NTNU
Norwegian University of Science and Technology

Recombinant expression and structural investigation of *Jd*LPMO10A using Nuclear magnetic Resonance (NMR) spectroscopy

Idd Andrea Christensen

Master in Biotechnology
Submission date: June 2019
Supervisor: Finn Lillelund Aachmann, IBT
Co-supervisor: Gaston Courtade, IBT

Norwegian University of Science and Technology,
Department of Biotechnology and Food Science

Abstract

Sustainable production of carbon-neutral energy is of paramount importance if the goals set by the Paris agreement in 2015 to limit temperature rise to 2.0 degrees is to be achieved. Fuel derived from renewable organic sources, such as abundant and recalcitrance polysaccharides has been an area of focus. However, effective exploitation of recalcitrance polysaccharides depends on a sustainable and cost-effective method for deconstructing the materials. Lytic Polysaccharide Monooxygenases (LPMOs) are a recently discovered group of copper-dependent enzymes found in numerous organisms, mainly bacteria, and fungi. LPMOs play an important part in the degradation of biomass such as cellulose and chitin, as the enzymes have been found to cleave the β1,4-bonds found in these crystalline substrates through an oxidative mechanism, boosting the effect of traditional glycosyl hydrolases. Although LPMOs have been the subject of intense research since their discovery in 2010, many aspects of the function of these enzymes are yet to be uncovered.

*Jd*LPMO10A is a small, chitin-active, LPMO produced by the Gram-negative bacterium *Jonesia denitrificans*. *Jd*LPMO10A is an interesting LPMO to study, as it could shed some light onto the minimal scaffold necessary for LPMO activity. The crystal structure of *Jd*LPMO10A was determined in 2015. An NMR structure of *Jd*LPMO10A is of interest, as NMR spectroscopy offers the opportunity to study protein function near physiological conditions. This thesis describes the structural investigation of *Jd*LPMO10A using NMR spectroscopy. Isotopically enriched samples of *Jd*LPMO10A were produced using the XylS/Pm based expression system pJB_nSP_*Jd*. In total, 91.6% of the backbone chemical shifts for $^{13}C_{\alpha}$, ^{13}CO, ^{15}N and $^{1}H^{N}$ were assigned by NMR, using the sequence-specific assignment method. Dihedral angles and secondary structural elements, predicted by the computational method TALOS-N, were in agreement with the crystal structure. However, some secondary structural elements were not predicted in segments of the primary structure lacking chemical shift information. Finally, a CS-ROSETTA model of *Jd*LPMO10A was generated by the computational method CS-ROSETTA. The immunoglobulin-like β-sandwich core in the CS-ROSETTA model were in good agreement with the crystal structure. However, residues without chemical shift information precluded prediction of certain structural elements, particularly in the *L2* region of the enzyme. The CS-ROSETTA model is an important step towards a complete NMR structure of *Jd*LPMO10A, which in turn can be used to better our chitinolytic LPMOs in general.

Sammendrag

En bærekraftig produksjon av karbonnøytral energi er ekstremt viktig dersom verdenssamfunnet skal oppnå 2,0 graders målet satt ved Parisavtalen i 2015. Et stort fokus har derfor blitt lagt på utviklingen av fornybare energikilder fra organiske ressurser som forekommende og motstandsdyktige polysakkarider. Derimot er effektiv utnyttelse av disse motstandsdyktige polysakkaridene avhengig av at man utvikler en bærekraftig og kostnads effektive metode for å bryte ned materialene. Lytisk Polysakkarid Monoksygenaser (LPMOer) er en relativt nyoppdaget gruppe kopper-avhengige enzymer som finnes i en rekke organismer, men hovedsakelig i bakterier og sopp. LPMOer spiller en viktig rolle i nedbrytningen av biomaterialer som cellulose og kitin da det er vist at enzymene spalter β1,4-bindingene som finnes i disse krystalline substratene via en oksidativ mekanisme, og øker dermed effekten av tradisjonelle glykosid hydrolaser. Selv om det har vært forsket mye på LPMOer siden de ble oppdaget i 2010 er det mange aspekter ved enzymene som fremdeles ikke er avklarte.

*Jd*LPMO10A er en liten kitinaktiv LPMO, produsert av den Gram-negative bakterien *Jonesia denitrificans*. *Jd*LPMO10A er interessant å studere da den kan belyse hvilke elementer som er essensielle for LPMO aktivitet. Krystalstrukturen til *Jd*LPMO10A ble bestemt i 2015 av Mekasha med flere. Det hadde derimot også vært interessant å ha en NMR struktur av *Jd*LPMO10A ettersom det er mulig å studere proteiner nær fysiologiske forhold med NMR. Denne oppgaven beskriver en strukturell studie av *Jd*LPMO10A ved hjelp av NMR spektroskopi. Isotopberikete prøver av *Jd*LPMO10A ble produsert ved å bruke det XylS/Pm basert vektorsystem pJB_nSP_*Jd*. NMR ble brukt til å tildele 91,6% av de kjemiske skiftene til $^{13}C_{\alpha}$, ^{13}CO, ^{15}N og $^{1}H^{N}$ in primærstrukturen til *Jd*LPMO10A ved ved a bruke den sekvens-spesifikke tilordnings metoden. Dihedrale vinklene og sekundær strukturen til *Jd*LPMO10A, forutsett av dataprogrammet TALOS-N, stemte overens med krystalstrukturen. Deler av primær strukturen uten kjemisk skift informasjon manglet derimot forutbestemte sekundærstruktur elementer. Avslutningsvis ble en foreløpig struktur for *Jd*LPMO10A generert ved å bruke dataprogrammet CS-ROSETTA. Deler av den foreløpige strukturen passet god overens med krystalstrukturen. Deler av primærstrukturen uten kjemisk skift informasjon gjorde at noen sekundærstruktur elementer manglet, spesielt i *L2* regionen til enzymet. Den foreløpige strukturen er et viktig steg mot en komplett NMR struktur av

*Jd*LPMO10A, som videre kan brukes til å forbedre vår generelle forståelse av kitinolytiske LPMOer.

Preface

This thesis concludes my Master of Science in Biotechnology at the Norwegian University of Science and Technology (NTNU). All the work described has been carried out at the Department of Biotechnology and Food Sciences in the period of August 2018 to June 2019.

I wish to extend my gratitude to all the members of the research group. Thank you to Finn Lillelund Aachmann, my main supervisor, for accepting me as a master student and always taking the time from our busy schedule to discuss my results and provide excellent feedback on my work. Thank you to Gaston Courtade, my co-supervisor, for guiding me both in the lab, NMR lab, and with the writing of my thesis. You have been a dedicated mentor, which a greatly appreciate!

I also wish to extend my gratitude to Gerd Inger Sætrom for helping me in the lab and providing support when everything in the lab was new and cumbersome. Knowing you were only a phone call away made all the expensive equipment in the lab seem much less scary!

Thank you to Edith Buchinger, Olav Andreas Tøndervik, Eva Madland, Margrethe Gaardløs, and Ina Sander Pedersen for providing interesting group meetings and discussions.

Thanks to the PEP-group for creating the PUCBB_*Jd* plasmid used to produce a ^{15}N-labeled protein sample in this project.

Finally, I wish to thank my family and my soon-to-be husband Frithof Bjørnstad for supporting me and providing endless amounts of tea.

Table of Contents

Symbols and Abbreviations

AA9	Auxiliary Activity Family 9
AA10	Auxiliary Activity Family 10
AA11	Auxiliary Activity Family 11
AA13	Auxiliary Activity Family 13
AA14	Auxiliary Activity Family 14
AA15	Auxiliary Activity Family 15
ANN	Artificial neural network
CARA	Computer Aided Resonance Assignment
CAZy	Carbohydrate-Active enZYmes Database
CMB	Carbohydrate Binding Module
CMB33	Carbohydrate Binding Module Family 33
CBP21	Carbohydrate Binding Protein 21 (also referred to as *Sm*LPMO10A)
CDH	Cellobiose Dehydrogenase
COSY	Correlation Spectroscopy
CS-ROSETTA	Chemical Shift-ROSETTA
C_α, C_β, CO	Alpha, Beta and Carbonyl carbon in an amino acid
FID	Free Induction Decay
FPLC	Fast Proton Liquid Chromatography
GH	Glyosidic Hydrolase
GH61	Glyosidic Hydrolase Family 61
HSQC	Heteronuclear Single Quantum Correlations
*Jd*LPMO10A	AA10 family member originating from the bacterium Jonesia denitrificans
LB	Lysogeny Broth
LPMO	Lytic Polysaccharide Monooxygenase
M9	Type of defined medium with solely one nitrogen and carbon source
MS	Mass Spectroscopy
mAU	Milli Absorbance Unit
NMR	Nuclear Magnetic Resonance Spectroscopy
NOE	Nuclear Overhauser Effect
NOSEY	Nuclear Overhauser Effect Spectroscopy
PDB	Protein Data Bank

ppm	parts per million
rmsd	root mean square deviation
r.t	Room temperature
SDS-PAGE	Sodium Dodecyl Sulphate PolyAcrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
T7lac/pRSET	Expression system relying on the lac operon and T7 RNA polymerase
TALOS-N	Torsion Angle Likeliness Obtained from Shift and Sequence Similarity
δ	The chemical shift (given in ppm)
φ	Phi (dihedral angle)
ψ	Psi (dihedral angle)
ϖ	Omega (dihedral angle)

1 Introduction

1.1 Chitin and chitin degradation

1.1.1 The structure of Chitin

Chitin is the second most abundant polysaccharide on earth. The polymer is the primary component in the exoskeleton of arthropods and crustaceans, and in the cell wall of fungi and yeast (Hamid *et al.*, 2013). Chitin is made up of β-1,4 linked N-acetyl-D-glucosamine monomers. The chain geometry of chitin is similar to cellulose. Both polymers have an extended chain structure, where each monomer is rotated 180° relative to its neighbors. Both cellulose and chitin adopts recalcitrant crystalline structures (Smidsrød and Moe, 2008). The crystalline structure of chitin is similar to that of cellulose I, with chain stabilizing intramolecular hydrogen bonds between O(5)-OH(3)´, and between the oxygen of the acetyl group (C=O) and O(6)´. The intramolecular hydrogen bonding found in crystalline chitin is shown in figure 1.1. The crystalline structure is further stabilized by several inter-sheet hydrogen bonds, resulting in strong fibers which exhibit high resistance towards degradation (Smidsrød and Moe, 2008).

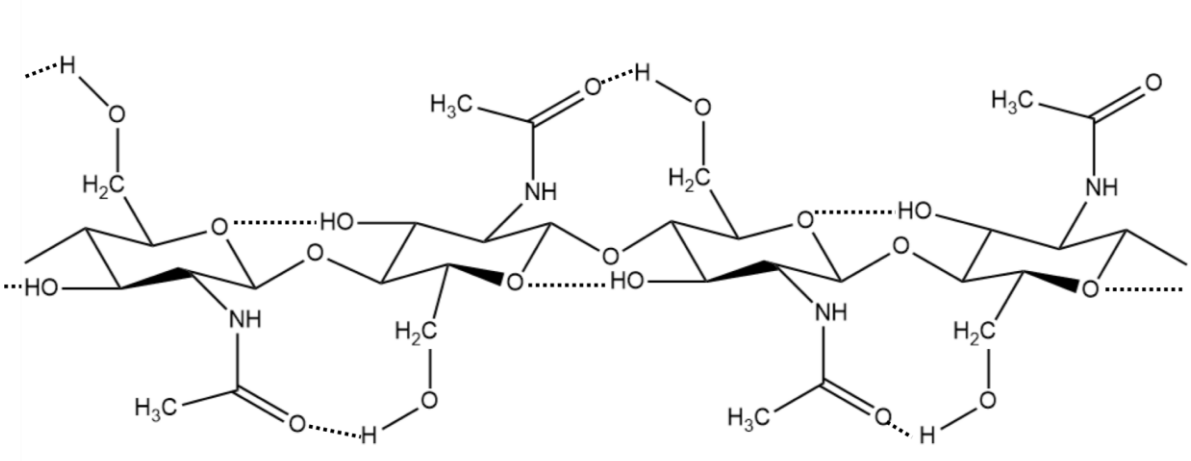

Figure 1.1: Diagram of the intramolecular hydrogen bonding found in crystalline chitin. Hydrogen bonds are found between O(5)-OH(3)´, and between the oxygen of the acetyl group (C=O) and O(6)´ of neighboring monomers. Each monomer in the chain is rotated 180° relative to its neighbors (Smidsrød and Moe, 2008). The figure was made by Christensen, I. A.

Crystalline chitin can be divided into the three structural allomorphs α-, β-, and γ-chitin, α-chitin being the most abundant of the three. The allomorphs are distinguished from one another by the way the individual chitin chains are oriented in the crystalline structure. The structure of α-chitin is characterized by having antiparallel chains, meaning that neighboring chains lie in the opposite direction. In β-chitin, the chains are organized in a parallel fashion. γ-chitin was long thought to be a variant of α-chitin but was later found to have its own "two up and one down" chain structure (Rinaudo, 2006), as shown in figure 1.2.

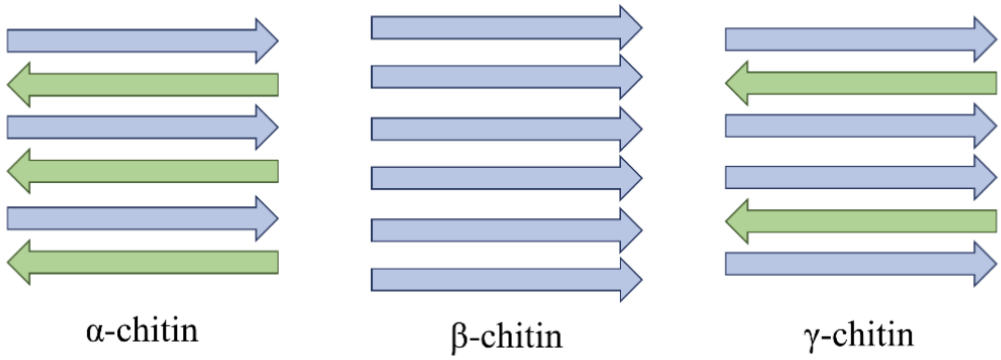

Figure 1.2: The crystalline structure of chitin can be divided into three allomorphs: α- β- and γ-chitin. Individual chitin chains are depicted as arrows, with the arrowhead representing the non-reducing end. In α-chitin, the chains are organized in an antiparallel fashion. β-chitin is characterized by having parallel chains. Lastly, γ-chitin have chains organized in a "two up one down" manner, where two parallel chains are interrupted by one antiparallel chain. The figure is collected from Brigham, 2017.

1.1.2 Chitin degradation by chitinases

Crystalline chitin is hydrolyzed by a number of enzymes collectively referred to as chitinases (Dahiya, Tewari and Hoondal, 2006). Chitinases, which are a subgroup of Glycoside Hydrolases (GHs), are produced by a variety of different organisms including bacteria, fungi, higher plants, viruses, and animals. The enzymes are broadly classified as either Endo- or Exo-chitinases based on their mode of action. Endochitinases cleave chitin chains at random internal sites, creating shorter oligomer fragments. Exochitinases cleaves off dimers, moving progressively from either the reducing or none-reducing end of the chitin chain (Sunny, Kumar and Kumar, 2018). N-acetylglucosaminidases produce N-acetylglucosamine (GlcNac) monomers by further cleaving the produced chitin oligomers and dimers (Sunny, Kumar and Kumar, 2018). Chitin degradation by chitinases is shown in figure 1.3. Most chitinolytic organisms produce multiple forms of chitinases which act together in synergy to break down crystalline chitin (Dahiya, Tewari and Hoondal, 2006).

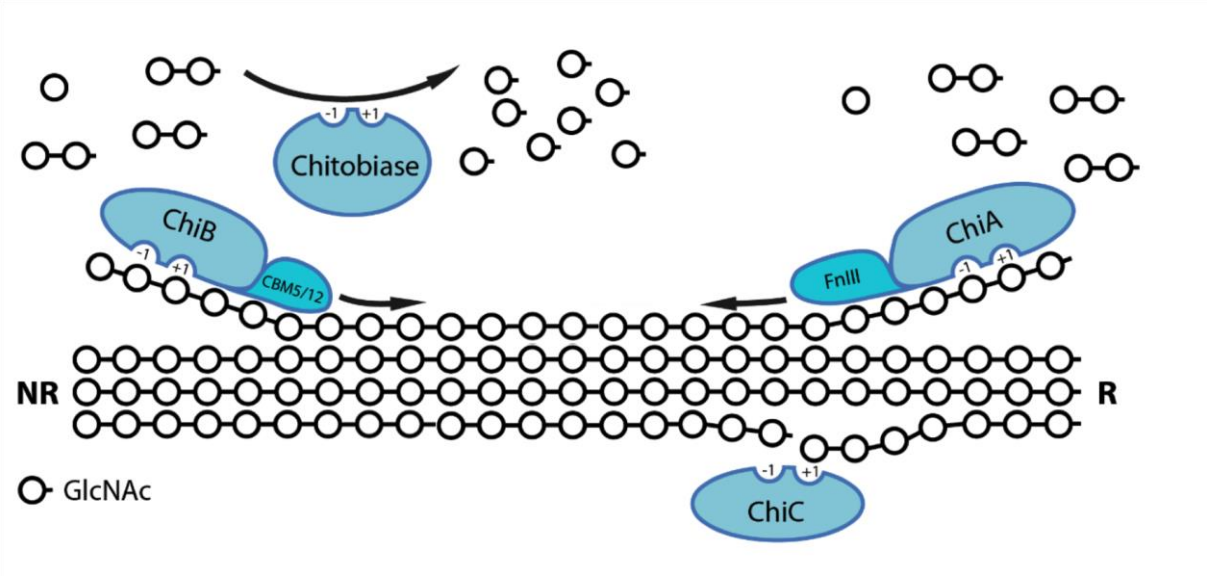

Figure 1.3: Schematic breakdown of crystalline chitin by traditional chitinases illustrated by the chitinolytic system of *S.marcescens*. The two progressive exochitinases ChiA and ChiB degrade the chitin chain for either the reducing end (R) or the non-reducing end (NR) releasing chitin dimers. The endochitinase ChiC cleaves the chitin chain at random internal sites releasing chitin oligomers. The released chitin di- and oligomers are further degraded into monomers by the N-acetylglucosaminidasen Chitobiase. The figure has been adapted from Vaaje-Kolstad *et al*., 2004.

1.2 Discovery of LPMOs

Despite much being known about the degradation of cellulose and chitin by GHs, the question of how these hydrolytic enzymes were able to degrade crystalline fibers, long remained unresolved (Walton and Davies, 2016). The active site of traditional GHs resides in the enzymes "carbohydrate binding cleft" which binds single, isolated, carbohydrate chains to induce chain cleavage. However, isolated chains are unavailable in crystalline substrates (Walton and Davies, 2016). In a forward-thinking paper Reese *et al.* proposed that an unknown group of enzymes was functioning to make crystalline substrates available to traditional GHs through a "C1" mechanism (Reese, Siu and Levinson, 1950; Walton and Davies, 2016).

It was not until 2010, when Vaaje-Kolstad *et al.* published a breakthrough paper, that the picture became clearer. Vaaje-Kolstad *et al.* reported that the enzyme Chitin Binding Protein 21 (CBP21) induced chain breaks in chitin through an oxidative mechanism (Vaaje-Kolstad *et al.*, 2010). At the time of this discovery, CBP21 was classified as a member of the non-catalytic Carbohydrate Binding Module 33 (CBM33) family. In 2004, Vaaje-Kolstad *et al*.

had observed the Gram-negative bacterium *Serratia marcescens* to excreted large amounts of CBP21 when grown on chitin (Vaaje-Kolstad *et al.*, 2004). Moreover, the presence of CBP21 was found to increase the efficiency of chitin degradation, presumably by boosting the effect of traditional chitinases (Vaaje-Kolstad *et al.*, 2005). In addition to discovering that CMB33 members were oxidases, Vaaje-Kolstad *et al.* revealed CBP21 to be a metalloenzyme with an immunoglobulin-like β-core structure and a flat substrate binding surface (Vaaje-Kolstad *et al.*, 2004).

When it was later discovered the enzyme *Tr*Cel61A, at the time classified as a member of the Glycoside Hydrolase family 61 (GH61), showed large structural similarities to CBP21 (Karkehabadi *et al.*, 2008), a possible link between the two enzyme families was suggested (Beeson *et al.*, 2015). The idea of a link between the enzyme families was reinforced when GH61s were found to be oxidative enzymes using copper as a metal co-factor (R Jason Quinlan *et al.*, 2011). The result was a reassignment of CBM33s and GH61s into members of the Auxiliary Activity (AA) Family in the Carbohydrate-Active EnZYmes (CAZy) database. The two enzyme families were joined and given the new name "Lytic Polysaccharide Monooxygenases" (LPMOs). Consequently, CMB33s and GHs are now reclassified as the AA10 family and the AA9 family, respectively (Levasseur *et al.*, 2013). The AA9s mainly consists of fungal LPMOs that act on cellulose, while most The AA10s are of bacterial origin, acting either on cellulose or chitin. The names of the LPMOs were also altered, renaming CBP21 to *Sm*LPMO10A (Hemsworth *et al.*, 2014).

The AA-family was later expanded when studies, using a module-walking approach, revealed LPMO-like domains that significantly diverged from the AA9s, giving rise to LPMO families AA11, AA13, AA14 and AA15 (Chou, 2007; Hemsworth *et al.*, 2014; Vu *et al.*, 2014; Lo Leggio *et al.*, 2015; Couturier *et al.*, 2018; Sabbadin *et al.*, 2018). AA11s have been found to act on chitin, AA13s on starch (Leggio *et al*., 2015), AA14 on xylan (Couturier *et al*., 2018), while the recently discovered AA15s are thought to be involved in food digestion in arthropods and algae (Sabbadin *et al*., 2018). In addition, LPMOs have been found to act on xyloglucan, glucomannan and other β-glucans (Hemsworth *et al*., 2014). The focus of this thesis will be on LPMOs found to cleave the β-1,4 bonds found in cellulose and chitin.

1.3 LPMO Structure

LPMOs are characterized by having few conserved sequences in their primary structure (Book *et al.*, 2014). This can in part be explained by large diversities between LPMOs when it comes to domain architecture. LPMOs are found as domains in large multimodular proteins, often associated with GHs or CMBs, and as independent catalytic units (Book *et al.*, 2014). Further, the size of LPMO domains also varies (Hemsworth *et al*, 2014). Despite significant variations in their primary structure, all LPMOs share the same overall fold in their tertiary structure (Beeson *et al*., 2016; Hemsworth *et al*., 2015). All LPMOs are translated as protein precursors with an N-terminal signal peptide. The signal peptide is cleaved off as the precursor is translocated into the periplasmic space of Gram-negative bacteria, resulting in the mature protein (Vaaje-Kolstad *et al*, 2005; Vaaje-Kolstad *et al*, 2010).

1.3.1 THE LPMO ACTIVE SITE:

The most conserved feature between LPMOs is the copper active site. In the copper active site, a copper atom is coordinated by three nitrogen ligands in a structural motif referred to as the "histidine brace" (Hemsworth *et al*., 2015). Quinlan *et al.* were the first to demonstrate that the metal ion found in the active site was, in fact, a type 2 copper ion using Isothermal Titration Calorimetry (R Jason Quinlan *et al.*, 2011). Their findings were later confirmed by NMR (Aachmann *et al.*, 2012). Previous to these studies, several metal ions including: Ca^{2+}, Ni^{2+}, Mn^{2+}, Co^{2+}, Mg^{2+}, and Zn^{2+} had been regarded as potential candidates responsible for LPMO activity (Aachmann *et al*., 2012). The presence of only one copper site separates LPMOs from other copper monooxygenases, which often utilize several copper atoms in a multi-electron reduction of O_2 to create reactive intermediates (Hemsworth *et al*., 2015).

Two of the three nitrogen atoms involved in the histidine brace originate from the N-terminal histidine, which is conserved between all LPMOs. The second nitrogen atom is located on another highly conserved histidine. Together they form an equatorial plane (Bacik *et al.*, 2017) where the copper atom is coordinated in a T-shaped arrangement (Hemsworth *et al*., 2015). Although conserved, there are some differences between the active site of LPMOs (Beeson *et al*., 2016), as illustrated for the active site of AA10s and AA9s in figure 1.4.

In AA9s, the copper ion is also coordinated by one water molecule found in the equatorial plane of the active site. Another water molecule occupies the axial position (Hemsworth *et*

al., 2015). The hydroxyl group of a conserved tyrosine residue is found in the buried axial position. This conserved tyrosine is also found in the active site of AA11s, AA13s, AA14s, and AA15s (Hemsworth *et al*., 2015; Couturier *et al*., 2018; Sabbadin *et al*., 2018). Further, AA9s also carries an unusual methylation on the τ-nitrogen of the N-terminal histidine. The methylation has been demonstrated to improve the enzymes resistance towards oxidative damage (Petrović *et al.*, 2018). However, non-methylated versions have been shown to function in near identical manner compared with the methylated version (Petrović *et al*., 2018).

In most AA10s, the conserved tyrosine residue found in the AA9 active site is replaced by a conserved phenylalanine residue. However, AA10s with a conserved tyrosine have also been described. AA10s also have a conserved alanine residue in the axial position of their active site. Interestingly, this alanine residue is also present in the active site of AA15s (Sabbadin *et al*., 2018). The alanine residue is thought, together with the phenylalanine, to block the axial positions of the copper ion offering a potential explanation to why no water molecule is found in the equatorial plane of the AA10 active site of AA10 (Hemsworth *et al.*, 2013).

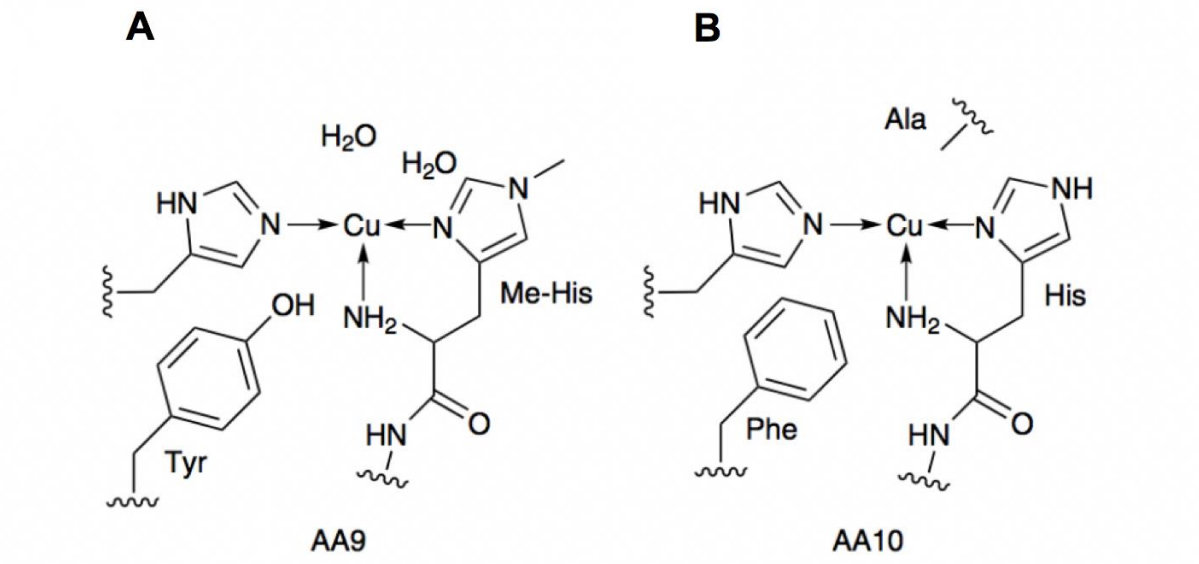

Figure 1.4: The structure of the LPMO active site. (A) The structure of the copper active site found in AA9s with its conserved tyrosine residue and the methylation on the τ-nitrogen of the N-terminal histidine. One water molecule occupies the equatorial position, and another the buried axial position. (B) The copper active site found in AA10s with the conserved phenylalanine and alanine residues. The figure has been adapted from Hemsworth *et al.*, 2013.

1.3.2 Tertiary structure of LPMOs

The overall fold of all LPMOs consists of a common immunoglobulin-like β-sandwich core structure made up of either 7 or 8 β-strands. Connecting the β-strands are a series of loops containing a variable number of short helixes. Variations in these loop segments are believed to give rise to much of the structural diversity between LPMOs (Loose, 2016; Vaaje-Kolstad *et al.*, 2017). To illustrate how these structural features are conserved both within- and between LPMO families, figure 1.5 shows a number of superimposed LPMO structures (Vaaje-Kolstad *et al.*, 2016). Figure 1.5 also shows the 3D-structure of *Jd*LPMO10A.

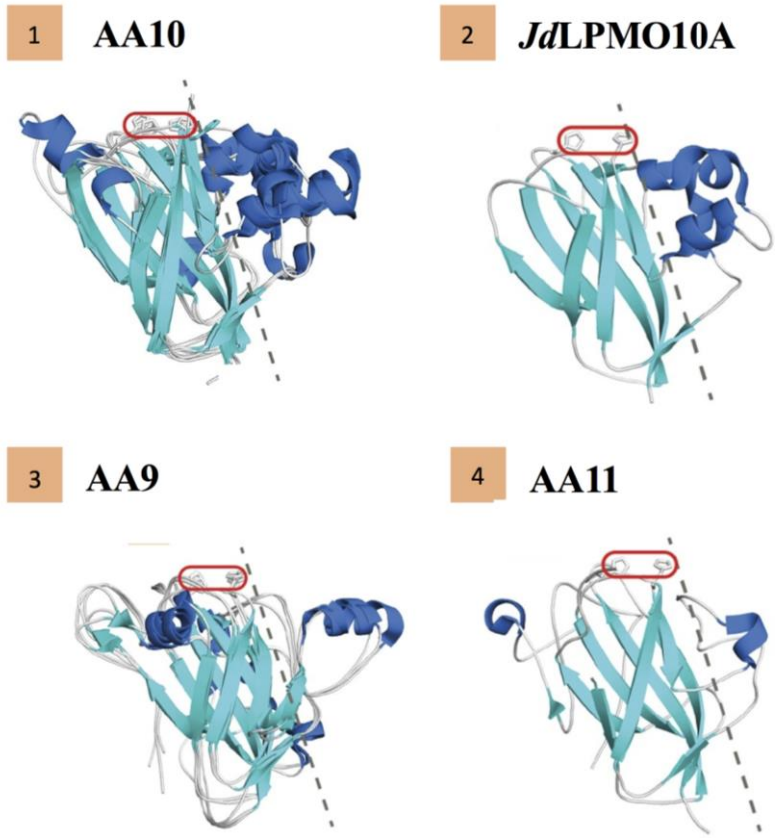

Figure 1.5: All LPMO structures have a common immunoglobulin-like β-sandwich core structure (light blue), and a flat binding surface that contains the copper-active site (Histidine residues are marked in red). The structural similarities become evident when LPMO structures are superimposed**.** **1)** Superimposed structures of: *Sm*LPMO10A (CBP21), *Vc*LPMO10B, *Ef*LPMO10A, and *Bp*LPMO10A **2)** *Jd*LPMO10A, **3)** Superimposed structures of: *Hj*LPMO9B, *Ta*LPMO9A, and *Nc*LPMO9M, **4)** *Ao*LPMO11A. The dashed line separates the immunoglobulin-like β-sandwich core from the areas known as "loop 2" region. The figure is been adapted from Vaaje-Kolstad *et al*., 2016.

Another conserved feature between LPMOs are the enzymes flat binding surface, which contains the copper active site. As shown in figure 1.5, the flat binding surface contains some of the loops connecting the β-strands in the core structure. The flat binding surface separates LPMOs from traditional GHs, which binds their substrate in a carbohydrate binding cleft (Beeson *et al*., 2016). An exception is found in the AA13s which display shallow carbohydrate binding cleft thought to bind amylose (Vaaje-Kolstad *et al.*, 2017).

The largest structural variations between LPMOs are found in an area dubbed the *loop 2* (*L2*) region, show in dark blue in figure 1.5. In AA10s the *L2* region is located between β-strands 1 and 3 in the β-sandwich core structure. A corresponding *L2* region has been described for AA9s, located between β-strands 1 and 2. The L2 region is believed to be involved in

substrate recognition and specificity as it comprises a substantial part of the substrate binding region and exhibits significant variations between different LPMOs (Vaaje-Kolstad *et al.*, 2017). Surface exposed aromatic residues found in the *L2* region, in particular, are believed to be linked with substrate binding and recognition. In addition, AA9s, AA13s, AA14s have variations in the L3 region located between β-strands 3 and 4 (Courtade *et al.*, 2016; Couturier *et al.*, 2018), and in the substrate binding surface opposite *L2* which includes the loop short (LS) and long C-terminal loop (LC) (Vaaje-Kolstad, *et al*., 2017). The *L2*, L3, LS and LC region is illustrated in figure 1.6.

A glutamate residue found about 5 Å from the active site is present in all chitin active LPMOs, and in C1 active AA10s acting on cellulose. In other LPMOs, the conserved glutamate is replaced by a conserved glutamine residue (Vaaje-Kolstad *et al*., 2017). This glutamate/glutamine residue does not appear to influence substrate specificity, but have been demonstrated to be essential for catalysis in *Sm*LPMO10A (Vaaje-Kolstad *et al*., 2005) and AA9s (Harris *et al.*, 2010). The glutamate/glutamine residue have been suggested to position reactive oxygen species-, or to be involved in H_2O_2 formation in the active site (Forsberg *et al.*, 2019). Interestingly, the Glutamine residue in *Sm*LPMO10A (Glu60) has been found to guard a narrow tunnel thought to connect the LPMO active site to the bulk solvent, while the enzyme is bound to its substrate (Bissaro *et al.*, 2018). Although the glutamate/glutamine appears in differ positions in the primary sequence of LPMOs, they are believed to play an important role in LPMO activity (Forsberg *et al*., 2019)

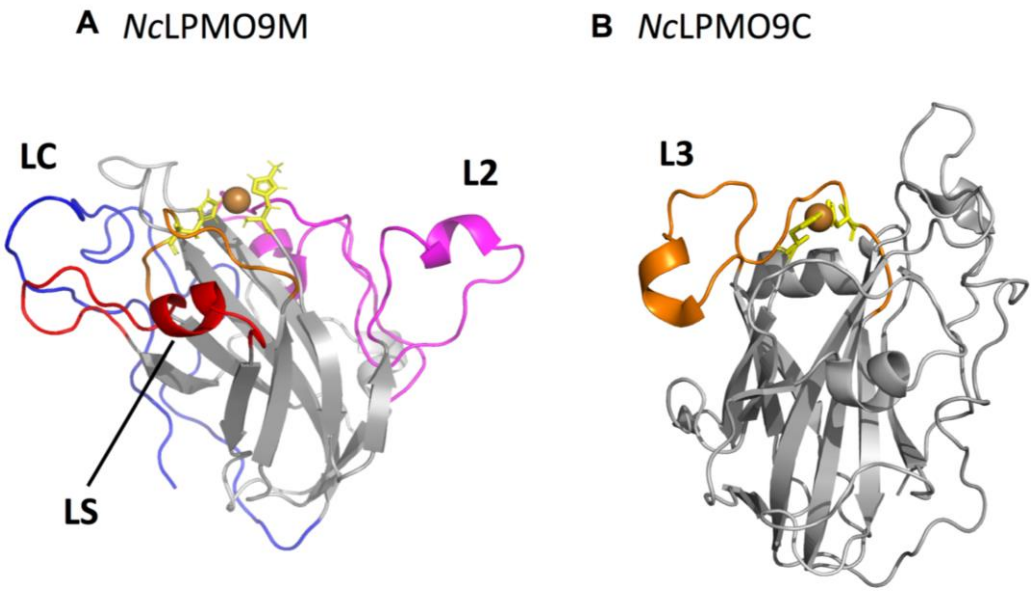

Figure 1.6: Structural diversity in LPMOs are often significant in the *L2*, L3, LS and LC regions. The conserved histidine residues constituting the active site are shown in yellow. A) The AA9 *Nc*LPMO9M (pdb 4eis), the *L2* region in colored in magenta, the LS region is colored in red, and the LC region is colored in blue. *Nc*LPMO9M does not have a L3 region, but the loop where the L3 region occur is colored in orange (Li *et al.*, 2012; Vaaje-Kolstad *et al.*, 2017). B) The AA9 *Nc*LPMO9C (pdb 4d7u), the L3 region is colored in orange (Borisova *et al.*, 2015). The figure was made by Christensen, I. A.

