Overexpression and structural studies on AA13 and AA10 LPMOs

M. L. A. Hanlon

A thesis submitted for the degree of Master of Science (by Dissertation)

in Biochemistry

School of Life Sciences

University of Essex

Date of submission for examination January 2023

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisor, Jonathan Worrall, who has continued to support and guide me through my project despite multiple delays. Secondly, I would like to thank my secondary supervisor, Mike Hough; post doc., Marina, who also helped guide me through my time as a post-graduate, and all the other School of Life Sciences staff that helped encourage me and push me forward during the stages of completion. I would also like to especially thank the Christine Desty Scholarship, who provided me with the opportunity to carry out this research. Finally, I thank my friends and family, who cheered me on continuously. I appreciate all the support I have been shown throughout my research.

List of Contents

List of Abbreviations

1G – 1st Generation

2G – 2nd Generation

A. oryzae – *Aspergillus oryzae*

*Ao*AA13 – *Aspergillus oryzae* AA13

AA – Auxiliary Activity

ASR – Ancestral Sequence Reconstruction

$CaCl_2$ – Calcium Chloride

CAZy – Carbohydrate-Active Enzymes

CBM – Carbohydrate Binding Module

CM – Chloramphenicol

Cu – Copper

Cys – Cystine

E. coli – *Escherichia coli*

GHs – Glycoside Hydrolases

His – Histidine

Kan – Kanamycin

LPMO – Lytic Polysaccharide Monooxygenase

MR – Molecular Replacement

N – Nitrogen

PDB – Protein Data Bank

Phe – Phenylalanine

RMS dev – Relative Mean Standard Deviation

S. lividans – *Streptomyces lividans*

*Sli*LPMO – *Streptomyces lividans* Lytic Polysaccharide Monooxygenase

Abstract

Lytic polysaccharide monooxygenases (LPMOs) are copper dependent metalloenzymes which oxidatively cleave recalcitrant polysaccharides. With the biofuel industry developing new ways of processing lignocellulosic biomass, LPMOs could play and important role in reducing production cost and increasing saccharification yield. This study reports the expression testing of two ancestral reconstructions, N231 and N237, from the starch degrading AA13 LPMO family and the characterisation of *Sli*LPMO10G from *Streptomyces lividans* in cellulolytic/chitinolytic AA10 LMPO family. N231 and N237 were both tested using the Tat and Sec signalling pathways and tested for expression in a variety of *Escherichia coli* cell lines. Results from the test expressions showed bands of expected mass for N231 with the Tat pathway in BL21-RIL and the Sec pathway in both BL21-DE3 and T7 Shuffle. However, no expression was seen for N237. *Sli*LPMO10G has a carbohydrate-binding module (CBM) from CBM5 appended to its LPMO domain via a linker. PCR was used to truncate the *Sli*LPMO10G and the LPMO domain, CBM5 domain and full *Sli*LPMO10G were all expressed and purified using ion exchange chromatography, immobilized metal affinity chromatography and gel filtration. Crystallographic trials were prepared for the CBM5 domain, whilst trials had previously been carried out for the LPMO domain. The structure of the *Sli*LPMO10G LPMO domain was analysed, and results indicated the same β-sandwich fold and Loop 2 region consistent with other AA10 structures. The conserved copper active site was also identified, with copper binding to the amino group Nt and side chain Nδ1 of the N-terminal histidine and the Nε2 of a secondary histidine side chain, in a T-shaped brace.

Chapter 1: Introduction

[1.1] – Biofuels

As worldwide economy and rapid industrialisation is driven forward, fossil fuels have become the primary energy source to generate electricity and power for domestic and industrial purposes (Nanda *et al.*, 2018). However, sustainable solutions are vital in order to lower the generation of environmental pollutants and emissions. Biofuels produced from renewable and biogenic materials have the capacity to mitigate greenhouse gas emissions, improve energy efficiency of existing fuel systems and supplement the growing energy needs (Nanda *et al.*, 2015). Unfortunately whilst increased production of biomass as a substitute for energy has the potential to offset the current dependency on fossil fuels, it also has the potential to risk sacrificing natural habitats and stimulate competition for land between biomass and agricultural farming (Field *et al.*, 2008; Simpson, 2009; Nanda *et al.*, 2015). Key factors should be considered for the prosperity of biomass usage as a global energy system, one of which is conversion technology and the prospects to use new plant and microbe varieties to improve upon the biomass-to-fuel process. This in turn could lead to the hope of increasing the yield of usable energy from each batch of biomass or reducing operational costs (Giampietro *et al.*, 1997; Field *at al.*, 2008).

In general, biofuels are liquid, solid or gaseous fuels derived from biomass such as ethanol, biodiesel and methanol (Kour *et al.*, 2019). Biofuels can be produced from various biomass feedstocks and can be categorised as first-, second-, and third-generation biofuels (Rodionova *et al.*, 2017; Nanda *et al.*, 2018). These categories cover a wide source of biomass which can originate from land, ocean and freshwater habitats. It can also include engineered microorganisms or energy gained through

photosynthetic cells (Field *et al.*, 2008). It is important to note that before the industrial revolution drove the usage of fossil fuels forward, biomass energy could be considered the dominant energy source of the world (Fernandes *et al.*, 2007).

The first-generation (1G) biomass mostly includes food crops, such as corn, wheat, sugarcane and food grains, which can relate back to controversy over food-versus-fuel debates (Ramos *et al.*, 2016; Nanda *et al.*, 2018). The second-generation biomass includes non-edible plant residues such as straw, wood and grasses. Unlike first-generation biofuel sources which are starch-based and can be directly used in biorefineries for fuel production, second-generation (2G) biomass requires pre-treatment (figure 1.1).

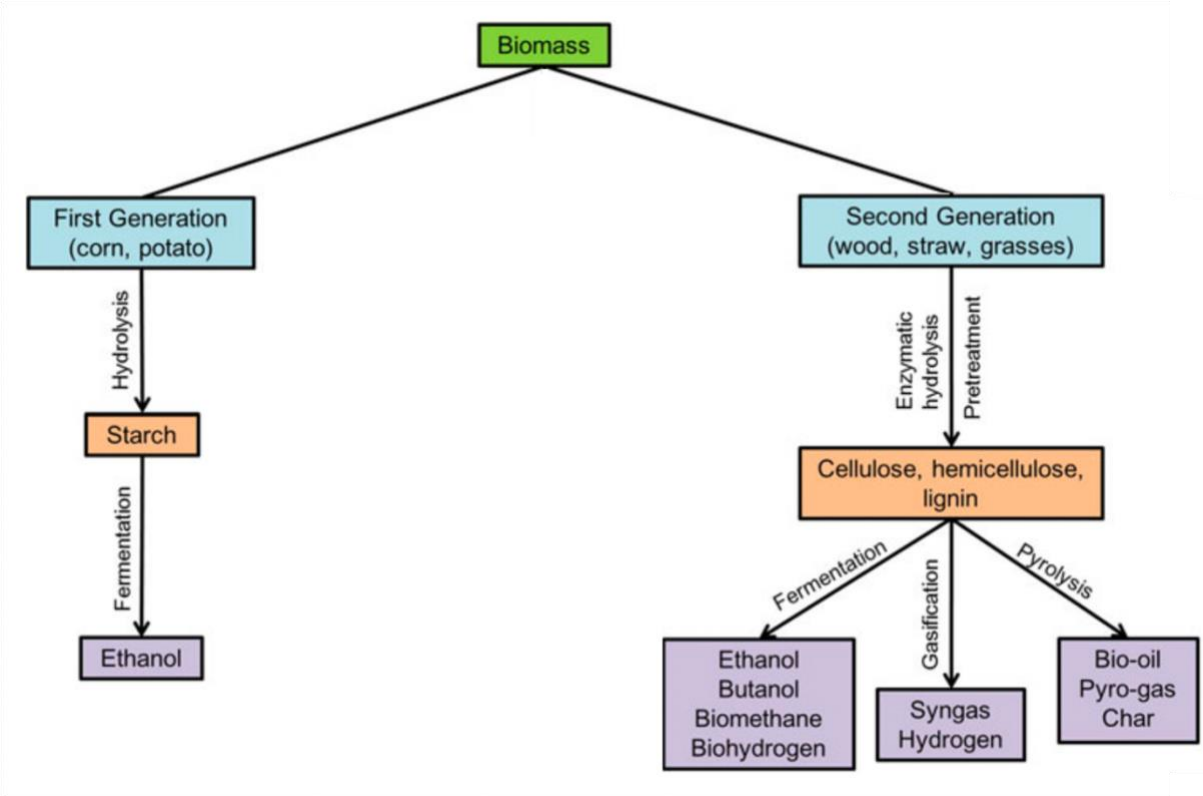


Figure 1.1: An overview of the basic production steps for first-generation and second-generation biofuels (Nanda *et al.*, 2018).

Production of 2G biofuels is a three-step process: i) pre-treatment, ii) enzymatic hydrolysis and iii) fermentation. It is required to pre-treat 2G biomass because they are mostly lignocellulosic materials comprised of cellulose, hemicellulose and lignin (Sorek *et al.*, 2014; Ramos *et al.*, 2016). Cellulose is a linear and crystalline homopolymer of repeating D-glucose subunits linked by β- 1,4 glycosidic bonds (figure 1.2)(Nanda *et al.*, 2018). These glucose polymers are largely insoluble and exist in crystalline microfibrils which makes extraction of the sugars difficult (Nanda *et al*., 2014). Hemicellulose is an amorphous short chain heteropolymer containing pentose sugar, hexose sugar and sugar acids (figure 1.3). Lignin is an amorphous, hydrophobic and aromatic polymer of ρ-hydroxyphenylpropanoid units linked via C-C and C-O-C bonds (figure 1.4)(Nanda *et al.*, 2018). Lignin forms a binding substance with cellulose and hemicellulose which can serve as a barrier for polysaccharide saccharification (Hu and Ragauskas, 2012). Harsh pre-treatment of the biomass is necessary due to the binding and allows for the accessibility of the carbohydrate polymers (Binod *et al.*, 2019).

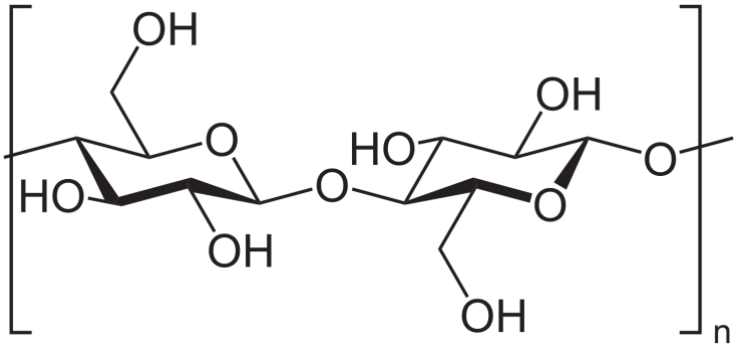


Figure 1.2: Cellulose chemical structure. A linear polymer of two D-glucose units linked by β-1,4 glycosidic bonds.

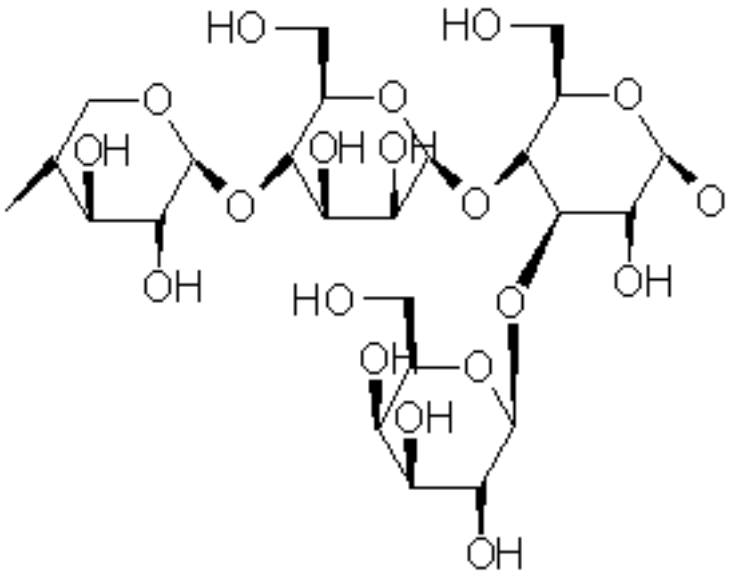


Figure 1.3: Hemicellulose chemical structure. Sugars include xylose - β-1,4 mannose - β-1,4 glucose - α-1,3 galactose.

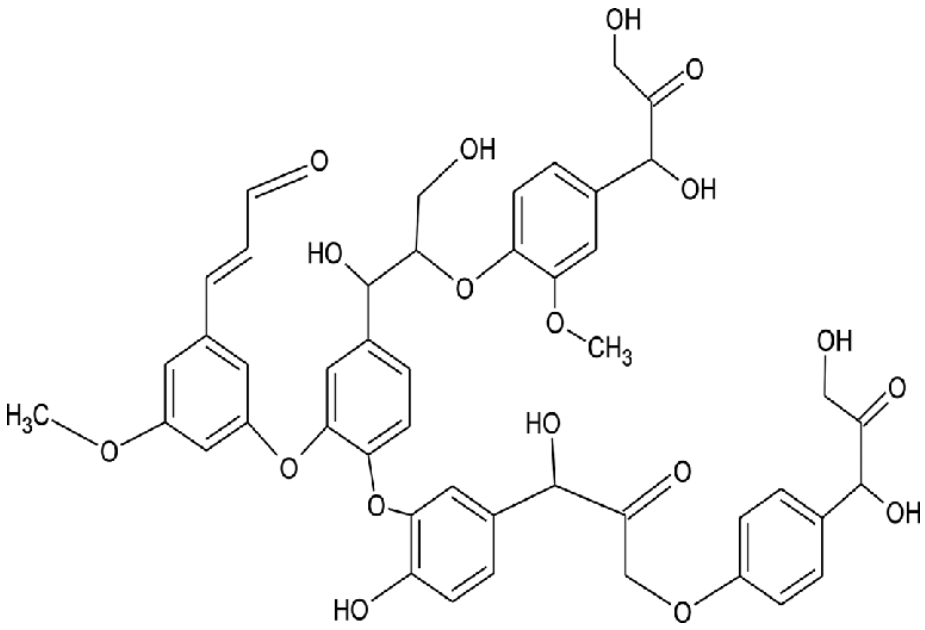


Figure 1.4: Lignin chemical structure (Mahmood *et al.*, 2018).

In the breakdown of 2G biofuel production (figure 1.5), pre-treatment opens the lignocelluloses in order to access the polymer chains of cellulose and hemicellulose, whilst the hydrolysis of the polymers achieves the release of the monomer sugars. These sugars are then fermented into an ethanol solution product “mash” by microorganisms which is then purified (Taherzadeh and Karimi, 2007; Ramos *et al.*, 2016). Binod *et al*. (2019) stated that due to the complex nature of lignocellulosic biomass, a single enzyme in unable to catalyse complete biomass hydrolysis. Pre-treatment also adds additional processing steps and can increase the operational cost for biofuel production (Nanda *et al.*, 2014; Nanda *et al.*, 2018). The cost-effective development of an enzyme cocktail which can be used in hydrolysis is a major challenge and most current strategies are not commercially viable. However, in nature, several bacterial genera including *Bacillus* and *Streptomyces* are known to produce cellulase and other biomass hydrolysing enzymes. Numerous fungi and other organisms also secrete enzymes which allows them to grow by metabolising lignocellulosic residues (Maki *et al*., 2009; Wilson, 2011; Ramos *et al.*, 2016; Binod *et al.*, 2019).

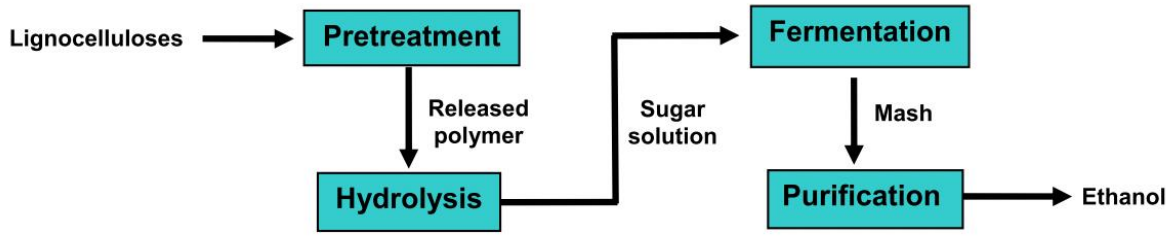


Figure 1.5: A further breakdown of the main steps of second-generation biofuel production of lignocellulosic materials into ethanol (Taherzadeh and Karimi, 2007).

In the last few years, considerable amounts of research has been directed towards the invigorated potential of biofuels, looking at biomass characterisation, more effective pre-treatment processes and biofuel upgrading (Nanda *et al.*, 2015). Renewable energy sources such as solar, wind and water can provide electricity with lower environmental costs. However, liquid fuels are still expected to remain necessary in the transport sector, despite the increase in electric vehicles, when looking at fuel for aviation, shipping and long-haul trucking (Correa *et al.*, 2019).

[1.2] – Lytic polysaccharide monooxygenases

Lytic polysaccharide monooxygenases (LPMOs) are a unique group of copper activated enzymes found across organisms such as bacteria, fungus and viruses. With their discovery in 2010, LPMOs revolutionised the process of enzymatic conversion of polysaccharides, being originally described as glycoside hydrolases (GH), (GH61 now known as AA9) or carbohydrate-binding-modules (CBMs), (CBM33 now known as AA10) both groups were demonstrated to be LPMOs (Vaaje-Kolstad *et al.*, 2010; Levasseur *et al.*, 2013; Filiatrault-Chastel *et al.*, 2019). LPMOs oxidatively cleave the glycosidic bonds of recalcitrant polysaccharide chains and represent a promising source for the development of industrial biomass processing in the production of biofuel (Vaaje-Kolstad *et al.*, 2017). LPMOs show a high diversity in their sequences and modulation, having been categorised into eight Auxiliary Activity (AA) families, on the Carbohydrate-Active Enzymes (CAZy) database (CAZy, 1998; Lombard *et al.*, 2014; Drula *et al.*, 2022). AA families are redox enzymes that act in conjunction with carbohydrate-active enzymes and based on sequence similarity, LPMOs are distributed across the AA families as AA9-AA11; AA13-AA17 (The CAZypedia Consortium, 2018). Each family is specific in the cleavage of

insoluble substrates such as cellulose (AA9, AA10, AA15, AA16) (Filiatrault-Chastel *et al.*, 2019), chitin (AA10, AA11, AA15), starch (AA13), hemicelluloses (AA9, AA14) and pectin (AA17) (Tamburrini *et al.*, 2021).

The catalytic domain of all LPMOs currently characterised adopts an immunoglobulin-like β-sandwich fold comprised of two β-sheets that are connected by loops and helices (Tandrup *et al.*, 2018). The active site is slightly exposed at the surface and a copper atom is coordinated at a strictly conserved histidine brace. The histidine brace is composed of three nitrogen ligands, two from the amino group Nt and side chain Nδ1 of the N-terminal histidine and one from Nε2 of a secondary histidine side chain, in a T-shaped brace (Quinlan *et al.*, 2011). Whilst LPMOs have a conserved active site, they have strong differences in the nature and arrangement of residues binding on the relatively flat substrate-binding surface (Forsberg *et al.*, 2018; Frandsen *et al.*, 2019). Due to the conserved copper-coordination site throughout all LPMOs, Forsberg *et al*. (2018) suggest that variation in enzyme performance, substrate specificity and oxidative regioselectivity must be driven in by the variation in the surrounding amino acids that constitute the substrate-binding surface.

The reaction mechanism of LPMOs is still relatively unclear due to this remarkable variety but there have been multiple suggestions. The shared view is that the resting redox state of the LPMO copper centre is Cu(II) which undergoes an initial reductive activation step to Cu(I). This then allows the enzyme to activate dioxygen (O_2) or hydrogen peroxide (H_2O_2) to hydroxylate the polysaccharide substrate. The redox state then alternates between Cu(II) and Cu(I) along the reaction pathway depending

on which mechanism is considered. The mechanisms require hydrogen abstraction from one of the carbons, C1 or C4, followed by the hydroxylation of the resulting substrate radical, which leads to destabilization of the glycosidic linkage and bond cleavage via an elimination reaction (Walton and Davies, 2016; Vaaje-Kolstad *et al.*, 2017). LPMOs have varying regioselectivity and some oxidise only the C1 atom of the glycosidic linkage, some only the C4 and others oxidise both C1 and C4 (Aachmann *et al.*, 2012; Bissaro *et al.*, 2017; Hedegård and Ryde, 2018).

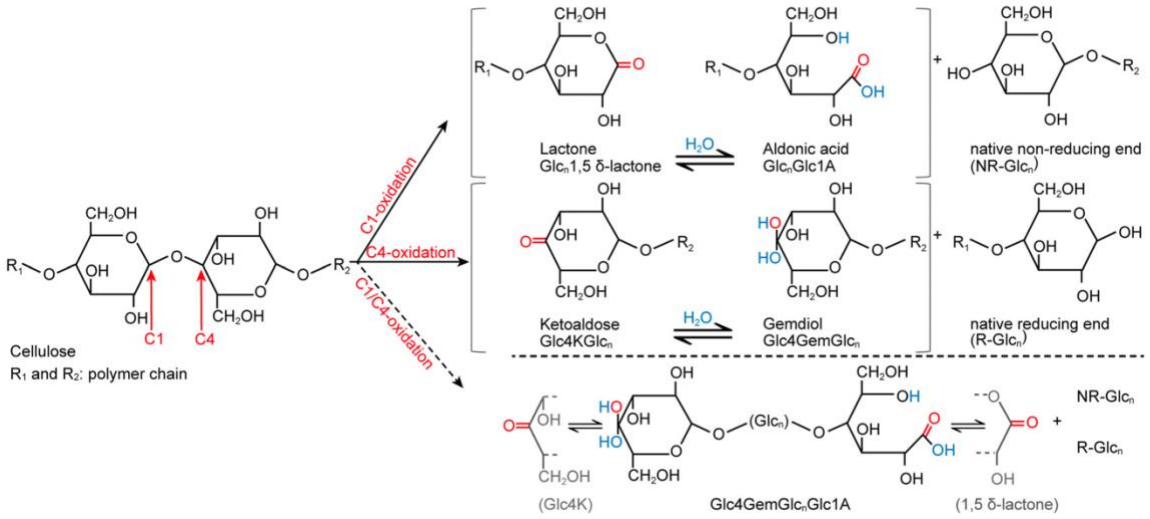


Figure 1.6: Oxidation pathways of cellulose active LPMOs with cleavage at C1, C4 or C1/C4. C1 oxidation (upper pathway) forms lactones that are hydrated to aldonic acids and generate native non-reducing ends. C4 oxidation (middle pathway) leads to formation of ketoaldoses and the corresponding hydrated gemdiols and generates native reducing ends. Mixed C1/C4 oxidation (lower pathway) can, in addition to the products from the upper and middle pathways, also produce oligosaccharides that are oxidised at both ends (Forsberg *et al.*, 2018).

Those that act on cellulosic substrates have been shown to either exclusively oxidise C1 or a mixture of C1 and C4 oxidised products. It is further shown that AA10 LPMOs that cleave at both C1 and C4 in cellulose are also capable of oxidising the C1 in chitin (figure 1.6) (Forsberg, Mackenzie, *et al.*, 2014; Forsberg, Røhr, *et al.*, 2014; Forsberg *et al.*, 2018).

$$\mathrm{RH} + \mathrm{O_2} + 2\mathrm{H^+} + 2\mathrm{e^-} \rightarrow \mathrm{ROH} + \mathrm{H_2O} \qquad (1)$$

The net oxidation of a substrate RH by O_2 by the LPMOs, proceeds under the consumption of two protons and two electrons as shown in equation (1) (Hedegård and Ryde, 2018). Bissaro *et al.* (2017), made the observation that this monooxygenase model is usually transposed to an enzyme on the basis that the observed catalysis depends on the presence of O_2 and on electron supply of a reductant. Monooxygenase catalysis typically involves cofactors, multinuclear metal centres, or the substrate itself as a means to store and deliver multiple electrons (Torres Pazmiño *et al.*, 2010; Solomon *et al.*, 2014; Bissaro *et al.*, 2017). The reaction condition of LPMOs typically lead to the production of H_2O_2 as a result of the reactions between O_2 and the reductant (Vaaje-Kolstad *et al.*, 2010; Caldararu *et al.*, 2019). However, Bissaro *et al*. showed that after priming reduction, LPMOs can use H_2O_2 in a controlled, substrate-associated manner. The proposed reaction mechanism (figure 1.7) offered H_2O_2 activation involving H_2O elimination and generation of a Cu(II)-oxyl species. When H_2O_2 is the co-substrate, two electrons are not needed, unlike the proposed O_2-based catalytic mechanism.

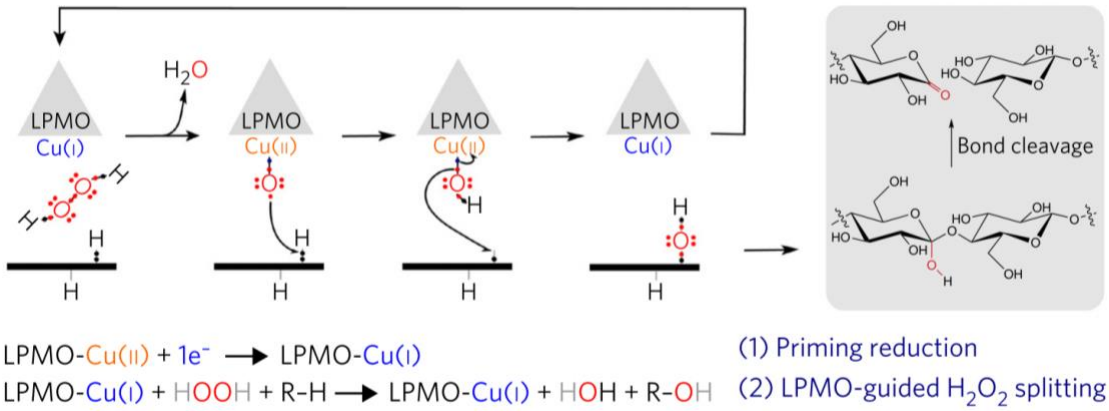


Figure 1.7: Suggested LPMO-H_2O_2 mechanism for enzymatic oxidative cleavage of polysaccharides. LPMO-Cu(II) is reduced to LPMO-Cu(I) (priming reduction), then reacts with H_2O_2 in the presence of a substrate. This reaction results in an elimination of a water molecule and generation of a Cu(II)-oxyl intermediate that can extract a hydrogen atom from the substrate. The resulting Cu(II)-associated hydroxide merges with a substrate radical through a rebound mechanism, which leads to hydroxylation of the substrate and regeneration of the Cu(I) centre, to once again begin the new catalytic cycle. The resulting hydroxylated polysaccharide undergoes molecular rearrangement and bond cleavage (Bissaro *et al.*, 2017).

In a new study by Chang *et al*., (2022) it was concluded that H_2O_2 was used as a co-substrate even in the presence of O_2 . There was no visible change in the O_2 concentration after the addition of LPMO, and H_2O_2-driven LPMO catalysis was observed for more than two hours. However, the high concentration of H_2O_2 produced by the auxiliary enzymes led to LPMO inactivation. The nature of the co-substrate was initially assumed to be O_2, but research has shown evidence of H_2O_2 as the actual co-substrate and that the catylytic efficiency of LPMOs is higher when adding H_2O_2 compared to the reduction of O_2 to H_2O_2 (Hangasky, Iavarone and Marletta, 2018; Müller *et al.*, 2018; Hedison *et al.*, 2021).

1.2.1 – AA13 LPMOs

The AA13 family of LPMOs are a fungal family active on starch. Starch is known to be broken-down mainly by α-amylase and glucoamylase, two hydrolytic enzymes used in starch-based biofuel processing (Vu *et al.*, 2014). Given the nature of LPMO biochemistry and their major industrial application to cleave recalcitrant polysaccharide chains, the use of starch active LPMOs is especially appealing. With starch being the most abundant storage glucan in plants, the production of starch-containing crops is greater than all other industrial or food substrates combined (Lo Leggio *et al.*, 2015; MacNeill *et al.*, 2017). Leggio *et al*., (2015) showed that activity on retrograded starch released aldonic acid malto-oligosaccharides, through lytic oxidative cleavage at the anomeric C1 of the polysaccharide chains. It also showed that the AA13 enzyme boosts the β- amylase catalysed release of maltose. Structurally the active site was observed within a shallow groove, unlike the binding surfaces of other known LPMOs.

Aspergillus oryzae (*A. oryzae*) or *Ao*AA13 is a fungal member of the AA13 LPMO family and the first structure characterised of an AA13 family member. Whilst this family has not received as much attention as the chitin/cellulose AA9-AA11 families, more structural studies have been carried out to develop advances in the methodology of starch-active AA13 LPMOs (Muderspach *et al.*, 2019). Five genomes in the genus *Aspergillus* have been determined and consequently it is now the best genus to clarify by comparative genome analysis (Kobayashi *et al.*, 2007). *A. oryzae* specifically has a long history of use in the food industry and has been supported by the World Health Organisation (WHO) and Food and Agriculture Organisation (FAD) (*FAD/WHO*, 2021). The Barbesgaard *et al*., (1992) review on *A.*

oryzae concluded with the statement that *A. oryzae* was an excellent host for the safe production of harmless products by recombinant strains. The fact that it can be manipulated genetically to enhance its protein production abilities (Christensen *et al.*, 1988) and that it secretes a variety of enzymes in solid state fermentation (Kitamoto, 2002) makes it a very interesting fungi for use in industrial applications.

