ABSTRACT

BODENHEIMER, ANNETTE MARIE. Structural Characterization of the Fungal Cellulose Degrading Enzymes Cel7A and Celllobiose Dehydrogenase IIA. (Under the direction of Flora Meilleur).

Lignocellulose obtained from biomass is highly recalcitrant, which impedes enzymatic degradation for downstream conversion to biofuels and value-added products. Fungal enzymes secreted in the presence of cellulose are currently utilized throughout industry for cellulose depolymerization. These proteins work synergistically to breakdown cellulose. Cellulase cocktails are industrially optimized through enzymatic engineering, adjusting ratios of enzymes, or mixing cellulases from different organisms. In this dissertation, a structural approach was taken to better understand carbohydrate active enzymes with high flexibility and to elucidate mechanistic details at the atomic and molecular levels to help guide future improvements.

One of the hydrolytic enzymes involved in depolymerization of cellulose is Cel7A. This cellulase loads onto the reducing-end of a cellodextrin strand and then processively hydrolyzes the strand into the disaccharide celllobiose. A single cellulose chain separated from the insoluble substrate is threatened into a tunnel within its catalytic domain for hydrolysis to occur. This active site tunnel is enclosed by eight loops. Crystallization of the full-length enzyme revealed two conformational states of the active site tunnel: an open state and a closed state. Crystallographic and molecular dynamics simulation analysis revealed that the catalytic domain loops synchronously move to promote unidirectional motion.

The auxiliary activity cellulase celllobiose dehydrogenase (CDH) has recently garnered attention due to newly characterized oxidative processes in cellulose degradation and its potential versatility. CDH oxidizes soluble sugars at the dehydrodegenase domain and
transfers the cytochrome domain. The cytochrome domain passes the electrons to the
electron acceptor lytic polysaccharide monooxygenase (LPMO). Together, CDH and LPMO
enhance the efficiency of the hydrolytic enzyme cocktail. A flexible linker connects the
cytochrome and dehydrogenase domains. The interdomain electron transfer (IET) rate is
affected by pH and the presence of cations. Structural characterization of cellobiose
dehydrogenase IIA (CDHIIA) from *Neurospora crassa* and *Myriococcum thermophilum*
using small angle scattering was conducted to probe CDHIIA conformational alteration in
various chemical environments. Small angle scattering (SAS) revealed that *Nc*CDHIIA
undergoes conformational changes when pH is shifted away from the optimal activity pH.
Modeling revealed that a closed conformation of *Nc*CDHIIA is favored at the optimal pH
promoting electron transfer between the cytochrome and dehydrogenase domains.

*Mt*CDHIIA structurally rearranges in the presence of cations, especially Ca$^{2+}$. Modeling
indicates that CDH is capable of adopting a more compact form in the presence of cations
that supports efficient transfer of electrons between the two domains.

Little is understood about the interaction between CDH and the electron acceptor
LPMO due to the transient nature of this redox complex. Small angle neutron scattering
(SANS) is uniquely suited to probe how LPMO elicits conformational changes within CDH
upon binding. This study unequivocally showed that the presence of LPMO induced
conformational changes within CDH leading to a compact conformer. These SANS studies
on CDH led to the conclusion that interaction with LPMO promotes the dehydrogenase and
cytochrome domains to come closer together allowing for the efficient transfer of electrons
between the two domains and subsequently to the LPMO.
Structural Characterization of the Fungal Cellulose Degrading Enzymes Cel7A and Cellobiose Dehydrogenase IIA

by
Annette Marie Bodenheimer

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

2016

APPROVED BY:

_______________________________  ______________________________
Dr. Flora Meilleur                  Dr. Robert Rose
Committee Chair

_______________________________  ______________________________
Dr. Gouzhou Xu                    Dr. Robert Kelly
DEDICATION

To John, Dad, Mom, and Thömaschen.
BIOGRAPHY

Annette Bodenheimer was born in Virginia to David and Rita Bodenheimer and has a younger brother, Tom. Her father is a government contract lawyer who still practices with the enthusiasm and excitement of a summer intern. David has shown Annette the importance of loving what you do and what true discipline looks like. Her mother worked in various branches of the government and is now happily retired. Rita managed to be both successful during her career and attend her kids’ recitals, sporting events, graduations (even the seemingly arbitrary ones); Annette is forever grateful for having this patient soul show her the importance of work-life balance. Tom now lives in Palo Alto, CA doing what she can only describe as marketing. Annette is fortunate to have a brother that taught her not to take life too seriously and can always make her laugh.

She went to North Carolina State University for her undergraduate studies and graduated in 2009. Upon graduating, she took care of her maternal grandparents and found a job performing histocompatibility testing for organ and bone marrow transplants. There was a sense of satisfaction with her job knowing that indirectly she was helping people, but knew that this was not where she would be happiest. Annette applied and was accepted to North Carolina State University’s PhD program. After completing her rotation in Dr. Flora Meilleur’s laboratory, she was fascinated by the structural biology work Dr. Meilleur was doing at Oak Ridge National Laboratory and joined her lab. In 2015, Annette married John Allen, the man who was the catalyst to her applying to graduate school. Annette will be graduating with her PhD in December 2016 with the hopes of continuing to answer biological questions through structural techniques.
ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Flora Meilleur for imparting her knowledge of crystallography and neutrons on me as well as allowing me to venture off into unfamiliar territory. We have had a not so typical PI-student arrangement and I cannot thank her enough for her trust and willingness to let me work on the majority of my PhD in a different state. She has helped me through stressful times and celebrated the joyful moments with me as well. Thank you.

To my committee members, Drs. Robert Rose, Robert Kelly, Guozhou Xu, and Clay Clark, your advice, encouragement, and support have been an immense help over the years. I would also like to thank Drs. Paul Swartz and Matthew Cuneo for mentoring me, sharing your wealth of knowledge with me, and random conversations to regain the little bit of sanity I had at times.

I would not have succeeded without the love and moral support from my family. John Allen, you will never know how much I appreciate all the times you were there for me when I felt run down and being right beside me to celebrate the joyous moments. I love you darling! To my parents, I appreciate your constant love and support throughout my educational endeavors. Tom, you help remind me of what is important in life and our conversations leave me feeling recharged. To the rest of my family, thank you for the various words of wisdom over the years, helpfulness, delicious meals, and wonderful times when we are able to get together and cut up. I love y’all.

To my friends, the family that I choose, your unwavering patience and kindness has meant the world to me. Thank you to Matthew Cuneo, Milka Vasquez, and Winston Joseph.
Cuneo for opening up your home to me anytime I needed to do experiments in TN and making life enjoyable during exhausting weeks. To my gym families at Forged Fitness and CrossFit Knoxville, I appreciate your patience with my strange schedule and always making me laugh even when I am riding the struggle bus. Thank you to my lab mates and the graduate students I have gotten to know throughout my time at NCSU and ORNL; the silliness, coffee and Mitch trips, and hallway conversations will be treasured memories. To Dr. Volker Urban and the Biological Soft Matter Division at Oak Ridge National Laboratory, thank you for the constant insightful conversations and welcoming atmosphere.
TABLE OF CONTENTS

LIST OF TABLES ....................................................................................................................... ix
LIST OF FIGURES .................................................................................................................... x

CHAPTER 1: Introduction to Structural Studies of Fungal Cellulases ...................................... 1
  1.1 Enzymatic Degradation of Cellulose .................................................................................. 1
     1.1.1 Introduction .................................................................................................................... 1
     1.1.2 Cellulose ...................................................................................................................... 2
     1.1.3 Cellulases ..................................................................................................................... 5
     1.1.4 Cel7A ............................................................................................................................ 10
     1.1.5 CDH11A ....................................................................................................................... 17
  1.2 Neutrons in Biology ........................................................................................................... 22
     1.2.1 Neutron Crystallography ............................................................................................ 22
     1.2.2 Small Angle Neutron Scattering (SANS) .................................................................... 25
  1.3 Research Outline .............................................................................................................. 30
  1.4 References ......................................................................................................................... 31