1.3.3 Structural basis for substrate binding and regioselectivity:

Interactions between LPMOs and their substrate can either be direct, or mediated by an associated Carbohydrate Binding Module (CMB). Studies have revealed several charged amino acids, found on the flat binding surface of LPMOs, to participate in substrate binding. This is in contrast to many other carbohydrate binding domains, that depend on interactions between aromatic side chains and the substrate through CH-π stacking (Vaaje-Kolstad *et al.*, 2004). The hydrophilic residues on the flat binding surface of LPMOs are on the other hand thought to interact with the substrate through hydrogen bonding (Loose, 2016).

Investigations by NMR have demonstrated that several residues in near proximity of the copper active site are involved in substrate binding (Aachman *et al*., 2012; Courtade *et al*., 2016). Early studies of *Sm*LPMO10A, using site-directed mutagenesis experiments, identified the residues Tyr54, Glu55, Glu60, His114, Asp182, and Asn185 as essential for substrate interaction (Vaaje-kolstad *et al*., 2010). Later, NMR-studies using a $^{2}H/^{1}H$ exchange experiment suggested that the surface exposed amino acids Tyr54, Glu55, Gln57, Ser58, and

Thr111 interacted strongly with β-chitin, while the residues Gln53, Leu110, Ala112, His114, and Thr116 interacted moderately with the substrate (Aachmann *et al*., 2012). Both experiments identified the only surface exposed aromatic residue (Tyr54) as important for surface binding. The other residues in question are a mixture of positively and negatively charged amino acids. A similar electrostatic profile has been found in other AA10s (Book *et al*., 2014). Using a combination of experimentally obtained data and molecular dynamics simulations, Bissaro *et al*. were in 2018, able to determine the interaction between *Sm*LPMO10A and crystalline β-chitin involving a spread of residues ranging from Tyr54 to Asp182, as shown in figure 1.7 (Bissaro *et al*., 2018).

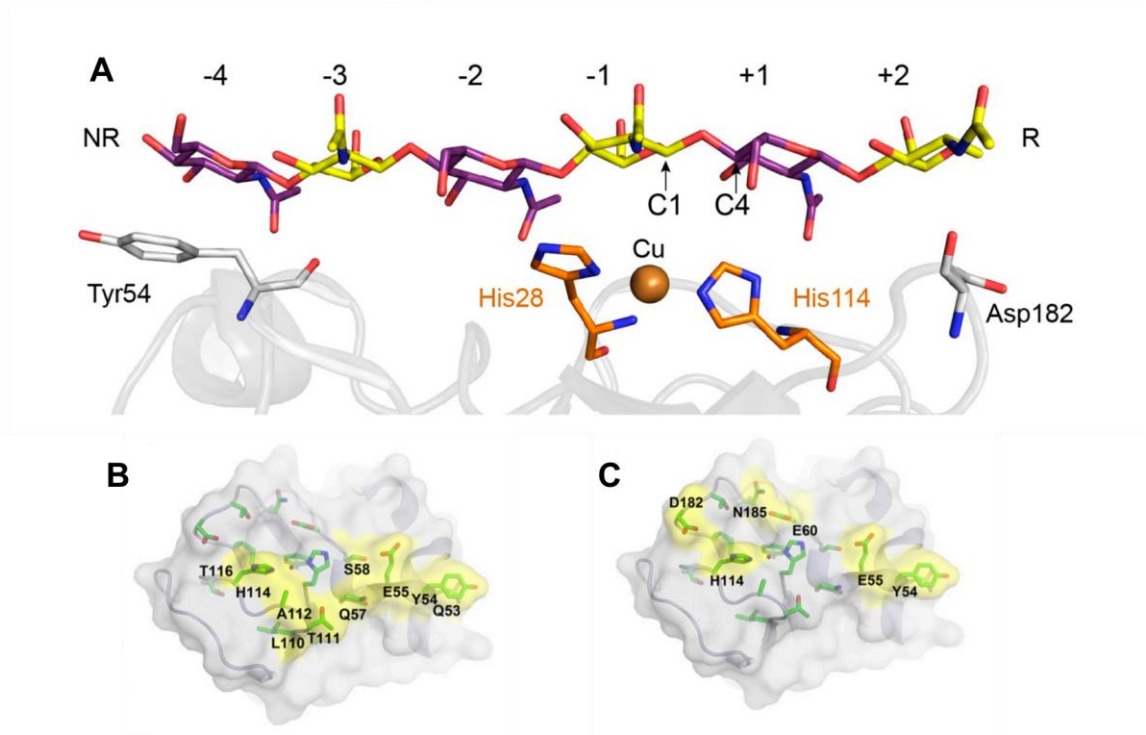

Figure 1.7: LPMO substrate interactions. **A)** Interactions between the LPMO SmLPMO10A and β-chitin the figure is collected from Bissaro *et al*., 2018. **B)** Residues on the flat binding surface of *Sm*LPMO10A found by NMR to be involved in substrate binding are marked in yellow, the figure is collected from Aachmann *et al*., 2012. **C)** Residues on the flat binding surface of *Sm*LPMO10A identified by site-directed mutagenesis experiments (Vaaje-kolstad, *et al*., 2005) are marked in yellow, the figure is collected from Aachmann *et al*., 2012.

Residues Gln53, Tyr54, Glu55, Gln57, and Ser58 are found in the *L2* region of *Sm*LPMO10A (Aachmann *et al.*, 2015). The *L2* region together with the LC region, is as previously described, believed to determine both substrate recognition and regioselectivity in AA10s (Wu *et al.*, 2013).

In AA9s, two additional loops, namely L3 and LS have been demonstrated to be important for substrate binding (Courtade *et al*., 2016; Wu *et al*., 2013). AA9s display some aromatic amino acids (Vaaje-Kolstad *et al.*, 2017) in both the LC and L2 region. The aromatic residues are thought to influence the positioning of the substrate relative to the active site, and the placement of the aromatic residues has been linked to the enzymes regioselectivity (Danneels, Tanghe and Desmet, 2019). LPMOs display different regioselective, oxidizing either the C1- or C4 position in the pyranose ring. Some LPMOs are able to oxidize both positions, generating so-called "double oxidation products". C1/C4-oxidizing LPMOs often carries N-glycans at their active site. The glycosylation is thought to play a role in regioselectivity. The assumption has to some degree been confirmed, as the amount of double-oxidize products have been found to decrease in response to site-directed mutagenesis of glycosylated residues (Danneels, Tanghe and Desmet, 2019). In addition, the regioselectivity of LPMOs can be altered by replacing their associate CBM with CBMs from other proteins (Crouch *et al.*, 2016).

1.4 Mechanism of chain scission

LPMOs were at the time of their discovery classified as monooxygenases found to utilize molecular oxygen (O_2) together with two externally delivered electrons to cleave the bonds of recalcitrant polysaccharides (Vaaje-Kolstad *et al*., 2010). When investigating the activity of CBP21 using isotopic labeling experiments, Vaaje-Kolstad *et al*. demonstrated that one of the oxygen atoms introduced into the oxidized chain by LPMOs originated from water, while the other originated from molecular oxygen (Vaaje-Kolstad *et al*., 2010).

In later studies, where the product profile of LPMO action was investigated by mass spectroscopy (MS), found the enzymes to display different regioselectivity (Quinlan *et al*, 2011). As illustrated in figure 1.8, LPMOs can either oxidize the C1- or C4-carbon in the pyranose ring (Quinlan, *et al*., 2011). The oxidation of C1 produces a soluble oligosaccharide with an δ-1,5 lactone, which in aqueous solutions spontaneously forms an aldonic acid, at the reducing end (Vaaje-Kolstad et al., 2010). The oxidation of C4 yields a ketoaldose, which exists in equilibrium with a germinal diol (hydrated ketoaldose), at the non-reducing end (Isaksen, T., et al., 2013). Moreover, some LPMOs have been found to oxidize both the C1 and C4 position, generating so-called "double oxidation products". Both AA9s and AA10s

have members able to oxidize either the C1- or C4-position, in addition to having members that generate double-oxidation products (Hemsworth *et al.*, 2015).

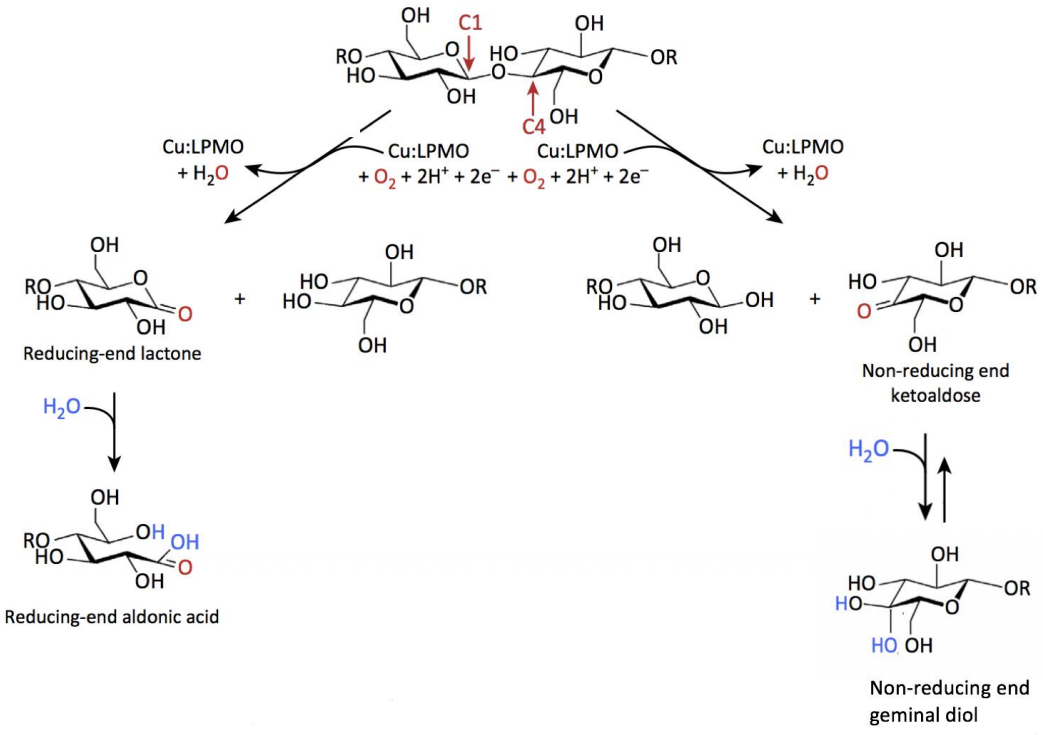

Figure 1.8: LPMOs are able to oxidize the C1- and C4-carbon in the pyranose ring. The oxidation of C1 results in an aldonic acid at the reducing end, while the oxidation of C4 gives a ketoaldose in equilibrium with a germinal diol at the non-reducing end (Quinlan *et al*, 2011; Isaksen *et al*, 2013). The figure has been adapted from Hemsworth *et al*. 2015.

As described in section 1.3.3, the observed differences in regioselectivity between LPMOs have, to some degree, been linked to phylogenetic variations in the substrate binding surface, and to glycosylation of residues on the flat substrate binding surface ((Book *et al.*, 2014; Hemsworth *et al.*, 2015; Danneels, Tanghe and Desmet, 2019).

There is currently no consensus regarding the putative reaction pathway used by LPMOs (Forsberg *et al.*, 2019). The absence of a consensus can in part be explained by the difficulties associated with investigating the interactions between LPMOs and their crystalline substrate, and in part by the possibility that different LPMOs might employ different reaction mechanisms (Ciano *et al.*, 2018). Several reaction mechanisms have been proposed, mainly based on experimental data obtained from other copper-dependent enzymes, revised in light of kinetic

studies, and computation simulations (Walton and Davies, 2016). Common to all of the putative pathways is the reduction of the copper active site from Cu(II) to Cu(I) by an external reductant (Beeson *et al*., 2016). The LPMO-Cu(I) is thought to activate oxygen, which is further reduced, producing a reactive oxygen species, that induces chain cleavage through an elimination reaction (Beeson *et al.*, 2012; Wang, Walton and Rovira, 2019).

Alternatively, H_2O_2 has been suggested to be the "true" co-substrate of LPMOs (Bissaro *et al.*, 2017; Forsberg *et al.*, 2019), making LPMOs peroxidases rather than monooxygenases. Bissaro *et al.* were able to demonstrate, not only that H_2O_2 can drive the LPMO reaction in absences of O_2, but that H_2O_2 seems to be the preferred co-substrate, even in conditions where the concentration of O_2 was 10-fold higher relative to the concentration of H_2O_2. Further, the addition of H_2O_2-scavenging enzymes has been demonstrated to inhibit LPMO activity under standard reaction conditions (Bissaro *et al*, 2017). The monooxygenase and peroxidase reaction of LPMOs is shown in figure 1.9.

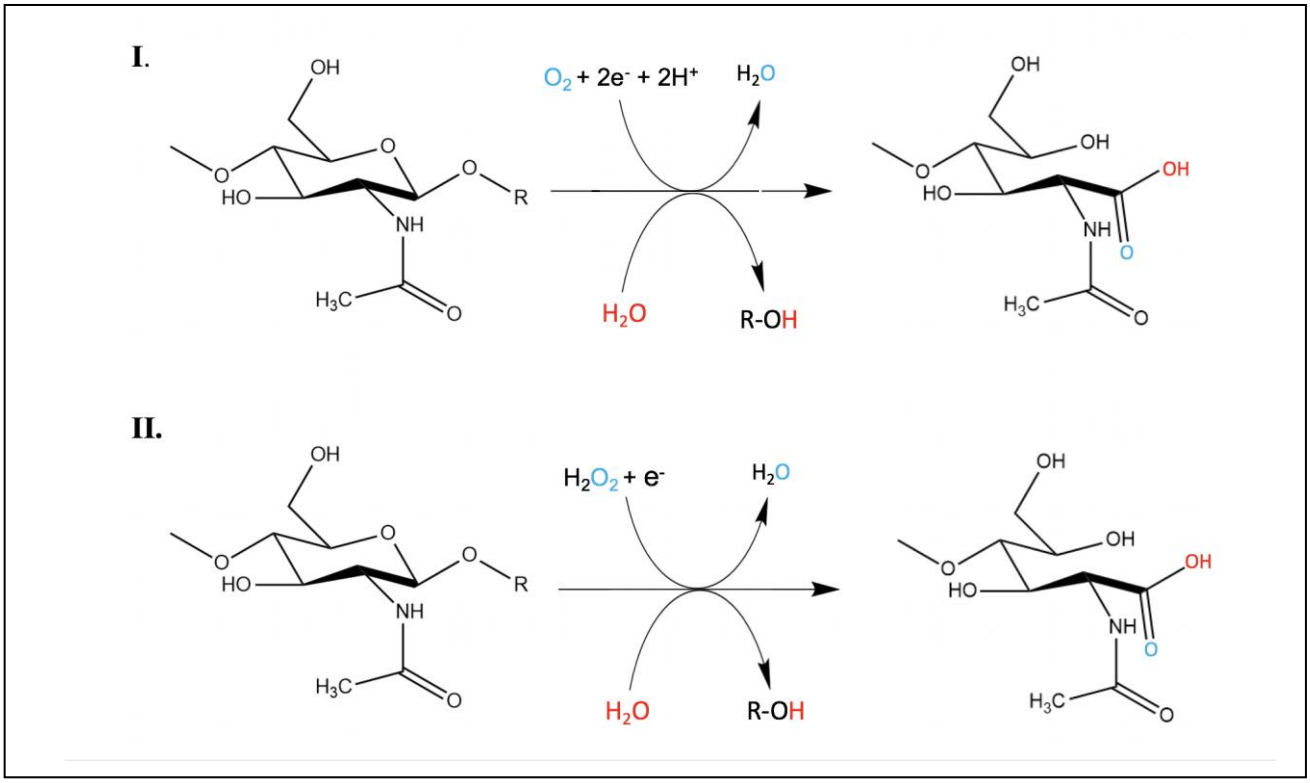

Figure 1.9: Overview of suggested mechanisms for LPMO enzyme reaction**. I)** Oxidative cleavage of chitin by a monooxygenase reaction. **II)** Oxidative cleavage of chitin by a peroxidase reaction. The figure was made by Christensen, I. A.

1.5 Electron donors:

As discussed in section 1.4, LPMO activity depends on the delivery of either two – or one externally provided election in order to perform a monooxygenase or peroxidase reaction, respectively (Vaaje-Kolstad *et al*, 2010; Bissaro *et al*, 2017). Studies have demonstrated that LPMOs are able to accept electrons from a wide array of reductants, including: ascorbic acid, reduced glutathione (Vaaje-Kolstad *et al.*, 2010)**,** gallic acid (R. J. Quinlan *et al.*, 2011)**,** lignin (Cannella *et al.*, 2012), and photocatalytic systems (Bissaro *et al*., 2017). In addition, studies have shown that LPMOs can be activated through interactions with other redox-enzymes. The activation of AA9s by the extracellular enzyme cellobiose dehydrogenase (CDH) in particular has been the subject of many studies (Beeson *et al*., 2016)

The first electron is believed to be delivered through direct contact between the external reductant and the copper-active site. Assuming LPMOs to be monooxygenases, it has yet to be uncovered how the second electron is delivered to the copper-active site. The delivery of the second electron presumably takes place after the LPMO is bound to its substrate, which

makes direct contact between the copper active site and the external reductant unlikely. The difficulty of understanding how the second elector in delivered is referred to as “the second electron conundrum”. Beeson *et al.* suggested that the second electron could be delivered through a potentially identified electron transfer pathway, with an associated CHD binding site (Beeson *et al*., 2016). However, NMR studies have demonstrated that CHD interacts directly with the copper active site in the AA9 *Nc*LPMO9A (Courtade *et al*., 2016), and the existence of such an electron transfer pathway has not been experimentally confirmed.

A tight tunnel has been identified in a model of *Sm*LPMO10A, connecting the active site to the bulk solvent while the enzyme is bound to its substrate. The tunnel is guarded by either a conserved glutamate or glutamine residue (Bissaro *et al.*, 2018). Although too narrow for larger reductants to pass, smaller molecules such as O_2, $O_2^{-\bullet}$, H_2O_2 and H_2O might be transported to the catalytic site through the tunnel (Bissaro *et al.*, 2018). Alternatively, the role of H_2O_2 as the preferred co-substrate for LPMOs offers a potential explanation of “the second electron conundrum”. In the case of a peroxidase reaction, only the first electron used to reduce the copper active site would be necessary for LPMO action (Forsberg *et al.*, 2019).

1.6 *Jd*LPMO10A

*Jd*LPMO10A is a small chitin active LPMO domain, located in the N-terminal end of the multimodular protein Jden1381 found in the Gram-negative bacterium *Jonesia denitirificans*. The function of Jden1381 has yet to be uncovered. However, *J.denitrificans* has been shown to express Jden1318 when grown on either cellulose or chitin (Mekasha *et al.*, 2016). The isolated domain of *Jd*LPMO10A has been shown to act on chitin, with a C1- mode of action, at a rate comparable to that of *Sm*LPMO10A (Mekasha *et al*., 2016).

The comparable enzyme activity is interesting, as *Jd*LPMO10A is an example of the small LPMOs found in clade I, subclade C, with a size of only 15.5 kDa (142 amino acids). For comparison, *Sm*LPMO10A has a size of 18.8 kDa. Its small size makes *Jd*LPMO10A compelling to study, as it could shed some insight into the “minimal scaffold” necessary for LPMO activity. The crystal structure of *Jd*LPMO10A is shown in figure 1.10 (Mekasha *et al.*, 2016). In addition, a neutron structure of *Jd*LPMO10A is available (Bacik *et al.*, 2017).

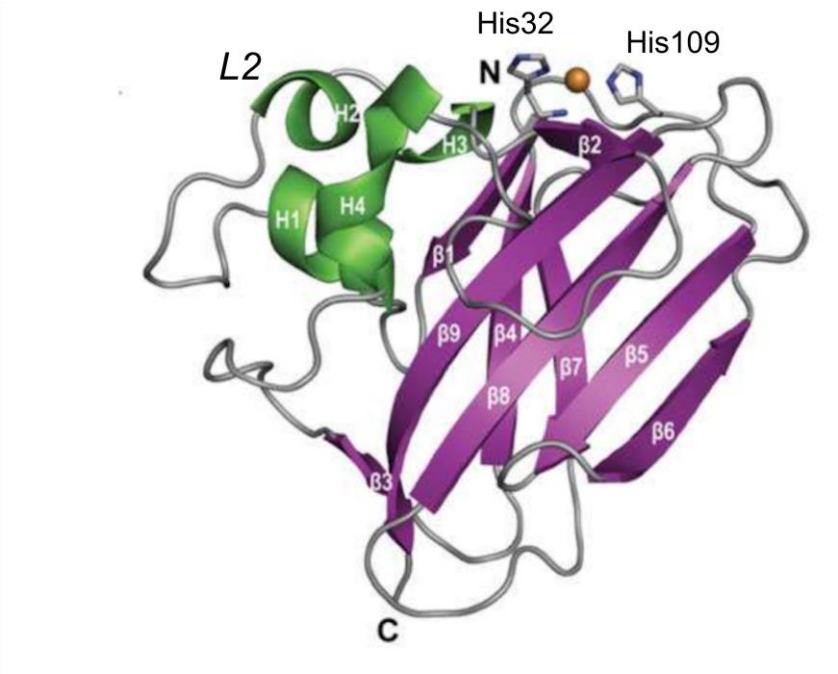

Figure 1.10: The 1.55 Å resolution structure of *Jd*LPMO10A (pdb: 5aa7) determined by Mekasha *et al.* using X-Ray crystallography. The structure shows an immunoglobulin-like β-sandwich made up of two β-sheets. Where one of the β-sheets are made up of β-strands β5, β6, β8 and β9, and the other of β-strands β1, β4 and β7. In addition, there a two more β-strands (β2 and β3) located on opposite sides of the β-sheet. Between β1 and β3, there is large protrusion made up of several loops and four helixes (H1-H4). The copper active site, with its T-shaped geometry, is located on the top of the enzyme. The structure also revealed two disulfide bridges between Cys45 and Cys54, and between Cys73 and Cys167. The figure has been adapted from Mekasha *et al.*, 2016.

The crystal structure reveals *Jd*LPMO10A to have the immunoglobulin-like β-sandwich core found in all LPMOs. The core of *Jd*LPMO10A is composed of seven β-strands organized into two β-sheets. The *L2* region, located between β-strands 1 and 3, was found to be made up of several loops, as well as four helixes. The T-shaped geometry of the "histidine brace" (His32 and His109) in the copper active-site was found to be consistent with findings from other LPMOs. Surface exposed amino acids found to influence substrate binding in *Jd*LPMO10A were in good agreement with findings from *Sm*LPMO10A. The sole difference between the two being that the conserved residue Asp182 in *Sm*LPMO10A was replaced with the residue Asn159 in *Jd*LPMO10A, as shown in figure 1.11 (Mekasha *et al.*, 2016).

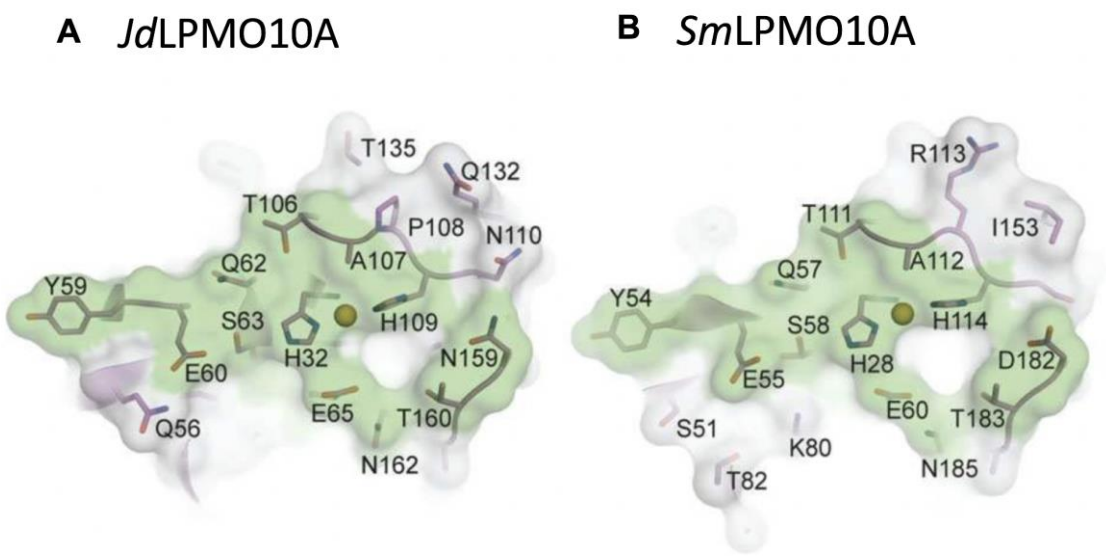

Figure 1.11: A comparison of the flat substrate binding surface of (A) *Jd*LPMO10A and (B) *Sm*LPMO10A found by Mekasha *et al.* (Mekasha *et al*, 2016). As with many chitin active LPMOs, the flat binding surfaces contains a mixture of both positively and negatively charged amino acids. The figure has been adapted from Mekasha *et al*., 2016.

All members of subclade C carries two major deletions compared to the sequence of larger AA10s. One is a deletion of 9 residues in the *L2* area, while the second is a deletion of 16 residues in the loop found between β-strand 5 and 6. A sequence alignment of *Jd*LPMO10A and *Sm*LPMO10A can be found in appendix A. The deleted regions were found to interact with each other in the structure of *Sm*LPMO10A, and the deletions could be therefore be correlated. Interestingly, neither of these deletions appear to have a significant effect on the enzyme activity of *Jd*LPMO10A (Bacik *et al.*, 2017). Further insight into why these two deletions does not appear to influence the activity of *Jd*LPMO10A would be interesting. NMR, being the method of choice for the dynamic study of proteins (Frueh, 2014), may provide further insight onto this minimal scaffold LPMO, and consequently, chitinolytic AA10s in general.

*Jd*LPMO10A is synthesized as a protein precursor consisting of the 142 amino acid residues found in the mature protein and an addition 31 residue long signal peptide. The 31 residues in the signal peptide will be included when referring to the primary structure of *Jd*LPMO10A in this thesis, meaning the N-terminal histidine will be denoted as His32, and so forth.

1.7 A brief mention of protein structure

Proteins can be briefly described as linear polymers, made up of simple repeating units, which assume three-dimensional structures in solution. This simplified explanation does, however, fail to illuminate the myriad of cellular functions these polymers perform. All proteins have a unique three-dimensional native fold, which enables them to carry out specific molecular functions. To study protein structure is an important step towards understanding its function. When describing the native fold of proteins, it is common to divide the structure into three layers of complexity called the primary-, secondary- and tertiary structure (Nelson, Cox, 2013).

1.7.1 Primary Structure:

The primary structure of a protein describes the amino acid sequence that makes up the polypeptide chain. A protein can be composed of any given combination of the 20 common amino acids, linked together by a peptide bond. The atoms surrounding the peptide bond are arranged in the sequence C_α-CO-NH-C_α. For a specific amino acid, the atoms are denoted as; N_i, C_i^α, and CO_i. Where *i* is the number of the residue in the amino acid sequence from the N-terminus end (Creighton, 1984).

The peptide bond has three defined dihedral angels called ω (omega), ϕ (phi) and ψ (psi). Together these dihedral angels describe the conformation of the peptide bond in a given segment of the polypeptide chain. In principle, free rotation about all three bonds is possible. However, the peptide bond is rigid due to its partial double bond-character which is the result of a resonance between the carbonyl (C=O) and the amide nitrogen (NH) bond. The peptide bond with the dihedral angels is shown in figure 1.12.

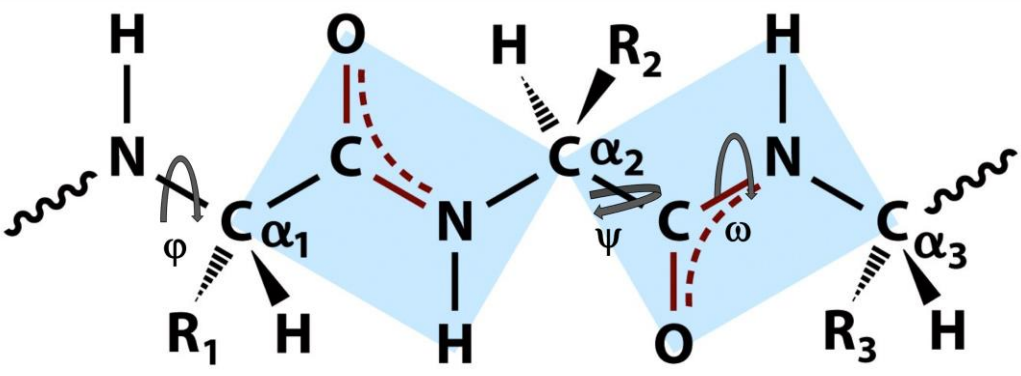

Figure 1.12: The peptide bond. All peptide bonds have some double bond-character due to resonance and do therefore have a limited range of conformations. Three dihedral angels define the conformation of the peptide. ϕ describes the rotation about N-C_α bond, ψ about the C_α-CO bond and ω about the C-N bond (where rotation is constrained). The double bond characteristic results in ϕ and ψ having a maximum distance of 180° (or - 180°) when the peptide backbone is fully extended. However, many of the values between 180° and - 180° are not possible because of steric hindrance. The figure has been modified from Sandwalk, 2008.

1.7.2 Secondary structure:

The secondary structure of a protein describes local arrangements in segments of the peptide chain were ϕ and ψ assume regular values. Two of the most stable secondary structures are α-helixes and β-sheets.

The α-helix is a right-handed coil with the amino acid side-chains protruding out from the structure. Each turn of the helix is 5.4 angstrom (Å), which corresponds to 3.6 amino acid residues. The structure is stabilized by hydrogen bonds between the amide proton and the oxygen of the carbonyl group of the backbone throughout the entire segment, so a total of four hydrogen bonds supports each turn of the helix. The typical dihedral angels found in the α-helix are -60° for ϕ and -50° for ψ (Nelson., Cox., 2013; Tramontano, 2006).

β-sheets are composed of two or more β-strands, linked together by hydrogen bonds between the carbonyl group and the amine proton in the polypeptide backbone (Tramontano, 2006). β-strands are extended conformations with a zigzag like character, where the side-chains of neighboring amino acid residues are positioned in the opposite direction, forming a layered structure (Tramontano, 2006). The β-strands can either lie in the same direction (parallel sheets) or the opposite direction (antiparallel sheets). Typical dihedral angles found in β-

sheets are -140° for ϕ and 130° for ψ (Tramontano, 2006). The structural features of both the α-helix and β-sheet are shown in figure 1.13.

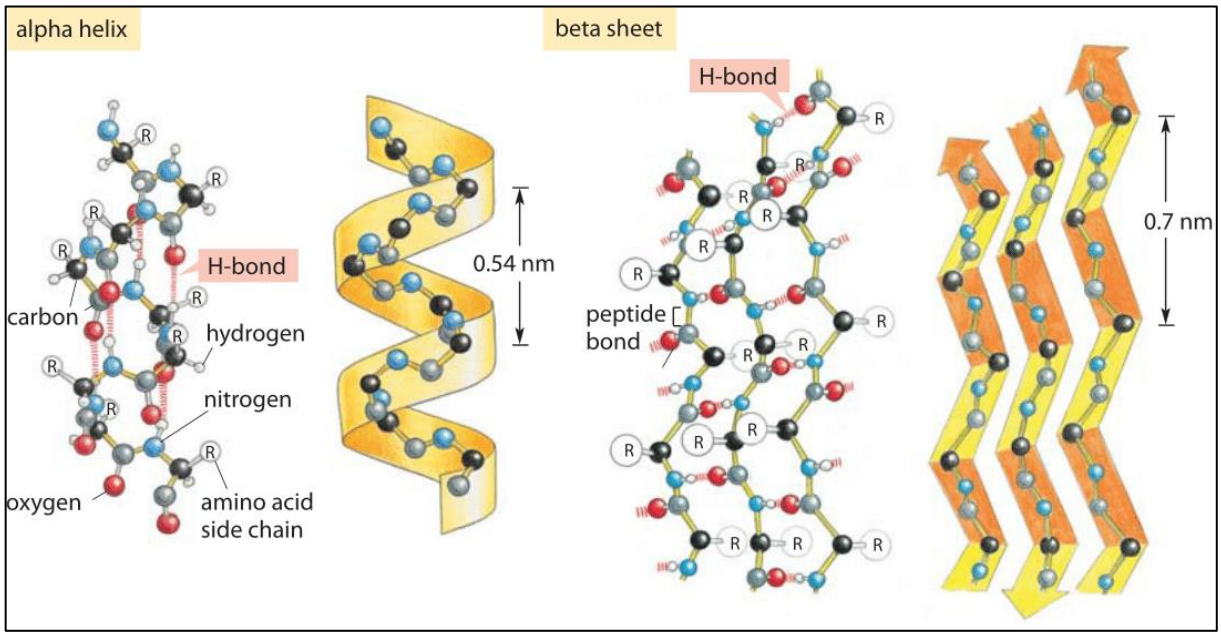

Figure 1.13: Two of the most common secondary structures are the α-helix and β-strand. The α-helix is a right-handed coil stabilized by intramolecular hydrogen bonds. The β-strand is composed of several β-sheet running organized in either a parallel or anti-parallel fashion, stabilized by intermolecular hydrogen bonds. The figure has been collected from cell biology of the numbers, no date.

Another reoccurring secondary structure is the β-turn which causes the polypeptide chain to change direction. β-turns are characterized as a region of four constitutive amino acids which are often stabilized by a hydrogen bond between the carbonyl oxygen of the first residue and the amine group of the fourth residue in the structure. Several types of β-turns exist (Nelson and Cox, 2013) When no regular structure can be found, the segment is referred to as a random coil (Nelson and Cox, 2013). Segments without local arrangements that connects two secondary structure elements are called loops (Tramontano, 2006).

1.7.3 Tertiary structure:

The tertiary structure of a protein describes the three-dimensional shape the protein assumes in solution. Consequently, the tertiary structure describes the association of all the secondary structure elements of the protein (Tramontano, 2006). Some combinations of secondary structures are common in many proteins, despite large variations in the primary structure of the proteins in question. These types of units are called “motifs” or supersecondary structures.

One example of a motif is the alpha-alpha unit, where two anti-parallel α-helixes are joined by a turn. Combinations of motifs are again referred to as protein domains or folds. A domain can be further defined as a semi-compartment of the polypeptide chain that interacts more with itself than other parts of the protein structure (Tramontano, 2006). Functional domains are often highly conserved, and can, therefore, indicate the evolutional relationship between related proteins. Larger proteins are often composed of several domains, such proteins are sometimes referred to as "mosaic-proteins" or multimodular proteins, while the domains are called "modules" (Tramontano, 2006).

1.8 NMR in protein structure elucidation

The tertiary structure of proteins has traditionally been determined by either X-ray crystallography or Nuclear Magnetic Resonance Spectroscopy (NMR) (Tramontano, 2006). The principle of NMR can be briefly explained as the detection of electrical currents produced by odd mass numbered nuclei, placed in an external magnetic field, after being subjected to a high-frequency pulse (Friebolin, 2011). The primary information provided by NMR is the chemical shift (δ), which is the resonance of a given nucleus relative to the external magnetic field, and couplings to neighboring nuclei. Couplings arise when the magnetic moment of one nucleus interacts with the magnetic moment of neighboring nuclei, influencing their chemical shifts. Couplings can occur either directly through bonds (scalar couplings) or indirectly through space (dipolar couplings, also called NOEs) (Friebolin, 2011). The simplest NMR experiments are one-dimensional, where the signals from one type of nuclei are observed. Correlations between nuclei are observed by expanding the experiments into two- or more dimensions (Friebolin, 2011).

The advantage of using NMR compared to X-ray crystallography is the ability to observe proteins in solution, near physiological conditions. NMR can therefore be used to obtain a more dynamic model of the protein in question, and is ideal for studying thermodynamic and kinetic aspects of proteins (Frueh, 2014). The main drawback of using NMR is the techniques inherently low sensitivity. This holds especially true for the NMR sensitive nuclei ^{13}C and ^{15}N, which has both small values for their gyromagnetic ratio (γ) and a low natural abundance (Friebolin, 2011). The advent of protein NMR was therefore focused on the so-called homonuclear method, where scalar couplings between ^{1}H-pairs were used to elucidate the proteins secondary- and tertiary structure. However, the homonuclear method is not well

suited to assign proteins >10 kDa, because spectra usually become too crowded to be useful for structural investigations when this threshold is exceeded (Wuthrich, 1990).

The challenge related to the sensitivity of NMR can be circumvented by using $^{15}N/^{13}C$-isotopically enriched protein samples, together with a combination of two- and three-dimensional heteronuclear experiments (Wuthrich, 1990). The chemical shifts of ^{13}C and ^{15}N are generally more dispersed compared with the chemical shifts of ^{1}H, resulting in fewer overlapping peaks (Edwards and Reid, 2000). Isotopically enriched proteins are often produced by over-expressing the protein in a suitable host cultivated in minimal medium with isotopically labeled feedstock. The emergence of triple resonance experiments in combination with isotopically enriched protein samples has made NMR one of the forefronts within structural investigations of proteins (Edwards and Reid, 2000).

