ABSTRACT

O’DELL, WILLIAM BRADLEY. Structural Investigations of the Fungal Lytic Polysaccharide Monooxygenase Reaction Mechanism using X-ray and Neutron Protein Crystallography. (Under the direction of Dr. Flora Meilleur).

Industrial production of biofuels and other value-added products derived from cellulosic feedstocks relies on enzymatic hydrolysis of cellulose to convert biomass to soluble sugars. The overall conversion efficiency and the cost of final products are influenced greatly by the efficiency of hydrolases which is diminished by cellulose recalcitrance arising from high degrees of polymerization, crystallinity and intra-/intermolecular hydrogen bonding. Efforts over the last 20 years to discover more efficient hydrolases have instead identified oxidative enzymes that boost the efficiency of cellulose conversion. These lytic polysaccharide monooxygenases (LPMOs) break cellulose chains by oxidizing carbon atoms involved in glycosidic bonds. LPMOs do not directly release soluble sugars but instead function as an important adjuvant to hydrolase function that increases industrial conversion efficiency by one or more orders of magnitude.

LPMOs are mononuclear copper metalloenzymes dependent upon an input of reducing equivalents that activate molecular O₂ for the net insertion of an oxygen atom into a carbon–hydrogen bond. Herein are described important new structural characterizations of this reactivity derived from crystallographic studies of the enzyme Neurospora crassa LPMO9D in resting and O₂-activated states. High resolution X-ray cryocrystallography has revealed that fungal LPMOs activate O₂ as an equatorial ligand of the active site Cu(I/II) ion. Molecular O₂ activation likely follows an initial “pre-binding” step in which molecular O₂ occupies a polar binding site adjacent to the active site ion. Analysis of the first reported LPMO structure determined by neutron protein crystallography provided new details of protonation states and hydrogen bonding around the active site and indicated a role of a conserved histidine in stabilizing “pre-bound” O₂. This role has been confirmed with quantum calculations. The impacts of this research for current mechanistic understanding and future experiments are also discussed.
Structural Investigations of the Fungal Lytic Polysaccharide Monooxygenase Reaction Mechanism using X-ray and Neutron Protein Crystallography

by
William Bradley O'Dell

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Biochemistry

Raleigh, North Carolina
2017

APPROVED BY:

_______________________________
Flora Meilleur
Committee Chair

_______________________________
Michael B. Goshe

_______________________________
Robert B. Rose

_______________________________
Amy M. Grunden
DEDICATION

For my loving parents who never once discouraged me from pursuing my education.
BIOGRAPHY

William Bradley “Brad” O’Dell was born in Newport, TN and grew up on his family’s small farm nearby. Brad completed grades K–8 at Bridgeport Elementary School and grades 9–12 at Cocke County High School from which he graduated as a valedictorian in 2005 with plans of pursuing a medical doctorate.

In Fall 2005, Brad enrolled at the University of Tennessee, Knoxville as a member of the University Honors program. Early in 2007, Brad was admitted to the UTK College of Arts and Sciences interdisciplinary College Scholars Program. Participation in College Scholars allowed a curricular flexibility that Brad used to become heavily involved in undergraduate research in the UTK Department of Chemistry. (At this point, having a career in basic science superseded Brad’s early goal of a career in medicine.) Brad received several honors for academics and research the most notable including admission to Phi Beta Kappa and receiving the Barry Goldwater Scholarship, both in spring of 2008, and he completed an undergraduate honors thesis. In May 2009 Brad graduated from UTK with a Bachelor of Arts in College Scholars Honors with Concentration in Structural Chemistry and University Honors.

Beginning Fall 2009, Brad spent an academic year abroad at the Department of Biochemistry, University of Cambridge as a participant in the National Institutes of Health–Oxford–Cambridge Partnership program with support from the NIH National Cancer Institute. In January 2011, Brad began a post-bachelor’s research assistantship at Oak Ridge National Laboratory that continued until August 2012.

In 2012 Brad was admitted to the PhD program of the Department of Molecular and Structural Biochemistry, NC State University and joined the laboratory of Dr. Flora Meilleur. Through participation in the ORNL GO! Program and an NSF IGERT studentship, Brad conducted his thesis research in residence at the Neutron Sciences Directorate, ORNL. During his studentship at NC State, Brad has received the Biochemistry Samuel Tove Award for Student Teaching and the Biochemistry A.R. Main–BD Award for Graduate Research. In 2017 Brad was awarded a National Research Council Post-Doctoral Associateship to join the Biomolecular Labeling Laboratory of the National Institute of Standards and Technology.
ACKNOWLEDGMENTS

I want to thank first Dr. Flora Meilleur for asking me to join her laboratory and in doing so giving me the opportunity to restart my academic career after an unplanned pause. Your enthusiastic and steady mentorship as the “boss lady” has been a driving force behind the science we accomplished together and has also balanced out my tendencies to be “bio-dramatic” when unexpected challenges arose. You facilitated unique opportunities and allowed intellectual freedom that far exceeded an average graduate experience. Thank you.

So many others deserve thanks for the help, guidance and support they have given to me along the way, and I consider myself lucky to have been welcomed both by Molecular and Structural Biochemistry at NC State and by Energy & Environment Group at Oak Ridge National Laboratory. At NC State, I thank my committee members, Drs. Robert Rose, Amy Grunden, Michael Goshe and Linda Hanley-Bowdoin (hon.) for their time, their insights and their commitment to making both my science and my graduate career a success. I also thank Drs. Paul Swartz and Trino Ascencio-Ibáñez for the long hours each spent teaching me and training me with the skills I needed to perform my research and become a structural biologist. NC State Biochemistry, faculty, students and staff alike, have been great instructors, colleagues and friends. At ORNL, I thank Dr. Volker Urban for welcoming me into his group and supporting numerous opportunities. I thank Dr. Kevin Weiss for his exceedingly abundant help with every aspect of my research, for his attention to every detail, and for his inexhaustible patience. ORNL has been a second home for me for several years now, and I thank each scientist, staff member, post-doc and student who has been a part of this community.

I thank Dr. Haskell Taub and the NSF IGERT program he led for supporting my Ph.D. studentship and for providing unique interdisciplinary learning opportunities in neutron science. I also thank Drs. Hugh O’Neill, Sylvia McLain, David C. Baker, Fred Schell and John Turner for fostering my interests in science and research so early in my career.

Finally, I thank my parents, family and friends for your love and encouragement through every day, every challenge and every success. I couldn’t have done this without you.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................................ viii
LIST OF FIGURES ........................................................................................................................................ ix
LIST OF ABBREVIATIONS ......................................................................................................................... xi
Chapter 1 ......................................................................................................................................................... 1
  1.1 Introduction to fungal lytic polysaccharide monooxygenases ................................................................. 1
    1.1.1 Cellulose as a renewable feedstock for fuels and chemical products .............................................. 1
    1.1.2 Enzymatic cellulose deconstruction ............................................................................................... 4
    1.1.3 Discovery of LPMOs and LPMO–cellulase synergy ........................................................................ 8
    1.1.4 Classes of LPMOs ....................................................................................................................... 17
    1.1.5 LPMOs in industrial lignocellulose conversion ............................................................................. 24
    1.1.6 LPMO electron donors ................................................................................................................ 26
    1.1.7 The LPMO reaction mechanism .................................................................................................... 29
    1.1.8 Neurospora crassa LPMO9D as a model for mechanistic studies .................................................... 33
    1.1.9 Overview of Research .................................................................................................................. 36
    1.1.10 REFERENCES .............................................................................................................................. 38
  1.2 Neutron protein crystallography: A complementary tool for locating hydrogens in proteins .......... 54
    1.2.1 Highlights ...................................................................................................................................... 54
    1.2.2 Abstract ......................................................................................................................................... 54
    1.2.3 Introduction .................................................................................................................................... 55
    1.2.4 Hydrogen/Deuterium isotopic substitution .................................................................................... 59
      1.2.4.1 H/D exchange ....................................................................................................................... 60
      1.2.4.2 Perdeuteration ..................................................................................................................... 63
    1.2.5 Crystal growth ............................................................................................................................... 65
    1.2.6 Instrumentation, data collection and processing .......................................................................... 69
    1.2.7 Structure refinement .................................................................................................................... 73
    1.2.8 Perspective .................................................................................................................................... 76
Chapter 2: Crystallization of a fungal lytic polysaccharide monooxygenase expressed from glyco-engineered *Pichia pastoris* for X-ray and neutron diffraction

2.1 Synopsis .......................................................... 102
2.2 Abstract .......................................................... 102
2.3 Introduction ...................................................... 103
2.4 Materials and methods ......................................... 105
  2.4.1 Macromolecule production ................................ 105
  2.4.2 Mass spectrometry ......................................... 108
  2.4.3 Crystallization ............................................. 108
  2.4.4 Data collection and processing ......................... 110
  2.4.5 Structure solution and refinement ..................... 113
2.5 Results and discussion ........................................ 114
2.6 Acknowledgements ............................................. 124
2.7 REFERENCES ..................................................... 125

Chapter 3 .................................................................... 129

3.1 Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase .................................................. 130
  3.1.1 Abstract ......................................................... 130
  3.1.2 Results and discussion ..................................... 130
  3.1.3 Experimental section ....................................... 136
  3.1.4 Acknowledgements .......................................... 137
  3.1.5 REFERENCES ................................................. 138
  3.1.6 Supporting information ...................................... 142
    3.1.6.1 Supporting information contents ................... 142
    3.1.6.2 Experimental details ................................... 143
      3.1.6.2.1 Crystallization and crystal soaking ............... 143
      3.1.6.2.2 Model Refinement .................................. 144
      3.1.6.2.3 Multiple sequence alignment .................... 144
### LIST OF TABLES

Table 1.2.1: H₂O and D₂O physico-chemical properties. ................................................................. 61
Table 1.2.2: Diffractometers for neutron protein crystallography.................................................. 72
Table 2.1: Macromolecule production information........................................................................... 108
Table 2.2: Crystallization conditions................................................................................................. 110
Table 2.3: Data collection and processing........................................................................................ 112
Table 2.4: Structure solution and refinement...................................................................................... 114
Table 3.1.1: Calculated thermodynamics of molecular oxygen “pre-binding” as a function of His157 conformation and protonation................................................................. 136
Table 3.1.2: (SI Table 1) X-ray and neutron dataset statistics............................................................. 149
Table 3.1.3: (SI Table 2) X-ray and neutron model refinement statistics.......................................... 150
Table 3.1.4: (SI Table 3) Copper–ligand distances and selected atomic displacement factors in the resting state active site structure (PDB 5TKG)............................................................. 151
Table 3.1.5: (SI Table 4) Copper–ligand distances and selected atomic displacement factors in the ascorbate-treated active site structure (PDB 5TKH)............................................................. 152
Table 3.1.6: (SI Table 5) Atomic coordinates of geometry optimized DFT active site models. ................................................................................................................................. 158
Table 3.2.1: Comparison of calculated and experimental CDHIIA’s $R_g$ (Å) and $d_{max}$ (Å). ................................................................................................................................. 170
Table 4.1: Data collection and processing statistics. ........................................................................ 184
Table A.1 Supporting Information Table S1 from “Neutron protein crystallography: A complementary tool for locating hydrogens in proteins”................................. 194
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>Hierarchical elements of lignocellulose structure.</td>
<td>3</td>
</tr>
<tr>
<td>1.1.2</td>
<td>Cellulase catalytic domains and their modes of interaction with cellulose.</td>
<td>7</td>
</tr>
<tr>
<td>1.1.3</td>
<td><em>T. reesei</em> GH61B compared with other endoglucanases.</td>
<td>11</td>
</tr>
<tr>
<td>1.1.4</td>
<td>Sites of single residue mutations introduced in <em>Tt</em>GH61E (PDB 3EII).</td>
<td>14</td>
</tr>
<tr>
<td>1.1.5</td>
<td>Structural variation of LPMOs from classes AA9, AA10, AA11 and AA13.</td>
<td>20</td>
</tr>
<tr>
<td>1.1.6</td>
<td>Proposed mechanisms for LPMO monooxygenation.</td>
<td>31</td>
</tr>
<tr>
<td>1.1.7</td>
<td><em>Nc</em>LPMO9D active site with superoxide in axial copper coordination.</td>
<td>35</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Incoherent neutron scattering cross sections and coherent neutron scattering lengths for selected elements.</td>
<td>56</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Neutron structure determination flow chart.</td>
<td>58</td>
</tr>
<tr>
<td>1.2.3</td>
<td>Neutron density maps from H/D-exchanged and perdeuterated beta-lactamase.</td>
<td>60</td>
</tr>
<tr>
<td>1.2.4</td>
<td>H/D exchange of a fully grown crystal in the capillary mount by vapor diffusion.</td>
<td>61</td>
</tr>
<tr>
<td>1.2.5</td>
<td>Backbone amide H/D exchange observed in two neutron structures of HIV-1 protease.</td>
<td>63</td>
</tr>
<tr>
<td>2.1</td>
<td>Structure of <em>Nc</em>PMO-2 and the “histidine brace”.</td>
<td>104</td>
</tr>
<tr>
<td>2.2</td>
<td>Analysis of <em>Nc</em>PMO-2 glycoforms by mass spectrometry.</td>
<td>116</td>
</tr>
<tr>
<td>2.3</td>
<td>Crystals and X-ray diffraction of <em>Nc</em>PMO-2 expressed from <em>P. pastoris</em> X-33.</td>
<td>118</td>
</tr>
<tr>
<td>2.4</td>
<td>Asparagine 60 N-linked glycan observed for <em>Nc</em>PMO-2 from <em>P. pastoris</em> X-33.</td>
<td>119</td>
</tr>
<tr>
<td>2.5</td>
<td>Crystals and X-ray diffraction of <em>Nc</em>PMO-2 expressed from <em>P. pastoris</em> SuperMans.</td>
<td>121</td>
</tr>
<tr>
<td>2.6</td>
<td>Asparagine 60 N-linked glycan observed for <em>Nc</em>PMO-2 from <em>P. pastoris</em> SuperMans.</td>
<td>122</td>
</tr>
<tr>
<td>2.7</td>
<td><em>Nc</em>PMO-2 crystal used for room temperature X-ray and neutron data collection as seen through the quartz capillary mount.</td>
<td>123</td>
</tr>
</tbody>
</table>
Figure 2.8: Neutron Laue diffraction recorded from $^1$H/$^2$H vapor-exchanged NcPMO-2. .............................................................. 123

Figure 3.1.1: NcPMO-2 active site for both NCS molecules in the enzymatic resting state and after treatment with ascorbic acid. .............................................................. 132

Figure 3.1.2: His157 conformation observed by room temperature neutron protein crystallography. ........................................................................................................ 135

Figure 3.1.3: (SI Figure 1) Electron density $F_0$–$F_c$ maps for oxygen species in untreated and ascorbate treated X-ray models. ............................................................. 153

Figure 3.1.4: (SI Figure 2) Superposition of copper–dioxo model (PDB 5TKH) with PDB 4EIR. ............................................................... 154

Figure 3.1.5: (SI Figure 3) Multiple sequence alignment of AA9 proteins with known crystal structures. ........................................................................................................ 155

Figure 3.1.6: (SI Figure 4) Fits of His157 conformational and protonation states with observed neutron scattering length density ............................................................................ 156

Figure 3.1.7: (SI Figure 5) Geometry optimized DFT active site models and the same superimposed with NCS molecule B resting state active site structure (PDB 5TKG)........................................................................ 157

Figure 3.1.8: (SI Figure 6) Superposition of copper–dioxo model (PDB 5TKH) with PDB 5ACJ and PDB 5ACI. .............................................................. 161

Figure 3.1.9: (SI Figure 7) X-ray fluorescence scan of a NcPMO-2 crystal near the copper K-edge ........................................................................................................ 162

Figure 3.2.1: Hydrogen bond between Tyr168 and Gln166. ........................................... 168

Figure 3.2.2: Water molecules involved in hydrogen bonds at the NcLPMO9D active site. ........................................................................................................ 169

Figure 3.2.3: SANS data and analysis for NcCDHIIA. ............................................... 170

Figure 3.2.4: $P(r)$ curve obtained by Fourier transformation of the SANS data. .......... 171

Figure 3.2.5: Models produced by SASSIE modeling.................................................... 172

Figure 4.1: Laue photograph of ascorbate-treated NcLPMO9D crystal at 20 K. ........ 182
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>amino acid</td>
</tr>
<tr>
<td>ASM</td>
<td>active site model</td>
</tr>
<tr>
<td>ASU</td>
<td>asymmetric unit</td>
</tr>
<tr>
<td>BMGY</td>
<td>buffered glycerol-complex medium</td>
</tr>
<tr>
<td>CAZY</td>
<td>Carbohydrate-Active Enzymes Database</td>
</tr>
<tr>
<td>CBH</td>
<td>cellobiohydrolase</td>
</tr>
<tr>
<td>CBM</td>
<td>carbohydrate binding module</td>
</tr>
<tr>
<td>CDH</td>
<td>cellobiose dehydrogenase</td>
</tr>
<tr>
<td>CYT</td>
<td>cytochrome or cytochrome domain</td>
</tr>
<tr>
<td>D₂O</td>
<td>deuterium oxide</td>
</tr>
<tr>
<td>DFT</td>
<td>density functional theory</td>
</tr>
<tr>
<td>DH</td>
<td>dehydrogenase or dehydrogenase domain</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrofolate reductase</td>
</tr>
<tr>
<td>DMAP</td>
<td>dimethylallyl pyrophosphate</td>
</tr>
<tr>
<td>d&lt;sub&gt;max&lt;/sub&gt;</td>
<td>maximum dimension</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPI</td>
<td>diffraction precision index</td>
</tr>
<tr>
<td>DβM</td>
<td>dopamine β-monooxygenase</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EG</td>
<td>endoglucanase</td>
</tr>
<tr>
<td>EPR</td>
<td>electron paramagnetic resonance</td>
</tr>
<tr>
<td>eV</td>
<td>electron volts</td>
</tr>
<tr>
<td>EXAFS</td>
<td>extended X-ray absorption fine structure</td>
</tr>
<tr>
<td>FAD</td>
<td>flavin adenine dinucleotide</td>
</tr>
<tr>
<td>FPPS</td>
<td>human farnesyl pyrophosphate synthase</td>
</tr>
<tr>
<td>GH</td>
<td>glycoside hydrolase</td>
</tr>
<tr>
<td>GH61</td>
<td>glycoside hydrolase family 61</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-d-glucose</td>
</tr>
</tbody>
</table>
GMC: glucose–methanol–choline or glucose–methanol–choline oxidoreductase
GO: galactose oxidase
GPP: geranyl pyrophosphate
GST: glutathione S-transferase
GTP: guanosine triphosphate
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV: human immunodeficiency virus
HIV Pr: HIV-1 protease
IET: interdomain electron transfer
IPP: isopentyl phosphate
ITC: isothermal titration calorimetry
$K_a$: equilibrium acid dissociation constant
$k_{cat}$: enzyme turnover number
$K_D$: equilibrium dissociation constant
kDa: kilodalton
LPMO: lytic polysaccharide monooxygenase
Man: D-mannose
micro-PIXE: microbeam proton-induced X-ray emission
mRNA: messenger ribonucleic acid
NADP: nicotinamide adenine dinucleotide phosphate
N-BP: nitrogen-containing bisphosphonate
NCBM: non-coupled binuclear monooxygenase
NCS: non-crystallographic symmetry or non-crystallographically symmetric
NMR: nuclear magnetic resonance
NPC: neutron protein crystallography
ORF: open reading frame
PAGE: polyacrylamide gel electrophoresis
PASC: phosphoric acid swollen cellulose
PDB: Protein Data Bank
PEG: polyethylene glycol
PES: polyethersulfone
PHM: peptidylglycine α-hydroxylating monooxygenase
PNGase F: peptide-N-glycosidase F
PTFE: polytetrafluoroethylene
QM/MM: quantum mechanics/molecular mechanics
$R_g$: radius of gyration
RMSD: root mean squared deviation
SANS: small-angle neutron scattering
SAXS: small-angle X-ray scattering
SHE: standard hydrogen electrode
SHF: separate hydrolysis and fermentation
SLD: scattering length density
SPR: surface plasmon resonance
SSF: simultaneous saccharification and fermentation
$T_4$: thyroxine
TF: translation function
TFA: trifluoroacetic acid
TLS: translation–libration–screw
TOF: time of flight
TTR: human transthyretin
TβM: tyramine β-monooxygenase
Chapter 1

1.1 Introduction to fungal lytic polysaccharide monooxygenases

1.1.1 Cellulose as a renewable feedstock for fuels and chemical products

As of June 2017 Earth retains known petroleum reserves of 240.7 gigatons, and human activity is consuming petroleum at a rate of 4.4 gigatons per year [1]. In contrast, Earth has been estimated to produce on the order of 100 gigatons of terrestrial cellulose per year in the form of plant biomass [2]. Therefore, considerable interest and motivation exist for satisfying global demand for products currently derived from petroleum, such as transportation fuels and plastics, by producing these materials from abundant, renewable supplies of biomass. Additionally, since cellulose is nonnutritive for human consumption, cellulosic biomass can be farmed and harnessed without directly competing with world food supplies. As recently reviewed by Kawaguchi et al. various polymerization monomers can already be accessed at industrial scale from cellulosic biomass feedstocks, and both the United States of America and the European Union have legislation in force setting targets for annual production of cellulose-derived ethanol for transportation use [3-5]

The total global cellulose production is certainly an unreasonable upper limit to the availability of cellulose feedstocks. However, realistic examinations of cellulose availability and usage have shown the potential for significantly replacing petroleum-derived products with cellulose-derived products. For example, Gelfand et al. used twenty-year agricultural data and current land use statistics to determine that the US Midwest region could produce sufficient and accessible cellulose necessary to yield ~18 gigaliters of ethanol per year from crops grown on marginal lands not suitable for agricultural food production [6].

Cellulose is the principal component of plant cell walls accounting for ~15-45% (wt./wt.) of dry plant matter across various species [7]. Cellulose is a linear polymer composed of glucose monomers with cellobiose (1-β-D-glucopyranosyl-4-D-glucopyranose) representing the true repeating unit as it retains the characteristic β-(1→4) glycosidic bond between glucose
subunits. The steric constraints of the β-(1→4) glycosidic bond organize glucose units within cellulose chains to lie within a shared plane unlike the helical structures formed by carbohydrate polymers composed of α-(1→4) linked sugars such as amylose and amylopectin. During cellulose synthesis, numerous chains are produced simultaneously by co-located copies of the transmembrane cellulose synthase complexes. As these cellulose chains elongate, they associate to form a microfibril structure in which adjacent chains in the hexose equatorial plane form regular hydrogen bonds and chains adjacent to each other in the perpendicular plane stack via strong van der Waals interactions [8]. Cellulose microfibrils can pack together to form microscopic domains of crystalline cellulose or exist in a less ordered amorphous form. In the structure of the plant cell wall, both crystalline and amorphous cellulose associate with non-cellulose carbohydrates collectively known as hemicellulose and the structurally heterogeneous polyphenolic polymer network of lignin. Figure 1.1.1 highlights these hierarchical structural elements of plant-derived cellulose.

Elements of this hierarchical structure contribute to an overall recalcitrance of lignocellulose toward chemical or biochemical conversion to products. Biorefineries, the cellulose-dependent equivalent to petroleum refineries, are organized to overcome lignocellulose recalcitrance largely in two processes called “pre-treatment” and “saccharification”. The conglomeration of lignin, hemicellulose and cellulose limits the accessibility of the cellulose fraction for chemical or enzymatic attack, and the dense packing of crystalline cellulose domains further restricts access to only cellulose chains on crystallite surfaces. Pre-treatment steps are designed to mitigate these recalcitrance factors by disrupting the association of lignin, hemicellulose and cellulose and, in some cases, selectively removing the lignin or hemicellulose fraction. In addition, certain chemical pre-treatments, such as ammonia fiber expansion, also directly decrease the total cellulose crystallinity [9].
Figure 1.1.1: Hierarchical elements of lignocellulose structure. Lignin (not pictured) associates with microfibrils around the hemicellulose and paracrystalline cellulose elements. Glucose is the fundamental unit of cellulose, but cellobiose is the repeating unit as it retains the β-(1→4) glycosidic bond. This figure is modified and reprinted with permission from reference [10]. Credit: Oak Ridge National Laboratory.

After successful pre-treatment, the major contributor to recalcitrance that remains is the stability of the glycosidic bond linking each glucose subunit. Glycosidic bonds in aqueous solution have been estimated to have a half-life of 5–8 million years representing a stability against hydrolysis greater than both peptide and phosphodiester bonds [11]. Catalysis is clearly a necessity in the saccharification of cellulose to glucose. Acid catalyzed hydrolysis is an effective route for cellulose conversion; however, acidic conditions catalyze the formation of undesirable side products including the furan analogs furfural and hydroxymethylfurfural [12]. The chemical specificity of cellulose hydrolyzing enzymes for producing cellobiose and then glucose as the sole final product avoids the accumulation of side products, and enzymatic hydrolysis occurs at biocompatible conditions. Both of these traits make enzymatic conversion
a biocompatible saccharification step that is advantageous for successive or simultaneous glucose fermentation to yield the desired chemical product. Enzymatic saccharification is the method of choice for all current industrial-scale and pilot-scale biorefineries producing cellulosic ethanol; however, the process demand of incubating each gram of biomass with 20–30 mg of enzyme to achieve efficient saccharification adds significant costs to ethanol produced. Process analyses suggest that achieving cellulosic ethanol market prices of $2.15 gal\(^{-1}\) will require enzymes to contribute no more than $0.34 gal\(^{-1}\) to the final cost [13]. Considering that ~10 kg of biomass are required to produce a gallon of ethanol, economic viability requires either producing cellulase enzymes at a cost less than $1.70 kg\(^{-1}\) or discovering new routes to achieve enzymatic hydrolysis with a lower enzyme-to-biomass ratio. These economic considerations have motivated biochemical, structural, evolutionary and protein engineering studies of cellulose hydrolyzing enzymes for decades, and the impetus for understanding the “cellulases” will only increase with increased global production of cellulose-derived products and fuels.

1.1.2 Enzymatic cellulose deconstruction

Biochemical understanding of and biotechnological approaches to enzymatic deconstruction of cellulose into glucose bear strong historical influences from studies of the filamentous fungus *Trichoderma reesei*, originally *T. viridae* QM6a, and the cellulolytic enzymes it secretes [14]. *T. reesei* was isolated and discovered to be an efficient degrader of high crystallinity cellulose around 1950 by Elwyn Reese and coworkers within the United States Army Quartermaster research program [15]. (Samples had been collected from decomposed cotton military articles deployed to the Solomon Islands during World War II [14, 16].) Since *T. reesei* could grow in media supplemented with cellulose in suspension as the sole carbon source, Reese and coworkers classified *T. reesei* as “cellulolytic”, and, like other contemporaneously recognized cellulolytic organisms, clarified *T. reesei* culture supernatant contained a soluble “cellulase” component that in isolation could degrade cellulose [17, 18]. *T. reesei* cellulase qualified as a “complete cellulase” under the “C\(_1\)–C\(_x\)” hypothesis of Reese
et al. [17]. The C1–Cx hypothesis described two types of enzymatic activity believed to be necessary for cellulolytic growth: Cx, a cellulose-hydrolyzing activity yielding soluble sugars, and C1, a cellulose-remodeling activity that converted naturally occurring cellulose into a form susceptible to Cx hydrolysis. The existence of separate but equally vital activities for cellulose deconstruction was inferred from the inability of some organisms to grow on native cellulose while being able to grow on soluble chemical derivates of cellulose, such as carboxymethyl cellulose; such organisms were presumed to lack the C1 activity and express only the cellulose-hydrolyzing Cx activity. Early fractionations of T. reesei culture supernatant proved that the cellulose-hydrolyzing Cx activity was attributable to multiple enzymes, and these enzymes were subject to inducible expression upon culture on cellulose, cellobiose, lactose or trace amounts of sophorose [19-22].

T. reesei became the cellulase source of choice for Reese and Mandels and coworkers because it produced a high Cx activity and a comparatively stable C1 activity that enabled fundamental investigations of cellulase activity and expression [23]. Given the advantages of deconstructing cellulose by enzymatic as opposed to chemical means and the recognition of T. reesei as the best known cellulase producer, researchers in the 1960s–70s conducted extensive efforts to maximize cellulase expression from T. reesei. This phenotypic discovery effort included mutagenesis induced by nitrosoguanidine and culture exposure to ionizing photon radiation followed by extensive screening of individual isolates. By 1979, the strain T. reesei Rut-C30 emerged as an excellent overproducer yielding ~2 grams cellulase per liter of culture while also being released from the catabolite repression of cellulase expression by simple carbon sources observed in parental T. reesei [24]. Abundant, stable, active cellulase which could be isolated from culture by simple filtration of the biomass promoted concepts of biorefineries based on cellulose feedstocks degraded by T. reesei cellulases to produce concentrated solutions of glucose to feed yeasts for conversion to single cell protein or fermentation products such as ethanol [23, 25, 26].

Levinson et al. reported in 1951 that cellobiose was the primary product of T. reesei cellulases and that the addition of an isolated β-glucosidase was necessary to yield glucose as the main reaction product [27]. As such, the T. reesei cellubiohydrolases (CBHs) became
subjects of intense research. The CBH enzyme Cel7A (originally CBH I) accounts for ~60% of the total protein secreted by \textit{T. reesei} and has been an archetype for understanding cellulases as evinced by its distinction as the first cellulase to have its gene cloned and to have a three-dimensional structure (catalytic domain only) solved by X-ray crystallography [28-30]. The protein consists of an \textit{N}-terminal catalytic domain exhibiting a hollow tunnel which becomes “threaded” with the reducing end of a cellulose chain terminus thus feeding the substrate toward the active site composed of one aspartate and two glutamate residues that act via an acid-base mechanism to catalyze hydrolysis [31, 32]. A long, glycosylated linker peptide connects the Cel7A catalytic domain to a cellulose-specific carbohydrate binding module (CBM) that greatly increases the affinity of the full-length protein for crystalline cellulose [33]. Both the cellosaccharide binding tunnel and the appended CBM contribute to processivity of Cel7A which experimental measurements have placed on the order of 50 hydrolysis reactions per chain threading event with the apparent hydrolysis \( k_{\text{cat}} = 0.1-1.45 \text{ s}^{-1} \) [34].

Despite the processivity and rapid rate of cellulose hydrolysis exhibited by \textit{TrCel7A}, the purified enzyme alone cannot digest most forms of naturally occurring cellulose. The additional activities of the co-secreted \textit{T. reesei} non-reducing end cellobiohydrolase Cel6A (originally CBH II) and of the non-processive, random glycosidic bond cleaving endoglucanase (EG) Cel7B (originally EG II). Figure 1.1.2 shows the structures of the catalytic domains of \textit{TrCel7A}, \textit{TrCel6A} and \textit{Melanocarpus albomyces} Cel7B along with a schematic representation of the different interaction modes of each with a cellulose substrate. Experimental reconstitution of the \textit{T. reesei} secretome from purified components has also demonstrated that differing amounts of the three activities are needed to digest cellulose substrates varying in their crystalline-to-amorphous content ratios [35]. In addition, mixtures of cellulases, such as those from \textit{T. reesei}, typically exhibit synergy. The overall conversion of cellulose to soluble sugars by a mixture of cellulases typically exceeds the conversion that would be expected by summing the contributions from the individual enzymes acting on the same substrate but in isolation [36]. One example of a cooperative effect is the cleavage of glycosidic bonds distant to the chain termini by EGs which creates new sites for CBHs to bind
Figure 1.1.2: Cellulase catalytic domains and their modes of interaction with cellulose. A–B: The CBHs \textit{Tr}Cel6A (PDB 1QJW) and \textit{Tr}Cel7A (PDB 4C4C) act from the non-reducing and reducing ends, respectively, of cellulose chains and require the removal of a single chain from the bulk cellulose structure. C: EGs, such as \textit{Melanocarpus albomyces} Cel7B (PDB 2RG0) act distant from cellulose chain termini but also require the decrystallization of a single cellulose chain. D: Lytic polysaccharide monooxygenases like \textit{Nc}LPMO9D (PDB 5TKG) can cleave cellulose chains oxidatively by direct action on cellulose surfaces. The cellulose monolayer shown is constructed from the crystal structure of cellulose I\textsubscript{\beta}, and the catalytic domain coordinates are taken from the respective PDB entries [37, 38]. (Placement of the domains on cellulose and of the chains undergoing hydrolysis was performed manually with reference to available structural data; the interactions depicted are only illustrative in nature.)

and begin processive hydrolysis. Another is the cellobiose-hydrolyzing activity of a \(\beta\)-glucosidase in a cellulase mixture. Despite not acting directly on the cellulose substrate, \(\beta\)-glucosidases reduce the concentration of free cellobiose in the reaction mixture and thereby decrease potent product inhibition of the CBHs by the disaccharide. The advantages of cellulase synergy have led to enzyme preparations consisting of the total \textit{T. reesei} secretome supplemented with a \(\beta\)-glycosidase from \textit{Aspergillus niger} (Cellulcast+Novozym 188, Novozymes, Inc.) being the enzymes most often applied in industrial enzymatic cellulose conversion.
During the last decade, a new cellulose deconstructing enzymatic activity, based not on hydrolysis but instead on oxidation, was discovered and quickly changed prior understanding of cellulase synergy. In fact, some have called these newly understood enzymes the C₁ cellulose remodeling component from Reese’s original C₁–Cₓ hypothesis for cooperative cellulose deconstruction [39, 40]. These enzymes are the lytic polysaccharide monooxygenases (LPMOs), of which an example is also shown in Figure 1.1.2. The history, classification, applications and current mechanistic understanding of LPMOs, particularly those of fungal origin, are discussed in detail below.

1.1.3 Discovery of LPMOs and LPMO–cellulase synergy

The first LPMO gene reported in the literature, cel1, was identified in *Agaricus bisporus* (button mushroom) by Raguz *et al.* in 1992 in a search for *A. bisporus* genes specifically induced by growth on cellulose as sole carbon source [41]. Cellulase genes were chosen as initial targets since both the levels of cellulase activity during the *A. bisporus* lifecycle were already known and a polyclonal anti-EG antibody was available at the time [42, 43]. *In vitro* translation of total polyadenylated mRNA isolated from cellulose-grown *A. bisporus* yielded three products that could be immunoprecipitated with the EG-reactive antibody. Immuno selection of colonies producing CEL1 from a cDNA fragment library constructed from the same mRNA pool and transformed into *E. coli* recovered the CEL1 open reading frame (ORF). Analysis of the *cell* sequence indicated that CEL1 includes a *N*-terminal secretion signal peptide and a *C*-terminal CBM which are common to fungal cellulose deconstructing enzymes. However, the 233-amino acid (AA) core of the protein could not be assigned any putative function or architecture due to a lack of homology with any contemporaneously known protein sequences.

