of Catalysis* 301:54-64.

130. Jang M-K, Kong B-G, Jeong Y-I, Lee CH, & Nah J-W (2004) Physicochemical characterization of α-chitin, β-chitin, and γ-chitin separated from natural resources. *Journal of Polymer Science Part A: Polymer Chemistry* 42(14):3423-3432.

131. Jee JG, *et al.* (2002) Solution structure of the fibronectin type III domain from *Bacillus circulans* WL-12 chitinase A1. *J. Biol. Chem.* 277(2):1388-1397.

132. Perrakis A, *et al.* (1994) Crystal structure of a bacterial chitinase at 2.3 A resolution. *Structure* 2(12):1169-1180.

133. Roberts SM & Davies GJ (2012) The crystallization and structural analysis of cellulases (and other glycoside hydrolases). *Cellulases, Methods in Enzymology*. (Elsevier Academic Press Inc.), pp 141-168.

134. Kabsch W (2010) Xds. *Acta crystallographica. Section D, Biological crystallography* 66(Pt 2):125-132.

135. Winn MD, *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* 67(Pt 4):235-242.

136. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* 53(Pt 3):240-255.

137. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 60(Pt 12 Pt 1):2126-2132.

138. Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030-1033.

139. Westereng B, *et al.* (2011) The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PloS one* 6(11):e27807.

140. BetaRenewables (2013) Here comes the green revolution. http://www.betarenewables.com/crescentino/green-revolution).

141. Wilson DB (2011) Microbial diversity of cellulose hydrolysis. *Current opinion in microbiology* 14(3):259-263.

142. Perez J, Munoz-Dorado J, de la Rubia T, & Martinez J (2002) Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *International microbiology : the official journal of the Spanish Society for Microbiology* 5(2):53-63.

143. Saloheimo M, Nakari-Setala T, Tenkanen M, & Penttila M (1997) cDNA cloning of a *Trichoderma reesei* cellulase and demonstration of endoglucanase activity by expression in yeast. *European journal of biochemistry / FEBS* 249(2):584-591.

144. Kim S, Stahlberg J, Sandgren M, Paton RS, & Beckham GT (2014) Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. *Proc Natl Acad Sci U S A* 111(1):149-154.

145. Yakovlev I, *et al.* (2012) Substrate-specific transcription of the enigmatic GH61 family of the pathogenic white-rot fungus *Heterobasidion irregulare* during growth on lignocellulose. *Applied microbiology and biotechnology* 95(4):979-990.

146. Bayer EA, Morag E, & Lamed R (1994) The cellulosome--a treasure-trove for biotechnology. *Trends Biotechnol* 12(9):379-386.

147. Shoham Y, Lamed R, & Bayer EA (1999) The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in microbiology* 7(7):275-281.

148. Ding SY, *et al.* (2008) A biophysical perspective on the cellulosome: new opportunities for biomass conversion. *Current opinion in biotechnology* 19(3):218-227.

149. Fontes CM & Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem* 79:655-681.

150. Stahl SW, *et al.* (2012) Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc Natl Acad Sci U S A* 109(50):20431-20436.

151. Arfi Y, Shamshoum M, Rogachev I, Peleg Y, & Bayer EA (2014) Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation. *Proc Natl Acad Sci U S A* 111(25):9109-9114.