CHAPTER 2: Crystal Structures of Wild-type *Trichoderma reesei* Cel7A Catalytic Domain in Open and Closed Sates ............................................................... 44
  Part I – Crystallization and Preliminary X-ray Diffraction Analysis of *Hypocrea jecorina* Cel7A in Two New Crystal Forms ................................................................. 45
  2.1a Abstract .......................................................................................................................... 45
  2.2a Introduction ..................................................................................................................... 46
  2.3a Materials and Methods ................................................................................................... 47
     2.3.1a *Hj*Cel7A purification ............................................................................................... 47
     2.3.2a Crystal growth .......................................................................................................... 48
     2.3.3a Data collection and refinement ............................................................................... 49
  2.4a Results & Discussion ....................................................................................................... 51
  2.5a References ....................................................................................................................... 53
  Part II – Crystal Structures of Wild-type *Trichoderma reesei* Cel7A Catalytic Domain in Open and Closed Sates ................................................................. 58
  2.1b Abstract .......................................................................................................................... 58
  2.2b Introduction ..................................................................................................................... 58
  2.3b Materials and Methods ................................................................................................... 62
     2.3.1b Crystal structures ...................................................................................................... 62
     2.3.2b Molecular dynamics analysis .................................................................................... 62
  2.4b Results ........................................................................................................................... 63
     2.4.1b Product interactions ................................................................................................. 63
     2.4.2b Loop analysis .......................................................................................................... 66
  2.5b Discussion ....................................................................................................................... 71
  2.6b References ....................................................................................................................... 75
  2.7b Supplementary Information ............................................................................................ 81
CHAPTER 3: Small Angle Scattering Reveals pH and Cation Induced Conformational Changes for *Neurospora crassa* and *Myriococcus thermophilum* Cellobiose Dehydrogenase ................................................................. 83

3.1 Abstract ............................................................................................................. 83
3.2 Introduction ....................................................................................................... 84
3.3 Materials and Methods ..................................................................................... 89
   3.3.1 Protein expression and purification ............................................................... 89
   3.3.2 Small angle neutron scattering (SANS) sample preparation, data collection, and reduction ........................................................... 91
   3.3.3 Small angle X-ray scattering (SAXS) sample preparation, data collection, and reduction ........................................................... 91
   3.3.4 Data evaluation and modeling ................................................................. 92
   3.3.5 Sedimentation velocity .............................................................................. 93
3.4 Results ............................................................................................................. 93
   3.4.1 pH Effect on *MtCDH* and *NcCDH* ......................................................... 94
   3.4.2 Effect of monovalent cations on CDH ...................................................... 99
   3.4.3 Effect of divalent cations on CDH .......................................................... 103
   3.4.4 Effect of cellobiose on CDH .................................................................. 109
3.5 Discussion ..................................................................................................... 110
3.6 Conclusion .................................................................................................... 114
3.7 References ..................................................................................................... 115

CHAPTER 4: Investigation of the *Neurospora crassa* Cellobiose Dehydrogenase and *Neurospora crassa* Lytic Polysaccharide Monoxygenase-2 Complex ......................... 118

4.1 Abstract ......................................................................................................... 118
4.2 Introduction ................................................................................................... 119
4.3 Materials and Methods ............................................................................... 123
   4.3.1 *Neurospora crassa* Cellobiose Dehydrogenase IIA expression and purification ........................................................................... 123
   4.3.2 Partially deuterated *Neurospora crassa* lytic polysaccharide monooxygenase-2 expression and purification ........................................... 124
   4.3.3 LPMO contrast variation series ................................................................ 125
   4.3.4 CDH and LPMO contrast matching series ............................................. 125
   4.3.5 CDH crowding experiment ...................................................................... 126
   4.3.6 Determination of *NcCDH*—*NcLPMO* binding constants .............. 126
4.4 Results ....................................................................................................... 127
   4.4.1 CDL-LPMO SANS measurements ......................................................... 127
   4.4.2 CDH crowding experiment .................................................................. 130
   4.4.3 Determination of CDH-LPMO binding constants ........................... 132
4.5 Discussion .................................................................................................. 133
4.6 Conclusion .................................................................................................. 137
4.7 References .................................................................................................. 138
CHAPTER 5: Conclusions and Perspectives ................................................................. 142
  5.1 Cel7A .................................................................................................................. 142
  5.2 CDHIIA ............................................................................................................. 146
  5.3 References ....................................................................................................... 150

APPENDIX ................................................................................................................ 153
  Pichia pastoris Protocols ...................................................................................... 154
    Electroporation .................................................................................................... 154
    Phenotyping ......................................................................................................... 155
    Growth and expression ....................................................................................... 156
    Media recipes ..................................................................................................... 158
  Small angle scattering .......................................................................................... 160
    Guinier analysis ................................................................................................. 160
    Pair distribution function .................................................................................. 160
    Kratky plots ....................................................................................................... 162
  Crystallography .................................................................................................... 163
  Vapor diffusion .................................................................................................... 163
    Molecular replacement ........................................................................................ 165
  References ............................................................................................................ 168
LIST OF TABLES

CHAPTER 2

Table 2.1 – X-ray data collection statistics................................................................. 50
Table 2.2 – Refinement statistics.................................................................................. 63
Table 2.3 – X-ray data collection and refinement statistics......................................... 80

CHAPTER 3

Table 3.1 – Buffer conditions tested for MtCDH (M) and NcCDH (N)...................... 90
Table 3.2 – Effect of pH on CDH.................................................................................. 94
Table 3.3 – Effect of K⁺ on NcCDH........................................................................... 100
Table 3.4 – Effect of K⁺ on MtCDH.......................................................................... 101
Table 3.5 – Effect of Ca²⁺ on NcCDH........................................................................ 106
Table 3.6 – Effect of Ca²⁺ on MtCDH....................................................................... 107

CHAPTER 4

Table 4.1 – Dimensions of CDH in complex with LPMO. Initial represents the data analyzed as described in the methods. Sub is the standard curve subtracted by the weighted dLPMO at 70% D2O curve................................................................. 134
CHAPTER 1

Figure 1.1 – Cellulose structure from biomass. Plant cell walls are comprised of a complex mixture of lignin, hemicellulose, and cellulose. Cellulose has crystalline and amorphous regions within the microfibrils allowing for different levels of ease for degradation into monomeric glucose. (US DOE, 2005) ........................................... 3

Figure 1.2 – Model of common fungal cellulose degrading enzymes. Exoglucanases load onto the reducing ends (red) or non-reducing ends (orange) to processively hydrolyze cellulose into cellobiose. Lytic polysaccharide monooxygenase (green) and endoglucanases (purple) oxidize and hydrolyze, respectively cellulose creating new loading sites for exoglucanases. Cellobiose dehydrogenase (cyan) is the redox partner for LPMO. β-glucosidase (pink) hydrolyzes cellobiose into glucose....................... 6

Figure 1.3 – Trichoderma reesei Cel7A catalytic domain loaded with cellulose. Tunnel loops are colored: A1, pink; A2, red; A3, orange; A4, yellow; B1, green; B2, cyan; B3, blue; B4, purple. Cellononaose is labeled for clarity of subsites. (PDB ID 4C4C) 12

Figure 1.4 – Structure of Neurospora crassa cellobiose dehydrogenase IIA. Cytochrome domain, red; dehydrogenase domain, yellow; CBM, cyan; linkers and cofactors, green. (PDB ID 4QI7) ........................................................................................................... 17

Figure 1.5 – Schematic of enzymatic fuel cells. CDH, cyan, is used at the bioanode to generate electrons from the oxidation of sugar molecules (PDB ID 4QI6). An enzyme such as bilirubin oxidase, orange, is used as the biocathode to utilize the electrons to convert molecular oxygen to water (PDB ID 2XLL). ................................. 19

Figure 1.6 – 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. ........................................................................................................................................... 25

Figure 1.7 – Scattering length density as a function of solvent deuteration. Values calculated with MULCh web server.(Whitten et al., 2008)........................................... 27

Figure 1.8 – Schematic of contrast variation series. As the ratio of H:D is varied (as illustrated by a gradient from blue to green), different particles become “visible”... 28
CHAPTER 2

Figure 2.1 – SDS-PAGE analysis of purified and crystallized \textit{Hj}Cel7A. The analysis confirms the integrity of full-length Cel7A. Lanes 1 and 2, crystals; lane 3, molecular weight marker; lanes 4 and 5, purified full-length \textit{Hj}Cel7A from two different preparations.