In 1994, Armesilla *et al.* reported an initial characterization of CEL1 enzymatic activity and cellulose binding [44]. A recombinant glutathione *S*-transferase (GST)–CEL1Δ130 fusion protein was expressed in *E. coli* and used to raise a CEL1-reactive polyclonal antibody. Immunoprecipitates from cellulose-fed *A. bisporus* cultures were separated by either native or
partially-denaturing polyacrylamine gel electrophoresis (PAGE) and then tested for in-gel enzymatic activity with either the EG substrate carboxymethyl cellulose, the CBH substrate 4-methylumbelliferyl-β-D-glucoside or the β-glucosidase substrate barley β-glucan. CEL1 migration did not overlap with any detectable EG, CBH or β-glucosidase in-gel activity; however, CEL1 did exhibit binding to crystalline cellulose. These findings led Armesilla et al. to draw a series of prescient conclusions about A. bisporus CEL1: CEL1 is not an EG, CBH or β-glucosidase, CEL1 may not exhibit any hydrolytic activity toward cellulose and CEL1 may represent a new class of cellulose-binding proteins. These authors also predicted that CEL1-like proteins may occur commonly within fungal genomes as their anti-CEL1 antibody cross-reacted with a similarly-sized protein from the brown-rot fungus Coniphora puteana.

Contrary to the conclusion that A. bisporus CEL1 did not exhibit any classical cellulase activity, subsequent studies of related proteins attributed endoglucanase activity to these enzymes. Saloheimo et al. cloned the Trichoderma reesei gene egl4 and heterologously expressed extracellular EGIV in Saccharomyces cerevisiae [45]. Assays with the supernatants of fermentation cultures indicated that the secretome included an endoglucanase activity that was attributed to overexpressed EGIV. These results along with sequence homology and hydrophobic cluster analysis led to classification of EGIV and CEL1 as glycoside hydrolase family 61 (GH61) enzymes in the Carbohydrate-Active Enzymes (CAZy) database [46]. Homologously overexpressed T. reesei EGIV–His(6) fusion protein (renamed Cel61A–His(6)) purified by affinity and gel filtration chromatographies also cleaved β-(1→4) glucosidic bonds with low specific activity as was true of GH61 enzymes from Aspergillus nidulans and Aspergillus kawachii, both heterologously expressed in Pichia pastoris [47-49]. The shared low endoglucanase activity of all known GH61s was considered paradoxical to the strong cellulose induction of GH61 expression that was demonstrated by transcript analysis in T. reesei and the white-rot fungus Phanerochaete chrysosporium [50, 51]. It was unclear why fungi would conserve GH61 expression when other higher-activity endoglucanses were also induced by cellulose growth.
The first strong indications that GH61 enzymes may not be hydrolytic endoglucanases were provided by the crystal structure of *T. reesei* GH61B (PDB 2VTC) as shown in Figure 1.1.3 [52]. The structure, which was refined against 1.60 Å resolution X-ray diffraction, revealed a single domain primarily comprised of β-strands assembled into two stacked β-sheets, also known as a “β-sandwich” motif. Such a fold is common among fungal cellulose-active endoglucanases as can be seen from the structures of *Fusarium oxysporum* EG1 (PDB 1OVW) and *Humicola grisea* Cel12A (PDB 1UU6), also shown in Figure 1.1.3 [53, 54]. The *F. oxysporum* EG1 and *H. grisea* Cel12A structures, which had been determined prior to that of *T. reesei* GH61B, show the β-sheets arranged approximately parallel to the cellulose-interacting surface and shaping a deep cleft. Because these and other structures had been determined for enzyme–(pseudo)substrate complexes, it was known that the cleft is the site of glucan chain binding and glycosidic bond hydrolysis catalyzed by two conserved carboxylic residues. *T. reesei* GH61B, in contrast, exhibited β-sheets approximately perpendicular to an extended planar surface that also contained an intramolecular metal binding site with coordinating ligands provided by an N-terminal histidine along with His89 and Tyr176. (One additional metal ion per asymmetric unit is involved in crystal packing contacts.) There is no appearance of a substrate-binding cleft, and the conserved residues Glu160 and Asn206, which had been proposed to serve as the catalytic carboxylates for hydrolysis, are separated by > 12 Å (Glu160:Cδ–Asn206:Cγ) [45, 55].

The intramolecular metal binding site discovered in the *TrCel61B* structure was considered significant since the residues involved in metal binding are highly conserved among GH61 proteins known at the time. This, in conjunction with the indeterminate location and structure of the GH61 active site, led Karkehabadi *et al.* to speculate that the metal binding site could be a catalytic center [52]. The bound metal ion was modeled in the structure as Ni$^{2+}$ due to both the presence of 10 mM NiCl$_2$ in the crystallization condition and the appearance of sufficient anomalous diffraction at incident X-ray energies around the Ni K-absorption edge to facilitate successful single-wavelength anomalous dispersion initial phase determination. Nickel ions had no known functional roles in glycosyl hydrolases, and this remains the case.
Figure 1.1.3: T. reesei GH61B compared with other endoglucanases. A: GH61B (PDB 2VTC) shown perpendicular to and parallel to the cellulose interacting surface. B–C: similar views for F. oxysporum EG1 (PDB 1OVW) and H. grisea Cell12A (PDB 1UU6), respectively. Active site residues and substrates, where present, are shown as cylinders.
among the presently known Ni-dependent metalloenzyme activities [56]. Elemental analysis of purified GH61B in solution by proton-induced X-ray emission (micro-PIXE) indicated calcium as the only metal present in solution. As such, the authors of the structure concluded that Ni\(^{2+}\) occupancy in the binding site was likely a crystallization artifact; however, a possible role for coordinated Ca\(^{2+}\) was equally unclear.

The TrGH61B structure allowed an important connection to be made between fungal GH61 enzymes and a family of bacterial and insect-virus proteins known as cellulose binding module family 33 (CBM33). In 2005, Vaaje-Kolstad et al. determined the structure of the CBM33 protein named chitin binding protein 21 from the soil bacterium Serratia marcescens (SmCBP21, PDB 2BEM) [57]. Karkehabadi et al. identified a high structural homology between TrGH61B and SmCBP21 despite the two proteins having < 50% sequence homology [52]. Also, the initial report of the SmCBP21 structure, which was refined against 1.55 Å resolution X-ray diffraction, did not discuss the presence of a tridentate metal binding site formed by N-terminal His28 and His114 and located within an extended planar surface. However, electron density consistent with a metal is present at this site in two of the three non-crystallographically symmetric (NCS) copies of the chain in the asymmetric unit (ASU), and the density is modeled as a coordinated Na\(^{+}\) ion in NCS molecule C. A follow-up report by Vaaje-Kolstad et al. reported that SmCBP21 alone does not hydrolyze chitin but that the presence of SmCBP21 is critical for efficient hydrolysis by other chitinases [58]. At this point, GH61 and CBM33 high expression by organisms grown on cellulose or chitin, the either poor or absent substrate hydrolase activity and the overall structural similarity suggested that the two families of proteins shared an important but enigmatic function in carbohydrate polymer deconstruction.

A cellulose-active enzyme discovery effort led by the biotechnology company Novozymes Inc. confirmed a hydrolysis efficiency-enhancing role for fungal GH61 enzymes [59]. The primary objective of this project was to identify enzymes exhibiting additional synergy with the cooperative CBH, EG and \(\beta\)-glucosidase enzymes from Trichoderma reesei already in use industrially for cellulose deconstruction. Researchers established the enzyme-to-biomass (wt./wt.) ratio of partially purified cellulolytic secretome of T. reesei needed to
release 80-90% of total available glucose from pretreated corn stover as a baseline measure of cellulose deconstruction efficiency. They then combined *T. reesei* secretome with culture supernatants from other cellulolytic fungi grown on cellulose as carbon source. Equal mixtures of *T. reesei* secretome and culture supernatant from the thermophile *Thielavia terrestris* could achieve the same endpoint glucose release at half the enzyme-to-biomass ratio of *T. reesei* enzymes alone. Fractionating the *T. terrestris* secreted proteins revealed that cellulase synergy could be attributed to GH61 proteins.

Harris *et al.* extended this work by demonstrating that three *T. terrestris* GH61 proteins, *Tt*GH61B, *Tt*GH61E and *Tt*GH61G, could each enhance cellulose deconstruction when combined with classical cellulases from *T. reesei* [60]. X-ray crystal structures of *Tr*GH61E (PDB 3EII, 3EJA) revealed an overall fold homologous with those of *Tr*GH61B and *Sm*CBP21 and a histidine-coordinating metal site modeled with either Zn$^{2+}$ or Mg$^{2+}$ bound, respectively. These authors established a functional role for metal binding by conducting assays for cellulose hydrolysis with combined cellulases and *Thermoascus aurantiacus* GH61A in the presence of the chelator ethylenediaminetetraacetic acid (EDTA) or divalent metal cations. Assays containing cations showed cellulase enhancement while those containing EDTA showed no measurable enhancement. Furthermore, single residue mutations introduced near the *Tt*GH61E active site, as shown in Figure 1.1.4, either abolished (H1N, H68A and Q151L) or measurably decreased (Q151N and Q151E) cellulase enhancement.

Experiments with the bacterial *Sm*CBP21 were the first to identify carbohydrate oxidation catalyzed by GH61-like proteins. Vaaje-Kolstad *et al.* discovered that *Sm*CBP21, when incubated with β-chitin, could convert the crystalline chitin substrate into chitooligosaccharides. *Sm*CBP21 synergy with chitinases was also enhanced in the presence of electron donor [61]. Mass spectrometric analysis of the reaction products identified oligosaccharides bearing a carboxylic group at C1’ (reducing end) presumably formed during cleavage of chitin glycosidic bonds. Similar analysis of products from reactions performed in the presence of H$_2^{18}$O or $^{18}$O$_2$ revealed that one oxygen atom of the carboxylic group is derived
Figure 1.1.4: Sites of single residue mutations introduced in TrGH61E (PDB 3EII). Mutation of either His1 or His68 abolished GH61 activity as did a non-conservative mutation of Gln151. Mutations of the other residues illustrated produced measurable decreases in GH61 activity.

from H₂O while the other is incorporated from atmospheric O₂. Such an oxidative activity promoting polysaccharide chain cleavage was previously unknown for any enzyme and was difficult to reconcile with the previous reports of redox-inactive ions (Zn²⁺ and Mg²⁺) occupying the SmCBP21 active site. Nonetheless, these findings began a rapid succession of new reports describing oxidative polysaccharide deconstruction [62-66].

Quinlan et al. demonstrated that the Thermoascus aurantiacus GH61 protein TaGH61A could catalyze oxidative cellulose chain cleavage with phosphoric acid swollen cellulose (PASC) as substrate when reactions were supplemented with an electron donor [62]. Given the variety of metals previously reported to bind to GH61 and CBM33 proteins, this group also performed binding assays with extensively demetallated TaGH61A and the ions Mg²⁺, Ca²⁺, Mn²⁺, Co²⁺, Ni²⁺, Zn²⁺ and Cu²⁺. Of all the ions tested, only Cu²⁺ showed evidence
of binding, with a 1:1 Cu$^{2+}$:protein stoichiometry as monitored by isothermal titration calorimetry (ITC). In fact, TaGH61A exhibited such high affinity for Cu$^{2+}$ that the dissociation constant ($K_D$) for Cu$^{2+}$ binding could not be determined by ITC measurements, and, as such, was estimated to be $< 1 \text{nM}$. Further experiments measuring the rate of EDTA chelation from holo-TaGH61A ($< 0.001 \text{s}^{-1}$) suggested that the Cu$^{2+} K_D$ could be $< 1 \text{pM}$. Quinlan et al. concluded that these findings explain the early ambiguity regarding GH61 and CBM33 metal requirements since trace amounts of copper, either remaining bound after attempted chelation or introduced to the apoprotein from reagents, could have permitted the activity detected in the presence of other metal ions. These authors also modeled the metal ion as Cu$^{2+}$ in the X-ray crystal structure determined for TaGH61A (PDB 2YET). Copper binding along with the known participation of atmospheric O$_2$ in the SmCBP21 oxidation reaction led to the first proposal that a reactive copper–dioxygen complex forms to facilitate glycosidic bond oxidation. Phillips et al. proposed that proteins having this enzymatic activity be designated as “polysaccharide monooxygenases” (PMOs) [64]. (The name has since been amended to “lytic polysaccharide monooxygenase” (LPMO) to reflect the importance of the activity in lysing carbohydrate polymer chains.)

Experiments by Langston et al. demonstrated that T. aurantiacus and T. telestris LPMOs, when supplemented with electron donor, act synergistically with CBHs, EGs and β-glucosidases in isolation and in mixtures [63]. Therefore, specific protein–protein interactions between LPMOs and a class of hydrolase could not explain the observed synergy. In addition, the amount of synergistic enhancement contributed by the LPMO varied according to the crystallinity of the cellulose substrate. These results suggested that LPMO activity directly on insoluble carbohydrate substrates increases accessibility for the classical cellulases leading to enhanced hydrolytic efficiency. The approximately planar surface observed surrounding the metal ion binding site in LPMO structures was envisioned as a substrate adsorption site promoting direct LPMO–surface interaction, and prior results supported the existence of the interaction. When reporting the crystal structure of SmCBP21, Vaaje-Kolstad et al. identified residues on this surface that both show conservation among CBM33 proteins and facilitate chitin binding as demonstrated by site-directed mutagenesis [57]. Li et al. suggested a similar
role for the planar surface in cellulose-active LPMOs by comparing the arrangements of aromatic side chains around the metal binding sites in the structures of TrGH61E, TaGH61A and NcPMO-3 along with the cellulose binding surface of the CBM from T. reesei [67]. For each structure, the aromatic residues could be aligned atop glucosyl rings arranged according to the cellulose Iβ (common plant allomorph) crystal structure.

Eibinger et al. revealed direct crystalline cellulose surface degradation by a fungal LPMO using confocal fluorescence and atomic force microscopies [68]. These authors developed a “mixed amorphous and crystalline” cellulose (MAC) substrate with a nano-flat surface (< 20 nm surface height variation per 4 μm²) suitable for measurement with atomic force microscopy (AFM). Thin sections of the MAC substrate were incubated with N. crassa LPMO9F in the presence of excess ascorbic acid, treated with the carboxylate-reactive fluorescent dye SYTO-62 and scanned through their depth using confocal scanning laser microscopy. Since NcLPMO9F has regiospecificity for oxidation of C1’ to yield aldonic acids, the SYTO-62 specifically stained sites of LPMO oxidation of the substrate. Localization of fluorescence revealed that oxidation was confined to the surfaces of cellulose crystallites and thereby confirmed that LPMOs adsorb to and react on crystalline substrate surfaces. AFM measurements of MAC substrates before and after LPMO incubation showed that LPMO action alone produced very little removal of material from crystalline surfaces with isolated nanoscale fibrils being the only features solubilized. In contrast, when LPMO-treated MAC substrate samples were then exposed to a CBH-I (T. reesei Cel7A), large regions of crystallites were solubilized.

However, Ebinger et al. did not attribute LPMO synergy in cellulose deconstruction solely to increased susceptibility of crystalline regions to CBH-I hydrolysis [68]. In their MAC substrate assays, free glucose release by CBH-I from LPMO-treated MAC substrates was approximately threefold reduced compared to the CBH-I mediated release from “as prepared” substrate. Simultaneous exposure of MAC substrate to LPMO and CBH-I also did not produce synergistic glucose release compared with CBH-I alone. But sequential and simultaneous LPMO and CBH-I treatment did produce synergistic glucose release from suspensions of cellulose nanocrystals and from the highly crystalline Avicel PH-101 substrate. The authors
rationalize these conflicting results as consequences of “mesoscopic” morphology differences in the distribution and accessibility of nano-crystalline regions within the substrates. It should be considered that simultaneous experimental choices of a C1′ regiospecific LPMO, D-glucose release as the synergy metric and glucose detection by a glucose oxidase–peroxidase coupled enzyme assay may have complicated the interpretation of results. Nonetheless, such a nuanced interpretation of the LPMO– and cellulase–substrate interactions necessary for synergy is consistent with experimental determinations of substrate effects and protein–substrate binding reported by Saddler and coworkers [69, 70]. In addition, very recent work by Villares et al. demonstrates that LPMOs can also effect ultrastructural changes in cellulose fibers at micrometer-to-millimeter length scales [71]. Ultimately, a detailed characterization of LPMO–cellulase synergy determinants awaits further multiscale (and likely multi-technique) experiments involving varied substrate types and LPMOs.

1.1.4 Classes of LPMOs

Currently, known LPMOs are organized into four known families within the “Auxiliary Activities” class of the CAZy database designated AA9–11 and AA13 [72-75]. LPMOs are often multi-domain proteins fused to C-terminal CBMs [76]. However, this domain architecture is highly modular with respect to enzyme function, as is the case for many families of carbohydrate-active enzymes [77]. Therefore, the presence or type of CBM is not considered for classification purposes.

Typical structural characteristics of the N-terminal catalytic domains of proteins from each class are compared in Figure 1.1.5. Shared structural features are presented with respect to AA9 LPMOs with differences specific to each class discussed separately. The mononuclear copper ion-coordinated active site is the most strictly conserved feature of LPMO structure across all classes. This coordination site, which is commonly described as having a “histidine-brace” motif, requires bidentate coordination by the amino and Nδ nitrogens of an N-terminal histidine along with the Nε nitrogen of another histidine residue typically occurring between 60–100 residues later in the protein sequence [62]. The histidine-brace establishes a roughly
“T-shaped” coordination in the copper equatorial coordination plane that is readily apparent in electron density maps from X-ray crystallography. AA9 and AA13 LPMOs expressed homologously or in filamentous fungal hosts, such as A. oryzae, show a characteristic N-methylation of $N_\varepsilon$ of the $N$-terminal histidine [62, 67, 75]. This post-translational modification is not observed in AA10 LPMOs, the one structure known for AA11 LPMOs or AA9 LPMOs expressed in yeasts [78]. The functional significance of His1 methylation is unknown, and heterologously-expressed AA9 enzymes lacking His1 methylation retain enzymatic activity [78]. Typically, the AA9 active site also includes a tyrosine hydroxy group at distances ~2.75 Å oriented toward an axial Cu coordination site perpendicular to the plane of the histidine-brace. This ligand arrangement resembles an octahedral $\text{Cu}^{2+}$ coordination environment with Jahn–Teller distortion producing elongation in the axial $\text{Cu}^{2+}$–ligand distances.

Kjaergaard et al. established that in an aerobic environment the LPMO copper ion predominantly exists in the $\text{Cu}^{2+}$ oxidation state due to the rapid oxidation of $\text{Cu}^+$ by molecular oxygen, but initial structural characterizations of LPMO active sites provided limited information about the additional ligands expected for (pseudo-) octahedral $\text{Cu}^{2+}$ due to the readiness with which the $\text{Cu}^{2+}$ ion reduces to $\text{Cu}^+$ upon exposure to synchrotron X-rays [79, 80]. Electron paramagnetic resonance (EPR) spectra of the paramagnetic $\text{Cu}^{2+}$ in AA9 LPMOs revealed a tetragonal coordination environment formed by nitrogen and oxygen ligands with spectral parameters satisfying the definition of a Peisach–Blumberg type II copper center [62, 81]. Type II copper coordination environments have low covalent character and exhibit only weak visible absorption around 600 nm ($\varepsilon \approx 100 \text{ cm}^{-1} \text{ M}^{-1}$). (Compare with the type I coordination of “blue copper” proteins which is roughly tetrahedral in geometry and exhibits strong visible absorption ($\varepsilon \approx 5,000 \text{ cm}^{-1} \text{ M}^{-1}$) due to a low energy ligand–metal charge transfer band formed by covalent interaction between $\text{Cu}^{2+}$ and a thiolate ligand from a cysteine side chain [82].) Low-dose extended X-ray absorption fine structure (EXAFS) spectra acquired for T. aurantiacus LPMO9A and density functional theory (DFT) calculations calibrated to the EXAFS spectra agreed with the presence of four nitrogen or oxygen ligands in an equatorial square planar coordination arrangement [80]. AA9 X-ray crystal structures collected under
dose limiting conditions, along with one joint X-ray/neutron crystal structure, discussed at length in later chapters, have confirmed that, in the absence of other exogenous ligands, the Cu\(^{2+}\) coordination sphere is completed by water molecules occupying the one equatorial and one axial coordination sites not occupied by the histidine-brace and the conserved active site tyrosine (see Figure 1.1.3) [83, 84].

As discussed in 1.1.3, Li et al. hypothesized that conserved aromatic residues, often tyrosines, on the AA9 extended planar surface mediate binding to crystalline cellulose and align the LPMO active site over glycosidic bonds via aromatic–pyranose ring stacking interactions [67]. This binding mode is somewhat supported by the X-ray crystal structure of *Lentinus similis* LPMO9A in complex with cellohexaose, which is currently the only known LPMO–substrate complex structure [85]. In this structure, a surface-exposed tyrosine stacks with a glucose unit that is three units away from the glycosidic bond positioned for monooxygenation. However, *Ls* LPMO9A is unique in that it is one of only two LPMOs known to act on soluble oligosaccharides (discussed below), and five potential hydrogen bonds between *Ls* LPMO9A and cellohexaose identified in the structure may influence or dominate the substrate binding interaction.

Classes AA9 and AA10 include proteins previously classified into GH61 or CBM33, respectively. Specifically, the AA9 class includes fungal enzymes with oxidative activity on carbohydrates found in plant cell walls. Cellulose originally was the only known substrate for AA9 activity. However, certain AA9 enzymes, such as *N. crassa* LPMO9C and *Podospora anserina* LPMO9H, have demonstrated activity on β-(1→4)-linked carbohydrate polymers that include sugars other than glucose, branching through additional glycosidic bonds, incorporation of β-(1→3) linkages or combinations of these modifications [86-88]. Two AA9 LPMOs, *Nc* LPMO9C and *L. similis* LPMO9A, also show activity on soluble cello-oligomers ranging from cellobriose to cellohexaose [85, 89]. Vu et al. used product identification, sequence alignment and structural homology of *N. crassa* LPMOs to categorize the regiospecificity of glycosidic carbon oxidation by AA9 LPMOs into four types: Type 1 AA9 LPMOs only oxidize C1’. Type 2 AA9 LPMOs only oxidize C4. Type 3 AA9 LPMOs can
Figure 1.1.5: Structural variation of LPMOs from classes AA9, AA10, AA11 and AA13. The active sites and solvent accessible surfaces of A: *N. crassa* LPMO9D (PDB 5TKG), B: *E. faecalis* LPMO10A (PDB 4ALC), C: *A. oryzae* LPMO11A (PDB 4MAI) and D: *A. oryzae* LPMO13A (PDB 4OPB). The active sites for AA9–11 are located in a relatively planar surface while the characteristic AA13 surface groove is apparent for AoLPMO13A.
oxidize either C1′ or C4. A subset of Type 3 with sequence and structural similarity only oxidizes C1′ and was designated Type 3* [90].

The AA10 class is comprised of enzymes primarily bacterial in origin that exhibit activity on cellulose, chitin or in some cases both substrates [91]. Fusolin and GP37 proteins from insect entomopoxviurses and nucleopolyhedroviruses, respectively, are notable exceptions to the bacterial origin of the AA10 class. Their presumed LPMO activity on chitin, based on the crystal structure of Anomala cuprea entomopoxvirus fusolin, is hypothesized to confer the ability of these proteins to enhance virulence in insect hosts [92, 93]. Phylogenetic analysis of bacterial AA10 LPMO sequences by Book et al. identified two clades that essentially separate the chitin-specific and the cellulose- or chitin-reactive activities [76]. Within the clades, the substrate binding preferences of fused CBMs, when present, were also enriched in specificity for chitin in the chitin-specific clade and for cellulose in the cellulose- or chitin-reactive clade. However, Forsberg et al. demonstrated that AA10s can bind to non-substrate carbohydrates as cellulose-active Streptomyces coelicolor CelS2 (ScLPMO10C) shows appreciable binding to different chitin allomorphs [94]. AA10 LPMOs predominantly perform regiospecific oxidation of C1′; however, products oxidized at both C1′ and C4 have been demonstrated for S. coelicor LPMO10B reacting with cellulose and proposed for the Streptomyces lividans LPMO10E reacting with β-chitin [91, 94, 95].

It has been suggested that differences both in AA10 active site structure and in the residues presented on the extended planar surface, relative to AA9 LPMOs, may give rise to the greater propensity for C1′ regiospecificity and the common occurrence of oxidation products having only even numbers of hexose units [96, 97]. From X-ray crystal structures of Enterococcus faecalis CBM33A (EfLPMO10A) determined at a series of total radiation doses, Gudmundsson et al. demonstrated that Tyr-to-Phe substitution in the AA10 active site creates a trigonal bipyramidal Cu²⁺ coordination geometry comprised of three ligands from the histidine-brace and two from water molecules [98]. This coordination geometry is also reflected in EPR studies of AA10 LPMOs as a deviation from classical Peisach–Blumberg spectral parameters for a type II copper center [91, 94]. Access to the Cu²⁺ site perpendicular to the plane of the histidine-brace is sterically hindered by the phenylalanine side chain and
the close proximity of an alanine residue as was observed previously in the X-ray crystal structure of *Bacillus amyloliquefaciens* CBM33 (BaLPMO10A) [97]. The steric hindrance imposed by these residues was hypothesized to limit substrate accessibility to Cu\(^{2+}\) and, thereby, limit substrate reactivity to C1'. In addition, Hemsworth *et al.* identified numerous highly conserved side chains capable of participating in hydrogen bonding on the extended planar surface of *Ba*LPMO10A instead of tyrosines or other aromatic sidechains. Hydrogen-bond mediated substrate interaction, as opposed to ring stacking interactions, was proposed as a possible explanation for the abundance of even-numbered AA10 oxidation products since the two-fold screw axis along the length of \(\beta-(1\rightarrow4)\) linked glycan chains would present hydrogen bonding groups at every other glycosyl unit [97].

Class AA11 consists of fungal chitin-active LPMOs of which currently one has been characterized biochemically and structurally, *Aspergillus oryzae* LPMO11A. Hemsworth *et al.* discovered the AA11 class by a “module walking” sequence analysis approach that examined *A. oryzae* genes containing a cysteine-rich domain of unknown function that occurs within some AA9 full-length sequences [73]. *Ao*LPMO11A heterologously expressed in *Escherichia coli* oxidized C1’ of chitin but showed no activity on cellulose, starch or mannans. The crystal structure of *Ao*LPMO11A, refined against 1.55 Å resolution X-ray diffraction, revealed a \(\beta\)-sandwich fold characteristic of LPMOs despite very low sequence homology between *Ao*LPMO11A and members of the AA9 and AA10 classes. If the characteristic active site features of AA9 and chitin-active AA10 enzymes are taken as extremes, the *Ao*LPMO11A active site (see Figure 1.1.5) has a structure with intermediate resemblance to both AA9 and AA10. Like the AA9 LPMOs, one axial coordination site of the *Ao*LPMO11A catalytic copper ion faces a tyrosine hydroxy group while the other can be occupied by exogenous ligands. However, like the chitin-active AA10 LPMOs, the axial exogenous ligand binding site is closely bounded by an alanine residue, and the Cu–O\(_{\text{Tyr}}\) distance of 3.1 Å essentially leaves the other axial site vacant as is often the case in AA10 LPMOs due to Tyr-to-Phe substitution [97]. This intermediate character with respect to the AA9 and AA10 classes may reflect evolutionary pressure for the fungal-derived AA11 LPMOs to gain activity on chitin and, as such, evolve toward an AA10-like active site structure.
Class AA13 LPMOs differ from classes AA9–11 in two contexts. The first is a much greater sequence homology among AA13 enzymes as identified from the gene sequence analyses that prompted the formation of the class [74, 75]. The second, and perhaps more consequential, distinction of class AA13 LPMOs is their substrate specificity for carbohydrates having α-(1→4) linkages including amylose, amylopectin and starch [74]. As revealed by the crystal structure of A. oryzae LPMO13A, refined against 1.50 Å X-ray diffraction, the overall fold and active site structure of AA13 LPMOs is highly similar to known structures from AA9–11 [75]. However, unlike the extended planar surfaces observed to encompass the active sites of AA9–11 LPMOs, AoLPMO13A surface residues shape a shallow groove that lies directly above the active site copper ion. Vu and Marletta demonstrated with manual coordinate docking that this groove could accommodate a double helix of amylose chains with glycosidic bond carbons positioned at the AA13 active site [99]. To date, AA13 enzymes are the only LPMOs known to have substrate specificity for the curved arrangement of hexose monomers produced by α-(1→4) glycosidic bonds. The higher-order structures that can result from α-(1→4) linkages, ranging from amylose helices to microscopic starch granules, starkly contrast with the extended planar surfaces formed by crystallization of the β-(1→4)-linked cellulose and chitin. As such, the discovery of class AA13 significantly extended the paradigm of LPMO activity while simultaneously expanding the potential applications of LPMOs to industries making use of amylose- and amylopectin-based carbohydrates [99].

It should be noted that all further mention of LPMOs (or PMOs) will refer to fungal, cellulose-specific AA9 LPMO enzymes unless stated otherwise. The research presented herein has exclusively focused on AA9 LPMOs as these enzymes have attracted the most attention to date for immediate application in industrial biomass conversion processes. Certain aspects of reactivity and applications can be generalized across classes AA9–11,13; however, the following discussion is not inclusive of all exceptional characteristics of classes AA10–11,13.
1.1.5 LPMOs in industrial lignocellulose conversion

Cannella et al. published the first literature report describing LPMO activity in commercial cellulase enzyme cocktails for lignocellulosic ethanol production in 2012, only two years after the seminal experiments of Vaaje-Kolstad et al. and Harris et al. [60, 61, 100]. Incubating the Cellic CTec2 enzyme cocktail (Novozymes, Inc.) with hydrothermally pretreated wheat straw yielded 85% cellulose conversion to soluble sugars with ~4% conversion of total glucose to gluconic acid by the action of a C1’-oxidizing LPMO. Despite the loss of glucose to gluconic acid, CTec2 outperformed the cellulose conversion of the prior generation cocktail of Cellulcast+Novozym 188 (Novozymes, Inc.), which did not produce measurable concentrations of oxidized sugars and yielded 60% cellulose conversion at the same enzyme activity:biomass loading and time endpoint. Even though a small molecule reducing agent was not added to these reactions, Cannella et al. concluded that CTec2 does not include a CDH enzyme to serve as electron donor to the LPMO(s). Mixing CTec2 with a solution of pure cellobiose did not produce cellobionic or gluconic acid which would be expected from CDH activity in the presence of a β-glucosidase. Follow-on experiments by the Müller et al. demonstrated that maximal conversion of pure forms of cellulose by CTec2 required the addition of a reducing agent [101]. However, oxidized sugar production and, more importantly, cellulose conversion from different types of pretreated lignocellulose positively correlated with the residual lignin content of the feedstock [101-103]. The necessity of lignin for LPMO-activated cellulase cocktails produces the added benefit that lignin-retaining pretreatments, such as steam explosion or hydrothermal treatment, are typically more biocompatible and less expensive than lignin-removing pretreatments such as organic solvent washing (organosolv) or alkaline treatment [103].

LPMO inclusion in lignocellulose hydrolysis has also prompted reevaluation of optimal conditions for fermentation of soluble sugars to ethanol or to other chemical products. As reviewed by Wyman, lignocellulose processing relying upon treatment with a cellulase cocktail and a separate fermentation organism typically proceeds via one of two approaches, namely, separate hydrolysis and fermentation (SHF) or simultaneous saccharification and
fermentation (SSF) [104]. For SHF, pre-treated biomass is incubated with cellulases under optimal conditions for enzyme activity until the desired cellulose conversion endpoint is reached; at this point, the glucose-rich broth is transferred to a separate bioreactor optimized for fermentation of the desired product. In contrast, under the SSF approach cellulases and an inoculum of the fermenting organism are introduced simultaneously to the pre-treated biomass and reaction conditions are optimized around the growth phenotype of the fermenting organism. Despite the requirement to perform SSF away from the temperature and pH where most cellulases exhibit maximum activity, SSF processes typically yield a greater final concentration of fermentation product since the continuous consumption of the released glucose prevents product inhibition of the cellulases and yields greater cellulose conversion.

Unlike hydrolytic cellulases, LPMO activity requires dissolved oxygen in the reaction mixture, and the LPMO oxygen demand competes with oxygen availability to the fermenting organism in SSF reactions. For processes employing Crabtree-positive organisms like *S. cerevisiae*, simultaneous aerobic and fermentative metabolism reduces available dissolved oxygen and limits LPMO-derived cellulase synergy [101, 102]. SSF processes reliant upon Crabtree-negative yeasts, such as *Pichia stipitis* which can ferment pentose sugars released from hemicellulose to ethanol, must optimize around anaerobic or microaerobic conditions to achieve fermentation which would also limit or eliminate LPMO synergy [105]. Also, some fermentation approaches to producing other chemical products such as bacterial production of lactic acid by *Pediococcus acidilactici* or *Bacillus coagulans* benefit in reducing side product generation by reducing dissolved oxygen concentration to limits that diminish LPMO activity [106]. As such, realizing increases in cellulose conversion due to LPMO activity in current commercial cellulase cocktails would seem to require a revival of the older and more process-intensive SHF approach so that optimal dissolved oxygen concentrations can be maintained for both LPMOs and fermenting organisms.
1.1.6 LPMO electron donors

LPMO enzymes require an input of two electrons to complete the monooxygenation reaction with substrate. Redox titrations of *N. crassa* LPMO9C with *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine and numerous quinones have determined the midpoint reduction potential for the LPMO-bound Cu\(^{2+}/Cu^{1+}\) couple to be in the range of 224–250 mV (versus standard hydrogen electrode, SHE) [107, 108]. Monooxygenase reactivity was initially discovered using a molar excess of ascorbic acid as reducing agent for the catalytic Cu\(^{2+}\) ion [61]. Various other small molecule reductants, including gallic acid and pyrogallol, can potentiate LPMO activity [62, 75, 108]. Currently, for all known LPMO electron donors, it is unclear whether intramolecular electron transfer through the LPMO structure is relevant and which sidechains might participate in such transfer.