The primary structure of the protein must be known prior to structural determination by NMR. The first step in the structural determination process is to record the desired NMR spectra of the $^{15}N/^{13}C$- labeled protein. Once the spectra have been recorded, the backbone of the protein is mapped by NMR, using the sequence-specific resonance assignment method (Quincy, 2005). The sequence-specific resonance assignment begins with correlating the NMR signals arising from the individual amino acids in the primary sequence into spin systems, using a combination of the ^{15}N-HSQC and triple resonance experiments. Once correlated, neighboring amino acids are assembled into fragments. Finally, the fragments are matched with segments of the protein sequence (Edwards and Reid, 2000; Quincy, 2005).

With the protein backbone completely mapped by NMR, it is possible to: (1) classify elements of the proteins secondary structure, (2) interpret changes in the chemical environment of the protein upon ligand binding, (3) investigate protein dynamics in solution, and (4) structural determination of the protein. After the primary structure of the protein has been mapped out, the secondary- and tertiary structure of the protein are identified by determining the position of the residues relative to each other. The distance between residues in the protein structure can be determined by dipolar couplings, also called NOEs. NOEs are only observed between nuclei that are close in space and provides information about the overall fold of the protein. In addition, characteristic NOEs are found between residues in secondary structural elements such as α-helix and β-sheet (Cavanagh J, 2006; Edwards and Reid, 2000), making it possible to determine the secondary structure in segments of the

polypeptide chain. The final step in the structural determination process is to use computer software to calculate the three-dimensional structure of the protein. The calculation is based on structural restrain information obtained by NMR data, in combination with covalent restrains (bond lengths etc.) and atomic properties (atomic radius, charge etc.) (Kwan *et al.*, 2011). The general method for protein structural determination by NMR, with an emphasis on the sequence-specific resonance assignment of the peptide backbone is shown in figure 1.14.

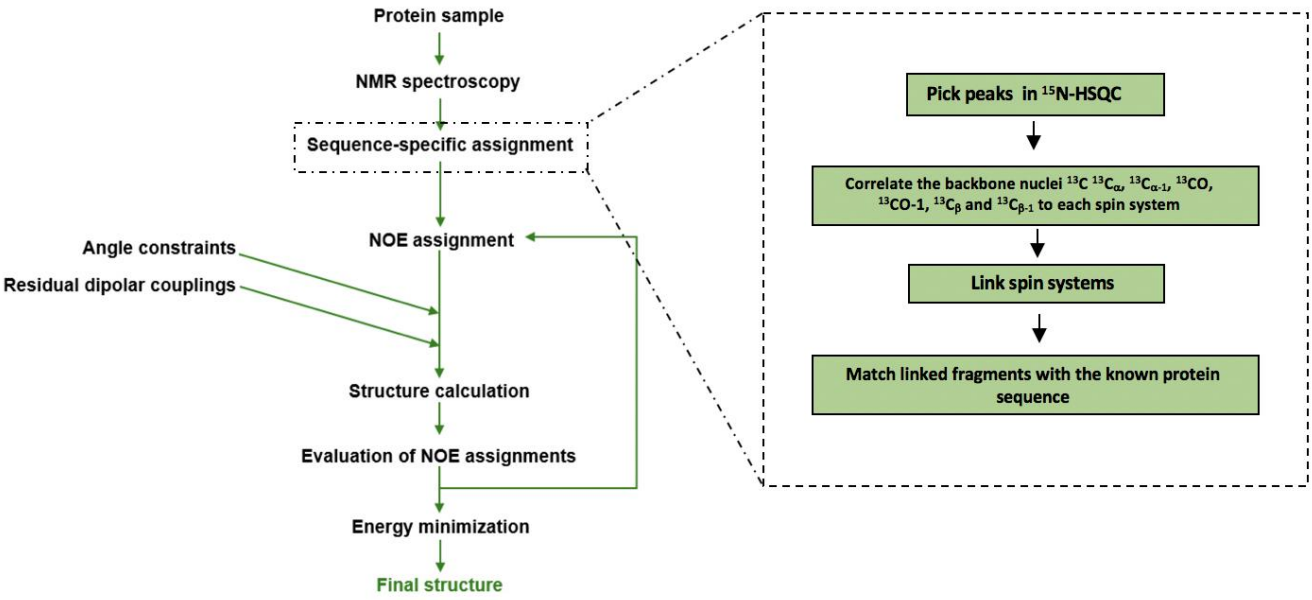

Figure 1.14: Overview of the general method for protein structural determination by NMR, with an emphasis on the sequence-specific resonance assignment of the protein backbone (Keller, 2004). The figure has been adapted from Aachmann, 2015 and Courtade, 2018.

The following sections will give a brief overview of the NMR-experiments used for the sequence-specific assignment of the proteins using ^{15}N- and ^{15}N/^{13}C- labeled protein samples.

1.8.1 Two-dimensional NMR experiments

Two-dimensional NMR experiments are frequently used to correlate nuclei. To understand the basis of a two-dimensional NMR experiment, the basis of one-dimensional NMR experiments must first be described.

The basic one-dimensional NMR-spectrum is generated by the two-step process of preparation and detection. In the preparation phase, the sample is irradiated with a single radiofrequency pulse and is immediately followed by the detection phase where the Free Induction Decay (FID) of the sample is recorded. The one-dimensional NMR spectrum is then obtained by Fourier transformation of the FID in a single time domain (Friebolin, 2011). Although excellent for structural elucidation of smaller organic molecules, the one-dimensional NMR spectrum of proteins is usually too complex to provide much structural information (Friebolin, 2011). The complexity of one-dimensional ^{1}H-NMR spectra of proteins is illustrated in figure 1.15, comparing the ^{1H}-spectrum of toluene to the ^{1}H-spectrum of egg white lysozyme.

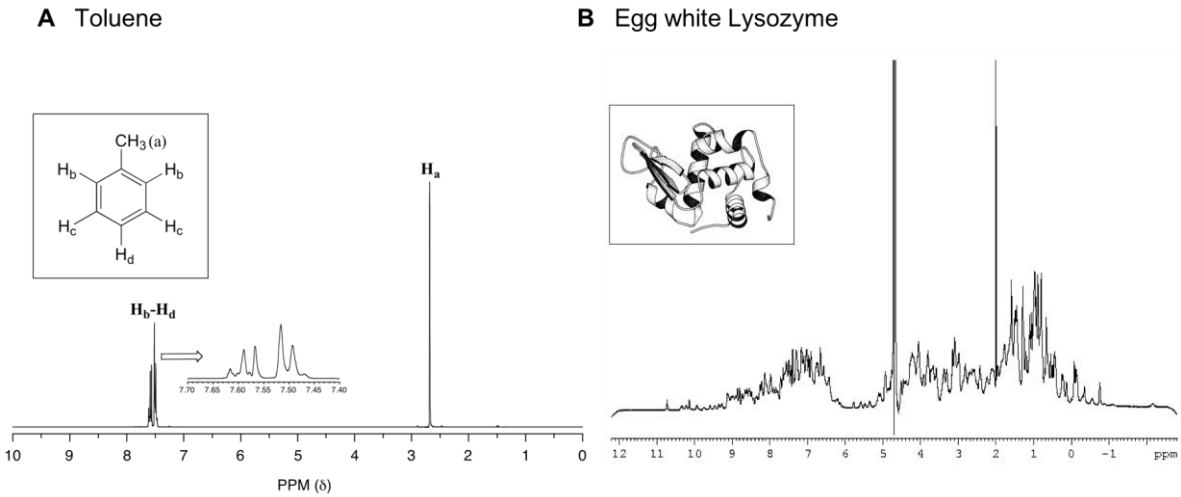

Figure 1.15: comparison of one-dimensional NMR spectra. **A)** The one-dimensional ^{1}H-NMR spectrum of toluene. **B)** The one-dimensional ^{1}H-NMR spectrum of egg white lysozyme. The figure was made by Christensen, I. A.

The problem with complexity can be overcome by using two-dimensional NMR experiments. In two-dimensional NMR experiments, a second dimension is obtained by Fourier transformation of two time domains. The second time domain is attained by introducing two new phases, *evolution* (t_1) and *mixing*, prior to the detection phase (t_2) (Friebolin., 2011). The length of the mixing period is systematically increased by an increment of $n\Delta t$, so that a series of n FIDs are obtained by the end of the experiment (Edwards and Reid, 2000). The Fourier transformation of t_2 followed by a second Fourier transformation of t_1 gives a spectrum that

displays the scalar-coupling between neighboring NMR sensitive nuclei in the protein (Friebolin, 2011).

1.8.1.1 ^{15}N-Heteronuclear Single-Quantum Coherence (^{15}N-HSQC)

The ^{15}N-Heteronuclear Single-Quantum Coherence experiment (^{15}N-HSQC) experiment detects couplings between ^{1}H-^{15}N pairs in protein (Friebolin., 2011). The signals from ^{1}H-^{15}N pairs play a central part in the sequence-specific resonance assignment of the peptide backbone. All of the residues (with the exception of proline) in primary structure features a ^{1}H-^{15}N pair, as shown in figure 1.16. The proton of the N-terminal end is rapidly exchanged with the solvent and is in most cases undetectable. ^{15}N-HSQC therefore functions almost like "a fingerprint" of a protein. Each backbone ^{1}H-^{15}N pair functions as the starting point for assigning all the atoms in the peptide backbone by NMR (Frueh, 2014).

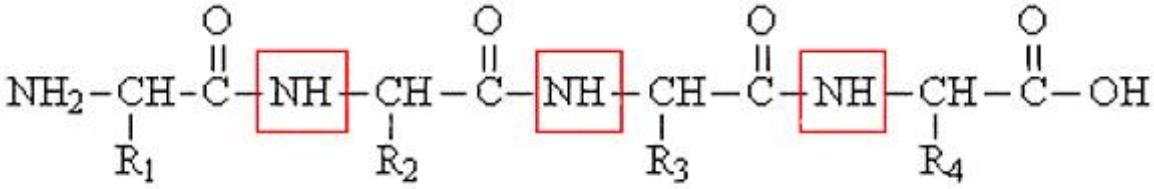

Figure 1.16: The peptide backbone with amide protons and nitrogen pairs marked in red. Sidechains are denoted as R-groups. The figure was made by Christensen, I. A.

It should be noted that that the ^{15}N-HSQC also displays individual signals for the ^{1}H-^{15}N pairs found in the side-chain of tryptophan, asparagine, glutamine, lysine, arginine and histidine residues (Kwan *et al.*, 2011). It is important to distinguish these side-chain peaks form the ones arising from the peptides backbone, as demonstrated in section 3.2.10.

1.8.2 Three-dimensional NMR experiments

Two dimensional NMR experiments can be expanded into three or more dimensions, making it possible to observe a number of nuclei simultaneously (Edwards and Reid, 2000). When working with ^{15}N/^{13}C-isotopically enriched protein samples, three-dimensional NMR experiments are frequently used, as they provide increased resolution compared with two-dimensional experiments (Cavanagh, 2007).

1.8.2.1 Heteronuclear Triple resonance-experiments

Heteronuclear Triple resonance-experiments are a group of experiments used to correlate backbone signals to nearby ^{15}N-^{1}H pairs in protein backbone through scalar couplings (Quincy., 2005; Edwards and Reid, 2000). A number of heteronuclear triple resonance experiments have been developed to assign the signals from backbone atoms. Common to all experiments is a transfer of magnetization from ^{1}H-^{15}N-pairs in the protein backbone, to other backbone atoms. The detected backbone atoms are either associated with the residue in question (*i*), or its sequential precursor (*i*-1) (Edwards and Reid, 2000). The names of the described triple resonance-experiments reflect the backbone atoms detected by the experiment (Quincy., 2005; Edwards and Reid, 2000).

HNCA and HN(CO)CA:

The HNCA and HN(CO)CA experiments are used to identify $^{13}C\alpha$ and $^{13}C_{\alpha\text{-}1}$ respectively. The correlations observed, as well as the pathways of the magnetization transfers, are shown in figure 1.17.

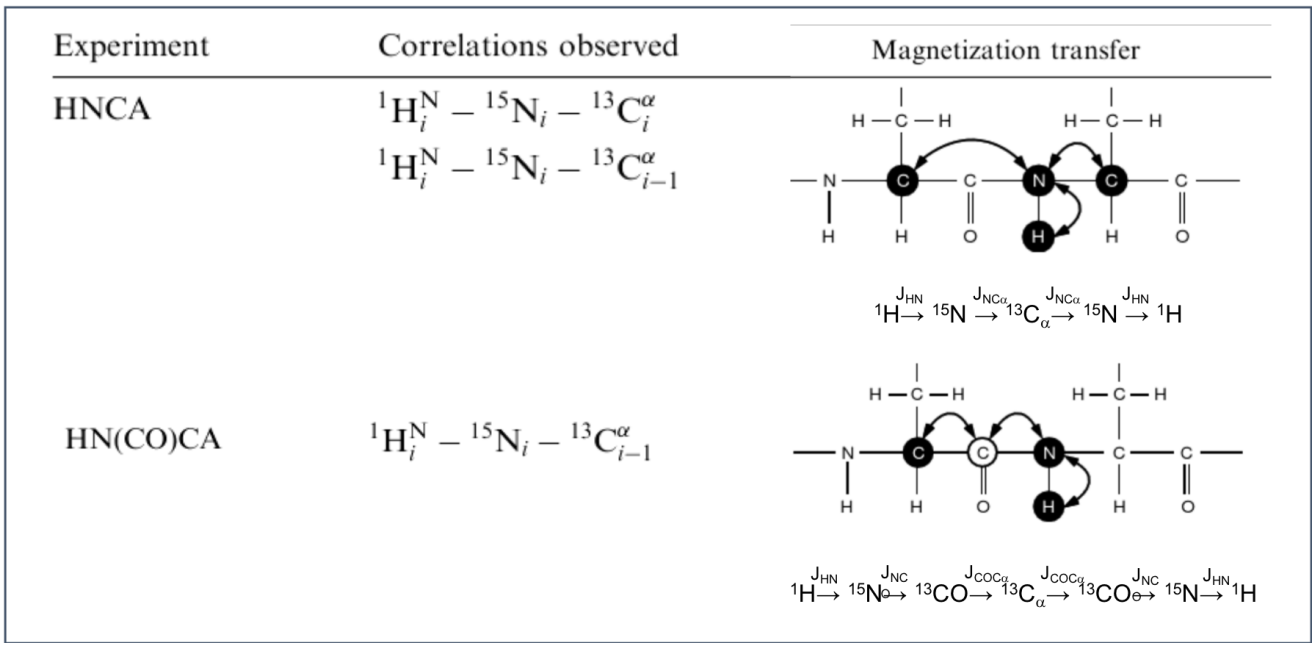

Figure 1.17: In the HN(CO)CA experiment, the magnetization is transferred from ^{1}H to ^{15}N, through ^{13}CO and then lastly to $^{13}C_{\alpha\text{-}1}$, before being transferred back to ^{1}H for detection. The result is one peak which corresponds to $^{13}C_{\alpha\text{-}1}$. In HNCA the magnetization is ultimately transferred to both $^{13}C_{\alpha\text{-}1}$ and $^{13}C_{\alpha}$. Generally, the coupling between N_i and $^{13}C_{\alpha i}$ is greater than the coupling between N_i and $^{13}C_{\alpha i\text{-}1}$.

The $^{13}C_{\alpha i}$ peak is therefore often observed as the most intense of the two. The figure has been adapted from Cavanagh, 2007.

HNCO and HN(CA)CO:

The HNCO and HN(CA)CO experiments are used to identify ^{13}C-1 and ^{13}C respectively. The correlations observed, as well as the pathways of magnetization transfers, are shown in figure 1.18.

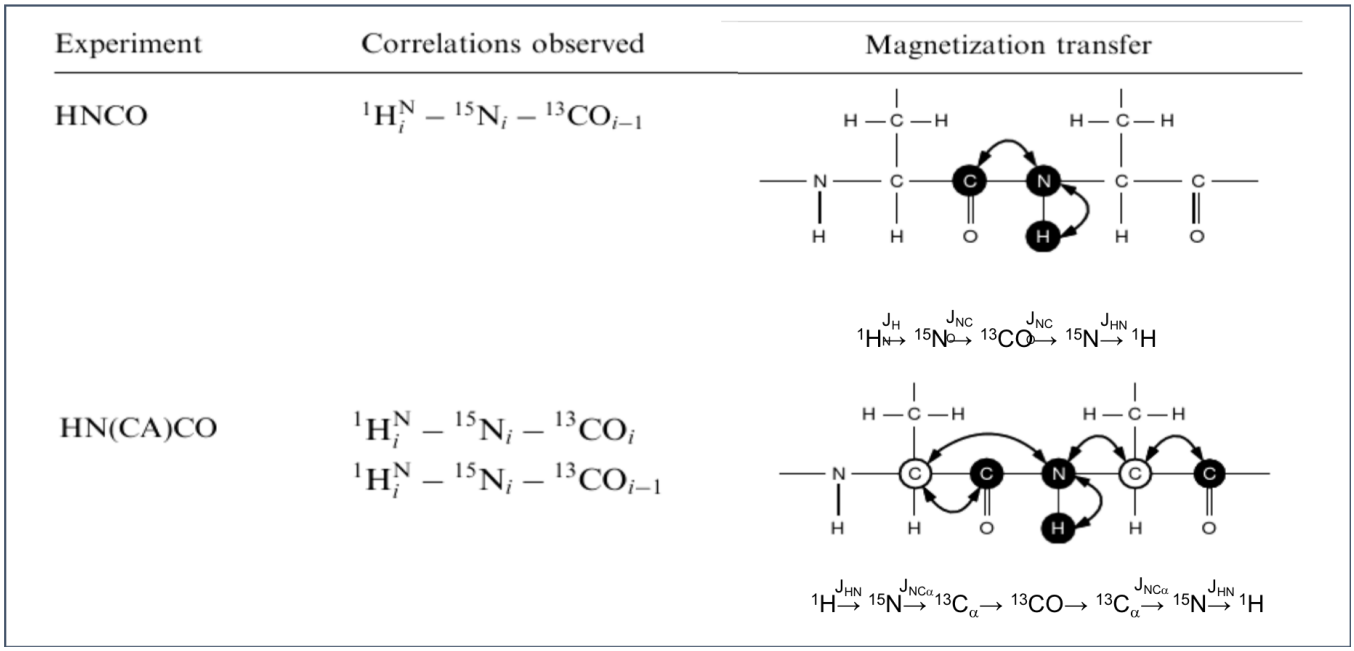

Figure 1.18: In HNCO, the magnetization is transferred from ^{1}H to ^{15}N and then selectively to ^{13}CO-1. In HN(CA)CO, the magnetization is transferred from ^{1}H to ^{15}N and then further to ^{13}Cα to both ^{13}CO and ^{13}CO-1. The magnetization is then transferred back to ^{1}H through the same pathway. Generally, the coupling between N_i and $^{13}CO_i$ is greater than the coupling between N_i and $^{13}CO_{i-1}$. The $^{13}CO_i$ peak is therefore often observed as the most intense of the two. The figure has been adapted from Cavanagh, 2007.

HNCACB and HN(CO)CACB:

The HNCACB and HN(CO)CACB experiments are used to identify C_α and C_β of both the amino acid in question (*i*) and its sequential precursor (*i*-1). The correlations observed, as well as the pathways of the magnetization transfers, are shown in figure 1.19.

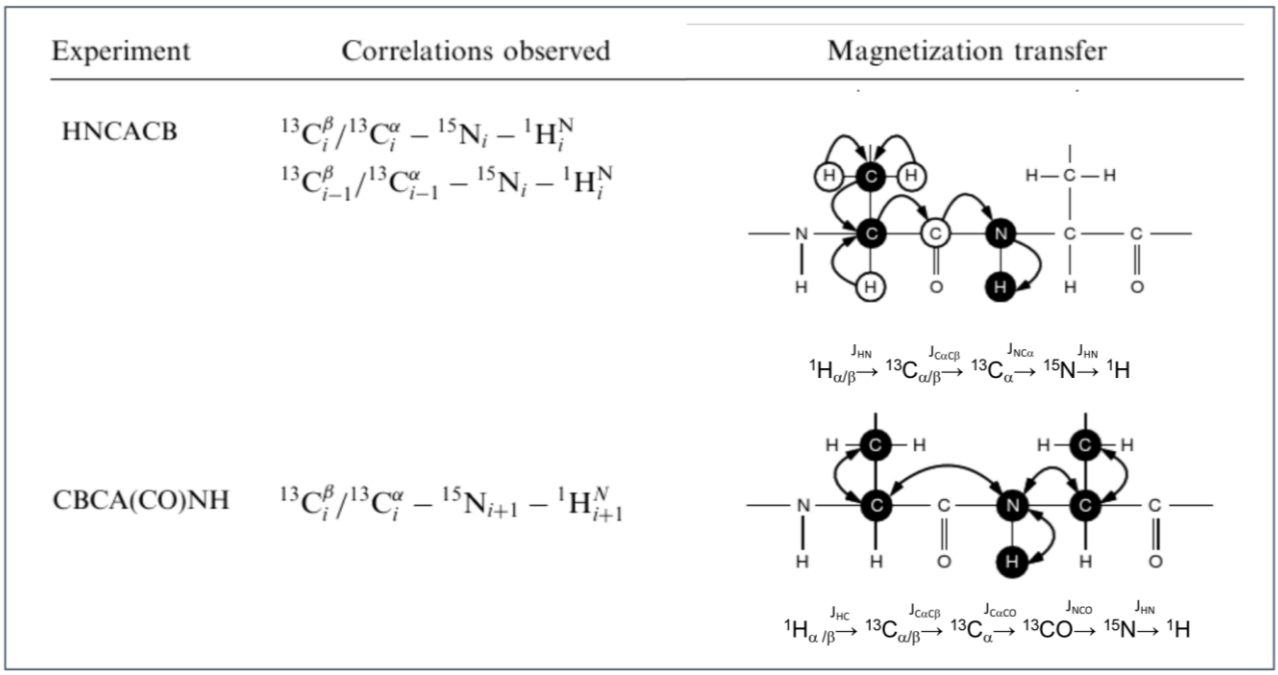

Figure 1.19: In HN(CO)CACB, the magnetization is transferred from $^{1}H_{\beta}$ to $^{13}C_{\beta}$ and further to $^{13}C_{\alpha}$, magnetization is also transferred from $^{1}H_{\alpha}$ to $^{13}C_{\alpha}$. From $^{13}C_{\alpha}$ through ^{13}CO to ^{15}N and finally to ^{1}H. In HNCACB, the magnetization is transferred by the same pathway as in HN(CO)CACB, but in addition, magnetization is transferred from $^{1}H_{\beta}$ to $^{13}C_{\beta}$ to $^{13}C_{\alpha}$, from $^{1}H_{\alpha}$ to $^{13}C_{\alpha}$, and finally to ^{15}N and ^{1}H. Generally, the coupling between N_i , $^{13}C_{\alpha I}$ and $^{13}C_{\beta}$ is greater than the coupling between N_i, and $^{13}C_{\alpha i-1}$ and $^{13}C_{\beta-1}$. The $^{13}C_{\alpha i}$ and $^{13}C_{\beta}$ peaks are therefore often observed as the more intense. The figure has been adapted from Cavanagh, 2007.

1.8.2.2 ^{1}H /^{15}N NOESY-HSQC

The Nuclear Overhauser Effect Spectroscopy (NOESY)-experiment is used to correlate nuclei that couple through space, providing information about the overall fold of the protein. NOEs are the results of mutual interactions between the magnetic moment spins that are closer than 5 Å in space. During relaxation, dipolar couplings might lead to an increase in the relaxation rate of the involved nuclei, a process known as the Nuclear Overhauser Effect (NOE). The NOESY experiment is useful for both investigations of protein secondary- and tertiary structure, as the experiment provides information about the spatial proximity of protons (Edwards and Reid, 2000).

The homonuclear NOSEY experiment can be combined with two-dimensional heteronuclear experiments to create triple resonance experiments such as the $^{1}H/^{15}N$ NOESY-HSQC (Edwards and Reid, 2000). In cases where the 2D NOESY spectrum of proteins is severely overlapped, the $^{1}H/^{15}N$ NOESY-HSQC disperse the signals because only correlations to

amide protons (NH) are visible in the spectrum. The $^{1}H/^{15}N$ NOESY-HSQC experiment can be view as a cube, with two ^{1}H and one ^{15}N frequency axes. To simplify matters, the cube can be divided into slices of two-dimensional spectrums, displaying ^{1}H-^{1}H NOESY information (Edwards and Reid, 2000), as shown in figure 1.20.

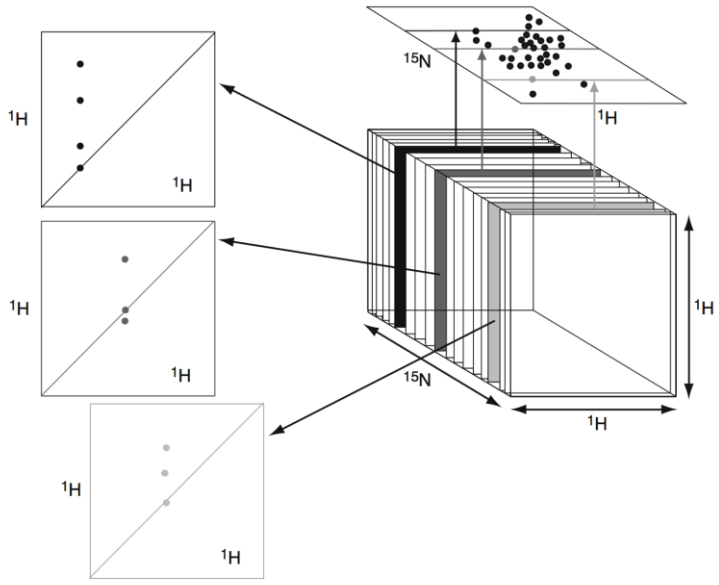

Figure 1.20: A schematic presentation of the $^{1}H/^{15}N$ NOSEY HSQC experiment. The three-dimensional experiments can be thought of as a cube, where each slice of the cube displays the ^{1}H-^{1}H NOSEY of amide protons with a specific chemical shift. The figure has been adapted from Edwards and Reid, 2000.

1.9 Prediction of secondary structural elements and dihedral angels using the computational method TALOS-N

The traditional chemical shift assignment protocol for proteins by NMR typically consists of two steps: (1) sequence-specific assignment of the polypeptide backbone atoms, and (2) assignment of side-chains. NOEs are then used to calculate the protein structure as described in section 1.8 (Shen and Bax, 2013). However, the increasing number of protein structures deposited in the Protein Data Bank (PDB), in addition to the chemical shift assignments found in the BioMagResBank (BMRB), has led to the development of computational methods to study the relationship between protein structure and chemical shifts (Shen and Bax, 2013). One such computational method is TALOS-N, the successor of the widely used programs TALOS (Cornilescu, Delaglio and Bax, 1999) and TALOS+ (Shen *et al.*, 2009). TALOS-N

predicts the backbone dihedral angels ϕ/ψ based on the assigned chemical shifts: $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, ^{15}N, $^{1}H^{N}$ and $^{1}H_{\alpha}$ by empirical comparison with tripeptides collected form protein deposited in the PDB with the help of an artificial neural network (ANN) (Shen and Bax, 2013). TALOS-N have been reported to predict the backbone torsion angles for ≥90% of the residues in the protein, with an error rate of less than 3.5% (Shen and Bax, 2013). TALOS-N is also able to predict secondary structural elements (helix, sheet, and coil) based on the assigned chemical shifts and the sequence of the protein in question. The accuracy of the prediction is about 88% when the chemical shifts $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, ^{15}N, $^{1}H^{N}$, and $^{1}H_{\alpha}$ are assigned, but drops to 81% if no chemical shifts are available (Shen and Bax, 2015).

1.10 Prediction of tertiary structure using computational method CS-ROSETTA

Chemical Shift (CS)-ROSETTA offers a protocol that generates the *de novo* tertiary structure of proteins. CS-ROSETTA generates protein structures using a combination of the assigned chemical shifts $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, ^{15}N, $^{1}H_{\alpha}$, and $^{1}H^{N}$, and the sequence of the protein in question. The assigned chemical shifts in combination with the primary structure is matched with peptide fragments found in the ROSETTA structural database (Shen *et al*., 2009). CS-ROSETTA thus proves an alternative to the traditional use of NOE constraints for structural calculations.

1.11 Recombinant gene expression

$^{15}N/^{13}C$ isotopically enriched proteins used for the structural determination by NMR are usually obtained through recombinant gene expression. The gene of the protein of interest is cloned into a vector, a small self-replicating artificial plasmid, which is further transformed into a suitable host cell (in most cases *E.coli*) (Simmons and Snustad, 2012).

One such vector system is the pUCBB vector. pUCBB vectors rely on a T7-promoter system. The T7-promoter system is named after a RNA polymerase derived from bacteriophage T7. The gene of interest is inserted downstream from the T7-promoter. Bacterial host, such as *E.coli*, does not encode T7 polymerase. Instead, the gene for T7 polymerase is separately cloned into the vector construct under the control of a *lac*P´ promoter (Cox and Nelson, 2013). In nature, the *lac* promoter is under negative regulation by the *lac* repressor protein.

However, the *lac* repression protein dissociates from the promoter upon binding allolactose, leading to gene expression.

In the laboratory, the inducer isopropylthiogalactopyranoside (IPTG) is used in place of allolactose (Vick *et al.*, 2011). Unlike allolactose, IPTG is not metabolized by the host and therefore proves an effective inducer for expression in overnight cultures. One issue with the T7-promoter system is the efficiency of T7 polymerase, meaning the promoter is regulated in an "on/off" manner, with little- or no control over protein expression levels. High expression of heterologous proteins has been linked with cell lysis, and mixtures of native- and precursor proteins (Afroz *et al.*, 2014). These problems can to some degree be circumvented by relying on a so-called "leaky-expression", where production expression occurs when the *lac* repressor protein transiently dissociates from the lac promoter (Cox and Nelson, 2013).

Further, the presence of glucose has long been known to repress the lac operon by a mechanism known as catabolic repression or the glucose effect. The mechanism is in place to assure that glucose is metabolized before lactose. The repression is mediated by Catabolic Activator Protein (CAP) and cyclic AMP (cAMP). In the absence of glucose, the CAP/cAMP complex binds to the lac promoter and stimulates binding of RNA-polymerase (Simmons and Snustad, 2012). However, the intracellular concentration of cAMP decreases sharply when glucose is present, impeding gene expression. The catabolic repression poses a problem when producing $^{15}N/^{13}C$-labeled protein samples, as ^{13}C-enriched glucose remains the most cost-effective way to obtain double-labeled samples (Frueh, 2014).

The problem of catabolic repression can be avoided by using an expression system that is independent of the host´s carbohydrate metabolism. One such expression system is the pJB_LPMO vector shown in figure 1.21 (Courtade *et al.*, 2017). The pJB_LPMO vector is based on the broad-hoast pJB656 vector developed by Sletta *et al*. and contains a regulatory *xylS*/*Pm* promoter, a multiple cloning site, *bla* (β-lactamase), and oriV and *trf* from the natural RK2 plasmid (Sletta *et al.*, 2004).

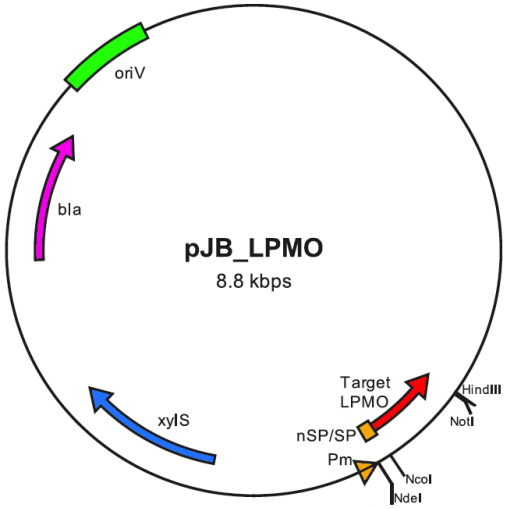

Figure 1.21: A vector map of the pJB_LPMO expression vector. The vector contains (1) Origin of replication (The RK2 replicon), (2) an antibiotics resistance gene β-lactamase, (3) the *xylS* gene, (4) the *Pm*-promoter followed by (5) a multiple cloning site. The figure has been modified from Courtade *et al*., 2017. A more thorough description of the pJB_LPMO expression system can be found in: (Sletta *et al.*, 2004).

The XylS/Pm promotor system is derived from the bacterium *Pseudomonas putida* TOL plasmid pWW0 (Sletta *et al.*, 2004). The native *Pm*-promoter controls the expression of an operon encoding enzymes involved in hydrocarbon degradation (Gawin, Valla and Brautaset, 2017). The promoter is under positive regulation of the transcription factor XylS. Upon binding benzoic acid derivatives, XylS becomes activated, dimerize, and binds to the Pm promoter, stimulating gene expression (Gawin, Valla and Brautaset, 2017). In the pJB_LPMO vector, the gene of interest is cloned into the multiple cloning site downstream from the Pm promoter. Gene expression is induced by addition of benzoic acid derivatives, such as m-toluic acid, which enters the cell through passive diffusion (Courtade *et al.*, 2017; Gawin, Valla and Brautaset, 2017). The main advantages offered by the XylS/Pm expression system is tight regulation of transcription, with a dose-depended induction. The ability to fine-tune protein production is of great importance when working with proteins that are translocated into the host´s periplasmic space (which is the case for LPMOs) because large amounts of translocated proteins have been linked with host toxicity (Sletta *et al.*, 2004). The XylS/Pm promoter system is therefore ideal for expressing isotopically enriched LPMO samples.

2 Aim of the project

This study is part of ongoing work in the research project OxyMod, focusing on redox enzymes and their potential in the depolymerization of recalcitrant polysaccharides like cellulose and chitin. The main objective of this thesis has been to study the small, chitin active, LPMO *Jd*LPMO10A by NMR spectroscopy. *Jd*LPMO10A is one of the smallest LPMOs to be discovered. This study, in combination with the determined crystal structure of *Jd*LPMO10A, can provide a good model of this "minimal scaffold" LPMO. The first aim of the thesis has been to recombinantly express ^{15}N- and ^{15}N/^{13}C-isotopically enriched samples for backbone assignment of *Jd*LPMO10A. Based on the assignment, a second aim is to evaluate the structural information obtained by NMR and compare it with the crystal structure of *Jd*LPMO10A. This data will form the basis for further structural and functional characterization of this "minimal scaffold" LPMO, in the OxyMod project.

3 Materials and Method

3.1 Materials

3.1.1 Culture Media

The culture media used in this thesis were prepared using Reverse Osmosis (RO) water. All of the culture media were autoclaved at 121 C° for 20 minutes, after preparation. Ampicillin (100 mg/mL, Sigma-Aldrich, sterile filtered) was added after the autoclaved culture media had cooled down to about 50 °C. In the case of M9 minimal medium, the base solution was prepared and autoclaved before adding ampicillin, $MgSO_4$, MEM-vitamins (Sigma-Aldrich, sterile filtered) trace metals (Sigma-Aldrich, sterile filtered), 98% Bioexpress (Cambridge Isotope Laboratories, Inc**)** and either D-glucose or 99% ^{13}C-isotopically enriched glucose (Cambridge Isotope Laboratories, Inc, sterile filtered). Agar plates were prepared by pouring Lysogeny Broth Agar (LBA) medium with ampicillin into 20-25 mL dishes. Table 3.1 lists the compositions of the culture media. The SOC-medium and listed in the table had previously been prepared by G. I. Sætrom. Unless stated otherwise, all the chemicals used to prepare the culture media are of analytical grade.

Table 3.1: The composition of the culture media used. In the case of the minimal medium (M9), the base solution was prepared and autoclaved before the addition of nutrients and trace metals (sterile filtered): 0.1 g/L $ZnSO_4$, 0.8 g/L $MnSO_4$, 0.5 g/L $FeSO_4$, 0.1 g/L $CuSO_4$, and 1 g/L $CaCl_2$, 1 M $MgSO_4$, and MEM-vitamins (100x) were purchased from Sigma-Aldrich, and 98% Bioexpress cell growth media (10x) (either ^{15}N or $^{15}N/^{13}C$ isotopically enriched) and 99% $^{15}N(NH_4)_2SO_4$ were purchased from Cambridge Isotope Laboratories, Inc. (^{15}N and $^{15}N/^{13}C$). Glucose ($C_6H_{12}O_6$) were purchased from VWR, 99% ^{13}C glucose was purchased from Cambridge Isotope Laboratories, Inc. Ampicillin was purchased from Sigma-Aldrich (SigmaA) and prepared in stock solutions of (100 mg/mL).