Figure 2.2 – Crystals of \textit{Hj}Cel7A. The C2 and P6322 crystal forms cannot be distinguished without collecting X-ray diffraction data.

Figure 2.3 – Crystal packing diagram of the \textit{Hj}Cel7A CD. Left, C2 crystal packing \textit{(ac} face). Right, P6322 crystal packing \textit{(ab} face).

Figure 2.4 – Crystal structure of TrCel7A with cellononaose molecule loaded into the tunnel. Active site tunnel loops are labeled. (PDB ID 4C4C)

Figure 2.5 – Celllobiose bound at the +1/+2 subsite (PDB ID 4P1J). T246, R251, D259 interact with the +1 glucosyl moiety. R394 and R267 directly interact with the +2 glucosyl moiety, while D262 and F338 form water-mediated interactions.

Figure 2.6 – Hydrogen bond network formed by R251 within the product-binding site. The orange dashed lines schematized the long range interaction between R251 and Y381. Top, open tunnel structure (4P1J). Celllobiose is not shown to simplify the figure. Bottom, closed tunnel structure (5TC9).

Figure 2.7 – Close up of loops from superposed 4P1H (dark shade) and 4P1J (light shade). Top, left to right: A1, A2, A3, A4. Bottom left to right: B1, B2, B3, B4.

Figure 2.8 – Superposition of 4P1J (green) and 5TC9 (blue) structures showing the Y247 and Y371 flip.

Figure 2.9 – Analysis of \textit{Tr}Cel7A conformational flexibility.

Figure 2.10 – Molecular dynamics simulation RMSD analysis for cellononaose (top, black), loop B4 (middle, cyan), and loop A4 (bottom, green).

Figure 2.11 – RMSD of different \textit{Tr}Cel7A crystal structures.

CHAPTER 3

Figure 3.1 – Crystal structure of Neurospora crassa CDHIIA. Cytochrome domain, red; dehydrogenase domain, yellow; CBM, orange. (PDB ID 4QI7)

Figure 3.2 – Electrostatic mapping of \textit{N. crassa} (PDB ID 4QI7) and \textit{M. thermophilum} (right, PDB ID 4QI6) cytochrome and dehydrogenase domain interface. Left and right are the electrostatic calculations done at pH 5.5 and 7.5, respectively. \textit{N. crassa} cytochrome \textit{(A-B)} and dehydrogenase domains \textit{(E-F)}. \textit{M. thermophilum} cytochrome
(C-D) and dehydrogenase (G-H) domains. Surface potentials are mapped from -10 kT/e (blue) to +10 kT/e (red). Electrostatics were calculated with PDB2PQR.(Dolinsky et al. 2004)...

Figure 3.3 – CDHIIA samples tested for purity and aggregates before SAS experiments. SDS-PAGE gels of NcCDHIIA (left) and MtCDHIIA (middle). SV showed NcCDH sample (right)...

Figure 3.4 – Guinier plot of Mt (pH 5.5, red; pH 7.4, orange) and NcCDH (pH 5.5, green; pH 7.5, blue). Curves are offset for clarity...

Figure 3.5 – Scattering curves of MtCDH (pH 5.5, red; pH 7.4, orange) and NcCDH (pH 5.5, green; blue) accompanied with the P(r) curves...

Figure 3.6 – Multistate modeling of NcCDHIIA. Top, pH 5.5 represented as a three state model with the most prevalent conformer in black (Rg 36.67 Å, w= 60%) lightening to the less visited states (gray: Rg= 46.36 Å, w=30.3%; light gray: Rg=31.85 Å, w=9.8%). Bottom, pH 7.5 represented as a two state model with black being the most visited conformer (Rg=38.48 Å, w= 75.9%) and light gray being the less visited state (Rg=37.20 Å, w=24.1%). To the right of both models are the accompanying calculated profiles overlayed with the experimental profiles. Glycans and cofactors not shown for ease of visualization.

Figure 3.7 – NcCDH SANS data on a log-log plot next to the respective P(r) curves. Red, 0 mM K+; orange, 20 mM K+; green, 40 mM K+; blue, 60 mM K+.

Figure 3.8 – MtCDH SAXS data plotted on a log-log plot next to the respective P(r) curves. Red, 0 mM K+; orange, 20 mM K+; green, 40 mM K+; blue, 60 mM K+.

Figure 3.9 – Best modeling solutions of MtCDHIIA resulted in 2 state solutions for pH 5.5. The most prevalent state is shown in black with the less prevalent state shown in gray. Top shows modeling results for 0 mM K+ (black: Rg 37.88 Å, w= 68.1%; light gray: Rg=33.76 Å, w=31.9%). Middle shows modeling results for 20 mM K+ (black: Rg 39.22 Å, w= 70.8%; light gray: Rg=35.07 Å, w=29.2%). Bottom, modeling results for 60 mM K+ (black: Rg=36.03 Å, w= 94.2%; light gray: Rg=50.76 Å, w=5.8%). To the right of the models are the accompanying calculated profiles overlaid with the experimental profiles. Glycans and cofactors not shown for ease of visualization.

Figure 3.10 – NcCDH Ca2+ series SANS data on a log-log plot with the respective P(r) curves. Red, 0 mM Ca2+; orange, 20 mM Ca2+; green, 40 mM Ca2+; blue, 60 mM Ca2+.

Figure 3.11 – MtCDH Ca2+ series SAXS data on a log-log plot next to the respective P(r) curves. Red, 0 mM Ca2+; orange, 20 mM Ca2+; green, 40 mM Ca2+; blue, 60 mM Ca2+.
Figure 3.12 – The most prevalent state is shown in black with the less prevalent state shown in lighter shades of gray. Top, best modeling result was a 2 state model for 20 mM Ca2+ (black: Rg 35.33 Å, w= 90.9%; light gray: Rg=47.93 Å, w=9.1%). Bottom shows 3 state model for 60 mM Ca2+ (black: Rg=35.89 Å, w=68.7%; gray: Rg=33.19 Å, w=17.2%; light gray: Rg=38.88 Å, w=14.2%). To the right of both models are the accompanying calculated profiles overlaid with the experimental profiles. Glycans and cofactors not shown for ease of visualization.

Figure 3.13 – NcCDH cellobiose SANS data plotted on log-log axis (top left). P(r) curve indicates little change in dimensions (top right). Kratky plot shows increase in flexibility with increase in substrate (bottom).

CHAPTER 4

Figure 4.1 – Partially deuterated LPMO contrast series. Data plotted on a linear–linear scale.

Figure 4.2 – Linear fitting of contrast variation series data from partially deuterated LPMO. 80% and 100% D2O arbitrarily made negative for ease of fitting.

Figure 4.3 – CDH-LPMO complex series SANS data plotted on a log-log scale.

Figure 4.4 – Normalized P(r) profiles for CDH-LPMO series. Ratio shown as molar ratio of CDH : deuterated LPMO. CDH alone, red; 1:1, orange; 1:5, green; 1:10, blue.

Figure 4.5 – Ratio of deuterated LPMO plotted as a function of deuterated LPMO. Error bars are from the P(r) analysis.

Figure 4.6 – Effect of crowders on NcCDHIIIA. The Rg (circle) and dmax (triangle) of CDHIIIA are plotted as a function of osmolality. Betaine series, gray; PEG 200 series, blue.

Figure 4.7 – AUC results from CDH-LPMO binding experiments. Left, measuring complex formation at wavelength of 280 nm. Right, measuring shifts in CDH in presence of LPMO at wavelength of 420 nm.