Langston *et al.* and Phillips *et al.* demonstrated early on that the often co-secreted fungal cellobiose dehydrogenases (CDHs) can potentiate LPMO reactivity, and CDHs have become accepted as biologically relevant electron donors for fungal LPMOs [63, 64]. (CDHs have not been observed in bacteria; however, CDH can provide electrons to bacterial LPMO10s [109].) CDHs are secreted flavocytochromes belonging to the glucose–methanol–choline (GMC) oxidoreductases and are dual classified in the CAZy AA3 and AA8 classes [72]. CDHs have a characteristic domain architecture of an *N*-terminal cytochrome (CYT) domain, a dehydrogenase (DH) domain including the active site and a prosthetic flavin adenine dinucleotide (FAD) and, in some cases, a *C*-terminal cellulose-specific CBM. The CYT and DH domains are connected by long (> 20 AAs), flexible linkers that permit considerable interdomain conformational flexibility [83]. CDHs preferentially catalyze the two-electron oxidation of cellobiose to cellubiono-1,5-lactone, which spontaneously hydrates to yield cellubionic acid, and the concomitant reduction of FAD to FADH\(_2\) [110]. Reoxidation of FADH\(_2\) can occur by a single two-electron transfer from the DH domain to a terminal electron acceptor or by two successive single-electron transfers mediated by internal electron transfer from the DH domain to the heme *b* cofactor of the CYT domain. The reduced CYT domain is, in turn, an effective one-electron donor to LPMOs with midpoint reduction potentials in the
range of 90–160 mV (versus SHE) [111]. Two-electron transfer from the CDH DH domain to LPMOs does not occur [64, 83].

The biological relevance of CDH-to-LPMO electron transfer is most strongly supported by genomic and biochemical results recently reported by Kracher et al. [108]. Analysis of 97 fungal genomes revealed that cdh genes are present only when one or more lpmo genes are also present. LPMO activity assays with each of 17 different potential small molecule electron donors or with CDH demonstrated 4–5 fold increased LPMO product generation in the presence of CDH. Despite the apparent preference of LPMOs for CDHs as electron donors when both enzymes are present, details about CDH–LPMO protein–protein interactions remain unknown. In their report of the X-ray crystal structures of NcLPMO9D and NcLPMO9M, Li et al. identified conserved surface residues spatially removed from the extended planar surface which they proposed could form a CDH recognition site [67]. In contrast, Courtade et al. report from nuclear magnetic resonance (NMR) chemical shift perturbations upon titration of NcCDHIIA into NcLPOM9C that only residues involved in or immediately adjacent to the LPMO histidine-brace show conformational changes upon CDH–LPMO interaction [112]. The uncertainty as to the nature of CDH–LPMO complexes is perpetuated by the apparently low affinity, and likely transient residence, of the binding interaction which cannot be characterized by thermodynamic techniques like ITC or kinetic techniques like surface plasmon resonance (SPR) [112, 113].

Kracher et al. have also revealed a potential for expanded involvement of additional GMC oxidoreductase enzymes in LPMO reactivity via regeneration of plant phenols that can serve as effective LPMO electron donors [108]. Of 16 plant-derived phenols tested, 11 had more negative reduction potentials than the LPMO Cu²⁺/Cu⁺ couple and could promote LPMO monooxygenation of substrate. Oxidized phenols could be reductively regenerated by co-incubated glucose oxidase, glucose dehydrogenase or pyranose dehydrogenase, and the LPMO–phenol–GMC three-component systems yielded greater amounts of LPMO reaction products than LPMO–phenol two-component systems. These findings led Kracher et al. to postulate that LPMO activity evolved with an initial dependence on plant-derived phenols as one-electron reductants. This activity was then augmented by participation of GMC
oxidoreductases for phenol redox cycling leading to the eventual emergence of CDH which, unlike other GMCs, combines the oxidoreductase activity of its DH domain with a covalently bound one-electron transfer mediator in the form of the CYT domain.

Prior to the elucidation of the LPMO–phenol–GMC system, Hu et al. suggested lignin-derived phenolic compounds as potentiators of LPMO activity on lignocellulosic biomass [69]. Experiments with pre-treated corn stover, softwood and hardwood materials demonstrated that as little as ~12% (wt./wt.) residual lignin in lignocellulose could saturate the LPMO reducing agent requirement as measured by synergistic enhancements in soluble sugar release by cellulases. Similar participation of lignocellulose-derived reducing agents explains the early observations of Harris et al. and Quinlan et al. that LPMOs, in the absence of added reducing agent, promoted greater cellulase synergy in the deconstruction of lignocellulosic materials as compared with relatively pure cellulose [60, 62]. Hu et al. also determined that residual or added hemicellulose content in lignocellulose could promote LPMO activity; however, this effect was primarily attributed to the existence of covalently cross-linked lignin–carbohydrate complexes existing within the residual hemicellulose or added birchwood xylan [69]. As discussed above in 1.1.5, commercial cellulase cocktails for lignocellulose hydrolysis that include LPMO(s) rely on residual lignin content as the LPMO electron donor.

Cannella et al. have explored innovative combinations of LPMOs with photosynthetic membranes and chlorophyll derivatives as sources of the necessary reducing electrons [114]. Both T. terrestris LPMO9E and T. aurantiacus LPMO9A when combined with isolated cyanobacterial thylakoids, isolated plant thylakoids or chlorophyllin and ascorbate (for chlorophyll re-reduction after photoexcitation) exhibited light-activated monooxygenase activity that exceeded the LPMO–ascorbate system activity by 20–100 fold as judged by rate of soluble oligosaccharide release, percent conversion of total glucose to gluconic acid or rate of O$_2$ consumption. In the absence of added ascorbate, residual lignin in biomass could effect photocenter reduction and maintain light-driven cellulose oxidation by the LPMOs. Furthermore, in the case of TtLPMO9E, the input of electrons from the strongly reducing activated photocenters potentiated monooxygenase activity on xylan that is not observed in the TtLPMO9E–ascorbate system. Cannella et al. hypothesize that the photoelectrons from the
photosynthetic systems, as compared with electrons from small molecule reductants or CDH, could possibly overdrive \( TtLPMO \) \( O_2 \) activation to the point that sufficient concentrations of reactive oxygen species exist to achieve Fenton-type reactions with non-substrate polysaccharides as was observed in the case of xylan.

1.1.7 The LPMO reaction mechanism

As discussed in 1.1.3, the LPMO active site requires a single copper ion for enzymatic activity. This 1:1 copper:protein stoichiometry situates LPMOs within the bioinorganic chemistry context of the mononuclear and the non-coupled binuclear copper proteins [82]. Other examples of these metalloenzymes include the mononuclear galactose oxidase (GO) and the non-coupled binuclear monooxygenases (NCBMs) dopamine \( \beta \)-monooxygenase (D\( \beta \)M), peptidylglycine \( \alpha \)-hydroxylating monooxygenase (PHM) and tyramine \( \beta \)-monooxygenase (T\( \beta \)M). Since reduction of Cu(II) to Cu(I) only supplies one electron for use in reducing molecular oxygen, each of these enzymes make use of prosthetic redox cofactors to provide the second electron needed for reducing molecular \( O_2 \). In the case of GO, the matured form of the enzyme catalyzes an oxygen-dependent biogenesis reaction that yields an ortho-thioether bond between a cysteine sidechain and a tyrosine involved in coordinating Cu(II/I). This Cys–Tyr moiety can act as a stabilized free radical one-electron sink [115]. The NCBMs use a second copper ion (Cu\( H \)) spatially separated (\( > 10\, \text{Å} \)) from the copper ion (Cu\( M \)) where oxygen and substrate binding occur. One-electron reduction of both ions to Cu(I) along with an intramolecular Cu\( H \)-to-Cu\( M \) electron transfer are necessary to complete the reaction cycle [116, 117]. As such, LPMOs appear to perform monooxygenation at a mononuclear Cu(II/I) site without the involvement of a prosthetic redox cofactor and must rely on sequential one-electron transfers to complete the reaction cycle.

Phillips et al. were the first to propose a LPMO reaction mechanism accounting for substrate oxidation and glycosidic bond cleavage [64]. According to this mechanism, a one-electron transfer to LPMO–Cu(II) reduces the copper ion to LPMO–Cu(I). Inner sphere oxidation of LPMO-Cu(I) proceeds by the addition of molecular oxygen which forms a
LPMO–Cu(II)–O–O' complex as shown in Figure 1.1.6 A. (Inner sphere oxidation of Cu(I) to Cu(II) by O_2 is consistent with the oxidation rate measured by Kjaergaard et al. [80].) Superoxide radical then abstracts a hydrogen atom from the glycosidic carbon subject to attack, either C4 or C1', forming a LPMO–Cu(II)–OOH complex (Figure 1.1.6 C). The distal oxygen then participates in “radical rebound” through homolytic cleavage of the O–O bond followed by coupling with the glycosidic carbon radical thus forming a new C–O bond. The input of an additional electron and addition of two protons (H^+) to the resulting LPMO–Cu(II)–O' species reduces the remaining oxygen from molecular O_2 to H_2O and regenerates LPMO–Cu(II). The additional oxygen atom at either C4 or C1’ electrophilically attacks the carbon leading to elimination of the glycosidic bond and the formation of a 4-ketoaldose and a free reducing end or an aldono-1,5-lactone and a free non-reducing end, respectively. (Both the 4-ketoaldose and aldono-1,5-lactone species equilibrate with hydrated forms in aqueous solution forming a population of either a 4-gemdiol or an aldonic acid species, respectively [89].)

This superoxide radical abstraction mechanism strongly resembles the accepted mechanism of the NCBMs DβM and PHM [118]. Extensive and elegant examination by Klinman and coworkers of various kinetic isotope effects on NCBM reaction rates yielded this mechanistic understanding. Miller and Klinman provided initial evidence that hydroxylation proceeded from the homolytic cleavage of the substrate C–H bond producing a carbon radical intermediate [119]. Later studies demonstrated that hydrogen abstraction from the substrate C–H must exhibit chemical reversibility with O–O bond cleavage in the activated Cu–O_2^2 complex [120, 121]. The possibility of Cu–OOH as the abstracting species was dismissed after experiments with a series of poorly reactive NCBM substrates showed no release of OOH’ or formation of H_2O which would be expected at a solvent-exposed coordination site if Cu–OOH could form prior to H abstraction [117, 122]. In addition, quantum chemical calculations predicted abstraction by Cu(II)–OOH’ to have an activation barrier ~23 kcal mol^{-1} greater than abstraction by Cu(II)–OO’. [123]. It should be noted that LPMOs do perform substrate uncoupled H_2O_2 production presumably via associative displacement of the OOH’ radical by a coordinated water molecule (Figure 1.1.6 B) [78, 80]. However, this experimental result lacks the same mechanisitic significance for LPMOs as the absence of uncoupled oxygen
Figure 1.1.6 Proposed mechanisms for LPMO monooxygenation. (A-B) Oxygen activation begins from a Cu(II) complex in aqueous oxygenated solution due to the futile H$_2$O$_2$ cycle that takes place in absence of substrate. (C-D) Both Cu–superoxy and Cu–oxyl radical complexes have been proposed as responsible for abstracting a hydrogen atom from substrate.
activation does for NCBMs. NCBMs exhibit ordered binding of substrate followed by molecular O$_2$ which implies that the activated copper–oxygen species could only gain hydrogen by abstraction from substrate [120]. In contrast, ordered binding has not been demonstrated for LPMOs, and H$_2$O$_2$ production appears to be a shunt for activated oxygen species that is detectable in substrate-free assays containing enzyme, oxygen and reducing agent but that cannot be detected with substrate added [78, 89].

An alternative LPMO reaction mechanism in which substrate hydrogen abstraction occurs after reductive cleavage of the superoxide O–O bond has received considerable attention after initial computational investigation by Kim et al. [124] By this mechanism (shown in Figure 1.1.6 D) the activated LPMO–Cu(II)–OO$^-$ complex receives two one-electron transfers reducing the superoxide radical such that with the concurrent input of two protons, a molecule of water is released leaving a LPMO–Cu(II)–O$^+$ species at the active site. Hydrogen abstraction from substrate produces LPMO–Cu(II)–OH, and then the hydroxyl radical “rebounds” to the substrate radical forming the new C–O bond. Glycosidic bond elimination proceeds as described above while the active site is regenerated as a LPMO–Cu(I) complex that can immediately bind molecular O$_2$ and initiate another reaction cycle.

As reviewed by Lee and Karlin, the Cu(II)–O$^+$ complex is expected to be more reactive toward hydrogen atom abstraction than Cu(II)–OO$^+$ [125]. Indeed, Kim et al. predicted LPMO–Cu(II)–O$^-$ abstraction to be energetically favored by 16 kcal mol$^{-1}$ compared with abstraction by LPMO–Cu(II)–OO$^-$ [124]. The AA9 active site models (with axial coordinating tyrosine) used for these DFT calculations were constructed with the respective activated oxygen species occupying the available axial coordination site as modeled in the structure of NcLPMO9D reported by Li et al. [67]. However, newer experimental results and quantum mechanics/molecular mechanics (QM/MM) calculations along with expected Cu–O$_x$ complex stabilities better agree with an equatorial coordination for the activated oxygen species, and it is unclear what effect the difference in coordination would have on the calculated energy barriers for the substrate H abstraction step [80, 84, 126]. Despite this possible complication for the calculations by Kim et al. the large homolytic bond dissociation energy expected for a glycosidic carbon C–H bond (estimated by DFT calculation to be ~ 95
kcal mol$^{-1}$) and possible additional C–H bond stabilization by the crystalline environment of a cellulose surface suggest that the greater oxidative power of LPMO–Cu(II)–O$^-$ may be necessary [125, 127]. A Cu(II)–O$^-$ species has not yet been observed structurally or spectroscopically in a biological system and would be intriguing if characterized. A high-oxidation LPMO–Cu(III)–OH species shown in parentheses in 1.1.6 D, which is also currently unknown to biology, has been proposed as a potential hydrogen abstraction agent based on the high reactivity of model Cu(III)–OH complexes [128, 129]. Going beyond the question of the identity of the hydrogen abstracting species, the order of addition of oxygen and substrate has the potential to influence the active site copper–oxygen chemistry, and the sequence and sources of the needed protons may involve residues adjacent to the active site. All of these questions are subjects of ongoing studies designed to elucidate the LPMO mechanism [130].

1.1.8 Neurospora crassa LPMO9D as a model for mechanistic studies

The experiments described herein have determined new structural details of the protein LPMO9D from the filamentous fungus Neurospora crassa (UniProt Q8WZQ2). (Due to the various nomenclature systems that have been used for LPMOs, this protein is known as NcLPMO9D, NcAA9D or NcPMO-2 both in the literature and in the following chapters.) The N. crassa genome contains fourteen LPMO genes, and nine of these genes are upregulated during growth on crystalline cellulose (Avicel) as sole carbon source [131, 132]. The NcLPMO9D gene, gh61-4, is the most differentially upregulated LPMO gene during Avicel culture; however, NcLPMO9D presumably accounts for < 1% (wt./wt.) of the total cellulase secretome as it was not identified in quantitative proteomics studies of Avicel-grown N. crassa which had sensitivity to this level [133]. NcLPMO9D is translated as a 238 AA product with the first 15 AAs forming a secretion signal peptide that targets NcLPMO9D to the endoplasmic reticulum for trafficking. Cleavage of the signal peptide yields a matured 223 AA single-domain protein with the N-terminal histidine residue that is requisite for formation of the histidine-brace Cu binding site. The sequence contains a single Asn-X-Ser sequence for
potential N-linked glycosylation, and X-ray crystal structures of NcLPMO9D, discussed below, confirm glycosylation at this site.

Using high performance anion-exchange chromatography coupled to electrospray mass spectrometry to analyze LPMO reaction products, Phillips et al. designated NcLPMO9D as having type 2 regiospecificity for oxidation at C4 of cellulose glycosidic bonds [64]. The same study established that NcLPMO9D requires a bound Cu ion for activity. To date, the NcLPMO9D Cu binding affinity has not been quantified; however, a \( K_D \leq 10 \text{ nM} \) is expected given the high Cu(II/I) binding affinities exhibited by other AA9 LPMOs (see 1.1.3). Similarly, kinetic parameters for NcLPMO9D activity have not been determined, and the requirement of an insoluble substrate complicates kinetic analysis. The two examples of AA9 LPMOs known to bind soluble cellosaccharides, NcLPMO9C and LsAA9A, have recorded \( k_{cat} \) values of 0.03-0.11 s\(^{-1}\) [85, 86]. This range may be considered a reasonable upper limit for the apparent rate of NcLPMO9D reactivity on crystalline cellulose since protein-to-substrate adsorption and desorption likely occur for each turnover.

Li et al. determined the first X-ray crystal structure of NcLPMO9D purified from the secretome of cellulose-grown \( N. \) crassa [67]. This structure, refined against 1.10 Å diffraction, exhibits the \( \beta \)-sandwich fold conserved among LPMOs along with an extended planar surface containing Tyr25 and Tyr204. Li et al. proposed that these tyrosine residues mediate interaction with substrate. In addition, Asn60 which lies within the predicted N-linked glycosylation sequence is in fact glycosylated with two \( N \)-acetylglucosamine residues included in the model. The structure of NcLPMO9D determined using homologously-expressed protein shows \( N \)-methylation at His1N\(_{\varepsilon} \), as shown in Figure 1.1.7 [67]. This presumably post-translational modification is common among structures of AA9 LPMOs expressed homologously or in fungal heterologous hosts but is not observed in protein expressed from yeasts or bacteria (cf [62], [67], [84], [85] and [107]).

The NcLPMO9D structure refined by Li et al. against \( d_{\text{min}} = 1.10 \) Å X-ray diffraction data included a dioxygen species present at 75% occupancy and apparently coordinated end-on to the available axial copper coordination site despite a Cu–O1 bond length of \(~2.94 \) Å [67]. (NcLPMO crystallized with two molecules per asymmetric unit, and the Cu–O1 bond lengths
varied between 2.92–2.96 Å for the two unique active sites.) The identity of the species was supported by unrestrained refinement of the O1–O2 bond length to 1.16 Å and independent atomic occupancy refinements that asymmetrically localized more electron density on the copper-proximal oxygen atom. This coordinated superoxide species was immediately discussed as a reactive intermediate in the fungal LPMO reaction mechanism and likely motivated Kim et al. to consider only axially-coordinated reactive oxygen species in their later DFT study of the LPMO mechanism [124, 134].

This presence and identity of this *axial* superoxide species is well supported by the diffraction data. The superoxide ligands in both chains also exhibited low isotropic atomic displacement parameters (< 22 Å²) and high model–density real space correlation coefficients (> 98%). However, the Cu–O1 distance exceeds coordination bond lengths for copper complexes even when accounting for the Jahn-Teller or tetragonal distortion observed in metalloprotein Type 2 copper centers [81, 82]. Oxygen binding at the available *equatorial* copper coordination site is consistent with both the EXAFS spectra presented by Kjaergaard

![Figure 1.1.7: NcLPMO9D active site with superoxide in axial copper coordination.](image)

NCS molecule A of the structure (PDB 4EIR) by Li et al. is shown [67]. The O1–O2 and Cu–O1 distances are 1.16 Å and 2.96 Å, respectively.
et al. and an associative–displacement mechanism for peroxide release after displacement of equatorially bound hydroperoxyl by an axially coordinated water molecule [80]. In addition, the QM/MM study by Hedegård and Ulf, discussed in 1.1.7, predicts axial oxygen activation to be less thermodynamically favorable than equatorial oxygen activation [126]. Given these inconsistencies with experimental data, computational results and expected metalloprotein chemistry, Marletta and coworkers have acknowledged that axially-bound superoxide may not have mechanistic significance for fungal LPMOs and that the electron density present for this species in their structure of NcLPMO9D may be an experimental artifact [135, 136]. Nonetheless, the potential to observe an activated LPMO copper–oxygen complex in protein crystal structures along with the high-resolution diffraction that can be obtained from crystals of NcLPMO9D strongly motivated the choice of this LPMO for structural studies designed to reveal mechanistic information about the fungal LPMO reaction mechanism.

1.1.9 Overview of Research

The research described herein was conducted to elucidate new structural details of oxygen activation at the active site of a fungal LPMO. Oxygen activation, introduced in 1.1.7, represents one of two central processes vital to LPMO reactivity that to date remains incompletely characterized. (Electron transfer to the LPMO active site is the second process.) The site of oxygen activation within the active site copper ion coordination sphere, the chemical identity of the copper–oxygen species responsible for hydrogen abstraction from substrate and the potential for involvement of non-active site residues in oxygen binding or activation are important questions. Complete understanding of the LPMO reaction mechanism clearly requires these details. In addition, the LPMO reaction is already inspiring attempts to synthesize small-molecule complexes that reproduce LPMO reactivity, and such efforts could undoubtedly benefit from a completed view of the LPMO mechanism. Furthermore, as discussed in 1.1.5, realizing the benefits of LPMO activity in industrial processes imposes new process design considerations to satisfy the oxygen requirements of LPMOs. This work addresses the vital process of fungal LPMO oxygen activation by reporting and discussing
crystal structures of *Neurospora crassa* LPMO9D determined by X-ray and neutron protein crystallography under varied conditions.

As neutron protein crystallography has been a choice technique for this research, the review article “Neutron protein crystallography: A complementary tool for locating hydrogens in proteins” is presented as 1.2 to provide background information on the technique and examples of its successful application to various questions of protein structure–function relationships requiring direct identification of hydrogen atoms [137]. **Chapter 2** is comprised of the article “Crystallization of a fungal lytic polysaccharide monooxygenase expressed from glycoengineered *Pichia pastoris* for X-ray and neutron diffraction” [138]. This article describes the work that was necessary to establish a heterologously-expressed form of NcLPMO9D as a robust platform for crystallographic studies and the chemical and structural bases of certain complications encountered during these experiments. **Chapter 3** is presented in two parts: 3.1 is comprised of the article “Oxygen Activation at the Active Site of a Fungal Lytic Polysaccharide Monooxygenase” which describes novel structures of NcLPMO9D that revealed previously unseen interactions between protein and oxygen both at and adjacent to the active site [84]. In addition, the first reported neutron protein crystal structure of a LPMO is presented with details discussed in the context of oxygen binding and activation. The article “Structural studies of *Neurospora crassa* LPMO9D and redox partner CDHIIA using neutron crystallography and small-angle scattering” discusses additional details of the NcLPMO9D neutron protein crystal structure and is presented as 3.2 [139]. **Chapter 4** is a brief prospectus placing the findings presented herein in context with current understanding of LPMOs and discussing what studies conceptually follow for further refinement of the LPMO reaction mechanism.
1.1.10 REFERENCES


1.2 Neutron protein crystallography: A complementary tool for locating hydrogens in proteins

The following work was reprinted with permission from: W.B. O’Dell, A.M. Bodenheimer and F. Meilleur. *Archives of Biochemistry and Biophysics*. **2016**. 602:48–60

1.2.1 Highlights

- Hydrogen atoms are central to enzyme chemistry.
- Neutron protein crystallography (NPC) locates hydrogen atoms in protein structures.
- NPC is a powerful tool for investigating enzyme chemistry.

1.2.2 Abstract

Neutron protein crystallography is a powerful tool for investigating protein chemistry because it directly locates hydrogen atom positions in a protein structure. The visibility of hydrogen and deuterium atoms arises from the strong interaction of neutrons with the nuclei of these isotopes. Positions can be unambiguously assigned from diffraction at resolutions typical of protein crystals. Neutrons have the additional benefit to structural biology of not inducing radiation damage in protein crystals. The same crystal could be measured multiple times for parametric studies. Here, we review the basic principles of neutron protein crystallography. The information that can be gained from a neutron structure is presented in balance with practical considerations. Methods to produce isotopically-substituted proteins and to grow large crystals are provided in the context of neutron structures reported in the literature. Available instruments for data collection and software for data processing and structure refinement are described along with technique-specific strategies including joint X-ray/neutron structure refinement. Examples are given to illustrate, ultimately, the unique scientific value of the neutron protein crystal structure.
1.2.3 Introduction

Proteins, and more specifically enzymes, are the lifeblood of cellular processes. Enzymes catalyze a broad array of chemical reactions essential to life. No less than 20,000 distinctive enzyme-catalyzed reactions are likely across all living organisms, controlling processes as diverse as the capture and conversion of energy from light, cell signaling, DNA translation and repair and metabolic pathways [1]. Most of these catalytic functions exhibit an exquisite chemical specificity, for both substrate and product, which is encoded in the precise three-dimensional organization of amino-acids at and leading to the active site of the enzyme. Enzyme activity is tightly orchestrated and controlled by a variety of post-translational modifications (phosphorylation, glycosylation, ubiquitination, acetylation, lipidation, etc.) and allosteric regulation mechanisms. Local environment, pH and chemical flux also trigger significant structural reorganization contributing to activity modulation [2-5].

Hydrogen atoms are central to enzyme chemistry, as ultimately, reaction rates and chemistry are dependent upon the coordinated changes in local electrostatics, hydrogen-bonding interactions and protonation states of catalytic residues along the reaction coordinate. Therefore, understanding enzyme chemistry at the atomic level requires the visualization of hydrogen atoms on active site and remote residues, co-factors, substrate and water molecules. This information can be challenging to obtain. While most structure-based knowledge arises from X-ray crystallography, hydrogen atoms are exceedingly difficult to visualize in X-ray structures. When data are available to ultra-high resolution, few hydrogen atoms may be visible or their positions may only be inferred from precise geometric parameter analyses [6]. The combination of atomic resolution X-ray crystallographic data with quantum chemistry or charge density analysis can then provide a further level of detail on the chemical profile of the enzyme [7-9]. However, even when such ultra-high resolution data can be obtained, a significant fraction (typically > 50%) of those more mobile or labile hydrogen atoms remain difficult to discern, leaving specific questions concerning catalytic mechanism unanswered.
Figure 1.2.1: Incoherent neutron scattering cross sections and coherent neutron scattering lengths for selected elements. Relative incoherent scattering cross sections are represented by the left hemispheres, and relative coherent scattering lengths are represented by the right hemispheres. The red hemisphere for hydrogen indicates the negative sign of its scattering length while the others shown in green are positive. Incoherent cross sections and coherent scattering lengths are not represented on the same scale.

The difficulty of locating hydrogen atoms using X-ray crystallography can be circumvented by neutron protein crystallography. This is because the coherent scattering lengths of hydrogen (H) and the hydrogen isotope deuterium (D) for neutrons are similar in magnitude to those of carbon, nitrogen and oxygen (Figure 1.2.1) [10]. A complication of neutron protein crystallography is that the scattering length of hydrogen is negative while that of carbon (C), nitrogen (N) and oxygen (O) are positive, which gives rise to density cancelation in Fourier maps that can hamper interpretation and analysis. In contrast, D has a neutron scattering length of the same sign as the heavier atoms and thus gives a clear positive peak in the nuclear density maps. While visibility of hydrogen atoms requires neutron crystallographic data at resolution of 2.0 Å or better, deuterium atoms are readily visible, in crystallographic structures identical to their hydrogenated counterparts, at typical resolutions of 2.5 Å or better [11-15].

Despite a number of significant technical challenges that we will discuss in this review, the ability to experimentally locate deuterium makes neutron diffraction a method of choice for visualizing the positions of hydrogen in enzymes, whether used as a standalone technique or in complementary studies along with nuclear magnetic resonance (NMR) spectroscopy or high-resolution X-ray crystallography. Neutron protein crystallography has, for example, been used to locate hydrogen atoms and determine the protonation states at the active site of HIV-1 protease (HIV Pr), beta-lactamase, dihydrofolate reductase (DHFR), carbonic anhydrase and...
xylose isomerase. [16-33]. These studies have been essential in identifying active site residues with perturbed pK_a’s and contribute to the growing body of knowledge on pK_a modulation at the active site of enzymes.

In this review, we focus on practical considerations to help guide the successful development of neutron protein crystallography projects by newcomers. (The interested reader is encouraged to find more historical perspectives on NPC in the works by Niimura and Podjarny (2011) and Blakeley (2009) [34, 35].) A flow chart of the steps involved in neutron crystallographic structure determination is presented in Figure 1.2.2. We specifically review H/D exchange protocols, crystal growth methods, data collection strategies and structure refinement options that have been used for the 85 neutron protein structures currently deposited in the Protein Data Bank (October 1, 2015). We also give a perspective on the future of NPC illustrated with recently determined neutron protein structures.
Figure 1.2.2: Neutron structure determination flow chart. The flow chart shows the steps in neutron protein crystallography. Most steps are similar to those of x-ray protein crystallography. D labeling or H/D exchange and growth of large crystals are specific steps for the neutron experiment.
1.2.4 Hydrogen/Deuterium isotopic substitution

As stated previously, replacing H for D in a protein crystal provides significant benefits to the neutron protein crystallography experiment. The greater magnitude of the D coherent scattering length (6.671 fm) compared to H (-3.741 fm) results in a larger $\sum |F|^2$ and, therefore, greater average reflection intensities. Simultaneously, the six-fold decrease in the D incoherent scattering length (3.99 fm) compared to H (25.27 fm) dramatically reduces the isotropic background intensity. The gain in signal and reduction in noise allow datasets to be collected from smaller crystals (< 0.5 mm$^3$) or with reduced total exposure times (< 10 days) [13, 24, 36-38]. Furthermore, the D coherent scattering length is positive as is that of common atoms found in proteins while the coherent scattering length of H is negative (Figure 1.2.1). At the typical resolutions (1.9–2.3 Å) obtained for neutron protein crystallographic data, D substitution simplifies the analysis of neutron density maps by minimizing the number of sites where the negative coherent scattering length of H atoms leads to density cancellation. Figure 1.2.3 demonstrates the greater extent of significant neutron density in the 2F_o–F_c map from a fully D-labeled, or perdeuterated, crystal of beta-lactamase as compared with the map from an H/D-exchanged crystal. Both datasets were measured on the LADI III diffractometer at Institut Laue–Langevin and extend to 2.0 Å resolution [39]. However, data for perdeuterated beta lactamase were collected from a smaller crystal and in half the measurement time than was required for the H/D-exchanged protein [22, 23]. In the case of aldose reductase, perdeuteration produced an improvement in diffraction resolution from 4.5–2.0 Å that enabled neutron data collection and structure determination [13, 40, 41].
Figure 1.2.3: Neutron density maps from H/D-exchanged and perdeuterated beta-lactamase. Top: the 2F₀–Fc map from the H/D-exchanged crystal (PDB 2WYX) is shown for residue Arg61. Bottom: the corresponding 2F₀–Fc map from the perdeuterated crystal (PDB 2XQZ) is shown. Both maps are contoured at \( \sigma = 1.5 \). Structures were drawn using CCP4MG with coordinates taken from the PDB [42].

1.2.4.1 H/D exchange

Bulk solvent, ordered water molecules and hydrogen atoms bound to protein oxygen, nitrogen or acidic carbon atoms such as histidine C₇₁ (i.e.: titratable hydrogen atoms) can all be replaced with deuterium by preparing all reagents in deuterium oxide (D₂O) either during or after complete crystal growth [43]. Of the neutron protein crystal structures currently deposited in the PDB (Appendix A, Supplementary Material Table S1 and www.rcsb.org), approximately half were determined from proteins with titratable hydrogen atoms exchanged during crystallization. In this approach, the protein sample is exchanged by dialysis or by repeated concentration and dilution into a D₂O-prepared buffer solution [44]. Similarly, crystallization salts and/or precipitants are dissolved in D₂O thus creating an H/D-exchanged mother liquor. The crystallization experiment is then setup in the same way as that performed with H₂O solutions. Re-optimization of crystallization conditions may be necessary for D₂O
solutions due to the reduced solubility of proteins in D$_2$O and the differences in physical properties of H$_2$O and D$_2$O (Table 1.2.1) [45-47]. The reduced solubility can be addressed by decreasing the crystallization agent concentration(s), reducing the protein concentration, increasing the crystallization temperature or by varying these parameters simultaneously (see 1.2.5 Crystal Growth) [11, 14, 22, 23, 48-51].

Table 1.2.1: H$_2$O and D$_2$O physico-chemical properties.

<table>
<thead>
<tr>
<th>Property</th>
<th>H$_2$O</th>
<th>D$_2$O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond energy, (kJ mol bond$^{-1}$, 0 K)</td>
<td>458.9</td>
<td>466.4</td>
</tr>
<tr>
<td>Boiling point (° C)</td>
<td>100.0</td>
<td>101.42</td>
</tr>
<tr>
<td>Density (kg m$^{-3}$, 25° C)</td>
<td>997.05</td>
<td>1104.36</td>
</tr>
<tr>
<td>Dynamic viscosity (mPa s, 25 ° C)</td>
<td>0.8909</td>
<td>1.095</td>
</tr>
<tr>
<td>Melting point (° C)</td>
<td>0.00</td>
<td>3.82</td>
</tr>
<tr>
<td>Molar mass (g mol$^{-1}$)</td>
<td>18.015265</td>
<td>20.027508</td>
</tr>
</tbody>
</table>

For the remaining structures deposited in the PDB crystals were grown to final size in H$_2$O mother liquor and then subjected to H/D exchange either by vapor equilibration in capillary (Figure 1.2.4) or soaking in an artificial H/D-exchanged mother liquor.

Figure 1.2.4: H/D exchange of a fully grown crystal in the capillary mount by vapor diffusion. (1) Crystals are drawn into the capillary from the crystallization setup. (2) Excess mother liquor is removed from the crystal (3) exposing the crystal surface for exchange. (4) Plugs of artificial mother liquor prepared in D$_2$O are added on each side of the crystal to drive the exchange in the closed environment provided by (5) the sealed capillary.
Exchange rates of individual hydrogen atoms in proteins in solution vary greatly depending upon factors including hydrogen bonding, solvent exclusion by protein folding, solvent pH and incubation temperature [52-55]. The crystal lattice further limits diffusion of water to the network of pores formed by the packing of protein molecules [45]. These factors make the choice of H/D equilibration time arbitrary. In practice crystals are exchanged from the time the crystal is fully grown until data collection. Reported times for H/D exchange after crystallization range from seven days for diisopropyl fluorophosphatase to ten years for a crystal of myoglobin [56-58]. By analyzing ten neutron structures deposited in the PDB, Bennett et al. concluded that H/D exchange by crystal soaking is likely to reach equilibrium after 30 days [59].

Two structures of HIV Pr allow for an interesting comparison of soaking and vapor diffusion exchange methods. Adachi et al. determined a structure of HIV Pr (PDB 2ZYE) from a crystal that was soaked for 14 days in artificial D2O mother liquor with pD 5.0 (pD = pH + 0.4) at 20 °C and observed approximately 54% exchange of backbone amide hydrogens [25]. Weber et al. also determined a structure of HIV Pr (PDB 4JEC) in which approximately 71% of the backbone amide hydrogens exchanged after vapor equilibration with D2O solution with pD 5.6 at 17 °C for 28 days [26]. Remarkably, as shown in Figure 1.2.5, the patterns of backbone amide exchange in the two structures are quite distinct despite very close agreement in atomic positions (< 0.5 Å root mean square deviation). Multiple experimental distinctions may have contributed to these differences. The crystals were grown under significantly different crystallization conditions. The structures were also obtained with different small molecule inhibitors bound at the HIV Pr active site, and it is known that different modes of ligand binding can influence backbone amide exchange patterns [60]. In addition, the 2ZYE model was refined using phenix.refine while the 4JEC model was refined using CNSsolve [61, 62].
Figure 1.2.5: Backbone amide H/D exchange observed in two neutron structures of HIV-1 protease. A: A ribbon diagram of the structure obtained after crystal soaking for 14 days is shown (PDB 2ZYE). B: The corresponding diagram for the structure obtained after vapor diffusion H/D exchange for 28 days is shown (PDB 4JEC). In both structures, residues are colored by the percent of D substitution observed (color bar) in the refined model. Structures were drawn using UCSD Chimera with coordinates from the PDB [63].