1.2.2 – AA10 LPMOs

The AA10 family of LPMOs, formerly referred to as CBM33, are active on cellulose or chitin. Chitin-binding proteins are abundant in bacteria and viruses but are rare in eukaryotes (Vaaje-Kolstad *et al.*, 2010). AA10s share a flat substrate-binding surface which is observed across LPMOs and the conserved copper activation site coordinated by three nitrogen atoms in a histidine brace arrangement (Quinlan *et al.*, 2011). Streptomyces are Gram-positive, soil dwelling, bacteria (Hodgson, 2000) and *Streptomyces lividans* (*S. lividans*) possesses seven genes that encode for AA10 LPMOs (*Sli*LPMO10A-G) (Cruz-Morales *et al.*, 2013). Three of these proteins have a carbohydrate-binding-module appended to the LPMO domain. CBMs are usually part of a multi-modular enzyme, contributing to substrate-binding of the LPMO whilst possibly affecting operational stability and enhancing plant biomass conversion. Although little work has been done on characterising LPMO-CBM systems, all chitin-active AA10 LPMOs with known structures oxidise C1 of the polysaccharide chain (Forsberg *et al.*, 2016).

[1.3] – Carbohydrate-binding modules

In the CAZy classification system (Lombard *et al.*, 2014; Drula *et al.*, 2022), carbohydrate-binding modules (CBMs) are classified as sugar-binding proteins that comprise amino acid sequences within a larger encoded protein sequence that fold into a structurally discrete module (CAZy, 1998; The CAZypedia Consortium, 2018). CBMs were initially classified as cellulose-binding domains (CBDs) based on the initial discovery of several cellulose-binding modules (Gilkes *et al.*, 1988; Tomme *et al.*, 1988). However additional modules are continuously being found that bind carbohydrates other than cellulose whilst meeting the criteria of a CBM, hence the reclassification. There are currently 94 CBM families listed on the CAZy database. Typically, the roll of a CBM is to bind to a carbohydrate ligand and direct the catalytic machinery onto the substrate, therefore enhancing the catalytic efficiency of the multi-modular enzyme (figure 1.8) (Taylor and Drickamer, 2014). CBMs are devoid of any catalytic activity and are primarily linked to catalytic modules such as glycoside hydrolases (GH) and polysaccharide oxidases (also categorised under AA) (Georgelis *et al.*, 2011; Levasseur *et al.*, 2013; The CAZypedia Consortium, 2018). They do not undergo conformational changes when ligand binding, although the multi-modular enzymes linked to CBMs may be quite flexible and have significant conformational changes when binding substrate (Ficko-Blean *et al.*, 2009).

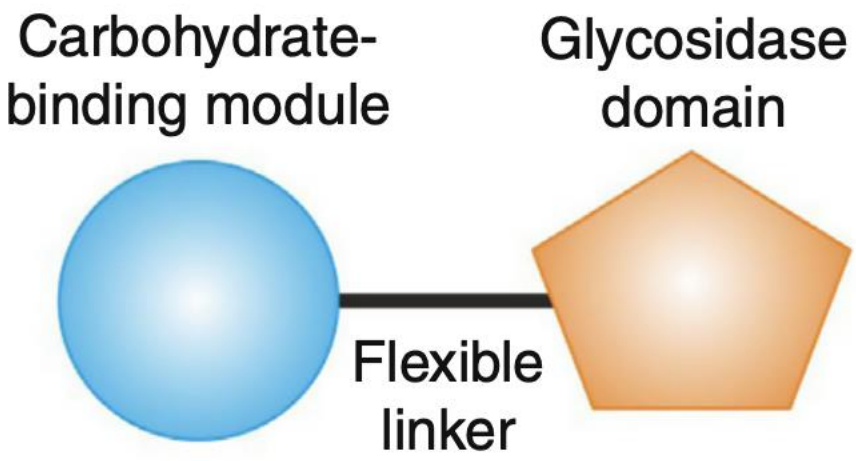


Figure 1.8: Association of the carbohydrate-binding domain with an enzymatically active domain. Carbohydrate-binding modules are often linked to glycoside hydrolases (glycosidases) and polysaccharide-degrading enzymes. The CBM localises the activity on substrates and enhances enzyme activity (Taylor and Drickamer, 2014).

The analysis of CBMs continues to provide new insight into the carbohydrate-based nature of some bacterial-host interactions while providing a novel group of carbohydrate-binding proteins for use as research tools (Ficko-Blean and Boraston, 2012). Their application in biotechnological areas ranges from exploitation of CBM features to create designer carbohydrate-active enzymes with enhanced or modified carbohydrate recognition features; to the enhancement of bioprocessing enzymes for industrial pulp processing and biofuel production (Cuskin *et al.*, 2012; Reyes-Ortiz *et al.*, 2013; Sidar *et al.*, 2020).

[1.4] – The use of LMPOs in biofuels

Traditional enzymatic cocktails used for the conversion of recalcitrant polysaccharides into fermentable sugars mainly consists of hydrolytic cellulases (Kapsokalyvas *et al.*, 2018). However, the discovery of LPMOs has offered a new way for biomass saccharification (Hemsworth *et al.*, 2015). Saccharification steps for the method of cellulose conversion involves three main enzyme types to synergistically convert cellulose to glucose; β-1,4 glucanases cleaving the glycosidic bonds in the cellulose chain at random; cellobiohydrolases which hydrolyse cellulose from the chain ends; and β- glucosidases for converting cellobiose and cello-oligosaccharides to glucose (Song *et al.*, 2018).

The presence of LPMOs in current cellulase cocktails increases efficiency due to their ability to boost enzyme activity and improve saccharification yields. However, substrate composition could affect LPMO activity due to LPMO dependence on the presence of a co-substrate, which must be factored into the designing of saccharification and fermentation processes (Hemsworth *et al.*, 2015; Müller *et al.*, 2015). Investigations by Müller *et al*., (2015) explored the effects of oxygen and a reductant on LPMO-containing cellulase efficiency in biomass saccharification. Results showed definitively that both oxygen and an electron donor being present are necessary to stimulate LPMO acitivty in commercial enzyme cocktails. With aerobic conditions and the subsequent stimulation of LPMO activity, an increased saccharification yield of 60 % was observed along with a increased rate of saccharification. This experiment sets the importance of designing processing conditions to promote LPMO activity for efficient enzymatic saccharification of lignocellulosic biomass. The benefits of adapting processing conditions, the

subsequent increase in production yield and the shortening of processing times, could translate to lower feedstock costs or reduced overall enzyme loadings during production. The data indicated that the addition of an external electron donor may not be necessary for a lignocelluosic substrate if the substrate is not washed after pre-treatment.

Lignocellulosic substrates generally contain, in addition to cellulose, hemicellulose and lignin. Some substrates like lignin which are involved in redox cycles (Felby *et al.*, 1997) may act as an electron donor for LPMOs depending on the pre-treatment applied (Rodríguez-Zúñiga *et al.*, 2015). Harris *et al*., (2010) observed that the effect of protein GH61 (now recongised as an AA9 LPMO) on cellulase activity, only occurred on lignocellulosic substrates over pure cellulose substrates. These observations could suggest that lignin or hemicellulose could play a role in LPMO activity (Agger *et al.*, 2014). Hu *et al*., (2014) demonstrated that LPMO enhancement is negatively correlated with cellulose accessibility.

Song *et al*., (2018) showed consistency with previous studies and further indicated that LPMOs react specifically to fragment large cellulose ribbons into small microfibrils with few oxidised sugar products. It was concluded that LPMOs breakdown the amorphous proportion of cellulose and increase surface accessibility for cellulase enzymes. Another study also demonstrated that LPMO enhancement is negatively correlated with cellulose accessibility (Hu *et al.*, 2014). Kim *et al*., (2016) also found that LPMOs increase the degradation of insoluble xylan, a type of hemicellulose, which suggests that LPMOs can oxidise sugars in different

amorphous polysaccharides; thus exposing cellulose microfibrils to be hydrolysed by cellulases.

[1.5] – Ancestral reconstructions

Ancestral reconstruction is an important application of phylogenetics, which enables the reconstruction and study of evolutionary relationships among an individual, species or populations, to their ancestors. In biology, it can be used to recover ancestral character states, including genetic sequences, known as ancestral sequence reconstruction (ASR) (Joy *et al.*, 2016). For protein engineering, two common objectives are i) to optimise the stability and solubility of a pre-existing protein whilst keeping or improving its primary function and ii) to identify or design protein sequences that present novel activity or ligand/substrate specificity. ASR has emerged as a valuable protein engineering tool to be used in industrial applications due to the desirable characteristics often displayed by ancestral proteins, including unique activity profiles or enhanced thermostability. By comparing the structure and function of modern proteins and ancestral proteins, ASR can highlight functionally important substitutions within families and provide direction in new variant design (Spence *et al.*, 2021). The study of ancestral protein structures and functions using ASR is likely to provide new insights into the physical and chemical determinants which have previously shaped the respective protein families evolution (Harms and Thornton, 2010). It has also been suggested that some ancestral enzymes had non-specific binding characteristics and could bind several different substrates during their initial steps of evolution (Khersonsky and Tawfik, 2010). However, there is also evidence to suggest that ancestral proteins were in general no more multifunctional than currently existing ones (Siddiq *et al*., 2017). Although there is uncertainty

regarding ASR to some extent, primarily regarding the technology's dependency on phylogenetic trees, studies continue to show that ASR is validated by the properties of the proteins reconstructed in the laboratory, with their capacities being consistent with what is expected based on physical science and paleogeology (Merkl and Sterner, 2016). Gumulya *et al*., (2018) were able to show that two unrelated groups of enzymes that ancestral reconstruction could yield thermostable and highly active forms. The approach of ASR to engineer proteins for the application in biomass processing will possibly play a pivotal role in the future of biofuels.

[1.6] – Aims and objectives

Chapter 2 of this study, Ancestral reconstructions of *Ao*AA13, aims to express and purify the ancestral sequence reconstructions ASR-N231 and ASR-N237, originating from *Aspergillus oryzae* of the LPMO AA13 family, through a bacterial cell line. It is hoped that the determination of the two ancestral proteins' functionality and characteristics and their eventual substrate binding tests will test their applications in industrial batch processes for starch-derived biofuels.

Chapter 3, Characterisation of AA10 *Sli*LPMO10G, aims to describe the truncation of *Sli*LPMO10G from *Streptomyces lividans* to study the binding characteristics of the full protein and the respective LPMO and CBM5 domains along with the subsequent structure determination. It is hoped that the separation of the LPMO and CBM5 domains and substrate binding test of the three protein forms will give a better understanding to the LMPO-CBM substrate binding relationship and specificity.

Chapter 2: Ancestral reconstructions of *Ao*AA13

[2.1] Introduction

2.1.1 – AA13 family of LPMOs

In the Carbohydrate-Active-enzymes (CAZy) database, LPMOs have been classified into families under the broadly termed Auxiliary Activities (AA) category (CAZy, 1998)(The CAZypedia Consortium, 2018). The fungal AA13 family is active on starch, with structural relatedness to the AA9, AA10 and AA11 families. The structure of the AA13 LPMOs conserves the typical active site seen across all other LPMOs consisting of a “T-shaped” histidine brace composed of three nitrogen ligands, two from histidine side chains and the third from the N-terminal amino group (Quinlan *et al.*, 2011). AA13s use a copper cofactor and an electron donor to oxidatively cause chain breaks in the α-1,4-linked glucose polymers that form starch (Hemsworth and Lo Leggio, 2019)(Vu *et al.*, 2014). Oxidation is shown to specifically occur at the C1 position of glycosidic bonds in starch and related substrates (Lo Leggio *et al.*, 2015).

2.1.2 – Ancestral reconstructions

Ancestral reconstruction (AR) is the extrapolation back in time from measured characteristics of individuals, or populations, to their common ancestors. It can be used to regain knowledge of different ancestral character states such as genetic sequence, the amino acid sequence of a protein and the composition of a genome (Joy *et al.*, 2016). Ancestral sequence reconstructions (ASR) has emerged as a valuable protein engineering tool due to ancestral proteins often displaying desirable characteristics (Spence *et al.*, 2021). The ability to now study the structure and functions of ancestral proteins found using ASR, is likely to provide new insights into

the thermostability and substrate specificity involved in the evolution of protein families (Harms and Thornton, 2010). By comparing the structure and function of modern proteins and ancestral proteins, ASR can highlight functionally important substitutions within families and provide direction in new variant designs (Spence *et al.*, 2021). The approach of ASR to engineer proteins for the application in biomass processing will possibly play a pivotal role in the future of biofuels (Gumulya *et al.*, 2018).

This study aims to express and purify the ancestral sequence reconstructions ASR-N231 and ASR-N237, originating from *Aspergillus oryzae* of the LPMO AA13 family, through a bacterial cell line. It is hoped that the determination of the two ancestral proteins' functionality and characteristics and their eventual substrate binding tests, will test their applications in industrial batch processes for starch-derived biofuels.

[2.2] Method

2.2.1 – Expression and purification of the AA13 ancestral reconstructions

Preparation of chemically competent cells

The following *Escherichia coli* (*E. coli*) strains were used; BL21-RIL, C43-DE3, BL21-DE3 and T7 SHuffle. Competent cells for BL21-RIL, C43-DE3, BL21-DE3 were prepared by inoculating 100 mL LB media with 2 mL overnight culture, before being grown at 37 °C, 180 RPM until an OD_{600} between 0.6 and 1.0 was reached. Cells were placed on ice for 20 minutes and the volume divided into two 50 mL falcon tubes. Centrifugation of the cells occurred at 4000 RPM for 10 minutes and the supernatant was discarded. The pellets were resuspended in 10 mL of ice-cold 0.1 M calcium chloride ($CaCl_2$) respectively, before being combined and incubated on ice for 30 minutes. A second centrifugation was carried out at 2500 RPM for 10 minutes and the supernatant discarded before resuspension in 4 mL ice-cold 0.1 M $CaCl_2$ containing 15 % glycerol. The final resuspension was transferred to pre-chilled 1.5 mL Eppendorf tubes in 100 μL aliquots. T7 SHuffle cells were prepared by a modified procedure whereby cells were incubated at 30 °C whilst grown. All cells were stored at -80 °C.

Plasmids

Four modified pET26b plasmids (Chaplin *et al.*, 2016) containing an AA13 ASR sequence with a Tat or Sec leader sequence were created by Dr Vijgenboom, University of Leiden, The Netherlands, and named Tat 231, Tat 237, Sec 231 and Sec 237. All plasmids contained a kanamycin (Kan) selection marker and were transformed to the desired chemically competent cells by heat shock at 42 °C for 45 seconds followed by a 75-minute incubation at 37 °C with the addition of 800 μL LB

before plating on LB agar plate containing final concentration of 50 μg mL^{-1} Kan. To create stock plasmid solutions, competent XL1-Blue cells were used and plasmid DNA was extracted from overnight cultures using a DNA mini-prep kit (ThermoFisher) to a final value of 50 μL.

Restriction digests and agarose gels

Restriction digests of the ASR plasmids were carried out using *HindIII* and *NdeI* restriction enzymes (Fermentas). 3 μL of DNA was added to 2 μL ddH_2O, 2 μL 10xRed Buffer (Fermentas), 2 μL *NdeI* and 1 μL *HindIII* to give a final volume of 10 μL and incubated for 2 hrs at 37 °C. Samples for agarose gels were prepared by adding 2 μL of 6 x loading dye and 5 μL GelRed® nucleic acid gel stain 10,000x stock (Biotium) to the sample. Samples were loaded to a 1 % (w/v) agarose (Fisher) gel and run in Tris/Borate/EDTA (TBE) buffer at 200V for 30-40 mins. Gels were viewed on a FastGene blue LED transilluminator.

Growth media and over-expression tests

LB and 2xYT (Melford Chemicals) were used for over expression tests. The desired plasmid was transformed to the selected *E. coli* strain. Overnight transformants were selected from an LB/agar/Kan plate and used to inoculate a 3 mL overnight LB culture containing a final concentration of 50 μg mL^{-1} of Kan. The LB culture was incubated at 37 °C, 180 RPM overnight. The BL21-RIL cell line contained a chloramphenicol (CM) resistance cassette and thus required the presence of 20 μg mL^{-1} CM in addition to Kan in LB agar plates and cultures. Flasks (250 mL) containing 100 mL LB or 2xYT, with Kan final concentration 50 μg mL^{-1} and CM final

concentration 20 μg mL^{-1} where appropriate, were inoculated with 1 mL from the overnight cultures and cultured throughout the day at 37 °C, 180 RPM. Incubation of the T7 SHuffle cells was carried out at 30 °C. At an OD_{600} of between 0.4 and 0.5, isopropyl β-D-1-thiogalactopyranoside (IPTG; Melford Chemicals) was added from a stock solution of 0.5 M to give a final concentration of 100 μM in LB and 250 μM in 2xYT. Growth continued overnight at 18 °C and 180 RPM.

SDS-PAGE

A combination of lab made, or commercial pre-cast gels (12%) were prepared following the volumes in table 2.1 to make two 12 % acrylamide gels. Lab made gels and commercial gels were run in a Mini-PROTEAN Tetra Cell (BioRad) system in a x1 dilution of 1 L x10 Running Buffer (30 g TRIS, 144.45 g Glycine, 5 g SDS) at 200 V for 40 mins.

Table 2.1: SDS PAGE 12 % gel volumes to make 2 gels.

Media	Resolving Gel	Stacking Gel
ddH_2O	6.4 mL	5.95 mL
1.5 M Tris pH 8.8	5.2 mL	-
0.5 M Tris pH 6.8	-	2.5 mL
Acrylamide/Bisacrylamide (30 % / 0.8 %, w/v)	8 mL	1.34 mL
10 % SDS	200 µL	100 µL
10% APS	200 µL	100 µL
TEMED	20 µL	10 µL

Once gels were run, they were stained in Coomassie blue, (455 mL ethanol, 455 mL ddH_2O, 90 mL glacial acetic acid and 2.5 g Coomassie blue in 1 L) for 30min followed by de-staining (455 mL ethanol, 455 mL ddH_2O and 90 mL glacial acetic acid in 1 L) until bands were visible. The gels were then left in water overnight to clean before images were taken.

Sample preparation for SDS-PAGE

Samples for SDS-PAGE from over-expression and purification were taken from the cultures at defined times. 20 μL of each prepared sample was loaded onto the gels. A 50 μL aliquot of the over-expression was initially taken and spun in a microfuge (13,000 RPM) for 1 min, followed by removal of the supernatant and the resuspension of the cell pellet in 50 μL cracking buffer (1 mL of 1mM TRIS pH 6.8, 2 mL of 10 % SDS, 5 mL glycerol, 11.8 mL ddH_2O, 10 mg 0.05 % bromophenol blue and containing a few crystals of DTT added before use). Samples taken during the over-expression before and after IPTG induction were prepared with 15 μL of the resuspended pellet sample and 5 μL cracking buffer. A sample of 6 μL was taken from the resuspension and 5 μL cracking buffer and 9 μL ddH_2O were added. Lysis supernatant samples were prepared with 20 μL supernatant and 5 μL cracking buffer. The lysis pellet sample was first mixed with 50 μL cracking buffer to prepare. A 4 μL sample was taken and added to 5 μL cracking buffer and 11 μL ddH_2O. Samples were denatured at 90 °C for 10 mins after preparation.

Sucrose shock

Overnight cultures were harvested by centrifugation at 4000 RPM, 4 °C for 20 mins in a Beckman Coulter Avanti JXN-26 centrifuge using a JLA-8.1000 rotor. The supernatant was discarded, and the cell pellet was resuspended in ice cold 50mM Tris/HCl, 1 mM EDTA, 20 % w/v sucrose, at pH 8. The cell resuspension was stirred at 4 °C for 1 hour and then 60 μL of 1 M $MgSO_4$ was added for every gram of the cell pellet. The cell paste was then stirred at 4 °C for a further 30 min. A second centrifugation step was carried out at 18,000 RPM for 20 minutes at 4 °C using a JA-25.50 rotor and the supernatant collected and stored at 4 °C. The pellet was then resuspended in ($1/30^{th}$ the volume of the culture) ice-cold water and stirred for 1 hr at 4 °C before a third centrifugation step at 18,000 RPM for 20 mins at 4 °C. The resulting supernatant was collected, and a sample of the pellet taken to run an SDS-PAGE.

Determination of physiochemical properties of the ASR sequences

The online programme ProtParam, part of the ExPASy Bioinformatics Resource Portal (Gasteiger *et al.*, 2005; Duvaud *et al.*, 2021), was used to calculate molecular weight, pI and extinction coefficients at 280 nm using the amino acids sequences of the ASR LPMOs as input.

[2.3] Results

2.3.1 – Sequence analysis of the AA13 ancestral sequence reconstructions (ASR)

A phylogenetic tree of the LPMO families has been published (Voshol, Punt and Vijgenboom, 2019) and has been used as a starting point to carry out ancestral reconstruction of the AA13 family. A simplified version of the tree is shown in figure 2.1. The AA9 are the largest family (41 %), followed by AA10s (27 %), AA11s (14 %), AA15 (7 %), AA16 (4 %), LMPO16s (4 %), AA13s (1 %) and AA14s (< 0.5%). To generate the tree of all LPMOs, associated module sequences like CBMs were removed if present, and all sequences started with the N-terminal His residue.

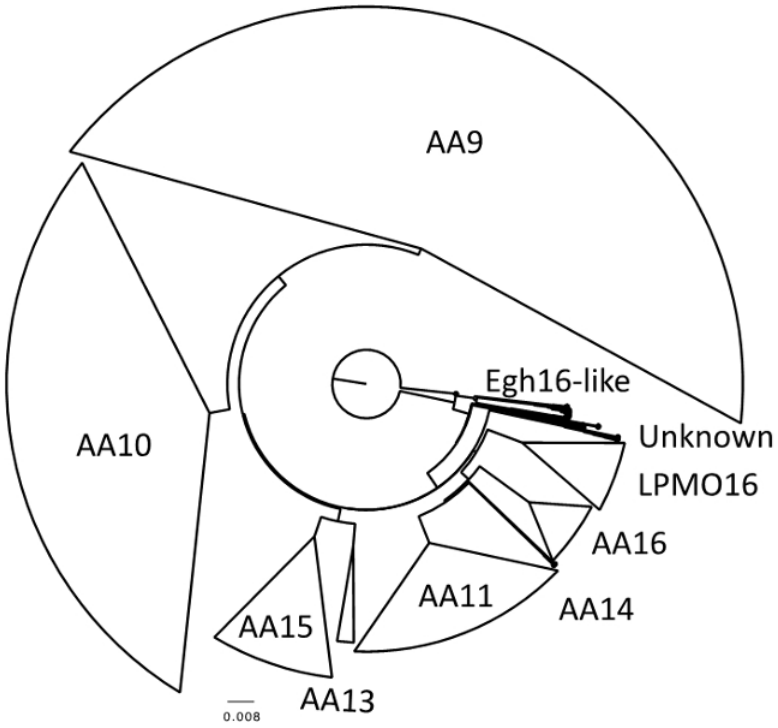


Figure 2.1: Simplified phylogenetic tree of the LPMO) (Voshol, Punt and Vijgenboom, 2019).

By using an approach outlined by Merkl and Sterner (2016), collaborators at the University of Leiden, the LPMO phylogenetic tree constructed ancestral sequences for the AA13 family by identifying the ancestral reconstruction sequences with the highest probability score for each node. Starting from *Aspergillus oryzae* (Q2U8Y3_ASPOR), the red line (figure 2.2) traces backwards to the node that is the ancestor of the large AA13 branch N237, which is derived following nodes N274, N273, N272, N271, N265, N264, N260, N259, N258, N251, N239. After N237, the node N231 is ancestor to a small branch of extant AA13 sequences. Going further back in time, the node N1 is reached which is the link between AA13 and AA15 sequences, and thus a potential common ancestor.

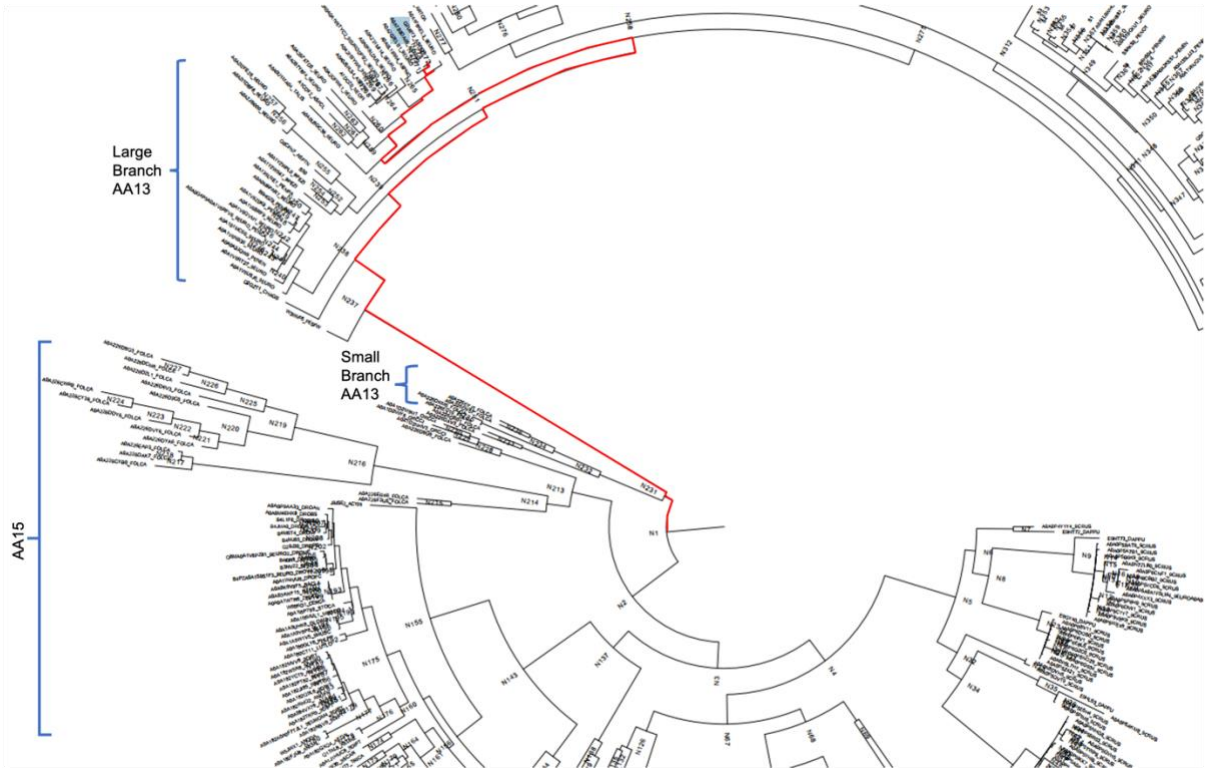


Figure 2.2: Close up section of the AA13 and AA15 LPMO families. The red line highlights the nodes, tracing back in time, from AA13 *Aspergillus oryzae* to the ancestor that links AA13 with AA15 sequences, N1.

Figure 2.3 shows the amino acid sequence alignment of the predicted ancestral sequences from *A. oryzae* to the node N1. The secondary structure of *A. oryzae* was included to help identify the sequence regions lost over time. Conserved regions of all sequences were seen across the His residues used for the histidine brace at the Cu binding site and the Cys regions for the 2nd, 4th and 7th disulphide bridges. Q2U8Y3_ASPOR to N237 are shown to have highly conserved regions in β1, β2, β4, β5, β6, β7 and β10. More distinct divergences can be seen in the sequences after N264 in node N260 onwards at residue 170, showing three amino acid deletions until N237. The divergence becomes apparent at node N231 to N1 with multiple gaps in the sequences compared to N237, showing the loss of α2, the loss of sequence between β7 and α4, the shortening of β8 and the loss of β9. The loop regions in LPMOs are often associated with substrate binding and many of the divergences seen are localised to loops. There is also a loss of 4 cysteines in the sequences with the total number of cysteines in N231 and N1 being 10 whilst N237 has 14. The disulphide bridges 1, 3, 5 and 6 were also lost for N231 and N1. Although, 4 cysteine residue positions in N231 and N1 sequences have changed to residues 30, 45, 167 and 176, reforming disulphide bridges at new locations. However, both N231 and N1 node still have highly conserved regions with the later sequences and between themselves.