LB	10 g/L	Tryptone (OXOID)	**LB Agar**	10 g/L	Tryptone (OXOID)
	5 g/L	Yeast extract (OXOID)		5 g/L	Yeast extract (OXOID)
	5 g/L	NaCl (VWR)		5 g/L	NaCl (VWR)
				15 g/L	Agar (OXOID)
2xLB	20 g/L	Tryptone (OXOID)	**M9**		
	10 g/L	Yeast extract (OXOID)	Base solution	6 g/L	Na_2HPO_4 (Merck)
	5 g/L	NaCl (VWR)		3 g/L	KH_2PO_4 (Merck)
				0.5 g/L	NaCl (VWR)
				1 g/L	$(NH_4)_2SO_4$ (SigmaA)
SOC	20 g/L	Tryptone (OXOID)	Glucose solution	4 g/L	
	5 g/L	Yeast extract (OXOID)	$MgSO_4$ solution (1M)	2 mL/L	
	0.5 g/L	NaCl (VWR)	MEM-vitamins	10 mL/L	
	0.186 g/L	KCl (Merck)	Bioexpress cell growth medium	10 mL/L	
	10 mM	$MgCl_2$ (VWR)	Ampicillin	1 mL/L	
Glucose (1M)	20 mL/L	$C_6H_{12}O_6$ (VWR)	Bioexpress cell growth medium (^{15}N or $^{15}N/^{13}C$)	10 mL/L	

3.1.2 Buffers and solutions

Table 3.2 lists the buffers used for protein extraction and purification. All buffers were prepared with RO-water. The pH of the buffers was adjusted using either HCl or NaOH, and the buffers were sterile-filtered using a 0.22 μm filter. The running buffer used in the first purification step has been dubbed "Buffer A", and the elution buffer "Buffer B". The 2xSDS loading buffer listed in the table had previously been prepared by G. I. Sætrom. Unless stated otherwise, all the chemicals used to prepare the buffers are of analytical grade.

Table 3.2: Buffers used for extraction and purification of *Jd*LPMO10A. The elution buffer used for purification has been dubbed "Buffer A", while the elution buffer is dubbed "Buffer B".

Buffer A (pH 8)	50 mM	Tris-HCl (Sigma)	**Spheroplast buffer (pH 8)**	100 mM	Tris-HCl (Sigma)
				0.5 M	Sucrose (VWR)
				0.5 mM	EDTA
Buffer B (pH 8)	50 mM	Tris-HCl (Sigma)	**SEC buffer (pH 8)**	25 mM	Tris-HCl (Sigma)
	1 M	NaCl (VWR)		250 mM	NaCl (VWR)
Phosphate Buffer (pH 5.5)	20 mM	NaH_2PO_4 (Merck)	**2xSDS loading Buffer**	0.15 M	Tris-HCL (PH 6.8, Sigma)
	10 mM	NaCl (VWR)		30%	Glycerol (VWR)
				4%	SDS (Bio-rad)
				30 mM	Bromophenylblue (Sigma)
				200 mM	β-mercaptoethanol (SigmaA)

3.1.3 Plasmid Vectors

*Jd*LPMO10A was recombinantly produced using transformed *E.coli* T7 express cells containing the vectors listed in table 3.3.

Table 3.3: Vectors used to recombinantly express *Jd*LPMO10A. The gene of *Jd*LPMO10A was either fused with its native signal peptide (nSP) or with the native signal peptide from the LPMO *Sm*LPMO10A.

Plasmid	Promoter	Description	Source
pJB_SP_*Jd*	Xyls/Pm	The pJB_LPMO expression system encoding the gene of *Jd*LPMO10A with SP	Courtade *et al*., 2017,
pJB_nSP_*Jd*	Xyls/Pm	The pJB_LPMO expression system encoding the gene of *Jd*LPMO10A with nSP	Courtade *et al*., 2017
pUCBB_*Jd*	T7	A pUCBB expression system encoding the gene of *Jd*LPMO10A with nSP	PEP group, Faculty of Chemistry, Biotechnology and Food Sciences at NMBU

3.1.4 NMR equipment

All NMR spectra (section 3.2.9) were recorded on a Bruker 800 MHz (18.9 Tesla) instrument equipped with a 5 mm cryogenic TCI Z-gradient probe and ^{1}H, ^{13}C, ^{15}N, and ^{2}H channels. All spectra were recorded at 298.15 K, and the samples were prepared as described in section 3.2.10.

3.1.5 Software

The software used for data processing and data analysis during the work on this thesis are listed in table 3.4.

Table 3.4**:** The Software used in the work on this thesis. The table contains the following abbreviations: Computer Aided Resonance Assignment (CARA), Chemical Shift (CS)-ROSETTA.

Software	**Version**	**Description**
Bruker TopSpin	3.5	Used to record the NMR spectra (Section 3.2.9)
Bruker TopSpin	4.0.6	Used to process the recorded NMR spectra (Section 3.2.9)
CARA	1.8.4	Used to analyze NMR spectra and perform backbone assignment (section 3.2.10)
CS-ROSETTA	Web server (https://csrosetta.bmrb.wisc.edu/csrosetta/submit)	Used to generate a preliminary 3D-structure of *Jd*LPMO10A
PyMOL	2.3	Used for visual investigations of protein structure
TALOS-N	Web server (https://spin.niddk.nih.gov/bax/nmrserver/talosn/)	Used to predict the dihedral angels-, and secondary structure of *Jd*LPMO10A

3.2 Methods

3.2.1 Plasmid Isolation

The LPMO expression vectors pJB_nSP_*Jd* (*Jd*LPMO10A + signal peptide from *Sm*LPMO10A) and pJB_SP_*Jd* (*Jd*LPMO10A + native signal peptide) were isolated from transformed E.coli DH5α cells using the Wizard® Plus SV Minipreps DNA Purification System (Promega).

25 mL LB-medium+ ampicillin (100 mg/mL) were inoculated with transformed DH5alpha cells, collected from a frozen glycerol stock, and incubated overnight at 30 °C and 225 rounds per minute (rpm). One milliliter of the overnight culture was transferred to an Eppendorf tube and pelleted by centrifugation at 12,100 xg at room temperature (r.t) for three minutes. The pellet was resuspended in Cell lysis solution (250 μL), and the tube was inverted four times to mix. Alkaline protease solution (10 μL) was added, the tube was inverted four times to mix, followed by 5 minutes of incubation at room temperature. Neutralizing solution (350 μL) was added, the solution was mixed and lastly centrifuged at 12,100 xg for 10 minutes at r.t.

The lysate was decanted into a wizard® plus SV minipreps DNA purification spin column, placed in a collection tube, and centrifuged for one minute at 12,100 xg at r.t. The flow-through was discarded. Wash Solution + ethanol (750 μL) was added, and the spin column was centrifuged at 12,100 xg for one minute at r.t. This step was repeated after addition of 250 μL of Wash Solution. The spin column was centrifuged at 12,100 xg for 2 minutes at r.t.

In the final step, the spin column was transferred to a sterile 1,5 mL Eppendorf tube and Nuclease-Free Water (100 μL) was added. The column was centrifuged at 12,100 xg at for one minute at r.t, eluating the isolated DNA from the column. The concentration of the isolated DNA was determined by analyzing 1 μL of the sample by single NanoDrop one (Thermo scientific) UV-spectrometer.

3.2.2 Transformation of competent cells

Competent E.*coli* T7 Express cells (New England Biolabs) were incubated for 10 minutes on ice, before adding 3 μL of the isolated plasmid (either pJB_nSP_*Jd* or pJB_SP_*Jd*). The cells were further incubated on ice with the isolated plasmid for 20 minutes. After incubation, a

heat shock was induced by placing the cells in a water bath holding 42 °C for 30 seconds. After the heat shock, the cells were incubated on ice for 2 minutes. Pre-heated SOC medium (1.5 mL, 37 °C) was added, and the cells were incubated at 37 °C and 225 rpm for one hour. Lastly, the cells were plated out on selective Amp LBA plates and incubated at 37 C° overnight.

3.2.3 General method for protein production and periplasmic extraction

Starter cultures of 25 mL LB-medium + ampicillin (100 mg/mL) were inoculated with freshly transformed *E.coli* T7 express cells (section 3.2.2) and incubated overnight at 30 °C and 225 rpm. One percent of the starter culture was transferred to a main culture of 500 mL 2xLB-medium + ampicillin (100 mg/mL) and incubated in a 2L baffled shake flask at 30 °C at 225 rmp until the OD_{600} ~ 0.8-1.2. The culture was cooled on ice for 5 minutes before adding 500 µL of the inducer, m-toluic acid (100 mM in 96% ethanol). Following induction, the culture was incubated at 16 °C and 225 rpm overnight.

The Osmotic shock method was used to extract *Jd*LPMO10A from the periplasmic space (Manoil, C., Beckwith J., 1986). The cells were harvested by centrifugation at 4000 xg and 4 C° for 10 minutes, and the supernatant was discarded. The remaining cell pellet was incubated for 5 minutes on ice, before being resuspended in 30 mL spheroplast buffer (table 3.1.2) together with half-a-tablet of cOmplete™ ultra protease inhibitor (Roche).

The resuspended cells were centrifuged at 5000 xg and 4 °C for 10 minutes and the supernatant was discarded. The remaining cell pellet was incubated for 10 minutes at room temperature before being resuspended in 25 mL ice-cold RO-water together with half-a-tablet of cOmplete™ ultra protease inhibitor (Roche). After resuspension, 1.5 mL $MgCl_2$ (20 mM, Sigma-Aldrich) was added to the sample. Finally, the resuspended cells were centrifuged at 15000 xg in 4 C° for 15 minutes. The resulting supernatant was kept and filtered through a 0.22 µm filter.

To achieve isotope-labeled protein samples, the transformed T7-express cells were incubated in M9 minimal media with 98% $(^{15}NH_4)_2SO_2$ and 98% ^{15}N-bioexpress (Cambridge Isotope Laboratories, Inc) to obtain ^{15}N-labeled samples, and 98% $(^{15}NH_4)_2SO_2$, 98% $^{15}N/^{13}C$-

bioexpress (Cambridge Isotope Laboratories, Inc), and 99% sterile filtered ^{13}C-glucose (Cambridge Isotope Laboratories, Inc) to obtain ^{15}N/^{13}C-labeled protein samples.

3.2.4 Protein purification

The periplasmic extract was purified by Fast Liquid Protein Chromatography (FPLC) in two steps. First using anion exchange chromatography and secondly by Size Exclusion Chromatography (SEC). Both purification steps were performed on either an ÄKTA*start* automatic liquid chromatography system or an ÄKTA Prime Plus FPLC system (GE Life Sciences).

The first purification step was performed using a 5 mL HiTrap DEAE FF anion exchange column (GE Healthcare Life Sciences). The column was equilibrated with five column volumes of equilibration Buffer A, with a flow rate of 4.5 mL/min. The periplasmic extract was loaded onto the column, and the column was washed with another five column volumes of equilibration Buffer A to wash out unbound species. *Jd*LPMO10A were eluted from the column by a linear salt gradient. The salt gradient was achieved by increasing the concentration of elution buffer B from 0-50% over a total of 60 column volumes, with the flow rate being kept at 4.5 mL/min. The eluates were collected in fractions of 5 mL. Finally, the column was washed with 100% elution Buffer B over five column volumes. The fractions containing *Jd*LPMO10A were pooled together and concentrated by spinning the fractions down in a 10 kDa cut-off spin column (Sartorius). After fractionation, the column was washed with five column volumes of RO-water, followed by five column volumes of ethanol (20%).

The second purification step was performed using a HiLoad® 16/600 Superdex® G-75 pg SEC column. The column was equilibrated with three column volumes of RO-water, followed by three column volumes of SEC-buffer, both with a flow-rate of 1 mL/min. The pooled fractions from the first purification step were loaded onto the column, and all the eluates were collected in fractions of 5 mL. After elution, the SEC column was washed with three column volumes of MQ-water, followed by three column volumes of ethanol (20%) both with a flow-rate of 0.5 mL/min.

3.2.5 Protein Assessment using SDS-PAGE

SDS-PAGE (Sodium Dodecyl Sulfate-PolyAcrylamide gel electrophoresis) was used to analyze the chromatographic fractions from both purification steps, to asses if recombinant expression of *Jd*LPMO10A had been achieved.

The chromatographic fractions of interest (10 μL) were transferred to individual Eppendorf tubes and 2x SDS-loading buffer (10 μL) was added to each sample. The samples were heated to 95 °C for five minutes, followed by one minute of centrifuged at 12,100 xg at r.t. The fractions were loaded onto a ClearPAGE™ 12% gel (Expedeon) as shown in figure 3.1. Precision Plus Protein™ Standards (Bio-Rad, 7 μL), was added to one of the wells on the gel as the standard. The gel was run for 85 minutes at 130 V. The SDS-PAGE gel was stained with InstantBlue™ (Expedeon) for one hour to make the protein bands visible. Finally, the gel was washed overnight in water.

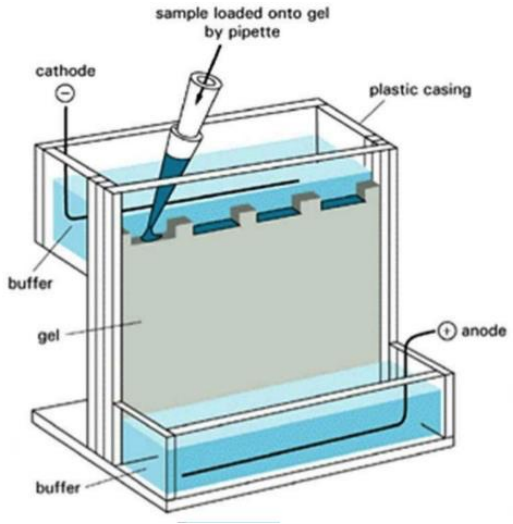

Figure 3.1: SDS-PAGE gel and electrophoresis apparatus. The samples are loaded vertically into the wells found on top of the SDS-PAGE gel. The bottom of the electrophoresis apparatus is filled with an electrophoresis buffer (Thermo Fisher Scientific), when the power is turned on an electric field is generated. The figure has been collected from Alberts *et al*., 2015.

3.2.6 Alternative strategies for protein production

Three alternative strategies for recombinant production of *Jd*LPMO10A was pursued. (1) The plasmid pJB_SP_*Jd* was tested in three different *E.coli* strains (section 3.2.6.1), (2) The plasmid pUCBB_*Jd* was tested (section 3.2.6.2), and (3) a high cell density cultivation method was tested (section 3.2.6.3).

3.2.6.1 Protein expression using the pJB_SP_*Jd* expression system in three *E.coil* strains

Competent *E.coli* T7-express, BL21 (D3) and RV308 cells (New England Biolabs) were transformed with the pJB_SP_*Jd* expression system using the method described in section 3.2.2.

Recombinant expression of *Jd*LPMO10A was induced in the transformed *E.coli* T7 express cells, using the general method, in both LB medium and in M9-medium with 2% glucose (section 3.2.3). The resulting periplasmic extract was purified using the two step purification protocol (section 3.2.4), and the results were analyzed by SDS-PAGE (section 3.2.5).

Recombinant expression of *Jd*LPMO10A was also attempted by cultivating the transformed *E.coli* T7 express, BL21 (D3) and RV308 cells in main cultures of M9-medium 2% glucose or 2% glycerol as the sole carbon source, following the general method for protein production and periplasmic extraction (3.2.3). The resulting periplasmic extracts were not purified. The results were instead analyzed directly by SDS-PAGE (section 3.2.5). Transformed *E.coli* DH5α cells containing the vector pAT64 was used as a negative control to assess protein production (Sletta *et al.,* 2007)

3.2.6.2 Protein expression using the pUCBB_*Jd* plasmid

The vector pUCBB_*Jd* (*Jd*LPMO10A + native signal peptide), constructed by the Protein Engineering and Proteomics group (PEP group), was isolated and transformed into competent *E.coli* T7-express cells by G. I. Sætrom, using the methods described in section 3.2.1 and 3.2.2. The transformed *E.coli* T7-express cells were incubated in starter cultures- and transferred to a main culture of M9-medium + 2% glycerol following the general method (section 3.2.3). The main culture was incubated overnight at 30 C° and 225 rpm. No inducer was added to the main culture, as the production of heterologous proteins was achieved through “leaky-expression” (section 1.11). The resulting periplasmic extract was purified using the two step purification protocol (section 3.2.4), and soluble fractions were analyzed by SDS-PAGE (section 3.2.5).

3.3.6.3 Protein expression in high-cell-density cultures

An alternative cultivation method was attempted to obtain higher protein production levels. The aim of the alternative cultivation method was to increase heterologous protein expression by inducing protein expression in a high-cell-density culture.

Starter cultures of 25 mL LB + ampicillin (100 mg/mL) were inoculated with Transformed *E.coli* T7-express cells containing the pJB_nSP_*Jd* plasmid and incubated overnight at 30 °C and 225 rpm. One percent of the starter cultures were transferred to two main cultures of 500 mL LB + ampicillin (100 mg/mL). The two main cultures were cultivated at 30 °C in a LEX bioreactor to a OD_{600} ~0.8. The cells from the two cultures were harvested by centrifugation at 5000 xg in 4 °C for 10 minutes. The supernatant was discarded, while the remaining cell pellets were resuspended, and combined, in a single main culture of M9-medium with 2% glucose. The culture was induced by addition of 500 μL m-toluic acid (100 mM in 96% ethanol) and incubated in a LEX bioreactor at 30 C° overnight. The periplasmic proteins were extracted using the general method (section 3.2.3), with the exception that the periplasmic extract was incubated for 1 hour in 5 mM EDTA to remove potential metal contaminations. The periplasmic extract was using the two step purification protocol (section 3.2.4), and the protein production was analyzed by SDS-PAGE (section 3.2.5).

3.2.7 Sequencing pJB_SP_*Jd*

The pJB_SP_*Jd* plasmid was sequenced by GATC Biotech which offers Sanger sequencing as a part of their LIGTHRun service.

The pJB_SP_*Jd* plasmid was isolated from *E.coli* transformed T7-express cells using the method described in section 3.2.1. The isolated plasmid (8 μL) was transferred to an Eppendorf tube with 2 μL of primer. Two samples were prepared for sequencing, one with the primer Pm/Xyls-forward, and one with the primer m/Xyls- Reverse.

3.2.8 Preparation of NMR-samples:

The chromatographic fractions from SEC found to contain pure *Jd*LPMO10A were pooled and concentrated by spinning the fractions down at 7000 xg and 10 °C in a 10 kDa cut-off spin column (Sartorius). The buffer was changed from SEC buffer to >99 % phosphate buffer (table 3.1.2). The fractions were concentrated to a volume of either < 0.5 mL or <0.15 mL and transferred to a 5 mm or 3 mm thin-wall NMR tube (Bruker), respectively. Deuterium (99.9%, Cambridge Isotope Laboratories) was added to the sample to a concentration of 10%, and the final volume of the sample was adjusted to 500 μL (5 mm NMR tube) or 160 μL (3 mm NMR tube) using phosphate buffer.

3.2.9 NMR-experiments:

The following NMR-experiments were all recorded on a Bruker 800 MHz (18.9 Tesla) instrument at 298.15 K.

- ^{15}N-HSQC was recorded with both the ^{15}N- and ^{15}N/^{13}C- labeled samples of *Jd*LPMO10A. The samples were prepared as described in section 3.2.8.

- HN(CO)CA, HNCA, HN(CA)CO, HNCO, HN(CO)CACB and HNCACB spectra were recorded with the ^{15}N/^{13}C-labled sample of *Jd*LPMO10A. The sample was prepared as described in section 3.2.8.

3.2.10 General method for backbone resonance assignment

The sequential-specific assignment of isotopically labeled proteins can be summarized in four steps: (1) Pick spin systems originating from backbone amine groups, (2) Assign associated backbone resonances to each spin system, (3) Correlate neighboring spin systems into sequential fragments, and (4) Assign the residue type of each spin system (Ser, Gly, Ala etc.), using the isotropic chemical shifts of C_α and C_β, and map the sequential fragments to segments of the proteins primary structure (Cavanagh, 2007; Frueh, 2014). All of the NMR spectra were analyzed in CARA (Computer Aided Resonance Assignment) (Keller, 2004) version 1.8.4, using a combination of 2D ^{15}N-HSQC and the three-dimensional triple resonance experiments listed in section 1.6.2.1.

1 Picking spin systems:

Peaks in the ^{15}N-HSQC spectrum of *Jd*LPMO10A, thought to belong to the ^{15}N-^{1}H pairs in the peptide backbone, were picked as spins-systems in CARA. As mentioned in section 1.6.2.1, all residues in the backbone, with the exception of proline and the N-terminal end, are visible in the ^{15}N-HSQC spectrum. When picking peaks, it is important to distinguish ^{1}H-^{15}N backbone resonances from ^{1}H-^{15}N side-chain resonances. This was done in CARA by simultaneously inspecting the ^{15}N-HSQC, spectrum and N/H and C/H strips for each potential spin system (Frueh, 2014). Clear cross-peaks were observed between backbone ^{1}H-^{15}N-peaks and other backbone atoms, such as CO and C_α. No such correlations were found between side-chain ^{1}H-^{15}N peaks and backbone atoms. The general method used to pick spin in CARA systems is shown in figure 3.2.

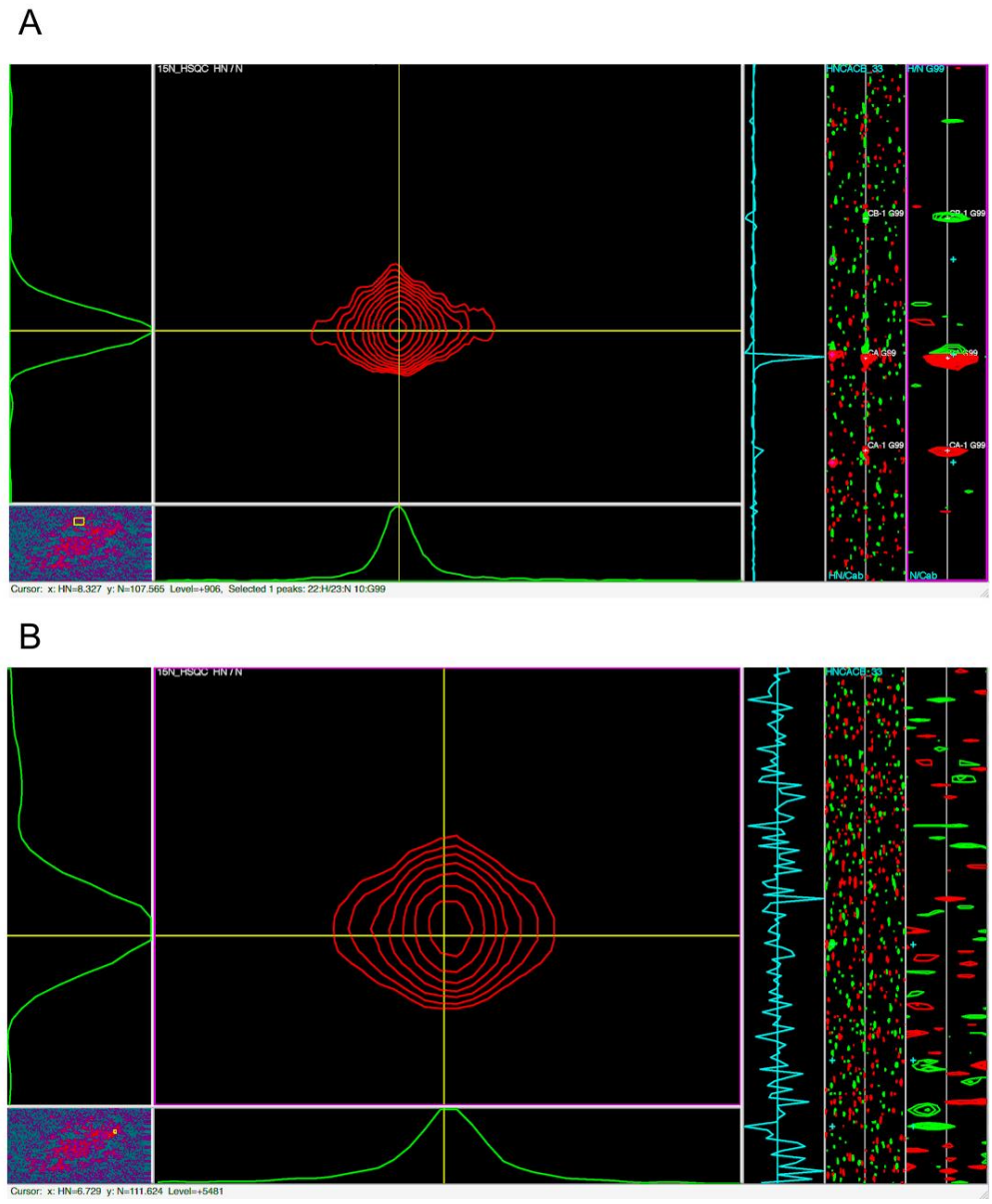

Figure 3.2: The general method used for picking peaks as spin-systems in the ^{15}N-HSQC spectrum of *Jd*LPMO10A. Each peak in the ^{15}N-HSQC spectrum corresponds to a ^{1}H-^{15}N pair found in peptide backbone or in the side-chains of Trp, Ans, Gln, Lys, Arg and His residues (Frueh, 2014). The backbone peaks can be distinguished from side-chain peaks by inspecting the potential spin system using CARA PolyScope. The main window shows the ^{15}N-HSQC spectrum of the sample. Looking at one peak, CARA allows for simultaneous inspection of the C/H strips, displaying correlations between backbone atoms recorded by the different triple-resonance experiments (far right-hand window). **A)** A clear observable correlation between the spin system and the backbone atoms $^{13}C_{\alpha}$, $^{13}C_{\alpha\text{-}1}$, $^{13}C_{\beta}$ and $^{13}C_{\beta\text{-}1}$ are visible when inspecting the strip of HNCACB. **B)** No correlations to backbone resonances are possible to detect in the strip of HNCACB. The Y-axis in the figure shows the ^{15}N-chemical shifts, and the X-axis the ^{1}H-chemical shifts in parts per million (ppm). The figure was made by Christensen, I. A.

2 Assigning backbone resonances:

Once the backbone ^{1}H-^{15}N resonances had been picked they were used as HN-anchors to correlate associated backbone resonances. The recorded three-dimensional triple resonance experiments HN(CO)CA, HNCA, HN(CA)CO, HNCO, HN(CO)CAC and HNCACB were used to assign $^{13}C_{\alpha}$, $^{13}C_{\alpha-1}$, ^{13}CO, ^{13}CO-1, $^{13}C_{\beta}$ and $^{13}C_{\beta-1}$ to each spin system using the method shown in figure 3.3. The observable signals provided by each experiment is described in section 1.6.2.1.

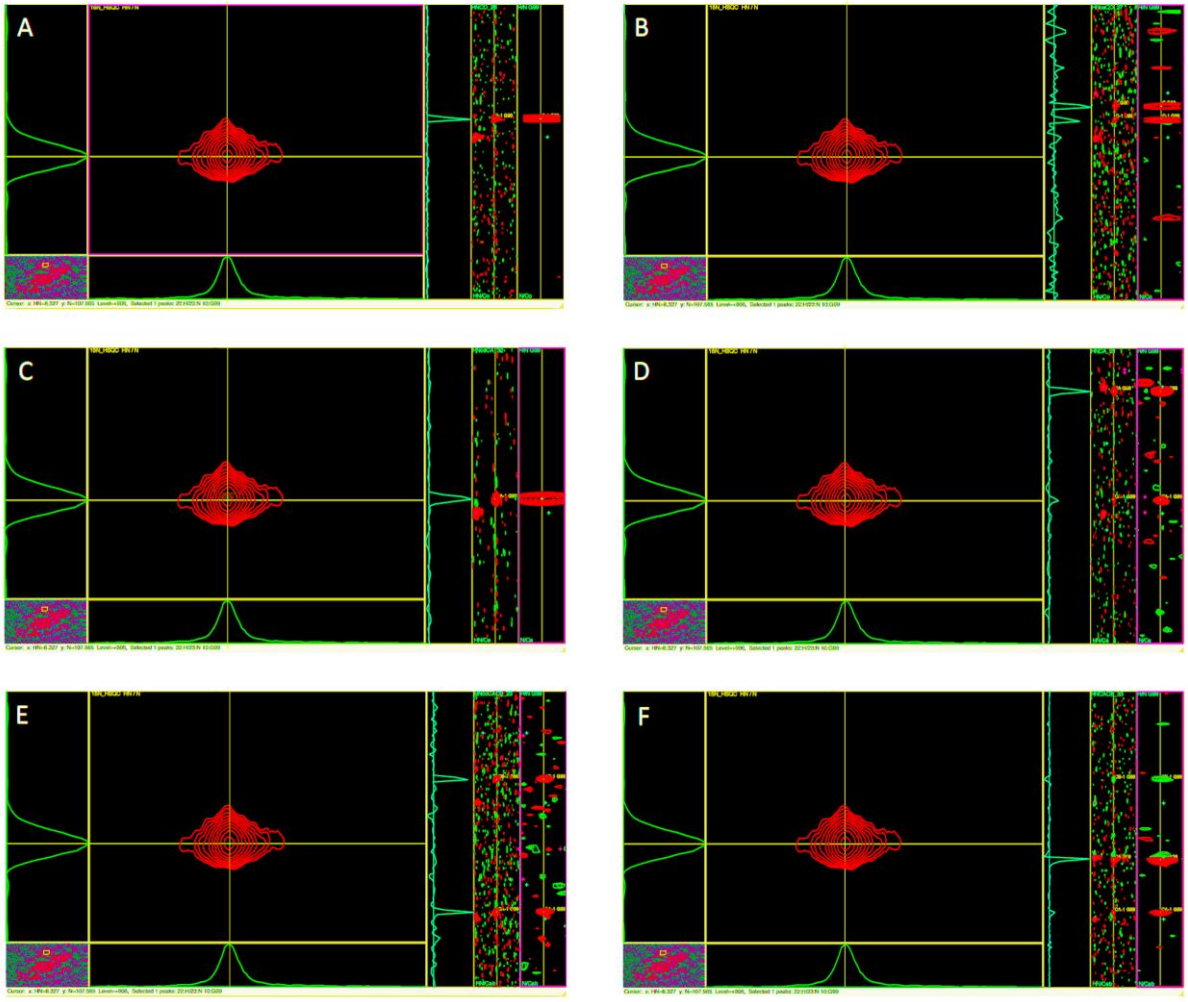

Figure 3.3: The general method used to assign $^{13}C_{\alpha}$, $^{13}C_{\alpha-1}$, ^{13}CO, ^{13}CO-1,$^{13}C_{\beta,}$ and $^{13}C_{\beta-1}$ to spin systems using CARA PolyScope. PolyScope makes it possible to view the 2D ^{15}N-HSQC spectrum of *Jd*LPMO10A simultaneously as the corresponding C/H strips of the triple resonance experiments. **A)** HNCO is used to assign ^{13}CO-1, **B)** in HN(CA)CO is used to assign ^{13}CO-1 (weaker intensity) and ^{13}CO (stronger intensity), **C)** HN(CO)CA is used to assign $^{13}C_{\alpha}$, **D)** HNCA is used to assign $^{13}C_{\alpha}$ (stronger intensity) and $^{13}C_{\alpha}$-1 (weaker intensity), **E)** HN(CO)CACB is used to assign $^{13}C_{\alpha}$- 1 and $^{13}C_{\beta}$-1, **F)** HNCACB is used to assign $^{13}C_{\alpha}$- 1, $^{13}C_{\beta}$-1 (weaker intensity) and $^{13}C_{\alpha}$- $^{13}C_{\beta}$ (stronger intensity). The figure was made by Christensen, I. A.

3. Correlating sequential spin systems:

Once $^{13}C_{\alpha}$, $^{13}C_{\alpha\text{-}1}$, ^{13}CO, $^{13}CO\text{-}1$, $^{13}C_{\beta}$, and $^{13}C_{\beta\text{-}1}$ had been assigned to each spin system, it was possible to correlate the spin systems into sequential fragments. Correlations are found by comparing the assigned backbone chemical shifts for the residue in question (*i*) and its sequential precursor (*i*-1) between spin systems. This was done using CARA StripScope (Keller, 2004) which makes it possible to compare the C/H strips of the assigned backbone chemical shifts between spin systems. An example of the method used for sequential correlation of spin systems is shown in figure 3.4. The figure shows how correlations between spin systems can be found using the HNCA experiment.

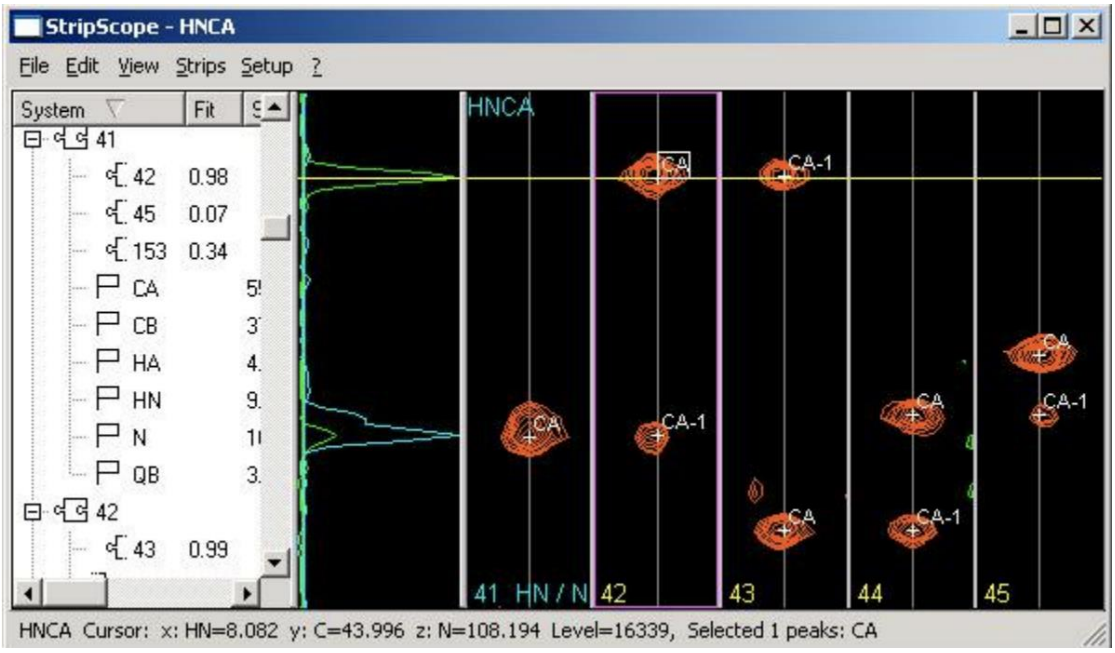

Figure 3.4: Linking of spin systems into fragments using CARA StripScope and the HNCA experiment. The HNCA experiments provides the chemical shifts of $^{13}C_{i}$ and $^{13}C_{i\text{-}1}$ of the residue in question. The chemical shift of $^{13}C_{i\text{-}1}$ of the residue in question matches the chemical shift of $^{3}C_{i}$ of its sequential precursor, making it possible to link the two spins systems together. The figure has been collected from Keller, 2004.

4. Assigning the spin systems as residues:

Lastly, the correlated fragments were matched with the sequence of *Jd*LPMO10A. To accomplish this, the identity of the spin systems was determined. The isotropic chemical shifts of C_{α} and C_{β} for each spin system were used to identify the amino acid in question. The assigned chemical shifts for C_{α} and C_{β} were compared with table values containing characteristic chemical shifts for the 20 common amino acids. The sequence-specific

assignment was completed by matching the correlated fragments with unique segments in the primary structure of *Jd*LPMO10A.

3.2.11 Prediction of secondary structural elements and dihedral angels using the computational method TALOS-N

TALOS-N (Shen and Bax, 2015) web server was used to predict the backbone dihedral angels phi (ϕ) and psi (ψ), and the secondary structural elements of *Jd*LPMO10A. A table with the assigned chemical shifts for the *Jd*LPMO10A backbone in NMRPipe table format were uploaded to the TALOS-N server, together with the pdb file (5aa7) of *Jd*LPMO10As crystal structure (Mekasha *et al.*, 2016).

3.2.12 Prediction of tertiary structure using the computational method CS-ROSETTA

Chemical Shift (CS)-ROSETTA (Shen *et al*., 2009) was used to generate a model of the 3D-structure of *Jd*LPMO10A based on NMR data. The TALOS file generated by the method described in section 3.2.11, was uploaded onto the CS-ROSETTA server. The root-mean-square deviation of the predicted structures compared to the crystal structure (Mekasha *et al.*, 2016; pdb file 5aa7) was calculated with respect to the $^{13}C_{\alpha}$ atoms in the protein backbone.