Figure 4.8 – Schematic of NcCDH electron transfer cycle. Top, cellobiose bound at the dehydrogenase domain of CDH does not allow IET. Middle, CDH bound to LPMO (triangle; Cu, blue circle) electron flow (dashed arrows) for secondary weak interaction site. Bottom, electron flow from CDH to a strong interaction site on LPMO. Cytochrome domain, red; dehydrogenase domain, yellow; CBM, orange.
CHAPTER 5

Figure 5.1 – Neutron diffraction images collected on IMAGINE of *T. reesei* Cel7A in hydrogenated buffer (top), 14 days (middle) and 100 days (bottom) of D2O-solvent exchanged crystal................................................................. 145

Figure 5.2 – Typical glycans present in *Pichia pastoris*. The core N-linked is showed on top with the starred that can have additional mannoses attached. Below is the average O-linked structure that is typically less than 5 mannoses in length. GlcNAc, blue square; Man, green circle................................................................. 146

Figure 5.3 – Full-length *N. crassa* (left) and *M. thermophilum* (right) with glycans shown in spheres. Glycans added with glycam web server.(Woods Group 2005-2016)......................................................................................... 148

APPENDIX

Figure A1 – Plates for phenotyping *Pichia pastoris*. The left is a minimal dextrose agar plate, where all colonies should grow. The right is a minimal methanol agar plate where only Mut+ strains will grow................................................................. 156

Figure A2 – P(r) curve examples for properly determining dmax. The correctly determined dmax is shown in red. The too large dmax, green, crosses the x-axes and becomes negative at high r values, while the too small dmax, blue, abruptly crosses the x-axis................................................................. 161

Figure A3 – Dimensionless Kratky plot. SAXS curves were generated from crystal structures using the FoXS server to create this plot. Closed conformation of a multi-domain enzyme (PDB ID 4QI6), magenta. Open conformation of a multi-domain enzyme (PDB ID 4QI7), orange. Linker belonging to 4QI7, teal.(Schneidman-Duhovny et al. 2016)......................................................................................... 162

Figure A4 – Schematics of common vapor diffusion methods. Hanging drop (left) involves adding precipitant (blue) to a protein drop (green) on a cover slip that is then sealed over top of a reservoir filled with precipitant. Sitting drop (right) utilizes pedestal to hold the drop that is set up similar to the hanging drop.......................... 163

Figure A5 – Vapor diffusion schematic for optimizing crystallization conditions. The metastable region is considered saturated, while the nucleation and precipitation areas are considered supersaturation........................................................................ 164
CHAPTER 1

Introduction to Structural Studies of Fungal Cellulases

1.1 Enzymatic Degradation of Cellulose

1.1.1 Introduction

It has been calculated that the world only has 50.7 years of oil, 52.8 years of natural gas, and 114 years of coal left in reserves at the current rate the world is consuming these fuels. The United States and China consumed over 40% of the globally consumed fossil fuels in 2015 and do not appear to be slowing down on energy consumption. With worldwide increasing energy usage, global efforts are being directed towards alternative energy sources, such as hydroelectricity, nuclear energy, and renewable energy (Renewable energy defined as wind, geothermal, solar, biomass, and waste utilization). These sources have increased by 195 million tonnes of oil equivalent in the past five years.

Biofuels are a potential alternative to fossil fuel based energy sources. From 2005 to 2015, global biofuel production increased by more than 400%. Biodiesel has become an extremely popular biofuel with consistent production increases over the past 10 years. Brazil has been the leader in bioethanol production. Many countries are attempting to emulate Brazil’s large-scale commercial production facilities.

One of the public concerns in the United States towards biofuel production is that it would cut into the USA’s food supply, specifically corn, and farming land. First generation biofuels did pursue the possibility of corn with genetically modified cell walls for bioethanol
production. (Himmel et al., 2007, Chang, 2007) Public concern about massive increases in food costs redirected the renewable carbon source for biofuels. Subsequent biofuel generations are investigating non-food crops, such as corn stover (the non-edible portion of corn), woody materials, and switchgrass, as a renewable carbon source. Genetically modified algae are another popular renewable source, due to their tunable ability to generate sugar or oil for various types of biofuels.

There are several hurdles that still need to be overcome to make lignocellulosic-based fuels more competitive with conventional fuels. Enzymes utilized to breakdown cellulose comprise a significant portion to the overall cost in biofuel production. (Humbird et al., 2011) The catalytic efficiency of these enzymes has been of concern with their narrow pH optima, slow kinetics, and inhibition from the complex environment. These problems are being addressed through creative biomass pretreatments and optimization of cellulase cocktails through mutagenesis of enzymes and mix-and-matching of cellulases from different organisms.

1.1.2 Cellulose

Plant cell walls are composed of multiple layers of polymeric sugars and phenols (Figure 1.1). The strata include the middle lamella, primary wall, and secondary wall (S1, S2, and S3). The middle lamella is comprised of pectins and gives the plant its stability. The primary cell wall surrounds the plant cell and is responsible for the determination and maintenance of the cell structure. The secondary wall provides structural support to the plant cell and can expand with a growing cell. Lignin, hemicellulose, and cellulose are found in both the primary and secondary walls, but in varying concentrations of each polymer.
The lignocellulose complex (LCC) is comprised of an intricate network of lignin, hemicellulose and cellulose. Lignin is a heterogeneous polymer of three hydroxyphenylpropanoid monomers, coniferyl, p-coumaryl, and sinapyl alcohol. Polymerization of lignin occurs through radical formation at various points within the monomer, which creates a random branched structure. Lignin is covalently linked to hemicellulose, a heterogeneous polymer of pentose sugars, such as xylose, and hexose sugars, such as mannose and galactose. Hemicellulose is a combination of linear and branched polymers with a degree of polymerization of approximately 70 to 200 in length adding to the complexity of this polymer. Within this matrix is cellulose, which is covalently linked to

![Figure 1.1 – Cellulose structure from biomass. Plant cell walls are comprised of a complex mixture of lignin, hemicellulose, and cellulose. Cellulose has crystalline and amorphous regions within the microfibrils allowing for different levels of ease for degradation into monomeric glucose. (US DOE, 2005)](image-url)
hemicellulose. The intricacy of the LCC makes extraction of cellulose difficult and requires pretreatment of biomass with harsh chemicals and/or extreme conditions.

Great efforts are being put into making cellulose a viable biofuel option since it is the most abundant biopolymer in the world. Current sources for biomass have focused on waste products, such as corn stover, grape marc from wineries, and logging residue (the unmerchantable portions of trees). (Perlack et al., 2005, Somerville et al., 2010, Corbin et al., 2015) Utilizing waste materials for biofuel production not only reduces usage of valuable raw materials, but provides bio-refineries a pre-processed starting material.

Cellulose is a linear polymer of glucose linked through a β-1,4 glycosidic bond containing a reducing and nonreducing end. While cellulose is technically a polymer of β-D-glucose, the repeating unit is cellobiose due to the orientation of glucose within the chain. (Nishiyama et al., 2003) Hydrogen bonding within and between the cellulose strands promotes the linearity of the strand and stacking of elementary fibrils during synthesis. (Moon et al., 2011) Elementary fibrils form higher order structures known as microfibrils with diameters ranging from 5 to 50 nm. The cellulose structure contains ordered and disordered regions, commonly referred to as crystalline and amorphous cellulose, respectively. The extensive hydrogen bonding between the strands gives cellulose its highly crystalline nature.