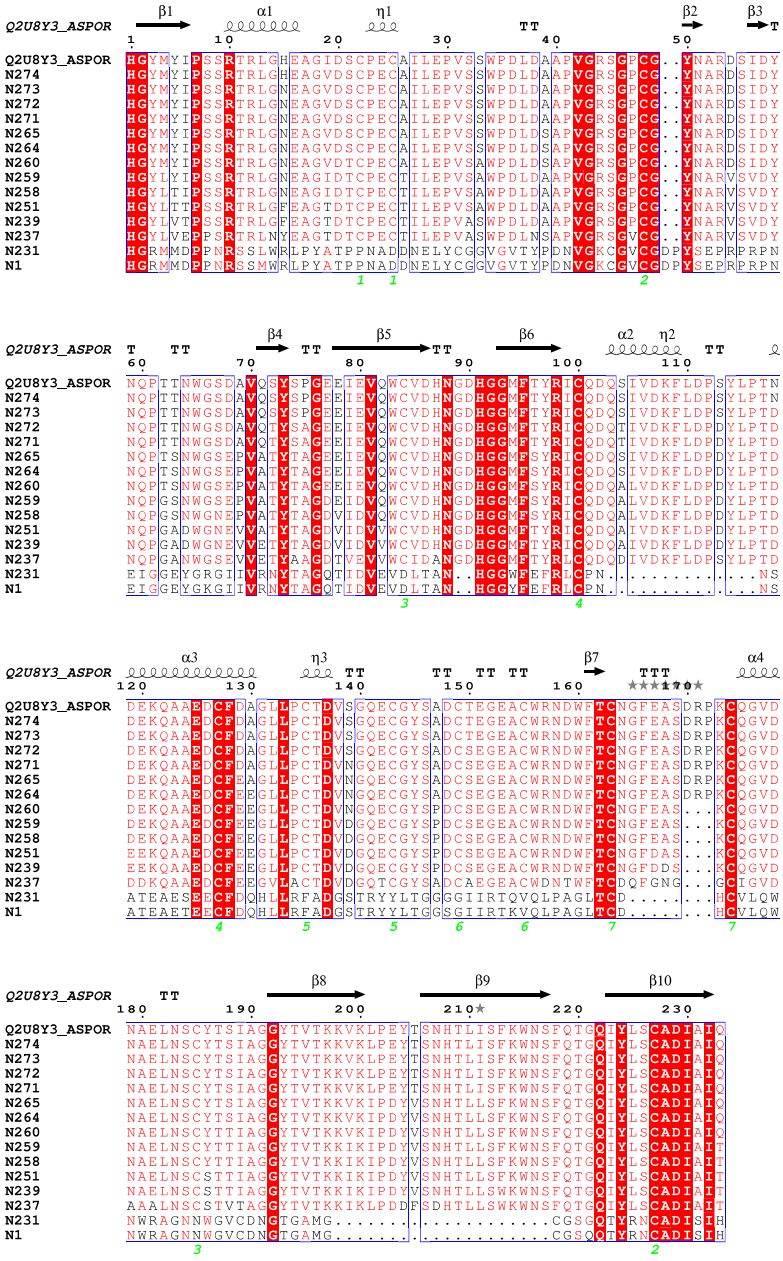


Figure 2.3: Sequence alignment of the predicted sequences with the highest probability score at the nodes dating back in time from AA13 *Aspergillus oryzae*, (Q2U8Y3_ASPOR) to the node that links AA13 with AA15 sequences, N1. The secondary structure of *A. oryzae* is shown above the sequence to identify the regions which are lost as the branches are followed back in time. Conserved regions are highlighted in red, regions with similar amino acids are grouped by a blue box and gaps are indicated by dots. α-helices, 3_{10}-helices and π-helices are displayed as medium, small and large squiggles respectively. 3_{10}-helices are indicated by η , β-strands are shown as arrows, strict β-turns as **TT** and strict α-turns as **TTT**. Disulphide bridges are indicated by green digits under each column with a bound cysteine and residues with alternate positions are marked by grey stars on the top of sequences blocks.

Selection of ancestral sequences for over-expression tests

The ancestral sequences at nodes N237 (ancestor for large family) and N231 (ancestor for small family) were selected for further investigation. Based on the amino acid sequence, codon optimised nucleotide sequences for over-expression in *E. coli* were generated. The synthetic DNA encoding the genes N231 and N237 were ligated into a pET26b plasmid at the *NcoI* and *HindIII* restriction sites. The pET26b plasmid used had previously been modified so that on cleavage of the signal sequence, a mature LPMO protein would be present, starting with the N-terminal His which is part of the His brace that coordinates the Cu at the active site (Chaplin *et al.*, 2016). Two types of signal sequence were employed, the first being the pelB sequence that is part of the pET26b plasmid (figure 2.4), which is recognised by the Sec translocation pathway and exports the protein across the membrane into the periplasm in an unfolded state. The second signal sequence to be used was an *E. coli* Tat sequence (figure 2.5), which is recognised by the twin-arginine translocation pathway, and exports proteins into the periplasm in their folded state. Previously, *S. lividans* LPMOs were shown to express well with Sec (Chaplin *et al.*, 2016).

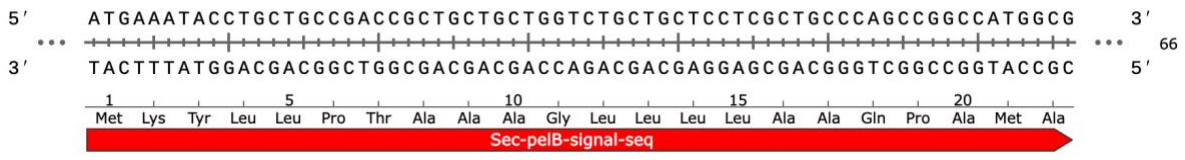


Figure 2.4: Sec signal sequence with amino acid sequence indicated.

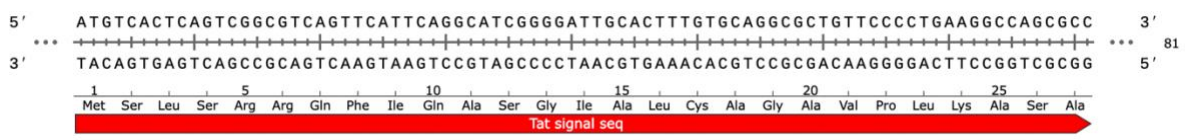


Figure 2.5: Tat signal sequence with amino acid sequence indicated.

In total four plasmids were constructed, two with Sec sequences (figure 2.6 and 2.7) and two with Tat sequences (figure 2.8 and 2.9) for testing over-expression in *E. coli.*

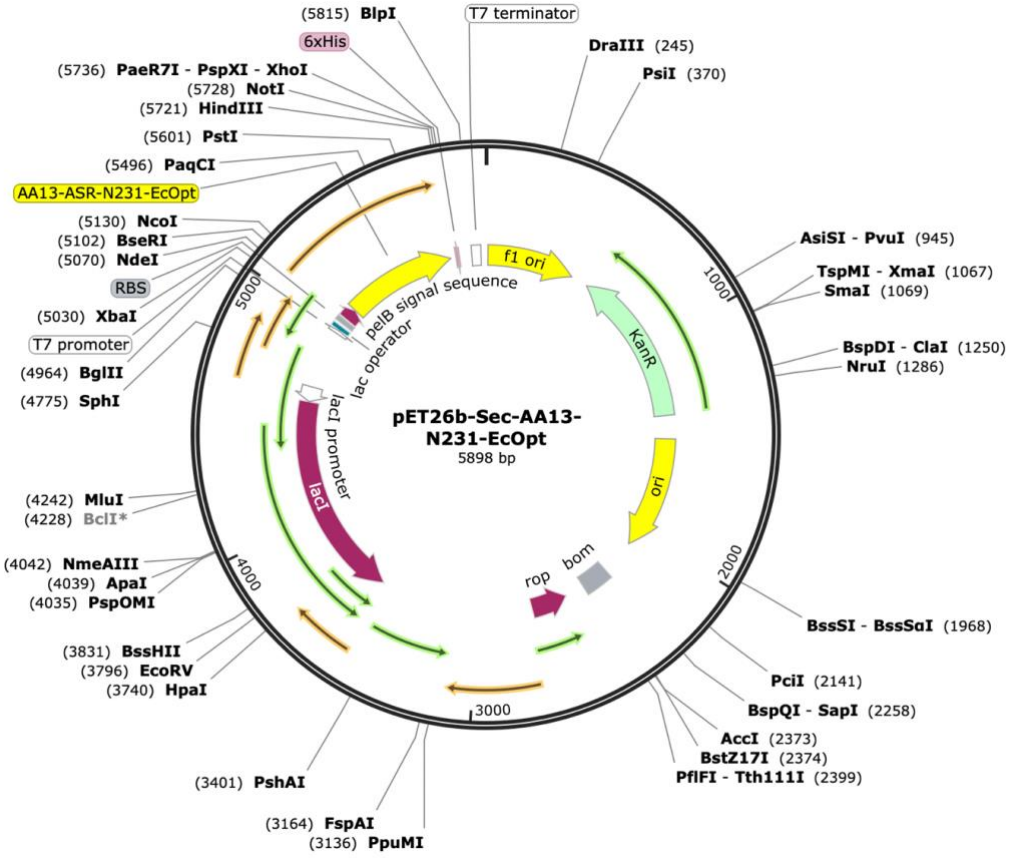


Figure 2.6: *E. coli* optimised pET26b plasmid of AA13 ASR-N231 with pelB Sec signal peptide.

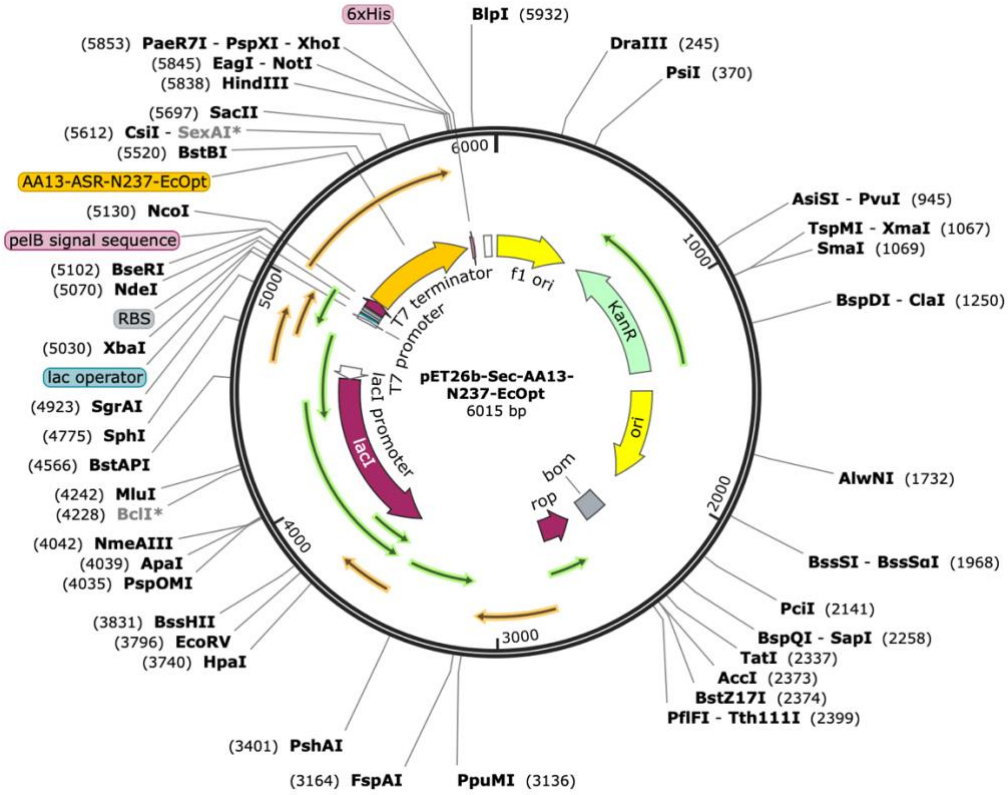


Figure 2.7: *E. coli* optimised pET26b plasmid of AA13 ASR-N237 with pelB Sec signal peptide.

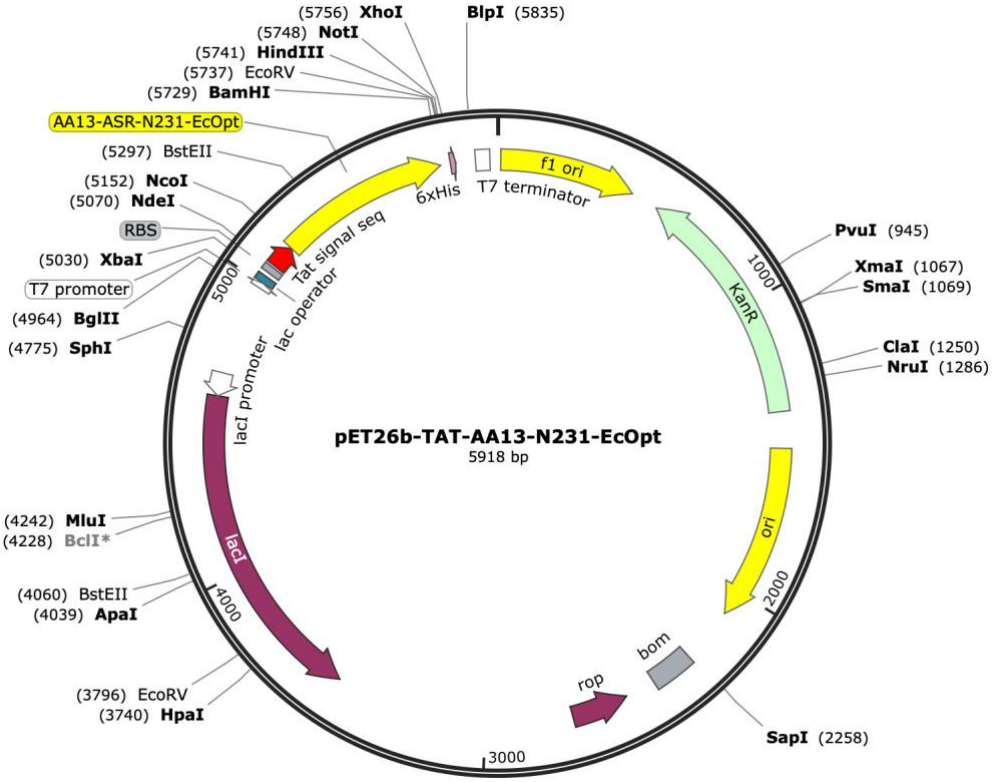


Figure 2.8: *E. coli* optimised pET26b plasmid of AA13 ASR-N231 with Tat signal peptide.

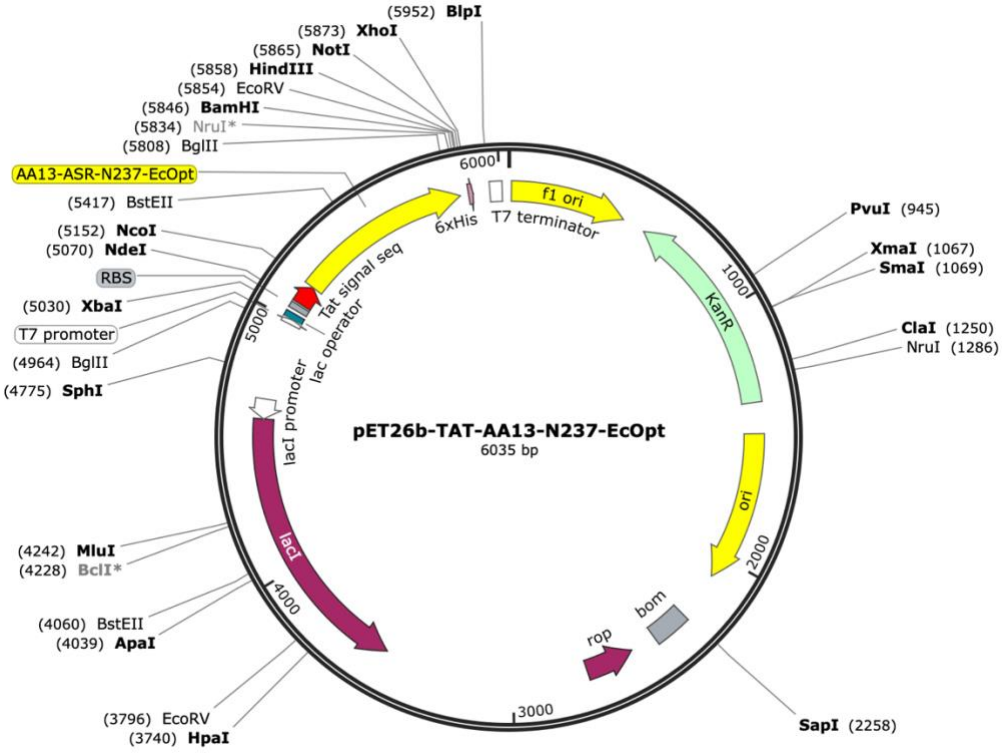


Figure 2.9: *E. coli* optimised pET26b plasmid of AA13 ASR-N237 with Tat signal peptide.

Diagnostic restriction digests of the N231 and N237 plasmids

Sec N231, Sec N237, Tat N231 and Tat N237 inserts were calculated to be 657 bp, 774 bp, 676 bp and 795 bp respectively. The restriction digest agarose gel (figure 2.10) indicated the inserts correlate to the expected size.

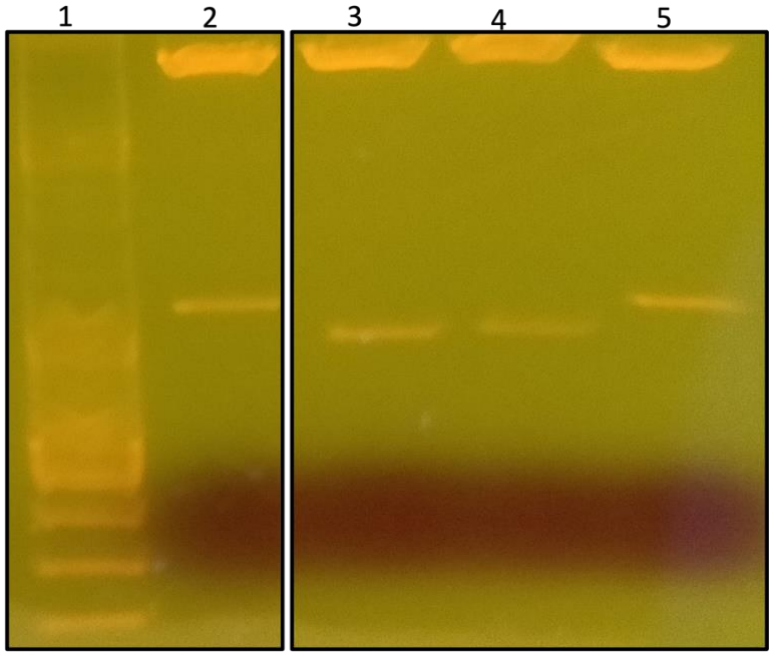


Figure 2.10: Agarose gel of AA13 DNA miniprep with the isolated plasmids restricted using *NdeI* and *HindIII*. 1. 1kb ladder, 2. Sec N237 plasmid, 3. Sec N231 plasmid, 4. Tat N231 plasmid, 5. Tat N327 plasmid.

2.3.2 – Over-expression tests for AA13 ASRs

ASR-N231 has a theoretical pI of 5.45 and an extinction coefficient of 41535 M^{-1} cm^{-1} at 280 nm (pH 6). ASR-N237 has a theoretical pI of 4.14 and an extinction coefficient of 54275 M^{-1} cm^{-1} at 280 nm (pH 4). Each plasmid was tested in Luria Broth (LB) and 2xYT media for every bacterial strain chosen (table 2.2). ASR-N231 was shown to express with Tat in BL21-RIL cells and Sec in BL21-DE3 and T7 SHuffle, whilst ASR-N237 did not express.

Table 2.2: Expression testing of AA13 plasmids, Tat N231, Tat N237, Sec N231 and Sec N237.

Plasmid	Bacterial Strain	Media	Band of Expected Mass*
Tat N231	BL21-(DE3)-RIL	LB	YES
		2xYT	YES
	C43-DE3	LB	NO
		2xYT	NO
	BL21-DE3	LB	NO
		2xYT	NO
	T7 SHuffle	LB	NO
		2xYT	NO
Tat N237	BL21-(DE3)-RIL	LB	NO
		2xYT	NO
	C43-DE3	LB	NO
		2xYT	NO
	BL21-DE3	LB	NO
		2xYT	NO
	T7 SHuffle	LB	NO
		2xYT	NO
Sec N231	BL21-(DE3)-RIL	LB	NO
		2xYT	NO
	C43-DE3	LB	NO
		2xYT	NO
	BL21-DE3	LB	YES
		2xYT	YES
	T7 SHuffle	LB	YES
		2xYT	YES
Sec N237	BL21-(DE3)-RIL	LB	NO
		2xYT	NO
	C43-DE3	LB	NO
		2xYT	NO
	BL21-DE3	LB	NO
		2xYT	NO
	T7 SHuffle	LB	NO
		2xYT	NO

*****Band of expected mass indicates whether a band on the SDS-PAGE at running is at or close to the mass corresponding to the expected molecular weight of the over-expressed LPMO (N231 Mw = 20,853 Da; N237 Mw = 24,847 Da).

Over-expression occurred in the ASR-N231 for both Tat and Sec. Tat N231 in the BL21-RIL (figure 2.11) strain for both Luria Broth (LB) and 2xYT media, and Sec N231 in the BL21-DE3 (figure 2.12) and T7 SHuffle (figure 2.13) strains for both media. Over-expression was not seen for the ASR-N237 in either Tat or Sec.

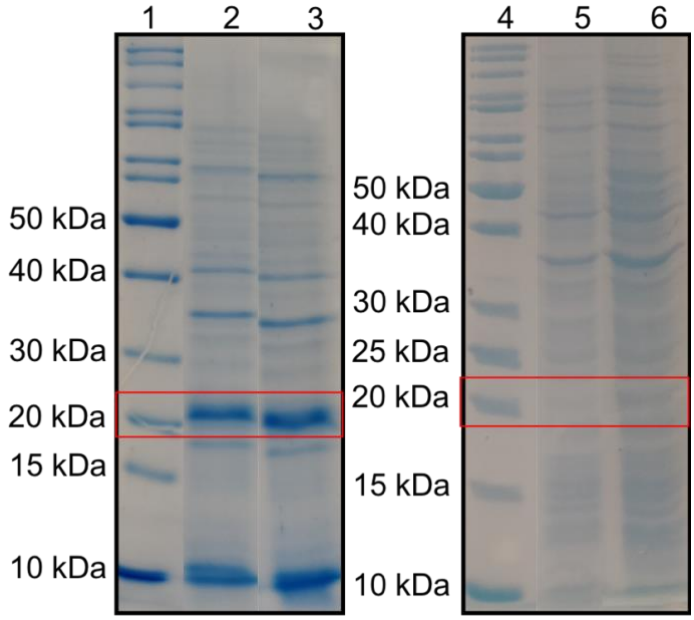


Figure 2.11: SDS-PAGE gel of Tat N231 plasmid in BL21-RIL using whole cell lysates. 1. Ladder, 2. Tat N231 LB at ~18 hours after induction, 3. Tat N231 2xYT at ~18 hours after induction, 4. Ladder, 5. pET28a LB control at ~18 hours after induction, 6. pET28a 2xYT control ~18 hours after induction. Gels were stained with Coomassie blue.

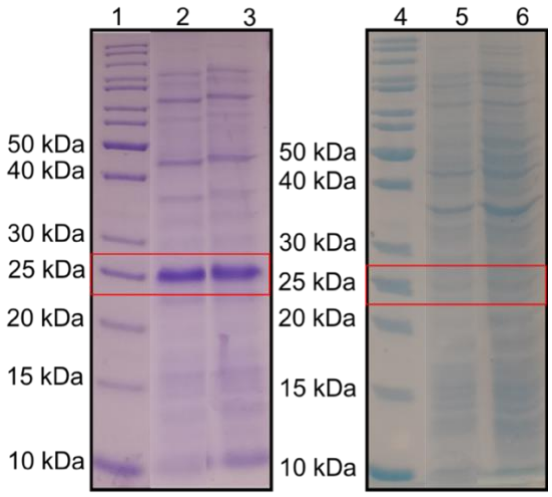


Figure 2.12: SDS-PAGE gel of Sec N231 plasmid in BL21-DE3 using whole cell lysates. 1. Ladder, 2. Sec N231 LB at ~18 hours after induction, 3. Sec N231 2xYT at ~18 hours after induction, 4. Ladder, 5. pET28a LB control at ~18 hours after induction, 6. pET28a 2xYT control ~18 hours after induction. Gels were stained with Coomassie blue.

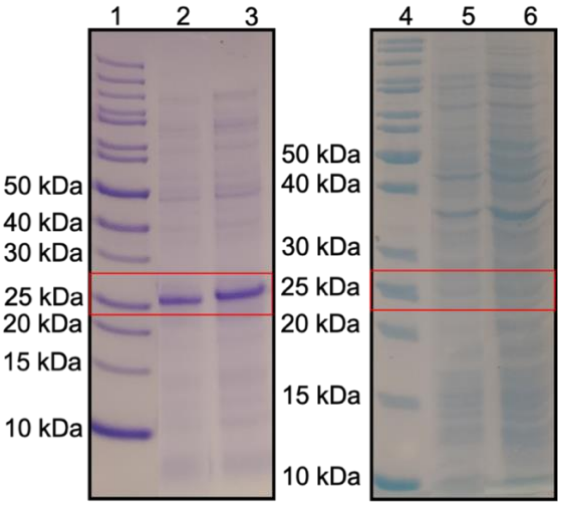


Figure 2.13: SDS-PAGE gel of Sec N231 plasmid in T7 SHuffle using whole cell lysates. 1. Ladder, 2. Sec N231 LB at ~18 hours after induction, 3. Sec 2N31 2xYT at ~18 hours after induction, 4. Ladder, 5. pET28a LB control at ~18 hours after induction, 6. pET28a 2xYT control ~18 hours after induction. Gels were stained with Coomassie blue.

Sucrose Shock

The dilution of the purified protein *Sli*LPMO10G was used as a control because the band of expected mass was known to be approximately 25 kDa.

Lysis of the successful Tat N231 over-expressed cells, indicated that the protein was lost after the resuspension in ice-cold water (figure 2.14). Initially the over-expressed band can be seen in lane 3 after the first centrifugation.

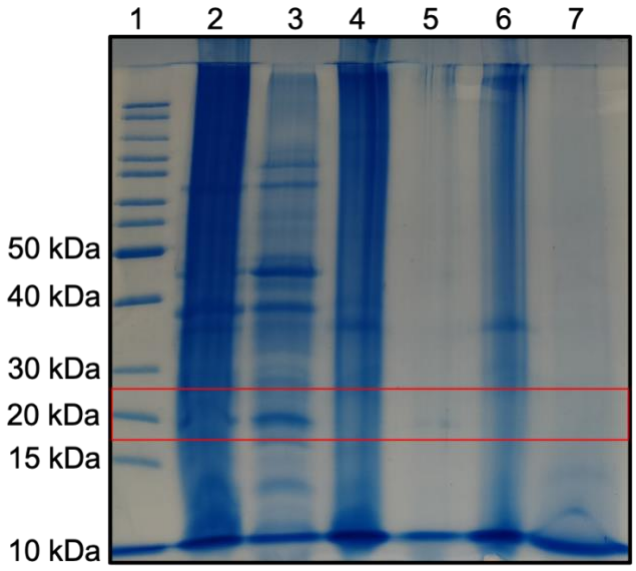


Figure 2.14: SDS-PAGE gel following the whole cell lysis of Tat N231 plasmid in BL21-RIL. The gel lanes correspond to 1) Protein ladder with molecular weight indicated, 2) cell paste after 1 hr of stirring, 3) supernatant after centrifugation, 4) resuspension in ice-cold water, 5) supernatant after final centrifugation, 6) pellet of final centrifugation, 7) concentrated *Sli*LPMO10G. Gels were stained with Coomassie blue.

Lysis of the successful Sec N231 over-expressed cells showed, that the protein was still in the pellet after the final centrifugation step. In BL21-DE3 (figure 2.15), the band of expected mass is shown in lane 3 from the initial over-expression at ~24 hrs. It can then be faintly seen in lane 4 during stirring of the cell paste. Lanes 5 and 6 indicated the protein was not present in the supernatant and a faint band near the expected mass is seen in the pellet after the final centrifugation.

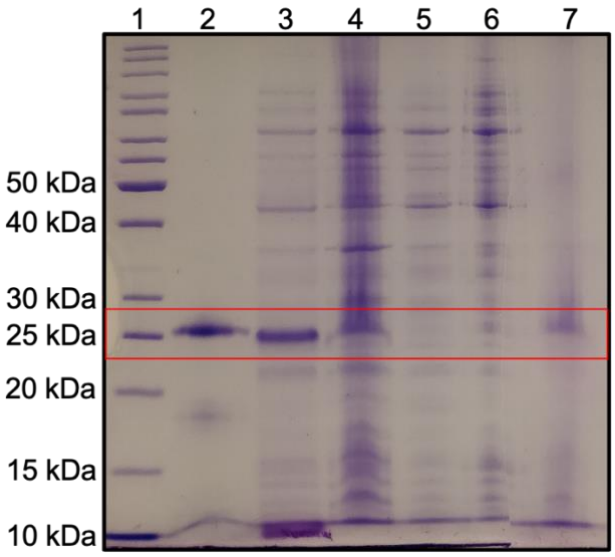


Figure 2.15: SDS-PAGE gel following the whole cell lysis of Sec N231 plasmid in BL21-DE3. The gel lanes correspond to 1) Protein ladder with molecular weight indicated, 2) 1:10 dilution of *Sli*LPMO10G protein 3) sample of LB from over expression at ~24 hrs, 4) cell paste after 1 hr of stirring, 5) supernatant after centrifugation, 6) supernatant after final centrifugation, 7) pellet of final centrifugation. Gels were stained with Coomassie blue.