1.2.4.2 Perdeuteration

Titratable hydrogen atoms represent approximately 25% of the total hydrogen atoms of any proteins. In order to exchange the remaining 75% attached to carbon atoms, proteins need to be perdeuterated during synthesis. Expression of perdeuterated protein in fully D-labeled growth media has yielded neutron protein crystal structures currently deposited in the PDB for six proteins, namely myoglobin, aldose reductase, transthyretin, type III antifreeze protein, beta lactamase and rubredoxin [3, 13, 23, 24, 36-38, 41, 64-69].

Each perdeuterated protein was expressed recombinantly in the host Escherichia coli. Cultures must be progressively adapted to grow in minimal media formulated with D$_2$O, H/D-exchanged salts and uniformly D-labeled carbon source prior to induction of expression either in shake flasks or in a bioreactor system. Methods for perdeuterated protein expression specifically for neutron protein crystallography have been described in detail by Meilleur et al. for cytochrome P40cam and by Golden et al. for cholesterol oxidase [44, 48, 70]. These studies describe two alternate approaches to growing of E. coli in fully deuterated medium in a bioreactor: the former used glycerol-limited fed-batch cultivation while the latter culture was grown until complete depletion of the culture broth’s initial carbon source content. Labeled proteins are generally isolated and purified using H$_2$O-formulated buffers due to the cost of
D$_2$O and D-labeled reagents. The protein is then either exchanged back to D$_2$O buffer for crystallization or crystallized in H$_2$O conditions and H/D-exchanged as discussed above.

One structure of HIV Pr has been determined from 85% uniform D-labeled protein since it was observed that fully deuterated HIV Pr would not form crystals with sufficient volume for neutron diffraction [26]. Uniformly D-labeled protein can be expressed with $\leq$ 95% D incorporation from *E. coli*, other prokaryotes, or even eukaryotic hosts using techniques largely developed for NMR experiments [71-74]. However, except for the case of HIV Pr, uniform partial deuterium labeling has not been applied in the field of neutron protein crystallography. Partial D-labeling does reduce incoherent background in the diffraction experiment, but it requires refinement of the H/D-occupancies for all sites in the model structure which considerably increases the number of parameters to refine (see 1.2.7 Structure Refinement) [26]. In addition, sites with H and D occupancies of approximately 60% and 40% respectively would also have null or insignificant density in Fourier maps due to the negative H and positive D coherent scattering lengths further complicating the structural analysis. More neutron diffraction experiments using crystals of partial uniform D-labeled protein still need to be conducted to fairly assess this approach.

Selective H-labeling in a perdeuterated background has been applied in neutron diffraction studies of rubredoxin (PDB 3KYY) and type III antifreeze protein (PDB 4NY6) with the ultimate goal of using neutron data to solve protein structures by direct methods [4, 64, 65, 75, 76]. Media supplementation with unlabeled α-ketoisovalerate during perdeuterated protein expression resulted in H-labeling of valine and leucine methyl groups producing (H-γ methyl)-valine and (H-δ methyl)-leucine uniformly throughout the proteins [77]. Fisher *et al.* compared neutron diffraction datasets for perdeuterated and CH$_3$-H-labelled antifreeze protein to demonstrate the negative effects of protein H atoms on the observability of structured water molecules in neutron density maps [4]. Theoretically specific H-labeling in a perdeuterated background can enable direct determination of model phases for structure solution; however, this application of neutron protein crystallography with selective labeling has yet to be demonstrated [75].
1.2.5 Crystal growth

The diffracted intensity in Bragg reflections in a single crystal experiment can be written as:

$$\langle I \rangle \propto I_0 |F|^2 V \lambda^2 \frac{1}{v_0^2} \quad (1.2.1)$$

where $\langle I \rangle$, $I_0$, $|F|$, $V$, $\lambda$, and $v_0$ are the diffracted intensity, incident beam intensity, structure factor magnitude, volume of the crystal, incident wavelength and volume of the unit cell, respectively [78]. Neutron beam fluxes are inherently many orders of magnitude weaker than synchrotron X-ray beams [79]. Therefore, neutron crystallography requires crystals that are at least three orders of magnitude larger than crystals suitable for X-ray crystallography because, as shown in Equation (1.2.1), the diffracted intensity is directly proportional to the incident beam intensity.

The neutron crystallography experiment typically follows X-ray crystallography experiments that, despite providing high-resolution diffraction data, fell short of unambiguously answering mechanistic questions due to the lack of information on hydrogen atom positions. All neutron crystallography projects have the starting advantage of known crystallization conditions, but these conditions most likely will need to be further optimized to sustain the growth of large crystals suitable for neutron data collection. Crystals typically need to be greater than 0.1 mm$^3$ in volume (Appendix A, Table S1 and references herein) and must be well ordered to diffract to sufficient resolution [80]. It is important to note that while the beam flux is the dominant requirement for large crystals, the volume of the unit cell is also a significant parameter. Equation (1.2.1) shows that the diffracted intensity is inversely proportional to $v_0^2$. In cases where several crystal forms can be grown, crystallization conditions promoting smaller unit cells should be favored.
Growing large crystals requires that the crystallization setup contains sufficient amount of protein. Equation (1.2.2) provides an estimation of the minimal volume of protein \( V_{\text{min}} \) solution that should be used:

\[
V_{\text{min}} \geq \frac{V_{\text{crystal}} \times N_{\text{asymmetric unit}} \times N_{\text{protein chain}} \times MW}{V_{\text{unit cell}} \times N_A \times C}
\]  

(1.2.2)

where \( V_{\text{crystal}} \) is the target volume of the fully grown crystal, \( N_{\text{asymmetric unit}} \) is the number of asymmetric units for the expected space group, \( N_{\text{protein chain}} \) is the number of protein chains in the asymmetric unit, \( MW \) is the molecular weight of the protein, \( V_{\text{unit cell}} \) is the volume of the unit cell, \( N_A \) is the Avogadro number and \( C \) is the protein concentration (in mg mL\(^{-1}\)). In using this equation to plan crystallization experiments, one assumes that only one crystal will grow per drop and that all of the protein in solution crystallizes. The estimation also relies on accurate protein concentration determination.

Readers are referred to the recent reviews by McPherson and Gavira (2014) and Gavira (2016) for detailed protein crystallization principles [81, 82]. The most commonly utilized crystallization method for neutron crystallography is vapor diffusion, with 35 out of the 85 neutron protein structures deposited in the PDB solved from crystals grown by vapor diffusion. Large 10–25 \( \mu \)L hanging drops were used only for aldose reductase and diisopropyl fluorophosphatase [13, 40, 41, 56, 57]. Sitting drop crystallization allows for larger volume drops to be setup easily compared to hanging drops. Sitting drop volumes reported range from 20–100s \( \mu \)L. Sandwich box setups (Hampton Research), consisting of a siliconized nine-well glass plate placed inside a sealed plastic box, are typically used to set up larger (> 50 \( \mu \)L) volume sitting drops. According to the manufacturer, each well is capable of holding drop sizes up to 800 \( \mu \)L, and the box, which serves as precipitant reservoir, can hold up to 25 mL of crystallization solution. The largest reported sitting drops were setup with a total volume of 400 \( \mu \)L [26, 83]. Neutron structures solved from crystals grown using the sitting drop vapor diffusion method include carbonic anhydrase II, DHFR, and transthyretin [38, 83, 84]. Despite the fairly high success rate of the sitting drop vapor diffusion techniques, it should be noted
that in about two thirds of the cases, the sitting drops are micro- or macroseeded as described below (also see Appendix A, Table S1).

Batch crystallization is a useful alternative method with about one third of the deposited structures reporting this technique (Appendix A, Table S1). Batch crystallization volumes for neutron crystallography span from 10s of mL in the early days of protein crystallization to volumes more compatible with modern protein production of 100s of μL and even down to 10 μL in a combined batch/seeding crystallization approach such as that used by Matsumura et al. to crystallize the HIV Pr–KNI-272 complex [24, 85-88]. The batch method is usually applied to proteins available in large quantities such as xylose isomerase [33, 46]. However, large batch crystallizations are not a requirement during optimization for this approach to be successful [89, 90]. In fact, the batch method presents the distinct advantage over other crystallization techniques that scaling to larger volumes does not affect the kinetic and thermodynamic factors that regulate the crystallization process. Crystallization conditions optimized using small batch volumes or even by high-throughput micro-batch screening can readily be scaled up.

Dialysis and counter diffusion have been used to a lesser extent for neutron protein crystallography. Dialysis with protein solution volumes of 200 μL and above has been performed for concanavalin A using dialysis tubing, insulin using a micro-dialysis chamber and P450cam using dialysis buttons (Hampton Research) [31, 91-94]. Smaller volumes (50–100 μL) were dialyzed using dialysis buttons for cytochrome c peroxidase and insulin.[95-98]. A customized counter diffusion apparatus was successfully used to grow large inorganic pyrophosphatase crystals suitable for neutron data collection [99, 100]. Dialysis, batch and counter diffusion setups can be supplemented with micro- and macroseeding and “crystal feeding”. In crystal feeding fresh protein solution is added to a crystallization drop containing a crystal that appears to have stopped growing. The crystal is typically “fed” several times. This method was applied to Achromobacter protease I [101]. Special attention should be given to the crystal habit response to the feeding process as the etching of the crystal surface that may occur can lead to further nucleation or growth of a multi-lattice crystal.
Seeding techniques are powerful as they offer the opportunity to decouple nucleation from growth. Decoupling these two steps can be instrumental in growing large single crystals. Just under a third of the crystals used to collect neutron data set deposited in the PDB used micro- or macroseeding illustrating the usefulness of this approach (Appendix A, Table S1). Seeding is used for both X-ray and neutron crystallography when initial drops do not produce crystals with sufficient volume or diffraction quality. For example, micro-seeding was necessary to grow a large single crystal of aldose reductase [13]. Macroseeding was successfully applied to the thrombin-bivalirudin complex, cholesterol oxidase, HIV Pr, rubredoxin, elastase and the recently determined structure of farnesyl pyrophosphate synthase (Appendix A, Table S1) [25, 48, 102-106]. In the last case, spontaneous nucleation and crystal growth was conducted in large 400 μL sitting drops that produced crystals with volumes up to 0.1 mm$^3$. These crystals were then used to macroseed 60 μL drops in which one crystal reached a final volume of 3.5 mm$^3$. Interestingly, Bennett et al. grew large crystals of DHFR–methotrexate by vapor diffusion using large sitting drop while Wan et al. combined both micro- and macroseeding with vapor diffusion to grow DHFR–folate–NADP (+) crystals suitable for neutron diffraction [21, 107].

While protein concentration, precipitant type and concentration and pH are the most varied crystallization parameters during optimization, temperature can also be an efficient parameter to vary [90]. By changing the crystal growth temperature, one can drive the protein solution from a nucleating state to a metastable state or maintain the solution in the metastable state to promote continued crystal growth. This can be performed using sophisticated equipment such as the temperature-controlled batch crystallization device developed and applied to urate oxidase by Budayova-Spano et al. [51, 108]. On the other hand, simply transferring crystals trays between incubators with different temperature set points or changing the incubator temperature by steps or continuous gradients provides straightforward options to vary temperature, though in a less controlled fashion [25, 88].

As discussed above, titratable hydrogen atoms present in the crystal should be exchanged to deuterium prior to data collection. Soaking, vapor diffusion or crystallization in solution prepared with D$_2$O can be used for H/D exchange. Crystallization in D$_2$O solutions
is a straightforward way to achieve H/D exchange, but this approach often does not produce crystals suitable for neutron diffraction or requires re-optimization of conditions. Soaking is the most efficient method for H/D exchange after crystal growth. However, crystals grown in H$_2$O solution may suffer damage (e.g.: cracks) from soaking in a fully deuterated solution. Gradual increases in D$_2$O concentration of the soak solution may help prevent such damage. A stepwise increase in D$_2$O concentration was successfully applied to DHFR crystals [107].

A gentler approach to H/D exchange is vapor diffusion. Vapor diffusion H/D exchange can be performed by replacing the hydrogenated mother liquor in the reservoir with an artificial deuterated mother liquor [109]. The well-to-drop ratio must be at least 10, and the D$_2$O solution should be replaced 3 times or more to maximize the exchange. If room temperature data collection is planned, H/D vapor exchange can occur inside the capillary in which the crystal is mounted as shown in Figure 1.2.4. The largest possible volume of deuterated solution “plug” must be used for exchange in capillaries, and this plug must also be repeatedly changed for new solution. For both the soak and vapor diffusion, longer exchange times permit exchange in buried area of the protein further and result in greater deuterium incorporation.

There is no systematic use of one exchange approach over another; the method used depends largely on the experimenter. However, data collection may restrict the choices. If data collection at cryogenic temperature is planned, soaking in or vapor diffusion with a deuterated reservoir will be the safest approaches. For room temperature data collection, the crystals should be mounted and exchanged in a quartz capillary since borosilicate glass absorbs neutrons (Figure 1.2.4). Uses of thin and thick wall capillaries are both reported in the literature. While thick wall capillaries generate higher background, they are sturdier. The choice of the type of capillary is again an experimenter’s preference.

### 1.2.6 Instrumentation, data collection and processing

Neutron crystallography is an inherently flux-limited technique. Available neutron fluxes are several orders of magnitude less than the X-ray flux available from modern in-house generators and synchrotrons. This limitation is the primary impetus for increasing the
scattering power of the protein crystal through the isotopic labeling and crystal growth techniques discussed in the previous sections. Low available flux also results in the long exposure times required to measure a neutron diffraction dataset to sufficient resolution and completeness. While complete X-ray datasets can be measured in seconds or less at high-intensity X-ray sources, neutron datasets typically require days-to-weeks for collection. However, through recent advances in instrumentation and labeling it has become possible to obtain useful neutron datasets in a single day or less in special cases [36].

Table 1.2.2 lists current single crystal diffractometers designed for neutron protein crystallography at neutron scattering facilities worldwide. In addition to these instruments, other multi-purpose single crystal diffractometers, such as the D19 instrument at ILL, have also been used for protein crystallography [5, 28, 37, 38]. Instruments can be divided into three groups based upon mode of operation, namely monochromatic, quasi-Laue (also known as “pink-Laue”) and time-of-flight (TOF) Laue modes. All instrument detector arrays are designed to maximize coverage of reciprocal space so that a large number of stimulated reflections that can be recorded simultaneously whether the instrument uses mono- or polychromatic incident radiation. Typically, neutron protein crystallography beam lines take advantage of longer wavelength neutrons to maximize the intensities of the diffraction peaks which are proportional to the square of the incident wavelength, as shown in Equation (1.2.1), and to reduce spatial overlap according to Bragg’s law \(2d \times \sin(\theta) = \lambda\). At the wavelength typically used for X-ray crystallography \(\lambda \leq 1.54 \, \text{Å}\), the diffracted beams lie on cones in the forward scattering direction, and two-dimensional detectors are installed downstream of the crystal. At the longer wavelengths \(2.3 \, \text{Å} < \lambda < 6.0 \, \text{Å}\) used by neutron diffractometers, high resolution diffraction also occurs in the backscattering direction. Instrument detector arrays approximate cylindrical or spherical geometries to record both forward- and backscattering reflections.

Existing monochromatic and quasi-Laue instruments are installed at reactor neutron sources and take advantage of the high time-integrated neutron flux provided by these facilities. Monochromatic instruments operate analogously to standard X-ray macromolecular diffractometers with the exception of longer incident wavelengths. Quasi-Laue instruments
accept a defined, broad range of incident neutron wavelengths ($\Delta \lambda/\lambda_{\text{mid}} \approx 25-60\%$) that greatly increases the number of stimulated reflections at each crystal setting relative to monochromatic mode. The broad wavelength bandpass, therefore, is a more effective use of neutrons available from the reactor source, and the configuration also avoids the need for a monochromator crystal upstream of the sample that would inevitably have less than 100% reflection efficiency at any selected wavelength [110]. Quasi-Laue represents a compromise on the amount of observed incoherent background with it being more than that observed with monochromatic radiation but less than that observed with truly “white-beam” radiation. While the monochromatic experiment produces data sets of higher quality compared to the quasi-Laue experiment (either for X-rays or neutrons) for neutron protein crystallography the quasi-Laue configuration has the unique advantage of allowing smaller crystals and/or shorter exposure times to be used [111].

Unlike reactor sources, accelerator-driven spallation sources provide neutrons in discrete pulses, and this time structure permits the use of a TOF Laue mode for macromolecular diffractometers. A relatively large wavelength range can be accepted by the instrument to maximize the number of stimulated reflections. The detectors record the positions and time of incidence for each scattered neutron from each pulse. The neutron time of flight can then be converted to the corresponding wavelength, and the observed quasi-Laue diffraction pattern can be “binned” or separated into reflections and background for each incident wavelength. TOF methods, therefore, provide both the increased number of reflections of quasi-Laue mode and the reduced background of monochromatic mode.

Data collection at two or more $\chi$ angles is essential when using longer wavelength radiation because the curvature of the Ewald sphere is such that the blind region is significant at any one $\chi$ setting. In addition, instruments using a cylindrical detector geometry (BIODIFF, BIO-C, BIX-3, BIX-4, IMAGINE, and LADI-III) have large, systematic vacancies in their coverage of reciprocal space along the cylindrical long axis. Reflections absent due to detector coverage can be recovered by using more than one $\chi$ angle.
Sample handling for neutron diffraction experiments is generally different from that for both in-house and synchrotron X-ray experiments due to the non-destructive nature of the neutron radiation. Neutrons having wavelengths of 2–10 Å (< 50 meV) do not induce ionization in biological samples. Therefore, cryogenic sample temperatures are not necessary to protect against radiation damage in the protein crystal, and the majority of experiments are performed at ambient temperature. While room temperature data collection is typical, cryocystallography does, however, reduce thermal disorder in protein crystals and make possible the study of reaction intermediates and temperature sensitive systems [121]. As reviewed by Coates et al. nitrogen cryostreams have been installed on a subset of neutron diffractometers to perform measurements at 100 K with sample mounting equipment standard for X-ray cryocrystallography [122]. For either temperature condition, crystal mounting and alignment are often performed manually; however, the newest instruments such as MaNDi do incorporate motorized sample handing and positioning under user control [118]. Given that data collection time exceeds days, manual centering is not a significant overhead. Upon completion of data collection, particularly at ambient temperature, the crystal is recovered. This same crystal can
then be used for additional neutron experiments or to collect an X-ray dataset for joint refinement of X-ray and neutron data (see 1.2.7 Structure Refinement).

Diffraction data from reactor-based instruments can be indexed and integrated using software packages developed for X-ray crystallography with appropriate modifications to account for the instrument’s detector geometry. Software packages used include Lauegen, d*TREK and DENZO [123-126]. Handling the additional TOF information recorded by spallation-source instruments requires the development of dedicated data reduction packages [127, 128]. Unlike monochromatic data, polychromatic datasets (either for X-rays or neutrons) must be wavelength-normalized to account for the spectral shape of the incident beam. Software used to wavelength normalize data are also adapted from X-ray Laue crystallography software packages [123, 129]. Data scaling is then performed with routine protein crystallography software such as SCALEPACK or SCALA [126, 130].

1.2.7 Structure refinement

Structure refinement is carried out against neutron data alone or simultaneously against both X-ray and neutron data. The latter protocol is referred to as joint refinement or X/N refinement [131]. The choice of one method of refinement over the other may be influenced by the neutron dataset resolution and completeness but is ultimately the experimenter’s choice. The number of parameters to be refined for a neutron structure is considerably greater than for an X-ray structure: the number of atoms is nearly doubled, and ~25% of the hydrogen positions require H/D-occupancy refinement. However, typical neutron datasets contain fewer unique reflections than their X-ray counterparts because of lower completeness and resolution. Refining a model against both X-ray and neutron data significantly increases the data-to-parameter ratio and reduces the risk of “over-fitting” the available data during refinement [132]. Joint refinement is widely used for neutron structure refinement against data collected from partially deuterated proteins at resolutions as high as 2.2 Å. At these moderate resolutions, cancellations of neutron density for CH, CH$_2$ and CH$_3$ groups may obscure the direct modeling of side chains [61, 132]. In joint refinement the electron density maps are used
to refine the positions of heavy atoms while the corresponding neutron density maps are only used to refine hydrogen atom positions [131, 132]. Examples of proteins for which both refinement against neutron data alone and joint refinement have been applied include photoactive yellow protein (PYP), transthyretin and DHFR. Yamagushi et al. refined the neutron structure of PYP at 1.5 Å against neutron data alone while Fisher et al. performed joint refinement to determine the PYP neutron structure at 2.5 Å. [133, 134] Both refinement strategies have been applied to neutron structures of transthyretin at equal resolution (2.0 Å) and to the neutron structures of DHFR at 2.0 Å and 2.17 Å resolutions. [20, 21, 38, 135] Protein perdeuteration prevents density cancellation by hydrogen atoms, but joint refinement is still broadly applied to data collected from perdeuterated crystals. The neutron structures of perdeuterated aldose reductase, transthyretin and type III antifreeze protein were jointly refined [38, 41, 69]. In contrast, structures of beta lactamase, myoglobin and rubredoxin were successfully solved using neutron data alone [23, 24, 36, 67].

Joint refinement requires that the X-ray data set be collected at the same temperature as the neutron data set. Collecting both data sets at the same temperature is essential as temperature alters aspects of the protein structure or of the crystal itself [136-140]. X-ray-induced radiation damage can be more severe in crystals measured at room temperature and can produce significant structural artifacts. Radiation damage can be minimized at room temperature by using short exposure times and strong attenuation of high-intensity incident beams. While the X-ray and neutron data set would ideally be collected from the same crystal for joint refinement, the X-ray dataset is most often measured from a smaller crystal grown in the same drop or under the same crystallization conditions as the neutron crystal. Large crystals typically produce lower quality X-ray data due to absorption, and radiation damage from X-ray exposure would make the large crystal unsuitable for subsequent neutron parametric studies (Knihtila, Mattos, and Meilleur, unpublished results).

Currently, there are three refinement programs that can refine neutron protein diffraction data: CNSsolve, phenix.refine, and SHELXL [61, 62, 132, 141-144]. The phenix.refine module of the PHENIX software package is capable of refining X-ray, neutron and joint X-ray/neutron data [61, 142]. The familiarity of phenix.refine to the crystallography
community has made it the most used package for neutron structure refinement. Crystallography and NMR system (CNS) was originally developed as an online server to aid structure determination from solution NMR and X-ray crystallography data [141, 145]. A neutron diffraction patch, termed nCNS, can be added to CNS to perform refinement of neutron or joint X-ray/neutron data with CNSsolve [62, 132]. nCNS has been used for 19 deposited neutron structures. SHELXL has been used to refine neutron structures of crambin, xylose isomerase, DHFR and endothiapepsin [20, 27, 30, 146, 147]. Recently Gruene et al. modified SHELXL to include specific instructions for neutron macromolecule structure refinement [144].

Refinement against neutron data closely follows X-ray refinement protocols, and neutron density maps can be visualized using the program COOT [148, 149]. The starting model is a structure determined at the same temperature using X-ray data alone that has been stripped of all ligands and water molecules. Deuterium or hydrogen atoms are added at all non-titratable positions for perdeuterated or hydrogenated proteins, respectively. Coordinate file editing can be performed manually or automatically by utilities such as *phenix.ready_set* [150]. Deuterium and hydrogen atoms, each with half occupancy, are usually added to the titratable positions when the crystal is H/D exchanged after crystallization. Deuterium atoms with full occupancy can be added if the protein was crystallized from deuterated buffer [67, 69, 87, 115, 116, 147, 151]. In this latter approach incompletely exchanged sites will show a negative density peak in the neutron F_0–F_c map indicating that the site is partially occupied by hydrogen [36]. At this initial state of refinement, titratable groups are assumed to be in their expected protonation state based on the “normal” pK_a value of the individual amino acid in solution.

An initial rigid body refinement and successive coordinate, atomic displacement parameter and occupancy refinements are performed in alternation with local rebuilding of the model including the addition of small molecules. Water molecules can be modeled as O, OD^−, DOD, and DO_3^+ depending on their degree of rotational anisotropy, neighboring groups and ionic state [32, 37, 152]. Water molecules can be oriented automatically by the refinement software. However, visual inspection of the model and neutron density maps should be
performed to confirm the automatically refined orientations. Experimenters should consider manual orientation of the water molecules to form hydrogen bonds with accepted geometries when possible [153]. Map observations will also inform the modeling of titratable group protonation states. Careful interpretation of the maps is essential to model “abnormal” or “perturbed” ionization states. Here the neutron F₀–F c maps are again used to model hydrogen and deuterium atoms. Positive map features directly indicate protonation of carboxylate, hydroxyl, amine and imine groups. Proper orientation of these groups, guided by the neutron F₀–F c maps, often clearly defines hydrogen-bond networks that may be ambiguous if only the electron density maps are considered. A negative F₀–F c feature at a titratable site indicates an overestimation of deuterium occupancy at this site. Deuterium atoms are readily visible as positive features in neutron 2F₀–F c maps at resolution ≤ 2.5 Å. It should be noted that while hydrogen atoms included in the coordinate file will contribute to negative features in the neutron 2F₀–F c map, hydrogen atoms are usually only visible at higher resolutions (≤ 2.0 Å).

1.2.8 Perspective

Atomic comprehension of protein chemistry requires knowing the simultaneous locations of numerous hydrogen atoms. Specifically for enzymes, the interactions driving chemistry generally cannot be unambiguously deduced solely from knowledge of the protonation states of individual titratable side chains within the active site [4]. It is widely accepted that remote amino acids or water molecules, whether buried in the protein structure or bridging the active site to the hydration shell, play essential roles in protein chemistry [154-161]. At the active sites of enzymes, both the chemical identities of ligands, substrates and water molecules (H₂O, OH⁻, or H₃O⁺) and the hydrogen bond networks formed must be known for a complete mechanistic understanding. Furthermore, the modulation of active site pKₐ’s by the local electrostatic environment must be understood.

The neutron crystal structures of human farnesyl pyrophosphate synthase (FPPS) in complex with the clinical inhibitor risedronate and of human transthyretin (TTR) illustrate these points. FPPS catalyzes condensation of dimethylallyl pyrophosphate (DMAPP) with
isopentyl pyrophosphate (IPP) yielding geranyl pyrophosphate (GPP) which further condenses with a second IPP molecule yielding farnesyl pyrophosphate. Nitrogen-containing bisphosphonates (N-BP), such as risedronate, inhibit FPPS by occupying the DMAPP/GPP binding site [162]. N-BPs bind the FPPS active site with a high affinity that was proposed to result from a protonated nitrogen mimicking a carbocation intermediate of the FPPS condensation reaction [163, 164]. The X-ray crystal structure of the FPPS–risedronate complex demonstrated that the risedronate pyridine nitrogen was positioned within hydrogen bonding distance of the Lys200 backbone carbonyl and the Thr201 hydroxyl group. The neutron crystal structure of H/D-exchanged FPPS–risedronate at pD 5.0 determined by Yokoyama et al. confirms the existence of these hydrogen-bond interactions [105]. A deuterium atom binds to the pyridine nitrogen and participates in a bifurcated hydrogen bond with Lys200 and Thr201. This protonation agrees with prior studies of p\(K_a\)'s for the risedronate titratable groups in solution. However, the neutron structure revealed that both risedronate phosphonate groups are fully deprotonated when the drug is bound to FPPS despite the first deprotonation of the phosphonates in free risedronate having a p\(K_a\) > 10.5 [165, 166]. The phosphonate groups together coordinate three Mg\(^{2+}\) ions and participate in hydrogen bonds with Arg112, Lys200 and Lys257 sidechains and four ordered water molecules. While unexpected, phosphonate deprotonation agrees well with a FPPS mechanism in which a basic pyrophosphate oxygen, stabilized by Arg112 and Lys200, abstracts the IPP 2\(R\) proton promoting nucleophilic attack of the carbocation intermediate [163, 164]. These two mechanistic insights resulted directly from unambiguous delineation of the residue, ligand and water protonation states at the enzyme active site by the neutron crystallography experiment.

The neutron structures of human transthyretin (TTR) reveal key hydrogen bond networks and pH sensitivity leading to a new understanding of substrate binding and amyloid formation. TTR is present in the blood, where it binds retinol and thyroxine (T\(_4\)), and in the cerebrospinal fluid, where it binds T\(_4\). The TTR protein is a homotetramer, or a dimer of dimers, held together through an extensive network of hydrogen bonds and hydrophobic interactions. Partial denaturation of TTR, which occurs in acidic conditions, leads to aggregate structures that include amyloid fibril formation. Previous TTR X-ray crystal structures did not
display significant conformational changes when the pH was altered [167, 168]. The neutron structure solved by Yokoyama et al. revealed an extensive hydrogen bond network connecting six residues and four water molecules located along both the monomer–monomer and dimer–dimer interfaces [135]. One of the residues involved in the hydrogen bond network is the singly protonated His88, which stabilizes the dimer–dimer interface. When the protonation state of His88 is altered, the interfacial hydrogen bond network is destabilized. Comparison of the neutron structure solved at pD 7.4 with a previously solved X-ray structure at pH 4.0 shows that the salt bridge between Lys76 and Glu89, the hydrogen bond between Asp74 and Ser77 and the hydrogen bond network including His88 are compromised at low pH. Haupt et al. also described the extensive hydrogen bond network between monomer–monomer interface in their structures leading to the conclusion that the dimer of TTR is the building block to amyloid fibril formation [38]. The neutron structures of TTR increased the understanding of pH sensitivity and suggested a potential mechanism of amyloid formation.

In the last year two neutron structures revealed unexpected protonation of key states of established enzymatic mechanisms. The neutron structure of the small GTPase Ras further illustrates the uniquely rich information content provided by a single neutron diffraction experiment [109]. This study suggests that the \(\gamma\)-phosphate group of the GTP substrate remains protonated at physiological pH when bound at the active site of Ras in the ground state. Efforts to elucidate the GTP hydrolysis mechanism of Ras have relied on the assumption that the \(\gamma\)-phosphate group of GTP bound at the active site of the enzyme would be deprotonated as is observed in solution. The neutron structure revealed for the first time the modulation of the \(pK_a\) of \(\gamma\)-phosphate group by the RAS active site microenvironment and sets a new stage for future investigation.

Neutron structures are usually determined at room temperature, but neutron data collection at cryogenic is necessary to characterize transient states. Intermediate freeze trapping was successfully applied to cytochrome c peroxidase to determine the structure of ferryl heme intermediate and unambiguously determined the protonation state of the iron–oxo species (also known as compound I) [95]. The study demonstrates that compound I is deprotonated supporting earlier data. Comparison of the room temperature neutron ferric
cytochrome c peroxidase structure and of the cryo neutron structure of compound I critically revealed that the heme distal histidine become protonated. Hence both the neutron studies of Ras and cytochrome c peroxidase state a new stage for future calculations and modelling.

Readers are referred to the recent review by Golden and Vrielink for additional examples of neutron crystallographic structures providing insights into hydronium ions, oxyanion holes, low barrier hydrogen bonds and mapping protein–ligand interactions [169].

NMR spectroscopy, ultra high-resolution X-ray crystallography and neutron crystallography contribute to a global understanding of protonation. Taken individually, each technique has specific limitations. Acid–base titrations can be monitored by NMR spectroscopy to determine pK_a values directly [170-172]. However, pH titration methods can only be applied to proteins that are stable over a wide range of pH, and the experiment does not simultaneously report on the residue conformation. Residues surrounding titratable sites also have chemical and structural behaviors sensitive to pH that may complicate chemical shift assignments and pK_a determination. In addition to probing pK_a's, long residence-time water molecules can be localized using NMR spectroscopy, but water molecule orientations are difficult to determine [173]. For X-ray crystallography, the weak X-ray scattering power of hydrogen atoms inherently limits the available experimental information. The protonation state of some titratable sites can be determined indirectly from X-ray structures by covalent bond geometry analysis; however, high-resolution X-ray crystallography proves not to be the most suitable technique to visualize hydrogens directly [40, 65]. Neutron protein crystallography directly identifies H/D atoms, but is inherently limited by the flux of the available neutron beams. However, when large crystals can be grown, a single neutron structure informs simultaneously on all aspects of catalytic activity: structural conformation and protonation state at the active site, conformation and protonation of remote residues, chemical identities and orientations of water molecules and the organization of hydrogen-bond networks.

Neutron crystallography is a powerful complementary approach to NMR spectroscopy and X-ray crystallography. Neutron crystallography and NMR spectroscopic measurements following the titration of tyrosine residues were elegantly combined in a study of carbonic
anhydrase by Michalczyk et al [19]. NMR revealed the perturbed pK$_a$ of Tyr7, and the neutron crystal structures provided a structural context. Similarly, sub-atomic resolution X-ray and neutron crystallography were combined to describe in further detail the catalytic mechanism of DHFR [21].

As we have discussed in this review, neutron protein crystallography does present a number of technical challenges. Deuterium incorporation is essential for neutron diffraction experiments, but the studies reviewed here illustrate that the incorporation required is readily achievable. The exchange of titratable hydrogen atoms suffices when full perdeuteration is not practical. Many crystals will likely endure vapor diffusion H/D exchange. Data collection and structure refinement methods are now well established, and various strategies can be selected to best suit the crystal system and scientific question being addressed.

Growing large protein crystals remains as the final hurdle for the widespread pursuit of neutron protein crystallography. Clearly, all proteins are not amenable to the growth of large crystals. However, a significant fraction of proteins that produce crystals with X-ray diffraction limit $\leq$ 1.5Å (currently, approximately 5000 proteins meet this criteria) is likely to be amenable to neutron diffraction. The foreseen brightness of neutron beams at facilities currently being built (European Spallation Source) or proposed (Second Target Station at the Spallation Neutron Source) combined with novel instrument concepts will contribute to further lowering the crystal size requirement. The application of polarization techniques for neutron protein crystallography will open up new opportunities in studying fully hydrogenated protein crystals [174, 175]. These developments will undoubtedly expand the opportunities to explore protein chemistry at the atomic level.