4 Results and Discussion

4.1 Plasmid purification and Transformation

The plasmids encoding the gene for *Jd*LPMO10A fused with the signal peptide from Smlpmo10A (pJB_SP_*Jd*) and the gene for JdLPMO10A fused with its native signal peptide (pJB_nSP_*Jd*) were successfully isolated at [32.5] and [45.2] ng/μL, respectively. The isolated plasmids were transformed into competent *E.coli* T7-express cells. transformed cells were used to produce both unlabeled, ^{15}N-labeled and ^{15}N/^{13}C-labeled samples of *Jd*LPMO10A. In addition, a ^{15}N-labeled sample of *Jd*LPMO10A was produced using the pUCBB_*Jd* plasmid through "leaky-expression".

4.2 Protein production and purification using the general method:

*Jd*LPMO10A was successfully purified from the periplasmic extract by the two step purification protocol of anion-exchange chromatography followed by SEC. In the first purification step, *Jd*LPMO10A eluted from the column at 12% [elution buffer B] during the salt gradient elution. In the second purification step by SEC, *Jd*LPMO10A eluted from the column after 60 minutes (Appendix C).

The second purification step by SEC was performed on the unlabeled, ^{15}N-labeled and one of the ^{15}N/^{13}C-labeled samples (section 4.4.3). The purity of the ^{15}N/^{13}C-labeled sample produced by the general method (section 3.2.3), was assessed to be sufficiently pure after the first purification step alone, based on visual inspection of the SDS-page gel, and was not purified by SEC. The purity of the soluble fractions from both purification steps was assessed by SDS-PAGE (section 3.2.5). The results from the analysis of the fractions from the first purification step are shown in figure 4.1.

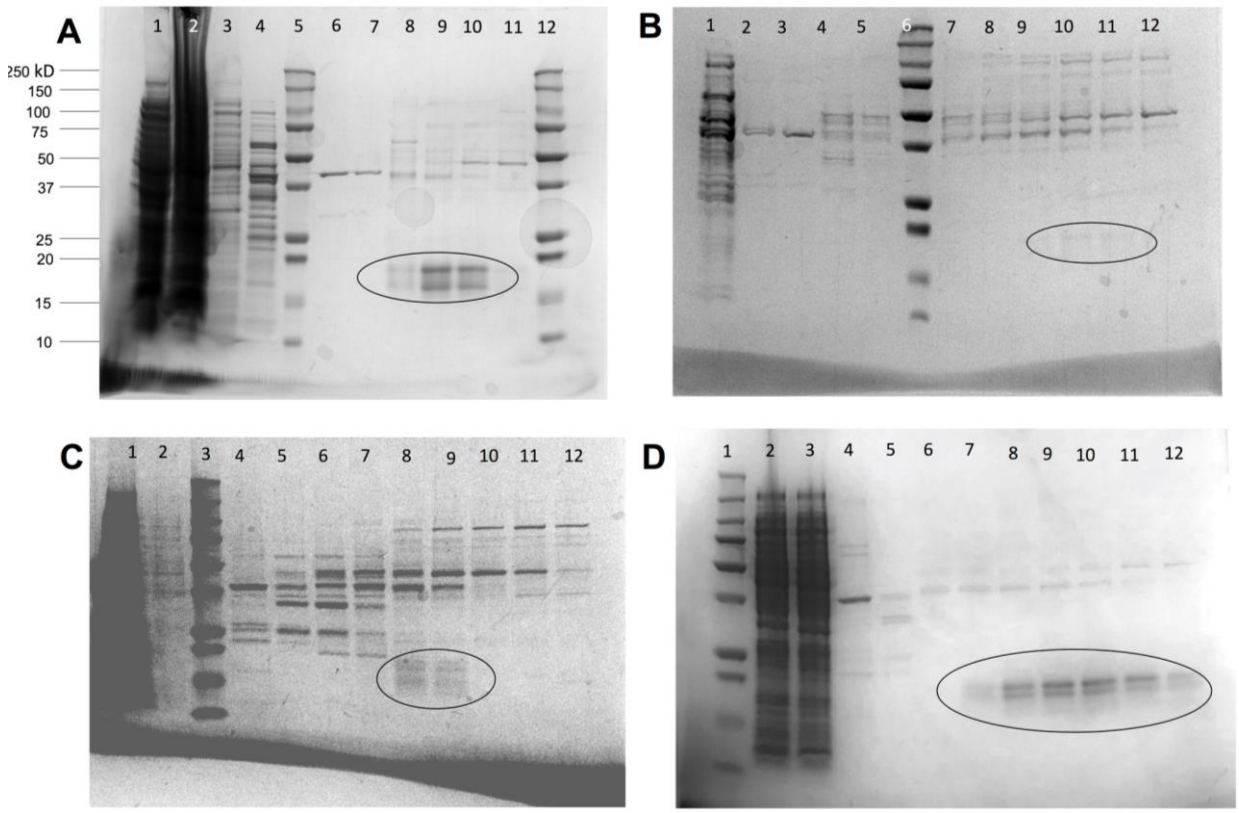

Figure 4.1: Analysis of the soluble fractions resulting after anion exchange chromatography of the periplasmic extract by SDS-PAGE. The encircled bands have a molecular weight similar to *Jd*LPMO10A. In all cases, a precision plus protein all blue standard (BioRad) was used to identify the molecular weight of the proteins present in the visible bands. **Gel A**: SDS-PAGE of the soluble chromatographic fractions of the unlabeled sample of *Jd*LPMO10A: (1) Uninduced cells, (2) Induced cells, (3) Cell pellet, (4) Periplasmic extract, (5) Standard, (6)-(12) Soluble fractions of interest. **Gel B:** SDS-PAGE of the soluble chromatographic fractions of the ^{15}N-labeled sample: (1) Uninduced cells, (2)-(5) Soluble fractions of interest, (6) Standard, (7)-(12) Soluble fractions of interest. **Gel C:** SDS-PAGE of the soluble chromatographic fractions of the ^{15}N/^{13}C-labeled sample: (1) Periplasmic extract, (2) Uninduced cells, (3) Standard, (4)-(12) Soluble fractions of interest. **Gel D:** SDS-PAGE of the soluble chromatographic fractions of the ^{15}N-labeled sample produced in M9 + 2% glycerol; (1) Standard, (2) Uninduced cells, (3) Induced cells, (4)-(12) Soluble fractions of interest. The figure was made by Christensen, I. A.

Two protein bands with molecular weights of ~ 18 and 15.5 kDa were observed after analysis by SDS-PAGE. The bands, which are encircled in figure 4.1, indicates that *Jd*LPMO10A (15.5 kDa) was recombinantly produced. Multiple other protein bands, in the range of 150-37 kDa, were also found in the soluble fractions containing *Jd*LPMO10A. These proteins are, in

most cases, not detected by NMR, as large molecules have short relaxation times. The presence of these impurities might, nevertheless, influence the sample, and illustrates the importance of a second purification step to achieve pure protein samples for structural investigations by NMR. A second purification step was therefore performed using SEC. The soluble fractions of interest were pooled and separated by SEC as described in section 3.2.4. The purity of the resulting chromatographic fractions was analyzed by SDS-PAGE, with the result of the analysis of the unlabeled sample is shown in figure 4.2.

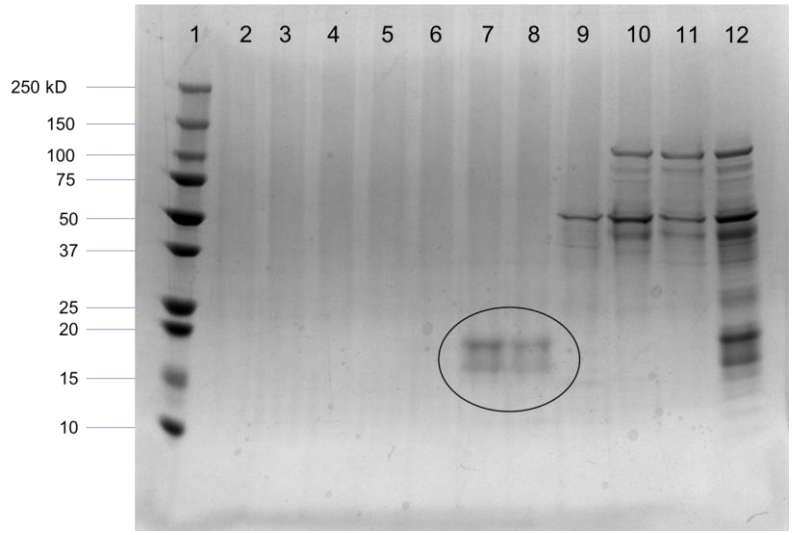

Figure 4.2: Analysis of the soluble chromatographic fractions resulting from SEC of the unlabeled sample by SDS-PAGE. The encircled bands have a molecular weight similar to that of *Jd*LPMO10A. A precision plus protein all blue standard (BioRad) was used to identify the Molecular weight of the proteins present in the visible bands. The wells contain the following samples: (1) Standard, (2)-(11) Soluble fractions of interest, (12) Pooled samples resulting from anion-exchange chromatography. The figure was made by Christensen, I. A.

Two bands with a molecular weight of ~ 18 and 15.5 kDa, similar to *Jd*LPMO10A (15.5 kDa), can be observed in the encircled areas in figure 4.2. The impurities found in the same fractions as *Jd*LPMO10A after the first purification step were no longer present after purification with SEC. This suggests the two-step purification protocol to be sufficient for obtaining pure protein samples from the periplasmic extract.

The observation of two protein bands with a similar molecular weight to *Jd*LPMO10A after analysis by SDS-PAGE is interesting. The two bands were present after both purification steps and were only detected in the instances where *Jd*LPMO10A had been successfully

produced. An explanation to why *Jd*LPMO10A seems to be present at two different molecular weights has not been experimentally uncovered. However, a potential explanation is that a mixture of the mature *Jd*LPMO10A and its precursor was isolated from the cells. The molecular weights of both the *Jd*LPMO10A precursors are ~18 kDa, which could correspond to the largest of the two encircled bands found by SDS-PAGE. However, no evidence of a *Jd*LPMO10A precursor was found when investigating the samples with NMR. Another explanation could be that the second band is the result of incomplete denaturation of *Jd*LPMO10A during SDS-PAGE. No further analysis was performed to investigate whether the second band was the result of partially denatured protein.

4.3 Assessment of protein production

Analysis of the soluble fractions by SDS-PAGE can give an indication of protein production levels, as the thickness and strength of the developed protein bands suggest the amount of protein present (Alberts, B., 2015). The weak protein bands, with molecular weights corresponding to *Jd*LPMO10A, found in the soluble fractions of the ^{15}N-labeled sample (Fig 4.2, gel B) therefore suggests suboptimal levels of protein production. In the case of the ^{15}N/^{13}C-labeled sample (Fig. 4.2, gel C), the bands corresponding to *Jd*LPMO10A are more visible because the contrast of the image has been increased to some extent, in order to better visualize the weak protein bands. Interestingly, protein production levels seemed to improve when the recombinant cells were cultivated in M9-medium with glycerol as the sole carbon source (Fig 4.2, gel D).

The suboptimal levels of protein production might be explained by a number of reasons, including: (1) acetate accumulation, (2) the pH of the growth medium, (3) protein aggregation, and (4) protein degradation.

E.coli has been demonstrated to produce acetic acid, which is present as acetate in solution, when grown in defined media with glucose as the sole carbon source (Eiteman and Altman, 2006). High concentrations of acetic acid or acetate (> 5g/L at pH 7) in the culture have been found to inhibit protein production by influencing both proteins and genes involved in the transcription and translation machinery of *E.coli*. The effects of acetate accumulation have further been shown to be more prominent for recombinant cells compared to wild-type cells (Eiteman and Altman, 2006), leading to a significant decrease in the produced amount of

recombinant proteins (Lee, 1996). Acetate is not produced when glycerol is used as the sole carbon source (Lee, 1996). The occurrence of acetate accumulation was not experimentally investigated for the cell cultures grown in M9-medium with 2% glucose but could have been interesting to investigate using an acetate colorimetric assay kits, or more extensive methods such as High Pressure Liquid Chromatography.

The production of acid by-products, such as acetate and lactic acid decreases the pH of the growth medium. Acidification of the growth medium has been demonstrated to negatively affect the production of recombinant proteins (Chou, 2007; Azatian, Kaur and Latham, 2019). In addition, the XylS/Pm system has been demonstrated to be strongly influenced by an increase in pH (Winther-Larsen *et al*, 2000). Winther-Larsen *et al*. found a 6-fold decrease in the performance of the system at pH 7.8 compared with pH 6.8. The decrease in performance at higher pH is believed to be linked with the dissociation state of the inducer m-toluic acid (Winther-Larsen *et al*, 2000). To minimize changes in pH, the recombinant bacteria were cultivated in M9-medium which functions as a buffer. The pH of the growth medium was not measured during cultivation, although it would have been interesting to measure if the pH remained stable during cultivation. Higher levels of protein production might also be achieved by using a medium with an even higher buffer capacity (Azatian, Kaur and Latham, 2019).

Incorrect folding can occur when recombinant proteins are expressed at a rate which exceeds the efficiency of the post-translational machinery of *E.coli* (Georgiou and segatori, 2005). Heterologous proteins undergo many post-translational modifications including translocation, correct folding, and disulfide bond formation (Baneyx and Mujacic, 2004). Incomplete disulfide bond formation, in particular, has been linked to protein aggregation and the formation of inclusion bodies. Inclusion bodies are accumulations of protein aggregates which can occur both the cytosol and the periplasmic space (Baneyx and Mujacic, 2004; Rosano and Ceccarelli, 2014). Over-expression of heterologous proteins can also result in increased rates of protein degradation. Large amounts of misfolded proteins trigger a number of stress responses in *E.coli,* including increased production of proteases. Protein degradation is a known problem when working with recombinant proteins. To limit protein aggregation and degradation, the transformed cells were incubated at 16 °C following induction. Decreased temperature (< 20 °C) has been found to influence both the elongation rate of ribosomes and protein degradation by proteases in *E.coli* (Murby, Uhlén and Ståhl, 1996; Farewell and

Neidhardt, 1998). In addition, *E.coli* T7-express cells, which are deficient in the two proteases Lon and OmpT, and cOmplete^{TM} ultra protease inhibitor, was uses to limit degradation.

The presence of protein degradation or aggregation can to an extent be assessed by SDS-PAGE. In the case of protein aggregation into insoluble inclusion bodies, *Jd*LPMO10A would reside in the cell pellet remaining after protein extraction (section 3.2.3). Fortunately, protein aggregates and inclusion bodies are solubilized after being heated in SDS-loading buffer (Wouter, et al., 2017). Given the formation of inclusion bodies, it would be possible to observe a protein band of ~ 15.5 kDA by analyzing the resulting cell pellet from the periplasmic extraction by SDS-PAGE (section 3.2.3). Although it was difficult to inspect the pellet samples visually on the gels, no bands of ~ 15.5 kDA could be observed in the pellet sample, that was not also present in the uninduced- and induced cell samples (example in figure 4.3). Protein degradation is typically characterized by the absence of clear protein bands after analysis by SDS-PAGE. Instead of a defined protein band, an elongated smear is often observed on the gel. No signs of protein degradation were evident in any soluble fraction. However, protein aggregation nor degradation can be ruled out completely based on visual inspection of the results from the analysis by SDS-PAGE alone.

4.4 Alternative strategies for protein production

Achieving recombinant expression of *Jd*LPMO10A in M9-medium was cumbersome. In addition to the general method (section 3.2.3), three alternative strategies for protein expression was, therefore, attempted to obtain expression of *Jd*LPMO10A. The three strategies included: (1) obtaining expression using the Pjb_SP_*Jd* vector, (2) obtaining protein expression using the pUCBB_*Jd* vector, and (3) obtaining high protein expression using a high-cell-density cultivation method. This section aims to summarize the three alternative strategies, the results obtained, and to discuss these results.

4.4.1 Protein expression using the pJB_SP_*Jd* expression system in three *E.coli* strains

An alternative strategy using pJB_SP_*Jd* vector was tested to obtain expression of *Jd*LPMO10A. In their 2017 paper, Courtade *et al*. found pJB_SP_LPMO vector constructs (vectors encoding LPMO genes fused with the signal peptide form *Sm*LPMO10A) to produce higher levels of mature, translocated LPMOs, compared with pJB_nsSP_LPMO vector

constructs (vectors encoding LPMO genes fused with their native signal peptide). Expression of *Jd*LPMO10A, using the pJB_SP_*Jd* expression system, encoding the gene for *Jd*LPMO10A fused with the signal peptide from *Sm*LPMO10A, was therefore investigated.

*Jd*LPMO10A was successfully produced in LB medium, using the general method (section 3.2.3). However, no observable amounts of translocated *Jd*LPMO10A were found when the transformed T7-express cells were cultivated in M9-medium with glucose.

Expression of *Jd*LPMO10A was, therefore, investigated in the additional *E.coli* strains RV308, and BL21 (D3). This was done to find an optimal vector/host combination. *E.coli* DH5α cells containing the vector pAT64, a XylS/Pm vector controlling the expression of a interferon-α-coding gene, was used as a negative control (Sletta *et al.*, 2007). All of the transformed *E.coli* strains were cultivated in main cultures of M9 with either glucose or glycerol as the sole carbon source. The resulting periplasmic extracts were not purified, but instead analyzed directly by SDS-PAGE (section 3.2.5). The results from the analysis of the periplasmic extracts from the transformed T7 express and BL21 (D3) cells are shown in figure 4.3.

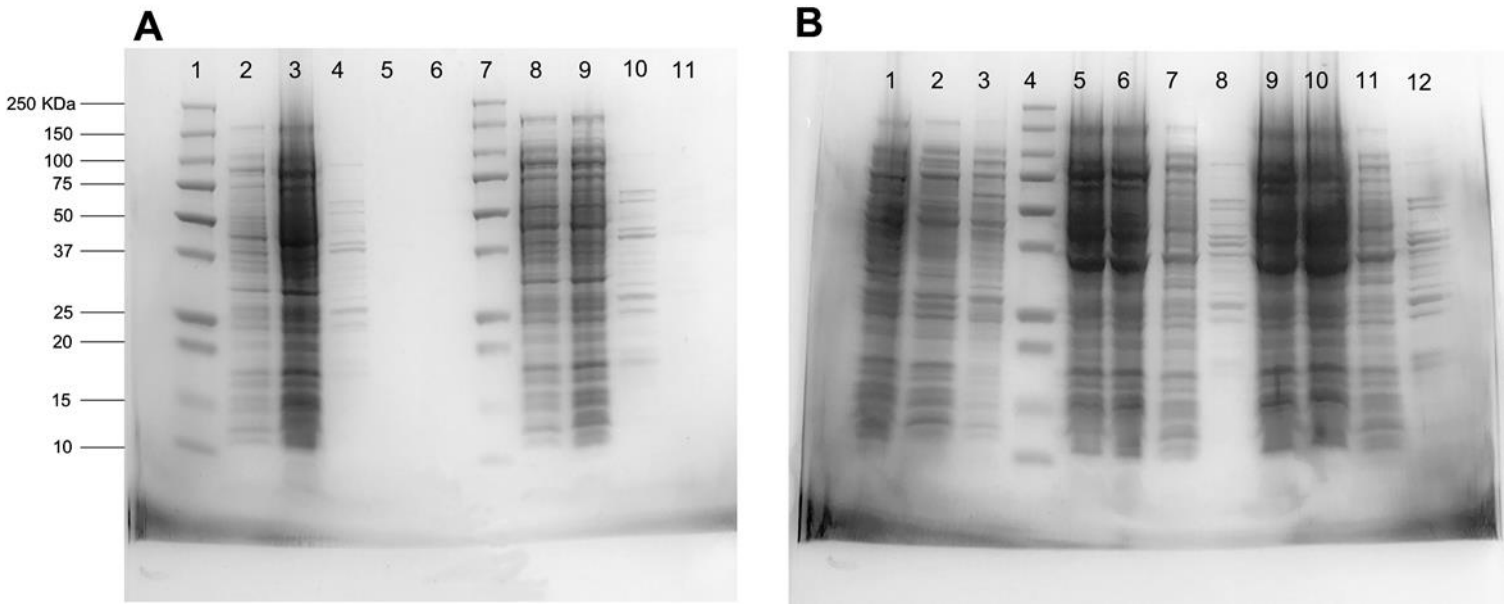

Figure 4.3: Analysis of *Jd*LPMO10A production in M9-minimal medium by SDS-PAGE. The encircled bands have a molecular weight corresponding to that of *Jd*LPMO10A. In all cases, a precision plus protein all blue standard (BioRad) was used to identify the molecular weight of the proteins present in the visible bands. **Gel A)** Transformed T7 express cells containing the pJB_SP_*Jd* plasmid cultivated in M9-medium with either 2% glucose or glycerol as the sole carbon source; (1) Standard, (2) Induced cells (glucose), (3) Cell pellet (glucose), (4) Periplasmic extract (glucose), Periplasmic extract 1:10 dilution (glucose), (6) Empty, (7) Standard, (8) Induced cells (glycerol), (9) Cell pellet (glycerol), (10) Periplasmic extract (glycerol), (11) Periplasmic extract 1:10 dilution (glycerol). **Gel B)** Transformed BL21 (D3) cells containing the pJB_SP_*Jd* plasmid cultivated in M9-medium with either 2% glucose or glycerol as the sole carbon source. Transformed DH5α cells containing pAT65 was used as control; (1) Uninduced cells (control), (2) Induced cells (control), (3) Periplasmic extract (control), (4) Standard, (5) Uninduced cells (glucose), (6) Induced cells (glucose), (7) Cell pellet (glucose), (8) Periplasmic extract (glucose), (9) Uninduced cells (glycerol), (10) Induced cells (glycerol), (11) Cell pellet (glycerol), (12) Periplasmic extract (glycerol). The figure was made by Christensen, I. A.

None of the periplasmic extracts obtained were purified, as it was believed that, it would be possible to detect bands corresponding to 15.5 kDa directly in the periplasmic extract sample after analysis by SDS-PAGE. These bands would not be present in the uninduced cell sample. However, looking at the resulting SDS-PAGE gels presented in figure 4.3, no such bands were observed. This could mean that observable amounts of *Jd*LPMO10A were not produced by any vector/host combination, or that the protein bands were too weak to be observed prior to purification of the periplasmic extract. Protein expression using the pJB_SP_*Jd* was not observed to be influenced by the carbon source in the culture.

4.4.2 Protein expression using the pUCBB_*Jd* plasmid

The T7-promoter based vector system pUCBB_*Jd* was investigated for the production of a ^{15}N-labeled *Jd*LPMO10A sample. The pUCBB_*Jd* vector was created by the Protein Engineering and Proteomics group (PEP group) at NMBU. The PEP group has used several vector constructs with variants of the lac promoter system to successfully express the genes of LPMOs (Vaaje-Kolstad *et al.*, 2005; Aachmann *et al.*, 2015; Mekasha *et al.*, 2016). Investigating the performance of a T7-promoter based system, therefore, seemed a natural alternative strategy to express heterologous *Jd*LPMO10A.

The expression of heterologous proteins by T7-promoter systems can be inhibited by catabolic repression (section 1.11). To circumvent this, the transformed *E.coli* T7 express cells were cultivated in a main culture of M9-medium with glycerol as the sole carbon source. In addition, the strong gene expression from T7-promoter systems can be unfavorable when working with proteins that are translocated into the periplasmic space, such as *Jd*LPMO10A. High levels of heterologous proteins have been linked with cell lysis and a mixture of mature- and protein precursors (Courtade *et al*., 2017). The ^{15}N-labeled sample of *Jd*LPMO10A was therefore obtained through so-called “leaky expression” (section 1.11). The resulting soluble fractions from purification of the periplasmic extract using the two step protocol (3.2.4), are shown in figure 4.4

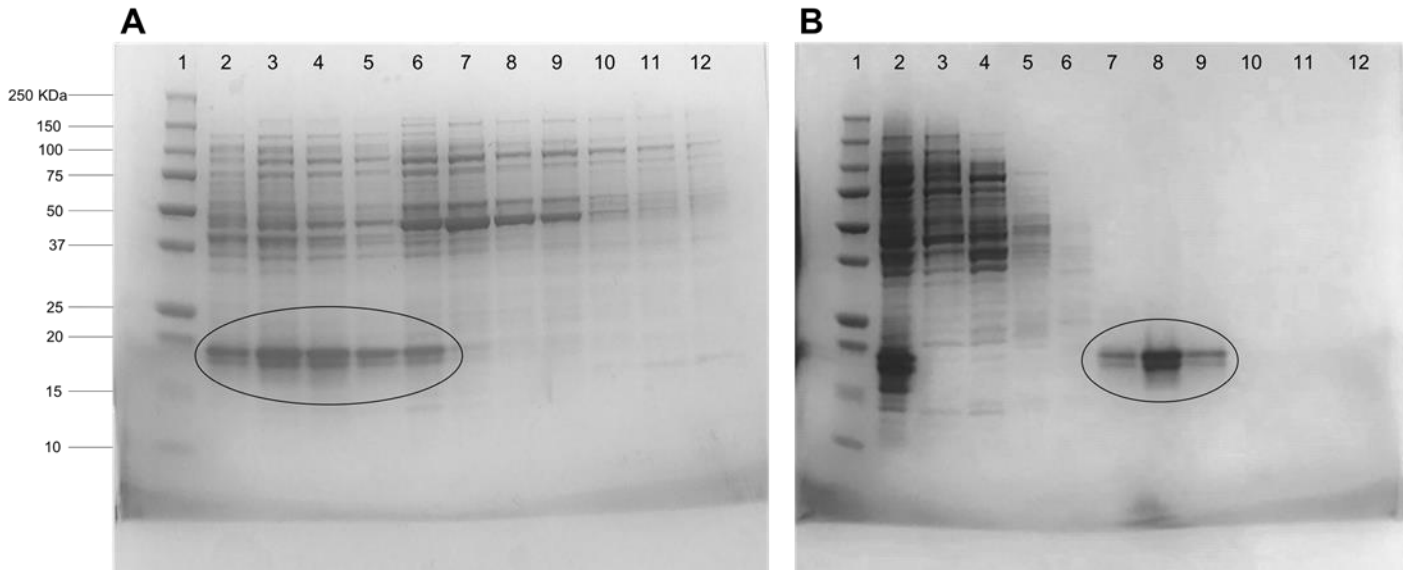

Figure 4.4: Analysis of recombinant production of *Jd*LPMO10A from transformed T7 express cells containing the pUCBB_*Jd* vector by SDS-PAGE. The encircled bands have a molecular weight corresponding to that of *Jd*LPMO10A. In both cases, a precision plus protein all blue standard (BioRad) was used to identify the molecular weight of the proteins present in the visible bands. The associated chromatograms of both purification steps are presented in Appendix C. **A)** The resulting SDS-PAGE gel after the first purification step using a anion-exchange column; **(1)** Standard, **(2)** Fraction T14, **(3)** Fraction T15, **(4)** Fraction T16, **(5)** Fraction T17, **(6)** Fraction T18, **(7)** Fraction T20, **(8)** Fraction T21, **(9)** Fraction T22, **(10)** Fraction T24, **(11)** Fraction T25 and **(12)** Fraction T26**. B)** The resulting SDS-PAGE gel after the second purification step using SEC; **(1)** Standard, **(2)** Pooled fractions of interest from purification step 1, **(3)** Fraction T10, **(4)** Fraction T11, **(5)** Fraction T12, **(6)** Fraction T13, **(7)** Fraction T14, **(8)** Fraction T15, **(9)** Fraction T16, **(10)** Fraction T17, (**11**) Fraction T18, (**12**) Fraction T20. . The figure was made by Christensen, I. A.

Two protein bands with a molecular weight of ~18 and 15.5 kDa, were found in the soluble fractions from both purification steps after analysis by SDS-PAGE. The two protein bands indicate that recombinant expression of ^{15}N-labeled *Jd*LPMO10A was successfully achieved using the pUCBB_*Jd* plasmid. Although significant amounts of impurities were present in the soluble fractions contain *Jd*LPMO10A after the first purification step, the second purification step using SEC was sufficient to obtain a pure ^{15}N-labeled sample.

Based on the results, the pUCBB_*Jd* system is an alternative strategy for heterologous expression of *Jd*LPMO10A. However, as ^{13}C isotopically enriched glucose remains the most cost-effective way to achieve double labeled protein samples, compared with ^{13}C glycerol. For this reason, the pUCBB_*Jd* vector was not deemed ideal for the production of ^{15}N/^{13}C-labeled *Jd*LPMO10A samples.

4.4.3 Protein expression in high-cell-density cultures.

An alternative cultivation method, using the pJB_nSP_*Jd* vector, was attempted to obtain higher protein expression levels. The objective of this alternative strategy for protein production, was to cultivate the transformed cells to a high cell density prior to induction. High cell densities result in a higher number of cells involved in heterologous protein production, meaning that although the amount of protein produced per cell remains the same, higher levels of protein production is obtained (Rosano and Ceccarelli, 2014).

High cell densities were obtained by combining two separate main cultures into a culture of M9-medium with glucose. In addition, the three cultures were cultivated in a LEX bioreactor, where air is pumped directly into the culture, leading to high growth rates. The protein production was analyzed by SDS-PAGE (section 3.2.5). The results from the analysis of the soluble fractions from the first purification step are shown in figure 4.5.

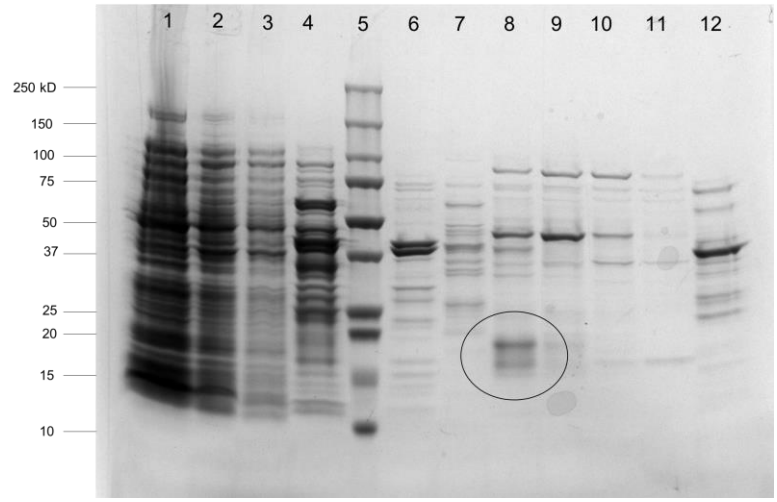

Figure 4.5: Analysis of the soluble fraction resulting from purification the periplasmic extract from recombinant T7 express cells cultivated in a LEX bioreactor by SDS-PAGE. The encircled bands have a molecular weight corresponding to that of *Jd*LPMO10A. A precision plus protein all blue standard (BioRad) was used to identify the molecular weight of the proteins present in the visible bands. The lanes contain the following soluble fractions: (1) Un-induced cells, (2) Induced cells, (3) cell pellet, (4) Periplasmic extract, (5) Standard, (6)-(12) Soluble fractions of interest. The figure was made by Christensen, I. A.

The two-step purification protocol was successful in providing a pure $^{15}N/^{13}C$-labeled sample of *Jd*LPMO10A as protein bands ~18 and 15.5 kDa, were found in the soluble fractions. Based on visual inspection of the protein bands, found by SDS-PAGE, the strength of the bands indicated that the novel protocol had indeed succeeded in producing high amounts of

*Jd*LPMO10A. However, this assumption was refuted when the sample was investigated with NMR. The peak intensities found in the proton spectrum of the $^{15}N/^{13}C$-labeled sample indicated that the amount of *Jd*LPMO10A was, in fact, comparable to the concentrations obtained using the general method (section 3.2.3). This indicated that the alternative cultivation method did not result in higher levels of protein production, compared to the general method.

This result is surprising, as high cell densities were expected to produce higher amounts of recombinant protein following induction. One possible explanation could be the acetate formation, discussed in section 4.3. Acetate accumulation is prominent when cultures are grown in conditions where oxygen is limited. Overnight cultivation in a LEX bioreactor at low temperatures stalls acetate accumulation, as air is pumped directly down into the culture. However, the production of acetate is not only related to the amount of oxygen present but also the maximum rate of oxygen uptake in *E.coli* (Eiteman and Altman, 2006). The maximum rate of oxygen assumption can, under high growth rates and high temperatures , be close to the threshold for acetate production, meaning that acetate might accumulate even under highly aerobic conditions (Eiteman and Altman, 2006). Acetate can also be produced under aerobic conditions when *E.coli* are grown in culture medium with an excess of glucose (Lee, 1996). Acetate accumulation can therefore not be factored out as a possible explanation. Alternatively, the high growth rates might have lead *E.coli* to rapidly exhaust the nutrients in the medium, which induces a number of stress responses that negatively effects recombinant expression (Palomares, et al., 2004).

4.5 Sequencing of pJB_SP_*Jd*

To investigate why no observable expression of *Jd*LPMO10A was produced using the pJB_SP_*Jd* expression system in M9-medium, the vector was sequenced with the primers Pm/Xyls-forward and Pm/Xyls-reverse, using the method described in section 3.2.4.

The Pm/Xyls-forward and Pm/Xyls-reverse primers were chosen to investigate if a mutation had occurred in the vector, impairing recombinant gene expression. A mutation in the T7-promoter region of the vector was considered as an expiation to why no protein expression was observed in M9 medium using pJB_SP_*Jd* expression system. Unfortunately, the

sequencing results were inconclusive, as the primes appear to have detached themselves shorty into the extension phase.

4.6 NMR-Experiments:

The NMR-spectra listed in section 3.2.9 were recorded for both unlabeled, ^{15}N-labeled, and ^{13}C/^{15}N-labeled samples of *Jd*LPMO10A. No NOE assignment of *Jd*LPMO10A was performed due to time constraints. The resulting ^{15}N-HSQC spectra of the ^{15}N-labeled sample produced with the pUCBB_*Jd* vector and the ^{15}N/^{13}C-labeled sample produced using the general method (Section 3.2.3) are shown in figure 4.6.

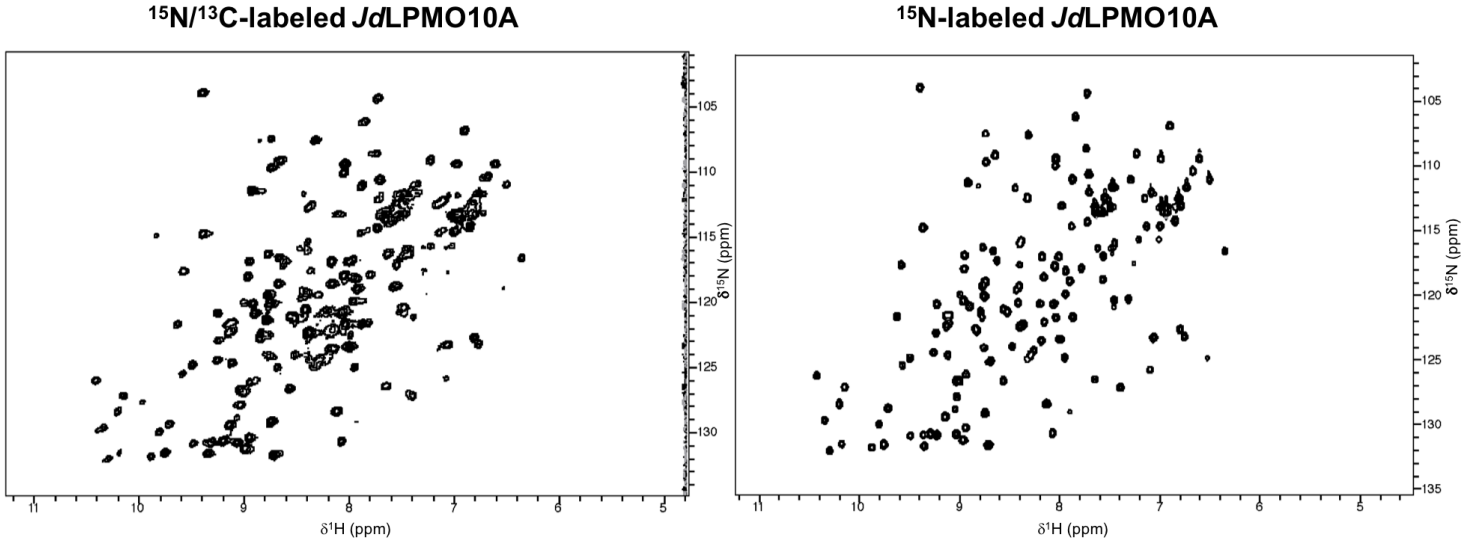

Figure 4.6: The recorded ^{15}N-HSQC spectra of the ^{15}N-labeled *Jd*LPMO10A sample produced using the pUCBB_*Jd* vector and the ^{15}N/^{13}C-labeled *Jd*LPMO10A sample produced using the general method (section 3.2.2). The X-axis shows the proton shift in parts per million (ppm), while the Y-axis shows the nitrogen shift in ppm. The samples were prepared in phosphate buffer (20 mM NaH_2PO_4, 10 mM NaCl) with 10% D_2O (99.9%), and recorded at 298.15 K. The figure was made by Christensen, I. A..