In T7 Shuffle (figure 2.16) the visual of the gel made it difficult to determine where the band of expected mass was. Initially the band was indicated to be present in the supernatant after the final centrifugation, however on comparison with the control, there is also a thick band at approximately 15 kDa. In the sample of the pellet after the final centrifugation the band of expected mass is present at approximately 20 kDa when considering the sloping of the SDS-PAGE lanes.

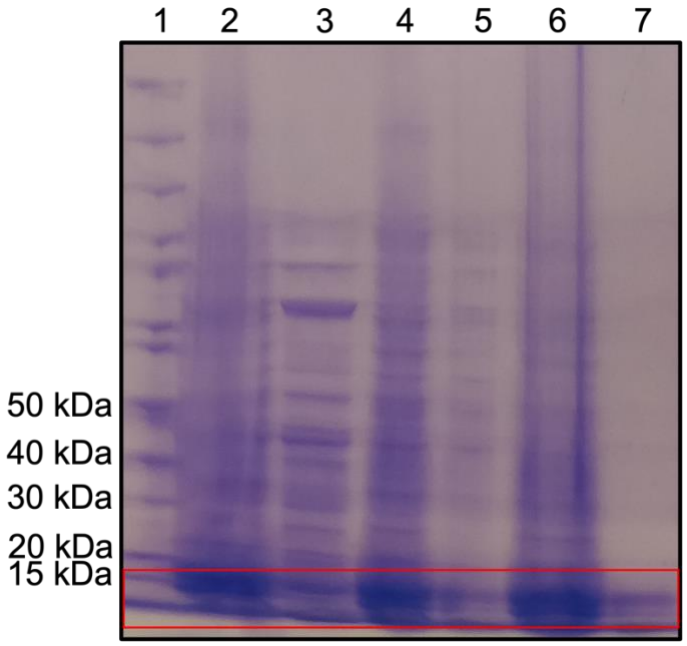


Figure 2.16: SDS-PAGE gel following the whole cell lysis of Sec N231 plasmid in T7 SHuffle. The gel lanes correspond to 1) Protein ladder with molecular weight indicated, 2) cell paste after 1 hr of stirring, 3) supernatant after centrifugation, 4) resuspension in ice-cold water, 5) supernatant after final centrifugation, 6) pellet of final centrifugation, 7) 1:10 dilution *Sli*LPMO10G. Gels were stained with Coomassie blue.

[2.4] Discussion

2.4.1 – Comparison of ASR-N231 and ASR-N237

The sequence alignment of the nodes N231 and N231 in figure 2.3, showed distinct differences in the deletion or addition of regions. ASR-N231 had a shorter sequence with the α2 region removed along with the regions between β7 and α4, and again with the shortening of β8 and removal of β9. The expression of ASR-N231 was possible with both Tat and Sec signal peptides in LB and 2xYT media and across three cell lines; i) BL21-RIL, ii) BL21-DE3 and iii) T7 SHuffle. ARS-N237 did not express in any of the cell lines chosen. As N237 had closer relation to the already characterised *A. oryzae* AA13 LPMO (Q2U8Y3_ASPOR), it would have been expected that expression would occur in ASR-N237 over ASR-N231. However, ASR-N237 also contains 14 cysteines over ASR-N231 containing only 10 cysteines. Whilst cysteine residues play an essential function in protein structure and function through disulphide bond formation, disulphide formation is difficult during protein expression which works more in favour of ASR-N231. Disulphide bonds provide structural support and stability for the protein and disulphide loss can cause mistargeting or malfunction (Morand *et al.*, 2004; Meitzler *et al.*, 2013; Namba *et al.*, 2022). With ASR-237 containing two more disulphide bridges than ASR-N231, it was expected to have greater structural stability. However, the alternate position of 4 cysteines in ASR-N231 could have had an impact on the protein structure, but it is hard to determine if this is truly the case.

2.4.2 – Method development

BL21-RIL cells were first used due to their success when over-expressing the AA10 LPMO, *Sli*LPMO10G, and successful over-expression was seen for Tat N231. C43-DE3 were then tried, as the strain was derived from BL21-DE3 cells and is efficient in expressing toxic proteins. As C43-DE3 did not express, BL21-DE3 cells were used, showing results for the Sec N231. Finally, T7 Shuffle cells were used as they are engineered to promote disulphide bond formation in the cytoplasm and ASR-N237 had so far not expressed even though it contained a higher number of cysteines than ASR-N231.

2.4.3 – Future development and limitations

After lysis, the SDS-PAGE gel of Tat N231 in BL21-RIL showed that the band of expected mass was present at approximately 20 kDa as predicted, after the first centrifugation but not during the cell pellet resuspension in ice-cold water. It could be that the protein was unstable and denatured during the resuspension. Alternative lysis methods should be trialled in future experiments. For Sec N231 in BL21-DE3, the SDS-PAGE gel after lysis had a band of expected mass present at approximately 25 kDa in the sample from the initial test expression, which was higher than the calculated mass expected of N231 of approximately 20 kDa. It was compared to a sample of another LPMO protein, *Sli*LPMO10G, from the AA10 family with a known mass of approximately 25 kDa. The Sec N231 in BL21-DE3 band was faintly present in the cell paste during resuspension. However, it was not present in the supernatant after centrifugation but again faintly seen in the cell pellet. This leads to the belief that a change in the method used for lysis should be considered. The use of the

Emulsiflex to better disrupt the cell membrane was discussed for when repeated. The SDS-PAGE gel of Sec N231 in T7 SHuffle after lysis also indicated that the protein was not present in the final supernatant but in the pellet, even though the gel was not very clear. The band of expected mass was however at approximately 20 kDa which was the expected mass of N231. The same alternative lysis method was discussed to use the Emulsiflex to better disrupt the cell membrane in the future.

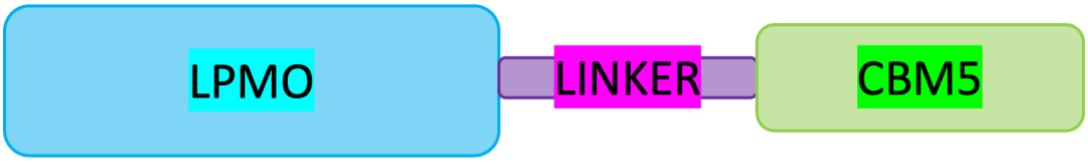
Time constraints meant that the repeat experiments and alternate lysis methods for Tat N231 and Sec N231 were not able to progress forward. If proper lysis of ASR-N231 were to occur, a full-scale expression and purification could be carried out. Further testing to express ASR-N237 may be possible, with changes to the cell line and expression method open to further development. As neither ASR-N231 or ASR-N237 were able to be fully purified, future work would include substrate binding for the fully purified ancestral reconstructions and their effect in starch-based biofuel batch processes.

Chapter 3: Characterisation of AA10 *Sli*LPMO10G

[3.1] Introduction

3.1.1 – AA10 family of LPMOs

In the CAZy database, the bacterial AA10 family of LPMOs is active on cellulose and chitin (Vaaje-Kolstad *et al.*, 2010). Like other LPMOs, AA10s use are activated by a copper cofactor and an electron donor to oxidatively cleave polysaccharide chains. *Streptomyces lividans* is a gram-positive soil dwelling bacteria which contains seven genes that encode an LPMO domain which are categorised in the AA10 family. Three of these genes contain an LPMO domain appended to a carbohydrate binding model (CBM) by a short linker. The gene encoding *Sli*LPMO10G (figure 3.1) is the protein of primary focus for this study and is one of the three LPMO genes containing an LPMO domain and CBM. *Sli*LPMO10G has the module CBM5 appended to it by a linker.



HGWVTSPGSRQDQCAAGIVDCGQIKYEPQSVEGPKGLSSCSGGN
SQFAELDDDGKGWRVTPVGTTHTFTWHHTARHATANWQYFIGSQ
KIAEFDGHRAQPAPDVSHQVNFGGFTGRQKVLAVWNVADTGNAF
YSCIDVNIGGSGGGDGGGDGGGDGDPGACDAPGWSAGSVYSGG
DTVSFGGHTWRAKWWVTGEEPGTTGEWGVWEDLGACAAA

Figure 3.1: Schematic of *Sli*LPMO10G with its LMPO domain (cyan), linker region (purple) and CBM5 domain (green). The corresponding amino acid sequence is shown below using one-letter-code. The yellow highlighted area indicates aromatic chitin/cellulose binding site residues.

The LPMO domain is found between residues 1 and 139; the linker between residues 140 and 163 and the CBM5 domain is found between residues 164 and 214. The CBM5 domain of *Sli*LPMO10G specifically binds to chitin. Currently, only one LPMO from *S. lividans* has been structurally characterised, *Sli*LPMO10E (Chaplin *et al.*, 2016). *Sli*LPMO10E did not have a CBM domain appended to the LPMO domain unlike *Sli*LPMO10G.

3.1.2 – Carbohydrate binding modules and the applications of AA10 LPMOs in biofuel processes

Second generation biofuel production uses non-edible biomass such as grasses, wood or straw. Pre-treatment opens the lignocelluloses in order to access the polymer chains of cellulose and hemicellulose, whilst the hydrolysis of the polymers achieves the release of the monomer sugars (Taherzadeh and Karimi, 2007; Ramos *et al.*, 2016). LPMOs breakdown the amorphous cellulose chains and therefore increase surface accessibility during the process (Song *et al.*, 2018). LPMOs also increase the degradation of insoluble xylan which suggests the ability to oxidise sugars in different polysaccharides (Kim *et al.*, 2016). When LPMO activity was simulated in a commercial enzyme cocktail, increased saccharification yield was observed when adapting current processing conditions to promote LPMO activity (Müller *et al.*, 2015). Carbohydrate binding modules are typically part of a multi-modular enzyme, contributing to substrate-binding of the LPMO whilst possibly affecting operational stability (Ficko-Blean and Boraston, 2012; Forsberg *et al.*, 2016, 2018). CBMs are devoid of any catalytic activity and are primarily linked to catalytic modules (Levasseur *et al.*, 2013; Lombard *et al.*, 2014). Currently, little work has

been carried out on characterising LPMO-CBM systems, however, their application in biotechnological areas have been noted by Cuskin *et al.*, 2012, Reyes-Ortiz *et al.*, 2013 and Sidar *et al.*, 2020. Investigating the relationship between the LPMO-CBM system could see the utilisation of CBM features to create designer LPMOs with enhanced or modified substrate binding recognition features and consequently the enhancement of LPMOs for industrial biofuel processing. With AA10 LPMO specificity towards cellulose and chitin, combined with the enhancement of substrate specificity using CBMs, there is the possibility to increase the yield and lower feedstock costs of biofuel production.

This study aims to describe the truncation of *Sli*LPMO10G from *Streptomyces lividans* to report the binding characteristics of the full protein and the respective LPMO and CBM5 domains, along with the subsequent structure determination. It is hoped that the separation of the LPMO and CBM5 domains and substrate binding test of the three protein forms will give a better understanding to the LMPO-CBM substrate binding relationship and specificity.

[3.2] Method

3.2.1 – Protein expression and purification of the full and truncated AA10 SliLPMO10G

Competent cells and plasmid transformation

Competent cells were prepared for the *E. coli* strain BL21-RIL. Pre-cultures of 10 mL LB (Luria Broth) were inoculated with scrapings of pre-made glycerol stock containing 750 μL overnight BL21-RIL pre-culture and 750 μL glycerol (50% glycerol) and incubated overnight. 100 mL LB (Luria Broth) media was inoculated with 2 mL from the overnight cultures and the optical density (OD) was checked. The cells were allowed to grow for 2 to 3 hours at 37 °C, 180 RPM until OD_{600} reached between 0.6 to 1.0. The cells were placed on ice for 20 minutes and the volume equally divided between two 50 mL falcon tubes. Centrifugation of the cells occurred at 4000 RPM for 10 minutes and the supernatant was discarded. Each pellet was resuspended in 10 mL ice-cold 0.1 calcium chloride ($CaCl_2$) and combined before being incubated on ice for 30 minutes. After incubation, the solution was centrifuged at 2500 RPM for 10 minutes and the supernatant discarded. The pellet was resuspended in 4 mL ice-cold 0.1 M $CaCl_2$ containing 15% glycerol and the final solution transferred to 1.5 mL Eppendorf tubes in 100 μL aliquots. The *Sli*LPMO10G was a modified pET26b plasmid (Chaplin *et al.*, 2016) containing a kanamycin (Kan) (Fisher) selection marker and the BL21-RIL cell line contained a chloramphenicol (CM) resistance cassette and thus required the presence of CM at final concentration 20 μg mL^{-1} along with 50 μg mL^{-1} Kan. The *Sli*LPMO10G construct was transformed into BL21-RIL *Escherichia coli* cells for protein overexpression. 2 μL of plasmid was mixed gently with one 100 μL aliquot of chemically competent cells and incubated on ice for

30 minutes. Transformation occurred via heat shock at 42 °C for 45 seconds, then placed on ice for a further 2 minutes. 800 μL of LB media was added to the cells and then incubated at 37 °C for 75 minutes. The heat shock transformation was grown on LB agar plate with a final concentration of 50 μg mL^{-1} Kan and 20 μg mL^{-1} CM at 37 °C overnight.

Over-expression and sucrose shock

Overnight cultures were made by picking individual colonies from the transformation plate and inoculating 3 mL LB medium (Melford) containing 50 μg mL^{-1} Kan and 20 μg mL^{-1} CM. The inoculated vials of pre-culture were incubated at 37 °C at 225 RPM overnight. Then 500 μL of 50 μg mL^{-1} Kan and 20 μg mL^{-1} CM were added to four 2 L flasks containing 500 mL LB media all of which were inoculated with 3 mL of pre-culture and incubated at 37 °C, 180 RPM until an optical density at 600 nm (OD_{600}) of 0.4 was reached when the temperature was lowered to 16 °C. When an OD_{600} at 0.6 was reached the cultures were induced with 500 μL from a stock solution of 0.5 M isopropyl-βD-1-thiogalactopyranoside (IPTG)(Melford) to final concentration of 1 mM and left to incubate at 16 °C with shaking overnight. The cells were harvested by centrifugation at 4000 RPM for 20 minutes at 4 °C in a Beckman Coulter Avanti JXN-26 centrifuge using a JLA-8.1000 rotor and the resulting pellets were combined and resuspended in ice-cold 50 mM Tris/HCl, 1 mM EDTA, 20% w/v sucrose, pH 8. Then at 4 °C the cell suspension was stirred for 1 hour and 60 μL of 1 M $MgSO_4$ was added for every gram of cell pellet. This cell paste was stirred at 4 °C for 30 minutes and centrifugation of the resulting cell paste was then carried out at 18,000 RPM for

20 minutes at 4 °C using a JA-25.50 rotor. The resulting supernatant was collected and combined with that from the previous sucrose fractionation. Dialysis overnight was carried out against 5 mM Bis Tris Propane/HCl 1 mM EDTA, pH 6 for the full *Sli*LPMO10G and pH 7 for the truncated *Sli*LPMO10G, at 4 °C with the combined supernatants.

Ion exchange chromatography and gel filtration

*Sli*LPMO10G (pI 4.85) was purified on a Q-Sepharose Hi trap column (GE Healthcare) via ion exchange chromatography. Before being loaded onto the column, the protein was diluted by half in Buffer A (10 mM Bis Tris Propane, pH 6) and then eluted using a salt gradient with Buffer B (10 mM Bis Tris Propane w/ 1 M NaCl, pH 6) using an AKTA Prime. The fractions were collected and analysed via SDS-PAGE and the factions containing the protein were concentrated to 2 mL at 4 °C, 4,200 RPM in Centricon tubes (Vivaspin) with 5 kDa cut-off. The truncated *Sli*LPMO10G (pI 6.13) was purified on a DEAE column (GE Healthcare) via ion exchange chromatography. In much the same way as the full *Sli*LPMO10G was purified, Buffer A (10 mM Bis Tris Propane, pH 7) was used to dilute the protein by half before being loaded onto the column and then eluted using a salt gradient with Buffer B (10 mM Bis Tris Propane w/ 1 M NaCl, pH 7) using an AKTA Prime. The 2 mL concentrated protein was injected onto a 120 mL Sephadex G75 column (GE Healthcare) and equilibrated in Buffer C (10 mM Sodium Acetate, 150 mM NaCl, pH 6) for the full *Sli*LPMO10G and Buffer C (10 mM Sodium Acetate, 150 mM NaCl, pH 7) for the truncated *Sli*LPMO10G. The selected fractions in the major elution peak at 280 nm

were examined for purity on 12% SDS-PAGE and were pooled, concentrated and stored at -20 °C.

SDS-PAGE

Lab made gels (12%) were prepared, the following volumes in table 3.1 were used to make two 12 % acrylamide gels. Samples for 20 μL samples were loaded onto the gels. Samples were prepared before and after IPTG induction with 15 μL of sample and 5 μL cracking buffer. Gels were run in a Mini-PROTEAN Tetra Cell (BioRad) system in a x1 dilution of 1 L x10 Running Buffer (30 g TRIS, 144.45 g Glycine, 5 g SDS) at 200 V for 40 mins. Once gels were run, they were stained in Coomassie blue, (455 mL ethanol, 455 mL ddH_2O, 90 mL glacial acetic acid and 2.5 g Coomassie blue in 1 L) for 30 min followed by de-staining (455 mL ethanol, 455 mL ddH_2O and 90 mL glacial acetic acid in 1 L) until bands were visible. The gels were then left in water overnight to clean before images were taken.

Table 3.1: SDS PAGE 12 % gel volumes to make 2 gels.

Media	Resolving Gel	Stacking Gel
ddH_2O	6.4 mL	5.95 mL
1.5 M Tris pH 8.8	5.2 mL	-
0.5 M Tris pH 6.8	-	2.5 mL
Acrylamide/Bisacrylamide (30 % / 0.8 %, w/v)	8 mL	1.34 mL
10 % SDS	200 µL	100 µL
10% APS	200 µL	100 µL
TEMED	20 µL	10 µL

Sample preparation for SDS-PAGE

Samples for SDS-PAGE from over-expression tests were taken from the cultures at defined times. A 50 μL aliquot of the over-expression was taken and spun in a microfuge (13,000 RPM) for 1 min, followed by removal of the supernatant and the resuspension of the cell pellet in 50 μL cracking buffer (1 mL of 1mM TRIS pH 6.8, 2 mL of 10 % SDS, 5 mL glycerol, 11.8 mL ddH_2O, 10 mg 0.05 % bromophenol blue and containing a few crystals of DTT added before use). A sample of 6 μL was taken from the resuspension and 5 μL cracking buffer and 9 μL ddH_2O were added. Lysis supernatant samples were prepared with 20 μL supernatant and 5 μL cracking buffer. The lysis pellet sample was first mixed with 50 μL cracking buffer to prepare. A 4 μL sample was taken and added to 5 μL cracking buffer and 11 μL ddH_2O. Samples were denatured at 90 °C for 10 mins after preparation.

3.2.2 – Copper Binding of the SliLPMO10G and truncated SliLPMO10G

The purified full and truncated *Sli*LPMO10G do not have copper bound. As *Sli*LPMO10G is a copper active protein, fluorescence spectroscopy was used to monitor the uptake of Cu(II) into the protein. For preparation, 200 mL of copper II sulphate (Cu(II)SO_4) buffer was made at 100 mM concentration using water to stop precipitation of the copper sulphate when mixed with Buffer C. This stock solution was diluted down to a 1 mM concentration for the titration. The sample buffer (10 mM sodium acetate, 150 mM NaCl, pH 6) was prepared and 990 μL of the buffer and 10 μL protein used to measure the absorbance and respective concentration of the stored protein samples. Using these concentrations, the amount of protein to be added at a concentration of 2 μM for the truncated *Sli*LPMO10G and the full

*Sli*LPMO10G in of total sample of 4 mL was calculated. The resulting samples were individually combined with 0.2 equivalents, 1.2 μL of $Cu(II)SO_4$ at 1 mM concentration, for titration. The protein samples were consistently mixed twice after the addition of the $Cu(II)SO_4$ aliquots. The Fluorimeter was left for 30 minutes before being used to warm up after turning on. During this time, samples were left to sit at room temperature. Measurement settings were changed with the emission spectra set between 300 nm and 400 nm with excitation at 295 nm. Excitation and emission slit widths were set at 1.5 nm and 5 nm respectively whilst sensitivity was set to low and data intervals set to 1.0 nm.

3.2.3 – Model building of the truncated SliLPMO10G and refinement using CCP4i2

Obtaining the truncated domain

The *Sli*LPMO10G is 214 residues in length with the LPMO domain consisting of the residues between 1 and 139. The LPMO domain was truncated during a Biochemical Society's Summer Studentship in 2017. Crystals for the domain were obtained in 0.1 M Tris pH 8, 32% PEG 6K conditions with protein concentration 13.905 mg mL^{-1} by Megan Straw, PhD and supervisor, Prof. Jonathan Worrall after the Biochemical Society Studentship. Batch crystallisation used a 1:1 protein to precipitant solution to give a final volume of 50 μL in a 1.5 mL Eppendorf tube. Crystals were grown at 18 °C for approximately 3 weeks and then frozen using a 40% sucrose cryoprotectant. These crystals diffracted to 1.04 Å resolution at Diamond Light Source.

Generating a Free R set and defining crystal contents

Using the CCP4i2 suite (Winn *et al.*, 2011; Potterton *et al.*, 2018), a new set of Free R flags was generated using the reflections data. Crystal content was defined by inputting the known amino acid sequence for the *Sli*LPMO10G LPMO domain.

Refmac5 and COOT

The model generated was refined using Refmac5 (Murshudov *et al*., 1997; Vagin *et al.*, 2004) and then opened in COOT (Emsley *et al.*, 2010) for model building and correction. Refinement was carried out at 1.04 Å resolution. B-factors were set to anisotropic with custom resolution limits of 38.89 to 1.04 and weight restraints were initially set for 0.1 until the 33rd job where it was changed to automatic. COOT density display was kept at approximately 1 σ and the amino acid sequence of both subunits were checked for errors and sections that needed to be rebuilt. The Model/Fit/Refine bar was used to remodel the structure to better fit the density. Mutate & Auto Fit changed the amino acid; Real Space Refine Zone moved the amino acid and Regularize Zone regulated the position within the model. Refmac5 was run regularly to ensure that progress was being made and the R_{free} was decreasing as the modelling progressed. After the amino acid sequence was built, waters were modelled into the density map through COOT and the checked for errors. Water was deleted in the model that had been added to sections with no density or in density too far for interaction and was manually added to the model in density that had been missed using Place Atom At Pointer and Rotate/Translate. The copper molecules seen as high density at the Cu binding sites of both subunits and within the interface between the subunits were also added to the model.

3.2.4 – Analysis of tertiary structure using PyMOL

The final Refmac5 output file was opened in PyMOL (v2.5.4) (Schrödinger LLC) and the tertiary structure analysed. The bond distances of Cu at the binding site were measured and figures were created to show the asymmetric unit, monomer, binding site and surface representation of the truncated *Sli*LPMO10G.

3.2.5 – Protein expression and purification of the AA10 SliLPMO10G CBM domain

PCR, restriction digest, ligation and agarose gels

Forward and reverse primers were prepared for the *Sli*LPMO10G CBM domain and diluted with H_2O, 10 μL of primer and 90 μL of H_2O. Samples for PCR were prepared using volumes as shown in table 3.2. Two programs of PCR were run at annealing temperatures 58 °C and 62 °C and compared using an agarose gel. The PCR program was set with initial denaturation at 95 °C for 3 minutes, then 95 °C for 1 minute, 58 °C / 62 °C at 1 minute for the respective runs, 72 °C for 1 minute, was cycled x35 and the final extension was set at 72 °C for 8 minutes. The final hold was 10 °C whilst the run was complete overnight. Agarose gels (1%) were prepared using 50 mL 1x TBE and 0.5 g Agarose. Sample consisted of 5 μL of the PCR sample added to 5 μL GelRed and 2 μL 6x loading dye. A control was made using 1 μL of the forward and reverse primers mixed with 50 μL H_2O.

Table 3.2: PCR sample volumes in μL.

Media	Volume (μL)
*Sli*LMPO10G DNA	1
Forward primer	1
Reverse primer	1
dNTP	2.5
DMSO	5
DNA polymerase	0.5
10x Buffer Pfn + $MgSO_4$	5
ddH_2O	34

Gel extraction was carried out on the two samples, all the restriction digest was loaded on to the gel and cut out as close to the expected bands as possible. The amount of gel cut that contained the digest was weighed and the appropriate volume of NEI buffer added, as per the gel and PCR clean up kit (NucleoSpin). However, ddH_2O was used for elution instead of the prepared elution buffer. Then a restriction digest was carried out using the primers *NdeI* and *HindIII* with *NdeI* at 2x concentration. The first digests and second digests were done following the amounts in table 3.3 to a total of 50 μL and 30 μL respectively.

Table 3.3: Restriction digest sample volumes for 50 and 30 µL using the pET28a plasmid vector, the PCR CBM insert, *NdeI*, *HindIII*, Buffer red x 10 and H_2O.

Total Volume (µL)	Template	Volume (µL)	
30	Plasmid vector pET28a	20	-
	PCR CBM insert	-	20
	NdeI	4	4
	HindIII	2	2
	Buffer red x10	3	3
	H_2O	1	1
50	Plasmid vector pET28a	25	-
	PCR CBM insert	-	25
	NdeI	4	4
	HindIII	2	2
	Buffer red x10	5	5
	H_2O	14	14

The samples were then incubated for 2 hours at 37 °C and run on an Agarose gel for gel extraction and PCR clean up (NucleoSpin). Ligation of the plasmid digest and insert fragment was carried out at 1 : 1 and 1 : 0.5 volumetric ratios. The 1 : 1 ratio included 5 µL pET28a plasmid, 5 µL CBM insert, 1.5 µL T4 ligase buffer 10x, 1 µL T4 ligase and 2.5 µL H_2O whilst the 1 : 0.5 ratio used 5 µL plasmid, 2.5 µL of the insert, 1.5 µL T4 ligase buffer 10x, 1 µL T4 ligase and 5 µL H_2O. The samples were incubated at 22 °C for 1 hour. Then 7.5 µL of each sample was transformed into DH5α cells. When plating the 1 : 1 and 1 : 0.5 transformations onto agar, a

concentrated plate was also plated. The remaining 7.5 µL of the ligation was left at room temperature overnight and transformed the next afternoon into XL1Blue cells. The overnight ligation was combined, denatured at 62 °C for 1 hour and stored at -20 °C. A DNA miniprep of the 1 : 1 ligation plate was carried out, then a restriction digest of the DNA miniprep before another agarose gel was run to check the expected band mass. The resulting DNA was then sent for sequencing, using a T7 forward primer. Two more ligations were carried out at 5 : 1 and 6 : 1 ratios, using 1 µL of plasmid and 5 µL or 6 µL of fragment respectively. 1.5 µL of T4 ligase buffer 10x and 1 µL T4 ligase was added to both and then each sample made up to 15 µL using H_2O.

Competent cells and plasmid transformation

Competent cells were prepared for the *Escherichia coli* strain BL21-DE3. Pre-cultures of 10 mL LB (Luria Broth) were inoculated with scrapings of pre-made glycerol stock, containing 750 µL overnight BL21-DE3 pre-culture and 750 µL glycerol (50% glycerol), and incubated overnight. 100 mL LB (Luria Broth) media was inoculated with 2 mL from the overnight cultures and the optical density (OD) was checked. The cells were allowed to grow for 2 to 3 hours at 37 °C, 180 RPM until OD_{600} reached between 0.6 to 1.0. The cells were placed on ice for 20 minutes and the volume equally divided between two 50 mL falcon tubes. Centrifugation of the cells occurred at 4000 RPM for 10 minutes and the supernatant was discarded. Each pellet was resuspended in 10 mL ice-cold 0.1 calcium chloride ($CaCl_2$) and combined before being incubated on ice for 30 minutes. After incubation, the solution was centrifuged at 2500 RPM for 10 minutes and the supernatant discarded. The pellet

was resuspended in 4 mL ice-cold 0.1 M $CaCl_2$ containing 15% glycerol and the final solution transferred to 1.5 mL Eppendorf tubes in 100 μL aliquots.

The *Sli*LPMO10G was a modified pET26b plasmid (Chaplin *et al.*, 2016) containing a kanamycin (Kan) selection marker and the BL21-DE3 cell line contained a chloramphenicol (CM) resistance cassette and thus required the presence of CM at final concentration 20 μg mL^{-1} along with 50 μg mL^{-1} Kan. The truncated *Sli*LPMO10G CBM construct was transformed into BL21-DE3 *E. coli* cells for protein overexpression. 2 μL of plasmid was mixed gently with one 100 μL aliquot of chemically competent cells and incubated on ice for 30 minutes. Transformation occurred via heat shock at 42 °C for 45 seconds, then placed on ice for a further 2 minutes. 800 μL of LB media was added to the cells and then incubated at 37 °C for 75 minutes. The heat shock transformation was grown on LB agar plate with a final concentration of 50 μg mL^{-1} Kan and 20 μg mL^{-1} CM at 37 °C overnight.