1.2.9 Acknowledgements

We thank Dean Myles for carefully reading our manuscript and providing input to improve the review. We also thank Steve Tomanicek and Leighton Coates for providing crystallographic data not available in the PDB. WBO and AMB are supported by the NSF foundation under Grant No. 1069091 and the Department of Energy Office of Basic Science.
(DOE BES) under the Graduate Opportunities (GO!) program at Oak Ridge National Laboratory. FM is supported by the College of Agriculture and Life Sciences of North Carolina State University and the DOE BES.
1.2.10 REFERENCES


120. NMX. Available from: https://europeanspallationsource.se/macromolecular-diffractometer-0.


Chapter 2: Crystallization of a fungal lytic polysaccharide monooxygenase expressed from glyco-engineered *Pichia pastoris* for X-ray and neutron diffraction

The following work was reprinted with permission from: W.B. O’Dell, P.D. Swartz, K.L. Weiss and F. Meilleur. *Acta Crystallographica, Section F: Structural Biology Communications*. 2017. F73.

2.1 Synopsis

Polysaccharide monooxygenase 2 from *Neurospora crassa* was expressed from a glyco-engineered strain of *Pichia pastoris*. Crystals of the protein bearing modified N-linked glycans showed improved X-ray diffraction at 100 K (1.20 Å resolution) and permitted room temperature neutron diffraction data collection to 2.12 Å resolution.

2.2 Abstract

Lytic polysaccharide monooxygenases (LPMOs) are carbohydrate disrupting enzymes secreted by bacteria and fungi that break glycosidic bonds via an oxidative mechanism. Fungal LPMOs typically act on cellulose and can enhance the efficiency of cellulose hydrolyzing enzymes that release soluble sugars for bioethanol production or other industrial uses. The enzyme PMO-2 from *Neurospora crassa* (NcPMO-2) was heterologously expressed in *Pichia pastoris* to facilitate crystallographic studies of the fungal LPMO mechanism. Diffraction resolution and crystal morphology were improved by expressing NcPMO-2 from a glyco-engineered strain of *P. pastoris* and by use of crystal seeding methods, respectively. These improvements resulted in high resolution (1.20 Å) X-ray diffraction data collection at 100 K and the production of a large NcPMO-2 crystal suitable for room temperature neutron diffraction data collection to 2.12 Å.
2.3 Introduction

Lytic polysaccharide monooxygenases (LPMOs) are enzymes that disrupt polysaccharide substrates including cellulose, chitin and starch by oxidizing carbon atoms involved in glycosidic bonds [1-3]. Molecular oxygen is activated for insertion into a substrate C–H bond at a mononuclear copper(I/II) active site contained within an extended planar surface that mediates substrate binding. LPMOs have attracted keen interest in the context of bioethanol production since cellulose-active LPMOs can increase the efficiency of cellulose conversion to soluble sugars by conventional hydrolases [4, 5].

X-ray crystallographic studies of both fungal and bacterial LPMOs have revealed that the active site copper(II) is ligated by a “histidine-brace” formed by the conserved N-terminal histidine and a second conserved histidine. The N-terminal histidine provides a bidentate ligand through the amino terminus and an imidazole ring nitrogen while the second histidine provides a monodentate ligand through an imidazole ring nitrogen as shown in Figure 2.1. In fungal LPMOs, the hydroxy group of a tyrosine residue occupies one axial coordination site; however, this site is unoccupied in bacterial LPMOs [6, 7]. Various chemical species including water, chloride ion and superoxide anion have been observed as equatorial and axial ligands to complete the copper coordination sphere [8-10]. Electron-donating proteins or small molecules can potentiate single electron reduction of copper(II) to copper(I) which has a high affinity for molecular oxygen. Upon oxygen binding the copper ion is oxidized to copper(II) by the formation of a copper–superoxide complex [2]. Despite multiple crystallographic, spectroscopic and computational studies, the identity of the activated oxygen species ultimately responsible for hydrogen abstraction from the substrate glycosidic carbon remains unknown. Furthermore, all subsequent mechanistic steps still require structural description [8-13]. Additional crystallographic studies of LPMO reactive intermediates are necessary, and direct interrogation of protonation states and hydrogen atom coordinates via neutron protein crystallography may ultimately reveal key details of the hydrogen abstraction reaction.

The structure of the extracellular LPMO enzyme polysaccharide monooxygenase 2 from Neurospora crassa (NcPMO-2, UniProt Q1K8B6) has been previously determined by X-
Figure 2.1: Structure of NcPMO-2 and the “histidine brace”. (A) Ribbon diagram of NcPMO-2 based on PDB 5TKF. Non-crystallographic symmetry (NCS) molecule A is shown. (B) The LPMO “histidine-brace” coordination consists of three nitrogen ligands here donated by His1 and His84. In fungal LPMOs, an additional copper coordination site is occupied by the hydroxy group of a conserved tyrosine residue (Tyr168).
ray crystallography of the protein as expressed by the source organism. Crystals of NcPMO-2 purified from the source organism diffracted to high resolution ($d_{\text{min}} = 1.10 \text{ Å}$) despite heterogeneity of the glycan conjugated to the single $N$-linked site (Asn60) [8]. In order to continue investigating the LPMO mechanism using both X-ray and neutron crystallography, we chose to produce NcPMO-2 by heterologous expression from the yeast *Pichia pastoris*. *P. pastoris* is a widely used expression host that tolerates high cell density cultivation and typically produces high yields of target proteins [14]. As a eukaryotic host, *P. pastoris* facilitates targeting pre-proteins for secretion, maturation by signal cleavage and conjugation of $N$-linked glycans which are all necessary for the expression of soluble and properly folded NcPMO-2. In addition, *P. pastoris* has been reported to be a suitable host for expressing deuterium ($^2\text{H}$)-labeled proteins which can be beneficial to neutron protein crystallography studies [15-18].

This communication reports our successful crystallization, high resolution X-ray data collection and room temperature neutron data collection of NcPMO-2 heterologously expressed from *P. pastoris*. This success was enabled by protein expression from the glyco-engineered strain *P. pastoris* SuperMan$_5$ that predominantly produces proteins bearing $N$-linked glycans with the composition (mannose)$_5$(N-acetylglucosamine)$_2$ ((Man)$_5$(GlcNAc)$_2$) instead of the high-mannose glycans typically observed in yeast expression products [19, 20]. The combined change of expression host and the use of crystal seeding techniques allowed us to crystallize in a higher-symmetry space group, obtain crystals that diffract to higher resolution and improve crystal morphology as compared with our initial attempts.

### 2.4 Materials and methods

#### 2.4.1 Macromolecule production

NcPMO-2 was heterologously expressed in the host *Pichia pastoris* and purified generally following the methods of Kittl *et al.* [21]. The amino acid sequence including the native N-terminal secretion signal peptide was used as the basis for DNA synthesis of the open
reading frame (ORF) with codon optimization for *Pichia* expression (GeneArt, Thermo Fisher Scientific) as shown in Table 2.1. BstBI and XbaI restriction sites introduced during synthesis were used for subcloning into the pPICZαA expression vector (Thermo Fisher Scientific). (Using the BstBI site removes the *Saccharomyces cerevisiae* α-mating factor secretion signal peptide from the vector.) The resulting plasmid was amplified in *Escherichia coli* DH5α using standard techniques and the bleomycin resistance marker. Purified plasmid was linearized by treatment with PmeI, and the linear product was transformed by electroporation into either the *Pichia pastoris* X-33 wild type or SuperMan5 (HIS+) strains (Thermo Fisher Scientific and BioGrammatics, respectively) according to the equipment manufacturer’s protocol (Bio-Rad). Transformants were selected for bleomycin resistance using media containing 500 μg/mL Zeocin® (Thermo Fisher Scientific). Positive colonies were screened for methanol utilization phenotype and inducible NcPMO-2 expression.

NcPMO-2 expression from the SuperMan5 strain was performed using a fed-batch approach in a bioreactor system with a 3 L starting volume. The bioreactor vessel was charged with 2700 mL fermentation basal salts medium (26.7 mL L⁻¹ 85% (wt./vol.) H₃PO₄, 0.93 g L⁻¹ CaSO₄·2 H₂O, 18.2 g L⁻¹ K₂SO₄, 14.9 g L⁻¹ MgSO₄·7 H₂O, and 4.13 g L⁻¹ KOH) and steam sterilized [22]. To the cooled media was added 33 mL of fermentation trace metals solution (6.0 g L⁻¹ CuSO₄·5 H₂O, 0.08 g L⁻¹ KI, 3.0 g L⁻¹ MnSO₄·H₂O, 0.2 g L⁻¹ Na₂MoO₄·2 H₂O, 0.05 g CoCl₂·6 H₂O, 0.02 g L⁻¹ boric acid, 0.5 g L⁻¹ CaSO₄·2 H₂O, 20 g L⁻¹ ZnCl₂, 65 g L⁻¹ FeSO₄·7 H₂O, 0.4 g L⁻¹ biotin and 19.2 mL L⁻¹ H₂SO₄ (concentrated), Amresco), 240 mL of 50% (wt./vol.) glycerol containing 12 mL L⁻¹ fermentation trace metals, 300 mg Zeocin®, and 5 mL antifoam 204 (Sigma Aldrich) [23]. Concentrated NH₄OH was added until the media equilibrated at pH 5.2. The bioreactor was inoculated at an initial optical density OD₆₀₀ = 4.3 from an overnight culture in buffered minimal glycerol media. The controller maintained culture temperature at 30 °C, dissolved oxygen concentration >30% using agitation and aeration with compressed air and pH > 5.0 by the intermittent addition of 6.5% (wt./vol.) NH₄OH. The culture progressed from batch phase upon depletion of initial glycerol concentration to fed-batch when feeding of 50% (wt./vol.) glycerol containing trace metals was automatically initiated by a spike in dissolved oxygen concentration. Glycerol feeding
continued at a rate of 18 mL h$^{-1}$ L$^{-1}$ until the culture optical density reached OD$_{600} = 110.0$. The vessel temperature was then lowered to 25 °C, and NcPMO-2 induction was induced by the addition of methanol (with 12 mL L$^{-1}$ fermentation trace metals added) at an initial rate of 3.7 mL h$^{-1}$ L$^{-1}$. During the 48-hour induction period, the methanol feeding rate was gradually increased stepwise to 18 mL h$^{-1}$ L$^{-1}$. NcPMO-2 expression from *Pichia* X-33 followed similar procedures with reagent amounts adjusted to a starting culture volume of 300 mL. Secreted NcPMO-2 was harvested by centrifuging the culture volume at 1500 x g and 4 °C for 10 minutes. The supernatant was collected and cleared further by centrifuging at 15000 x g and 4 °C for 30 minutes. Supernatant from the second centrifugation was then filtered through 0.22 μm pore-size polyethersulfone (PES) membranes and stored at 4 °C until purification.

NcPMO-2 was purified by sequential hydrophobic interaction, anion exchange and size exclusion chromatography. Filtered supernatant was warmed to 20 °C prior to slow addition of 176 g L$^{-1}$ (NH$_4$)$_2$SO$_4$ with stirring. This sample was applied to a phenyl sepharose FF high substitution column (GE Healthcare Life Sciences, 150 mL column volume) equilibrated with 20 mM Na(CH$_3$COO$^-$) pH 5.0 containing 176 g L$^{-1}$ (NH$_4$)$_2$SO$_4$. NcPMO-2 eluted at the end of a 7.5 column-volume linear gradient to 100% 20 mM Na(CH$_3$COO$^-$) pH 5.0. Fractions containing NcPMO-2 were pooled, concentrated using 5 kilodalton (kDa) molecular weight cutoff (MWCO) PES centrifugal concentrators and then diafiltered for buffer exchange to 20 mM Tris-HCl pH 8.0. The sample was applied to a MonoQ column (GE Healthcare Life Sciences, 8 mL column volume) equilibrated with the same buffer. NcPMO-2 eluted in the sample loading flow through while multiple contaminant proteins bound to the anion exchange media and were eluted with 20 mM Tris-HCl pH 8.0, 250 mM NaCl. The flow through was concentrated as before and loaded onto a Superdex 75 column (GE Healthcare Life Sciences, 320 mL column volume) equilibrated with 20 mM Na(CH$_3$COO$^-$) pH 5.0. Fractions containing NcPMO-2 were collected and concentrated as before. The final yield of purified NcPMO-2 from the SuperMan$_5$ strain was approximately 50 miligrams protein per liter of starting culture volume.
### Table 2.1: Macromolecule production information.

<table>
<thead>
<tr>
<th>Source organism</th>
<th>Neurospora crassa OR74A</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA source</td>
<td>Synthesis</td>
</tr>
<tr>
<td>Expression vector</td>
<td>pPICZαA</td>
</tr>
<tr>
<td>Expression host</td>
<td>Pichia pastoris X-33 and Pichia pastoris SuperMan5</td>
</tr>
<tr>
<td>Complete amino acid sequence of the construct produced</td>
<td>HTIFSSLEVNGVQGLGEGVRVTYNGPIE</td>
</tr>
<tr>
<td></td>
<td>DVTSASIACNPSNTVASTSVKTQAGTN</td>
</tr>
<tr>
<td></td>
<td>VTAIWRMLSTTGDPADVMDSSHKGPTIA</td>
</tr>
<tr>
<td></td>
<td>YLKKVDNAATASGVNGWFKIQQDGMSSG</td>
</tr>
<tr>
<td></td>
<td>VWGTERVINGKGRHSIKIPECIAPQYLLR</td>
</tr>
<tr>
<td></td>
<td>AEMIALHAASNYPGAQFYMecaQLNVGGT</td>
</tr>
<tr>
<td></td>
<td>GAKTPTTSFGAYSGDPGVK1IYWPPV</td>
</tr>
<tr>
<td></td>
<td>TSYTVPGPSVFTC</td>
</tr>
</tbody>
</table>

#### 2.4.2 Mass spectrometry

Intact mass analysis of purified NcPMO-2 expressed from both X-33 and SuperMans5 strains was performed by MS Bioworks, LLC. Samples of NcPMO-2 in 20 mM Na(CH3COO–) pH 5.0 were diluted with 0.1% (vol./vol.) formic acid and injected onto a Symmetry C4 column (Waters) equilibrated with 0.1% (vol./vol.) trifluoroacetic acid (TFA). NcPMO-2 eluted during a linear gradient to 100% acetonitrile, 0.1% (vol./vol.) TFA and was analyzed by a Q Exactive mass spectrometer (Thermo Scientific) operating with positive mode electrospray ionization. Data were acquired from 600–2500 m/z at resolution 17,500 FWHM (400 m/z) with ten microscans per spectrum. Mass spectra were visualized with XCalibur (Thermo Scientific) and deconvoluted using MagTran 3.0 [24].

#### 2.4.3 Crystallization

The crystallization conditions were adapted from those reported by Li et al. [8]. NcPMO-2 was crystallized by the vapor diffusion technique under similar conditions (Table
2) for protein expressed from both the X-33 and SuperMan strains of \textit{Pichia pastoris}. Hanging drops of 2 μL total volume were setup on siliconized cover slips by mixing equal amounts of protein and reservoir solutions (Table 2.2). Plate-like crystals appeared within a week (Figure 2.3). Crystal morphology could be improved by microseeding 600 μL sitting drops composed of 200 μL protein solution and 400 μL 25 % (wt./vol.) polyethylene glycol (PEG) 3350, 100 mM HEPES pH 6.0 and equilibrated for 1–2 weeks against a 22 % (wt./vol.) PEG 3350, 100 mM HEPES pH 6.0 reservoir. Microcrystals were prepared using PTFE Seed Beads (Hampton Research) and vortexing, and volumes ranging from 1–10 μL were added to the sitting drops. (The pH’s of the drops were not measured before or after equilibration.) Crystals for cryogenic X-ray data collection were harvested and cryoprotected with either reservoir solution with added 25% (vol./vol.) glycerol or with low viscosity cryo-oil (MiTeGen) prior to flash freezing in liquid nitrogen.

A crystal obtained by microseeding was used as a macroseed to obtain the crystal used for room-temperature X-ray and neutron data collection. The macroseed was transferred twice through 22% (wt./vol.) PEG 3350, 100 mM HEPES pH 6.0 and placed in an equilibrated 750 μL sitting drop. The crystal reached final volume within 10 weeks of growth. Once grown, the crystal was mounted in a quartz capillary containing artificial mother liquor formulated in deuterium oxide. Vapor exchange over 30 days promoted replacement of solvent and titratable protein hydrogen atoms with deuterium atoms to increase neutron diffraction intensity and decrease the incoherent neutron scattering background.
Table 2.2: Crystallization conditions.

<table>
<thead>
<tr>
<th>PDB</th>
<th>5TKF</th>
<th>5TKG</th>
<th>5TKI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Expression host</strong></td>
<td><em>Pichia X-33</em></td>
<td><em>Pichia SuperMan</em></td>
<td><em>Pichia SuperMan</em></td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td>hanging drop</td>
<td>sitting drop microseeding</td>
<td>sitting drop microseeding</td>
</tr>
<tr>
<td><strong>Plate type</strong></td>
<td>VDX</td>
<td>9-well glass plate</td>
<td>9-well glass plate</td>
</tr>
<tr>
<td><strong>Temperature (K)</strong></td>
<td>291</td>
<td>291</td>
<td>291</td>
</tr>
<tr>
<td><strong>Protein concentration (mg mL⁻¹)</strong></td>
<td>12.6</td>
<td>12.5</td>
<td>12.2</td>
</tr>
<tr>
<td><strong>Buffer composition</strong></td>
<td>20 mM Na(CH₃COO⁻) pH 5.0</td>
<td>20 mM Na(CH₃COO⁻) pH 5.0</td>
<td>20 mM Na(CH₃COO⁻) pH 5.0</td>
</tr>
<tr>
<td><strong>Composition of reservoir solution</strong></td>
<td>22% (wt./vol.) PEG 3350 pH 6.4</td>
<td>22% (wt./vol.) PEG 3350</td>
<td>22% (wt./vol.) PEG 3350</td>
</tr>
<tr>
<td><strong>Volume and ratio of drop</strong></td>
<td>2.0 μL 1 protein : 1 reservoir</td>
<td>2.25% (wt./vol.) PEG 3350, 100 mM HEPES pH 6.0</td>
<td>2.25% (wt./vol.) PEG 3350, 100 mM HEPES pH 6.0</td>
</tr>
<tr>
<td><strong>Volume of reservoir (mL)</strong></td>
<td>1.0</td>
<td>50.0</td>
<td>50.0</td>
</tr>
</tbody>
</table>

2.4.4 Data collection and processing

X-ray cryocrystallography data were collected using the SER-CAT beamlines of the Advanced Photon Source, Argonne National Laboratory. Room temperature X-ray data were collected using a copper rotating anode home source. All X-ray data were indexed, integrated and scaled using the CCP4 suite [25]. Quasi-Laue neutron diffraction data were collected using the CG-4D IMAGINE beamline at the High Flux Isotope Reactor, Oak Ridge National Laboratory [26]. A total of 26 still diffraction images were collected with 10° phi rotation
steps with three unique crystal orientations to maximize reciprocal space coverage in the combined diffraction dataset. Quasi-Laue data were indexed and integrated using LAUEGEN prior to wavelength normalization and scaling with LSCALE and Scala [25, 27, 28]. Dataset statistics shown in Table 2.3 were calculated from scaled but unmerged intensities using the program phenix.table_one in the PHENIX suite [29]. Datasets 5TKF and 5TKG represent the highest diffraction resolution observed from screening ten or more crystals grown under similar conditions. Refined models are deposited in the Protein Data Bank, and the high-resolution X-ray structure (PDB 5TKG) and the joint X-ray/neutron structure (PDB 5TKI) are described in detail by O’Dell et al. [30].
### Table 2.3: Data collection and processing.

<table>
<thead>
<tr>
<th>PDB</th>
<th>5TKF</th>
<th>5TKG</th>
<th>5TKI (X-ray)</th>
<th>5TKI (neutron)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diffraction source</td>
<td></td>
<td>RIGAKU</td>
<td>HFIR CG-4D IMAGINE</td>
</tr>
<tr>
<td>APS 22-BM</td>
<td>APS 22-ID</td>
<td></td>
<td>MICROMAX-007 HF</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.54</td>
<td>2.8–4.5</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>100</td>
<td>ambient</td>
<td>ambient</td>
</tr>
<tr>
<td>Detector</td>
<td>Rayonix MAR225</td>
<td>Rayonix 300HS</td>
<td>Rigaku RAXIS IV++</td>
<td>Arinax image plate</td>
</tr>
<tr>
<td>Crystal-detector distance (mm)</td>
<td>200</td>
<td>110</td>
<td>90</td>
<td>199</td>
</tr>
<tr>
<td>Rotation range per image (°)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0a</td>
</tr>
<tr>
<td>Total rotation range (°)</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>120a</td>
</tr>
<tr>
<td>Number of exposures</td>
<td>180</td>
<td>180</td>
<td>180</td>
<td>90</td>
</tr>
<tr>
<td>Exposure time per image (s)</td>
<td>1</td>
<td>1</td>
<td>120</td>
<td>26</td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P21</td>
<td>P21</td>
<td>P21</td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>43.30, 67.04, 84.15</td>
<td>67.35, 42.21, 69.48</td>
<td>68.12, 42.23, 70.29</td>
<td>68.12, 42.23, 70.29</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>97.57, 97.61, 97.43</td>
<td>90.00, 98.96, 90.00</td>
<td>90.00, 98.33, 90.00</td>
<td>90.00, 98.33, 90.00</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>35.45 – 2.10</td>
<td>31.82 – 1.20</td>
<td>36.10 – 1.50</td>
<td>32.87 – 2.12</td>
</tr>
<tr>
<td>(2.18 – 2.10)</td>
<td>(1.24 – 1.20)</td>
<td>(1.55 – 1.50)</td>
<td>(2.19 – 2.12)</td>
<td></td>
</tr>
<tr>
<td>Total No. of reflections</td>
<td>102814 (10204)</td>
<td>228097 (21939)</td>
<td>224095 (20772)</td>
<td>31698 (2251)</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>52131 (5173)</td>
<td>117125 (11360)</td>
<td>58482 (5409)</td>
<td>18141 (1381)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.0 (95.9)</td>
<td>97.0 (94.6)</td>
<td>90.2 (80.3)</td>
<td>79.0 (61.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>2.0 (2.0)</td>
<td>1.9 (1.9)</td>
<td>3.8 (3.8)</td>
<td>1.7 (1.6)</td>
</tr>
<tr>
<td>&lt; I/σ(I) &gt;</td>
<td>8.8 (3.5)</td>
<td>7.2 (4.2)</td>
<td>13.8 (4.2)</td>
<td>4.9 (3.6)</td>
</tr>
<tr>
<td>Rmeas</td>
<td>0.1232 (0.4136)</td>
<td>0.1034 (0.2138)</td>
<td>0.0633 (0.3963)</td>
<td>0.1806 (0.2691)</td>
</tr>
<tr>
<td>Overall B factor from Wilson plot (Å²)</td>
<td>16.04</td>
<td>8.77</td>
<td>16.04</td>
<td>8.08</td>
</tr>
</tbody>
</table>

Values for the outer shell are given in parentheses.

*aStill quasi-Laue diffraction images were collected with 10° steps in phi rotation at three unique crystal orientations that were combined into a single dataset. Angular ranges reported are those covered at each crystal orientation.*
2.4.5 Structure solution and refinement

Initial phases were determined by molecular replacement with Phaser as implemented in PHENIX using the prior NcPMO-2 structure by Li et al. [8](PDB 4EIR) which has 99% sequence identity to heterologous NcPMO-2 [29, 31]. Molecular replacement in the P1 space group for 5TKF placed four non-crystallographically symmetric (NCS) molecules of NcPMO-2 within the asymmetric unit with a translation function (TF) Z score = 83.8. Molecular replacement in the P21 space group for 5TKG placed two NCS molecules of NcPMO-2 per asymmetric unit with TF Z score = 135.3. PHENIX AutoBuild was used for automated structure rebuilding to minimize initial phase bias from the replacement model [32]. Further structure refinement was performed using phenix.refine with manual model building performed in Coot [33, 34]. For model 5TKF, non-hydrogen protein atoms were refined in 20 translation-libration-screw (TLS) groups per protein molecule as parametrized by the TLSMD server [35, 36]. For model 5TKG, all non-hydrogen atoms were refined with anisotropic atomic displacement parameters. Final models were analyzed using the Online_DPI web server to determine the Cruickshank diffraction precision index (DPI) [37, 38]. Refinement statistics for both models are shown in Table 2.4.
Table 2.4: Structure solution and refinement.

<table>
<thead>
<tr>
<th>PDB</th>
<th>5TKF</th>
<th>5TKG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Resolution range (Å)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35.45 – 2.10 (2.18 – 2.10)</td>
<td>28.73 – 1.20 (1.24 – 1.20)</td>
</tr>
<tr>
<td></td>
<td>No. of reflections, working set</td>
<td></td>
</tr>
<tr>
<td></td>
<td>52110 (5173)</td>
<td>117101 (11354)</td>
</tr>
<tr>
<td></td>
<td>No. of reflections, test set</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1999 (199)</td>
<td>1997 (202)</td>
</tr>
<tr>
<td></td>
<td>Final $R_{\text{cryst}}$</td>
<td>0.1452 (0.1795)</td>
</tr>
<tr>
<td></td>
<td>Final $R_{\text{free}}$</td>
<td>0.1813 (0.2193)</td>
</tr>
<tr>
<td>Cruickshank DPI</td>
<td>0.142</td>
<td>0.030</td>
</tr>
<tr>
<td>No. of non-H atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6667</td>
<td>3705</td>
</tr>
<tr>
<td>Ligand</td>
<td>183</td>
<td>69</td>
</tr>
<tr>
<td>Water</td>
<td>1008</td>
<td>921</td>
</tr>
<tr>
<td>Total</td>
<td>7858</td>
<td>4699</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds (Å)</td>
<td>0.003</td>
<td>0.009</td>
</tr>
<tr>
<td>Angles (°)</td>
<td>0.73</td>
<td>1.18</td>
</tr>
<tr>
<td>Average $B$ factors (Å$^2$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>19.03</td>
<td>11.30</td>
</tr>
<tr>
<td>Ligand</td>
<td>42.48</td>
<td>23.56</td>
</tr>
<tr>
<td>Water</td>
<td>30.34</td>
<td>26.34</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favoured (%)</td>
<td>96.04</td>
<td>97.0</td>
</tr>
<tr>
<td>Allowed (%)</td>
<td>3.51</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Values for the outer shell are given in parentheses.

2.5 Results and discussion

Mass spectrometry measurements of the protein expressed from *P. pastoris* X-33 (Figure 2.2 A–B) gave intact masses corresponding to water adducts of NcPMO-2 bearing three glycan chains: (Man)$_9$(GlcNAc)$_2$ (predicted 25159.8 Da, observed 25159.0 Da), (Man)$_{10}$(GlcNAc)$_2$ (predicted 25321.9 Da, observed 25321.0) and (Man)$_{11}$(GlcNAc)$_2$ (predicted 25484.13, observed 25482.0). Mass spectrometry measurements of NcPMO-2 expressed from *P. pastoris* SuperMan$_5$ (Figure 2.2 C–D) gave the expected (Man)$_5$(GlcNAc)$_2$ species (predicted 24993.3 Da, observed 24492.0 Da) along with the presence of (Man)$_7$(GlcNAc)$_2$ (predicted 24817.6 Da, observed 24817.0 Da) and (Man)$_{10}$(GlcNAc)$_2$
(predicted 25304.0 Da, observed 25302.0 Da) species. As reported by Vervecken et al. a (Man)_5(GlcNAc)_2 glycan accounts for 90-95% of all N-linked glycosylation of target proteins expressed by this strain [19].
Figure 2.2: Analysis of NcPMO-2 glycoforms by mass spectrometry. (A–B) Electrospray mass spectrum and deconvolution obtained for NcPMO-2 expressed from P. pastoris X-33 showing the presence of (Man)$_6$(GlcNAc)$_2$, (Man)$_8$(GlcNAc)$_2$ and (Man)$_{11}$(GlcNAc)$_2$ glycoforms (as water adducts) along with an unidentified species (star). (C–D) The spectrum and deconvolution for NcPMO-2 expressed from P. pastoris SuperMan5 showing the presence of (Man)$_5$(GlcNAc)$_2$, (Man)$_7$(GlcNAc)$_2$, and (Man)$_{10}$(GlcNAc)$_2$ glycoforms.
Initial crystallization attempts were conducted with NcPMO-2 heterologously expressed from the wild type *P. pastoris* X-33 strain. Crystals could be obtained with a morphology of either thin or stacked plates as shown in Figure 2.3. Synchrotron X-ray diffraction at 100 K could only be observed to a highest resolution of 2.10 Å and revealed a triclinic space group (Table 2.3, 5TKF). Examination of the electron density maps revealed densities corresponding to N-linked glycans attached to Asn60 of each of the four protein molecules in the asymmetric unit [8]. Different lengths of glycans ranging up to approximately seven sugar units could be modeled into the electron densities (Figure 2.4 A). Increases in the B-factor of each subsequent sugar residue from the Asn60 linkage suggest that additional sugar residues could not be observed due to disorder. Notably, the glycans for NCS molecules A, B and D each face toward solvent channels within the crystal and only show interactions with the parent protein chain as shown in Figure 2.4 B; however, the (Man)$_2$(GlcNAc)$_2$ glycan observed for NCS molecule C participates in crystal contacts with Ser48 and Ser50 of NCS molecule A via two water molecules (Figure 2.4 C). This contact is not observed in the prior structure of NcPMO-2 and disrupts the $P_{2_1}$ symmetry of the homodimer of NCS molecules A and B yielding a triclinic space group.
Figure 2.3: Crystals and X-ray diffraction of NcPMO-2 expressed from *P. pastoris* X-33. (A) Crystals grew as thin plates that often overlaid each other forming stacks. (B) Representative synchrotron X-ray diffraction shows a highest resolution of ~ 2.10 Å for these crystals.
Figure 2.4: Asparagine 60 N-linked glycan observed for NcPMO-2 from P. pastoris X-33. (A) A glycan corresponding to \((\text{Man})_5\text{(GlcNAc)}_2\) could be modelled into the electron density \((2F_o-F_c\text{ at } \sigma = 1.0)\) adjacent to Asn60 of NCS molecule A. (B) The four NCS molecules in the \(5\text{TKF} P1\) unit cell are shown as \(C_\alpha\) traces as viewed from the \(bc\) crystal face. The observed glycans are shown as gray sticks while protein and solvent atoms within 3.0 Å of the glycan are shown as spheres. The \(N\)-linked glycan of NCS molecule C (blue) appears to participate in crystal packing with NCS molecule A (red). (C) An enlarged view of the interface between NCS molecule C (blue) and NCS molecule A (orange) shows three possible hydrogen bonding interactions between Ser48 and Ser50 of NCS molecule A with the glycan from NCS molecule C.
As this packing contact between NCS molecules A and C is mediated by the core GlcNAc residues of the N-linked glycan, we initially attempted complete removal of the glycan from purified NcPMO-2 using peptide-N-glycosidase F (PNGase F) to cleave the Asn60–GlcNAc amide bond. Enzymatic glycan cleavage yielded protein with very poor solubility as did the expression of a NcPMO-2 construct bearing an N60D mutation that removed the single N-glycosylation site. As these trials showed the N-linked glycan to be necessary for NcPMO-2 solubility, reducing the length of the conjugated glycan was tried. Prior to attempting treatment of NcPMO-2 from wild-type P. pastoris with exo- or endoglycosidases, we tested the crystallization behaviour of NcPMO-2 expressed from the glyco-engineered P. pastoris SuperMan5. Crystallization with the largely truncated NcPMO-2 glycoforms obtained from this strain under similar conditions also gave crystals in a thin or stacked plate morphology. However, these crystals diffracted synchrotron X-rays to 1.20 Å resolution at 100 K in the $P2_1$ space group (Figure 2.5). This diffraction resolution and space group matches that observed from crystals of natively expressed protein [8]. Microseeds prepared from these crystals and added to pre-equilibrated vapor diffusion drops yielded single crystals with equal diffraction resolution (Table 3, 5TKG). Examination of the electron density maps showed density corresponding to $(\text{GlcNAc})_2$ and $(\text{Man})_1(\text{GlcNAc})_2$ species attached to Asn60 of the two molecules in the asymmetric unit as shown in Figure 2.6 A. (The inability to model the entire glycan is again presumed to result from disorder of the sugar residues farthest from the Asn60 linkage.) In addition, the glycans are not observed to participate in crystal contacts as was the case in the NcPMO-2 structure of Li et al. [8].
Figure 2.5: Crystals and X-ray diffraction of NePMO-2 expressed from *P. pastoris* SuperMan5.

(A) Spontaneously nucleating crystals grew as thin plates that typically formed stacks or clusters. (B) Microseeds prepared from spontaneous crystals and added to pre-equilibrated drops nucleated primarily single crystals with increased growth in all three dimensions. (C) Representative synchrotron X-ray diffraction shows reflections exceeding a resolution of 1.20 Å for the seeded crystals.
Figure 2.6: Asparagine 60 N-linked glycan observed for NcPMO-2 from *P. pastoris* SuperMans. (A) A glycan corresponding to (GlcNAc)₂ could be modelled into the electron density (2F₀–Fᵣ at σ = 1.0) adjacent to Asn60 of NCS molecule A. (B) The truncated N-linked glycans of NcPMO-2 (sticks) in the 5TKG P2₁ unit cell do not participate in intermolecular crystal contacts as all atoms within 3.0 Å of either glycan (spheres) correspond to the NCS molecule to which the glycan is conjugated.

Having improved the diffraction resolution and crystal morphology of heterologous NcPMO-2, we attempted to grow crystals with sufficient volume for neutron protein crystallography. Single crystals grown by microseeding were transferred to larger pre-equilibrated vapor diffusion drops to obtain, ultimately, a NcPMO-2 single crystal measuring ~0.35 mm³ in total volume (Figure 2.7). This crystal was used for successful room temperature neutron and X-ray diffraction data collection to resolutions of 2.12 Å and 1.50 Å, respectively (Table 2.3, 5TKI and Figure 2.8).
Figure 2.7: *NcPMO-2 crystal used for room temperature X-ray and neutron data collection as seen through the quartz capillary mount.* The crystal volume is approximately 0.35 mm³. Many of the fissures in the crystal appeared during room-temperature X-ray data collection which followed the neutron data collection.

Figure 2.8: *Neutron Laue diffraction recorded from \(^{1}\)H/\(^{2}\)H vapor-exchanged NcPMO-2.* Neutron Laue diffraction recorded from \(^{1}\)H/\(^{2}\)H vapor-exchanged NcPMO-2. One hour exposure of the ~0.35 mm³ NcPMO-2 to neutrons with wavelengths 2.8 ≤ \(\lambda\) ≤ 10 Å produced the above pattern. For data collection, the incident wavelengths were restricted to 2.8 ≤ \(\lambda\) ≤ 4.5 Å to minimize spatial and harmonic overlaps.