The ^{15}N-HSQC spectrum functions like a "fingerprint" for a given protein, with resonances originating from the backbone and side-chain ^{1}H-^{15}N pairs. This "fingerprint" can be used to evaluate whether the protein in the sample is correctly folded (Kwan *et al.*, 2011). Generally, the ^{15}N-HSQC spectrum of a correctly folded protein will display a larger signal dispersion compared to a disordered protein. The amide protons of unfolded proteins exist in a more similar chemical environment that they would if the proton was correctly folded, as the chemical groups in the three dimensional structure influence each other. The result is fewer

variations in the chemical shift of unordered proteins, resulting in poor signal dispersion (Kwan *et al.*, 2011). The position of the observable peaks is similar between the two ^{15}N-HSQC spectra, indicating that purified samples of *Jd*LPMO10A were indeed correctly folded. As the samples were believed to be correctly folded, heteronuclear tipple resonance experiments were recorded with the ^{15}N/^{13}C-labeled sample to be used for structural investigations of *Jd*LPMO10A.

Additional peaks are found in the ^{15}N-HSQC spectrum of the ^{15}N/^{13}C-labeled sample compared with the ^{15}N-HSQC spectrum of the ^{15}N-labeled sample. The additional peaks could arise from impurities, as the 15N/13C-labeled sample sample was only purified by anion-exchange chromatography.

4.7 Resonance Assignment of the *Jd*LPMO10A backbone:

The backbone resonances of *Jd*LPMO10A were identified and assigned using the method described in section 3.2.10. In total, 151 spin systems were picked in CARA (Keller, 2004) as potential backbone resonances in the ^{15}N-HSQC spectrum of *Jd*LPMO10A. Mature *Jd*LPMO10A is composed of 142 amino acids after the 31-residue signal peptide has been cleaved off during translocation. However, 10 of the residues are prolines which are not detected in the ^{15}N-HSQC experiment. The N-terminal residue is also undetectable in most cases. The theoretical number of observable backbone spin-systems in the *Jd*LPMO10A backbone is 131. In addition, resonances from side-chain ^{1}H-^{15}N pairs are observed in the ^{15}N-HSQC spectrum.

For these 131 observable spin systems, 91.6% of the backbone chemical shifts for ^{13}CO, ^{13}Cα, N, and H^N were assigned. Figure 4.7 shows an example of how the sequence-specific assignment method was used to assign the unique fragment Ala46-Ser47-Gly48-Glu49-Thr50 in the primary structure of *Jd*LPMO10A, using the HN(CA)CACB experiment. The ^{15}N-HSQC spectrum of *Jd*LPMO10A with the assigned spin systems is presented in figure 4.8.

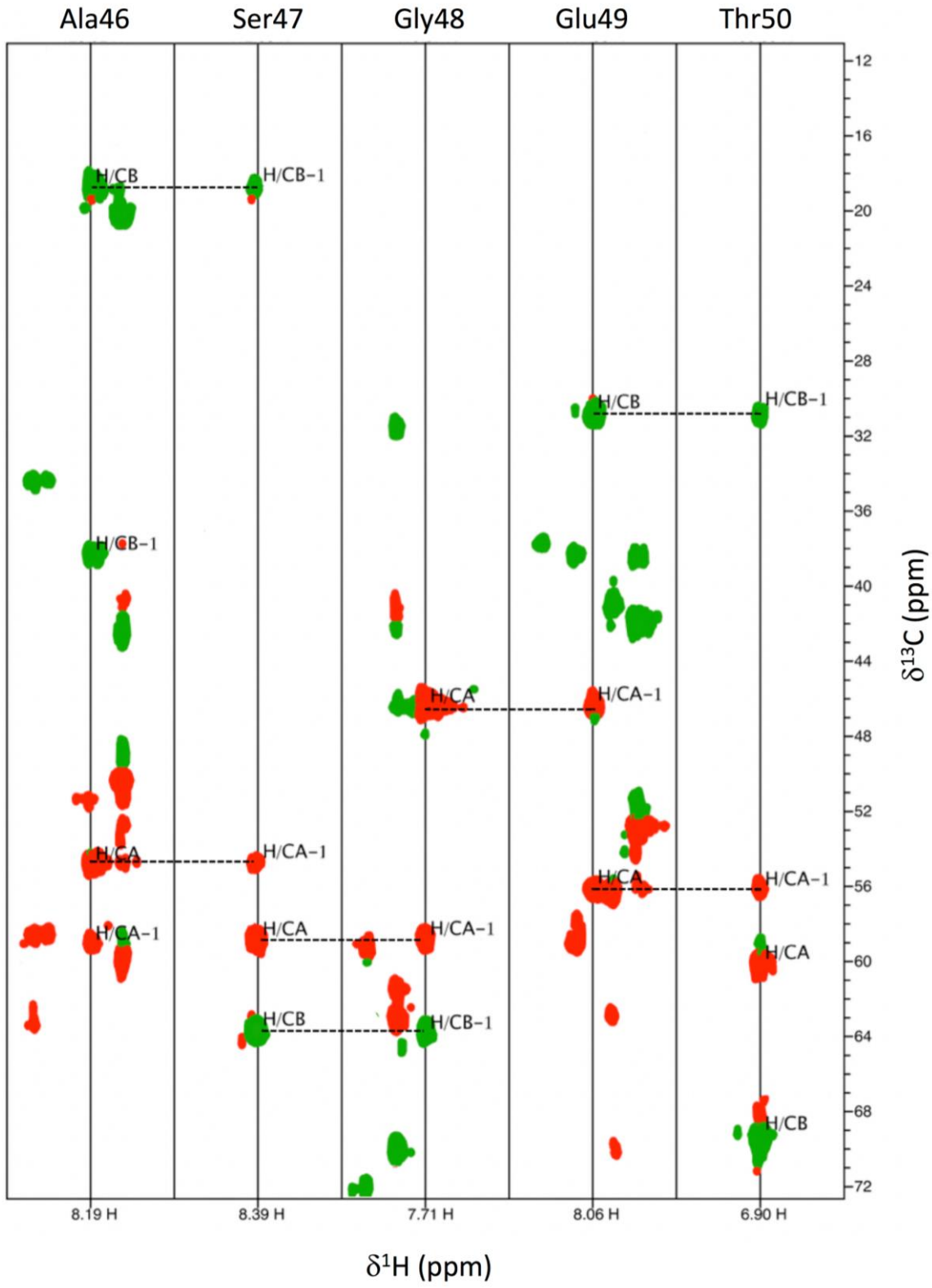

Figure 4.7: The method used for mapping the unique sequence Ala46-Ser47-Gly48-Glu49-Thr50 found in the primary structure of *Jd*LPMO10A. Cross-peaks are evident when comparing the chemical shifts for $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ to the shifts for $^{13}C_{\alpha}$-1 and $^{13}C_{\beta}$-1 between the sequential spin systems. The assignment was performed using CARA StripeScope (Keller, 2004). The bottom X-axis shows the ^{1}H-chemical shift, the top X-axis the ^{15}N-chemical shift, and the Y-axis, the ^{13}C-chemical shift, all in parts per million (ppm). The figure was made by Christensen, I. A.

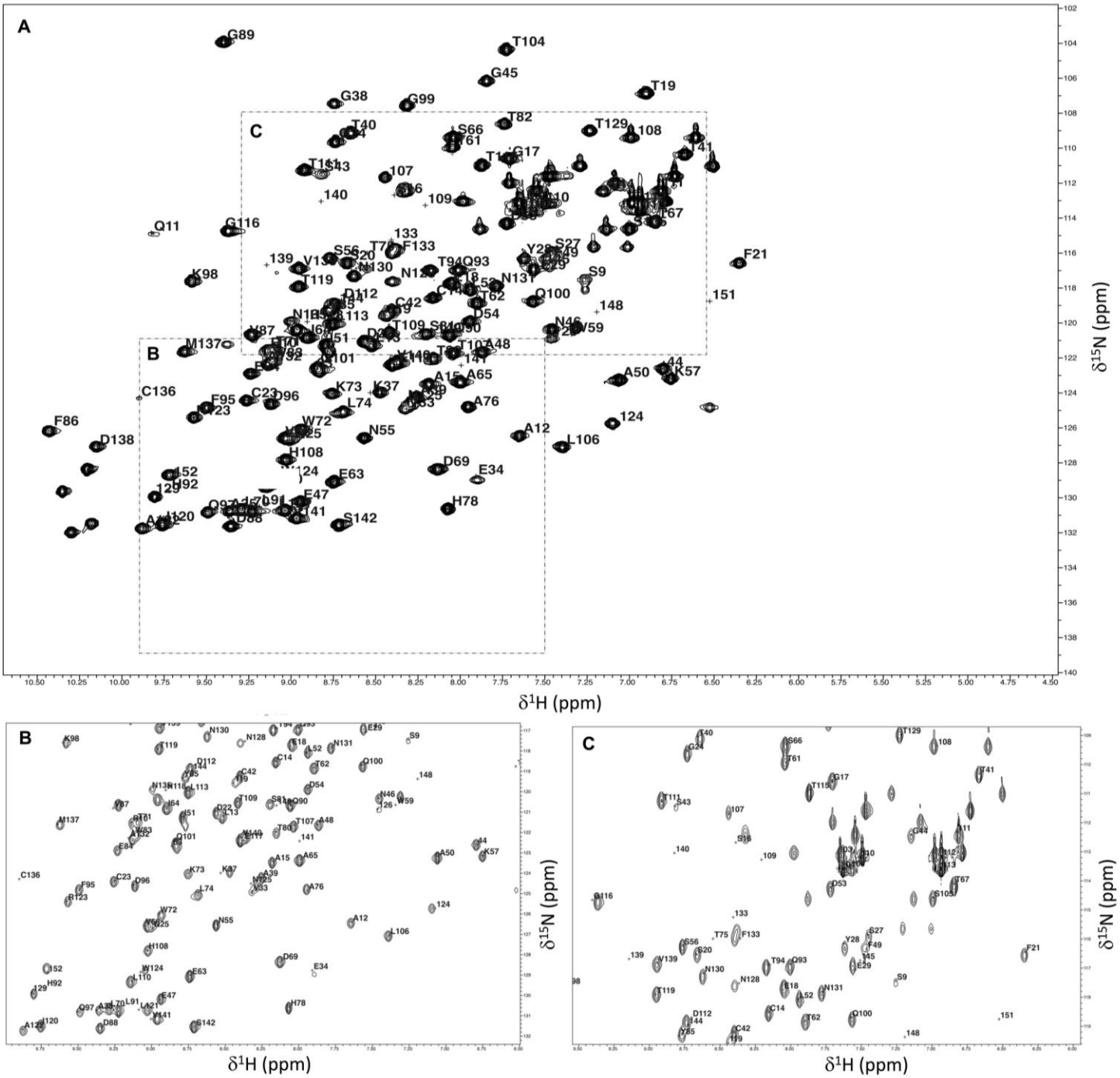

Figure 4.8: Assigned residues in the *Jd*LPMO10A backbone in the ^{15}N-HSQC of the ^{15}N-labeled sample of *Jd*LPMO10A. **A)** The full ^{15}N-HSQC of the ^{15}N-labeled sample, **B)** enlargement of area B, **C)** enlargement of area C. The sample was recorded in phosphate buffer (20 mM NaH2PO4, 10 mM NaCl, pH 5,5) with 10% D_2O (99.9%) at 298.15 K. The assigned residues are indicated by the type of residue (one letter abbreviation) and the number of the residue (*i*) from the N-terminus end (signal peptide excluded). The X-axis shows the ^{1}H chemical shift, while the Y-axis shows the ^{15}N chemical shift. The figure was made by Christensen, I. A.

No connectivity to sequential residues was found for the remaining 19 picked spin systems, which remained unassigned. The unassigned spin systems were most likely either side chain resonances mistaken for backbone resonances or artifacts in the ^{15}N-HSQC spectrum. To better illustrate which residues were assigned spin systems, the primary sequence of *Jd*LPMO10A with the assigned residues marked in red is shown in figure 4.9.

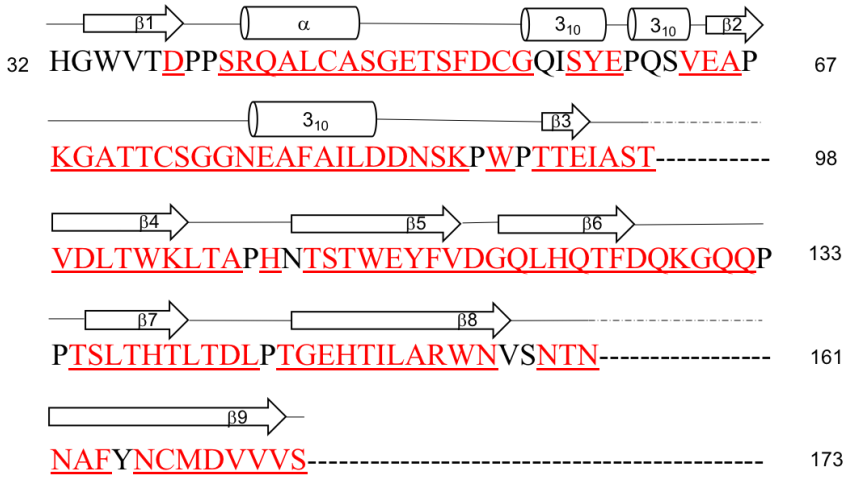

Figure 4.9: Overview of residues with assigned spin systems in the primary sequence of *Jd*LPMO10A. Assigned residues are marked in red, and unassigned residues in black. A diagram of the secondary structure of *Jd*LPMO10A found by X-ray crystallography (Mekasha *et al*., 2016; pdb 5aa7) is placed over the corresponding segments of the primary structure. The figure was made by Christensen, I. A.

As illustrated in figure 4.9, it was not possible to assign the first five residues (His32-Tyr36) found in the N-terminal end of *Jd*LPMO10A. Moreover, it was not possible to assign the residues adjacent to His109. This is interesting because both the N-terminal histidine and His109 constitute the histidine brace in the enzymes active site. Trace amounts of copper bound to the active site of the samples was considered as a potential explanation to why these residues were not detected. Cu(II) is a paramagnetic ion and its presence in solution shortens the relaxation time of nuclei within a 10-12 Å distance from the copper center (Bertini and Pierattelli, 2004). The final ^{15}N/^{13}C-labeled sample of *Jd*LPMO10A to be produced (section 4.3.3) was therefore incubated for 1 hour with 5 mM EDTA. EDTA is a chelator and was added to bind any divalent metal ions present in the sample. However, no new resonances were observed in the ^{15}N-HSQC spectrum of the sample as a result of incubation with EDTA. Suggesting that trace amounts of copper was not the cause of the missing signals.

Potential spin systems were identified for residues Val35, Thr36, Gln56, Ser158, and Tyr165. However, as the sequential connectivity of the spin systems were ambiguous, and the candidates were, therefore, not included for structural prediction. Further, some of the unassigned residues, marked in black in figure 4.9, could have been assigned indirectly by using the chemical shift information from their sequential successors. However, as the chemical shift information from the sequential successor could not be validated, indirectly assigned chemical shift information was not included for structural prediction.

4.8 Prediction of *Jd*LPMO10A secondary- and tertiary structure using computational methods:

No three-dimensional structure determination using distance restraints based on NOEs was performed for *Jd*LPMO10A. Instead, the computational methods TALOS-N (Shen and Bax, 2015) and CS-ROSETTA (Shen, *et al.*, 2009) were used to predict the dihedral angels ϕ/ψ and secondary structural elements, and to generate a tertiary structure of *Jd*LPMO10A, respectively.

4.8.1 Secondary structure prediction using the computational TALOS-N:

The secondary structural elements of *Jd*LPMO10A was predicted using TALOS-N (Shen and Bax, 2015). The overall secondary structure consisted of two helixes and seven β-strands, organized in the following manner: $helix_1$(41-47)- $helix_2$(78-83)- β_1(92-96)- β_2(98-104)- β_3(112-118)- β_4(121-128)- β_5(137-142)- β_6(148-156)- β_7(162-172). The numbers inside the parenthesis denote the residues in the primary structure comprising each secondary structural element. Illustrations of the secondary structural elements predicted by TALOS-N, together with the secondary structural elements found in the crystal structure of *Jd*LPMO10A (Mekasha *et al*, 2016), are shown in figure 4.10.

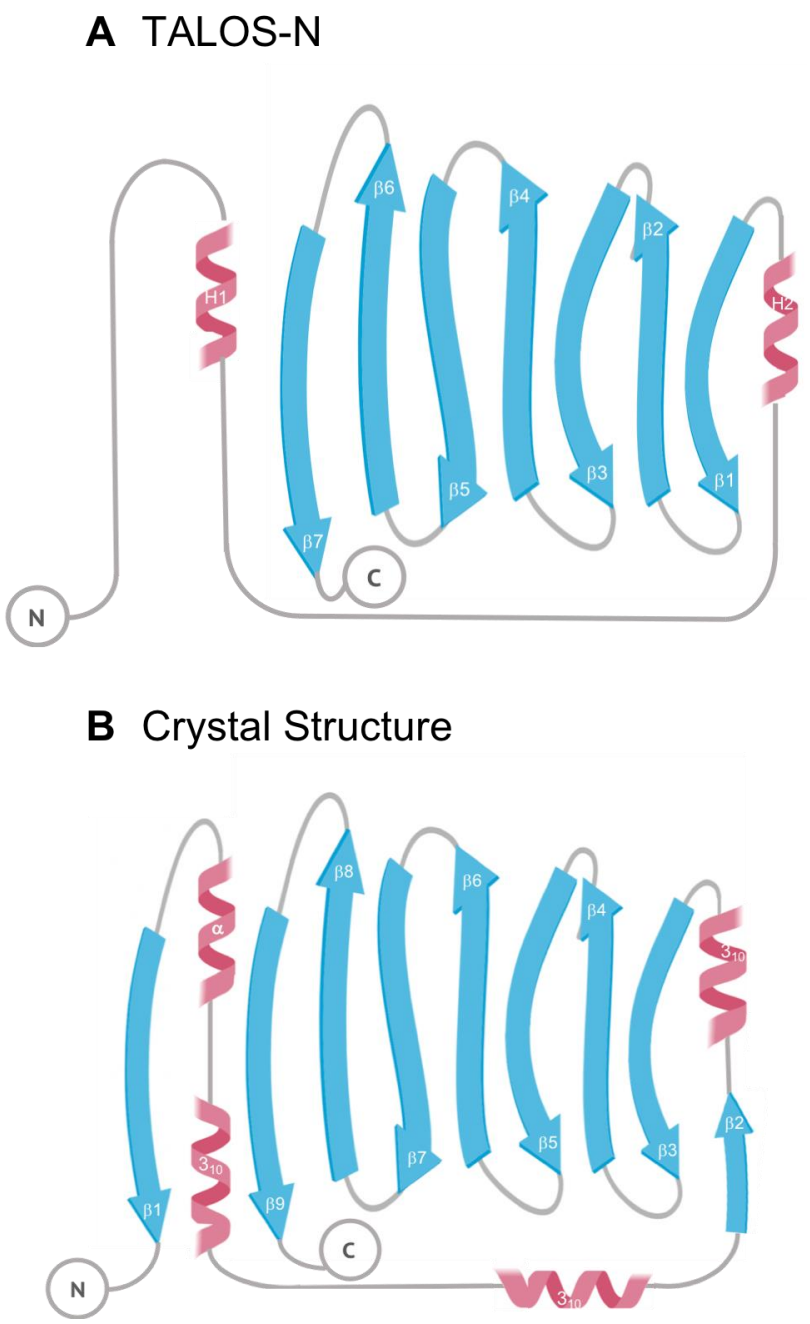

Figure 4.10: Illustration of the secondary structural elements of *Jd*LPMO10A. In the figure, helixes are depicted as pink coils, and β-sheets are depicted as blue arrows. A) The secondary structural elements predicted by TALOS-N includes seven β-strands and two helixes. B) the secondary structural elements from the crystal structure includes nine β-strands and four helixes (Mekasha *et al*, 2016; pdb file 5aa7). The figure was made by Christensen, I. A.

The secondary structural elements of *Jd*LPMO10A predicted by TALOS-N are in agreement with the crystal structure found by Mekasha *et al*. (Mekasha *et al*., 2016). However, as seen in figure 4.10, the following secondary structural elements found in the crystal structure were not predicted by TALOS-N: β-strand 1, β-strand 2, 3_{10}-helix 1, and 3_{10}-helix 2. The absence

of these secondary structural elements can be explained by residues without assigned spin systems in the primary sequence of *Jd*LPMO10A.

TALOS-N utilizes a combination of the assigned chemical shifts for $^{13}C_{\alpha}$, $^{13}C_{\beta}$, ^{13}CO, N, H^N, and $^{1}H_{\alpha}$ for each spin system, in combination with the primary sequence of the protein, to predict secondary structural elements. Consequently, the certainty of the prediction decreases for segments in the polypeptide chain without assigned chemical shifts (Shen and Bax, 2015). Looking at figure 4.10, the missing secondary structural elements are found in segments of the primary structure which are either completely unassigned or only partly assigned. In the segment constituting β-strand 1 (Gly33-Asp37) only one of the residues were assigned a spin system. No residues were assigned in 3_{10}-helix 2 (Gly54-Tyr58). With regard to 3_{10}-helix 1 (Pro60-Ser63) and β-strand 2 (Glu74-Pro76), 40% and 33.3% of the residues were left without assigned spin systems, respectively. In addition, the three-dimensional triple resonance experiment H(CA)NH, used to assign the $^{1}H_{\alpha}$ atoms in the peptide backbone was not recorded with *Jd*LPMO10A. Further decreasing the certainty of the prediction by TALOS-N.

As no NOE correlations were obtained for *Jd*LPMO10A, it is not possible to validate the secondary structure predictions obtained using TALOS-N. TALOS-N has been demonstrated to predict secondary structures with an accuracy of 88% when supplied with assigned chemical shifts in addition to the primary structure of the protein (Shen and Bax, 2015). However, the accuracy of the method decreases to 81% if no chemical shifts are assigned (Shen and Bax, 2015). In addition to providing a prediction of secondary structural elements, TALOS-N also evaluates the certainty of the prediction. The certainty of the predicted secondary structural element for each residue is shown in figure 4.11. The figure demonstrates how the certainty of the secondary structure prediction by TALOS-N decreases in segments where few – or none of the residues were assigned a spin system.

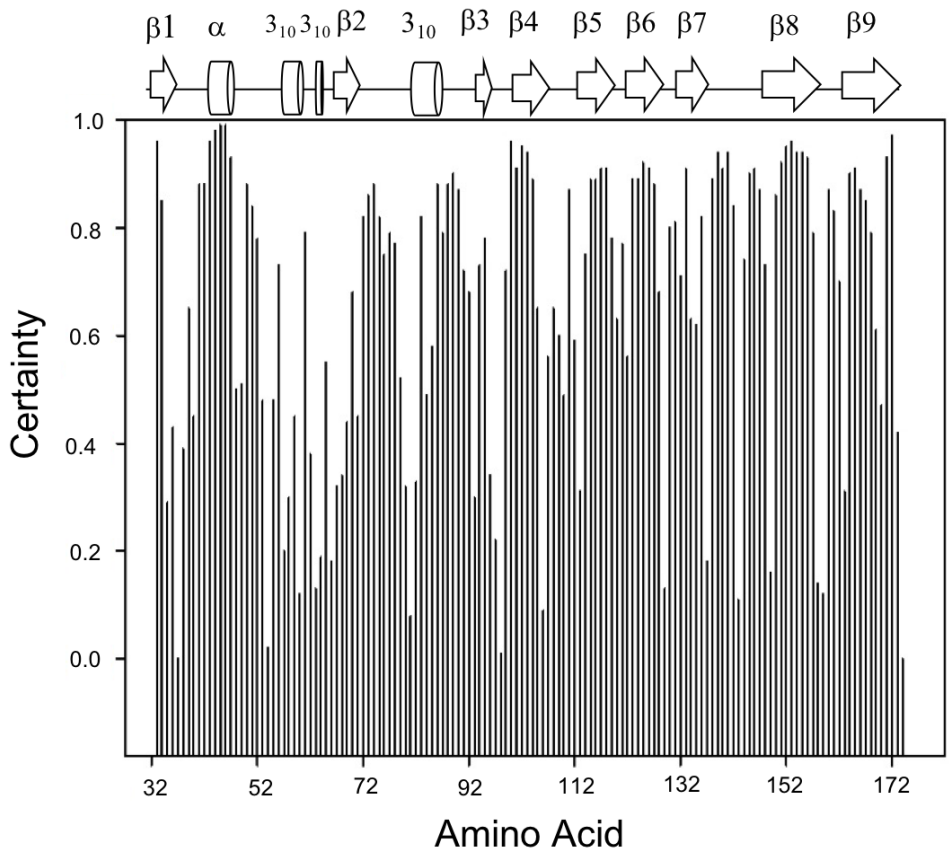

Figure 4.11: Certainty of the secondary structure prediction by TALOS-N for each amino acid residue. The secondary structural elements of *Jd*LPMO10A obtained from the crystal structure (Mekasha *et al.*, 2016; pdb file 5aa7) is placed over the corresponding segments of the primary structure. The figure was made by Christensen, I. A.

4.8.2 Prediction, and comparison, of Dihedral angles

In addition to prediction of secondary structural elements, TALOS-N was used to predict the dihedral angels ϕ/ψ for *Jd*LPMO10A, with a mean standard deviation calculated to be $12.09 \pm 2*5.76$ and $12.02 \pm 2*6.65$, respectively (Appendix B). To validate the structural information obtained using NMR, the dihedral angels ϕ/ψ predicted by TALOS-N were compared with the corresponding dihedral angels obtained from the crystal structure (pdb file 5aa7), as shown in figure 4.12.

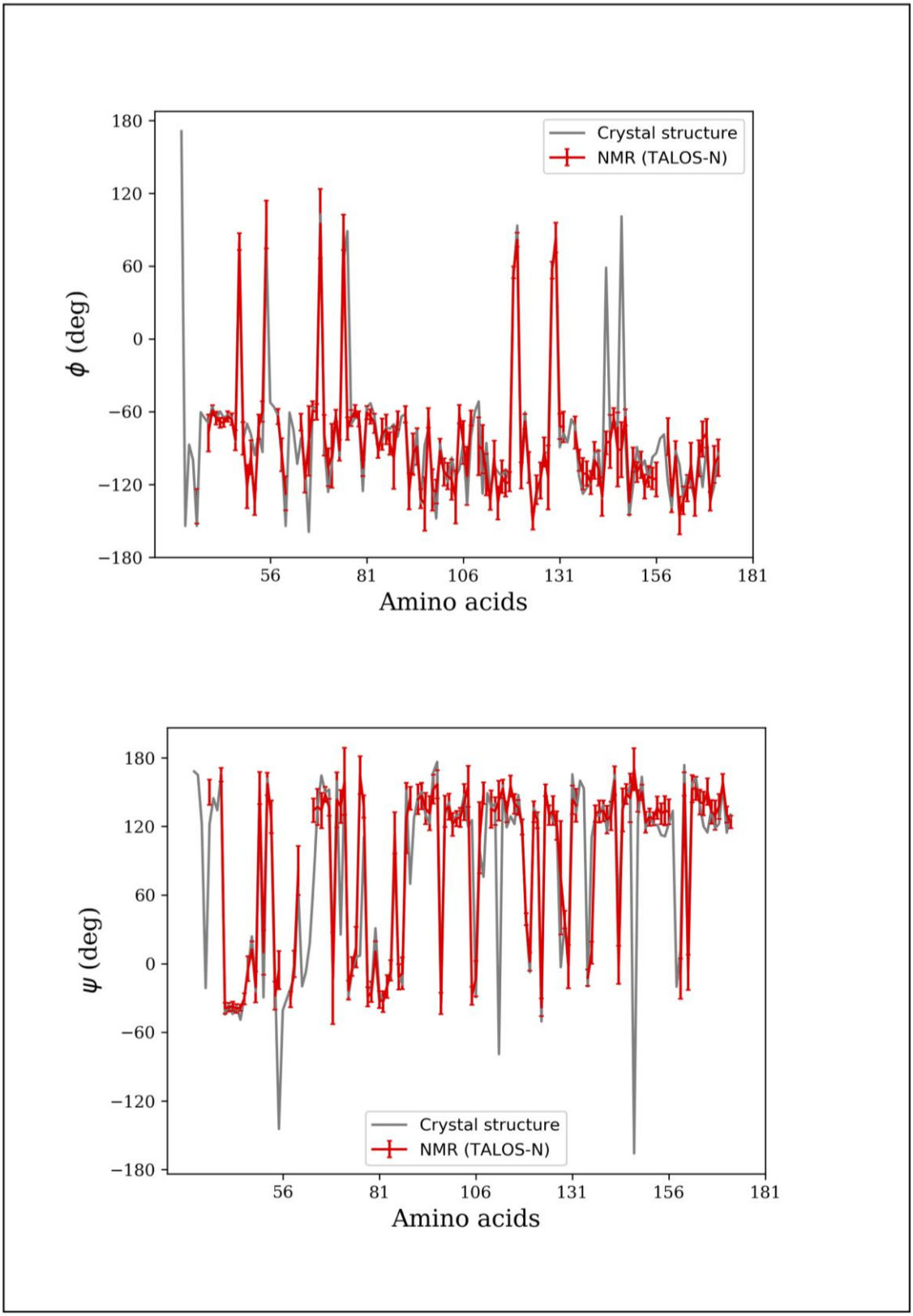

Figure 4.12: Comparison of the dihedral angles phi (ϕ) and psi (ψ) obtained for *Jd*LPMO10A using TALOS-N (based on chemical shifts determined by NMR) and X-ray crystallography (Mekasha *et al.*, 2016; pdb file 5aa7) plotted against amino acid residue number (*i*). The dihedral angles determined using TALOS-N are presented together with the corresponding standard deviation estimated for each residue. The figure was made by Christensen, I. A.

The root-mean square deviation (RMSD) difference of ϕ/ψ predicted by TALOS-N compared with the dihedral angels obtained from the crystal structure was calculated to be 16.65 and 28.87 degrees, respectively (Appendix B). The RMSD is frequently used to assess the difference between predicted- and observed values. In general, low RMSD values indicate a good fit between the predicted and observed values. However, as the RMSD is proportional to the square difference between the predicted and observed values, it is sensitive to outliers (Tramontano., 2006).

The effect of outliers might explain the large RMSD difference found for ψ. Large differences were found between the values of ψ for residues Gly55, Thr71, Gly76, Trp89, Leu105, Pro108, and Ser136. In addition, a large difference was found for Ser66 with regard to ϕ. The differences between the estimated values by TALOS-N and the corresponding values obtained from the crystal structure are shown in figure 4.13.

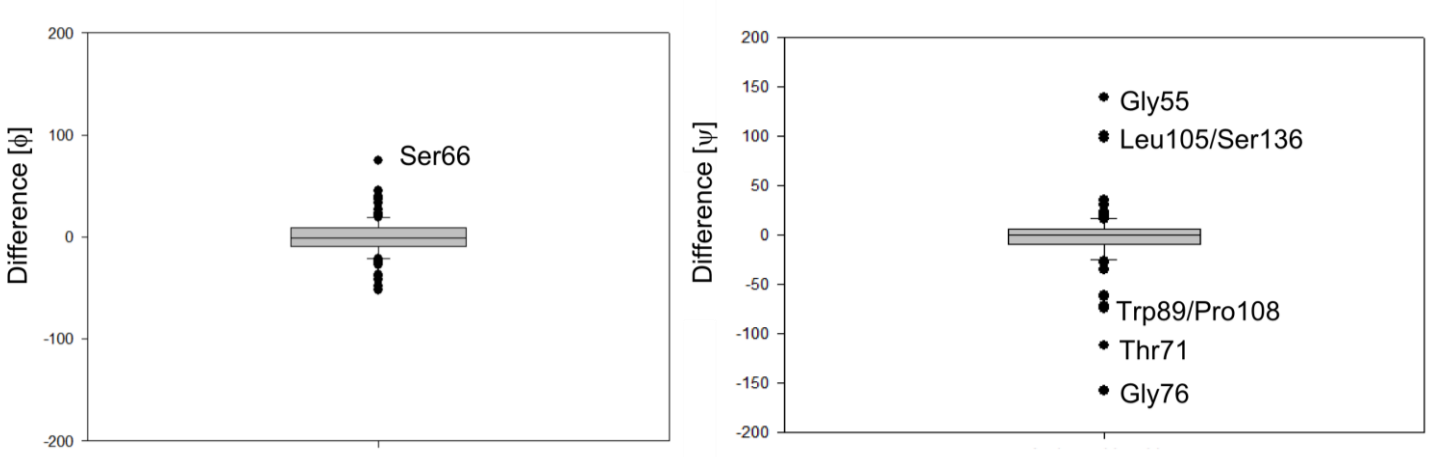

Figure 4.13: Box plots of the differences between the predicted dihedral angles ϕ/ψ by TALOS-N and the corresponding dihedral angels obtained from the crystal structure (pdb 5aa7). The figure was made by Christensen, I. A.

Visual inspection showed that the differences between the predicted values, and the corresponding values from the crystal structure, was more than 60 degrees for residues Gly55, Ser66, Thr71, Gly76, Trp89, Leu105, Pro108, and Ser136. The residues were therefore defined as outlier values. The deviations found for residues Ser66, Gly76, Thr71, Trp89, and Pro108 can be explained by the fact that the residues are found in segments of *Jd*LPMO10A without any regular secondary structure, where the chemical shift signature of the structure can be ambiguous (Shen and Bax, 2015). In addition, glycine has the greatest conformational flexibility of the 20 common amino acids, which might explain the differences found for

Gly55 (Nelson and Cox, 2013). For the remaining residues, it is not possible to comment with the current results whether the deviations are the result of erroneous assignments of residues in the protein, the quality of the reference structure, or the quality of the prediction by TALOS-N.

Removing the outlier differences the modulated RMSD of ϕ/ψ was found to be 15.53 and 16.08 degrees, respectively (Appendix B). Suggesting that bulk of the dihedral angels predicted by TALOS-N are in good agreement with the corresponding dihedral angels obtained from the crystal structure, when considering, the mean RMSD obtained using TALOS-N is 12 degrees (Shen and Bax, 2013).

4.8.3 Prediction of tertiary structure using the computational method CS-ROSETTA

CS-ROSETTA (Shen *et al*., 2009) was used to generate a model of *Jd*LPMO10As tertiary structure based on chemical shift assignment. In total, 3000 de novo structures were generated for *Jd*LPMO10A. To identify the generated structure best matching the crystal structure, the root-mean-square deviation (RMSD) was calculated for all the predicted structures with respect to the $^{13}C_{\alpha}$ atoms found in the polypeptide backbone. Of the 300 structures, structure number 2492 to was found to be the best match, with an RMSD of 3.21. In addition, structure number 2492 was found to be the lowest energy structure by CS-ROSETTA. Structure number 2492, referred to as the CS-ROSETTA model of *Jd*LPMO10A from here onwards, is shown in figure 4.14. Figure 4.14 also shows the CS-ROSETTA model of *Jd*LPMO10A superimposed on the crystal structure (Mekasha *et al*, 2016; pdb 5aa7).