Over-expression and Purification

Over-expression tests were run for the *Sli*LPMO10G CBM domain in BL21-DE3 and BL21-RIL in Luria Broth (LB; Melford) and 2xY (Melford Chemicals). The BL21-RIL cell line contained a chloramphenicol (CM) resistance cassette and thus required the presence of 20 μg mL^{-1} CM in addition to 50 μg mL^{-1} Kanamycin (Kan) in LB agar plates and cultures. The desired plasmid was transformed to the selected *E. coli* strain. Overnight transformants were selected from an LB/agar/Kan plate, or LB/agar/Kan/CM where appropriate and used to inoculate a 3 mL overnight LB culture containing a final concentration of 50 μg mL^{-1} of Kan, and 20 μg mL^{-1} CM for

BL21-RIL. The inoculated vials of pre-culture were incubated at 37 °C, 225 RPM overnight. Flasks (100 mL) containing 40 mL LB or 2xYT, with Kan final concentration 50 μg mL^{-1} and CM final concentration 20 μg mL^{-1} where appropriate, were inoculated with 400 μL from pre-cultures and grown throughout the day at 37 °C, 200 RPM. At an OD_{600} of between 0.4 and 0.5, isopropyl β-D-1-thiogalactopyranoside (IPTG; Melford Chemicals) was added from a stock solution of 0.5 M to give a final concentration of 100 μM in LB and 250 μM in 2xYT. Growth continued overnight at 25 °C and 200 RPM.

Over-expression for the *Sli*LPMO10G CBM domain was carried out using BL21-DE3 cells in five 2 L flasks of 750 mL LB and two 2 L flasks of 750 mL 2xYT and 750 μL of 50 μg mL^{-1} Kan. All flasks were inoculated with 3 mL of pre-culture and incubated at 37 °C, 180 RPM until an optical density at 600 nm (OD_{600}) of 0.4 was reached when the temperature was lowered to 25 °C. When an OD_{600} at 0.6 was reached, the cultures were induced with a stock solution of 0.5 M isopropyl-βD-1-thiogalactopyranoside (IPTG)(Melford) to final concentration of 100 μM in LB and 250 μM in 2xYT and both then left to incubate at 25 °C with shaking overnight. The cells were harvested by centrifugation at 4000 RPM for 20 minutes at 4 °C in a Beckman Coulter Avanti JXN-26 centrifuge using a JLA-8.1000 rotor and the resulting pellets were combined and stored at -80 °C. The stored pellets were thawed and stirred, then passed through the Emulsiflex to lyse the cells using ice-cold 50 mM Tris/HCl, 1 mM EDTA, 20% w/v sucrose, pH 8 buffer. Centrifugation of the resulting cell paste was then carried out at 18,000 RPM for 30 minutes at 4 °C using a JA-25.50 rotor.

Immobilized metal affinity chromatography and gel filtration

The truncated *Sli*LPMO10G CBM domain was purified using a Nickel Column via immobilized metal affinity chromatography (IMAC). Buffer A (50 mM Tris/HCl, 500 mM NaCl, 20 mM Imidazole, pH 7.5) was used to dilute the protein by half before being loaded onto the column and then eluted using Buffer B (50 mM Tris/HCl, 500 mM NaCl, 500 mM Imidazole, pH 7.5) using an AKTA Prime. The fractions were collected and analysed via SDS-PAGE and the factions containing the protein were concentrated to 2 mL at 4 °C, 4,200 RPM in Centricon tubes (Vivaspin) with 5 kDa cut-off. The 2 mL concentrated protein was injected onto a 120 mL Sephadex G75 column (GE Healthcare) and equilibrated in Buffer C (50 mM Tris/HCl, 100 mM NaCl, pH 7). The selected fractions in the major elution peak at 280 nm were examined for purity on 15% SDS-PAGE and were pooled, concentrated to approximately 2 µL and stored at -20 °C.

SDS-PAGE and SDS-PAGE sample preparation

Lab made gels (15%) were used, following the volumes in table 3.4 to make two 15 % acrylamide gels. Samples of 20 µL were loaded onto the gels and prepared before and after IPTG induction with 15 µL of sample and 5 µL cracking buffer. Gels were run in a Mini-PROTEAN Tetra Cell (BioRad) system in a x1 dilution of 1 L x10 Running Buffer (30 g TRIS, 144.45 g Glycine, 5 g SDS) at 200 V for 40 mins.

Table 3.4: SDS PAGE 15 % gel volumes to make 2 gels.

Media	Resolving Gel	Stacking Gel
ddH_2O	2.3 mL	2.25 mL
1.5 M Tris pH 8.8	2.5 mL	-
0.5 M Tris pH 6.8	-	1 mL
Acrylamide/Bisacrylamide (30 % / 0.8 %, w/v)	5 mL	666 µL
10 % SDS	100 µL	40 µL
10% APS	100 µL	40 µL
TEMED	10 µL	5 µL

Once gels were run, they were stained in Coomassie blue, (455 mL ethanol, 455 mL ddH_2O, 90 mL glacial acetic acid and 2.5 g Coomassie blue in 1 L) for 30 min followed by de-staining (455 mL ethanol, 455 mL ddH_2O and 90 mL glacial acetic acid in 1 L) until bands were visible. The gels were then left in water overnight to clean before images were taken.

Crystallisation trials

Following protein purification, the CBM was set up for crystallisation trials using a Gryphon crystallisation robot. Trials were set up in 96-well sitting-drop plates (figure 3.2).

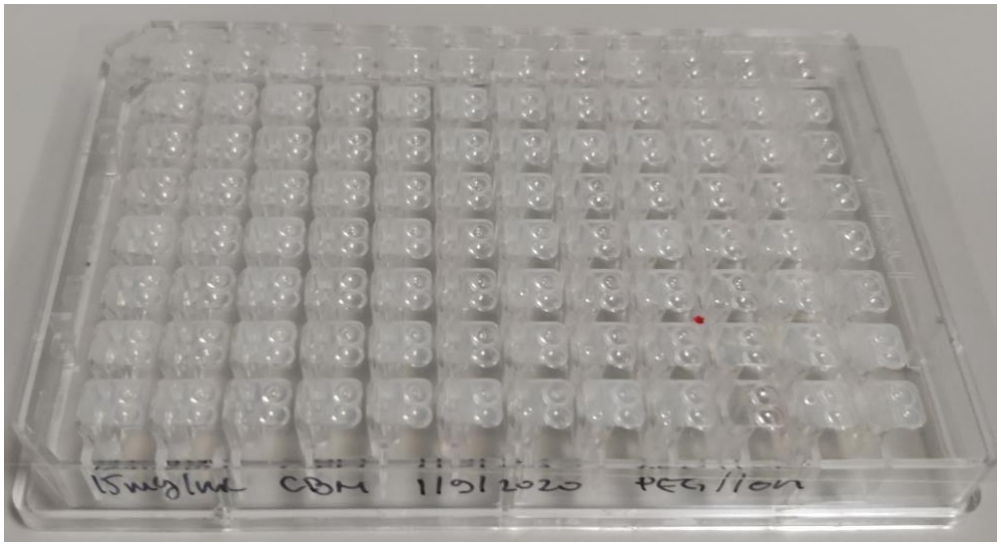


Figure 3.2: CBM crystal screen PEG/Ion 96-well plate.

Seven 96-well crystallisation screens were prepared; pH clear, PGA screen, PACT, ProPLex, Peg/Ion, JCSG – plus and Structure screen. The plates were incubated at 20 °C and each well was checked for crystallisation using a light microscope. The ratio of protein to reservoir was 1 : 1, the reservoir volume 75 nL and the drop size 20 nL.

3.2.6 – Determine the physiochemical properties of the SliLPMO10G domains

The online programme ProtParam (Gasteiger *et al.*, 2005) part of the ExPASy Bioinformatics Resource Portal (Duvaud *et al.*, 2021) was used to calculate molecular weight, pI and extinction coefficients at 280 nm using the amino acids sequences of the respective *Sli*LPMO10G domains as input.

[3.3] Results

3.3.1 – Characterisation of the full length and truncated LPMO domain

Purification of full length SliLPMO10G

The full length *Sli*LPMO10G was over-expressed in BL21-RIL and purified using ion exchange chromatography and gel filtration. The ion exchange absorbance was relatively messy with multiple peaks seen, so fractions were taken from the column and run on SDS-PAGE gels. Based on the band of expected mass, the full length *Sli*LPMO10G was observed to be present in the peak labelled A (figure 3.3), corresponding to between fractions 12 and 14 taken between 40 and 50 minutes (figure 3.4). However, the band of expected mass could also be seen across fractions 17 to 22 collected between 50 and 70 minutes, although there was more interference from other protein bands (figure 3.5). The protein began to unbind from the column when the buffer B salt gradient reached 50 %. All fractions collected between 40 and 70 minutes were then pooled and prepared through concentration of the collected fractions down to 2 mL for gel filtration.

The gel filtration absorbance spectra only showed one peak, labelled A (figure 3.6) between 70 to 110 mL. The corresponding fractions between 25 and 38 had samples taken to run on an SDS-PAGE in order to confirm the presence of the full length *Sli*LPMO10G protein (figure 3.7).

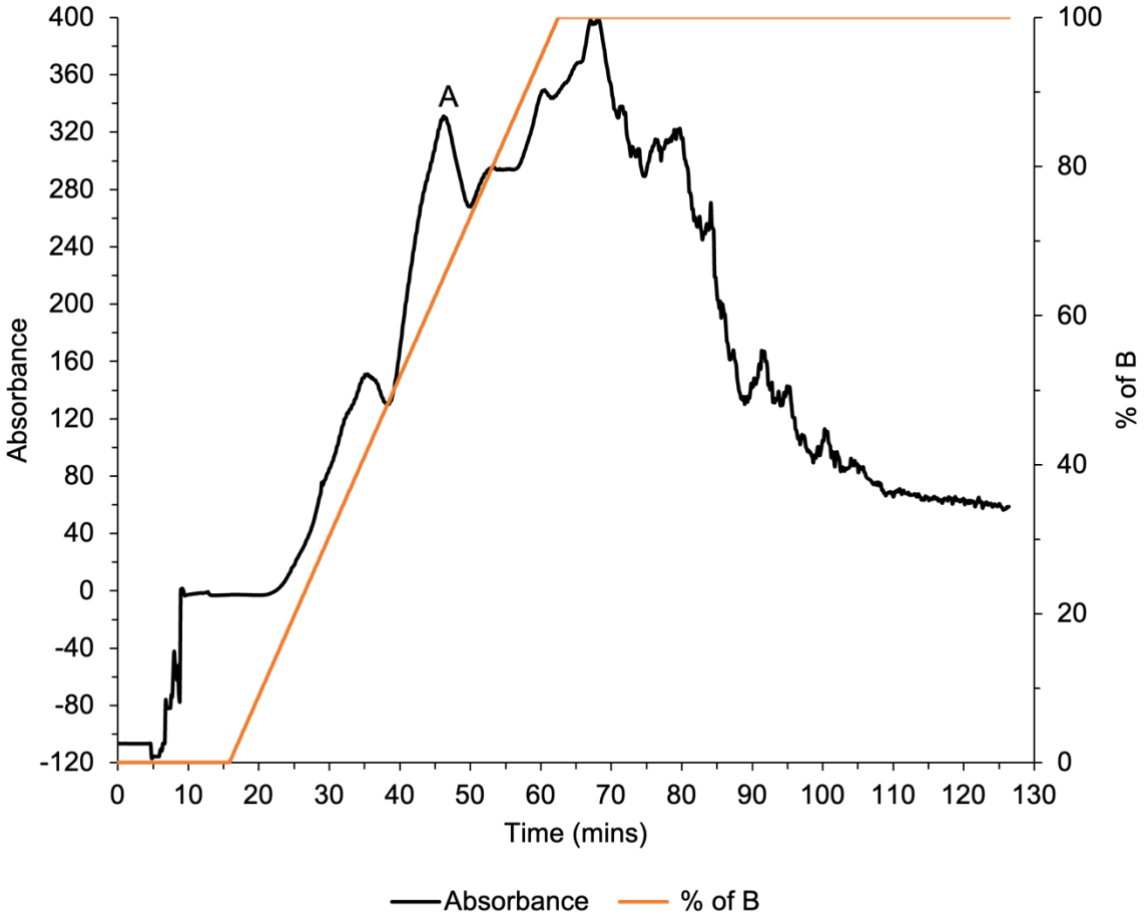


Figure 3.3: Ion exchange absorbance and buffer B salt gradient percentage of the full *Sli*LPMO10G domain using a Q Sepharose column on the AKTA Prime Plus. Note the X-axis is in Time not Volume. The peak area label A between 35 to 65 mins was determined to be the *Sli*LPMO10G protein using an SDS page gel. The black line indicates the absorbance over time and the orange line the percentage of buffer B.

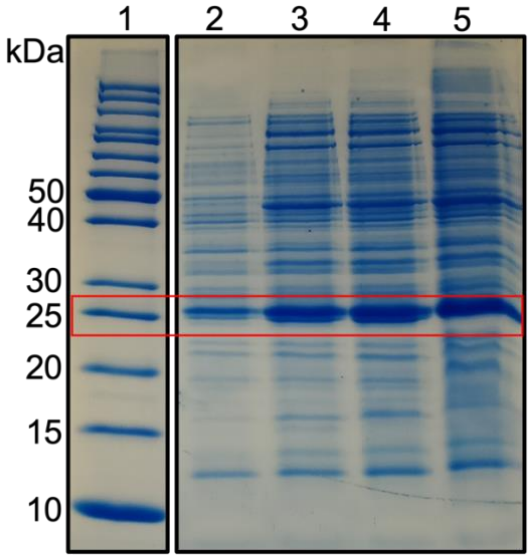


Figure 3.4: SDS page gel from the ion exchange fractions of the full *Sli*LPMO10G; 1. Ladder, 2. Frac6, 3. Frac9, 4. Frac12, 5. Frac14 with red box indicating the protein present at ~25 kDa across the lanes. Gels were stained with Coomassie blue.

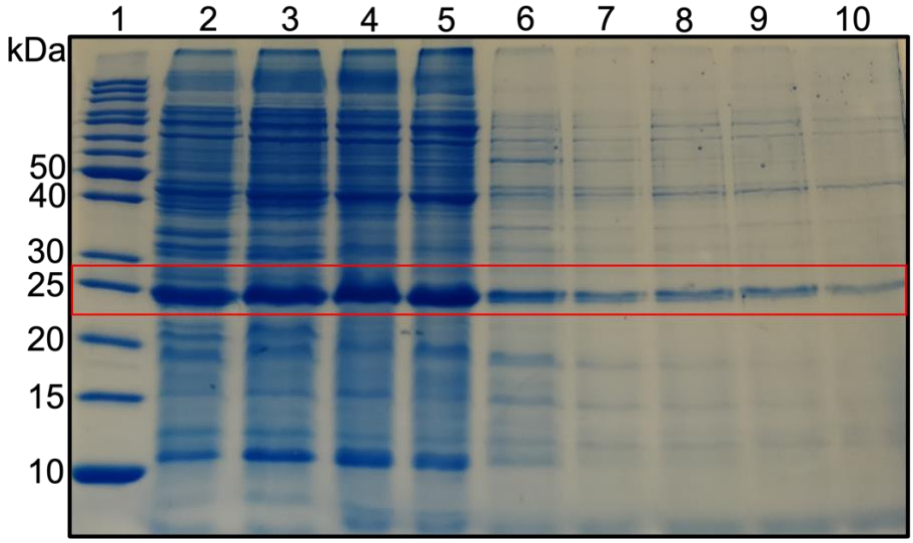


Figure 3.5: SDS page gel from the ion exchange fractions of the full *Sli*LPMO10G; 1. Ladder, 2. Frac17, 3. Frac20, 4. Frac21, 5. Frac22, 6. Frac23, 7. Frac25, 8. Frac27, 9. Frac28, 10. Frac30 with red box indicating the protein present at ~25 kDa across the lanes. Gels were stained with Coomassie blue.

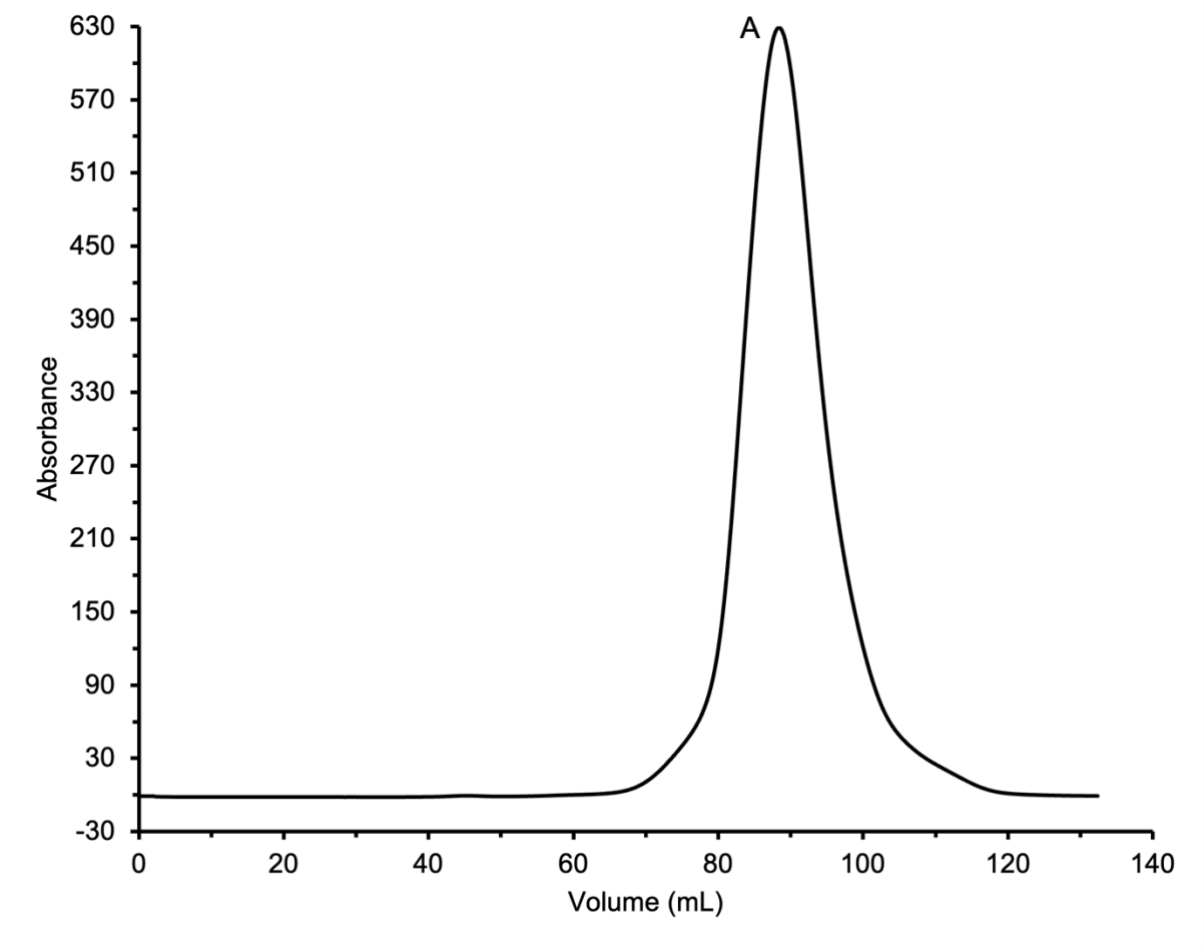


Figure 3.6: Gel filtration absorbance of the full *Sli*LPMO10G using a G75 column on the AKTA Pure. The peak between 70 to 110 mL was determined to be the *Sli*LPMO10G protein using an SDS page gel.

The band of expected mass for the full *Sli*LPMO10G was expected to be at approximately 25 kDa, which can be seen clearly after gel filtration in figure 3.7. The protein was found predominantly in lanes 5 and 6, covering fractions 29 and 32 and less noticeably in lane 4. fraction 25. A slight contamination was shown at approximately 40 kDa in lane 4. Fractions between 29 and 32 were pooled for concentrating, before being frozen.

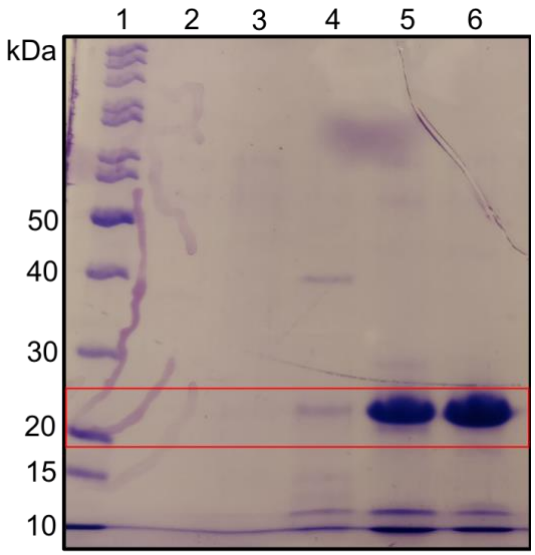


Figure 3.7: SDS page gel from the gel filtration fractions of the full *Sli*LPMO10G; 1. Ladder, 2. Frac17, 3. Frac23, 4. Frac25, 5. Frac29, 6. Frac32 with red box indicating the protein present at ~25 kDa in lanes 5 & 6. Gels were stained with Coomassie blue.

Purification of truncated SliLPMO10G LPMO domain

The truncated *Sli*LPMO10G LPMO domain was over-expressed in BL21-RIL and purified using ion exchange chromatography and gel filtration. The ion exchange absorbance spectra indicated a small peak across 40 to 80 minutes and a large peak across 80 to 112 minutes (figure 3.8). In order to check both peaks, an SDS-PAGE gel was run (figure 3.9). Fractions 1 to 16 related to the small peak and the band of expected mass of *Sli*LPMO10G was seen present on the SDS-PAGE gel within those fractions. The protein began to unbind from the column when the buffer B salt gradient reached 15 %. The fractions between 22 to 25 corresponding to the large peak did not contain the protein and therefore the fractions past 16 were not included when pooled and prepared for gel filtration.

The gel filtration absorbance showed only one peak labelled A (figure 3.10), between 70 to 135 mL. The corresponding fractions between 28 and 46 had samples taken to run on an SDS-PAGE in order to confirm the presence of the truncated *Sli*LPMO10G LPMO domain (figure 3.11).

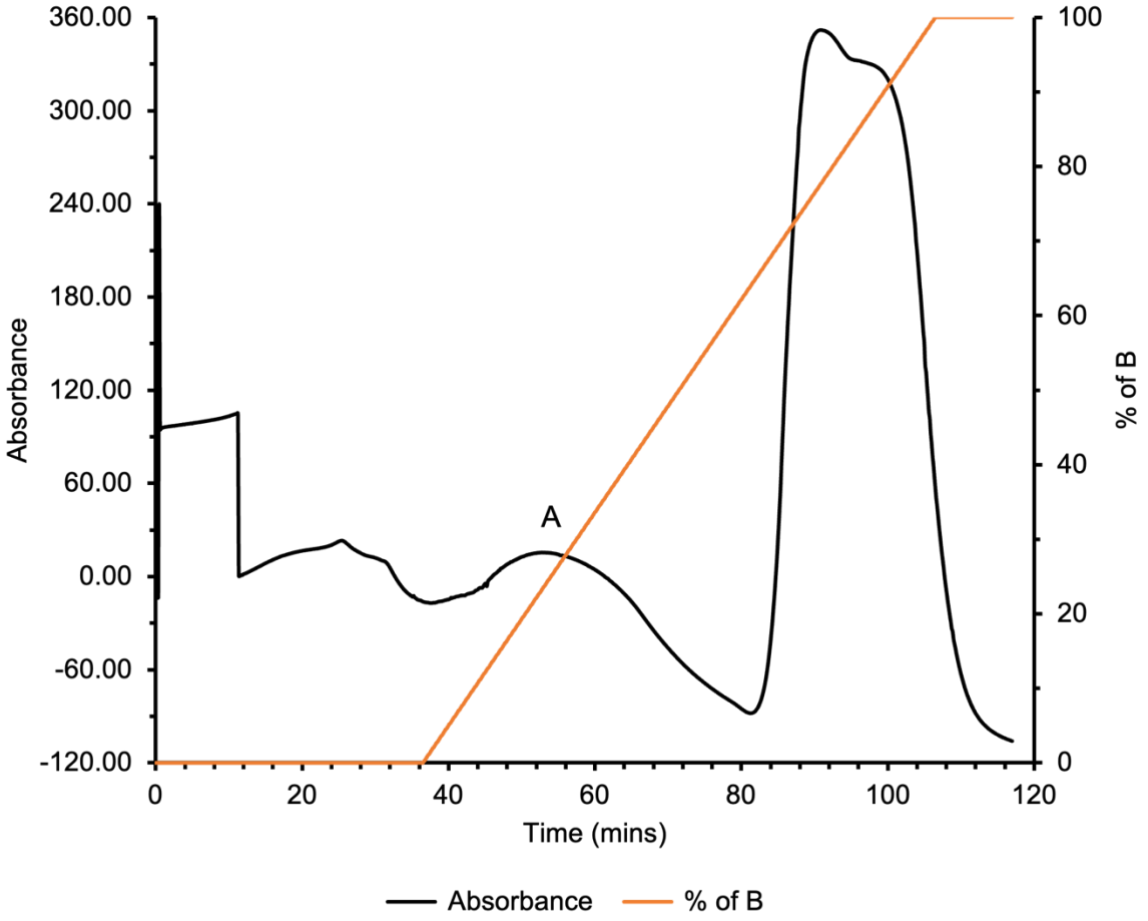


Figure 3.8: Ion exchange absorbance and buffer B salt gradient percentage of the truncated *Sli*LPMO10G domain using a DEAE column on the AKTA Prime Plus. Note the X-axis is in Time not Volume. The peak labeled A between 36 to 76 mins was determined to be the truncated *Sli*LPMO10G protein using an SDS page gel. The black line indicates the absorbance over time and the orange line the percentage of buffer B.

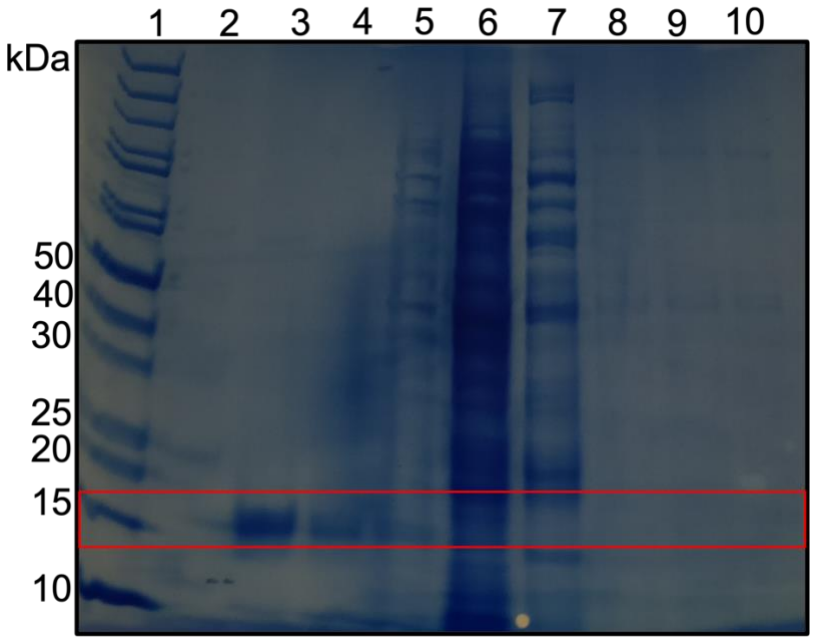


Figure 3.9: SDS page gel from the ion exchange fractions of truncated *Sli*LPMO10G; 1. Ladder, 2. Flow Through, 3. Frac1-16, 4. Frac18, 5. Frac19, 6. Frac22, 7. Frac25, 8. Frac29, 9. Frac30, 10. Frac31 with red box indicating the protein present at ~15 kDa with expression strongest in lane 3. Gels were stained with Coomassie blue.