Glycosylation constitutes a paradoxical situation for protein crystallization. While glycosylation can be required for protein stability or solubility, the chemical and/or
Conformational heterogeneity of glycan chains may prevent crystallization or lead to crystals with limited diffraction resolution [39]. Here we described how the same target protein, NcPMO-2, expressed from wild type and glyco-engineered *P. pastoris* showed an improvement in X-ray diffraction resolution from 2.10–1.20 Å and improvement in crystallographic symmetry under similar crystallization conditions. In addition, successful neutron data collection was ultimately possible with this protein. We attribute the improved crystal symmetry and diffraction resolution to the elimination of an intermolecular protein–glycan crystal contact. Conformational changes in the N-linked glycan caused by truncating to a (Man)$_6$(GlcNAc)$_2$ structure likely allowed the change in crystal packing as the core GlcNAc residues involved in the contact remained in the truncated form. It should also be noted that expression from glyco-engineered *P. pastoris* produced amounts of glycan-modified protein sufficient for growing large crystals and precluded the need for potentially expensive and time consuming post-expression treatment with glycan modifying enzymes. These results demonstrate the utility and potential for improved crystallization behavior of heterologously expressed glycoproteins when expressed from eukaryotic hosts with engineered glycosylation pathways producing shortened and more chemically homogeneous N-linked glycans.

2.6 Acknowledgements

Protein expression and purification experiments were conducted at the Center for Structural Molecular Biology, a DOE Office of Biological and Environmental Research User Facility. Diffraction data were collected at Southeast Regional Collaborative Access Team 22-ID beamline at the Advanced Photon Source and at the CG-4D IMAGINE beamline at the High Flux Isotope Reactor, DOE Office of Science User Facilities. Student support was provided by the ORNL Graduate Opportunities (GO!) Program and by NSF Integrative Graduate Education and Research Training program award 1069091. The authors acknowledge M. B. Goshe for useful discussions during manuscript preparation.
2.7 REFERENCES


Chapter 3

3.1 Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase


3.1.1 Abstract

Lytic polysaccharide monooxygenases have attracted vast attention due to their abilities to disrupt glycosidic bonds via oxidation instead of hydrolysis and to enhance enzymatic digestion of recalcitrant substrates including chitin and cellulose. We have determined high resolution X-ray crystal structures of an enzyme from Neurospora crassa in the resting state and of a copper(II)–dioxo intermediate complex formed in the absence of substrate. X-ray crystal structures also revealed “pre-bound” molecular oxygen adjacent to the active site. An examination of protonation states enabled by neutron crystallography and density functional theory calculations identified a role for a conserved histidine in promoting oxygen activation. These results provide a new structural description of oxygen activation by substrate free lytic polysaccharide monooxygenases and provide insights that can be extended to reactivity in the enzyme–substrate complex.

3.1.2 Results and discussion

Lytic polysaccharide monooxygenases (LPMOs) are copper-dependent metalloenzymes that activate molecular O$_2$ for the net insertion of one oxygen atom into the carbon–hydrogen bond of a glycosidic carbon in polymeric carbohydrates [1, 2]. This oxygenation leads to spontaneous elimination of the glycosidic bond, polysaccharide chain cleavage and, in the case of crystalline substrates, local disruption of chain ordering [3, 4]. Fungal LPMOs are commonly involved in the saprotrophic metabolism of cellulosic biomass,
and this biological role has inspired industrial use of LPMOs to enhance the efficiency of cellulose deconstruction by hydrolytic enzymes that release soluble sugars for biofuels production [5]. Despite industrial use and rapidly growing scientific interest in LPMOs, the reaction mechanism, which overall requires the input of molecular O₂, two electrons and two protons, remains poorly understood. It is generally accepted that the mechanism begins with one-electron reduction of the Cu(II) LPMO resting state to the Cu(I) state that has a high affinity for molecular O₂. O₂ binding to the Cu(I) state has been shown to yield rapid oxidation to Cu(II) with the presumed formation of a copper–superoxide complex [6]. Multiple mechanistic pathways exist from this intermediate complex. Superoxide, hydroperoxyl and oxyl, reflecting different redox and protonation states, have each been proposed as the reactive oxygen species responsible for abstracting hydrogen from a glycosidic carbon of the polysaccharide substrate [5]. Additionally, in the absence of substrate the superoxide species can dissociate from Cu(II) leading to the generation of H₂O₂ [7, 8]. This communication presents new atomistic details of oxygen activation by a fungal LPMO in the absence of substrate and provides insights for understanding oxygen reactivity in both the productive cellulose disrupting and non-productive H₂O₂ generating reactions.

We have determined high resolution (1.20 Å) X-ray crystal structures of a heterologously expressed LPMO from *Neurospora crassa*, NcPMO-2 (NCU01050; UniProt Q1K8B6), which is a family AA9 LPMO (CAZy database) [9, 10]. As was the case for the previously reported structure of natively expressed NcPMO-2 (PDB 4EIR), crystallization occurred in the P₂₁ space group with two protein molecules with non-crystallographic symmetry (NCS) per asymmetric unit [11]. Figure 3.1.1 A-B shows the NcPMO-2 active site in the Cu(II) oxidation resting state as determined from synchrotron X-ray diffraction data collected at 100 K (PDB 5TKG). (Dataset and model refinement statistics for all structures are listed in Table 3.1.6.1 and Table 3.1.6.2, respectively. Uncertainties were derived from diffraction precision index analysis of individual atomic coordinate precision; see SI “Experimental Details” 3.1.6.2.) The single Cu ion is coordinated by the amino (NH₂) nitrogen and N₅ of the N-terminal His1 and Nε of His84 in a T-shaped “histidine brace” motif that is conserved in all known LPMOs. The hydroxy group of Tyr168 is oriented toward an axial Cu
coordination site though with an elongated Cu–O distance. The remaining axial and equatorial Cu coordination sites are occupied by water molecules. The presence of the axial and equatorial waters in the Cu coordination sphere (Cu–H$_2$O$_{eq}$ = 2.00 Å and 1.96 Å; Cu–H$_2$O$_{ax}$ = 2.44 Å and 2.36 Å) is indicative of the Cu(II) oxidation state despite the readiness with which reduction to Cu(I) can occur during synchrotron X-ray exposure [12]. Copper–ligand distances and selected atomic displacement factors and occupancies for both NCS molecules are listed in Table 3.6.1.3.

Figure 3.1.1: NcPMO-2 active site for both NCS molecules in the enzymatic resting state and after treatment with ascorbic acid. A) NCS molecule A resting state B) NCS molecule B resting state C) NCS molecule A ascorbate-treated D) NCS molecule B ascorbate-treated. Electron densities are shown as 2F$_0$-F$_C$ maps contoured at $\sigma = 1.0$. 

132
In contrast with the resting state, treating crystals with an excess of ascorbic acid in the presence of atmospheric \( \text{O}_2 \) prior to flash freezing reduced the Cu(II) site to Cu(I) and permitted the formation of a Cu(II)–dioxo complex in NCS molecule A, shown in Figure 3.1.1C (PDB 5TKH). The ligand, present at 59\% occupancy, coordinates to Cu with \( \eta_1 \) “end-on” geometry at the equatorial site with a Cu–O1 distance of 1.90 ± 0.05 Å and a Cu–O1–O2 angle of 141° (see Table 3.6.1.4). The equatorial oxygen ligand is modeled as a peroxo species (\( \text{O}_2^{2−} \)) due to the observed O1–O2 bond length of 1.44 ± 0.06 Å; however, the precision of the bond length does not conclusively indicate “peroxo” as opposed to “superoxo” as the activated dioxo species observed. Omit maps for this species are shown in Figure 3.1.6.1. The axial water molecule shows a reduced occupancy of 48\%. Incomplete reduction of the Cu ion by ascorbate may account for the partial occupancies observed for the coordinating species.

The activated copper–dioxo complex observed in this structure has a geometry similar to that of a theoretical copper–superoxide active site model of *Thermoascus aurantiacus* LPMO9A; Kjaergaard *et al.* determined this model from density functional theory (DFT) calculations calibrated to solution-state extended X-ray absorption fine structure measurements of *Ta*-LPMO9A in the Cu(II) resting state [6]. Dioxo coordination in the available equatorial coordination site is also consistent with a classical associative–displacement mechanism for hydroperoxo release by an axially coordinated water molecule in the absence of polysaccharide substrate [13]. The equatorial dioxo coordination observed is different from that modeled in the previous *Ne*PMO-2 crystal structure (see Figure 3.1.6.2) which shows superoxide instead occupying the available axial Cu coordination sites of both NCS molecules [11]. However, despite crystallographic evidence supporting these previous axial coordination models, the complexes exhibited Cu–O1 distances = 2.92–2.96 Å that are too distant for a Cu–O coordination interaction [5, 14, 15]. Furthermore, the recently reported crystal structures of *Lentinus similis* (AA9)A in complex with either cellotriose or cellohexaose substrates (PDB 5ACJ and 5ACI, respectively) clearly show that the axial coordination site of this cellulose active LPMO is occluded by substrate binding while the equatorial site faces a cavity presumed to be suitable for \( \text{O}_2 \) binding [16].
Electron density for the alternate NCS molecule in the ascorbate treated crystals, shown in Figure 3.1.1 D, did not support the presence of a copper–dioxo complex but instead revealed the partial occupancy of ordered, “pre-bound” molecular O₂ adjacent to the equatorial Cu coordination site. The presence of molecular O₂ at partial occupancies was also apparent in the electron density maps for both NCS molecules of the resting state crystals (Figure 3.1.1 A-B). Residues His157 and Gln166, which are conserved among all AA9 LPMOs with known crystal structures (see Figure 3.6.1.3), along with Glu30 from the alternate NCS molecule (not shown) are immediately adjacent to the molecular O₂ molecule. As such, this site lacks the hydrophobic character commonly observed for oxygen “pockets” identified by xenon binding studies and instead resembles the more polar oxygen binding sites that have been identified in chloride binding or high pressure O₂ studies [17, 18].

We further characterized the NcPMO-2 active and molecular O₂ sites by determining a room-temperature crystal structure jointly refined against 1.6 Å X-ray diffraction and 2.1 Å neutron diffraction data (PDB 5TKI). For joint X-ray/neutron refinement, heavy atoms are refined against the X-ray dataset while hydrogen atoms are simultaneously refined independently against the neutron diffraction data. The relatively large coherent neutron scattering lengths of hydrogen (and deuterium, ²H) atoms permit accurate modeling even at medium resolutions (dₘᵋᵣₐᵣ₉ ≈ 2 Å) [19-21]. In the NcPMO-2 structure, obtained at pH 5.6 (pD 6.0), neutron scattering length density is apparent in 2F₀–Fᵣ maps at sites occupied by ²H atoms introduced by vapor exchange (Figure 3.1.2). For both NCS molecules His157 is modeled best as neutral and as the Nₑ protonated (²H) tautomer with torsion angle χ₂ = -75.0° (C₅–C₆–C₇–N₈) (see Figure 3.1.6.4). In this side chain conformation protonated Nₑ points directly toward the molecular O₂ pre-binding site observed in the higher resolution 100 K X-ray structures. It would be expected that at more acidic conditions (pH ≈ 5.0) where many cellulases show maximum activity, His157 would become doubly protonated and positively charged. The proximity of His157 to the O₂ pre-binding site (Hₑ–O distances ranging 1.80–2.30 Å) suggests that this doubly protonated histidine may promote O₂ binding or activation.
Figure 3.1: His157 conformation observed by room temperature neutron protein crystallography. Neutron scattering length densities are shown as 2F_{O}-F_{C} maps contoured at σ = 1.0. NCS molecule A is shown. Covalently bonded hydrogen atoms are shown in gray, and hydrogens subject to ¹H/²H exchange are shown in white.

The energetic consequences of the His157 sidechain conformation and Nε protonation that we observe from neutron diffraction have not been evaluated in prior DFT studies of LPMO oxygen activation [6, 22]. We performed DFT geometry optimizations of three active site models (ASMs) derived from our high-resolution resting state X-ray structure and calculated the free energy of molecular O₂ “pre-binding” when His157 exhibits different side chain conformations and protonation states. (All atoms of the ASMs are shown in Figure 3.1.6.5, and optimized atomic coordinates are given in Table 3.1.6.5.) Molecular O₂ addition near neutral His157 in the observed conformation is approximately thermoneutral (ΔG_{add} = -1.1 kcal mol⁻¹) which may explain the partial occupancy of O₂ in each of the X-ray crystal structures. However, upon second protonation of His157 to form the positively charged sidechain, O₂ addition becomes strongly thermodynamically favored (ΔG_{add} = -17.2 kcal mol⁻¹). This suggests that under acidic conditions O₂ may readily occupy the pre-binding site and be poised for activation immediately upon reduction of the resting state Cu(II) ion.
Table 3.1.1: Calculated thermodynamics of molecular oxygen “pre-binding” as a function of His157 conformation and protonation.

<table>
<thead>
<tr>
<th>His157</th>
<th>$\chi_2$ ($^\circ$)[a]</th>
<th>protonation</th>
<th>$\Delta E$ (kcal mol$^{-1}$)</th>
<th>$\Delta H$ (kcal mol$^{-1}$)</th>
<th>$\Delta S$ (cal mol$^{-1}$)</th>
<th>$\Delta G$ (kcal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>-70.2</td>
<td>Ne</td>
<td>0.7</td>
<td>0.7</td>
<td>5.6</td>
<td>-1.1</td>
</tr>
<tr>
<td>positive</td>
<td>-70.2</td>
<td>Ne, N$\delta$</td>
<td>-22.8</td>
<td>-22.8</td>
<td>-18.5</td>
<td>-17.2</td>
</tr>
<tr>
<td>neutral, flipped</td>
<td>106.2</td>
<td>N$\delta$</td>
<td>8.2</td>
<td>8.2</td>
<td>-19.6</td>
<td>14.1</td>
</tr>
</tbody>
</table>

[a] Reported torsion angles are those of the starting coordinates prior to geometry optimization.

Overall, our results provide a new structural perspective on fungal LPMO reactivity with unique crystallographic and computational characterization of O$_2$ binding and activation and of protein protonation states at and near the active site. The structures reported herein have greatest consequence for understanding the H$_2$O$_2$ generating reaction that occurs when LPMOs are exposed to O$_2$ and reducing agent in the absence of substrate. It is also intriguing to consider the activated copper–dioxo complex in the context of substrate oxidation by superimposing its active site structure with those of the Ls-(AA9)–substrate complexes reported by Frandsen et al., as is shown in Figure 3.1.6.6 [16]. (Both Ls(AA9)-A and NcPMO-2 show regiospecificity for oxygenation at C4 of cellosaccharide glycosidic bonds [23].) Considering the positions of O1 and O2 relative to C4 and C1′, C4 regiospecificity would result only from O1 acting as the hydrogen abstracting atom implying that the O1–O2 bond is broken beforehand. However, this structural comparison does not consider possible changes in atomic positions or oxygen species identity that could occur when a LPMO is bound to substrate and activated dioxygen simultaneously. Ultimately, combined X-ray and neutron crystallographic studies of an LPMO exhibiting, in crystallo, both O$_2$ activation and carbohydrate substrate binding may reveal currently elusive details including the chemical identity of the copper–oxygen species responsible for lytic polysaccharide monooxygenation.

3.1.3 Experimental section

Heterologous expression of NcPMO-2 from the host Pichia pastoris, protein purification and protein crystallization were performed according to the methods of O’Dell et
al. [24]. Additional experimental and computational details are provided in 3.1.6 Supporting Information.

3.1.4 Acknowledgements

Protein expression and purification experiments were conducted at the Center for Structural Molecular Biology, a DOE BER User Facility. Diffraction data were collected at SER-CAT 22-ID at the Advanced Photon Source and at CG-4D IMAGINE (NSF MRI 09229719) at the High Flux Isotope Reactor, DOE BES User Facilities. WBO acknowledges student support from NSF IGERT 1069091. FM acknowledges support from USDA NIFA Hatch 211001. PKA acknowledges support from NIH GM105978.
3.1.5 REFERENCES


3.1.6 Supporting information
To accompany “Oxygen activation at the active site of a fungal lytic polysaccharide monooxygenase”

3.1.6.1 Supporting information contents

Experimental details
Table S1. X-ray and neutron crystallographic data statistics
Table S2. X-ray and neutron model refinement statistics
Table S3. Copper–ligand distances, atomic displacement factors and occupancies in the resting state active site structure (PDB 5TKG)
Table S4. Copper–ligand distances, atomic displacement factors and occupancies in the ascorbate-treated active site structure (PDB 5TKH)
Figure S1. Electron density omit maps for oxygen species in untreated and ascorbate-treated X-ray models
Figure S2. Superposition of copper–dioxo model (PDB 5TKH) with PDB 4EIR
Figure S3. Multiple sequence alignment of AA9 proteins with known crystal structures
Figure S4. Fits of His157 conformational and protonation states with observed neutron scattering length density
Figure S5. Geometry optimized DFT active site models and the same superimposed with NCS molecule B resting state active site structure (PDB 5TKG)
Table S5. Atomic coordinates of geometry optimized DFT active site models
Figure S6. Superposition of copper–dioxo model (PDB 5TKH) with PDB 5ACJ and 5ACI
Figure S7. X-ray fluorescence scan of a NcPMO-2 crystal near the copper K-edge
3.1.6.2 Experimental details

3.1.6.2.1 Crystallization and crystal soaking

Heterologous protein production of NcPMO-2 from Pichia pastoris host, crystallization and data collection methods used are described by O’Dell et al. [1]. Protein was crystallized for these studies as expressed with respect to the bound metal ion. X-ray fluorescence scans of crystals from “as expressed” protein, Figure S7, showed an edge centered at 8984 electron volts (eV) corresponding to copper(I) which can be attributed to the facile photoreduction of the copper ion in X-ray fluorescence studies [2]. (Scans near absorption edges of other metals were not performed.) The crystallization behaviors of metal-exchanged forms of NcPMO-2 were used to confirm further that the active site of the crystallized protein predominantly contains a copper ion as opposed to nickel or zinc which have been observed in other AA9 LPMO structures [3, 4]. Apo-NcPMO-2 was prepared by treating the protein with agarose-immobilized nitrilotriacetic acid [5, 6]. Reconstituting the metal coordination site by diluting apo-NcPMO-2 to a 100-fold molar excess of 10 mM CuCl₂ in 20 mM Na(CH₃COO⁻), incubation for two hours, concentration and chromatographic desalting resulted in Cu holo-NcPMO-2 with identical crystallization behavior of the “as expressed” protein. Reconstituting apo-NcPMO-2 with either 10 mM NiSO₄ or 10 mM ZnCl₂ yielded holo-NcPMO-2 which would not crystallize under the precipitant and pH conditions used to crystallize “as expressed” NcPMO-2 for this study.

For ascorbate-treated structures, crystals were harvested and transferred to a soaking solution containing 100 mM HEPES, 100 mM ascorbic acid pH = 6.0 and 25% (wt./vol.) PEG 3350. Crystals were incubated at room temperature for two hours prior to flash freezing for data collection. All crystal manipulations were conducted in air.
3.1.6.2.2 Model Refinement

X-ray diffraction intensities and starting atomic coordinates derived from PDB 4EIR were used for molecular replacement with Phaser as implemented in PHENIX [7-9]. Automated structure rebuilding of the resulting model by PHENIX AutoBuild was used to minimize phase bias from the replacement model [10]. Models were refined against X-ray data using phenix.refine [11]. For PDB 5TKI neutron data were added to the model after several rounds of refinement against the X-ray data so that the model was jointly refined against X-ray data for heavy atoms and neutron data for hydrogen and deuterium atoms [12]. The conformation and protonation state of His157 was determined by refining single Nδ protonated, single Nε protonated and double Nδ,Nε protonated forms against the neutron scattering length density in the two sidechain conformations that could be fitted to the electron density. (Here, “protonated” refers to 100% occupancy of ²H due to the ¹H/²H exchange of the protein crystal prior to neutron and X-ray diffraction measurement.)

Dataset and final refinement statistics shown for all models in Tables S1 and S2 were calculated using phenix.table_one [13]. The precision of atomic coordinates in the models 5TKG (0.030 Å average) and 5TKH (0.035 Å average) were estimated by evaluating the Cruickshank diffraction precision index for the model as implemented in the web server Online_DPI [14-16]. Selected coordinate precision estimates are reported in Tables S3 and S4. Simple omit maps for the dioxygen species observed in 5TKG and 5TKH were calculated by removing the oxygen atom proximal to copper, distal to copper or both oxygen atoms and calculating F₀–Fₐ difference electron density. Positive contours of the F₀–Fₐ density extending for 10.0 Å around the omitted atoms are shown in Figure 3.1.3.

3.1.6.2.3 Multiple sequence alignment

*aurantiacus* GH61-A, *Thelavia terrestris* GH61-E and *Trichoderma reesei* CEL61B were obtained from PDB entries 5TKG, 4EIS, 4QI8, 4D7U, 5ACF, 4B5Q, 2YET, 3EII, and 2VTC, respectively. Multiple sequence alignment was performed with the T-coffee server, and results were formatted using ESPript [17, 18].

### 3.1.6.2.4 Density functional theory calculations

The three active site models (ASMs) used for DFT calculation were derived by extracting coordinates from the resting state X-ray structure (PDB 5TKG) including residues His1, His84, His157, Gln166 and Tyr168 along with the active site copper(II) ion, pre-bound molecular O$_2$ and the axial and equatorial water molecules as shown in Figure 3.1.7. To reduce the total number of atoms in the ASMs, residues His84, His157, Gln166 and Tyr168 were truncated at $C_\beta$ which was modeled as a methyl group. DFT calculations were performed with Gaussian 09 using the uB3LYP functional with the 6-31g** basis set applied to all atoms [19]. The model was implicitly solvated using the polarizable continuum model as implemented in Gaussian 09 with a dielectric constant of 4.24. For geometry optimization, the coordinates of one heavy atom per residue, indicated by a star in Figure 3.1.7, were constrained to their starting values to maintain the relative conformation imposed by the protein backbone. Molecular O$_2$ “pre-binding” energies were estimated by the difference between two forms of each ASM: one with molecular O$_2$ occupying the pre-binding site and the other with molecular O$_2$ moved to a distance >10 Å from other atoms in the model and the pre-binding site left vacant.
3.1.6.3 REFERENCES


Table 3.1.2: (SI Table 1) X-ray and neutron dataset statistics.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>5TKG</th>
<th>5TKH</th>
<th>5TKI</th>
<th>5TKJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident Radiation</td>
<td>X-ray</td>
<td>X-ray</td>
<td>X-ray</td>
<td>Neutron</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.54</td>
<td>2.85-4.50</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>28.73 – 1.20</td>
<td>31.82 – 1.20</td>
<td>36.10 – 1.50</td>
<td>32.87 – 2.12</td>
</tr>
<tr>
<td></td>
<td>(1.24 – 1.20)</td>
<td>(1.243 – 1.20)</td>
<td>(1.55 – 1.50)</td>
<td>(2.19 – 2.12)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 1 2 1 1</td>
<td>P 1 2 1 1</td>
<td>P 1 2 1 1</td>
<td>P 1 2 1 1</td>
</tr>
<tr>
<td>Unit cell (a, b, c) (Å)</td>
<td>67.35, 42.21, 69.48</td>
<td>67.17, 42.24, 69.37</td>
<td>68.12, 42.23, 70.29</td>
<td>68.12, 42.23, 70.29</td>
</tr>
<tr>
<td>(α, β, γ) (°)</td>
<td>90, 98.96, 90</td>
<td>90, 98.65, 90</td>
<td>90, 98.33, 90</td>
<td>90, 98.33, 90</td>
</tr>
<tr>
<td>Total reflections</td>
<td>228097 (21939)</td>
<td>227779 (22042)</td>
<td>224095 (20772)</td>
<td>31698 (2251)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>117125 (11360)</td>
<td>115269 (11178)</td>
<td>58482 (5409)</td>
<td>18141 (1381)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>1.9 (1.9)</td>
<td>2.0 (2.0)</td>
<td>3.8 (3.8)</td>
<td>1.7 (1.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.0 (94.6)</td>
<td>95.7 (93.4)</td>
<td>90.2 (80.3)</td>
<td>79.0 (61.0)</td>
</tr>
<tr>
<td>Mean intensity/sigma(intensity)</td>
<td>7.21 (4.18)</td>
<td>5.86 (4.23)</td>
<td>13.8 (4.2)</td>
<td>4.9 (3.6)</td>
</tr>
<tr>
<td>Wilson B-factor (Å²)</td>
<td>8.77</td>
<td>8.37</td>
<td>16.04</td>
<td>8.08</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.0732 (0.1512)</td>
<td>0.1091 (0.1416)</td>
<td>0.0540 (0.3406)</td>
<td>0.1277 (0.1903)</td>
</tr>
<tr>
<td>Rfree</td>
<td>0.1034 (0.2138)</td>
<td>0.1543 (0.2003)</td>
<td>0.0633 (0.3963)</td>
<td>0.1806 (0.2691)</td>
</tr>
<tr>
<td>CC1/2</td>
<td>0.982 (0.897)</td>
<td>0.950 (0.895)</td>
<td>0.999 (0.892)</td>
<td>0.926 (0.754)</td>
</tr>
</tbody>
</table>
Table 3.1.3: (SI Table 2) X-ray and neutron model refinement statistics.

<table>
<thead>
<tr>
<th>PDB ID</th>
<th>5TKG</th>
<th>5TKH</th>
<th>5TK1 (X-ray)</th>
<th>5TK1 (neutron)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflections used in refinement</td>
<td>117101 (11354)</td>
<td>115249 (11177)</td>
<td>57483 (5061)</td>
<td>18134 (1381)</td>
</tr>
<tr>
<td>Reflections used for R-free</td>
<td>1997 (202)</td>
<td>1992 (203)</td>
<td>1967 (174)</td>
<td>630 (54)</td>
</tr>
<tr>
<td>R-work</td>
<td>0.1142 (0.1275)</td>
<td>0.1279 (0.1174)</td>
<td>0.1476 (0.2550)</td>
<td>0.2142 (0.3143)</td>
</tr>
<tr>
<td>R-free</td>
<td>0.1399 (0.1650)</td>
<td>0.1540 (0.1619)</td>
<td>0.1790 (0.2901)</td>
<td>0.2508 (0.3562)</td>
</tr>
<tr>
<td>CC(work)</td>
<td>0.977 (0.964)</td>
<td>0.953 (0.956)</td>
<td>0.980 (0.893)</td>
<td>0.886 (0.485)</td>
</tr>
<tr>
<td>CC(free)</td>
<td>0.968 (0.926)</td>
<td>0.912 (0.960)</td>
<td>0.980 (0.837)</td>
<td>0.830 (0.105)</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
<td>4699</td>
<td>4503</td>
<td>3841</td>
<td>3841</td>
</tr>
<tr>
<td>Macromolecules</td>
<td>3705</td>
<td>3578</td>
<td>3404</td>
<td>3404</td>
</tr>
<tr>
<td>Ligands</td>
<td>73</td>
<td>62</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>Solvent</td>
<td>921</td>
<td>863</td>
<td>379</td>
<td>379</td>
</tr>
<tr>
<td>Protein residues</td>
<td>446</td>
<td>446</td>
<td>446</td>
<td>446</td>
</tr>
<tr>
<td>RMS(bonds)</td>
<td>0.009</td>
<td>0.008</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>RMS(angles)</td>
<td>1.18</td>
<td>1.10</td>
<td>1.45</td>
<td>1.45</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
<td>97.0</td>
<td>96.0</td>
<td>96.0</td>
<td>96.0</td>
</tr>
<tr>
<td>Ramachandran allowed (%)</td>
<td>2.9</td>
<td>2.9</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Ramachandran outliers (%)</td>
<td>0.23</td>
<td>0.68</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Rotamer outliers (%)</td>
<td>1.7</td>
<td>1.3</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Clash score</td>
<td>2.01</td>
<td>1.67</td>
<td>4.40</td>
<td>9.99</td>
</tr>
<tr>
<td>Average B-factor</td>
<td>14.44</td>
<td>13.66</td>
<td>21.54</td>
<td>21.54</td>
</tr>
<tr>
<td>Macromolecules</td>
<td>11.30</td>
<td>10.89</td>
<td>20.04</td>
<td>20.04</td>
</tr>
<tr>
<td>Ligands</td>
<td>23.56</td>
<td>22.02</td>
<td>36.09</td>
<td>36.09</td>
</tr>
<tr>
<td>Solvent</td>
<td>26.34</td>
<td>24.53</td>
<td>32.75</td>
<td>32.75</td>
</tr>
<tr>
<td></td>
<td>Molecule A</td>
<td>Molecule B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------</td>
<td>------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atomic displacement factor (Å^2)^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copper(II) ion</td>
<td>8.40</td>
<td>7.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>equatorial water</td>
<td>13.68</td>
<td>12.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axial water</td>
<td>17.33</td>
<td>16.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-bound oxygen</td>
<td>19.68</td>
<td>18.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupancies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copper(II) ion</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>equatorial water</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axial water</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-bound oxygen</td>
<td>0.49</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coordinate error (DPI estimate)^b (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>copper(II) ion</td>
<td>0.023</td>
<td>0.021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>equatorial water</td>
<td>0.029</td>
<td>0.027</td>
<td></td>
<td></td>
</tr>
<tr>
<td>axial water</td>
<td>0.032</td>
<td>0.031</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pre-bound oxygen (O1, O2)</td>
<td>0.035, 0.035</td>
<td>0.034, 0.034</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distances (Å)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–H1-Nδ</td>
<td>1.96</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–H1-N_amine</td>
<td>2.13</td>
<td>2.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–H84-Nε</td>
<td>2.00</td>
<td>1.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–Y168-OH</td>
<td>2.64</td>
<td>2.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–H2O_equatorial</td>
<td>2.00</td>
<td>1.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–H2O_axial</td>
<td>2.44</td>
<td>2.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu–O1</td>
<td>4.61</td>
<td>4.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1–O2</td>
<td>1.21</td>
<td>1.20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a All non-hydrogen atoms in the model were refined with anisotropic atomic displacement parameters. Converted isotropic B-factors are shown here.

^b Diffraction precision index
Table 3.1.5: (SI Table 4) Copper–ligand distances and selected atomic displacement factors in the ascorbate-treated active site structure (PDB 5TKH).

<table>
<thead>
<tr>
<th></th>
<th>Molecule A</th>
<th>Molecule B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atomic displacement factor (Å²)(^a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>copper(II) ion</td>
<td>8.39</td>
<td>9.07</td>
</tr>
<tr>
<td>equatorial water</td>
<td>--</td>
<td>34.60</td>
</tr>
<tr>
<td>axial water</td>
<td>28.14</td>
<td>29.21</td>
</tr>
<tr>
<td>peroxide</td>
<td>18.98</td>
<td>--</td>
</tr>
<tr>
<td>pre-bound oxygen</td>
<td>--</td>
<td>26.27</td>
</tr>
</tbody>
</table>

| Occupancies (%)          |            |            |
| copper(II) ion           | 1.00       | 1.00       |
| equatorial water         | --         | 0.60       |
| axial water               | 0.48       | 0.34       |
| peroxide                 | 0.59       | --         |
| pre-bound oxygen          | --         | 0.60       |

| Coordinate error (DPI estimate)\(^b\) (Å) |            |            |
| copper(II) ion           | 0.027       | 0.028      |
| equatorial water         | --          | 0.055      |
| axial water               | 0.050       | 0.051      |
| peroxide (O1, O2)         | 0.041, 0.041| --         |
| pre-bound oxygen (O1, O2) | --          | 0.048, 0.048|

| Distances (Å)            |            |            |
| Cu–H1-N\(_{\delta}\)    | 1.93        | 1.95       |
| Cu–H1-N\(_{\text{amino}}\) | 2.19        | 2.20       |
| Cu–H84-N\(_{\varepsilon}\) | 1.97        | 1.90       |
| Cu–Y168-OH               | 2.66        | 2.73       |
| Cu–H\(_{2}\)\(_{\text{equatorial}}\) | --          | 2.06       |
| Cu–H\(_{2}\)\(_{\text{axial}}\)  | 2.36        | 2.23       |
| Cu–O1                    | 1.90        | 3.57       |
| O1–O2                    | 1.44        | 1.20       |

\(^{a}\) All non-hydrogen atoms in the model were refined with anisotropic atomic displacement parameters. Converted isotropic B-factors are shown here.

\(^{b}\) diffraction precision index
Figure 3.1.3: (SI Figure 1) Electron density $F_O - F_C$ maps for oxygen species in untreated and ascorbate treated X-ray models. Panels A–B show NCS molecules A and B of 5TKG (resting state) with omit density calculated for molecular oxygen. Panels C–D show NCS molecules A and B of 5TKH (ascorbate treated) with omit density calculated for peroxide and molecular oxygen. Omit maps are contoured at $\sigma = 3.5$ with density (positive $F_O - F_C$ contours) from omitting both oxygens, the proximal oxygen and the distal oxygen shown in blue, red and green, respectively. Maps extend 10.0 Å radially around the omitted oxygen atoms.
Figure 3.1.4: (SI Figure 2) Superposition of copper–dioxo model (PDB 5TKH) with PDB 4EIR. Panel A shows PDB 5TKH NCS molecule A superimposed with PDB 4EIR NCS molecule A (orange) along with the identity of the oxygen species in the respective models and the Cu–O1 distance. Panel B shows the same comparison for PDB 4EIR NCS molecule B (purple).
Figure 3.1.5: (SI Figure 3) Multiple sequence alignment of AA9 proteins with known crystal structures. Sequences are identified by PDB and Uniprot accession codes. Strictly conserved residues are highlighted in red, and highly conserved residues are highlighted in yellow. Residues discussed in the main text are denoted by a star.
Figure 3.1.6: (SI Figure 4) Fits of His157 conformational and protonation states with observed neutron scattering length density. Six forms of His157 were refined for both NCS molecules. (NCS molecule A is shown.) Neutron scattering length density (SLD) is shown in gray as 2Fo–Fc maps contoured at $\sigma = 1.0$. Positive and negative excess neutron SLD from Fo–Fc maps contoured at $\sigma = 2.5$ is shown in green and red, respectively. The form shown at upper left was used in the joint X-ray/neutron refined model (PDB 5TKI).
Figure 3.1.7: (SI Figure 5) Geometry optimized DFT active site models and the same superimposed with NCS molecule B resting state active site structure (PDB 5TKG). For the superimposed structures, the active site models are colored in magenta. Panel A–B: “neutral” active site model. Panel C–D: “positive” active site model. Panel E–F: “neutral, flipped” active site model.
**Table 3.1.6:** (SI Table 5) Atomic coordinates of geometry optimized DFT active site models.