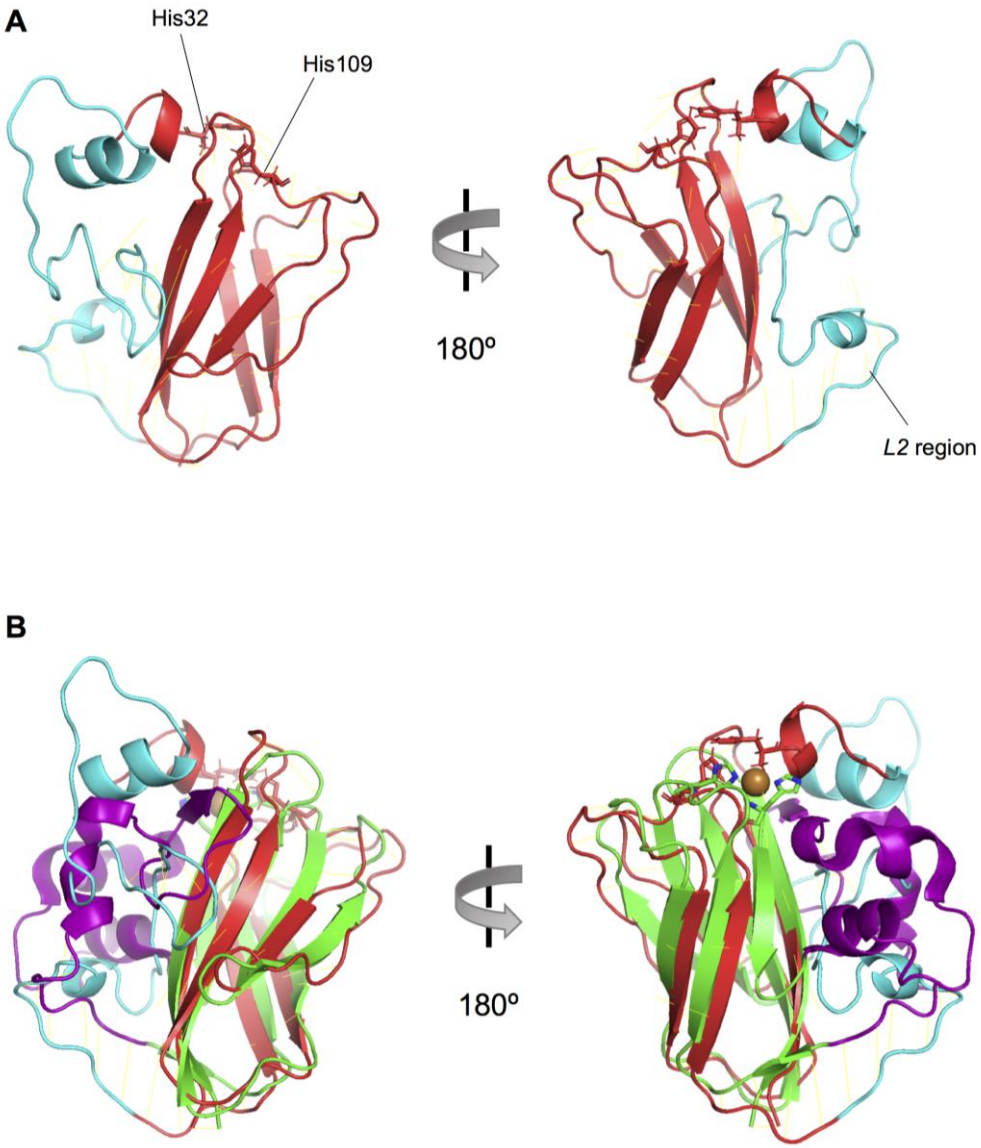

Figure 4.14: Cartoon representation of**; A)** The CS-ROSETTA model of *Jd*LPMO10A generated by CS-ROSETTA using experimentally obtained data from NMR. **B)** The CS-ROSETTA model of *Jd*LPMO10A (immunoglobulin-like β-sandwich core in red, *L2* region in aqua) superimposed on the crystal structure of *Jd*LMO10A (immunoglobulin-like β-sandwich in green, *L2* region in purple) found by Mekasha *et al*. (Mekasha *et al*, 2016; pdb file 5aa7). The figure was made by Christensen, I. A.

The CS-ROSETTA model of *Jd*LPMO10A displays a segment similar to the *L2*-region found in the crystal structure (His32-Ala96), succeeding the N-terminal histidine (His32). The *L2*-region consists of three helixes, H_1(Gly33-Val35), H_2(Arg41-Ser47), and H_3(Ala81-Ile82), connected by several long loop segments. The structure further extends into a

immunoglobulin-like β-sandwich core structure, which is composed of two β-sheets made up of five β-strands connected by loops (Thr98-Val171). The second conserved histidine residue (His109), is found in the loop segment connecting β-strand 1 with β-strand 2 (Leu105-Trp114). Disulfide bonds are found between residues Cys45 and Cys54, -and Cys73 and Cys167, all of which reside in loop segments. The CS-ROSETTA model of *Jd*LPMO10A ends with the C-terminal residues (Val172-Ser173).

Comparing the CS-ROSETTA model of *Jd*LPMO10A with the crystal structure (Mekasha *et al*, 2016), both have the same overall fold of their immunoglobulin-like β-sandwich core. However, while the β-sandwich core of the crystal structure is made up of seven β-strands, only five β-strands are present in the CS-ROSETTA model which is missing β-strand 1 and β-strand 6. In addition, no secondary structural elements corresponding to β-strand 2 and β-strand 3 is present in the CS-ROSETTA model. The segment corresponding to β-strand 5 in the crystal structure (Thr111-Val118) is shorter in the CS-ROSETTA model (Glu115-Val118). Further, only two of the four helixes found in the *L2* region can be observed, with 3_{10}-helix 1 and 2 found in the crystal structure absent from the CS-ROSETTA model. The segment corresponding to 3_{10}-helix is also notably shorter in the CS-ROSETTA model (Ala81-Ile82) compared with the crystal structure (Glu78-Asp84). Moreover, the overall shape of the *L2* region is dissimilar between the structures, being largely composed of loops in the CS-ROSETTA model.

The positions of the two conserved histidine residues found in the copper active site (His32 and His109), are similar between the CS-ROSETTA model and the crystal structure, as shown in figure 4.15.

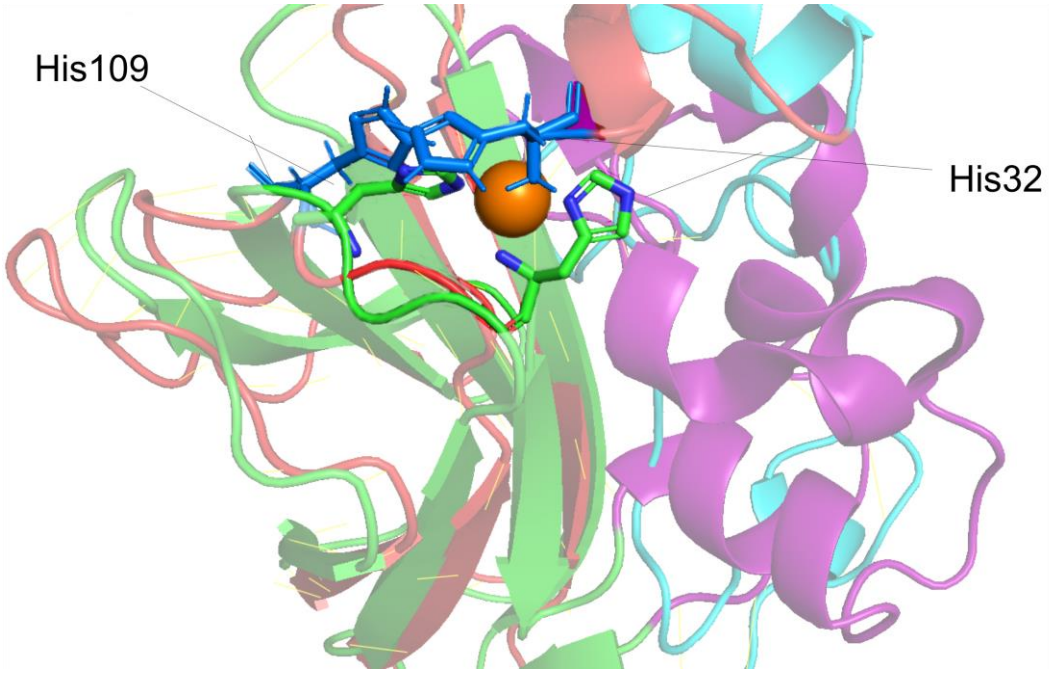

Figure 4.15: Stick illustration of the active site histidine brace motif found in the CS-ROSETTA model of *Jd*LPMO10A generated using CS-ROSETTA (Histidine residues marked in blue) data superimposed on the active site found in the crystal structure of *Jd*LPMO10A (Histidine residues marked in green) found by Mekasha *et al* (Mekasha *et al*, 2016; pdb file 5aa7). The figure was made by Christensen, I. A.

Mekasha *et al*. identified the surface exposed residues His32, Tyr59, Glu60, Gln62, Ser63, Glu65, Thr106, Ala107, His109, Asn159, Thr160, and Asn162, as candidates involved in substrate binding. All of which were located on *Jd*LPMO10A flat, substrate binding surface (Mekasha *et al*., 2016). The location of these residues in both the crystal- and CS-ROSETTA model is shown in Figure 4.16. Comparing the crystal structure to the CS-ROSETTA model, residues His32, Thr106, Ala107, His109, Asn159, Thr160, and Asn162 are found on the flat binding surface of the CS-ROSETTA model. However, the displacement of the N-terminal region (His32-Ala96) have led to residues Tyr59, Glu60, Ser63, Glu65, and, Asn77 to reside adjacent to the three stranded β-sheet in the core structure.

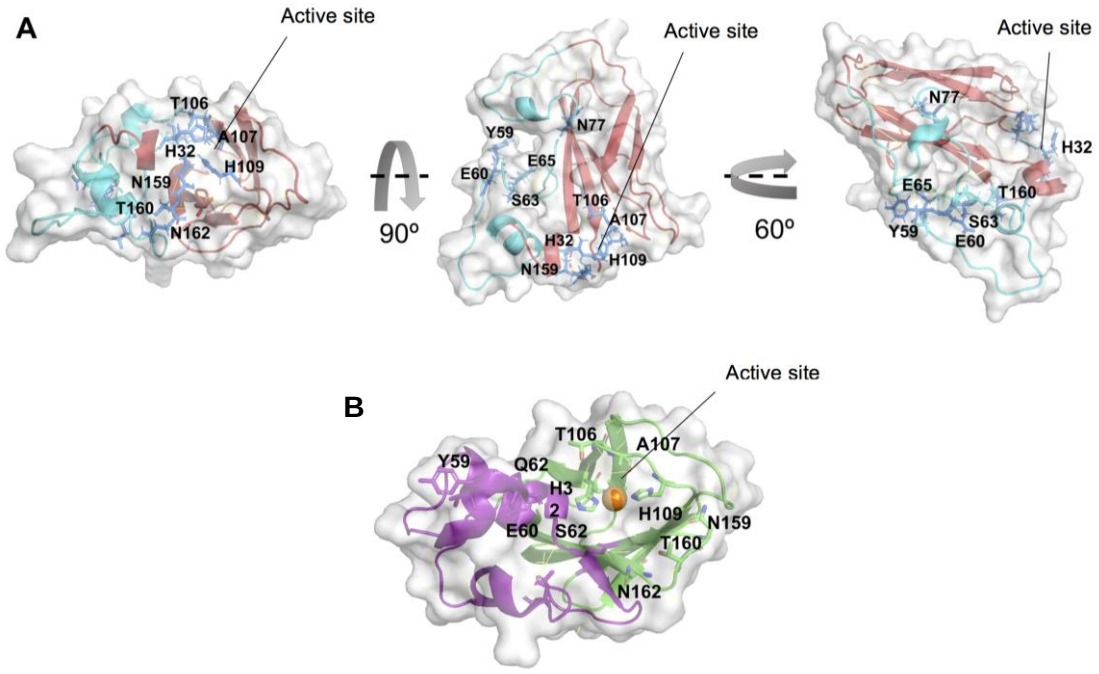

Figure 4.16: Comparison of the CS-ROSETTA model and crystal structure (Mekasha *et al*., 2016; pdb file 5aa7) of *Jd*LPMO10A. **A)** The location of residues His32, Tyr59, Glu60, Gln62, Ser63, Glu65, Thr106, Ala107, His109, Ans159, Thr160, and Ans162 in the CS-ROSETTA model of *Jd*LPMO10A generated by CS-ROSETTA. Residues His32, Thr106, Ala107, His109, Asn159, Thr160, and Asn162 are found on the flat binding surface of the enzyme, while residues Tyr59, Glu60, Ser63, Glu65, and Asn77 are found adjacent to the three stranded β-strand in the immunoglobulin-like β-sandwich core. **B)** The location of the residues His32, Tyr59, Glu60, Gln62, Ser63, Glu65, Thr106, Ala107, His109, Ans159, Thr160, and Ans162 in the crystal structure of *Jd*LPMO10A, all the residues are found on the enzyme the flat binding surface. The figure was made by Christensen, I. A.

The CS-ROSETTA protocol uses a combination of the assigned chemical shifts $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, ^{15}N, $^{1}H\alpha$ and H^N and the ROSETTA algorithm to generate protein structures (Shen *et al* 2009). The ROSETTA algorithm relies on the assumption that fragments of similar amino acid residues, with similar chemical shifts, will assume similar orientations in solution. In short, the ROSETTA algorithm predicts protein structure by matching small fragments of the proteins primary structure with similar protein fragments found in the ROSETTA database. The fragments are used as an indication of the conformation in a given segment of the primary structure and is further used to generate low energy structures of the protein in question. In CS-ROSETTA, the assigned chemical shifts $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO, ^{15}N, $^{1}H\alpha$, and H^N are used to increase the accuracy of the fragments by the ROSETTA algorithm (Shen, *et al*., 2009).

Consequently, the accuracy of the structure generated by CS-ROSETTA decreases for segments of the primary structure without any assigned chemical shift information, as sequence similarities become the only remaining criterion for structural predictions (Shen *et al., 2009)*. This holds especially true for shorter fragments (3-residues), where one residue without an assigned spin system corresponds to a 33% loss of information about the segment. An even more challenging situation occurs when residues without assigned spin systems are clustered along the protein sequence (Shen *et al*., 2009). As in the case of TALOS-N, no chemical shift information was obtained for any of the $^{1}H_{\alpha}$ atoms in the backbone of *Jd*LPMO10A, which might have further affected the accuracy of the generated CS-ROSETTA model.

The importance of assigned chemical shift information becomes evident considering that CS-ROSETTA generated a small helix segment (Tyr34-36) in place of β-strand 1 (Gly33-Asp37) in the CS-ROSETTA model. As shown in figure 4.10, only 20% of the residues comprising β-strand 1 was assigned. Incomplete- or no assigned chemical shifts for residues in the segments of the polypeptide chain comprising 3_{10}-helix 1, 3_{10}-helix 2, and β-strand 2 might also explain why these secondary structural elements are absent in the CS-ROSETTA model, as in case of the secondary structural element prediction by TALOS-N.

Neither β-strand 3 or β-strand 6 found in the crystal structure were present in the CS-ROSETTA model. This is interesting, as the segments of the primary structure comprising these secondary structural elements were all assigned spin systems. Moreover, secondary structural elements corresponding to β-strand 3 and β-strand 6 were predicted by TALOS-N (section 4.8). Although, erroneously assigned residues or interchange short fragments during the assignment cannot be discarded as a potential explanation, the missing secondary structural elements suggest that the CS-ROSETTA protocol can provide suboptimal result when used in combination with datasets without complete chemical shift assignment.

Future work:

The production of recombinant expressed *Jd*LPMO10A was suboptimal in M9-medium with glucose as the sole carbon source. The suboptimal production can be explained by a number of reasons, with acetate accumulation and changes in the pH of the growth medium being prominent candidates. It could, therefore, be beneficial to attempt new variations of the cultivation method in order to minimize the effects of acetate accumulation and changes in the pH of the growth medium. Acetate accumulation might result from oxygen depletion, glucose overfeeding and high growth rates (Chou, 2007). Cultivating the cells in a LEX bioreactor at 16 °C with reduced amounts of glucose (< 4 g/L) could therefore be interesting. Changes in pH could be minimalized by cultivating the transformed cells in variants of the traditional M9-medium with higher buffer capacities (Azatian, Kaur, Latham, 2019). Higher levels of protein production might, in turn, make it possible to assign the remaining 8.4% chemical shifts of ^{13}CO, $^{13}C\alpha$, ^{15}N, and $^{1}H^{N}$ in *Jd*LPMO10As backbone. Further, the three-dimensional tipple resonance experiment H(CA)NH could be recorded to assign the $^{1}H\alpha$ atoms in the backbone of *Jd*LPMO10A.

It would also be interesting to complete the structural elucidation of *Jd*LPMO10A using the traditional method described in figure 1.14, beginning with NOE assignment, and structural calculations. Further structural information obtained by NMR would also make it easier to comment on the performance of the computational methods TALOS-N and CS-ROSETTA. Once a complete NMR structure of *Jd*LPMO10A has been obtained, it would be interesting to study the activity of *Jd*LPMO10A by NMR spectroscopy. As discussed in section 1.5 (and Appendix A), *Jd*LPMO10A carries two major deletions compared with the much studied AA10 *Sm*LPMO10A. However, Mekasha *et al*., found the two enzymes to operate at comparable rates on β-chitin. Further insight into how this small LPMO interacts with its substrate could better our understanding of how AA10 functions in general.

Conclusion

Both ^{15}N- and ^{15}N/^{13}C-labeled samples of *Jd*LPMO10A were successfully produced using the recombinant expression vector pJB_SP_*Jd*. *Jd*LPMO10A was successfully purified using a two-step purification protocol of anion-exchange chromatography followed by Size Exclusion Chromatography to obtain pure sample for structural investigations by NMR. Both two-dimensional and three-dimensional NMR experiments were recorded with isotopically labeled samples of *Jd*LPMO10A. Using the sequence-specific assignment method, 91.6% of the backbone chemical shifts ^{13}CO, $^{13}C_{\alpha}$, ^{15}N and $^{1}H^{N}$ for *Jd*LPMO10A were assigned. The dihedral angles phi (ϕ) and psi (ψ), and the secondary structural elements found in *Jd*LPMO10A were predicted using the computational method TALOS-N. The predicted dihedral angels were generally in good agreement with the dihedral angels obtained from the crystal structure. Furthermore, the predicted secondary structural elements by TALOS-N were similar to findings from the crystal structure. However, TALOS-N was not able to predict secondary structural elements in the primary structure of *Jd*LMO10A with incomplete chemical shift assignment. This was especially true in the *L2* region of the enzyme.

A preliminary NMR structure of *Jd*LPMO10A was generated using CS-ROSETTA.
The overall fold of the immunoglobin like β-sandwich core was similar between the CS-ROSETTA model and the crystal structure. However, the several secondary structural elements found in the crystal structure were missing from the CS-ROSETTA model, both in the *L2* region and in the immunoglobin like β-sandwich core.

The CS-ROSETTA model is an important step towards an NMR structure of *Jd*LPMO10A. The assigned backbone chemical shifts can be used as a basis for further experimentation by NMR, which can be used to better our understanding of this minimal scaffold LPMO, and chitinolytic LPMOs in general.

References

Papers:

Aachmann, F. L. (2005) '*The NMR Structure of the R-module – A Parallel β-roll Subunit from Azotobacter vinelandii Alginate C-5 Epimerase'* Ph.D. thesis, Aalborg University Department of Life Sciences Section of Biotechnology and Norwegian University of Science and Technology Department of Biotechnology.

Aachmann, F. L. *et al* (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.

Aachmann, F. L. *et al*. (2015) 'Lytic Polysaccharide Monooxygenase', *Encyclopedia of Inorganic and Bioorganic Chemistry*, pp. 1-13. doi: 10.1002/9781119951438.eibc2270

Afroz, T. *et al*. (2014) 'Bacterial sugar utilization gives rise to distinct single-cell behaviours', Molecular Microbiology 93(6), pp. 1093-1103. doi: 10.111/mmi.12695.

Azatian, S. B., Kaur, N, and Latham, M. P. (2019) 'Increasing the buffering capacity of minimal media leads to higher protein yield', *Journal of Biomolecular NMR* 73(1-2), pp. 11-17. doi: 10.1007/s10858-018-00222-4.

Bacik, J. P. *et al* (2017) 'Neutron and Atomic Resolution X-ray Structures of a Lytic Polysaccharide Monooxygenases Reveal Copper-Mediated Dioxygen Binding and Evidence for N-Terminal Deprotonation', 56(20), pp. 2529-2532. doi: 10.1021/acs.biochem.7b00019.

Baneyx, F. and Mujacic, M. (2004) 'Recombinant protein folding and misfolding in Escherichia coli', *Nature Biotechnology* 22(11), pp. 1399-1408. doi: 10.1038/nbt1029

Beeson, W. T. *et al.* (2012) 'Oxidative Cleavage of Cellulose by Fungal Copper-Dependent Polysaccharide Monooxygenases', *Journal of the American Chemical Society*, 134(2), pp. 890-892. doi: 10.1021/ja2110657t.

Beeson, W. T. *et al*. (2015) 'Cellulose Degradation by Polysaccharide Monooxygenases', *Annual Review of Biochemistry*, 84(1), pp. 923-946. doi: 10.1146/annurev-biochem-060614-034439.

Bertini, I. and Pierattelli, R. (2004) 'Copper(II) proteins are amenable for NMR investigations', *Pure and Applied Chemistry*, 76(2). pp. 321-333. doi: 10.1351/pac200476020321

Bissaro, B. *et al*. (2017) 'Oxidative cleavage of polysaccharides by monocopper enzymes

depends on H_2O_2', *Nature Chemical Biology* 13(10), pp. 1123-1128. doi: 10.1038/nchembio.2470.

Bissario, B. *et al*. (2018) 'How a Lytic Polysaccharide Monooxygenase Binds Crystalline Chitin', *Biochemistry*, 57, pp. NA. doi: 10.1021/acs.biochem.8b00138

Book, A. J. *et al*. (2014) 'Evolution of substrate specificity in bacterial AA10 lytic polysaccharide monooxygenases', *Biotechnology for Biofuels* 7(1), pp. NA. doi: 10.1186/1754-6834-7.109.

Borisova, A. S. *et al*. (2015) 'Structural and Functional Characterization of a Lytic Polysaccharide Monooxygenase with Broad Substrate Specificity', The Journal of biological chemistry 290(38), pp 22955-22969. doi: 10.1074/jbc.M115.660183.

Brigham, C. J. (2017) 'Chitin and Chitosan: Sustainable, Medically Relevant Biomaterials', *International Journal of Biotechnology for Wellness Industries.* Available at: http://www.fao.org/fishery/facp/USA/en (Accessed: 11 June 2019)

Cannella, D. *et al*. (2012) 'Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content', *Biotechnology for Biofuels*, 5(1). Pp NA. doi: 10.1186/1754-6834-5-26

Chou, C. P. (2007) 'Engineering cell physiology to enhance recombinant protein production in Escherichia coli', *Applied Microbiology and Biotechnology* 76(3), pp. 521-532. doi : 10.1007/s00253-007-1039-0.

Ciano, L. *et al*. (2018) 'Bracing copper for the catalytic oxidation of C-H bonds', *Nature Catalysis* 1(8), pp. 571-577. doi 10.1038/s41929-018-0110-9.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) 'Protein backbone angle restraints from searching a database for chemical shifts and sequence homology' *Journal of Biomolecular NMR* 13(3), pp. 289-302. doi: 10.1023/A:1008392405740.

Courtade, G (2018) '*An NMR investigation of lytic polysaccharide monooxygenases'*, Ph.D. Thesis, Norwegian University of Science and Technology Department of Biotechnology and Food Science.

Courtade, G. *et al*. (2017) 'A novel expression system for lytic polysaccharide monooxygenases', *Carbohydrate Research*. Elsevier, 448, pp. 212-219. doi: 10.1016/J.CARRES.2017.02.003.

Courtade, G. *et al*. (2016) 'Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase' *Proceedings of the National Academy of Sciences* 133(21), pp. 5922-5927. doi: 10.1073/pnas.1602566113.

Couturier, M. *et al*. (2018) 'Lytic xylan oxidases from wood-decay fungi unlock biomass degradation', *Nature Chemical Biology* 14(3), pp.306-310. doi: 10.1038/nchembio.2558.

Crouch, L. I. *et al.* (2016) 'The Contribution of Non-catalytic Carbohydrate Binfing Modules to the Activity of Lytic Polysaccharide Monooxygenases', *Journal of the American Chemical Society* 291(14), pp. 7439-7449. doi: 10.1074/jbc.M115.702365.

Dahiya, N., Tewari, R. and Hoondal, G. S. (2006) 'Biotechnological aspects of chitinolytic enzymes: a review', *Applied Microbiology and Biotechnology* 71(6), pp. 773-782. doi: 10.1007/s002553-005-0183-7.

Danneels, B., Tanghe, M. and Desnemet, T. (2019) 'Structural Features on the Substrate Binding Surface of Fungal Lytic Polysaccharide Monooxygenases Determine Their Oxidative Regioselectivity', *Biotechnology Journal* 14(3), pp. NA. doi: 10.1002/biot.201800211.

Edwards, A. J. and Reid, D. (2000) 'Introduction to NMR of Proteins', *Current Protocols in Protein Science*, pp. 17.5.1-17.5.39. doi: 10.1002/0471140864.ps1705s19.

Eiteman, M. A. and Altman, E. (2006) 'Overcoming acetate in Escherichia coli recombinant protein fermentations', *Trends in Bioechnology* 24(11), pp. 530-536. doi: 10.1016/J.TIBTECH.2006.09.001.

Farewell, A. and Neidhardt, F. C. (1998) 'Effect of Temperatur on In Vivo Protein Synthetic Capacity in Escherichia coli', *Journal of Bacteriology* 180(17), pp. 4704-4710. Available at: http://jb.asm.org/content/180/17/4704.full (Accessed: 5 June 2019).

Fathi-Roudsari, M., Akhavian-Tehrani, A. and Maghsoudi, N. (1998) 'Comparison of Three Escherichia coli Strains in Recombinant Production of Reteplace', *Journal of Bacteriology*, pp. 4704-4710. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4717461/pdf/AJMB-8-16.pdf (Accessed: 17 April 2019).

Forsberg, Z. *et al.* (2019) 'Polysaccharide degradation by lytic polysaccharide monooxygenases', *Current Opinion in Structural Biology*, 59, pp 54-64. doi: 10.1016/J-SBI.2019.02.05.

Frueh, D. P. (2004) 'Practical aspects of NMR signal assignment in larger and challenging proteins', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 78, pp. 47-75. doi: 10.1016/J.PNMRS.2013.12.001.

Gawin, A., Valla, S. and Brautaset, T. (2017) 'The XylS/*Pm* regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production

and metabolic engineering', *Microbial Biotechnology* 10(4), pp. 702-718. doi: 10.1111/1751-7915-12701.

Georgiou, G. and Segatori, L. (2005) 'Preparative expression of secreted proteins in bacteria: status report and future prospects' *Current Opinion in Biotechnology* 16(5), pp. 538-545. doi: 10.1016/j.copbio.2005.07.008.

Hamid, R. *et al*. (2013) 'Chitinases: An update', *Journal of pharmacy & bioallied sciences* 5(1), pp. 21-29. doi: 10.4103/0975-7406.106559.

Harris, P. V. *et al.* (2010) 'Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Engimatic Family', *Biochemistry* 49(15), pp. 3305-3316. doi: 10.1021/bil100009p.

Hemsworth, G. R. *et al.* (2014) 'Discovery and characterization of a new famliy of lytic polysaccharide monooxygenases', *Nature Chemical Biology* 10(2), pp. 122-126. doi: 10.1038/nchembio.1417.

Hemsworth, G. R. *et al.* (2015) 'Lytic Polysaccharide Monooxygenases in Biomass Conversion', *Trends in Biotechnology* 33(12), pp. 747-761. doi: 10.1016/J.TIBTECH.2015.09.006.

Karkehabadi, S. *et al.* (2008) 'The First Structure of a Glycoside Hydrolase Family 61 Member, Cel61B from Hypocrea jecorina, at 1.6 Å Resolution', *Journal of Molecular Biology* 383(1), pp, 144-154. doi: 10.1016/j,jmb.2008.08.016.

Keller, R. L. J. (2004) 'The Computer Aided Resonance Assignment Tutorial', Avaliable at: http://cara.nmr-software.org (Accessed: 10 June 2019).

Kwan, A. H. *et al*. (2011) 'Macromolecular NMR spectroscopy for the non-spectroscopist', *FEBS Journal* 278(5), pp. 687-703. doi: 10.1111/j.1742-4658.2011.08004.x.

Lee, S. Y. (1996) 'High cell-density culture of Escherichia coli', *Trends in Biotechnology* 14(3), pp. 98-105. doi: 10.1016/0167-7799(96)80930-9.

Lo Leggio, L. *et al.* (2015) 'Structure and boosting activity of a strach-degrading lytic polysaccharide monooxygenase', *Nature Communications* 6(1) pp. 5961-5970. doi: 10.1038/ncomms6961.

Levasseur, A. *et al.* (2013) 'Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes', *Biotechnology for biofuels* 6(1), pp 41-55. doi: 10.1186/1754-6834-6-41.

Li, X. *et al.* (2012) 'Structural Basis for Substrate Targeting and Catalysis by Fungal

Polysaccharide Monooxygenases', *Structure* 20(6), pp. 1051-1061. doi: 10.1016/j.str.2012.04.002.

Loose, J. S. M. (2016) '*Biochemical investigation of catalysis by lytic polysaccharide monooxygenases*', Ph.D. thesis, Norwegian University of Life Sciences, Ås. Available at: http://brage.bibsys.no/xmlui/handle/11259/24977230 (Accessed: 23 November 2018).

Marisch, K. *et al.* (2013) 'A Comparative Analysis of Industrial Escherichia coli K-12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Levels' *PLos ONE* 8(8), pp. e70516. doi: 10.1371/journal.pone.0070516.

Mekasha, S. *et al.* (2016) 'Structural and functional characterization of a small chitin-active lytic polysaccharide monooxygenase domain of a multi-modular chitinase from Jonesia denitrificans', *FEBS Letters* 590(1), pp.34-42. doi: 10.1002/1873-3468.12025.

Murby, M., Uhlén, M. and Ståhl, S. (1996) 'Upstream Strategies to Minimize Proteolytic Degradation upon Recombinant Production in Escherichia coli', *Protein Expression and Purification* 7(2), pp. 129-136. doi: 10.1006/prep.1996.0018.

Palmer, I. and Wingfield, P. T. (2004) 'Preparation and Extraction of Insoluble (9nclusion-Body) Proteins from Escherichia coli', *Current Protocols in Protein Science*, pp. 6.3.1-6.3.18. doi: 10.1002/0471140864.ps0603s38.

Palomares L. A. *et al.* (2004) 'Production of Recombinant Proteins', *Recombinant Gene Expression. Methods in Molecular Biology* (267), pp. 15-51. doi: 10.1385/1-59259-774-2:015

Petrovic, D. M. *et al.* (2018) 'Methylation of the N-terminal histidine protects a lytic polysaccharide monooxygenase from auto-oxidative inactivation', *Protein Science* 27(9), pp. 1636-1650. doi 10.1002/pro.3451.

Quinlan, R. J. *et al.* (2011) 'Insight 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.

Reese, E. T., Siu R. G. H. and Levinson, H. S. (1950) 'The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis'. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC385789/pdf/jbacter00620-0053.pdf (Accessed 3 June 2019).

Rinaudo, M. (2006) 'Chitin and chitosan: Properties and applications', *Progress in Polymer Science* 31(7), pp. 603-632. doi: 10.1016/J.PROGPOLYMSCI.2006.06.001.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in Escherichia

coli: advances and challenges', Frontiers in Microbiology 5, pp. 172. doi: 10.3389/fmicb.2014.00172.

Sabbadin, F. *et al*. (2018) 'An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion', *Nature Communications* 9(1), pp. 756. doi: 10.1038/s41467-018-03142-x.

Schoepfer, R. (1993) 'The pRSET family of T7 promoter expression vectors for Escherichia coli', *Gene* 124(1), pp. 83-85. Doi:10.1016/0378-1119(93)90764-T.

Shen, Y. *et al*. (2009) 'TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts', *Journal of Biomolecular NMR* 44(4), pp. 213-223. doi: 10.1007/s10858-009-9333-z.

Shen, Y. and Bax, A. (2013) 'Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks', *Journal of Biomolecular NMR* 56(3), pp. 227-241. doi: 10.1007/s10858-013-9741-y.

Shen, Y. and Bax, A. (2015) 'Protein Structural Information Derived from NMR Chemical Shift with the Neural Network Program TALOS-N', *Cartwrigth H. (eds) Artifical neural networks. Methods in molecular biology* (1260), pp. 17-32. doi: 10.1007/978-1-4939-2239-0_2.

Shen, Y. and Bax A. (2009) '*DE novo* protein structure generation from incomplete chemical shift assignments', *Journal of Biomolecular NMR* 43(63), pp 63-78. doi: 10.1007/s10858-008-9288-5.

Sletta, H. *et al*. (2004) 'Broad-host-range plasmid pJB658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in Escherichia coli', *Applied and environmental microbiology* 70(12), pp. 7033-7039. doi: 10.1128/AEM/70.12.7033-7039.2004.

Sletta, H. *et al*. (2007) 'The Presence of N-Terminal Secretion Signal Sequences Leads to Strong Stimulation of the Total Expression Levels of Three Tested Medically Important Proteins during High-Cell-Density Cultivations of Escherichia coli', *Applied and environmental* microbiology 73(3), pp.906-912. doi: 10.1128/AEM.01804-06.

Sunny, N. E., Kumar, S. R. and Kumar, S. V. (2018) 'A Review on Chitinase Synthesis from varied sources and its Applications towards Environment', *Research J. Pharm. And Tech* 11(9). pp. NA. doi: 10.5958/0974-36X.2018.00770.9.

Vaaje-Kolstad, G. *et al.* (2004) 'Crystal Structure and Binding Properties of the Serratia marcescens Chitin-binding Protein CBP21', *Journal of Biological Chemistry* 280(12), pp. 11313-11319. doi: 10.1074/jbc.M407175200.

Vaaje-Kolstad, G. (2005) 'The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation', *Journal of Biological Chemistry* 280(31), pp. 28492-28497. 10.1074/jbc.M54468200.

Vaaje-Kolstad, G. (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. *et al.* (2017) 'Structural diversity of lytic polysaccharide monooxygenases', *Current Opinion in Structural Biology*, 44, pp. 67-76. doi:10.1016/j.sbi.2016.12.012.

Vick, J. E. *et al.* (2011) 'Optimized compatible set of BioBrick™ vectors for metabolic pathway engineering', *Applied Microbiology and Biotechnology* 92(6), pp. 1275-1286. doi: 10.1007/s00253-011-3633-4.

Vu, V. V. *et al.* (2014) 'Determinants of Regioselective Hydroxylation in the Fungal Polysaccharide Monooxygenases', *Journal of the American Chemical Society* 136(2), pp. 562-565. doi: 10.1021/ja409384b.

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.

Wang., B. Walton, P. H. and Rovira, C. (2019) 'Molecular Mechanisms of Oxygen Activation and Hydrogen Peroxide Formation in Lytic Polysaccharide Monooxygenases', *ACS Catalysis* 9, pp. 4958-4969. doi: 10.1021/acscatal.9b00778.

Wiithrich, K. (1990) 'Protein Structure Determination in Solution by NMR Spectroscopy', *Journal of Biological Chemistry*. Available at: http://www.jbc.org/content/265/36/22059.full.pdf (Accessed: 26 February 2019).

Winther-Larsen, H. C., Josefsen, K. D., Brautaset, T., Valla, S. (2000) 'Parameters Affecting Gene Expression from the Pm Promotor in Gram-Negative Bacteria', *Metabolic Engineering* 2(2), pp. 79-91. doi: 10.1006/mben.1999.0142.

Wo, J. *et al.* (2017) 'Application of an *E.coli* signal sequence as a versatile inclusion body tag', *Microbial Cell Factories* 16(1), pp. 50. doi: 10.1186/s12934-017-0662-4

Wu, M. *et al.* (2013) 'Crystal Structure and Computational Characterization of the Lytic Polysaccharide Monooxygenase GH61D from the Basidiomycota Fungus *Phanerochaete chrysosporium'*, 288(18), pp. 12828-12839. doi: 10.1074/jbc.M113.459396.

Books:

Alberts, B., Johnson, A., Lewis J., Morgan D., Raff, M., Roberts, K., Walter P., *Molecular Biology of The Cell* (6. Edition) United States of America, Garland Science.

Cavanagh, J., Fairbrother, W. J., Palmer III, A. G., Rance, M., Skelton, N. J. (2007) *Protein NMR Spectroscopy principle and practice* (2. Edition) United states of America; Elsevier Academic Press.

Creighton, T. E. (1984) *Proteins Structures and Molecular Properties* (1. Edition) United States of America; W. H. Freeman and Company.

Friebolin, H. (2011) *Basic One- and Two-Dimensional NMR Spectroscopy* (5. Edition). Gemany: Wiley-VCH Verlag GmbH & Co.

Nelson, D. L., Cox, M. M., (2013) *Lehninger Principles of Biochemistry* (6. Edition) United States of America; W. H. Freeman and Company.

Quincy, T. (2005) *Structural Biology practical NMR Applications* (1. Edition). United States of America: Springer Science + Business Media, Inc.

Smidsrød, O., & Moe, S. T. (2008) *Biopolymer chemistry* (1. Edition) Norway: Tapir Academy Press.

Snustad, D. P., Simmons, M. J., (2012) *Genetics* (6. Edition) United States of America, John Wiley & Sons, Inc.

Tramontano, A. (2006) *Protein Structure Prediction concepts and Applications* (1. Edition). Germany: Wiley-VCH Verlag GmbH & Co.