Crystalline cellulose has four main polymorphs. Cellulose I is subdivided into two morphologies, Iα (triclinic) and Iβ (monoclinic), and is the form of crystalline cellulose found in nature, such as plants, algae, and bacteria. While these two forms are primarily found in different specimens (Iα is high in algal and bacterial cellulose, while Iβ rich samples are wood and cotton), these two forms can be found within the same microfibril. (Sugiyama et
Cellulose II is a monoclinic structure generated by mercerization and regeneration of cellulose I. Cellulose I or II can be converted into cellulose III, specifically III\textsubscript{I} or III\textsubscript{II} depending upon the originating cellulose morphology, through liquid ammonia treatments. Heat treatments to cellulose III will create cellulose IV. Treating the cellulose disrupts the crystallinity making it more accessible to enzymatic attack. (Chundawat et al., 2011)

1.1.3 Cellulases

Cellulose degrading enzymes, commonly referred to as cellulases, have been extensively studied for over 65 years. (Reese et al., 1950) Cellulases are secreted by fungi to breakdown cellulose into glucose for energy. These enzymes are also found in bacteria. Two different mechanisms are employed by organisms to facilitate the breakdown of cellulose: cellulosomes and free enzymes. Cellulosomes are massive complexes consisting of hundreds of proteins noncovalently held together through dockerin-cohesin interaction on protein scaffolds. This method of cellulose degradation is typically utilized by bacteria and was initially discovered in Clostridium thermocellum. (Bayer et al., 1998) The other common method of cellulose degradation is free enzymes, where several, separate enzymes are secreted to work together to facilitate the breakdown of cellulose into monomeric glucose. (Payne et al., 2015) Filamentous fungi typically deploy these enzyme suites (Figure 1.2).

Cellulases have been grouped into families based upon their amino acid sequence with the idea that similar sequences have conserved structures and catalytic
processes (Lombard et al., 2014) The CAZy database currently treats enzymes in a modular fashion, so that each component, whether that domain is catalytically active or not, of an enzyme is assigned to a family within the database.(Lombard et al., 2014) Currently, over 330,000 sequences are organized into six protein classes within the CAZy database.(Lombard et al., 2014) These protein classes are: glycoside hydrolase (GH), glycosyltransferase (GT), polysaccharide lyase (PL), carbohydrate esterase (CE), carbohydrate-binding module (CBM), and auxiliary activity (AA).

Figure 1.2 – Model of common fungal cellulose degrading enzymes. Exoglucanases load onto the reducing ends (red) or non-reducing ends (orange) to processively hydrolyze cellulose into cellobiose. Lytic polysaccharide monooxygenase (green) and endoglucanases (purple) oxidize and hydrolyze, respectively cellulose creating new loading sites for exoglucanases. Cellobiose dehydrogenase (cyan) is the redox partner for LPMO. β-glucosidase (pink) hydrolyzes cellobiose into glucose.
Hydrolytic enzymes have received a majority of the attention over the past 65 years. GHs utilize a water molecule to break the glycosidic bond between two or more carbohydrates. (Lombard et al., 2014) Typically this reaction is facilitated by two acidic amino acids, one acting as an acid/base catalyst and the other as the nucleophile. The classical hydrolytic enzymes encompass β-glucosidases, endoglucases, and cellobiohydrolases. These enzymes have been utilized within industrial cellulase cocktails for decades to breakdown cellulose. (Peterson & Nevalainen, 2012)

β-glucosidase hydrolyzes the disaccharide cellobiose into two glucose molecules, but have also been observed to cleave β-D-galactosides, α-L-arabinosides, β-D-xylosides, and β-D-fucosides. (Sørensen et al., 2013) These enzymes fall into the GH1 and GH3 families and are present both extra- and intracellularly. Hydolysis of sugar molecules occurs through a retaining double-displacement mechanism. (Sinnott, 1998) Retaining, double-displacement mechanisms preserve the anomeric carbon configuration and involves the formation of a glycosyl-enzyme intermediate. β-glucosidases contain a pocket for substrate binding where the active site residues are located. These enzymes are the final step in the conversion of cellulose to glucose and are essential to alleviating product inhibition of other cellulases in the system. (Payne et al., 2015) Due to the accumulation of glucose in the reaction vessel, β-glucosidase becomes inhibited after cellulose degradation has occurred for extended periods of time without clearance of glucose.

Endoglucanases (EGs) randomly hydrolyze the amorphous regions of cellulose. Depending upon the GH family, EGs utilize either a retaining or inverting mechanism. (Payne
et al., 2015) All EGs contain a structurally conserved substrate-binding cleft. Some EGs contain CBMs, which typically, but not always, improves catalytic efficiency. (Tang et al., 2014, Wang et al., 2012, Cheng et al., 2012) EGs not only create new loading sites for cellobiohydrolases, but alter the morphology of the cellulose. (Wang et al., 2012) By digesting amorphous regions, EGs roughen the surface of the cellulose followed by a swelling of the microfibrils due to a loosening of the fiber. (Wang et al., 2012)

Cellobiohydrolases (CBH) processively hydrolyzes both amorphous and crystalline cellulose. A structurally conserved tunnel within the catalytic domain (CD) is responsible for hydrolysis of cellulose into cellobiose. These enzymes typically have CBMs that enhance crystalline cellulose degradation, but are not required for enzymatic activity. (Igarashi et al., 2009) CBHs have directional specificity. They can only load at a reducing or a non-reducing end of a cellulose strand depending on the family. (Igarashi et al., 2011, Ghattyvenkatakrishna et al., 2013) The reducing end and non-reducing end acting CBHs synergistically work together to degrade crystalline cellulose. (Boisset et al., 2000, Igarashi et al., 2011) These enzymes have been proposed to exhibit exo-initiation, where a cellulose strand is threaded through the tunnel, as well as endo-mode, where the cellulose chain is positioned within the tunnel through collective loop motions similar to EGs. (Ståhlberg et al., 1991) This has been recently demonstrated in Humicola grisea CBH. (Oliveira et al., 2013)

Initial oxidative enzymes were discovered from complementary DNA libraries and initially classified as GHs. (Beeson et al., 2015) In 2010, oxidative enzymes were isolated and shown to break glycosidic bonds presenting new opportunities for increasing cellulose degradation. (Vaaje-Kolstad et al., 2010) These enzymes were reclassified from GHs and
CBMs to AA to account for their oxidative properties. (Levasseur et al., 2013) The redox pair lytic polysaccharide monooxygenases (LPMO) and cellobiose dehydrogenases (CDH) have gained interest due to their ability to significantly boost hydrolytic cellulases’ ability to depolymerize cellulose. (Wang & Lu, 2016, Bulakhov et al., 2016) Other carbohydrate active enzymes with oxidative properties are glucose 1-oxidase, pyranose oxidase, alcohol oxidase.

LPMO binds to the cellulose surface and oxidizes the cellodextrin strand. Organization of LPMOs is based upon the site of oxidation of the cellulose. Type 1 LPMOs oxidize C1 creating an aldonolactone. (Phillips et al., 2011) Type 2 LPMOs oxidize at the C4 position generating a 4-ketoaldose. Type 3 LPMOs are capable of oxidizing at both the C1 and C4 position along a cellulose strand. These enzymes typically contain a copper ion coordinated by a histidine brace. In order to cleave the glycosidic bond two electrons and protons are required. (Phillips et al., 2011, Kim et al., 2014) CDH or small molecules, such as ascorbic acid or gallic acid, can supply electrons to LPMO. (Kittl et al., 2012, Dimarogona et al., 2012, Vaaje-Kolstad et al., 2010) CBMs have been shown to increase the retention of LPMOs to the cellulose surface, but changes due to the addition of a CBM to catalytic rates are LPMO and substrate dependent. (Crouch et al., 2016)

CDH is a multi-domain enzyme that oxidizes small, soluble sugars, primarily cellobiose. All CDHs contain a dehydrogenase domain with a flavin adenine dinucleotide (FAD) cofactor and a cytochrome domain with a heme b cofactor. A flexible linker connects the two domains. Some variants of CDH have a CBM attached to the dehydrogenase domain by a short linker. Sugar oxidation occurs within a pocket in the dehydrogenase domain where FAD is reduced to FADH₂. Electrons are passed one at a time from the dehydrogenase
domain to the cytochrome domain. The cytochrome domain transfers electrons to an acceptor, which within the cellulase suite is LPMO.