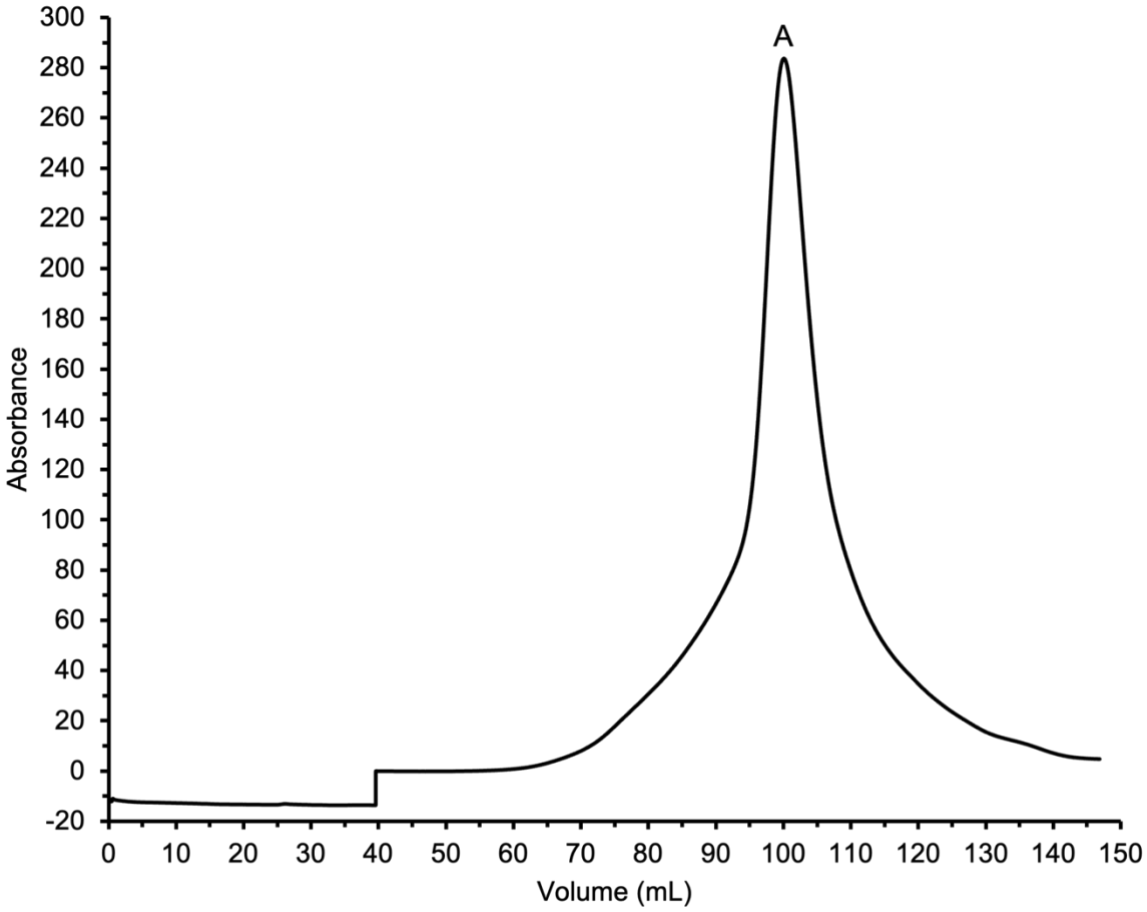


Figure 3.10: Gel filtration absorbance of the truncated *Sli*LPMO10G using a G75 column on the AKTA Pure. The peak between 70 to 135 mL was determined to be the truncated *Sli*LPMO10G protein using an SDS page gel.

For the truncated *Sli*LPMO10G, the band of expected mass was expected to be seen at approximately 15 kDa. As shown in figure 3.11, after gel filtration the protein is seen clearly across lanes 5 to 10, covering fractions 28 to 46, with the strongest expression in the peak of fractions 35 and 38. Fractions between 28 and 46 were pooled for concentrating, before being frozen.

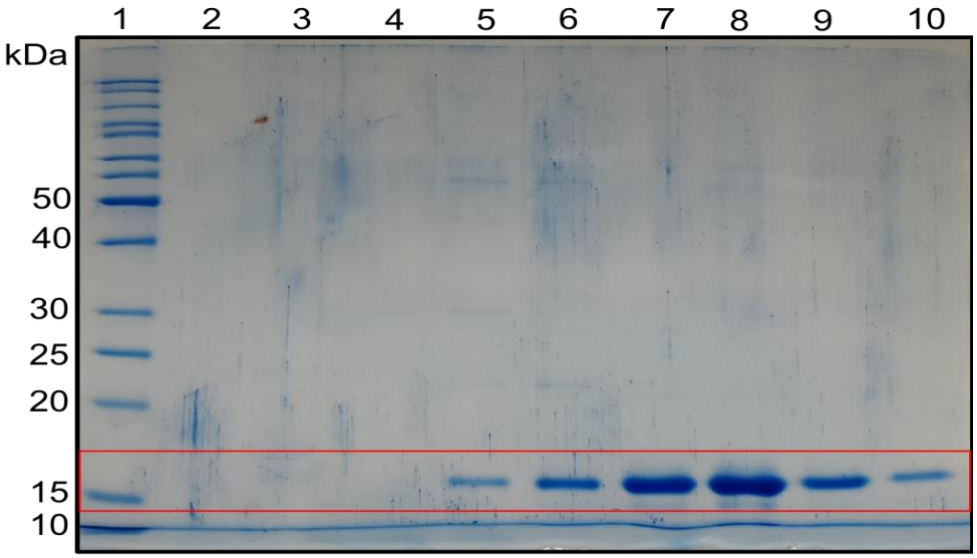


Figure 3.11: SDS page gel from the gel filtration fractions of truncated *Sli*LPMO10G; 1. Ladder, 2. Concentrating Flow Through, 3. Gel Filtration Flow Through, 4. Frac23, 5. Frac28, 6. Frac31, 7. Frac35, 8. Frac38, 9. Frac42, 10. Frac46 with red box indicating the protein present at ~15 kDa in lanes 5-10 with expression strongest in lanes 7 & 8. Gels were stained with Coomassie blue.

Copper Binding Titrations

Using fluorescence spectroscopy, the stoichiometry of copper binding for the full and truncated *Sli*LPMO10G was obtained. Upon the addition of Cu(II) to the respective LPMO samples, the emission fluorescence at 328 nm decreased (figure 3.12A&C), indicating that copper was having a quenching effect. By plotting the copper to protein stoichiometry as a function of the decrease in emission fluorescence (figure 3.12B&D), it is apparent that copper binding is occurring. From these plot it is apparent that a binding stoichiometry of 2 : 1 (Cu : Protein) for the full length *Sli*LPMO10G whereas for the truncated LPMO domain only has a stoichiometry of 1 : 1 (Cu : Protein).

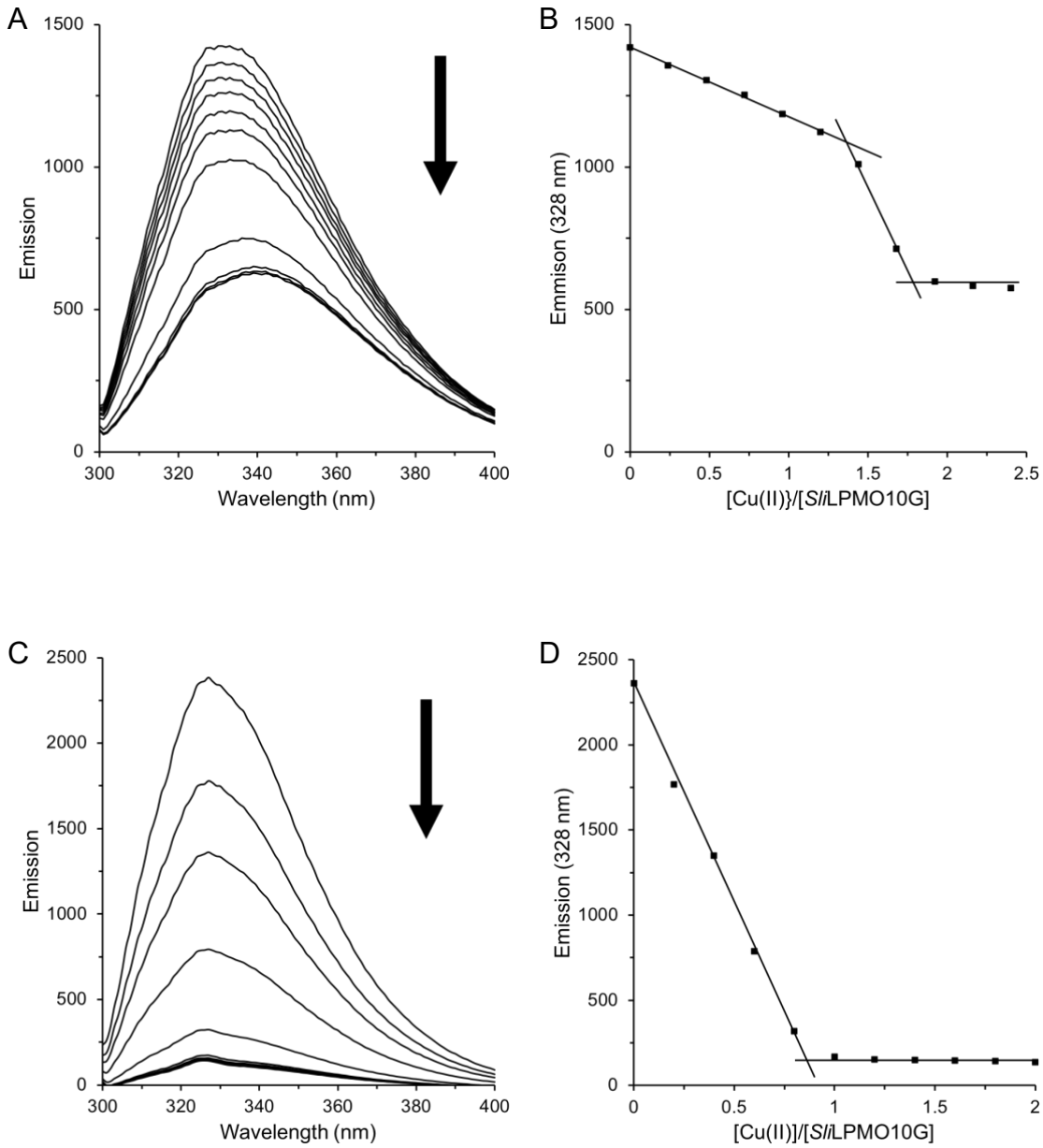


Figure 3.12: Cu(II) binding to the full length *Sli*LPMO10G (A & B) and the truncated LPMO domain (C & D) at pH 6. Changes in emission spectrum with arrow indicating the direction of emission (A & C) upon titration of 1 mM Cu(II)SO_4 to *Sli*LPMO10G (2 µM) and plotted as a function of [Cu(II)]/[*Sli*LPMO10G] (B & D), the stoichiometry of the reaction is indicated by the intersection of the solid lines.

3.3.2 – Structure determination of the truncated SliLPMO10G LPMO domain

Model building and refinement cycling

Initial inspection of the reflection data showed well-defined electron density across many residues in both subunit chains A and B, following molecular replacement using MR.BUMP (Keegan and Winn, 2007). The initial refinement using Refmac5 gave an R_{factor} valuing 0.27 and an R_{free} valuing 0.29. COOT showed multiple areas of broken and incorrect structure, such as the region at the start of the chain between residues 18 to 21. Additional amino acid residues had been modelled into the density giving two unknown (UKN) residues at 19 and 20 in both chains (figure 3.15A). This was also shown between residues 111 and 115 with residue 113 missing from the region as well as the phenylalanine sidechain (figure 3.14A). Partially modelled amino acids were also clear in both subunits, with residue 83 showing a clear tryptophan which needed its conformation changed (figure 3.13). Inaccuracies were also discovered in the modelling of water molecules and meant they had to be checked during rebuilding. Water molecules were found in density where an amino acid residue should be or where no density was present. At residue 12 and 13, water molecules were initially modelled into the density where the amino acid sidechains would be present (figure 3.16). High green density spheres were apparent where water molecules were needed to be modelled into the density and were re-checked at the end of residue building. Cu was finally modelled into the binding sites of both chains and at a point between the two subunits which was identified as an intermediate.

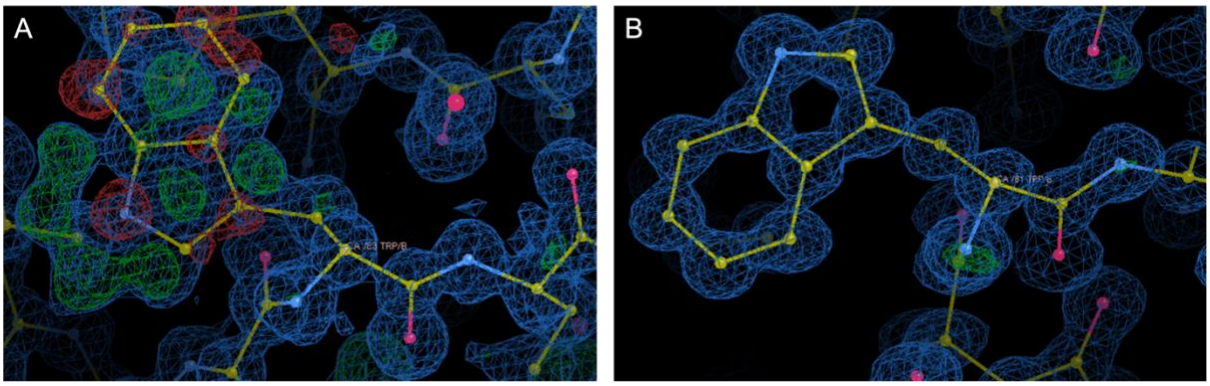


Figure 3.13: $2F_o$-F_c electron density maps contoured at 1 σ. A) The red and green density illustrates where the experimental density and atomic model disagree. B) The remodelled section of tryptophan 83, now tryptophan 81, after the configuration change and modelling fit to the electron density data.

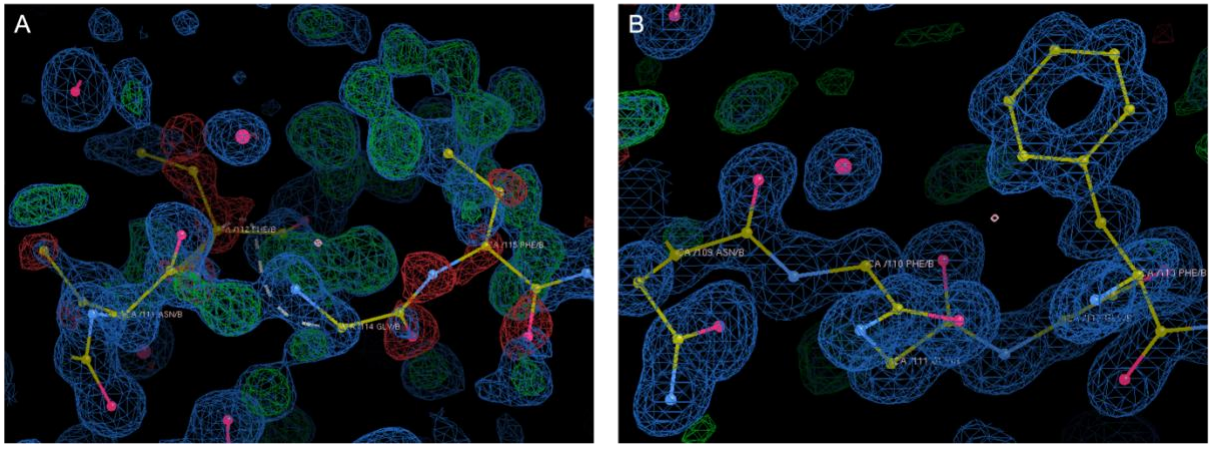


Figure 3.14: $2F_o$-F_c electron density maps contoured at 1 σ. A) The red and green density illustrates where the experimental density and atomic model disagree. B) The remodelled phenylalanine 115, now phenylalanine 113, as well as the rebuilt section of residues 111-115, now residues 109-113, after remodelling the chain.

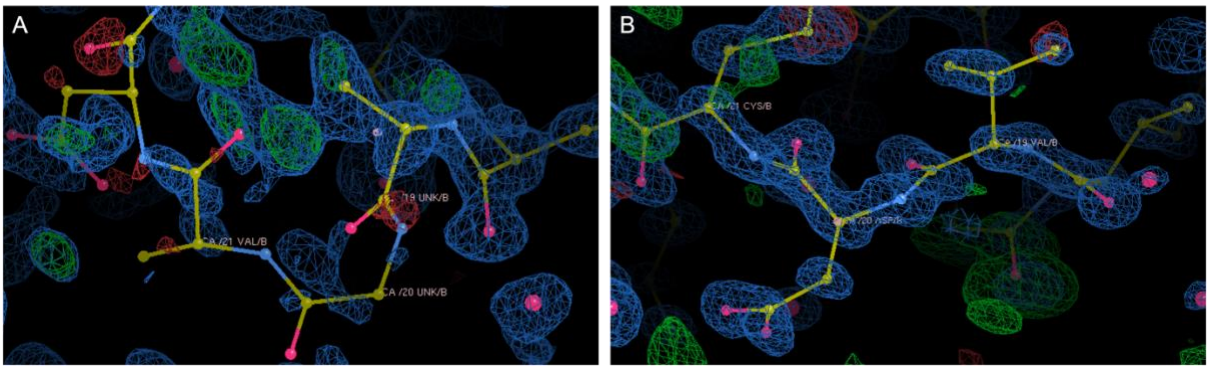


Figure 3.15: $2F_o$-F_c electron density maps contoured at 1 σ. A) The red and green density illustrates where the experimental density and atomic model disagree. B) The rebuilt region of valine 19 to cysteine 21. The unknown amino acid residues were deleted, and the residues remodelled into the electron density.

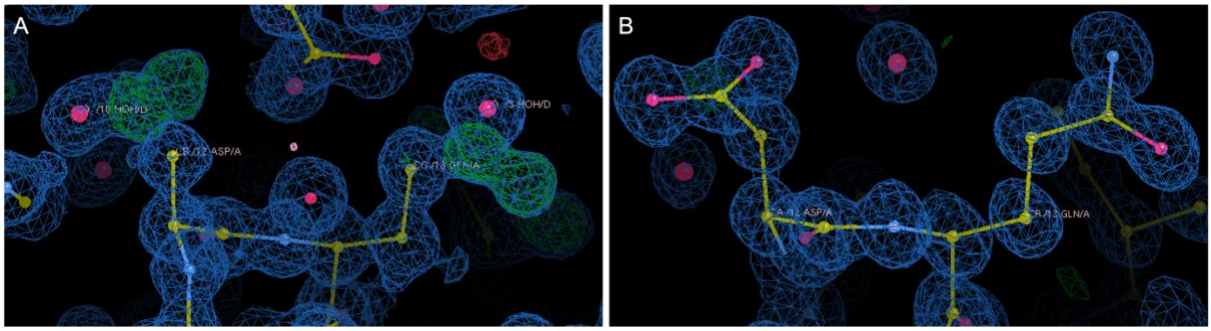


Figure 3.16: $2F_o$-F_c electron density maps contoured at 1 σ. A) The red and green density illustrates where the experimental density and atomic model disagree. Water molecules can be seen in the electron density where the amino acid side chains should be present. B) The rebuilt region of aspartic acid 12 and glutamine 13 with the side chains in place and the water molecules deleted.

Structural validation of SliLPMO10G

The refinement statistics were summarised into a table (table 3.5). The final R_{factor} and R_{free} were 0.124 and 0.143 respectively. Relative Mean Standard Deviation (RMS dev.) bond lengths were 0.0127 Å and RMS dev. bong angles were 1.781° with a Ramachandran favoured percentage of 99.27%.

Table 3.5: Crystallographic and refinement statistics for *Sli*LPMO10G.

Space group	P 1 21 1
Unit cell (Å)	60.8, 30.7, 70.1, 90.0, 106.5, 90.0
Resolution (Å)	38.89 – 1.04
R_{factor}	0.124
R_{free}	0.143
RMS dev. bond lengths (Å)	0.0127
RMS dev. bond angles (°)	1.781
Ramachandran favoured (%)	99.27

The Ramachandran plot (figure 3.17) indicated 2 amino acids in the allowed region constituting 0.73% of the 274 amino acids analysed across the asymmetric unit and 0 outliers.

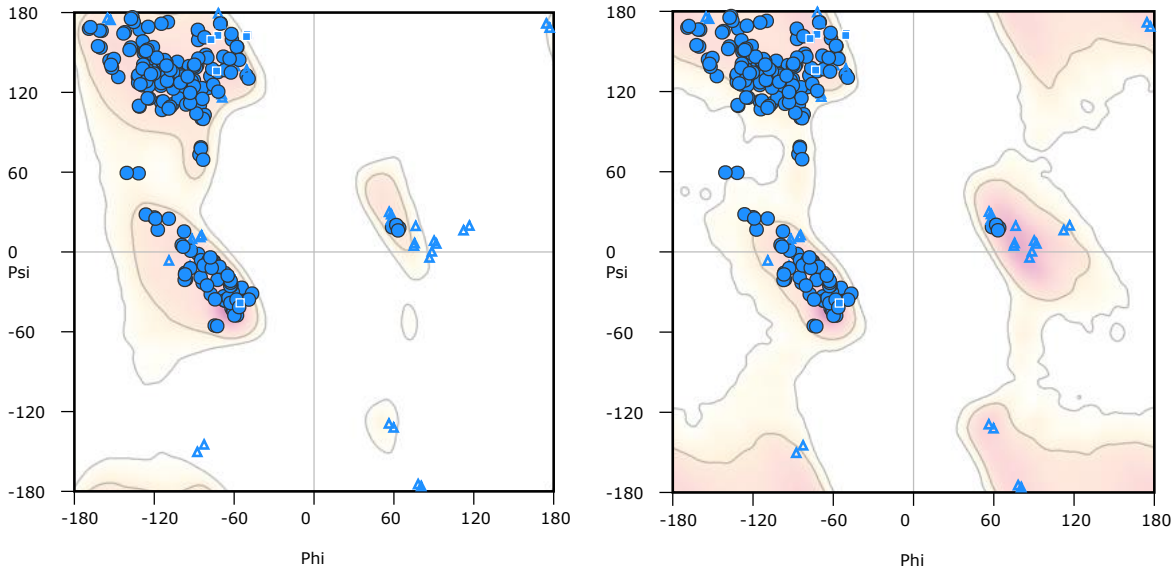


Figure 3.17: Ramachandran plot. Red regions indicate preferred conformations and yellow regions correspond to acceptable conformations. Triangles are GLY, circles are not-GLY-or-PRO and those with grey outlines are PRO.

Description of the asymmetric unit

Two identical molecules are visible in the crystallographic asymmetric unit (figure 3.18) which has a space group of P 1 21 1. The recognised substrate binding surface of *Sli*LPMO10G is flat with the Cu site between the N-terminal His residue and His 76. The electron density indicated that Cu was bound at the binding site and present in the interface between the two asymmetric units.

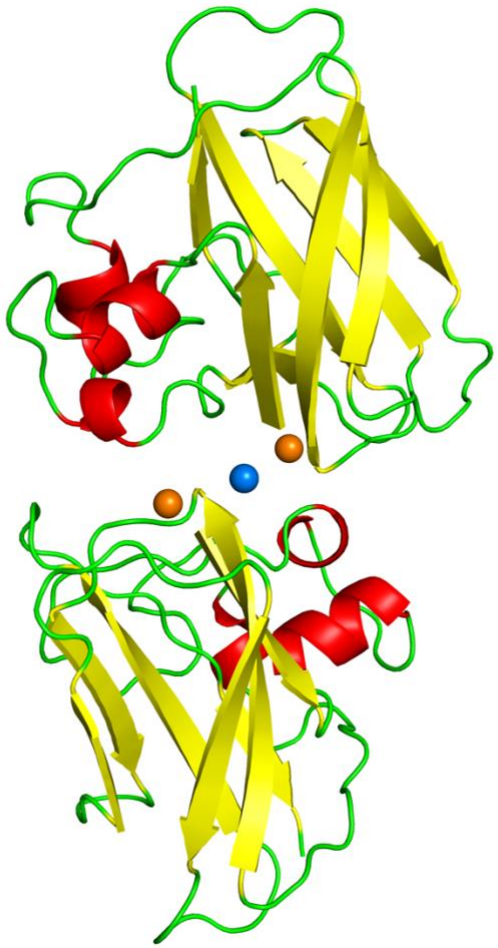


Figure 3.18: Cartoon representation of the *Sli*LPMO10G asymmetric unit. The core β-sheet fold shown in yellow, the Loop 2 (α-helical) region shown in red, the disulphide bonds shown in magenta and loops connecting the secondary structure in green. The active site Cu is represented as orange spheres and the interface Cu is shown in blue.

Description of the overall fold

The overall structure (figure 3.19) comprises 7 anti-parallel β-strands arranged in two β-sheets; one strand containing β1, β2 and β5, the other strand containing β4, β3, β6 and β7. Between strands one and two is the Loop 2 region, which consists of three

small α-helices, α1, α2 and α3. Two disulphide bonds are present at Cys14/Cys21 and Cys40/Cys135, with the latter connecting the Loop 2 region to the C-terminal β-strand (β7). The surface representation (figure 3.20) shows that Cu bound in the Cu active site protrudes slightly from the flat substrate binding site.

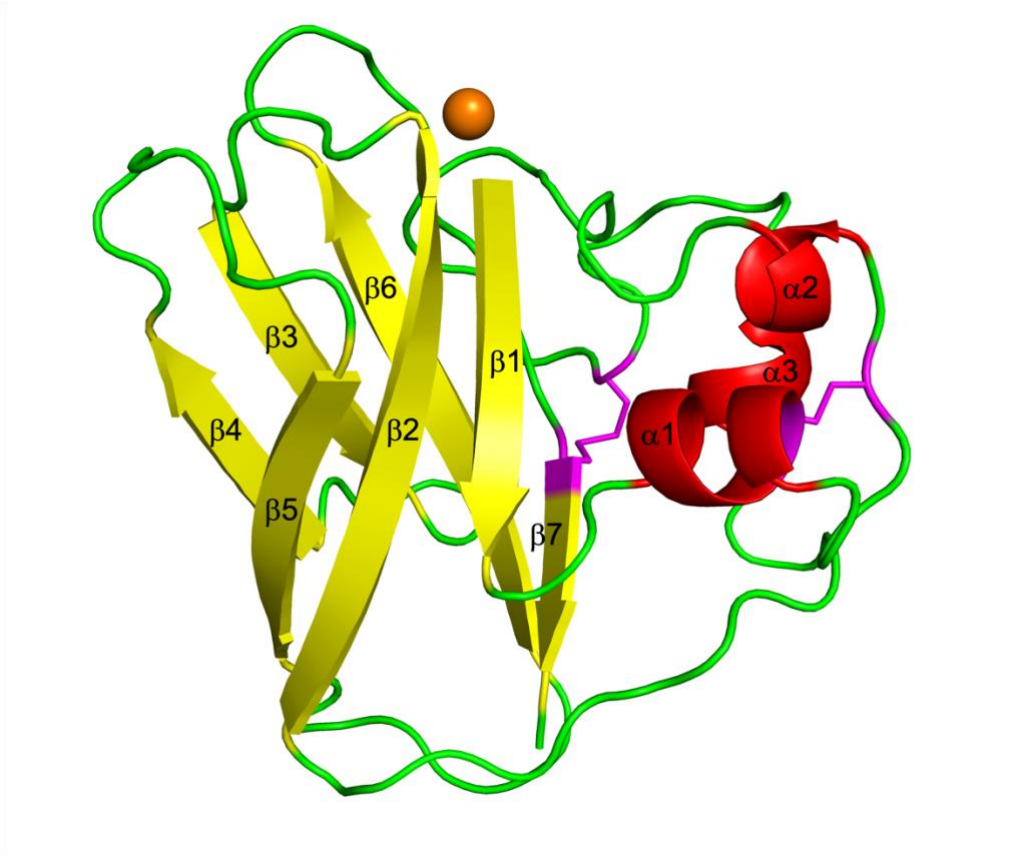


Figure 3.19: Cartoon representation of the *Slii*LPMO10G monomer. The core β-sheet fold shown in yellow, the Loop 2 (α-helical) region shown in red, the disulphide bonds shown in magenta and loops connecting the secondary structure in green. The active site Cu is represented as an orange sphere.

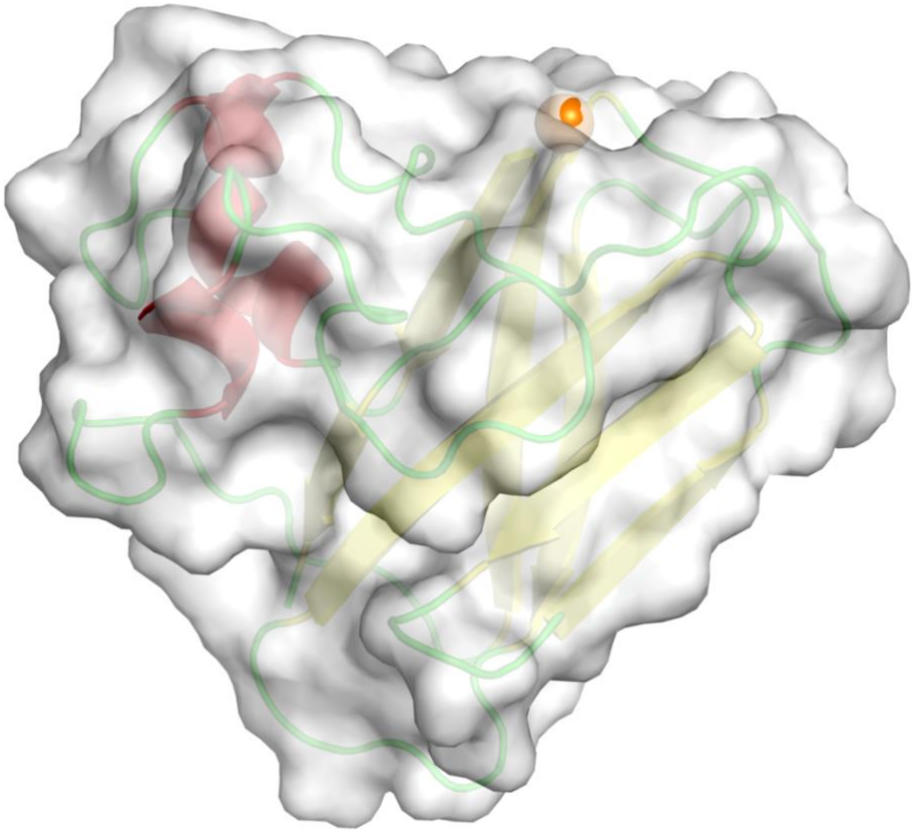


Figure 3.20: Surface representation of the *Sli*LPMO10G monomer. Bound Cu is visible at the substrate binding site as an orange sphere in the Cu binding site.