Additional online sources for figures:

Figure 1.12: The peptide bond, Moran, L. A., (published 3 April 2008), The peptide bond, available at: https://sandwalk.blogspot.com/2008/04/peptide-bond.html (Accessed March 2019).

Figure 1.13: Secondary structures, author unknown (year unknown), *Cell Biology By The Numbers. What is the Energy of a Hydrogen Bond*, available at: http://book.bionumbers.org/what-is-the-energy-of-a-hydrogen-bond/ (Accessed March 2019).

Appendix A: Comparison of *Jd*LPMO10A and *Sm*LPMO10A

*Jd*LPMO10A and the other "small" LPMOs found in Clad I, subclade C carries two major deletions in their primary structure compared to *Sm*LPMO10A (CBP21). The two deletions were found by Mekasha *et al*. by performing a structure-based sequence alignment of the structure of the AA10s: *Sm*LPMO10A (pdb file 2bem), *Ba*LPMO10A (pdb file 2yoy), *Ef*LPMO10A (pdb file 4alc), and *Jd*LPMO10A (pdb file 5aa7), using PyMOD (Mekasha *et al*., 2016). Figure A1 shows the structure-based sequence alignment of *Sm*LPMO10A and *Jd*LPMO10A. The two major deletions are colored in gray in the primary structure of *Sm*LPMO10A.

```
SmLPMO10A    28  HGYVESPASRAYQCKL-QLN-TQCGSVQYEPQSVEGLKGFP  67
JdLPMO10A    32  HGWVTDPPSRQALCASG-ETSFDCGQISYEPQSVEAPKG--  69

SmLPMO10A    68  QAG-PADGHIASADKSTFFELDQQTPTRWNKLNLKTGPNS  106
JdLPMO10A    70  --AT-----TCSGGNEAFAILDDNS-KPWPTTEIAS-TVD  100

SmLPMO10A   107  FTWKLTARHSTTSWRYFITKPNWDASQPLTRASF--DLTP  144
JdLPMO10A   101  LTWKLTAPHNTSTWEYFV----------------DGQL--  125

SmLPMO10A   145  FCQFNDGGAIPAAQVTHQCNIPADRSGSHVILAVWDIADT  184
JdLPMO10A   126  HQTFDQKGQQPPTSLTHTLTDL--PTGEHTILARWNVSNT  163

SmLPMO10A   185  ANAFYQAIDVNLSK                            198
JdLPMO10A   164  NNAFYNCMDVVVS-                            174
```

Figure A1: Structure-based sequence alignment of the AA10s *Sm*LPMO10A (pdb file 2bem) and *Jd*LPMO10A (pdb file 5aa7) adapted from results obtained by Mekasha *et al*. (Mekasha *et al*, 2016). The primary structure of *Jd*LPMO10A carries to major deletions compared to the primary structure of *Sm*LPMO10A. One is a deletion of 9 nine residues (or four and five consecutive deletions) after Gly69, the other a deletion of 16 residues after Val118 (Mekasha *et al*, 2016).

To further illustrate the two deletions in the primary structure of *Jd*LPMO10A, figure A1 shows the three-dimensional structure of *Jd*LPMO10A superimposed on the three-dimensional structure of *Sm*LPMO10A. The two deletions found in *Jd*LPMO10A compared to *Sm*LPMO10A are colored in yellow in the three-dimentional structure of *Sm*LPMO10A.

Figure A1: The structure of *Jd*LPMO10A (pdb file 5aa7) superimposed on structure of *Sm*LPMO10A (pdb file 2bem). The primary structure of *Jd*LPMO10A carries to major deletions compared to the primary structure of *Sm*LPMO10A. One is a deletion of 9 nine residues (or four and five consecutive deletions) after Gly69, the other a deletion of 16 residues after Val118 (Mekasha *et al*, 2016). The two deletions are colored yellow in the three-dimensional structure of *Sm*LPMO10A.

Appendix B: Calculations

As described in section 4.8.1 and 4.8.2, the computational method TALOS-N (Shen and Bax, 2015) was used to predict the dihedral angels ϕ and ψ in the primary structure of *Jd*LPMO10A.
The TALOS-N server is found at the following website:
(https://spin.niddk.nih.gov/bax/software/TALOS-N/)

TALOS-N provides an associated standard deviation for each predicted value of ϕ and ψ. The mean of the associated standard deviations provided by TALOS-N and its 95% confidence interval was calculated in order to evaluate the precision of the predicted dihedral angels. The mean of the associated standard deviations was calculated using equation I.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x1+x2+\cdots+xn}{n} \qquad \text{I}$$

$\bar{X}$ is the mean, n is the sample size and x_1, $x_2 \ldots x_n$ are the associated standard deviations of the sample.

The standard deviation of the mean standard deviation was calculated using equation II.

$$\sigma = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \qquad \text{II}$$

σ is the standard deviation, n is the sample size and $\bar{x}$ is the mean value of the sample.

To investigate the accuracy of the dihedral angels predicted by TALOS-N, the root-mean square deviation (RMSD) of the predicted dihedral angels relative to the dihedral angels obtain from the crystal structure was calculated using equation III.

$$RMSD = \frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}$$

Where y_t is the reference value, $\hat{y}_t$ is the predicted value, and T is the number of predictions.

Appendix C: Chromatograms

As described in section X, the periplasmic extract was purified using a two-step protocol of anion-exchange chromatography with a gradient elution, followed by size exclusion chromatography (SEC). The two-step purification protocol was sufficient to provide pure protein samples for NMR. Figure C1 shows the resulting chromatograms from the purification of the ^{15}N-labeled *Jd*LPMO10A sample, produced using the pUCBB_*Jd* plasmid

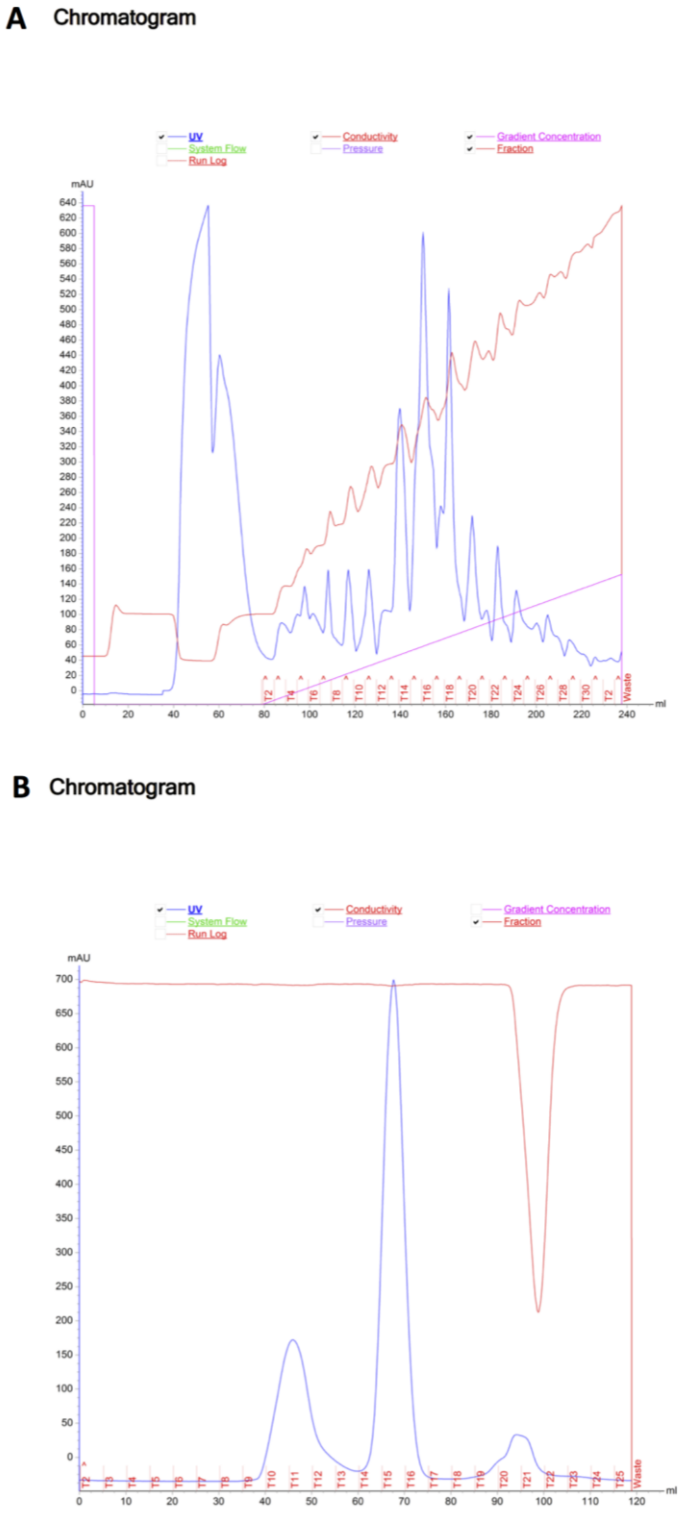

Figure C1: Figure 4.1: The chromatograms from the purification of the ^{15}N-labeled *Jd*LPMO10A sample. **A**) The chromatogram from the first purification step using a HitTrap- anion exchange column. The sampled eluted at approximately [12-16%] of elution Buffer B, which corresponds to fraction T14-T18 in the chromatogram. **B)** The chromatogram from second purification step using SEC chromatography. The sample eluted after 60 minutes, which corresponds to fraction T14-T16 in the chromatogram. In both chromatograms, the y-axis shows the UV-absorption signal in mAU (in blue), and the x-axis the fractionated volume in mL. The red lines in both spectra show the conductivity. The pink line in chromatogram **A** shows the % buffer B in the gradient elution.

Appendix D: Data

The data used to calculate the difference, and root-mean square deviation, of the dihedral angels phi (ϕ) and psi (ψ) predicted by TALOS-N and ϕ/ψfrom the crystal structure, and to calculate the mean standard deviation of the dihedral angels predicted by TALOS-N, are presented in table D1 and table D2.

Table D1: The data used to calculate the difference, and root-mean square difference between phi (ϕ) predicted by TALOS-N and ϕ from the crystal structure (pdb file 5aa7), and the mean standard deviation for the predicted values of ϕ by TALOS-N.

Amino acid	PHI CRYSTOLOGRAPHY	abs crystolography	PHI NMR	abs NMR	Standard deviation P	Difference TALOS-N/CRYS	Difference modulatec	Standard deviation differance
1								
2	171,0241102	171,0241102		0				17,89667544
3	-154,2207384	154,2207384		0				
4	-87,28444289	87,28444289		0				
5	-99,7737291	99,7737291		0				
6	-154,1542525	154,1542525	-137,942	137,942	14,23	16,21225251	-16,21225251	mean differance TALOS-N/CRYSTOLOGRAPHY
7	-60,60077574	60,60077574		0				
8	-65,46434318	65,46434318		0				-0,66
9	-69,51975822	69,51975822	-77,625	77,625	15,364	-8,105241781	8,105241781	
10	-55,6695813	55,6695813	-59,058	59,058	4,212	-3,388418702	3,388418702	
11	-65,75708772	65,75708772	-65,28	65,28	5,29	0,477087717	-0,477087717	
12	-59,86632965	59,86632965	-69,108	69,108	4,156	-9,241670347	9,241670347	mean standard deviation TALOS-N
13	-65,19927878	65,19927878	-68,986	68,986	3,562	-3,786721223	3,786721223	
14	-62,81724177	62,81724177	-64,161	64,161	4,301	-1,343758234	1,343758234	12,10
15	-64,83652612	64,83652612	-66,281	66,281	4,753	-1,444473882	1,444473882	
16	-84,45263234	84,45263234	-83,674	83,674	8,313	0,778632344	-0,778632344	
17	68,90835523	68,90835523	80,183	80,183	7,079	-11,27464477	-11,27464477	
18	-82,85040527	82,85040527	-82,105	82,105	13,331	0,745405268	-0,745405268	Standard deviation (SD) TALOS-N
19	-69,98696356	69,98696356	-121,848	121,848	17,969	-51,86103644	51,86103644	
20	-78,63537517	78,63537517	-97,698	97,698	14,141	-19,06262483	19,06262483	5,76
21	-95,89473952	95,89473952	-134,526	134,526	10,666	-38,63126048	38,63126048	
22	-78,70401478	78,70401478	-72,439	72,439	10,105	6,265014779	-6,265014779	
23	-93,41416274	93,41416274	-59,981	59,981	8,391	33,43316274	-33,43316274	
24	79,03549443	79,03549443	94,263	94,263	19,658	-15,22750557	-15,22750557	Root-mean square deviation
25	-52,51759304	52,51759304		0				
26	-56,37209559	56,37209559		0				16,65
27	-63,10047357	63,10047357	-64,058	64,058	6,428	-0,957526426	0,957526426	
28	-94,32954407	94,32954407	-95,621	95,621	13,488	-1,291455932	1,291455932	
29	-154,2889576	154,2889576	-127,494	127,494	13,97	26,79495763	-26,79495763	Root-mean square deviation modulated
30	-60,76671606	60,76671606		0				
31	-74,74492374	74,74492374		0				15,53
32	-103,0417173	103,0417173		0				
33	-81,7239976	81,7239976	-68,665	68,665	6,99	13,0589976	-13,0589976	
34	-104,281164	104,281164	-115,104	115,104	11,788	-10,82283595	10,82283595	
35	-159,136202	159,136202	-84,177	84,177	28,705	74,95920201		
36	-57,00316663	57,00316663	-59,292	59,292	7,642	-2,28883337	2,28883337	
37	-56,37817316	56,37817316	-60,233	60,233	6,425	-3,854826843	3,854826843	
38	102,5066562	102,5066562	94,838	94,838	28,599	7,668656243	7,668656243	
39	-86,86172428	86,86172428	-79,513	79,513	16,63	7,348724276	-7,348724276	
40	-126,2597252	126,2597252	-105,854	105,854	15,551	20,40572519	-20,40572519	
41	-102,8088648	102,8088648	-96,494	96,494	26,303	6,314864846	-6,314864846	
42	-64,68538959	64,68538959	-63,448	63,448	5,057	1,237389588	-1,237389588	
43	-96,5352802	96,5352802	-91,814	91,814	8,64	4,721280204	-4,721280204	
44	64,34398767	64,34398767	87,686	87,686	14,64	-23,34201233	-23,34201233	
45	88,74199158	88,74199158	-74,007	74,007	9,297	14,73499158		
46	-70,34554365	70,34554365	-65,369	65,369	6,622	4,97654365	-4,97654365	
47	-64,94846049	64,94846049	-59,562	59,562	4,938	5,386460495	-5,386460495	
48	-65,98586489	65,98586489	-65,98	65,98	6,257	0,005864889	-0,005864889	
49	-125,5147592	125,5147592	-106,43	106,43	6,81	19,08475917	-19,08475917	
50	-56,39380465	56,39380465	-60,911	60,911	5,421	-4,517195347	4,517195347	
51	-53,20186651	53,20186651	-64,112	64,112	6,042	-10,91013349	10,91013349	
52	-64,2043336	64,2043336	-69,722	69,722	7,926	-5,517666404	5,517666404	
53	-89,21165477	89,21165477	-92,391	92,391	5,759	-3,179345228	3,179345228	
54	-84,04898109	84,04898109	-78,23	78,23	12,894	5,818981095	-5,818981095	
55	-73,05858968	73,05858968	-74,729	74,729	12,222	-1,670410316	1,670410316	
56	-74,24315296	74,24315296	-86,434	86,434	9,257	-12,19084704	12,19084704	
57	-70,48386575	70,48386575	-98,022	98,022	25,587	-27,53813425	27,53813425	
58	-80,70003277	80,70003277	-67,212	67,212	7,5	13,48803277	-13,48803277	
59	-63,96419904	63,96419904		0				
60	-62,2579442	62,2579442	-62,215	62,215	6,774	0,042944195	-0,042944195	
61	-111,8297396	111,8297396	-125,662	125,662	14,969	-13,83226036	13,83226036	
62	-108,8125306	108,8125306	-95,072	95,072	17,074	13,74053061	-13,74053061	
63	-75,79234386	75,79234386	-88,926	88,926	15,085	-13,13365614	13,13365614	
64	-135,5365261	135,5365261	-131,661	131,661	8,058	3,875526104	-3,875526104	
65	-87,8675887	87,8675887	-136,189	136,189	22,055	-48,3214113	48,3214113	
66	-74,39894682	74,39894682	-65,391	65,391	8,113	9,007946821	-9,007946821	
67	-114,923434	114,923434	-124,527	124,527	17,063	-9,603566042	9,603566042	
68	-148,096024	148,096024	-126,038	126,038	9,776	22,05802398	-22,05802398	
69	-86,87090014	86,87090014	-92,608	92,608	10,2	-5,737099861	5,737099861	
70	-107,9093198	107,9093198	-109,777	109,777	11,369	-1,867680249	1,867680249	

71	-117,8823107	117,8823107	-113,281	113,281	10,941	4,601310733	-4,601310733	
72	-97,96201388	97,96201388	-116,131	116,131	16,521	-18,16898612	18,16898612	
73	-114,0482822	114,0482822	-130,598	130,598	21,619	-16,54971781	16,54971781	
74	-107,5056434	107,5056434	-62,107	62,107	7,763	45,39864336	-45,39864336	
75	-85,62732719	85,62732719	-85,415	85,415	17,676	0,212327188	-0,212327188	
76	-135,9347038	135,9347038	-113,013	113,013	23,91	22,92170376	-22,92170376	
77	-80,81154428	80,81154428	-65,218	65,218	6,137	15,59354428	-15,59354428	
78	-60,29167346	60,29167346		0				
79	-51,67625588	51,67625588	-88,628	88,628	23,992	-36,95174412	36,95174412	
80	-127,5630428	127,5630428	-90,911	90,911	19,992	36,6520428	-36,6520428	
81	-85,88140793	85,88140793	-112,115	112,115	17,101	-26,23359207	26,23359207	
82	-129,5748453	129,5748453	-133,352	133,352	7,8	-3,777154668	3,777154668	
83	-104,9207681	104,9207681	-98,172	98,172	14,151	6,748768122	-6,748768122	
84	-109,8340612	109,8340612	-135,223	135,223	13,612	-25,38893876	25,38893876	
85	-113,1262263	113,1262263	-113,664	113,664	13,434	-0,537773688	0,537773688	
86	-107,6326745	107,6326745	-119,241	119,241	10,671	-11,60832552	11,60832552	
87	-117,7691344	117,7691344	-117,791	117,791	8,021	-0,02186564	0,02186564	
88	50,05371411	50,05371411	54,762	54,762	4,823	-4,708285886	-4,708285886	
89	93,33794	93,33794	81,692	81,692	5,826	11,64594	11,64594	
90	-109,6951475	109,6951475	-112,641	112,641	10,826	-2,94585251	2,94585251	
91	-60,53111792	60,53111792	-68,278	68,278	6,065	-7,746882083	7,746882083	
92	-110,9365975	110,9365975	-105,965	105,965	12,774	4,971597453	-4,971597453	
93	-145,8229307	145,8229307	-149,788	149,788	7,301	-3,965069265	3,965069265	
94	-129,0923556	129,0923556	-124,468	124,468	12,009	4,624355642	-4,624355642	
95	-114,4184251	114,4184251	-116,818	116,818	14,825	-2,399574939	2,399574939	
96	-93,51651722	93,51651722	-93,426	93,426	12,065	0,09051722	-0,09051722	
97	-94,14084239	94,14084239	-114,233	114,233	26,32	-20,09215761	20,09215761	
98	57,98199722	57,98199722	56,614	56,614	7,018	1,367997219	1,367997219	
99	76,45982428	76,45982428	83,94	83,94	12,16	-7,480175717	-7,480175717	
100	-89,44608087	89,44608087	-72,165	72,165	10,496	17,28108087	-17,28108087	
101	-78,18902146	78,18902146	-72,708	72,708	12,583	5,481021458	-5,481021458	
102	-85,40559142	85,40559142		0				
103	-66,35427467	66,35427467		0				
104	-72,01728361	72,01728361	-77,325	77,325	13,165	-5,307716387	5,307716387	
105	-113,1734583	113,1734583	-100,166	100,166	8,843	13,00745832	-13,00745832	
106	-127,7967467	127,7967467	-111,294	111,294	12,966	16,50274673	-16,50274673	
107	-122,7460472	122,7460472	-111,083	111,083	10,057	11,66304725	-11,66304725	
108	-118,9731952	118,9731952	-118,424	118,424	9,536	0,549195209	-0,549195209	
109	-94,57844053	94,57844053	-100,235	100,235	15,128	-5,65655947	5,65655947	
110	-91,50640877	91,50640877	-106,626	106,626	14,414	-15,11959123	15,11959123	
111	-131,1020554	131,1020554	-130,075	130,075	15,82	1,027055372	-1,027055372	
112	58,58863093	58,58863093	-85,735	85,735	9,464	-27,14636907		
113	-74,80201715	74,80201715	-84,583	84,583	21,821	-9,780982849	9,780982849	
114	-67,2557793	67,2557793	-65,511	65,511	8,349	1,744779298	-1,744779298	
115	-71,53438472	71,53438472	-91,262	91,262	30,359	-19,72761528	19,72761528	
116	100,8888095	100,8888095	-91,443	91,443	22,581	9,445809483		
117	-84,36873205	84,36873205	-64,766	64,766	6,587	19,60273205	-19,60273205	
118	-145,667896	145,667896	-134,613	134,613	10,341	11,05489599	-11,05489599	
119	-120,2021282	120,2021282	-99,613	99,613	10,132	20,58912815	-20,58912815	
120	-89,20194647	89,20194647	-109,759	109,759	8,268	-20,55705353	20,55705353	
121	-106,8957536	106,8957536	-103,778	103,778	10,948	3,117753644	-3,117753644	
122	-100,334465	100,334465	-121,8	121,8	9,881	-21,465535	21,465535	
123	-116,0433507	116,0433507	-113,55	113,55	8,625	2,493350728	-2,493350728	
124	-97,53788303	97,53788303	-115,141	115,141	8,909	-17,60311697	17,60311697	
125	-94,28761528	94,28761528	-115,898	115,898	14,031	-21,61038472	21,61038472	
126	-82,21553598	82,21553598		0				
127	-78,86939844	78,86939844		0				
128	-118,2556127	118,2556127	-80,699	80,699	15,281	37,55661275	-37,55661275	
129	-139,2395201	139,2395201	129,516	129,516	13,53	9,723520057		
130	-92,55517083	92,55517083	-96,688	96,688	12,238	-4,132829167	4,132829167	
131	-103,5803259	103,5803259	-145,697	145,697	15,645	-42,11667406	42,11667406	
132	-130,5532307	130,5532307	-133,329	133,329	11,407	-2,775769288	2,775769288	
133	-112,7237583	112,7237583	-123,124	123,124	10,825	-10,40024166	10,40024166	
134	-120,1245372	120,1245372	-104,169	104,169	18,545	15,9555372	-15,9555372	
135	-132,4293653	132,4293653	134,797	134,797	11,092	-2,367634712		
136	-96,26810463	96,26810463	-98,468	98,468	14,912	-2,199895372	2,199895372	
137	-122,1411255	122,1411255	-82,506	82,506	14,581	39,63512548	-39,63512548	
138	-88,62038113	88,62038113	-77,931	77,931	11,74	10,68938113	-10,68938113	
139	-133,3425341	133,3425341	-134,064	134,064	7,591	-0,721465885	0,721465885	
140	-121,3492783	121,3492783	-103,637	103,637	15,223	17,71227832	-17,71227832	
141	-94,39381586	94,39381586	-98,001	98,001	15,07	-3,607184137	3,607184137	
142	-140,9241899	140,9241899		0				

Table D2: the data used to calculate the difference, and root-mean square difference between psi (ψ) predicted by TALOS-N and ψ from the crystal structure (pdb file 5aa7), and the mean standard deviation for the predicted values of ψ by TALOS-N.

Amino acid residue	PSI CRYSTALOGRAPHY	ABS crystolography	PSI TALOS-N	Abs Talos-n	Standard DEVIATION PSI NM	differance	Difference modulated	Standardavik differanse
1	131,0044961	131,0044961		0				
2	168,0560994	168,0560994		0				30,92880228
3	164,9263302	164,9263302		0				
4	122,0423254	122,0423254		0				
5	-21,38588024	21,38588024		0				
6	121,7235513	121,7235513	149,843	149,843	11,045	-28,11944867	-28,11944867	Mean differance TalosN/crystologarphy
7	144,6487119	144,6487119		0				
8	134,1461015	134,1461015		0				8,981713411
9	166,0614051	166,0614051	165,324	165,324	5,968	0,737405075	0,737405075	
10	-44,60432125	44,60432125	-39,073	39,073	4,81	5,531321246	-5,51321246	
11	-35,52407064	35,52407064	-38,021	38,021	3,891	-2,496929359	2,496929359	
12	-44,08511797	44,08511797	-37,28	37,28	4,207	6,805117968	-6,805117968	mean standard deaviation TALOS-N
13	-37,16834764	37,16834764	-39,181	39,181	3,816	-2,012652359	2,012652359	
14	-49,12016602	49,12016602	-38,474	38,474	2,774	10,64616602	-10,64616602	12,02586179
15	-29,94003978	29,94003978	-30,747	30,747	4,81	-0,80696022	0,80696022	
16	-4,742789097	4,742789097	-4,353	4,353	10,958	0,389789097	-0,389789097	
17	23,85600594	23,85600594	12,325	12,325	7,314	11,53100594	11,53100594	
18	-24,36606912	24,36606912	-20,483	20,483	13,5	3,883069115	-3,883069115	standard deviation (SD) TALOS-N
19	148,196708	148,196708	153,676	153,676	14,03	-5,479292	-5,479292	
20	-29,56492418	29,56492418	9,937	9,937	19,591	19,62792418	-39,50192442	6,646977015
21	161,4776805	161,4776805	157,262	157,262	9,857	4,215680523	4,215668052	
22	128,395759	128,395759	128,44	128,44	14,268	-0,044241022	-0,044241022	
23	-11,94333962	11,94333962	-27,914	27,914	12,351	-15,97066038	-15,97066038	
24	-144,5204875	144,5204875	-5,552	5,552	16,746	138,9684875		Root mean square deviation
25	-40,80533458	40,80533458		0				
26	-31,12673264	31,12673264		0				28,869981
27	-21,74438133	21,74438133	-31,304	31,304	6,701	-9,559618671	9,55961871	
28	-10,89185051	10,89185051	-0,139	0,139	11,607	10,75285051	-10,75285051	
29	66,51521194	66,51521194	81,532	81,532	21,362	-15,01678806	-15,01678806	
30	-19,87134695	19,87134695		0				modulated root mean squarre deviatio
31	-6,886124952	6,886124952		0				
32	18,36358052	18,36358052		0				16,08186799
33	73,34801144	73,34801144	134,05	134,05	10,389	-60,70198856	-60,70198856	
34	136,4614487	136,4614487	137,112	137,112	15,88	-0,65055131	-0,65055131	
35	164,4373796	164,4373796	134,031	134,031	15,462	30,40637959	30,40637959	
36	149,6387829	149,6387829	148,097	148,097	6,685	1,541782916	1,541782916	
37	152,3222477	152,3222477	136,596	136,596	7,397	15,7262477	15,72622477	
38	-3,165433765	3,165433765	-12,88	12,88	40,185	-9,714566235	9,714566235	
39	159,3231348	159,3231348	143,138	143,138	24,076	16,18513483	16,18513483	
40	25,43465983	25,43465983	137,417	137,417	14,545	-111,9823402		
41	158,4386555	158,4386555	159,317	159,317	29,368	-0,878344518	-0,878344518	
42	-28,31357968	28,31357968	-23,161	23,161	8,32	5,152579676	-5,12379676	
43	-6,194131669	6,194131669	-2,498	2,498	8,188	3,696131669	-3,696131167	
44	5,805328747	5,805328747	14,404	14,404	17,629	-8,598671253	-8,598677125	
45	6,951151269	6,951151269	164,767	164,767	16,475	-157,8158487		
46	110,5756397	110,5756397	137,307	137,307	9,914	-26,73136035	-26,73136035	
47	-21,72299513	21,72299513	-29,021	29,021	8,42	-7,298004872	7,298004872	
48	-21,75744127	21,75744127	-24,71	24,71	8,957	-2,952558734	2,952558734	
49	30,91576612	30,91576612	10,363	10,363	9,258	20,55276612	20,55276612	
50	-24,37299869	24,37299869	-31,605	31,605	7,403	-7,232001307	7,232001307	
51	-33,63815915	33,63815915	-33,264	33,264	9,29	0,374159149	-0,374159149	
52	-19,77620845	19,77620845	-19,932	19,932	10,043	-0,155791546	0,155791546	
53	-8,792045722	8,792045722	-5,935	5,935	8,981	2,857045722	-2,857045722	
54	97,63175524	97,63175524	114,475	114,475	18,145	-16,84324476	-16,84322448	
55	-9,818343428	9,818343428	-11,649	11,649	11,048	-1,830656572	1,830656572	
56	-18,28757745	18,28757745	-8,354	8,354	13,82	9,933577449	-9,933577499	
57	151,2275378	151,2275378	127,48	127,48	30,822	23,74753779	23,7453779	
58	69,82808931	69,82808931	144,677	144,677	10,015	-74,84891069		
59	127,1225948	127,1225948		0				
60	145,2106199	145,2106199	143,796	143,796	13,056	1,414619857	1,41461985	
61	150,291803	150,291803	146,592	146,592	11,729	3,699803033	3,699803033	
62	129,6815801	129,6815801	135,61	135,61	13,682	-5,928419854	-5,928419854	
63	122,7290041	122,7290041	131,53	131,53	15,006	-8,800995945	-8,800995945	
64	167,1945084	167,1945084	152,419	152,419	12,687	14,77550836	14,77550836	
65	176,4395049	176,4395049	156,801	156,801	12,382	19,63850486	19,63850486	
66	-20,68132642	20,68132642	-34,923	34,923	9,298	-14,24167358	14,24167358	
67	129,4010095	129,4010095	133,171	133,171	13,854	-3,769990498	-3,76999045	
68	141,0563933	141,0563933	137,275	137,275	11,511	3,781393257	3,7813257	
69	122,4828138	122,4828138	121,839	121,839	10,058	0,643813763	0,643813763	
70	129,6842339	129,6842339	126,811	126,811	6,741	2,87323386	2,87323238	
71	130,9548751	130,9548751	127,821	127,821	8,422	3,133875071	3,133875071	
72	147,7084685	147,7084685	141,273	141,273	15,721	6,435468492	6,435468492	
73	119,8866639	119,8866639	154,986	154,986	18,072	-35,09933615	-35,09933615	
74	125,4545253	125,4545253	-28,013	28,013	7,895	97,44152529		
75	-29,52446499	29,52446499	-13,348	13,348	15,486	16,17646499	-16,17646499	

76	98,52896306	98,52896306	110,286	110,286	31,123	-11,75703694	-11,75703694	
77	75,89231418	75,89231418	148,924	148,924	9,575	-73,03168582		
78	149,004152	149,004152		0				
79	136,9380944	136,9380944	135,26	135,26	16,368	1,678094419	1,678094419	
80	139,7833294	139,7833294	133,22	133,22	12,159	6,563329414	6,563329414	
81	-79,21842835	79,21842835	142,644	142,644	17,405	-63,42557165		
82	155,7541651	155,7541651	152,862	152,862	8,14	2,892165119	2,892116512	
83	119,1927015	119,1927015	132,477	132,477	8,973	-13,28429846	-13,28429846	
84	128,8377475	128,8377475	155,188	155,188	9,404	-26,35025246	-26,35025246	
85	122,0791	122,0791	139,695	139,695	8,658	-17,61589999	-17,61589999	
86	147,4955121	147,4955121	136,576	136,576	6,582	10,91951215	10,91955122	
87	118,2643313	118,2643313	119,51	119,51	6,772	-1,245668702	-1,245668702	
88	38,97044369	38,97044369	38,752	38,752	5,309	0,218443694	0,218443694	
89	-7,484417723	7,484417723	1,715	1,715	8,023	5,769417723	-9,199417723	
90	135,9643274	135,9643274	132,777	132,777	9,259	3,187327397	3,187327397	
91	128,022652	128,022652	125,427	125,427	6,95	2,595651963	2,595651963	
92	-50,5098753	50,5098753	-38,158	38,158	7,657	12,3518753	-12,3518754	
93	150,1932108	150,1932108	151,582	151,582	5,169	-1,388789222	-1,388789222	
94	133,0874079	133,0874079	129,845	129,845	8,548	3,242407858	3,242407858	
95	125,4296995	125,4296995	133,89	133,89	12,778	-8,460300488	-8,460300488	
96	119,2282344	119,2282344	120,598	120,598	12,431	-1,369765624	-1,369765624	
97	-3,147869135	3,147869135	75,059	75,059	49,62	-71,91113087	-78,20686914	
98	34,50037734	34,50037734	38,51	38,51	7,792	-4,009622663	-4,009622663	
99	7,43633497	7,43633497	-2,573	2,573	18,977	4,86333497	10,00933497	
100	165,5275211	165,5275211	143,227	143,227	12,363	22,30052113	22,30052113	
101	131,8360939	131,8360939	138,532	138,532	14,346	-6,69590615	-6,69590615	
102	159,9529988	159,9529988		0				
103	153,1223797	153,1223797		0				
104	-19,91957745	19,91957745	-12,3	12,3	7,368	7,619577446	-7,619577446	
105	110,4367126	110,4367126	9,663	9,663	9,68	100,7737126		
106	129,3843751	129,3843751	131,26	131,26	7,193	-1,875624911	-1,875624911	
107	129,5552436	129,5552436	133,043	133,043	9,777	-3,487756419	-3,487756419	
108	138,6283424	138,6283424	133,504	133,504	9,265	5,124342434	5,124434244	
109	116,2380237	116,2380237	125,91	125,91	12,172	-9,671976308	-9,671976308	
110	141,8636013	141,8636013	129,252	129,252	12,528	12,61160129	12,61160129	
111	164,6714645	164,6714645	160,058	160,058	12,381	4,613464532	4,613346453	
112	36,19232979	36,19232979	-1,046	1,046	16,54	35,14632979	37,23832979	
113	137,4774252	137,4774252	134,371	134,371	18,709	3,106425169	3,106425169	
114	144,9593089	144,9593089	148,37	148,37	8,758	-3,41069109	-3,41069109	
115	156,558799	156,558799	144,912	144,912	21,368	11,64679905	11,64679905	
116	-165,9354375	165,9354375	169,877	169,877	18,56	-3,941562483	-3,941562483	
117	126,9920603	126,9920603	142,237	142,237	9,567	-15,24493969	-15,24493969	
118	163,3831854	163,3831854	149,829	149,829	6,73	13,55418544	13,55418544	
119	118,3007312	118,3007312	123,542	123,542	9,135	-5,241268847	-5,241268847	
120	131,0626322	131,0626322	128,058	128,058	7,547	3,004632178	3,004632178	
121	121,2881779	121,2881779	126,677	126,677	5,752	-5,38882207	-5,38882207	
122	121,4563884	121,4563884	136,344	136,344	9,783	-14,8876116	-14,8876116	
123	112,4161949	112,4161949	131,063	131,063	9,032	-18,64680514	-18,64680514	
124	111,3680829	111,3680829	133,682	133,682	12,99	-22,31391711	-22,31391711	
125	122,4815989	122,4815989	132,806	132,806	10,883	-10,32440111	-10,32440111	
126	133,6119557	133,6119557		0				
127	-20,19189489	20,19189489		0				
128	11,62075446	11,62075446	-13,061	13,061	17,563	-1,440245542	24,68175446	
129	173,653252	173,653252	157,129	157,129	10,264	16,52425195	16,52425195	
130	19,29148696	19,29148696	-7,555	7,555	15,512	11,73648696	26,8464487	
131	156,8693464	156,8693464	152,757	152,757	11,967	4,112346385	4,112346385	
132	162,6538141	162,6538141	153,315	153,315	10,215	9,338814085	9,33881485	
133	139,0590288	139,0590288	141,404	141,404	10,904	-2,344971239	-2,344971239	
134	119,8426428	119,8426428	139,978	139,978	10,393	-20,13535723	-20,13535723	
135	114,7456646	114,7456646	150,516	150,516	11,579	-35,77033538	-35,77033538	
136	134,8962165	134,8962165	132,779	132,779	10,416	2,117216458	2,117216458	
137	118,4458213	118,4458213	129,79	129,79	12,804	-11,34417875	-11,34417875	
138	122,2716357	122,2716357	137,357	137,357	10,965	-15,08536426	-15,08536443	
139	156,3188591	156,3188591	158,017	158,017	8,002	-1,698140901	-1,698114901	
140	114,5223275	114,5223275	130,292	130,292	7,114	-15,76967251	-15,76967251	
141	130,2194628	130,2194628	123,769	123,769	5,463	6,450462757	6,450462757	
142	19,27433437	19,27433437		0				