1.1.4 Cel7A

The CBHI, Cel7A, has been extensively studied to better understand how this two-domain enzyme is processively depolymerizes cellulose. Cel7A is composed of a glycosylated, GH7 catalytic domain (CD), where cellulose hydrolysis occurs, and a family 1 CBM connected by a glycosylated linker. Cel7A CD loads onto the reducing-end of a cellulose strand that is threaded through the structurally conserved active site tunnel for hydrolysis to occur. (Divne et al., 1994, Divne et al., 1998) The CD tunnel is subdivided into nine subsites that correspond to the position of the glucosyl moiety within the tunnel. (Divne et al., 1998) The cellodextrin chain enters at the -7 subsite and is threaded through to the +2 subsite which is located at the back of the tunnel. While Cel7A is threading the cellodextrin chain through the tunnel, the complex is in the “pre-slide” mode. (Beckham et al., 2014) The following step, “slide” mode, involves a fully loaded tunnel with each glucosyl moiety in a chair conformation. The glycosidic bond is positioned too far from the catalytic residues at the slide mode, so distortion of the -1 glucosyl moiety occurs through a rotation of the cellulose strand creating the Michaelis complex. The enzyme and substrate are now primed for hydrolysis. Immediately after cleavage of the glycosidic bond, Cel7A is in the “unprimed” glycosyl enzyme intermediate (GEI) mode. Cellobiose shifts towards the tunnel’s exit site forming the “primed” GEI mode with the glycosyl moieties resembling the slide mode. (Beckham et al., 2014) The movement of the cellobiose allows the movement of a nucleophilic water molecule to deglycosylate the cellulose strand from E212. The
cellobiose product is released from the +1/+2 subsites and Cel7A returns to the “pre-slide” mode.

Eight loops enclose the CD active site tunnel (Figure 1.3). Variations in the length of these loops are observed throughout organisms and have been reported to alter activity and/or processivity of the enzyme. (von Ossowski et al., 2003, Textor et al., 2013, Momeni et al., 2013) Loops A1 and B1 interact with the incoming cellulose strand at the tunnel entrance. Residues within loop A1 form hydrogen bonds with the celldextrin strand at the -6/-5/-4 subsites, while the backbone of the residues of loop B1 have water-mediated interactions with the -6/-5 glucosyl moieties. Loop B1 hydrogen bonds with loop B2, which forms direct and water-mediated interactions with the -5/-4 glucose residues. Loop B3, previously referred to as the exo-loop, forms direct and water-mediated interactions with the -2/-1/+1/+2 subsites. This loop has been shown to exhibit conformational variability in previously reported crystal structures and has been the subject of mutational studies. (von Ossowski et al., 2003, Knott, Haddad Momeni, et al., 2014) Loop B4 has only been observed to interact with cellobiose during product release at the +4 site. (Divne et al., 1998) The opposite side of the tunnel is composed of loops A2, A3, and A4 that are connected as a large, well-structured loop. Loop A2 does not interact with the threaded cellulose strand, but reorients upon complexation with cellulose. (Granum et al., 2014) The backbone of residues within loop A3 is part of the -3 subsite. Loop A4 is situated above the product-binding site, but, similarly to loop B4, is only observed to make water-mediated interactions with cellobiose during product release at the +3 subsite. (Divne et al., 1998) Computational work has noted the high
flexibility at loop A4 may play a role in product binding and/or release. (Momeni et al., 2013, Bu et al., 2013)

Flexibility of the loops along the tunnel have been investigated through molecular dynamics (MD) simulations and inferred through other techniques. Small angle neutron scattering (SANS) explored the effect of pH on full-length *Trichoderma reesei* Cel7A. (Pingali et al., 2011) It was noted that as pH decreases from 7.0 to 5.3, there were subtle changes in the distance between the CD and CBM. However, when the pH decreased to 4.2 a significant increase in disorder was observed. This indicates an increase in conformational variability of the tunnel loops that occurs when Cel7A is at the optimal
pH. (Pingali et al., 2011) Subsequent MD simulations on *T. reesei* Cel7A and Cel6A, a non-reducing end acting CBH, determined that pH affected the flexibility of the tunnel loops as well as supporting the correct ring pucker of the loaded substrate. (Bu et al., 2013) For Cel7A, loops B2 and B3 were the most dramatically affected by pH. (Bu et al., 2013) Simulations on a different CBH, Cel7B, investigated the role of pH on loop motions within an open tunnel and substrate loaded tunnel. (Granum et al., 2014) pH was observed to affect the loop motions corresponding to *T. reesei* loop B3, while loops A2 and A4 were noted to consistently have high fluctuations. In the substrate bound simulations, A2 and A4 increased in mobility, while loop A3 decreased in fluctuations. The overall loop fluctuations led to loops B2 and B3 moving in towards the tunnel, while loop A2 moved away from the active site. Loops around the active site were observed to form an open conformation when the cellulose strand was loaded, but the tunnels forms a closed conformation upon hydrolysis. (Silveira & Skaf, 2015)

In a study attempting to engineer *T. reesei* Cel7A loop B3 to mimic another CBH, it was noted that there is a delicate balance between substrate binding at the -7 to -2 subsites and the product site binding. (von Ossowski et al., 2003) The product site has the strongest affinity for cellobiose substrates, while the entrance only has a mild affinity in comparison. (Colussi et al., 2015) The high affinity for cellobiose substrates towards the back of the tunnel assists in threading the cellulose strand to the back of the tunnel. (Knott, Crowley, et al., 2014) However this property of the tunnel leads to some of the worst product inhibition present within cellulases. (Bu et al., 2012, Teugjas & Väljamäe, 2013) Alleviating product inhibition has been a topic of great interest for CBHs, but has been a difficult
problem to overcome. Computational studies mutated residues throughout the product-binding site of *T. reesei* Cel7A to investigate their structural implications on ligand interactions. (Silveira & Skaf, 2015) By altering the product-binding site, consistent changes in the entrance of the tunnel occur. The only tested mutant that did not significantly affect the entrance or structural dynamics was R394, which is a mutation that has been patented by BP. (Hanson *et al.*, 2013) Enzymatic assays on *Talaromyces emersonii* Cel7A revealed that an alanine mutation to the residue corresponding to *TrCel7A* Y381 alleviated product inhibition with minimal negative effects to hydrolytic activity. (Atreya *et al.*, 2016) Other mutations made throughout the product-binding site that led to significant decreases in cellobiose affinity also had significant reductions in hydrolytic activity.

CBMs, previously referred to as cellulose binding domains, are prevalent throughout the different cellulase classes. *T. reesei* CBM is made up of $\beta$-sheets forming a wedge shape and is stabilized by disulfide bonds. (Kraulis *et al.*, 1989) The CBM has a flat surface with three tyrosine residues (Y474, Y492, and Y493, also referred to as Y5, Y31, and Y32, respectively) that interact with the cellulose surface. (Kraulis *et al.*, 1989, Reinikainen *et al.*, 1992) These tyrosine residues are evenly spaced out so that they interact with every second pyranose ring. (Mattinen, Kontteli, *et al.*, 1997) Three hydrophobic residues along the flat surface, typically comprised of tyrosines, are conserved throughout family 1 CBMs with variation sometimes seen with tryptophan in either or both of the first two positions (WWY, WYY, YWY). (Mattinen, Linder, *et al.*, 1997, Beckham, Matthews, *et al.*, 2010, Linder, Mattinen, *et al.*, 1995) Mutations to Q34 and N29 affected cellulose binding indicating hydrogen bonding along the flat surface influenced cellulose interactions. (Linder, Mattinen,
et al., 1995) T. reesei CBMs have a higher binding affinity to increased cellulose crystallinity. (Guo & Catchmark, 2013)

There have been evolving hypotheses on how the CBM interacts with and disrupts the cellulose surface. Initial mutation studies found that alteration to the flat surface and rough face resulted in a significant decrease in crystalline cellulose degradation. (Reinikainen et al., 1992) This led to the initial conclusion that the CBM disrupted the cellulose surface by wedging itself between cellodextrin chains to disrupt the hydrogen bonding network. (Reinikainen et al., 1992) Bacterial, family 2 CBMs from Cellulomonas fimi diffused into the cellulose surface supporting the hypothesis that the CBM buries itself within crystalline to aid in cellulose disruption. (Jervis et al., 1997) Computational studies confirmed Reinikainen et al.’s findings that both the flat and rough face interacted with the cellulose during cellulose degradation. (Mulakala & Reilly, 2005)