<table>
<thead>
<tr>
<th>residue</th>
<th>atom</th>
<th>neutral, flipped</th>
<th>positive</th>
<th>neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>His1</td>
<td>N</td>
<td>-2.38986</td>
<td>-0.69716</td>
<td>1.61173</td>
</tr>
<tr>
<td>His1</td>
<td>H2</td>
<td>-2.22257</td>
<td>0.29737</td>
<td>1.44727</td>
</tr>
<tr>
<td>His1</td>
<td>H3</td>
<td>-2.27977</td>
<td>-0.82816</td>
<td>2.61709</td>
</tr>
<tr>
<td>His1</td>
<td>CA</td>
<td>-3.81072</td>
<td>-0.98818</td>
<td>1.28475</td>
</tr>
<tr>
<td>His1</td>
<td>HA</td>
<td>-4.00859</td>
<td>-0.59507</td>
<td>0.28575</td>
</tr>
<tr>
<td>His1</td>
<td>H</td>
<td>-4.46188</td>
<td>-0.46536</td>
<td>1.99203</td>
</tr>
<tr>
<td>His1</td>
<td>CB</td>
<td>-4.08216</td>
<td>-2.49356</td>
<td>1.33212</td>
</tr>
<tr>
<td>His1</td>
<td>HB1</td>
<td>-3.72929</td>
<td>-2.89382</td>
<td>2.29132</td>
</tr>
<tr>
<td>His1</td>
<td>HB2</td>
<td>-5.16135</td>
<td>-2.66392</td>
<td>1.30371</td>
</tr>
<tr>
<td>His1</td>
<td>CG</td>
<td>-4.44471</td>
<td>-3.20118</td>
<td>0.17938</td>
</tr>
<tr>
<td>His1</td>
<td>CD2</td>
<td>-3.98001</td>
<td>-4.10987</td>
<td>-0.69341</td>
</tr>
<tr>
<td>His1</td>
<td>HD2</td>
<td>-4.95741</td>
<td>-4.56154</td>
<td>-0.74456</td>
</tr>
<tr>
<td>His1</td>
<td>ND1</td>
<td>-2.13881</td>
<td>-2.93457</td>
<td>-0.21020</td>
</tr>
<tr>
<td>His1</td>
<td>CE1</td>
<td>-1.90023</td>
<td>-3.66386</td>
<td>-1.29676</td>
</tr>
<tr>
<td>His1</td>
<td>HE1</td>
<td>-0.98781</td>
<td>-3.66236</td>
<td>-1.86802</td>
</tr>
<tr>
<td>His1</td>
<td>NE2</td>
<td>-2.98747</td>
<td>-4.39018</td>
<td>-1.60869</td>
</tr>
<tr>
<td>His84</td>
<td>CB</td>
<td>1.77894</td>
<td>1.76665</td>
<td>4.17943</td>
</tr>
<tr>
<td>His84</td>
<td>HB1</td>
<td>2.65120</td>
<td>4.12845</td>
<td>4.74869</td>
</tr>
<tr>
<td>His84</td>
<td>HB2</td>
<td>0.93541</td>
<td>1.84011</td>
<td>4.86808</td>
</tr>
<tr>
<td>His84</td>
<td>H</td>
<td>1.99629</td>
<td>2.7718</td>
<td>3.80289</td>
</tr>
<tr>
<td>His84</td>
<td>CG</td>
<td>1.44659</td>
<td>0.83251</td>
<td>3.07244</td>
</tr>
<tr>
<td>His84</td>
<td>CD1</td>
<td>0.31440</td>
<td>0.11216</td>
<td>2.80832</td>
</tr>
<tr>
<td>His84</td>
<td>HD1</td>
<td>-0.57654</td>
<td>0.07381</td>
<td>3.41260</td>
</tr>
<tr>
<td>His84</td>
<td>ND2</td>
<td>2.30227</td>
<td>0.5369</td>
<td>2.02865</td>
</tr>
<tr>
<td>His84</td>
<td>HD2</td>
<td>3.23819</td>
<td>0.89985</td>
<td>1.88360</td>
</tr>
<tr>
<td>His84</td>
<td>CE1</td>
<td>1.68827</td>
<td>-0.31355</td>
<td>1.18593</td>
</tr>
<tr>
<td>His84</td>
<td>HE1</td>
<td>2.15850</td>
<td>-0.66865</td>
<td>0.28751</td>
</tr>
<tr>
<td>His84</td>
<td>NE2</td>
<td>0.46612</td>
<td>-0.59857</td>
<td>1.62508</td>
</tr>
<tr>
<td>His157</td>
<td>CB</td>
<td>7.23766</td>
<td>1.78495</td>
<td>1.45161</td>
</tr>
</tbody>
</table>

158
Table 3.1.6 Continued

<table>
<thead>
<tr>
<th>residue</th>
<th>atom</th>
<th>neutral, flipped</th>
<th>positive</th>
<th>neutral</th>
</tr>
</thead>
<tbody>
<tr>
<td>HisL157</td>
<td>HB1</td>
<td>7.42337</td>
<td>1.33241</td>
<td>2.43169</td>
</tr>
<tr>
<td>HisL157</td>
<td>HB2</td>
<td>6.87685</td>
<td>2.80284</td>
<td>1.61126</td>
</tr>
<tr>
<td>HisL157</td>
<td>H</td>
<td>8.19699</td>
<td>1.84588</td>
<td>0.92676</td>
</tr>
<tr>
<td>HisL157</td>
<td>CG</td>
<td>6.23347</td>
<td>1.00781</td>
<td>0.66469</td>
</tr>
<tr>
<td>HisL157</td>
<td>CD2</td>
<td>5.02149</td>
<td>1.34887</td>
<td>0.10781</td>
</tr>
<tr>
<td>HisL157</td>
<td>HD2</td>
<td>4.54106</td>
<td>2.31781</td>
<td>0.11791</td>
</tr>
<tr>
<td>HisL157</td>
<td>ND1</td>
<td>6.40773</td>
<td>-0.31769</td>
<td>0.31660</td>
</tr>
<tr>
<td>HisL157</td>
<td>HD1</td>
<td>7.20562</td>
<td>-0.89158</td>
<td>0.55320</td>
</tr>
<tr>
<td>HisL157</td>
<td>CE1</td>
<td>5.33417</td>
<td>-0.71623</td>
<td>-0.41525</td>
</tr>
<tr>
<td>HisL157</td>
<td>HE1</td>
<td>5.23689</td>
<td>-1.71728</td>
<td>-0.80808</td>
</tr>
<tr>
<td>HisL157</td>
<td>NE2</td>
<td>4.46618</td>
<td>0.26869</td>
<td>-0.56372</td>
</tr>
<tr>
<td>HisL157</td>
<td>HE2</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Gin166</td>
<td>CB</td>
<td>0.55719</td>
<td>3.92806</td>
<td>-2.52961</td>
</tr>
<tr>
<td>Gin166</td>
<td>HB1</td>
<td>0.08141</td>
<td>4.68934</td>
<td>-3.15259</td>
</tr>
<tr>
<td>Gin166</td>
<td>HB2</td>
<td>1.56932</td>
<td>4.26870</td>
<td>-2.29229</td>
</tr>
<tr>
<td>Gin166</td>
<td>H</td>
<td>-0.01214</td>
<td>3.85176</td>
<td>-1.60022</td>
</tr>
<tr>
<td>Gin166</td>
<td>CG</td>
<td>0.58291</td>
<td>2.58317</td>
<td>-2.82387</td>
</tr>
<tr>
<td>Gin166</td>
<td>HG1</td>
<td>-0.43585</td>
<td>2.25997</td>
<td>-3.51066</td>
</tr>
<tr>
<td>Gin166</td>
<td>HG2</td>
<td>1.12080</td>
<td>2.69406</td>
<td>-4.22899</td>
</tr>
<tr>
<td>Gin166</td>
<td>CD</td>
<td>1.23078</td>
<td>1.50111</td>
<td>-2.44610</td>
</tr>
<tr>
<td>Gin166</td>
<td>NE2</td>
<td>2.54476</td>
<td>1.31728</td>
<td>-2.57603</td>
</tr>
<tr>
<td>Gin166</td>
<td>HE1</td>
<td>3.06406</td>
<td>1.88312</td>
<td>-3.23192</td>
</tr>
<tr>
<td>Gin166</td>
<td>HE2</td>
<td>3.09010</td>
<td>0.75144</td>
<td>-1.91163</td>
</tr>
<tr>
<td>Gin166</td>
<td>OE1</td>
<td>0.54639</td>
<td>0.83307</td>
<td>-1.62363</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CB</td>
<td>-0.05149</td>
<td>5.25240</td>
<td>0.81138</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HB1</td>
<td>-0.05845</td>
<td>6.09622</td>
<td>0.11080</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HB2</td>
<td>-0.64160</td>
<td>5.60806</td>
<td>1.74746</td>
</tr>
<tr>
<td>Tyr168</td>
<td>H</td>
<td>-0.16261</td>
<td>4.99070</td>
<td>1.00557</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CG</td>
<td>-4.31159</td>
<td>4.07747</td>
<td>0.24848</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CD1</td>
<td>-4.90569</td>
<td>3.20030</td>
<td>-0.67323</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HD1</td>
<td>-5.94427</td>
<td>3.35074</td>
<td>-0.95710</td>
</tr>
<tr>
<td>residue</td>
<td>atom</td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CD2</td>
<td>-2.97634</td>
<td>3.83872</td>
<td>0.59448</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HD2</td>
<td>-2.48625</td>
<td>4.49036</td>
<td>1.31372</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CE1</td>
<td>-4.19594</td>
<td>2.14911</td>
<td>-1.24769</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HE1</td>
<td>-4.66207</td>
<td>1.49153</td>
<td>-1.97480</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CE2</td>
<td>-2.24886</td>
<td>2.78174</td>
<td>0.03524</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HE2</td>
<td>-1.21079</td>
<td>2.62162</td>
<td>0.31717</td>
</tr>
<tr>
<td>Tyr168</td>
<td>CZ</td>
<td>-2.85211</td>
<td>1.94073</td>
<td>-0.91008</td>
</tr>
<tr>
<td>Tyr168</td>
<td>OH</td>
<td>-2.19885</td>
<td>0.89636</td>
<td>-1.50544</td>
</tr>
<tr>
<td>Tyr168</td>
<td>HH</td>
<td>-1.22169</td>
<td>1.01003</td>
<td>-1.47009</td>
</tr>
<tr>
<td>Copper</td>
<td>CU</td>
<td>-0.87737</td>
<td>-1.71927</td>
<td>0.68548</td>
</tr>
<tr>
<td>Water1</td>
<td>O</td>
<td>0.28036</td>
<td>-1.71539</td>
<td>-1.04726</td>
</tr>
<tr>
<td>Water1</td>
<td>H</td>
<td>0.50046</td>
<td>-0.79102</td>
<td>-1.34916</td>
</tr>
<tr>
<td>Water1</td>
<td>H</td>
<td>1.06738</td>
<td>-2.26390</td>
<td>-1.16686</td>
</tr>
<tr>
<td>Water2</td>
<td>O</td>
<td>0.02505</td>
<td>-3.50609</td>
<td>1.67091</td>
</tr>
<tr>
<td>Water2</td>
<td>H</td>
<td>0.57356</td>
<td>-3.43619</td>
<td>2.46425</td>
</tr>
<tr>
<td>Water2</td>
<td>H</td>
<td>-0.43397</td>
<td>-4.35538</td>
<td>1.72701</td>
</tr>
</tbody>
</table>
Figure 3.1.8: (SI Figure 6) Superposition of copper–dioxo model (PDB 5TKH) with PDB 5ACJ and PDB 5ACI. The active site of NCS molecule A from model 5TKH is shown in green with corresponding residues and substrates from 5ACJ and 5ACI shown in orange and purple, respectively. Non-hydrogen protein atoms were superimposed with RMSD = 0.25 Å (5THK–5ACJ) and RMSD = 0.29 Å (5TKH–5ACI).
Figure 3.1.9: (SI Figure 7) X-ray fluorescence scan of a NcPMO-2 crystal near the copper K-edge. The center of the absorption edge is near 8984 eV which is indicative of a copper(I) species.
3.2 Structural studies of *Neurospora crassa* LPMO9D and redox partner CDHIIA using neutron crystallography and small–angle scattering

The following work was reprinted with permission from: A.M. Bodenheimer*, W.B. O’Dell*, C.B. Stanley and F. Meilleur. *Equal contributors. *Carbohydrate Research. (In press.) DOI:10.1016/j.carres.2017.03.001

3.2.1 Abstract

Sensitivity to hydrogen/deuterium and lack of observable radiation damage makes cold neutrons an ideal probe for the structural studies of proteins with highly photosensitive groups such as the copper center of lytic polysaccharide monooxygenases (LPMOs) and flavin adenine dinucleotide (FAD) and heme redox cofactors of cellobiose dehydrogenases (CDHs). Here, neutron crystallography and small-angle neutron scattering are used to investigate *Neurospora crassa* LMP09D (NcLPMO9D) and CDHIIA (NcCDHIIA), respectively. The presence of LPMO greatly enhances the efficiency of commercial glycoside hydrolase cocktails in the depolymerization of cellulose. LPMOs can receive electrons from CDHs to activate molecular dioxygen for the oxidation of cellulose resulting in chain cleavage and disruption of local crystallinity. Using neutron protein crystallography, the hydrogen/deuterium atoms of NcLPMO9D could be located throughout the structure. At the copper active site, the protonation states of the side chains of His1, His84, His157 and Tyr168, and the orientation of water molecules could be determined. Small-angle neutron scattering measurements provided low resolution models of NcCDHIIA with both the dehydrogenase and cytochrome domains in oxidized states that exhibited elongated conformations. This work demonstrates the suitability of neutron diffraction and scattering for characterizing enzymes critical to oxidative cellulose deconstruction.
3.2.2 Introduction

Efficient enzymatic degradation of cellulose is essential to the viability of cellulosic ethanol production. Fungal lytic polysaccharide monooxygenases (LPMOs), which are classified as the Auxiliary Activity 9 (AA9) family of the CAZy database, have recently emerged as synergistic partners of glycoside hydrolase (GH) enzymes in the efficient degradation of cellulose [1, 2]. Cellulose-active LPMOs oxidize glycosidic bonds leading to chain cleavage and disruption of local cellulose crystallinity. This new paradigm for cellulose degradation relies on the activation of molecular oxygen at the copper(I/II) containing LPMO active site. X-ray crystallographic studies have revealed that the Cu ion is ligated by a “histidine brace” formed by two conserved histidine residues, including the N-terminal histidine. In AA9 LPMOs, the hydroxy group of a tyrosine occupies an axial coordination site. Cellobiose dehydrogenases (CDHs) or small molecules can potentiate single electron reduction of Cu(II) to Cu(I) which has a high affinity for molecular oxygen. We have recently solved the high-resolution X-ray structures of *Neurospora crassa* LPMO9D (NcLPMO9D) in resting state and O2-activated state. The resting state revealed the pre-binding of molecular dioxygen in a hydrophilic pocket bordered by conserved residue His157. In the activated state, molecular dioxygen has migrated to the copper center and occupies an equatorial coordination site [3]. The lack of information on the positions of hydrogen atoms in the X-ray crystallographic structures prevented the discrimination between “peroxo” and “superoxo” species. Atomic structures of the subsequent reaction intermediates including the determination of the protonation states of titratable residues remain to be determined for a complete description and understanding of the LPMO mechanism [4].

CDHs are flavocytochrome proteins. CDHs oxidize cellobiose at the dehydrogenase domain using an FAD cofactor. Electrons are shuttled one at a time to the N-terminal cytochrome domain which can further donate to an LPMO [5, 6]. The dehydrogenase (DH) and cytochrome (CYT) domains are connected by a long, flexible, and glycosylated linker. In addition to the DH and CYT domains, NcCDHIIA has a C-terminal carbohydrate binding module (CBM) attached to the DH domain. The recently solved crystallographic structures of
full length *Neurospora crassa* and *Myriococcus thermophilum* CDHs revealed that the proteins adopt closed and extended conformations and small-angle X-ray scattering (SAXS) confirmed the presence of these conformations in solution [6]. The closed state is proposed to promote intramolecular electron transfer between the DH and CYT domains, while the open conformations could facilitate electron transfer from the CYT domain to a LMPO [6]. Given the inherent structural flexibility, further structural investigation is required to obtain a complete picture of CDH intra- and intermolecular electron transfer mechanism.

Metalloproteins, including *Nc*PMO9D and *Nc*CDHIIA, are inherently sensitive to photoreduction induced by the high energy (~10 keV) X-rays typically used in crystallography and small-angle scattering studies. For example, X-ray absorption spectra recorded concurrently with X-ray diffraction data at 100 K demonstrated metal ion reduction in studies of putidaredoxin and T-6 bovine insulin [7, 8]. Metal reduction is often accompanied by structural changes within the metal ion coordination sphere. Such structural effects of photoreduction of the LPMO copper center have been methodically characterized in a series of X-ray structures solved against crystallographic data sets collected with increasing X-ray doses [9]. The most noticeable signature of photoreduction of the LPMO active site from Cu(II) to Cu(I) is the absence of exogeneous ligands, which are most often water molecules, within coordinating distance of the copper ion [10]. In our earlier crystallographic work, we highly attenuated the incident synchrotron X-ray beam to prevent photoreduction and preserve an unperturbed Cu(II) active site as evidenced by the presence of well-ordered (low B-factor) coordinating axial and equatorial water molecules.

X-ray induced photoreduction does not only produce local chemical disturbance. In low-resolution structural techniques, radiation damage can induce long range reorganization including aggregation or unfolding as characterized by SAXS [11, 12]. Methods to mitigate radiation damage during SAXS measurements are well documented; however, fully preventing photoreduction of redox centers can be challenging [11]. In a series of studies on cytochrome P450 reductase combining nuclear magnetic resonance (NMR), SAXS, SANS and optical spectroscopy, Roberts *et al.* demonstrated that partial photoreduction of the redox centers
during SAXS data collection generated a mixture of conformational states that mislead the modeling of the oxidized state done in an earlier study [13, 14].

In marked contrast with X-ray radiation, cold neutrons (< 25 meV) used in structural studies of biological molecules do not cause measurable damage to the samples. The oxidation states of metal centers or redox cofactors are not altered during neutron crystallography data collection, so the geometric and electronic coordination environments can be correctly assigned to the known redox state of the center. In addition to providing a radiation damage-free model of the protein studied, neutron protein crystallography has the added benefit over X-ray crystallography of exquisite sensitivity to hydrogen (\(^{1}\)H or \(^{2}\)H) atoms. This allows the determination of the protonation states of titratable residues and the orientation of functional and structural water molecules [15]. In small-angle neutron scattering (SANS), the lack of photoreduction ensures that the protein dimensions, described by the radius of gyration (\(R_g\)) and maximum dimension (\(d_{\text{max}}\)) derived from the measurement are not affected by photoreduction and correspond to single redox state population. Combined with deuterium labeling and contrast variation techniques, SANS in its typical use, further allows the structural investigation individual components within protein-protein complexes [16].

Here we have applied neutron protein crystallography and SANS to investigate the structures of \(\text{NcLPMO9D}\) and \(\text{NcCDHIIA}\), respectively. The neutron crystallographic structure of \(\text{NcLPMO9D}\) in the resting state allowed the determination of protonation states of titratable groups in the oxidized protein and orientation of important water molecules such as the two waters coordinating the Cu(II) ion which are clearly visible in the neutron scattering length density (SLD) Fourier maps. The SANS analysis of \(\text{NcCDHIIA}\) revealed that, in the fully oxidized state, \(\text{NcCDHIIA}\) adopts an extend conformation without indication of a closed conformation being present in solution.
3.2.3 Results and discussion

3.2.3.1 Protonation states, water orientations and hydrogen bonding at the active site of \textit{Nc}LPMO9D

A neutron crystal structure has recently been determined for the \textit{N. crassa} LPMO \textit{Nc}LPMO9D (PDB 5TKI) \cite{3, 17}. (It should be noted that this structure was determined from \textit{Nc}LPMO9D in which non-exchangeable hydrogens had natural isotopic abundance while titratable groups were exchanged to contain $^2$H \textit{in crystallo}. As such, neutron SLD is only apparent in Fourier maps for $^2$H atoms due to cancellations arising from the negative scattering length of $^1$H.) This structure of the Cu(II) resting state under ambient conditions demonstrates how neutron protein crystallography is an important complementary tool for structurally characterizing the LPMO enzyme mechanism as many currently proposed intermediates differ only in the locations of hydrogen atoms \cite{4, 18}. Protonation states of substrates, intermediates and amino acid groups can often be observed directly by neutron diffraction. For example, neutron SLD is clearly present for the $^2$H atoms of the Gln166 amido and Tyr168 hydroxy groups as shown by a 2F\textsubscript{O}–F\textsubscript{C} map in Figure 3.2.1. The extent to which these atoms contribute to the total observed diffraction is apparent from the significant ($\sigma > 4.0$) excess density that appears in an F\textsubscript{O}–F\textsubscript{C} map calculated with these $^2$H atoms omitted. The neutron SLD maps confirm the presence of the expected hydrogen bond between Gln166 and Tyr168 and rule out the possibility of a deprotonated tyrosinate participating in copper coordination as was considered in early discussions of LMPO active site coordination \cite{19, 20}.

In addition to revealing protonation, neutron SLD maps permit unambiguous orientation of hydroxy groups, asparagine/glutamine/histidine side chains and ordered solvent molecules that provide a clear description of hydrogen bonding. Figure 3.2.2 shows numerous unambiguous hydrogen bonds around the \textit{Nc}LPMO9D active site. The hydrogen bond network connecting the active site Cu(II) ion to Gln166 and His157 \textit{via} the coordinating Tyr168 and/or the equitorial-coordinated water is conserved among cellulose active LPMOs from both fungi and bacteria \cite{4}. Hydrogen bond donation from Gln166 and His157 to a single
Figure 3.2.1: Hydrogen bond between Tyr168 and Gln166. Hydrogen atoms that have undergone solvent exchange with $^2$H are shown in white and non-exchangeable hydrogens ($^1$H) are shown in gray. Neutron SLD is shown as a $2F_O-F_C$ map scaled at $\sigma = 1.0$ (gray mesh). Excess neutron SLD, shown as a $F_O-F_C$ map scaled at $\sigma = 4.0$ (green mesh), results when the hydroxy and amide $^2$H atoms are omitted from the model. The hydrogen bond is indicated by the dashed line.

water molecule is consistent with the recently-characterized role of this site in “pre-binding” molecular O$_2$ prior to the formation of a reactive copper–oxygen complex [3]. Hydrogen bonding can also be confirmed between the His1 backbone carbonyl oxygen and the “pocket water” molecule shown by Frandsen et al. in X-ray structures of LsAA9A to interact with carbohydrate substrate [21]. Furthermore, a hydrogen bond is apparent between protonated His1 $N_\varepsilon$ and a water molecule that is also hydrogen bonded to an alternate conformation of Tyr25 (not shown). Fungal LPMOs expressed natively or heterologously from Aspergillus oryzae show post-translational methylation of $N_\varepsilon$ [10, 22]. In the case of NcLMO9D, His1$N_\varepsilon$ methylation would prevent this hydrogen bond which could, through the conformational flexibility of Tyr25, increase the interaction of the active site with bulk solvent.
Figure 3.2.2: Water molecules involved in hydrogen bonds at the NcLPMO9D active site. Solvent-exchangeable hydrogen atoms are shown in white and non-exchangeable hydrogens are shown in gray. Subscripts “e” and “p” denote the equatorially-bound and “pocket” water molecules, respectively. Neutron SLD is shown as a 2F₀–Fₐ map scaled at σ = 1.0 (gray mesh). Dashed lines indicate hydrogen bonds. The water molecule axially-coordinated to Cu(II) is omitted for clarity.

3.2.3.2 Solution structure of NcCDHIIA

The SANS intensity profile measured for full-length glycosylated NcCDHIIA at pD 5.5 is shown in Figure 3.2.3 A with the associated Guinier fit displayed in Figure 3.2.3 B. The linear trend observed in the Guinier plot indicates that the NcCDHIIA sample was free of interparticle interactions and aggregation. The radius of gyration (Rₕ) value determined from the Guinier fit is 36.72 ± 1.34 Å. The P(r) profile has a typical bell shape curve with the maxima at 38.22 Å and dₘₐₓ of 142 Å (Figure 3.2.4).

The crystal structures of NcCDH (PDB ID 4QI7) and MtCDH (PDB ID 4QI6) were recently solved in open and closed states, respectively [6]. The theoretical Rₕ and dₘₐₓ of the crystallographic structures were determined with calculated scattering curves using the FoXS server [23, 24]. The results are presented in Table 3.2.1.
Table 3.2.1: Comparison of calculated and experimental CDHIIA’s $R_g$ (Å) and $d_{\text{max}}$ (Å).

<table>
<thead>
<tr>
<th>Protein</th>
<th>$R_g$ (Å)</th>
<th>$d_{\text{max}}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>N. crassa; open (PDB 4QI7)</em></td>
<td>35.47</td>
<td>122</td>
</tr>
<tr>
<td><em>M. thermophilum (PDB 4QI6)</em></td>
<td>28.27</td>
<td>101</td>
</tr>
<tr>
<td>Neutron–Experiment**</td>
<td>38.22</td>
<td>142</td>
</tr>
<tr>
<td>Neutron–Model 1***</td>
<td>38.68</td>
<td>134</td>
</tr>
<tr>
<td>Neutron–Model 2***</td>
<td>37.69</td>
<td>129</td>
</tr>
</tbody>
</table>

*calculated from crystallographic structures.
**experimentally determined from SANS data
***calculated from SANS models

Figure 3.2.3: SANS data and analysis for *Nc*CDHIIA. Top: Reduced data are shown in a double logarithmic plot. Bottom: Markers are the data shown in a Guinier plot and, the line is a linear fit to the SANS data for determining $R_g$. 
The SANS data suggest that under the conditions of the experiment, NcCDHIIA adopts a conformation that is more extended than that captured by the crystallographic study. This result agrees with a complementary SAXS study performed by Tan et al. that demonstrated that a wide range of conformations, encompassing the crystallographic models, are present in solution [6]. SAXS data from glycosylated NcCDHIIA were modeled with seven clusters, with $R_g$ and $d_{max}$ ranging from 29.98–41.38 Å and 97.43–132.57 Å, respectively. The SANS data presented here were modeled using SASSIE [25]. The data could be best modeled with two conformations ($\chi^2 < 2.28$). The models resemble the most extended clusters from the SAXS data with the most noticeable difference occurring at the carbohydrate binding module (CBM) (Figure 3.2.5). In the SAXS modeling calculations the CBM is tucked into the dehydrogenase domain, but it is free to move in the models derived from the SANS profile. The small number of clusters from our modeling analysis suggests that under the condition of the SANS experiment CDH adopts exclusively an extended form. The closed state has been suggested to be the interdomain electron transfer (IET) competent conformation while the open form would be inactive [26]. The SANS data presented here support this view. The DH and CYT domains of CDHs are sensitive to photoreduction. Under the conditions of the SAXS experiments it is therefore not unexpected that a fraction of the protein molecules became
photoreduced and adopted an IET competent conformation. In the neutron beam, the protein remained fully oxidized with the DH and CYT domain moving freely.

**Figure 3.2.5: Models produced by SASSIE modeling.** The two-full length *NcCDHIIA* models are represented in grey and green, respectively. The DH domains are superimposed. The glycans are not shown for clarity.

### 3.2.4 Conclusions

Structural studies of *Neurospora crassa* LPMO9D and CDHIIA enzymes conducted using neutron crystallography and small-angle scattering, respectively, revealed new structural details about these enzymes. Neutron crystallography is well suited to the structural investigation of enzymatic mechanisms at the atomic level because of the visibility of hydrogen/deuterium atoms in the resulting neutron SLD maps [15]. Small-angle neutron scattering is typically used to study biological complexes where variation in the hydrogen atom content of the components can be used to sequentially model individual molecules [27]. In the neutron crystallography and SANS study of *NcLPMO9D* and *NcCDHIIA* reported here, the unique sensitivity of neutrons to hydrogen/deuterium atoms was leveraged further by the lack
of radiation damage in the study of two photoreduction sensitive redox proteins. The body of structural knowledge on LPMOs and CDHs gained from X-ray techniques is growing rapidly shedding light on their enzymatic mechanisms. It is important for X-ray photoreduction of these systems to be considered and mitigated. Neutron crystallography and SANS provide alternative approaches.

3.2.5 Experimental

3.2.5.1 NcLPMO9D expression, purification and crystallography

NcLPMO9D was expressed heterologously from *Pichia pastoris* and purified as previously described [17]. X-ray and neutron diffraction data collection, processing and model refinement have been described by O’Dell *et al.*[3, 17]. Briefly, neutron diffraction was recorded from a single ~0.35 mm³ crystal of NcLPMO9D that had undergone ¹H/²H exchange by vapor diffusion from an artificial mother liquor formulated in ²H₂O. Twenty-six neutron quasi-Laue diffraction frames (λ = 2.85–4.50 Å) were collected at ambient temperature using the IMAGINE instrument at the High Flux Isotope Reactor, Oak Ridge National Laboratory [28]. Subsequently, monochromatic X-ray diffraction was recorded at ambient temperature from the same crystal. Diffraction data were processed, and a single structural model was jointly refined against the 2.12 Å resolution neutron and 1.60 Å resolution X-ray datasets using the program *phenix.refine*; parameters for non-hydrogen atoms were refined solely against the X-ray data while hydrogen atom parameters were refined solely against the neutron data [29].

3.2.5.2 NcCDHIIA expression and purification

NcCDHIIA was expressed and purified following a protocol adapted from Sygmund *et al.*[30]. NcCDHIIA was expressed in *Pichia pastoris* KM71H cells (Invitrogen). Cells were cultivated in 2 L buffered glycerol-complex medium (BMGY) at 30 °C. BMGY was removed and cells were resuspended into buffered methanol-complex medium to induce NcCDHIIA
expression. Additional methanol was added 24 h after induction. Centrifugation separated the cells from the supernatant containing secreted $Nc$CDHIIA. Slow addition of ammonium sulfate to a final concentration of 20% saturated solution was followed by centrifugation. A phenyl sepharose HP column (GE Healthcare) was equilibrated with 50 mM sodium acetate, pH 5.5, containing 20% ammonium sulfate saturated solution. The filtered supernatant was applied to the phenyl sepharose HP column and eluted with a linear gradient of 20% to 0% ammonium sulfate. The eluted $Nc$CDHIIA buffer was exchanged into 50 mM sodium acetate, pH 5.5. $Nc$CDHIIA was applied to a Mono Q HR column (Pharmacia) pre-equilibrated with 50 mM sodium acetate, pH 5.5. A linear gradient from 0–1 M NaCl eluted CDHIIA. A Superdex 200 column (GE Healthcare) was used as final size exclusion and buffer exchange step into 50 mM sodium acetate, pH 5.5.

### 3.2.5.3 Small-angle scattering sample preparation, data collection and analysis

The $Nc$CDHIIA sample was exchanged into a 70% D$_2$O sodium acetate buffer using a 30 kDa molecular weight cutoff concentrator (Vivaspin) to reduce the incoherent background arising from hydrogen atoms [16], and subsequently spin-filtered with a 0.2 μm nylon membrane filter (VWR). The protein concentration of the sample was 2 mg mL$^{-1}$. Buffer and protein samples were measured on the Extended Q-range SANS beamline at the Spallation Neutron Source, Oak Ridge National Laboratory [31]. Samples were measured in 1 mm banjo quartz cuvettes (Hellma USA, Plainville, NY) at 18 °C. A detector distance of 4 m with wavelength bands of 2.5–6.1 Å and 9.4–13.4 Å provided a $q$ range of 0.005–0.2 Å$^{-1}$. Data were reduced and corrected for transmission, dark current, detector sensitivity and sample background using MantidPlot [32]. Guinier and distance distribution plots were analyzed in Primus and GNOM [33, 34]. Glycans were added to the crystallographic structure of $Nc$CDHIIA using CHARMM-GUI [35]. $N$-linked glycans were modeled with 2 $N$-acetylglucosamines and 8 mannoses at glycosylation sites identified from the crystal structure PDB 4QI7. For $O$-linked glycosylation, 4 mannoses were attached at sites based on the crystal
structure. The glycan patterns were chosen based on the core structures produced by *P. pastoris* [36]. Modeling was performed using SASSIE [25].

### 3.2.6 Acknowledgements

Protein expression and purification experiments were conducted at the Center for Structural Molecular Biology, a DOE BER User Facility. Diffraction data were collected at CG-4D IMAGINE (NSF MRI 09229719) at the High Flux Isotope Reactor, DOE BES User Facilities. Small-angle scattering data were collected at EQ-SANS at the Spallation Neutron Source, DOE BES User facility. WBO and AMB acknowledge student support from NSF IGERT 1069091 and DOE BES under the Graduate Opportunities (GO!) program at ORNL. FM acknowledges support from USDA NIFA Hatch 1010523.
3.2.7 REFERENCES


Chapter 4: Prospectus of future research

4.1 Protonation states of an activated LPMO copper–oxygen complex

As discussed in Chapter 3, we have determined and reported three new mechanistically-relevant crystal structures of NcLPMO9D under different conditions and with different information content. The resting state structure from X-ray diffraction at 100 K (PDB 5TKG) identified molecular O$_2$ in a previously unknown “pre-binding” site adjacent to the active site Cu ion. The activated state structure from X-ray diffraction at 100 K (PDB 5TKH) confirmed the site of O$_2$ activation within the Cu coordination sphere with the first direct structural observation of a coordinated LPMO Cu–O$_2$$^2$- complex. The resting state structure jointly refined against room temperature X-ray and neutron diffraction collected on the same crystal (PDB 5TKI) describes unambiguously the conformation and protonation state of His157 and prompted our structurally calibrated DFT study that predicted double protonation of His157 to increase the thermodynamic stability of molecular O$_2$ occupying the “pre-binding” site. Each of these structures provides new insight into LPMO oxygen activation. However, determining a structure of the LPMO copper–oxygen activated state using the complementary information provided by both X-ray and neutron diffraction would yield an integrated perspective on this important intermediate in the LPMO reaction. To that end, we have collected neutron and X-ray diffraction on an ascorbate treated crystal of NcLPMO9D under cryocooled conditions and are currently refining a structural model.