Description of the Cu site

A Cu ion was modelled with coordination at the Cu active site by three nitrogen ligands (figure 3.21A), the Nt of His 1 and the Nδ1 of His 1 and Nε2 of His 76 of the mature sequence. These are His 30 and His 106 in the pre-mature sequence as the signal peptide is cleaved between residue 29 and 30. This formation results in the typical T-shaped His-brace coordination geometry with bond lengths to the Cu ion of 1.95 Å, 2.16 Å and 1.89 Å (figure 3.21B).

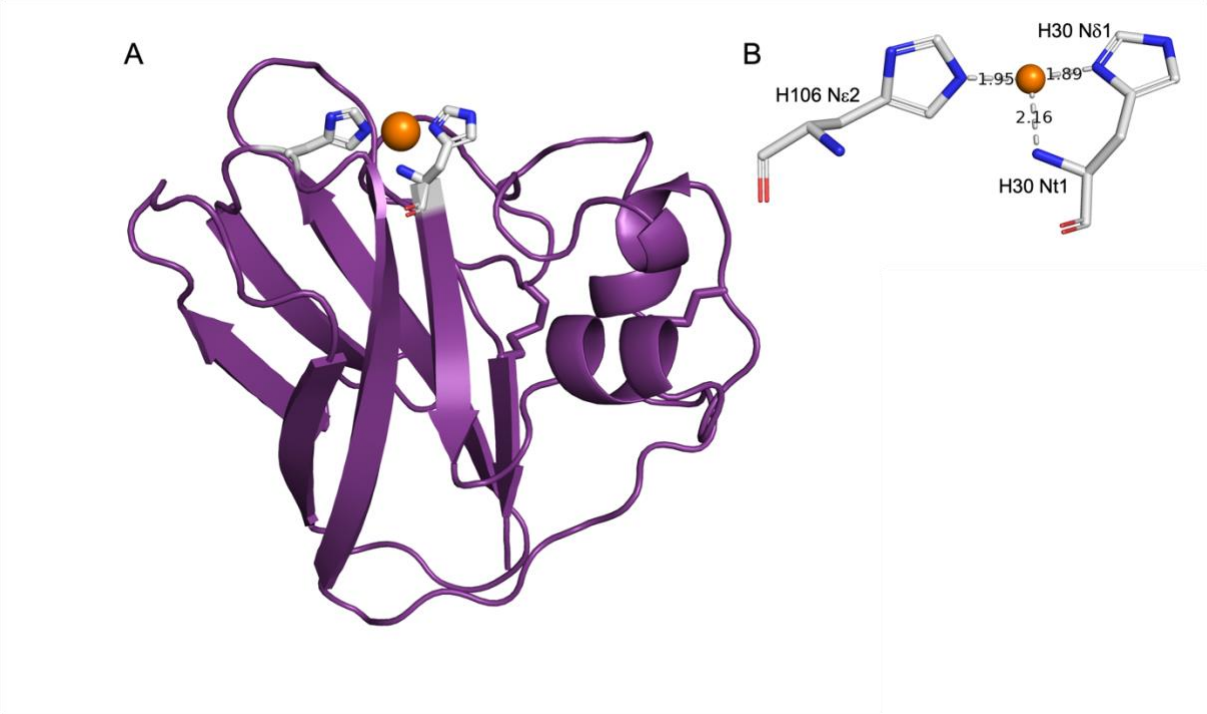


Figure 3.21: A) Cartoon representation of the *Sli*LPMO10G monomer in purple with a stick representation of the Cu active site coloured silver. Cu is represented as an orange sphere. B) Stick representation of the Cu active site. Coordinating His residues H30 Nt, H30 Nδ1 and H106 Nε2, were coloured silver with bond lengths to the Cu ion indicated by dashed lines.

Description of the Cu in the interface between the LPMO domains

Whilst the molecule has been modelled as copper based on the density, it is possible that the interface molecule could be something else such as magnesium. The occupancy of the intermediate copper was changed to 0.55 in order to have a positive density, as an occupancy of 1.00 showed a negative density. The distances between the interface and the surrounding amino acids were measured to identify the binding interactions of the Cu (figure 3.22). The interface Cu had bond lengths measured to at 3.08 Å and 3.10 Å to threonine residue 73 of subunits A and B in the

asymmetric unit. The distance to glutamate residue 29 of subunits A and B were also measured with bond lengths 3.39 Å and 3.44 Å respectively.

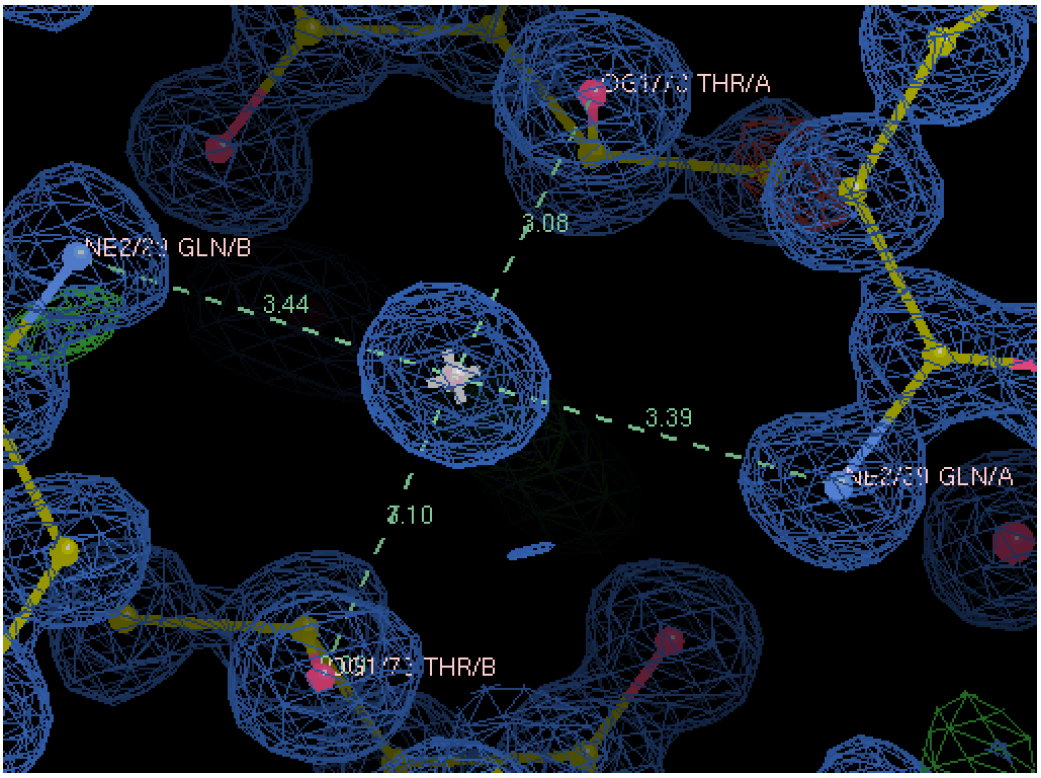


Figure 3.22: A $2F_o$-F_c electron density maps contoured to 1 σ. The Cu modelled into the interface region between the asymmetric A and B units. Bond lengths are shown between the interface Cu and the threonine residue 73 of both subunits A and B as well as the glutamate residue 29 of both subunits A and B.

3.3.3 – Characterisation of CBM5 domain

Purification

The truncated *Sli*LPMO10G CBM5 was over-expressed in BL21-DE3 and purified using nickel immobilized metal affinity chromatography (IMAC) and gel filtration. The absorbance of the Ni column (figure 3.23) shows the absorbance over time in relation

to the percentage gradient of buffer B. The peak between 19 and 39 minutes, labelled A, indicates the CBM protein peak. The related fractions, 6 to 27, were run on an SDS-PAGE to confirm. The protein began to unbind from the column when the buffer B gradient reached 18 %.

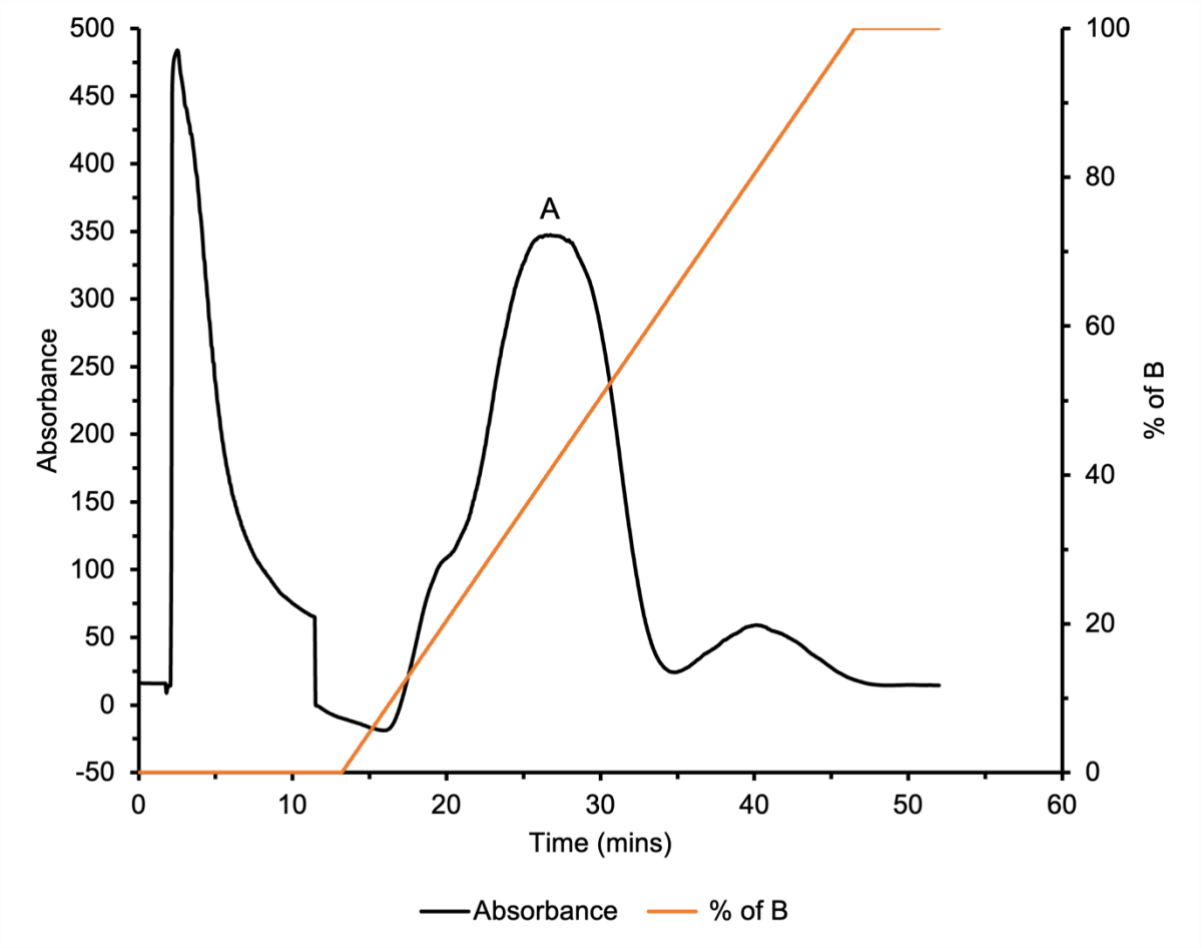


Figure 3.23: Immobilized metal affinity chromatography (IMAC) absorbance and buffer B imidazole percentage of the *Sli*LPMO10G CBM domain using a Nickel column on the AKTA Prime Plus. Note the X-axis is in Time not Volume. The black line indicates the absorbance over time and the orange line the percentage of buffer B.

The SDS-PAGE gel of the nickel column for the *Sli*LPMO10G CBM5 domain (figure 3.24) showed bands of expected mass to be slightly tilted across the gel. However, the protein was expected to be approximately 10 kDa and a predominant band was shown across lanes 6 to 10, with the bands in 7, 8 and 9 having the most expression. Fractions 6 to 27 from the IMAC were collected and concentrated for gel filtration.

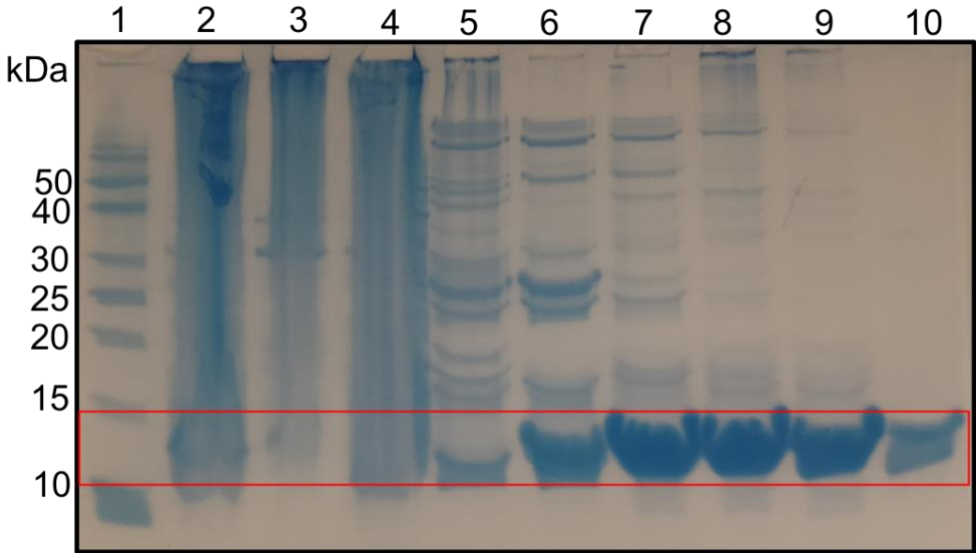


Figure 3.24: SDS page gel of the *Sli*LPMO10G CBM domain from the Ni column fractions collected; 1. Ladder, 2. Final supernatant after lysis, 3. Final cell pellet after lysis, 4. Concentrator flow through, 5. Frac.6, 6. Frac9, 7. Frac12, 8. Frac14, 9. Frac17, 10. Frac27, with the red box indicating the protein present at ~10 kDa in lanes 7 to 9. Gels were stained with Coomassie blue.

The gel filtration absorbance spectra showed 3 peaks (figure 3.25). The peak labelled A between 130 and 170 minutes was determined to be the CBM protein due to the proteins small size and it being the last peak eluted. The fractions were pooled and concentrated to 3 mL before a sample was prepared for running on an SDS-PAGE gel. Using an SDS-PAGE gel (figure 3.26) the fractions of the peaks were also tested, fraction 16 at 30 minutes and fractions 22 and 23 between 71 and 74 minutes. Fractions between 29 and 32 were pooled for concentrating, before being frozen.

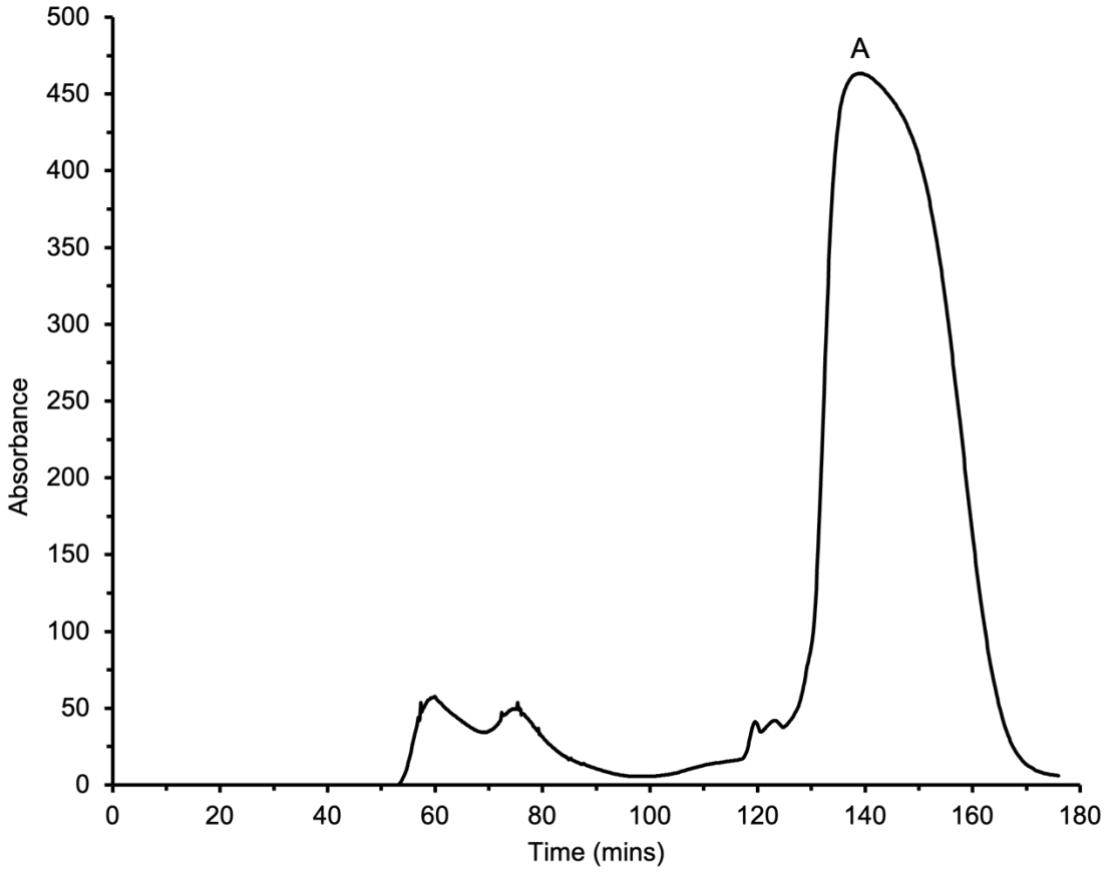


Figure 3.25: Gel filtration absorbance of the *Sli*LPMO10G CBM domain using a G75 column on the AKTA Prime Plus. Note the X-axis is in Time not Volume. The peak labelled A, between 130 to 170 minutes was determined to be the CBM protein.

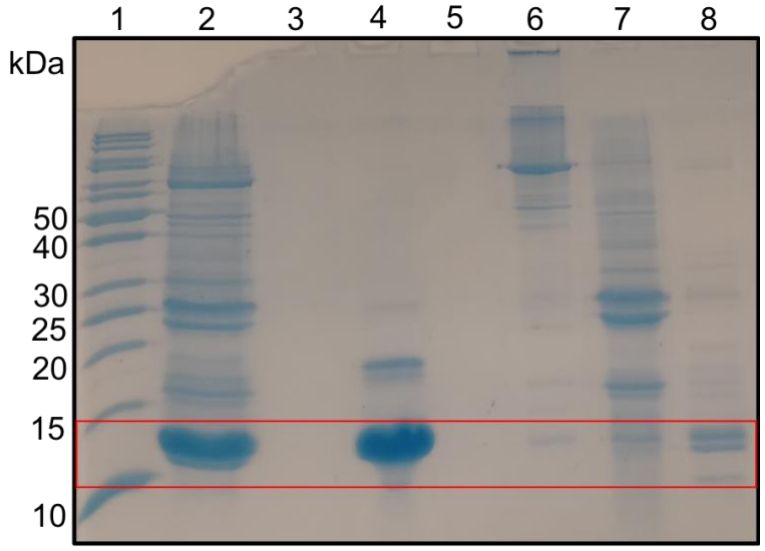


Figure 3.26: SDS page gels of the *Sli*LPMO10G CBM domain from the G75 gel filtration fractions; 1. Ladder, 2. After Ni column concentrated CBM, 3. Ni column concentrator flow through, 4. Concentrated CBM after G75, 5. G75 concentrator flow through, 6. G75 Frac16, 7. G75 Frac22/23, 8. G75 other pooled peaks, with the red box indicating the protein present at ~10 kDa in lanes 2 & 4. Gels were stained with Coomassie blue.

The UV absorbance spectrum of the purified and concentrated CBM domain was tested (figure 3.27) and the peak at 280 nm indicates the CBM protein. The peak labelled b at 290 nm is indicative of tryptophan absorbance, and the CBM domain of *Sli*LPMO10G contains 6 Trp. Using the extinction coefficient of 34,490 M^{-1} cm^{-1} at 280 nm the concentration of the CBM was determined to be 15 μg mL^{-1}.

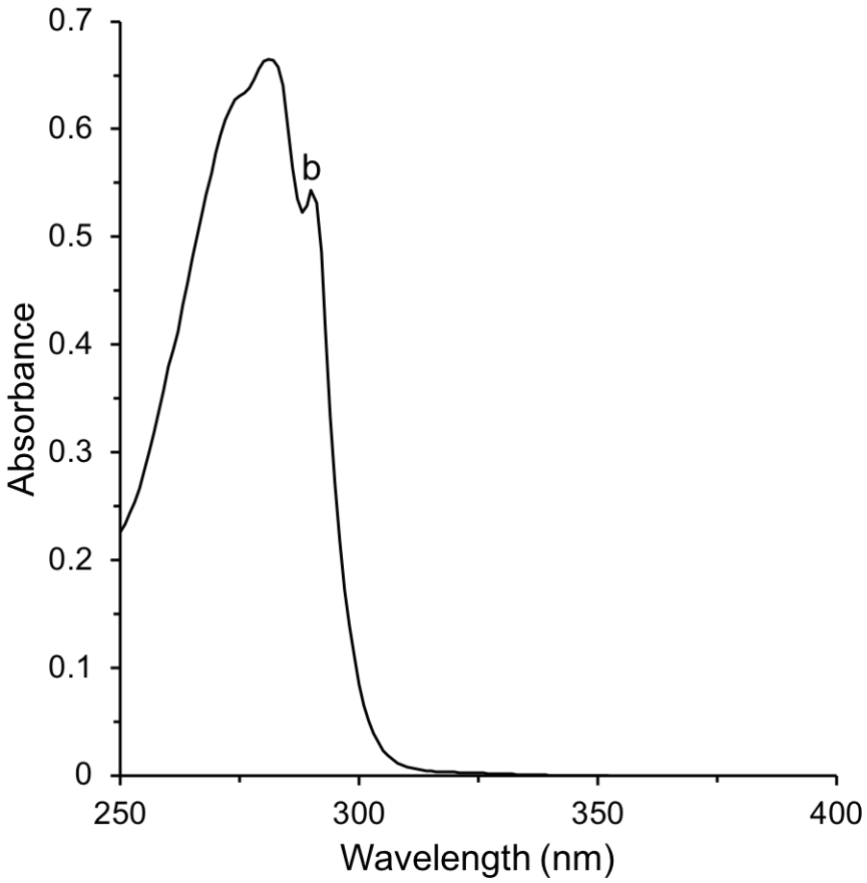


Figure 3.27: UV absorbance spectrum of the purified CBM domain recorded using a Carry 60 spectrophotometer. The distinct shoulder (b) at 290 nm is indicative of tryptophan absorbance. The CBM contains 6 Trp.

Crystallisation of the CBM domain

Crystal trials plates were periodically checked under a microscope. A crystal was seen in well G6 of the JCSG – plus screen under conditions 20 % w/v PEG 3350, 0.2 M sodium malonate pH 7.0 (figure 3.28). Optimisations were unable to be carried out due to time constraints and further testing to determine if the crystal was protein or salt were also unable to be investigated.

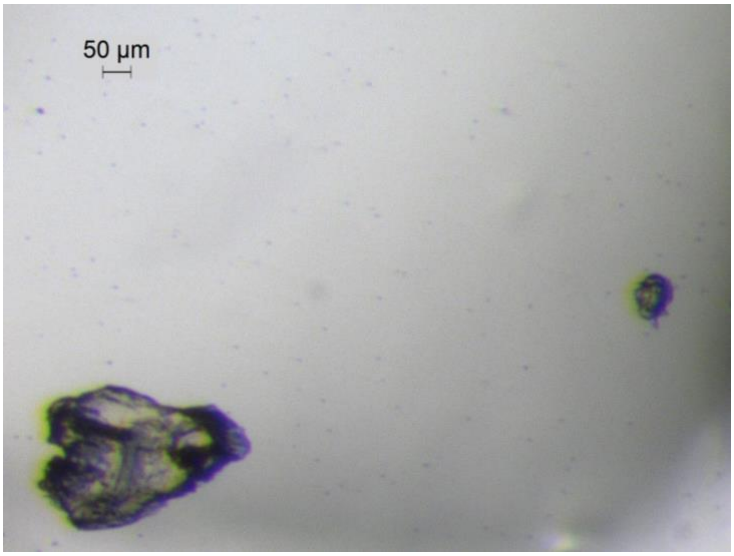


Figure 3.28: CBM crystal from JCSG – plus in well G6 with condition 20 % w/v PEG 3350, 0.2 M sodium malonate pH 7.0.

[3.4] Discussion

3.4.1 – Structural comparison

The structure of *Sli*LPMO10G contains seven β-strands with a Loop 2 region consisting of three α-helices and a β-sandwich fold consistent with other AA10 structures. The *Sli*LPMO10G LPMO domain was compared with *Sli*LPMO10E, one of the seven *S. lividans* enzymes which is also chitin specific and has previously been characterised (Chaplin *et al.*, 2016). *Sli*LPMO10E contains seven β-strands with a Loop 2 region consisting of two α-helices and a small β-strand (Chaplin *et al.*, 2016). The x-ray crystal structure of *Sli*LPMO10G (figure 3.21B) revealed the Cu active site to include the characteristic His-brace coordination associated with all other LPMOs (Vaaje-Kolstad *et al.*, 2010). The cysteines highlighted in figure 3.19 indicate the disulphide bridges relate to structural stability, linking the Loop 2 region to the C-terminal β-strand.

3.4.2 – Cu active sites of the full SliLPMO10G and truncated LPMO domain

It was predicted before the Cu binding experiments that both the full *Sli*LPMO10G and truncated LPMO domain would show a stoichiometry ratio of 1 : 1 (Cu : Protein) indicating one Cu active site, due to the active site being observed on the flat substrate binding surface of the LPMO domain. However, whilst the LPMO domain of *Sli*LPMO10G did indeed have a 1 : 1 ratio of Cu to protein, the ratio of the full *Sli*LPMO10G was 2 : 1 when observed. This suggested the possibility that the full *Sli*LPMO10G had two Cu binding sites and therefore needed two coppers for substrate binding.

3.4.3 – Limitations and future projects

Time constraints of the project meant that Cu titration experiments to determine the stoichiometry of the *Sli*LPMO10G CBM domain were unable to be carried out. As the results of the stoichiometric ratio was against predictions, it would be interesting to determine if the two Cu present in the activation of the full *Sli*LPMO10G was related to the presence of the CBM domain. Further crystallisation trials for the CBM domain were also unable to be continued. Crystallisation trials of the full *Sli*LPMO10G had been attempted previously by Megan Straw, PhD (2017/2018) before the start of the project but had been unsuccessful. It was theorised that the linker may have been an issue in the process of crystal formation (Boraston *et al.*, 2001; Smyth *et al.*, 2003). Design of substrate binding assays had also started to compare the substrate binding activity of the full *Sli*LPMO10G, LPMO domain and CBM domain. As the CBM domain was specifically from the CBM5 family, it has specificity to chitin and cellulose and deletion of CBM5 has previously shown considerable reduction in chitin binding (Forsberg *et al.*, 2016; Manjeet *et al.*, 2019). Future projects would include the completion of the crystallisation trials for the CBM5 domain of *Sli*LPMO10G and continuation of the Cu titration experiments. Substrate binding assays would also be carried out for the full *Sli*LPMO10G, LPMO domain and CBM domain. With continued research into the relationship between LPMOs and CBM domains, along with their influence in cellulolytic and chitinolytic biomass processing, there are many opportunities to further explore the topics in this study.

Chapter 4: Conclusion

Lytic polysaccharide monooxygenases have continued to become more prevalent in the biofuel production process as research continues to examine and learn about their characteristics. With their ability to oxidatively break down polysaccharide chains, LPMOs have the potential to be a more efficient and cost-effective option than current processing method used in biofuel production. As insight into their mechanism increases; their cooperative relationship with carbohydrate-binding modules is studied and the use of ancestral reconstruction as a tool becomes more widespread, LPMO families will ensure there are many new avenues to investigate moving forward.

The ancestral sequence reconstruction of LPMOs is a tool which would help explore the nature of possible LPMO ancestors and help inform us of other potential uses for LPMOs outside of biofuel production. Determining LPMO functions and structural links to substrate binding through ancestral reconstruction, will be hugely beneficial in designing LPMOs for usage in industry. It is also likely to provide insights into thermostability and substrate specificity. The use of ASR alongside other protein engineering tools could potentially optimise LPMO usage for varying biofuel production batches, regardless of first- or second- generation biomass feedstock.

By further deciphering the relationship between the *Sli*LPMO10G's LPMO domain and CBM domain along with their respective effects on substrate binding, it would help to broaden our understanding of carbohydrate-binding module interaction with the LPMO domain and the relation to substrate specificity. If it is possible to further

increase or decrease substrate specificity, LPMOs could be more accurately adapted for biofuel batch processing needs. Tying this adaption to understanding of the LPMO structure and binding through their Loop regions, could make them an ever-versatile component in biofuel production.

In conclusion, whilst research is primarily focused on LPMO use in biofuel production processes, their potential use could be much wider. The combination of using ancestral sequence reconstruction and carbohydrate-binding modules as tools to further understand and adapt LPMOs, will open many new avenues of research. As we further explore the characteristics and functions of LPMOs, we may be able to discover alternative possibilities for them, consolidating their diverseness as proteins that are worth investigating.