Extensive mutations of a synthesized CBM along both the flat surface and rough face concluded that binding only occurred on the flat surface. (Linder, Lindeberg, et al., 1995) In one of their initial NMR binding studies, Mattinen et al. observed that the CBM interacted, albeit weakly, with cellohexaose only at the flat surface. (Mattinen, Linder, et al., 1997) This was potentially due to preferential binding to a crystalline substrate rather than a soluble sugar, where it had been previously found that CBMs have no additional benefit to amorphous cellulose degradation. (Van Tilbeurgh et al., 1986) MD simulations later observed that the CBM preferentially binds to the hydrophobic face of cellulose. (Nimlos et al., 2012) If binding at a hydrophilic face occurs, thermodynamic forces drive the CBM to the hydrophobic face. Isothermal titration calorimetry and adsorption isotherms were utilized to
characterize differences in \textit{TrCel7A} and \textit{TrCel6A} CBMs that were expressed in \textit{Escherichia coli}.\cite{Guo2013} These experiments showed that CBM binding for both cellulases were enthalpically driven revealing that the CBM adsorbs onto the cellulose surface interacting with only the flat face.

Previous studies investigating the CBM on its own were obtained through peptide synthesis or \textit{E. coli} expression, both leading to CBMs lacking glycosylation. Taylor \textit{et al.} found that glycosylation significantly alters the binding of the CBM to the cellulose surface by increasing hydrophobic stacking and hydrogen bonding.\cite{Taylor2012} The addition of O-linked glycans to the CBM at a naturally occurring site, Ser3, increased binding by three-fold, whereas the addition of a non-native glycan at Ser14 increased binding by 140-fold. Glycans increasing CBM binding is an intriguing result, since previous attempts to increase binding affinity through amino acid substitution only led to a two-fold increase.\cite{Linder1995}

The linker between the CBM and CD has been investigated to gain insight into its role for Cel7A cellulose degradation mechanism. Linker truncations led to decreases in cellulose degradation and Cel7A adsorption.\cite{Srisodsuk1993} Thus signifying that the linker length allows for an optimal distance between the two domains and that it allows the CBM to adsorb onto and laterally diffuse along the cellulose surface so the CD can find a loading site. Further analysis into the role glycosylation plays on linker flexibility found that the glycans did not alter the stiffness.\cite{Beckham2010} The results of the MD simulation showed the linker to behave like an intrinsically disordered protein and that the glycans increased the working range of Cel7A.\cite{Beckham2010} A
subsequent MD simulation revealed that the linkers were optimized based on the CD and CBM to which they were attached. (Sammond et al., 2012) Glycans were confirmed to act as a linker extension by consistently increasing the linker length with succeeding additions of monosaccharides and disaccharides at each O-linked glycosylation site. (Sammond et al., 2012) These studies showed that this seemingly random linker connecting the two functional domains, could be another optimization site and is not a portion of the enzyme to be overlooked.

1.1.5 *Cellobiose Dehydrogenase IIA*

CDH is one of the oxidative enzymes organisms utilize to breakdown biomass. The CDHIIA variant is a glycosylated protein comprised of three domains: an N-terminal cytochrome domain with a heme b cofactor; a dehydrogenase domain with an FAD cofactor; C-terminal family 1 CBM (Figure 1.4). The enzyme has a natural affinity towards cellobiose,
a disaccharide of glucose linked by a β-1,4 glycosidic bond, but is capable of oxidizing other soluble saccharides, such as glucose, lactose, and xylobiose. (Harreither et al., 2011, Sygmund et al., 2012) Oxidation of sugars occurs at the dehydrogenase domain where electrons are stored within the FAD cofactor that is reduced to FADH$_2$. Electrons are passed from the dehydrogenase domain to the cytochrome domain one at a time. The cytochrome domain will then transfer electrons to an electron acceptor, such as an LPMO.

The ability for CDHs to transfer electrons and oxidize a wide-range of substrates has made these enzymes an intriguing target for a wide-range of applications outside of biofuel production. One of the application areas for CDH is sensor technology. CDH-immobilized onto electrodes has been used to monitor cellulose degradation reactions and is a potential candidate for third-generation biosensors for glucose monitoring.(Cruys-Bagger et al., 2012, Cruys-Bagger et al., 2013, Olsen et al., 2015, Zafar et al., 2012) Third-generation biosensors work by immobilizing an enzyme, such as CDH, onto an electrode surface capable of direct electron transfer eliminating the need for mediating substances.(Felice et al., 2013)

Enzymatic fuel cells (EFC) are one avenue of interest for CDHs (Figure 1.5).(Shao et al., 2013, Krikstolaityte et al., 2014) EFCs are of increased interest due to their ability to self-regenerate power and ease of miniaturization to implement into biomedical devices. CDH, the bioanode, and bilirubin oxidase, the biocathode, are immobilized onto a surface, such as silver or graphite.(Krikstolaityte et al., 2014, Shao et al., 2013) A sugar is oxidized at the dehydrogenase domain of CDH and the electrons are passed to the electrode either through the cytochrome domain or directly from the dehydrogenase domain.(Ludwig et al., 2013) The electrons generated from the oxidation of the sugars produce an electric current. Once
the electrons reach the enzyme at the biocathode, the enzyme converts molecular oxygen to water. CDH capability to produce hydrogen peroxide ($H_2O_2$) has further expanded possible application areas. (Pricelius et al., 2009, Kittl et al., 2012, Flitsch et al., 2013, Sygmund et al., 2013, Tegl et al., 2016) Production of $H_2O_2$ initiates upon reduction of FAD into $FADH_2$.

The electrons stored within $FADH_2$ react with molecular oxygen to form $H_2O_2$. *Myriococcum thermophilum* CDH has been shown to effectively bleach cotton and oxidize standard cotton flavonoids. (Flitsch et al., 2013) CDH has been shown to effectively prevent microbial growth making it an attractive addition to biomedical devices that are common sites for infection. (Thallinger et al., 2014, Vaterrodt et al., 2016, Tegl et al., 2016, Lipovsky et al., 2015)
Each of these applications requires CDH to maintain activity at various pHs and under less than favorable enzymatic conditions, thus requiring a robust, tunable enzyme to operate under a spectrum of environments. Previous work grouped CDHs from ascomycetes into acidic, neutral, and alkaline based on their activity pH optimum. (Harreither et al., 2011) The alkaline category of CDHs has activity pH optima within a neutral to alkaline region. Acidic CDHs have a pH optimum around pH 5.0 and a narrow pH range where in they are capable of continuing to operate. (Harreither et al., 2011) MtCDH is characterized as an acidic CDH. (Harreither et al., 2007) Intermediate CDHs, NcCDH falls into this category, have a pH optima between 5.0 and 6.0, but are capable of maintaining their activity at a broader pH range than acidic CDHs. (Harreither et al., 2011) In the context of interdomain electron transfer (IET) rate, it has been observed that the rate limiting step, FAD reduction or IET, is pH and substrate concentration dependent. (Igarashi et al., 2002)

Recent experiments have been able to increase or regain IET in CDHIIA at pH conditions away from their optimal activity pH. (Kielb et al., 2015, Kracher et al., 2015, Schulz et al., 2012) Initial work investigated immobilized class I CDH from Phanerochaete chrysosporium, and class II CDHs from both Humicola insolens and M. thermophilum. (Schulz et al., 2012) There was a two-fold increase in catalytic current for MtCDH in the presence of monovalent cation salt, KCl, whereas a five-fold increase was observed when the divalent cation salt, CaCl₂, was added to the buffer. Kracher et al further investigated this phenomenon by observing effects of different atomic radii of cations CaCl₂ on CDHs from basidiomycota and ascomycota. (Kracher et al., 2015) It was found that the atomic radius did not affect the IET rate. All CDHs had increased IET rates when away from optimal pH
conditions in the presence of CaCl\(_2\) except for the Sclerotium rolfsii and N. crassa CDHs. Based on these activity studies and docking models, it was hypothesized that the cations undergo complexation with CDHIIA at the interface of the cytochrome and dehydrogenase domains, allowing the domains to more easily come together and electrons to be transferred at a more efficient rate. (Schulz et al., 2012, Kracher et al., 2015)