The joint X-ray/neutron cryocooled diffraction experiment was performed using the same single crystal of natural isotopic abundance NcLPMO9D as described in Chapter 2. After allowing the crystal to undergo an additional 90 days of $^1$H/$^2$H exchange by vapor diffusion, the crystal was removed from the quartz capillary and soaked for two hours in an artificial $^2$H$_2$O-formulated mother liquor containing an increased concentration of PEG 3350 (25% wt./vol.) and 100 mM ascorbic acid. After the ascorbate soak, the crystal was retrieved onto a mounted loop and flash frozen by plunging into liquid nitrogen. While keeping the crystal under liquid nitrogen using a cryovial, the mount was attached to the fixed $\chi$ angle cryo-
Figure 4.1: Laue photograph of ascorbate-treated NcLPMO9D crystal at 20 K. Diffraction intensities were accumulated for one hour using incident neutrons having $2.8 \, \text{Å} < \lambda < 10.0 \, \text{Å}$ at the IMAGINE diffractometer (CG-4D). The image is scaled to ~18% of the original size of the IMAGINE detector (1200 x 450 mm).
goniometer of the IMAGINE diffractometer. This goniometer is in thermal contact with the cold finger of a closed cycle refrigerator which was operating to maintain a sample temperature of 20 K. The cryovial was removed, and the goniometer was sealed with two thin-walled aluminum cans to permit evacuating the space around the crystal to high vacuum and to minimize heat transfer to the sample from the environment. Nineteen still quasi-Laue frames were recorded with twenty hours of exposure and 10° φ rotation between each. The cryogoniometer was then opened briefly to move the mounted crystal 15° along the χ circle and resealed. An additional eight quasi-Laue frames were recorded with the same exposure and φ rotation step. Data were integrated, wavelength normalized, scaled and merged as described for the room temperature neutron diffraction in 2.4.4. Figure 4.1 shows a Laue photograph from one hour of exposure to 2.8 Å < λ <10 Å neutrons with the crystal in the χ = 15° setting. After neutron diffraction measurement, the crystal was transferred under liquid nitrogen to a copper rotating anode X-ray diffractometer equipped with a cryostream operating at 100 K. X-ray diffraction was measured through 270° φ rotation with five minute exposures. Table 4.1 summarizes the data collection parameters and dataset processing statistics for the X-ray and neutron measurements. At present, joint refinement of a model against both the X-ray and neutron data is underway.
Table 4.1: Data collection and processing statistics.

<table>
<thead>
<tr>
<th></th>
<th>X-ray data</th>
<th>Neutron data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffractometer</td>
<td>RIGAKU</td>
<td>IMAGINE (HFIR)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54</td>
<td>2.85–4.5</td>
</tr>
<tr>
<td>Temperature (K)</td>
<td>100</td>
<td>20</td>
</tr>
<tr>
<td>Detector</td>
<td>Rigaku RAXIS IV++</td>
<td>Arinax image plate</td>
</tr>
<tr>
<td>Crystal–detector</td>
<td>90</td>
<td>199</td>
</tr>
<tr>
<td>distance (mm)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rotation range per</td>
<td>1.2</td>
<td>0b</td>
</tr>
<tr>
<td>image (°)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total rotation range</td>
<td>0-150, 222-270a</td>
<td>190, 80c</td>
</tr>
<tr>
<td>(°)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of exposures</td>
<td>165</td>
<td>27</td>
</tr>
<tr>
<td>Exposure time per</td>
<td>300</td>
<td>72000</td>
</tr>
<tr>
<td>image (s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>( P_{21} )</td>
<td>( P_{21} )</td>
</tr>
<tr>
<td>( a, b, c (\text{Å}) )</td>
<td>68.12, 42.23, 70.29</td>
<td>68.12, 42.23, 70.29</td>
</tr>
<tr>
<td>( \alpha, \beta, \gamma (°) )</td>
<td>90.00 98.33 90.00</td>
<td>90.00 98.33 90.00</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>36.11–1.50 (1.55–1.50)</td>
<td>16.12–2.50 (2.59–2.50)</td>
</tr>
<tr>
<td>Total No. of</td>
<td>224095 (20772)</td>
<td>16897 (1123)</td>
</tr>
<tr>
<td>reflections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of unique</td>
<td>58482 (5409)</td>
<td>9591 (686)</td>
</tr>
<tr>
<td>reflections</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>91.8 (85.8)</td>
<td>71.2 (51.9)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>3.8 (3.8)</td>
<td>1.8 (1.6)</td>
</tr>
<tr>
<td>( \langle I/\sigma(I) \rangle )</td>
<td>13.8 (4.2)</td>
<td>4.0 (1.8)</td>
</tr>
<tr>
<td>( R_{\text{merge}} )</td>
<td>0.054 (0.341)</td>
<td>0.140 (0.275)</td>
</tr>
</tbody>
</table>

Values for the highest resolution shell are given in parentheses.

\( ^a \) Data were collected continuously from 0–270°, but only frames from listed rotations were included in processing.

\( ^b \) Data were collected as still frames with phi rotations of 10° after each frame.

\( ^c \) After collecting 190° of data (19 frames) the crystal was reoriented and then measured for an additional 80° of data (8 frames).

4.2 Influences of the fungal LPMO secondary coordination sphere

The X-ray and neutron protein crystal structures of \( \text{NcLPMO9D} \) described in 3.1 and 3.2 demonstrate an important contribution to fungal LPMO reactivity arising from residues in the “secondary” Cu coordination sphere through shaping the molecular \( \text{O}_2 \) “pre-binding” site. The roles of His157 conformational and tautomeric states were examined in detail, and proximity and high conservation both suggest importance for Gln166 at this site as well. In a report immediately succeeding the communication of our \( \text{NcLPMO9D} \) structures, Span et al.
confirmed the functional importance of the equivalent His and Gln residues in *Myceliophthora thermophila* (or *Thermothelomyces thermophila*) PMO-3* [1]. (The authors use the designation *MtPMO-3* for this AA9 protein classified as Type 3* due to sequence similarity with Type 3 AA9 sequences but regiospecificity for monooxygenation only at C1'. This designation does not yet have an accepted translation into the standardized LPMO enzyme nomenclature.)

Span *et al.* generated a set of single AA mutants changing either His161 or Gln167 (*MtPMO3* numbering) and compared the activities of these mutants with that of the wild-type enzyme in separate assays measuring oxidized product accumulation from PASC substrate, oxygen consumption or H$_2$O$_2$ production in absence of substrate [1]. Each tested mutant of His161 (H161A, H161E and H161Q) and Gln167 (Q167A and Q167E) showed reduced activity by all three metrics which functionally demonstrates that His161 and Gln167 contribute to fungal LPMO reactivity. From the results of substrate-free assays, Span *et al.* conclude that His161 in *MtPMO3* may stabilize the Cu–O$_2^-$ intermediate and/or donate a proton for the formation of Cu–OOH$^-$ in the substrate uncoupled reaction to produce H$_2$O$_2$. *MtPMO3* His161 mutants showed increased H$_2$O$_2$ production when the substrate-free assays were supplemented with a superoxide dismutase. The effect of superoxide dismutase addition implies that free superoxide accumulates instead of H$_2$O$_2$ when this histidine residue is absent. The conformation and tautomeric state of His157 observed in our room-temperature neutron protein crystal structure of resting-state *NcLPMO9D* and the equatorial coordination of the activated copper–oxygen complex observed in our 100 K X-ray structure agree well with these conclusions [2]. In our neutron structure at pD 6.0, *NcPMO9D* His157 is singly protonated at N$_\varepsilon$ with this N–H group pointed toward the equatorial Cu$^{2+}$ coordination site which would position this proton in proximity to the distal oxygen of a Cu–O$_2^-$ species.

In contrast, Span *et al.* did not consider molecular O$_2$ pre-binding adjacent to His161 and Gln167 like what we observed for *NcLPMO9D* in their proposed explanations for the further mechanistic importance of these residues but instead considered possible effects from hydrogen bonding interactions. His161 and Gln167 participate in an extended hydrogen bonding network connected to the active site copper ion *via* a hydrogen bond between the
carbonyl of the Gln167 amide group and the hydroxy group of the axially coordinated Tyr169. The authors consider that this network may produce subtle tuning of the Cu$^{2+}$ ion electronic structure to promote oxygen activation. EPR spectroscopy experiments demonstrated that Gln167–Tyr169 hydrogen bonding does affect the Cu$^{2+}$ electronic structure as the Q167A mutation induced expected changes in the Cu$^{2+}$ EPR spectrum that were returned to the wild-type state by the Q167E mutant which can restore the hydrogen bond to Tyr169. Hydrogen bonding effects on oxygen activation are also consistent with the observation that each His161 and Gln167 mutant diminished but did not abolish MtPMO3* activity if it is assumed that, particularly in the case of H161A and Q167A, the space vacated by the truncated sidechains was filled by water molecules that reconnected the hydrogen bonding network and permitted some activity. X-ray and neutron protein crystal structures of equivalent His-to-Ala and Gln-to-Ala mutants could provide vital information regarding both potential changes to the hydrogen bonding network and to the occupancy of molecular oxygen in the pre-binding site. In addition, DFT calculations using active site models based on these structures could predict the consequences of observed structural, occupancy and protonation state changes to the electronic structure of the Cu$^{2+}$ ion.

4.3 The structure of an LPMO, substrate and activated oxygen ternary complex

The structures we have reported for resting-state and activated NcLPMO9D provided new insights into oxygen activation by fungal LPMOs [2, 3]. However, the mechanistic conclusions directly supported by these structures is limited to substrate-uncoupled production of H$_2$O$_2$ which is expected to share the same initial steps for oxygen activation. Direct conclusions on the reaction with substrate will require crystallographic characterization using both X-rays and neutrons of a ternary complex of LPMO, substrate and an activated copper–oxygen species. Such structures could permit unambiguous identification of the copper–oxygen intermediate responsible for substrate hydrogen abstraction and reveal protonation changes occurring during the reaction that could be correlated with charge and oxidation states.
The structures of substrate-free \textit{Ls}LPMO9A and \textit{Ls}LPMO9A complexes with the soluble substrates cellotriose and cellohexaose reported by Frandsen \textit{et al.} demonstrate a potential for observing an LPMO ternary complex if the chosen LPMO shows enzymatic activity on soluble substrates [4]. (LPMOs showing activity on soluble substrates were unknown until early 2014 and uncharacterized crystallographically until the report by Frandsen \textit{et al.} in mid-2016 [4, 5].) Comparisons between the substrate-free and substrate-bound structures also underscore the necessity of characterizing oxygen activation in the ternary complex as substrate binding induced structural changes in the \textit{Ls}LPMO9A active site. In substrate bound structures, the copper ion was found to be 0.2 Å closer to the axially-coordinated Tyr164 hydroxy group which alters the Cu$^{2+}$ electronic structure as observed by EPR spectroscopy. Also, substrate binding appears to cause displacement of the water molecule axially-coordinated to Cu$^{2+}$ in the resting state due to steric exclusion by the proximity of a substrate C6 hydroxymethyl group to the axial coordination site. These substrate-induced changes could be consequential to formation and/or stabilization of activated copper–oxygen species and warrant attempts to determine structures of an LPMO ternary complex.

\section{4.4 Oxygen activation by class AA10, AA11 and AA13 LPMOs}

The studies detailed herein focused on the enzyme \textit{Nc}LPMO9D as a representative of fungal, cellulose-active AA9 LPMOs. As discussed at various points in 1.1, fungal enzymes for cellulose deconstruction are of intense research and industrial interest due to the vital roles in saccharification that these enzymes fulfill in modern biorefineries [6]. Therefore, studies of AA9 LPMOs are perhaps the most immediately consequential to global interests and initiatives for augmenting supplies of fuels and products historically derived from petroleum with renewable bio-derived fuels and products. However, mechanistic investigations of LPMOs from classes AA10, AA11 and AA13 warrant pursuit both to characterize any differences that may exist when compared with AA9 LPMOs and to assess implications for the functions of
AA10–11,13 LPMOs. A real potential exists for benefits from AA10–11, 13 LPMOs utilizing other carbohydrate polymers including hemicellulose, xylan, chitin and starch. Also, characterizing AA10 LPMOs expressed by pathogenic bacteria including Vibrio cholerae and Listeria monocytogenes could have impacts for human health [7, 8].

Bacik et al. have recently reported X-ray and neutron diffraction studies of an AA10 LPMO from the bacterium Jonesia denitrificans (JdLPMO10A) that strongly parallel our studies of NcLPMO9D [9]. The JdLPMO10A model refined against 1.10 Å X-ray data collected at room temperature and under X-ray dose-limited conditions includes peroxide species coordinated “equatorially” to the active site Cu ions of both non-crystallographically symmetric protein molecules. (Since this AA10 protein, like others discussed in 1.1.4, exhibits a trigonal bipyramidal coordination geometry the coordination site equivalent to the AA9 equatorial site is not strictly “equatorial” in geometry.) Interestingly, this activated species was observed to be highly occupied and ordered, particularly in NCS molecule B, without the need for freeze-trapping at cryogenic temperatures. Also, the species formed without the addition of a small molecule reducing agent to the crystal leaving X-ray induced photoelectrons as the only possible source of reducing equivalents. In contrast with the one Cu–OO²⁻ species observed in our activated NcLPMO9D X-ray structure, one peroxide appears to be coordinated to the Cu ion of JdLPMO10A NCS molecule A in an η2 side-on fashion. The peroxide coordinated to the Cu ion of NCS molecule B instead coordinates by an η1 end-on geometry.

The corresponding neutron protein crystal structure was refined against 2.10 Å resolution neutron diffraction measured from a crystal having undergone ¹H/²H (H/D) exchange at titratable groups. Bacik et al. do not discuss the protonation state of the Cu–OO²⁻ species observed coordinated to the Cu ion of NCS molecule B [9]. Neutron SLD 2Fo–Fc and Fo–Fc maps generated from the deposited model and structure factors (PDB 5VG1) do not show obvious density corresponding the proton (deuteron) of a Cu–OOH⁻ species. However, the authors did draw conclusions from the intensity and shape of excess neutron SLD apparent in Fo–Fc maps (contoured at σ = 3.0) omitting contributions from the amino terminus NH₂ (ND₂). For NCS molecule A, the omit density was interpreted as reflecting the presence of both deuterons at the N-terminal and copper coordinating ND₂ group; however, for NCS
molecule B, the omit density was interpreted as reflecting the presence of a single deuteron “hydrogen” bonded to a nearby backbone carbonyl group forming a ND⁻ group. It has been suggested that the formation of ND⁻ at the N-terminus could result due to tautomerization of Cu(II)–O⁻ to Cu(III)–OH and stabilize this high-valency species shown to be highly reactivity for hydrogen abstraction in model complexes [10, 11]. (See 1.1.7 and Figure 1.1.6 for additional discussion of Cu(III) species in the LPMO mechanism.)

The argument for the ND⁻ species would be strengthened by determining a neutron protein crystal structure from a deuterium labeled protein crystallized under deuterium enriched conditions. It is well known that incomplete H/D exchange, particularly in instances where exchange is performed by vapor diffusion for short periods of time (14 days in the present case), can easily obscure neutron SLD for H/D atoms. Partial occupancy by H atoms having a negative coherent neutron scattering length (-3.74 fm) sums with the contribution from the positive (6.67 fm) scattering length of D atoms so that ~50-70% H and ~30-50% D at the same position can effectively cancel the contribution of the H/D atom to Fourier maps from moderate (~2 Å) resolution diffraction data [12]. We will certainly examine the data from our 20 K neutron diffraction experiment discussed in 4.1 to determine if we can provide additional information on either protonation of the Cu–OO²⁻ species or deprotonation of the amino terminus.
4.5 REFERENCES


APPENDICES
Appendix A

The following work was reprinted with permission from: W.B. O’Dell, A.M. Bodenheimer and F. Meilleur. *Archives of Biochemistry and Biophysics*. 2016. 602: 48–60.
Table A.1 Supporting Information Table S1 from “Neutron protein crystallography: A complementary tool for locating hydrogens in proteins”

<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProt</th>
<th>PDB</th>
<th>Deuterium Labeling</th>
<th>Crystallization Method</th>
<th>H/D Exchange Method</th>
<th>Crystal Volume (mm$^3$)</th>
<th>Instrument</th>
<th>Measurement Time (days)</th>
<th>Neutrons (Å)</th>
<th>X-ray (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIV-1 protease</td>
<td>P03367</td>
<td>2ZYE</td>
<td>none</td>
<td>macroseeded batch</td>
<td>soak 14 days vapor 28 days soak 240 days</td>
<td>3.6</td>
<td>BIX-4</td>
<td>46</td>
<td>1.9</td>
<td>1.4</td>
<td>[1]</td>
</tr>
<tr>
<td>myoglobin</td>
<td>P02185</td>
<td>2MB5</td>
<td>uniform</td>
<td>sitting drop</td>
<td>n/r</td>
<td>0.2</td>
<td>LADI III</td>
<td>17</td>
<td>2.0</td>
<td>2.0</td>
<td>[2]</td>
</tr>
<tr>
<td>1CQ2</td>
<td>D uniform</td>
<td></td>
<td>batch</td>
<td>macroseeded sitting drop</td>
<td>2.5</td>
<td>HFBIR*</td>
<td>not reported</td>
<td>1.8</td>
<td>n/a</td>
<td>[3, 4]</td>
<td></td>
</tr>
<tr>
<td>1L2K</td>
<td>none</td>
<td></td>
<td>batch</td>
<td>sitting drop</td>
<td>6.25</td>
<td>BIX-3</td>
<td>24</td>
<td>1.5</td>
<td>n/a</td>
<td>[6]</td>
<td></td>
</tr>
<tr>
<td>transhyretin</td>
<td>P02766</td>
<td>3U2J</td>
<td>none</td>
<td>sitting drop</td>
<td>crystallization</td>
<td>2.5</td>
<td>iBIX</td>
<td>30</td>
<td>2.0</td>
<td>n/a</td>
<td>[7]</td>
</tr>
<tr>
<td>beta lactamase</td>
<td>Q40766</td>
<td>2WyX</td>
<td>none</td>
<td>sitting drop</td>
<td>crystallization</td>
<td>3.4</td>
<td>LADI III</td>
<td>5</td>
<td>2.0</td>
<td>1.95</td>
<td>[8, 9]</td>
</tr>
<tr>
<td>4PVM</td>
<td>D uniform</td>
<td></td>
<td>batch</td>
<td>crystallization</td>
<td>3.4</td>
<td>D19</td>
<td>10</td>
<td>2.3</td>
<td>1.95</td>
<td>n/a</td>
<td>[10]</td>
</tr>
<tr>
<td>4PVN</td>
<td>D uniform</td>
<td></td>
<td>batch</td>
<td>crystallization</td>
<td>3.4</td>
<td>LADI III</td>
<td>12</td>
<td>2.1</td>
<td>n/a</td>
<td>[11]</td>
<td></td>
</tr>
<tr>
<td>type III antifreeze protein</td>
<td>P19614</td>
<td>3QF6</td>
<td>D uniform</td>
<td>crystallization</td>
<td>1.13</td>
<td>LADI III</td>
<td>21</td>
<td>1.85</td>
<td>1.05</td>
<td>[14-16]</td>
<td></td>
</tr>
<tr>
<td>rubredoxin</td>
<td>P24297</td>
<td>1VCX</td>
<td>selective H CH$_3$</td>
<td>macroseeded sitting drop</td>
<td>0.23</td>
<td>LADI III</td>
<td>30</td>
<td>1.80</td>
<td>1.05</td>
<td>[15, 17]</td>
<td></td>
</tr>
<tr>
<td>1U6</td>
<td>none</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.9</td>
<td>LADI III</td>
<td>5</td>
<td>1.68</td>
<td>1.6</td>
<td>[18]</td>
<td></td>
</tr>
<tr>
<td>3KYX</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>BIX-3</td>
<td>30</td>
<td>1.6</td>
<td>1.5</td>
<td>[19]</td>
<td></td>
</tr>
<tr>
<td>3KYY</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>4.1</td>
<td>LADI III</td>
<td>3</td>
<td>1.66</td>
<td>1.1</td>
<td>[20, 21]</td>
<td></td>
</tr>
<tr>
<td>3RZT</td>
<td>D</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.9</td>
<td>LADI III</td>
<td>0.5</td>
<td>1.75</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
<tr>
<td>3RZ6</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.9</td>
<td>LADI III</td>
<td>1.5</td>
<td>1.75</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
<tr>
<td>3SS2</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.9</td>
<td>LADI III</td>
<td>2</td>
<td>1.75</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
<tr>
<td>3RGV</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.9</td>
<td>LADI III</td>
<td>5</td>
<td>1.75</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
<tr>
<td>4AR3</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>6.9</td>
<td>D19</td>
<td>8</td>
<td>1.05</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
<tr>
<td>4AR4</td>
<td>D uniform</td>
<td></td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>2.0</td>
<td>D19</td>
<td>3</td>
<td>1.38</td>
<td>n/a</td>
<td>[22, 23]</td>
<td></td>
</tr>
</tbody>
</table>

Table A.1 Continued
<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProt</th>
<th>PDB</th>
<th>Deuterium Labeling</th>
<th>Crystallization Method</th>
<th>H/D Exchange Method</th>
<th>Crystal Volume (mm²)</th>
<th>Instrument</th>
<th>Measurement Time (days)</th>
<th>Neutron data (Å)</th>
<th>X-ray data (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rubredoxin</td>
<td>P24297</td>
<td>4K9F</td>
<td>D uniform</td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>0.7</td>
<td>IMAGINE</td>
<td>3</td>
<td>1.75</td>
<td>n/a</td>
<td>[25]</td>
</tr>
<tr>
<td>aldose reductase</td>
<td>P15121</td>
<td>2R24</td>
<td>D uniform</td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>0.15</td>
<td>LADI</td>
<td>not reported</td>
<td>2.2</td>
<td>1.75</td>
<td>[26, 27]</td>
</tr>
<tr>
<td>amicyanin</td>
<td>P22364</td>
<td>3L45</td>
<td>none</td>
<td>macroseeded hanging drop</td>
<td>crystallization</td>
<td>2.0</td>
<td>PCS</td>
<td>21</td>
<td>1.8</td>
<td>1.5</td>
<td>[28, 29]</td>
</tr>
<tr>
<td>bovine pancreatic trypsin inhibitor</td>
<td>P00974</td>
<td>5PTI</td>
<td>none</td>
<td>macroseeded sitting drop</td>
<td>soak</td>
<td>8.0</td>
<td>NIST</td>
<td>not reported</td>
<td>1.8</td>
<td>0.94</td>
<td>[30]</td>
</tr>
<tr>
<td>carbonic anhydrase</td>
<td>P00918</td>
<td>3KXX</td>
<td>none</td>
<td>sitting drop</td>
<td>crystallization</td>
<td>1.2</td>
<td>PCS</td>
<td>55</td>
<td>2.0</td>
<td>1.5</td>
<td>[31]</td>
</tr>
<tr>
<td>TMJ</td>
<td>P02933</td>
<td>4Q9C</td>
<td>none</td>
<td>sitting drop</td>
<td>vapor 90 days</td>
<td>1.7</td>
<td>PCS</td>
<td>20</td>
<td>2.0</td>
<td>1.65</td>
<td>[32]</td>
</tr>
<tr>
<td>concanavalin A</td>
<td>P02866</td>
<td>1C57</td>
<td>none</td>
<td>dialysis</td>
<td>crystallization</td>
<td>15.0</td>
<td>LADI</td>
<td>20</td>
<td>2.4</td>
<td>1.7</td>
<td>[35, 36]</td>
</tr>
<tr>
<td>1XQN</td>
<td>P22034</td>
<td>2Y24</td>
<td>none</td>
<td>dialysis</td>
<td>crystallization</td>
<td>5.6, 1.6</td>
<td>LADI III</td>
<td>34</td>
<td>2.5</td>
<td>n/a</td>
<td>[37]</td>
</tr>
<tr>
<td>crambin</td>
<td>P01542</td>
<td>4FC1</td>
<td>none</td>
<td>vapor diffusion</td>
<td>vapor 30 days</td>
<td>21.0</td>
<td>LADI</td>
<td>15</td>
<td>2.2</td>
<td>n/a</td>
<td>[38]</td>
</tr>
<tr>
<td>cytochrome c peroxidase</td>
<td>P00431</td>
<td>4CV1</td>
<td>none</td>
<td>dialysis</td>
<td>crystallization</td>
<td>1.0</td>
<td>LADI III</td>
<td>18</td>
<td>2.4</td>
<td>2.1</td>
<td>[41]</td>
</tr>
<tr>
<td>dihydrofolate reductase</td>
<td>P0ABQ4</td>
<td>2INQ</td>
<td>none</td>
<td>sitting drop</td>
<td>crystallization</td>
<td>0.72</td>
<td>PCS</td>
<td>23</td>
<td>2.17</td>
<td>n/a</td>
<td>[42]</td>
</tr>
<tr>
<td>diisopropylfluorophosphatase</td>
<td>Q7SIG4</td>
<td>3DYC</td>
<td>none</td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.6</td>
<td>IMAGINE</td>
<td>17</td>
<td>2.0</td>
<td>1.6</td>
<td>[43, 44]</td>
</tr>
<tr>
<td>dissimilatory sulfite reductase</td>
<td>Q46582</td>
<td>1WQ2</td>
<td>none</td>
<td>hanging drop</td>
<td>crystallization</td>
<td>1.7</td>
<td>BIX-3</td>
<td>70</td>
<td>2.4</td>
<td>n/a</td>
<td>[47]</td>
</tr>
<tr>
<td>elastase</td>
<td>P00772</td>
<td>31GN</td>
<td>none</td>
<td>macroseeded sitting drop</td>
<td>crystallization</td>
<td>3.3</td>
<td>BIX-3</td>
<td>93</td>
<td>1.65</td>
<td>1.2</td>
<td>[48]</td>
</tr>
<tr>
<td>endothiapepin</td>
<td>P11838</td>
<td>1GKT</td>
<td>none</td>
<td>batch</td>
<td>vapor 5.5, 3.0</td>
<td>LADI</td>
<td>not reported</td>
<td>2.1</td>
<td>n/a</td>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td>hemoglobin</td>
<td>P69905</td>
<td>2DXM</td>
<td>none</td>
<td>batch</td>
<td>crystallization</td>
<td>36.0</td>
<td>BIX-3</td>
<td>120</td>
<td>2.1</td>
<td>n/a</td>
<td>[52]</td>
</tr>
<tr>
<td>Protein</td>
<td>UniProt</td>
<td>PDB</td>
<td>Deuterium Labeling</td>
<td>Crystallization Method</td>
<td>H/D Exchange Method</td>
<td>Crystal Volume (mm³)</td>
<td>Instrument</td>
<td>Measurement Time (days)</td>
<td>Neutron dose (Å)</td>
<td>X-ray dose (Å)</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>------------</td>
<td>------------------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>hemoglobin</td>
<td>P69905</td>
<td>3KMF</td>
<td>none</td>
<td>batch</td>
<td>crystallization</td>
<td>20.0</td>
<td>PCS</td>
<td>18</td>
<td>2.0</td>
<td>n/a</td>
<td>[53]</td>
</tr>
<tr>
<td>hen egg white lysozyme</td>
<td>P00698</td>
<td>1IOS</td>
<td>none</td>
<td>concentration gradient</td>
<td>crystallization</td>
<td>6.0</td>
<td>LADI</td>
<td>10</td>
<td>2.0</td>
<td>n/a</td>
<td>[54]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>batch</td>
<td>vapor 180 days</td>
<td>2.0</td>
<td>LADI</td>
<td>14</td>
<td>1.7</td>
<td>n/a</td>
<td>[55]</td>
</tr>
<tr>
<td>human lysozyme</td>
<td>P61626</td>
<td>2ZW B</td>
<td>none</td>
<td>batch</td>
<td>crystallization</td>
<td>not reported</td>
<td>BIX-3</td>
<td>not reported</td>
<td>1.8</td>
<td>n/a</td>
<td>unpublished</td>
</tr>
<tr>
<td>inorganic pyrophosphatase</td>
<td>H21L23 L</td>
<td>3Q3L</td>
<td>none</td>
<td>counter diffusion</td>
<td>vapor 30 days soak</td>
<td>5.0</td>
<td>LADI III</td>
<td>21</td>
<td>2.5</td>
<td>n/a</td>
<td>[56]</td>
</tr>
<tr>
<td>insulin</td>
<td>P01315</td>
<td>3INS</td>
<td>none</td>
<td>batch</td>
<td>crystallization</td>
<td>126 days soak 2.7</td>
<td>BIX-4</td>
<td>19</td>
<td>2.7</td>
<td>n/a</td>
<td>[57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>batch</td>
<td>crystallization</td>
<td>20.8</td>
<td>BIX-3</td>
<td>12</td>
<td>2.3</td>
<td>n/a</td>
<td>[58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>batch</td>
<td>crystallization</td>
<td>30 days soak 2.0</td>
<td>BIX-4</td>
<td>41</td>
<td>2.11</td>
<td>n/a</td>
<td>[59]</td>
</tr>
<tr>
<td>photoactive yellow protein</td>
<td>P16113</td>
<td>2Z0 I</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>1.9</td>
<td>BIX-4</td>
<td>45</td>
<td>1.5</td>
<td>1.25</td>
<td>[60]</td>
</tr>
<tr>
<td>protease I</td>
<td>P15636</td>
<td>4GP G</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>1.0</td>
<td>BIX-4</td>
<td>69</td>
<td>1.98</td>
<td>1.9</td>
<td>[61]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>0.95</td>
<td>LADI III</td>
<td>17</td>
<td>2.2</td>
<td>1.76</td>
<td>[62]</td>
</tr>
<tr>
<td>phycocyanobilin:ferredoxin oxidoreductase</td>
<td>Q13976</td>
<td>4QX K</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>2.7</td>
<td>IBIX</td>
<td>9</td>
<td>1.95</td>
<td>1.55</td>
<td>[63]</td>
</tr>
<tr>
<td>ribonuclease A</td>
<td>P61823</td>
<td>5RSA</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>25.0 30.0 21 days 2.7</td>
<td>NIST</td>
<td>not reported</td>
<td>2.0</td>
<td>2.0</td>
<td>[64]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>30.0</td>
<td>NIST</td>
<td>not reported</td>
<td>2.0</td>
<td>2.0</td>
<td>[65-67]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>60 days 14.0</td>
<td>BIX-4</td>
<td>25</td>
<td>1.7</td>
<td>n/a</td>
<td>[66]</td>
</tr>
<tr>
<td>α-thrombin/bivalirudin</td>
<td>P00734</td>
<td>5YX F</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>8.0</td>
<td>BIX-4</td>
<td>48</td>
<td>2.75</td>
<td>1.6</td>
<td>[67]</td>
</tr>
<tr>
<td>complex</td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>365 days 5.0</td>
<td>HFBR</td>
<td>not reported</td>
<td>1.8</td>
<td>n/a</td>
<td>[68]</td>
</tr>
<tr>
<td>trypsin</td>
<td>P00760</td>
<td>NTP</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>14 days 2.2</td>
<td>BIX-4</td>
<td>46</td>
<td>2.15</td>
<td>1.6</td>
<td>[69]</td>
</tr>
<tr>
<td>trypsin:bovine pancreatic</td>
<td>P00760</td>
<td>3OT J</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>60 days 4.0</td>
<td>LADI III</td>
<td>6</td>
<td>1.9</td>
<td>1.92</td>
<td>[70]</td>
</tr>
<tr>
<td>trypsin inhibitor</td>
<td>P00760</td>
<td>3OT J</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>60 days 4.0</td>
<td>LADI III</td>
<td>25</td>
<td>2.3</td>
<td>2.0</td>
<td>[71]</td>
</tr>
<tr>
<td>urate oxidase</td>
<td>Q00511</td>
<td>5N3M</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>LADI III</td>
<td>6</td>
<td>1.9</td>
<td>1.92</td>
<td>[72]</td>
</tr>
<tr>
<td>xylanase</td>
<td>P36271</td>
<td>4S2 D</td>
<td>none</td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>PCS</td>
<td>not reported</td>
<td>2.0</td>
<td>1.6</td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>MaNDI</td>
<td>13</td>
<td>2.0</td>
<td>1.7</td>
<td>[74]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>BIODIFF</td>
<td>19</td>
<td>2.0</td>
<td>1.6</td>
<td>[75]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>microseeded hanging drop</td>
<td>crystallization</td>
<td>4.0</td>
<td>PCS</td>
<td>not reported</td>
<td>1.7</td>
<td>1.6</td>
<td>[76]</td>
</tr>
</tbody>
</table>

Table A.1 Continued
<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProt</th>
<th>DR</th>
<th>Deuterium Labeling</th>
<th>Crystallization Method</th>
<th>H/D Exchange Method</th>
<th>Crystal Volume (mm³)</th>
<th>Instrument</th>
<th>Measurement Time (days)</th>
<th>Neutron dspacing (Å)</th>
<th>X-ray dspacing (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>xylose isomerase</td>
<td>P24300</td>
<td>4LN</td>
<td>none</td>
<td>batch</td>
<td>counter soak</td>
<td>8.0</td>
<td>LADI</td>
<td>27</td>
<td>2.2</td>
<td>1.84</td>
<td>[77-79]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20VE</td>
<td>none</td>
<td>diffusion</td>
<td>counter soak</td>
<td>8.0</td>
<td>PCS</td>
<td>not reported</td>
<td>1.8</td>
<td>1.60</td>
<td>[80, 81]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3CWH</td>
<td>none</td>
<td>diffusion</td>
<td>soak</td>
<td>4.0, 6.0</td>
<td>PCS</td>
<td>36</td>
<td>2.2</td>
<td>n/a</td>
<td>[80, 82]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3KCL</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>50.0</td>
<td>D19</td>
<td>14</td>
<td>2.0</td>
<td>2.0</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3KCO</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>50.0</td>
<td>PCS</td>
<td>14</td>
<td>1.8</td>
<td>1.53</td>
<td>[83]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3XCI</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>28.0</td>
<td>D19</td>
<td>14</td>
<td>2.0</td>
<td>2.0</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3QZA</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>9.0</td>
<td>PCS</td>
<td>14</td>
<td>2.0</td>
<td>1.7</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4DVO</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>50.0</td>
<td>D19</td>
<td>14</td>
<td>2.0</td>
<td>1.55</td>
<td>[85]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4QDP</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>10.0</td>
<td>PCS</td>
<td>14</td>
<td>2.0</td>
<td>1.6</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4QDW</td>
<td>none</td>
<td>batch</td>
<td>vapor</td>
<td>10.0</td>
<td>D19</td>
<td>14</td>
<td>1.8</td>
<td>1.6</td>
<td>[86]</td>
</tr>
</tbody>
</table>

(a) Data collected at the High Flux Beam Reactor, Brookhaven National Laboratory which is now decommissioned.
(b) X-ray data was not used for joint model refinement.
(c) Data collected at the Neutron Beam Split Core Reactor, National Institute of Standards and Technology which no longer operates a macromolecular single crystal diffractometer.

Table S1. Comprehensive list of neutron and joint x-ray/neutron structures of protein deposited in the Protein Database. Protein deuterium labelled during expression are listed first. The non labelled protein are then listed in alphabetical order.
REFERENCES