References

Aachmann, F. L., Sørlie, M., Skjåk-Bræk, G., Eijsink, V. G. H. and Vaaje-Kolstad, G., (2012) 'NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions', Proceedings of the National Academy of Sciences, 109(46), pp. 18779–18784. doi: 10.1073/pnas.1208822109.

Agger, J. W., Isaksen, T., Várnai, A., Vidal-Melgosa, S., Willats, W. G. T., Ludwig, R., Horn, S. J., Eijsink, V. G. H., Westereng, B., (2014) 'Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation', Proceedings of the National Academy of Sciences of the United States of America, 111(17), pp. 6287–6292. doi: 10.1073/pnas.1323629111.

Barbesgaard, P., Heldt-Hansen, H. P. and Diderichsen, B., (1992) 'On the safety of *Aspergillus oryzae*: a review', Applied Microbiology and Biotechnology, 36(5), pp. 569–572. doi: 10.1007/BF00183230.

Binod, P., Gnansounou, E., Sindhu, R. and Pandey, A., (2019) 'Enzymes for second generation biofuels: Recent developments and future perspectives', Bioresource Technology Reports, 5, pp. 317-325. doi: 10.1016/j.biteb.2018.06.005.

Bissaro, B., Røhr, Å. K., Müller, G., Chylenski, P., Skaugen, M., Forsberg, Z., Horn, S. J., Vaaje-Kolstad, G. and Eijsink, V. G. H., (2017) ‘Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2’, Nature Chemical Biology, 13(10), pp. 1123–1128. Doi: 10.1038/nchembio.2470.

Boraston, A. B., McLean, B. W., Guarna, M. M., Amandaron-Akow, E. and Kilburn, D. G., (2001) ‘A family 2a carbohydrate-binding module suitable as an affinity tag for proteins produced in Pichia pastoris’, Protein Expression and Purification, 21(3), pp. 417–423. doi: 10.1006/prep.2001.1393.

Caldararu, O., Oksanen, E., Ryde, U. and Hedegård, E. D., (2019) ‘Mechanism of hydrogen peroxide formation by lytic polysaccharide monooxygenase’, Chemical Science, 10(2), pp. 576–586. doi: 10.1039/c8sc03980a.

CAZy (1998) Auxiliary Activities. Available at: http://www.cazy.org/Auxiliary-Activities.html (Accessed: 29 October 2018).

Chang, H., Gacias Amengual, N., Botz, A., Schwaiger, L., Kracher, D., Scheiblbrandner, S., Csarman, F. and Ludwig, R., (2022) ‘Investigating lytic polysaccharide monooxygenase-assisted wood cell wall degradation with microsensors’, Nature Communications, 13(1), pp. 6258. doi: 10.1038/s41467-022-33963-w.

Chaplin, A. K., Wilson, M. T., Hough, M. A., Svistunenko, D. A., Hemsworth, G. R., Walton, P. H., Vijgenboom, E. and Worrall, J. A., (2016) 'Heterogeneity in the histidine-brace copper coordination sphere in auxiliary activity family 10 (AA10) lytic polysaccharide monooxygenases', Journal of Biological Chemistry, 291(24), pp. 12838–12850. doi: 10.1074/jbc.M116.722447.

Christensen, T., Woeldike, H., Boel, E., Mortensen, S. B., Hjortshoej, K., Thim, L., Hansen, M. T., (1988) 'High level expression of recombinant genes in *Aspergillus oryzae*', Bio/Technology, 6(12), pp. 1419–1422. doi: 10.1038/nbt1288-1419.

Correa, D. F., Beyer, H. L., Fargione, J. E., Hill, J. D., Possingham, H. P., Thomas-Hall, S. R. and Schenk, P. M., (2019) 'Towards the implementation of sustainable biofuel production systems', Renewable and Sustainable Energy Reviews, 107(February), pp. 250–263. doi: 10.1016/j.rser.2019.03.005.

Cruz-Morales, P., Vijgenboom, E., Iruegas-Bocardo, F., Girard, G., Yanez-Guerra, L. A., Ramos-Aboites, H. E., Pernodet, J. L., Anne, J., van Wezel, G. P. and Barona-Gomez, F., (2013) 'The Genome Sequence of *Streptomyces lividans* 66 Reveals a Novel tRNA-Dependent Peptide Biosynthetic System within a Metal-Related Genomic Island', Genome Biology and Evolution, 5(6), pp. 1165–1175. doi: 10.1093/gbe/evt082.

Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Baslé, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J. and Gilbert, H. J., (2012) 'How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity', Proceedings of the National Academy of Sciences of the United States of America, 109(51), pp. 20889–20894. doi: 10.1073/pnas.1212034109.

Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B. and Terrapon, N., (2022) 'The carbohydrate-active enzyme database: Functions and literature', Nucleic Acids Research, 50(D1), pp. D571–D577. doi: 10.1093/nar/gkab1045.

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V. and Durinx, C., (2021) 'Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users', Nucleic Acids Research, 49(W1), pp. W216–W227. doi: 10.1093/nar/gkab225.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K., (2010) 'Features and development of Coot', Acta Crystallographica Section D Biological Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Evaluation of certain food additives: eighty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. World Health Organization. (2021) World Health Organization - Technical Report Series 1027. World Health Organization, Food and Agriculture Organization of the United Nations & Joint FAO/WHO Expert Committee on Food Additives.

Felby, C., Nielsen, B. R., Olesen, P. O. and Skibsted, L. H., (1997) ‘Identification and quantification of radical reaction intermediates by electron spin resonance spectrometry of laccase-catalyzed oxidation of wood fibers from beech (Fagus sylvatica)’, Applied Microbiology and Biotechnology, 48(4), pp. 459–464. doi: 10.1007/s002530051080.

Fernandes, S. D., Trautmann, N. M., Streets, D. G., Roden, C. A. and Bond, T. C., (2007) ‘Global biofuel use, 1850-2000’, Global Biogeochemical Cycles, 21(2). doi: 10.1029/2006GB002836.

Ficko-Blean, E., Gregg, K. J., Adams, J. J., Hehemann, J. H., Czjzek, M., Smith, S. P. and Boraston, A. B., (2009) ‘Portrait of an enzyme, a complete structural analysis of a multimodular β-N-acetylglucosaminidase from Clostridium perfringens’, Journal of Biological Chemistry, 284(15), pp. 9876–9884. doi: 10.1074/jbc.M808954200.

Ficko-Blean, E. and Boraston, A. B., (2012) ‘Insights into the recognition of the human glycome by microbial carbohydrate-binding modules’, Current Opinion in Structural Biology, 22(5), pp. 570–577. doi: 10.1016/j.sbi.2012.07.009.

Field, C. B., Cambell, J. E. and Lobell, D. B., (2008) ‘Biomass energy: the scale of the potential resource’, Trends in Ecology & Evolution, 23(2), pp. 65–72. doi: 10.1016/j.tree.2007.12.001.

Filiatrault-Chastel, C., Navarro, D., Haon, M., Grisel, S., Herpoël-Gimbert, I., Chevret, D., Fanuel, M.,, Henrissat, B.,Heiss-Blanquet, S., Margeot, A., Berrin, J. G., (2019) 'AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes', Biotechnology for Biofuels, 12(1), p. 55. doi: 10.1186/s13068-019-1394-y.

Forsberg, Z., Røhr, Å. K., Mekasha, S., Andersson, K. K., Eijsink, V. G. H., Vaaje-Kolstad, G. and Sørlie, M., (2014) 'Comparative Study of Two Chitin-Active and Two Cellulose-Active AA10-Type Lytic Polysaccharide Monooxygenases', Biochemistry, 53(10), pp. 1647–1656. doi: 10.1021/bi5000433.

Forsberg, Z., Mackenzie, A. K., Sørlie, M., Røhr, Å. K., Helland, R., Arvai, A. S., Vaaje-Kolstad, G. and Eijsink, V. G. H., (2014) 'Structural and functional characterization of a conserved pair of bacterial cellulose-oxidizing lytic polysaccharide monooxygenases', Proceedings of the National Academy of Sciences, 111(23), pp. 8446–8451. doi: 10.1073/pnas.1402771111.

Forsberg, Z., Nelson, C. E., Dalhus, B., Mekasha, S., Loose, J. S., Crouch, L. I., Røhr, Å. K., Gardner, J. G., Eijsink, V. G. H. and Vaaje-Kolstad, G., (2016) 'Structural and functional analysis of a lytic polysaccharide monooxygenase important for efficient utilization of chitin in *Cellvibrio japonicus*', Journal of Biological Chemistry, 291(14), pp. 7300–7312. doi: 10.1074/jbc.M115.700161.

Forsberg, Z., Bissaro, B., Gullesen, J., Dalhus, B., Vaaje-Kolstad, G. and Eijsink, V. G. H., (2018) ‘Structural determinants of bacterial lytic polysaccharide monooxygenase functionality’, Journal of Biological Chemistry, 293(4), pp. 1397–1412. doi: 10.1074/jbc.M117.817130.

Frandsen, K. E., Tovborg, M., Jørgensen, C. I., Spodsberg, N., Rosso, M. N., Hemsworth, G. R., Garman, E. F., Grime, G. W., Poulsen, J. C. N., Batth, T. S. and Miyauchi, S., (2019) ‘Insights into an unusual Auxiliary Activity 9 family member lacking the histidine brace motif of lytic polysaccharide monooxygenases’, Journal of Biological Chemistry, 294(45), pp. 17117-17130. doi: 10.1074/jbc.RA119.009223.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., Bairoch, A., (2005) ‘Protein identification and analysis tools on the ExPASy server’, in The Proteomics Protocols Handbook, pp. 571–607. doi: 10.1385/1592598900.

Georgelis, N., Tabuchi, A., Nikolaidis, N. and Cosgrove, D. J., (2011) ‘Structure-function analysis of the bacterial expansin EXLX1’, Journal of Biological Chemistry, 286(19), pp. 16814–16823. doi: 10.1074/jbc.M111.225037.

Giampietro, M., Ulgiati, S. and Pimentel, D., (1997) ‘Feasibility of Large-Scale Biofuel Production’, BioScience, 47(9), pp. 587–600. doi: 10.2307/1313165.

Gilkes, N. R., Warren, R., Miller Jr, R. C. and Kilburn, D. G., (1988) ‘Precise excision of the cellulose binding domains from two Cellulomonas fimi cellulases by a homologous protease and the effect on catalysis.’, The Journal of biological chemistry, 263(21), pp. 10401–10407. doi: 10.1016/s0021-9258(19)81530-2.

Gumulya, Y., Baek, J. M., Wun, S. J., Thomson, R. E., Harris, K. L., Hunter, D. J., Behrendorff, J. B., Kulig, J., Zheng, S., Wu, X. and Wu, B., (2018) ‘Engineering highly functional thermostable proteins using ancestral sequence reconstruction’, Nature Catalysis, 1(11), pp. 878–888. doi: 10.1038/s41929-018-0159-5.

Hangasky, J. A., Iavarone, A. T. and Marletta, M. A., (2018) ‘Reactivity of O2 versus H2O2 with polysaccharide monooxygenases’, Proceedings of the National Academy of Sciences of the United States of America, 115(19), pp. 4915–4920. doi: 10.1073/pnas.1801153115.

Harms, M. J. and Thornton, J. W. (2010) ‘Analyzing protein structure and function using ancestral gene reconstruction’, Current Opinion in Structural Biology, 20(3), pp. 360–366. doi: 10.1016/j.sbi.2010.03.005.

Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S. and Xu, F., (2010) ‘Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family’, Biochemistry, 49(15), pp. 3305–3316. doi: 10.1021/bi100009p.

Hedegård, E. D. and Ryde, U. (2018) ‘Molecular mechanism of lytic polysaccharide monooxygenases’, Chemical Science, 9(15), pp. 3866–3880. doi: 10.1039/c8sc00426a.

Hedison, T. M., Breslmayr, E., Shanmugam, M., Karnpakdee, K., Heyes, D. J., Green, A. P., Ludwig, R., Scrutton, N. S. and Kracher, D., (2021) ‘Insights into the H2O2-driven catalytic mechanism of fungal lytic polysaccharide monooxygenases’, FEBS Journal, 288(13), pp. 4115–4128. doi: 10.1111/febs.15704.

Hemsworth, G. R., Johnston, E. M., Davies, G. J. and Walton, P. H., (2015) ‘Lytic Polysaccharide Monooxygenases in Biomass Conversion’, Trends in Biotechnology, 33(12), pp. 747–761. doi: 10.1016/j.tibtech.2015.09.006.

Hemsworth, G. R. and Lo Leggio, L. (2019) ‘Auxiliary Activity Family 13’ in CAZypedia. Available at: https://www.cazypedia.org/index.php/Auxiliary_Activity_Family_13 (Accessed: 18 February 2021).

Hodgson, D. A. (2000) ‘Primary metabolism and its control in Streptomycetes: A most unusual group of bacteria’, Advances in Microbial Physiology, 42, pp. 47–238. doi: 10.1016/s0065-2911(00)42003-5.

Hu, F. and Ragauskas, A. (2012) ‘Pretreatment and Lignocellulosic Chemistry’, Bioenergy Research, 5(4), pp. 1043–1066. doi: 10.1007/s12155-012-9208-0.

Hu, J., Arantes, V., Pribowo, A., Gourlay, K. and Saddler, J. N., (2014) ‘Substrate factors that influence the synergistic interaction of AA9 and cellulases during the enzymatic hydrolysis of biomass’, Energy Environ. Sci., 7(7), pp. 2308–2315. doi: 10.1039/C4EE00891J.

Joy, J. B., Liang, R. H., McCloskey, R. M., Nguyen, T. and Poon, A. F., (2016) ‘Ancestral Reconstruction’, PLOS Computational Biology, 12(7), p. e1004763. doi: 10.1371/journal.pcbi.1004763.

Kapsokalyvas, D., Wilbers, A., Boogers, I. A., Appeldoorn, M. M., Kabel, M. A., Loos, J. and Van Zandvoort, M. A., (2018) ‘Biomass pretreatment and enzymatic hydrolysis dynamics analysis based on particle size imaging’, Microscopy and Microanalysis, 24(5), pp. 517–525. doi: 10.1017/S1431927618015143.

Keegan, R. M. and Winn, M. D., (2007) ‘Automated search-model discovery and preparation for structure solution by molecular replacement’, Acta Crystallographica Section D Biological Crystallography, 63(4), pp. 447–457. doi: 10.1107/S0907444907002661.

Khersonsky, O. and Tawfik, D. S., (2010) ‘Enzyme promiscuity: A mechanistic and evolutionary perspective’, Annual Review of Biochemistry, 79, pp. 471–505. Doi: 10.1146/annurev-biochem-030409-143718.

Kim, I. J., Youn, H. J. and Kim, K. H., (2016) ‘Synergism of an auxiliary activity 9 (AA9) from *Chaetomium globosum* with xylanase on the hydrolysis of xylan and lignocellulose’, Process Biochemistry, 51(10), pp. 1445–1451. doi: 10.1016/j.procbio.2016.06.017.

Kitamoto, K., (2002) ‘Molecular Biology of the Koji Molds’, in Advances in Applied Microbiology, pp. 129–153. doi: 10.1016/S0065-2164(02)51004-2.

Kobayashi, T., Abe, K., Asai, K., Gomi, K., Juvvadi, P. R., Kato, M., Kitamoto, K., Takeuchi, M. and Machida, M., (2007) ‘Genomics of *Aspergillus oryzae*’, Bioscience, Biotechnology and Biochemistry, 71(3), pp. 646–670. doi: 10.1271/bbb.60550.

Kour, D., Rana, K. L., Yadav, N., Yadav, A. N., Rastegari, A. A., Singh, C., Negi, P., Singh, K. and Saxena, A. K., (2019) ‘Technologies for Biofuel Production: Current Development, Challenges, and Future Prospects’, in Springer International Publishing, pp. 1–50. doi: 10.1007/978-3-030-14463-0_1.

Lo Leggio, L., Simmons, T. J., Poulsen, J.C.N., Frandsen, K. E., Hemsworth, G. R., Stringer, M. A., Von Freiesleben, P., Tovborg, M., Johansen, K. S., De Maria, L. and Harris, P. V., (2015) 'Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase', Nature Communications, 6(1), p. 5961. doi: 10.1038/ncomms6961.

Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. and Henrissat, B., (2013) 'Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes', Biotechnology for Biofuels, 6(1), pp. 1–14. doi: 10.1186/1754-6834-6-41.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. and Henrissat, B., (2014) 'The carbohydrate-active enzymes database (CAZy) in 2013', Nucleic Acids Research, 42(D1), pp. 490–495. doi: 10.1093/nar/gkt1178.

MacNeill, G. J., Mehrpouyan, S., Minow, M. A., Patterson, J. A., Tetlow, I. J. and Emes, M.J., (2017) 'Starch as a source, starch as a sink: The bifunctional role of starch in carbon allocation', Journal of Experimental Botany, 68(16), pp. 4433–4453. doi: 10.1093/jxb/erx291.

Mahmood, Z., Yameen, M., Jahangeer, M., Riaz, M., Ghaffar, A. and Javid, I., (2018) 'Lignin as Natural Antioxidant Capacity', in Lignin - Trends and Applications. InTech. doi: 10.5772/intechopen.73284.

Maki, M., Leung, K. T. and Qin, W., (2009) ‘The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass’, International Journal of Biological Sciences, 5(5), pp. 500–516. doi: 10.7150/ijbs.5.500.

Manjeet, K., Madhuprakash, J., Mormann, M., Moerschbacher, B. M. and Podile, A. R., (2019) ‘A carbohydrate binding module-5 is essential for oxidative cleavage of chitin by a multi-modular lytic polysaccharide monooxygenase from *Bacillus thuringiensis* serovar *kurstaki*’, International Journal of Biological Macromolecules, 127, pp. 649–656. doi: 10.1016/j.ijbiomac.2019.01.183.

Meitzler, J. L., Hinde, S., Bánfi, B., Nauseef, W. M. and de Montellano, P. R. O., (2013) ‘Conserved cysteine residues provide a protein-protein interaction surface in dual oxidase (DUOX) Proteins’, Journal of Biological Chemistry, 288(10), pp. 7147–7157. doi: 10.1074/jbc.M112.414797.

Merkl, R. and Sterner, R., (2016) ‘Ancestral protein reconstruction: Techniques and applications’, Biological Chemistry, 397(1), pp. 1–21. doi: 10.1515/hsz-2015-0158.

Morand, S., Agnandji, D., Noel-Hudson, M. S., Nicolas, V., Buisson, S., Macon-Lemaitre, L., Gnidehou, S., Kaniewski, J., Ohayon, R., Virion, A. and Dupuy, C., (2004) ‘Targeting of the dual oxidase 2 N-terminal region to the plasma membrane’, Journal of Biological Chemistry, 279(29), pp. 30244–30251. doi: 10.1074/jbc.M405406200.

Muderspach, S. J., Tandrup, T., Frandsen, K. E., Santoni, G., Poulsen, J. C. N. and Lo Leggio, L., (2019) ‘Further structural studies of the lytic polysaccharide monooxygenase Ao AA13 belonging to the starch-active AA13 family’, Amylase, 3(1), pp. 41–54. doi: 10.1515/amylase-2019-0004.

Müller, G., Várnai, A., Johansen, K. S., Eijsink, V. G. H. and Horn, S. J., (2015) ‘Harnessing the potential of LPMO-containing cellulase cocktails poses new demands on processing conditions’, Biotechnology for Biofuels, 8(1), p. 187. doi: 10.1186/s13068-015-0376-y.

Müller, G., Chylenski, P., Bissaro, B., Eijsink, V. G. H. and Horn, S. J., (2018) ‘The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail’, Biotechnology for Biofuels, 11(1), pp. 1–17. doi: 10.1186/s13068-018-1199-4.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J., (1997) ‘Refinement of macromolecular structures by the maximum-likelihood method’, Acta Crystallographica Section D: Biological Crystallography, 53(3), pp. 240–255. doi: 10.1107/S0907444996012255.

Namba, S., Kato, H., Shigenobu, S., Makino, T. and Moriya, H., (2022) ‘Massive expression of cysteine-containing proteins causes abnormal elongation of yeast cells by perturbing the proteasome’, G3: Genes, Genomes, Genetics, 12(6). doi: 10.1093/g3journal/jkac106.

Nanda, S., Azargohar, R., Dalai, A. K. and Kozinski, J. A., (2015) ‘An assessment on the sustainability of lignocellulosic biomass for biorefining’, Renewable and Sustainable Energy Reviews, 50, pp. 925–941. doi: 10.1016/j.rser.2015.05.058.

Nanda, S., Rana, R., Sarangi, P. K., Dalai, A. K. and Kozinski, J. A., (2018) ‘A Broad Introduction to First-, Second-, and Third-Generation Biofuels’, in Recent Advancements in Biofuels and Bioenergy Utilization. Singapore: Springer Singapore, pp. 1–25. doi: 10.1007/978-981-13-1307-3_1.

Nanda, S., Dalai, A. K. and Kozinski, J. A., (2014) ‘Butanol and ethanol production from lignocellulosic feedstock: biomass pretreatment and bioconversion’, Energy Science & Engineering, 2(3), pp. 138–148. doi: 10.1002/ese3.41.

Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K. and Lebedev, A., (2018) ‘CCP4i2: the new graphical user interface to the CCP 4 program suite’, Acta Crystallographica Section D Structural Biology, 74(2), pp. 68–84. doi: 10.1107/S2059798317016035.

Quinlan, R. J., Sweeney, M.D., Lo Leggio, L., Otten, H., Poulsen, J. C. N., Johansen, K. S., Krogh, K. B., Jørgensen, C.I., Tovborg, M., Anthonsen, A. and Tryfona, T., (2011) ‘Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components’, Proceedings of the National Academy of Sciences, 108(37), pp. 15079–15084. doi: 10.1073/pnas.1105776108.

Ramos, J. L., Valdivia, M., García-Lorente, F. and Segura, A., (2016) 'Benefits and perspectives on the use of biofuels', Microbial Biotechnology, 9(4), pp. 436–440. doi: 10.1111/1751-7915.12356.

Reyes-Ortiz, V., Heins, R. A., Cheng, G., Kim, E. Y., Vernon, B.C., Elandt, R. B., Adams, P. D., Sale, K. L., Hadi, M. Z., Simmons, B. A. and Kent, M. S., (2013) 'Addition of a carbohydrate-binding module enhances cellulase penetration into cellulose substrates', Biotechnology for Biofuels, 6(1), p. 1. doi: 10.1186/1754-6834-6-93.

Rodionova, M. V., Poudyal, R. S., Tiwari, I., Voloshin, R. A., Zharmukhamedov, S. K., Nam, H. G., Zayadan, B. K., Bruce, B. D., Hou, H. J. and Allakhverdiev, S. I., (2017) 'Biofuel production: Challenges and opportunities', International Journal of Hydrogen Energy, 42(12), pp. 8450–8461. doi: 10.1016/j.ijhydene.2016.11.125.

Rodríguez-Zúñiga, U. F., Cannella, D., de Campos Giordano, R., Giordano, R. D. L. C., Jørgensen, H. and Felby, C., (2015) 'Lignocellulose pretreatment technologies affect the level of enzymatic cellulose oxidation by LPMO', Green Chemistry, 17(5), pp. 2896–2903. doi: 10.1039/c4gc02179g.

Schrödinger LLC (no date) 'The PyMOL Molecular Graphics System, Version 2.5'. Available at: https://pymol.org/2/ (Accessed: 2 March 2019).

Sidar, A., Albuquerque, E. D., Voshol, G. P., Ram, A. F., Vijgenboom, E. and Punt, P. J., (2020) ‘Carbohydrate Binding Modules: Diversity of Domain Architecture in Amylases and Cellulases from Filamentous Microorganisms’, Frontiers in Bioengineering and Biotechnology, 8, pp. 871. doi: 10.3389/fbioe.2020.00871.

Siddiq, M. A., Hochberg, G. K. and Thornton, J. W., (2017) ‘Evolution of protein specificity: insights from ancestral protein reconstruction’, Current Opinion in Structural Biology, 47, pp. 113–122. doi: 10.1016/j.sbi.2017.07.003.

Simpson, T., (2009) ‘Biofuels: The Past, Present, and a New Vision for the Future’, BioScience, 59(11), pp. 926–927. doi: 10.1525/bio.2009.59.11.2.

Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P. and Kobe, B., (2003) ‘Crystal structures of fusion proteins with large-affinity tags’, Protein Science, 12(7), pp. 1313–1322. doi: 10.1110/ps.0243403.

Solomon, E.I., Heppner, D. E., Johnston, E. M., Ginsbach, J. W., Cirera, J., Qayyum, M., Kieber-Emmons, M. T., Kjaergaard, C. H., Hadt, R. G. and Tian, L., (2014) ‘Copper active sites in biology’, Chemical Reviews, 114(7), pp. 3659–3853. doi: 10.1021/cr400327t.

Song, B., Li, B., Wang, X., Shen, W., Park, S., Collings, C., Feng, A., Smith, S. J., Walton, J. D. and Ding, S.Y., (2018) ‘Real-time imaging reveals that lytic polysaccharide monooxygenase promotes cellulase activity by increasing cellulose accessibility’, Biotechnology for Biofuels, 11(1), p. 41. doi: 10.1186/s13068-018-1023-1.

Sorek, N., Yeats, T. H., Szemenyei, H., Youngs, H. and Somerville, C. R., (2014) ‘The implications of lignocellulosic biomass chemical composition for the production of advanced biofuels’, BioScience, 64(3), pp. 192–201. doi: 10.1093/biosci/bit037.

Spence, M. A., Kaczmarski, J. A., Saunders, J. W. and Jackson, C. J., (2021) ‘Ancestral sequence reconstruction for protein engineers’, Current Opinion in Structural Biology, 69, pp. 131–141. doi: 10.1016/j.sbi.2021.04.001.

Taherzadeh, M. J. and Karimi, K., (2007) Enzymatic-Based Hydrolysis Processes for Ethanol, BioResources. doi: 10.15376/biores.2.4.707-738.

Tamburrini, K. C., Terrapon, N., Lombard, V., Bissaro, B., Longhi, S. and Berrin, J. G., (2021) ‘Bioinformatic analysis of lytic polysaccharide monooxygenases reveal the pan-families occurrence of intrinsically disordered C-terminal extensions’, Biomolecules, 11(11). doi: 10.3390/biom11111632.

Tandrup, T., Frandsen, K. E., Johansen, K. S., Berrin, J. G. and Lo Leggio, L., (2018) ‘Recent insights into lytic polysaccharide monooxygenases (LPMOs)’, Biochemical Society Transactions, 46(6), pp. 1431–1447. doi: 10.1042/BST20170549.

Taylor, M. E. and Drickamer, K., (2014) ‘Convergent and divergent mechanisms of sugar recognition across kingdoms’, Current Opinion in Structural Biology, 28(1), pp. 14–22. doi: 10.1016/j.sbi.2014.07.003.

The CAZypedia Consortium (2018) ‘Ten years of CAZypedia: A living encyclopedia of carbohydrate-active enzymes’, Glycobiology, 28(1), pp. 3–8. doi: 10.1093/glycob/cwx089.

Tomme, P., Van Tilbeurgh, H., Pettersson, G., van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T. and Claeyssens, M., (1988) ‘Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: Analysis of domain function in two cellobiohydrolases by limited proteolysis’, European Journal of Biochemistry, 170(3), pp. 575–581. doi: 10.1111/j.1432-1033.1988.tb13736.x.

Torres Pazmiño, D. E., Winkler, M., Glieder, A. and Fraaije, M. W., (2010) ‘Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications’, Journal of Biotechnology, 146(1–2), pp. 9–24. doi: 10.1016/j.jbiotec.2010.01.021.

Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M. and Eijsink, V. G. H., (2010) ‘An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides’, Science, 330(6001), pp. 219–222. doi: 10.1126/science.1192231.

Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B. and Eijsink, V. G. H., (2017) ‘Structural diversity of lytic polysaccharide monooxygenases’, Current Opinion in Structural Biology, 44, pp. 67–76. doi: 10.1016/j.sbi.2016.12.012.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G. N., (2004) ‘REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use’, Acta Crystallographica Section D: Biological Crystallography, 60(12 I), pp. 2184–2195. doi: 10.1107/S0907444904023510.

Voshol, G. P., Punt, P. J. and Vijgenboom, E., (2019) ‘Profile comparer extended: Phylogeny of lytic polysaccharide monooxygenase families using profile hidden markov model alignments’, F1000Research, 8, pp. 1–15. doi: 10.12688/f1000research.21104.1.

Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R. and Marletta, M. A., (2014) ‘A family of starch-active polysaccharide monooxygenases’, Proceedings of the National Academy of Sciences of the United States of America, 111(38), pp. 13822–13827. doi: 10.1073/pnas.1408090111.

Walton, P. H. and Davies, G. J., (2016) ‘On the catalytic mechanisms of lytic polysaccharide monooxygenases’, Current Opinion in Chemical Biology, 31, pp. 195–207. Doi: 10.1016/j.cbpa.2016.04.001.

Wilson, D. B., (2011) ‘Microbial diversity of cellulose hydrolysis’, Current Opinion in Microbiology, 14(3), pp. 259–263. doi: 10.1016/j.mib.2011.04.004.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A. and McNicholas, S. J., (2011) ‘Overview of the CCP 4 suite and current developments’, Acta Crystallographica Section D Biological Crystallography, 67(4), pp. 235–242. doi: 10.1107/S0907444910045749.