Cellulase optimization studies have found that adding CDH and LPMO significantly improves cellulose degradation. (Bulakhov et al., 2016, Langston et al., 2011, Phillips et al., 2011, Wang & Lu, 2016, Igarashi et al., 1998, Igarashi et al., 2002) The ability for LPMO to create new reducing or non-reducing ends within crystalline cellulose strands is of major importance, since this allows increased loading sites for CBHs. (Quinlan et al., 2011) While CDH is the enzymatic electron donor for LPMO, several studies have found that lignin within the cell wall and small molecules, such as ascorbic acid and gallic acid, are capable of restoring the active site copper ion. (Quinlan et al., 2011, Kittl et al., 2012, Bulakhov et al., 2016, Vaaje-Kolstad et al., 2010, Dimarogona et al., 2012) Lignin regeneration of LPMO is an intriguing route to explore, since it is already present within the biomass; however, lignin is known to inhibit cellulases. (Berlin et al., 2006) Gallic acid, a phenol, has also been demonstrated to inhibit cellulases. (Ximenes et al., 2011) While it seems more convenient and cost-effective to replace an enzyme in a cellulase cocktail, it has been argued that the presence of CDH also prevents product inhibition through the oxidation of cellulose degradation products. (Igarashi et al., 1998, Wang & Lu, 2016)
1.2 Neutrons in Biology

Structural biology involves the study of biological materials at the molecular level where length scales ranging from the Ångstrom ($1 \text{ Å} = 10^{-10} \text{ m}$) to the micron ($1 \text{ µm} = 10^{-6} \text{ m}$) are investigated. Within these scales, structures of proteins or nucleic acids can be elucidated through techniques such as electron microscopy, nuclear microscopy, crystallography and small angle scattering. Neutrons are capable of probing questions that cannot be answered with X-rays. With neutron crystallography, neutrons are capable of “seeing” hydrogen. This allows researchers to investigate enzymatic mechanisms and gain insight on protonation states that would not be observed with X-rays. Small angle neutron scattering enables users to study complexes by isotopically labeling components and/or taking advantage of natural contrast. This technique allows for the mapping of massive biological complexes.

1.2.1 Neutron Protein Crystallography


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 (Purich, 2010),
controlling processes as diverse as the capture and conversion of energy from light, cell signaling, DNA translation and repair, and metabolic pathways. 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) and allosteric regulation mechanisms. Local environment, pH, and chemical flux also trigger significant structural reorganization contributing to activity modulation. (Whittier et al., 2013, Hanoian et al., 2015, Harris & Turner, 2002, Gutteridge & Thornton, 2005)

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, cofactors, 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 be inferred from precise geometrical parameter analyses. (Neumann & Tittmann, 2014) 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. (Jelsch et al., 2000, Liebschner
et al., 2009, Zarychta et al., 2015) 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.

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 1.6). (Sears, 1992) 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 (Di Costanzo et al., 2007, Fisher & Helliwell, 2008, Hazemann et al., 2005, Meilleur et al., 2005, Arteiro et al., 2005), at typical resolutions of 2.5 Å or better.

Despite a number of significant technical challenges, 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 stand-alone technique or in complementary 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) (Adachi et al., 2009, Weber et al., 2013), beta-lactamase (Tomanicek et al., 2010, Tomanicek et al., 2011, Tomanicek et al., 2013), dihydrofolate reductase (DHFR) (Bennett et al., 2006, Wan, Bennett, et al., 2014), carbonic anhydrase II (Fisher et al., 2010, Fisher et al., 2011, Fisher et al., 2012, Michalczyk et al., 2015), and xylose isomerase (Kovalevsky et al., 2008, Kovalevsky et al., 2010, Kovalevsky et al., 2012, Katz et al., 2006, Meilleur, 2004, Meilleur, Snell, et al., 2006, Munshi et al., 2014). These studies have been essential in identifying active site residues with perturbed pKa's and contribute to the growing body of knowledge on pKa modulation at the active site of enzymes.

1.2.2 Small Angle Neutron Scattering

Small angle scattering (SAS) of a macromolecule gives information about the particle’s size, shape, and molecular weight. Monodisperse samples are studied in dilute aqueous environments allowing for molecules to be studied in their native state. SAS is one of many tools in a structural biologist’s toolbox. Crystallography enables experimenters to

![Figure 1.6 – 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.](image-url)
obtain atomic resolution of molecules without size restrictions on the macromolecule of interest, but obtaining crystals can sometimes be difficult and time consuming. When structures are obtained of multi-domain proteins, concerns of trapping non-biologically relevant conformations are raised. Nuclear magnetic resonance (NMR) can also provide atomic resolution structures of macromolecules with the added advantage of being in solution, but is limited in the size of the particle that can be studied. SAS compliments both of these methods by providing structural information on domain and subunit arrangements, determining or confirming oligomeric states in solution, or probing conformational changes. Electron microscopy allows for the direct imaging of a macromolecule, unlike the previously listed techniques, but is limited in its resolution (which is continuously improving).

Neutron or X-ray radiation sources can be used for SAS to study biological samples with their own distinct advantages and disadvantages. With small angle X-ray scattering (SAXS), small sample volumes (10 – 70 µL with protein concentrations 2 – 5 mg mL\(^{-1}\)) are needed for measurements, while small angle neutron scattering (SANS) require larger sample volumes (about 200 – 300 µL). SAXS instruments are much easier to gain access to, whether at a home-source or synchrotron. While many biological questions can be answered with SAXS, SANS is beneficial for samples that are sensitive to X-ray radiation and is most commonly used for contrast variation experiments.

SAS measures the intensity of radiation scattered by a sample and can be described by equation 3:

\[
I(q) = N(\Delta \rho V)^{2} P(q)S(q)
\]

Equation 3
where \( N, \Delta \rho, V, P(q), \) and \( S(q) \) are the number of particles per unit volume, the contrast, the volume of each macromolecule, form factor (provides information on the average structure of the particle), and structure factor (gives information on correlation distances between particles, which should not be present in dilute systems), respectively. X-rays scatter from the electrons of atoms, while neutrons interact with the nucleus. (Whitten & Trewhella, 2009)

While the X-ray scattering length of atoms directly varies with the amount of electrons, the neutron scattering length does not change based on the size of the nucleus or isotope. The neutron scattering length of H and D are \(-3.74 \text{ fm}\) and \(6.67 \text{ fm}\), respectively, which allows for the scattering length of a component to be varied through isotopic labeling (Figure 1.7).

Contrast variation allows for an experimenter to study multi-component systems by manipulating the solvent to gradually mask a component while highlighting another to determine its location in relation to another molecule or determine a molecule’s match point,

![Scattering length density as a function of solvent deuteration](image)

**Figure 1.7** – Scattering length density as a function of solvent deuteration. Values calculated with MULCh web server. (Whitten *et al.*, 2008)
where the scattering at zero angle (I(0)) is zero. Contrast variation SANS experiments are carried out by varying the H$_2$O:D$_2$O ratio in the sample. By altering the deuterium content of the solvent, the experimenter is able to highlight different portions of the system being studied (Figure 1.8).

When a contrast matching experiment is performed, the match point of a particle is determined through an initial contrast variation experiment. This type of experiment is typically performed to observe the effects of a macromolecule on another macromolecule. The contrast variation experiment is set up by measuring the particle to be matched out, or made “invisible”, in five to six different percentages of D$_2$O buffers (for example, 0%, 20%, 40%, 60%, 80%, and 100% D$_2$O). The square root of I(0) from each of these points is plotted against the percent D$_2$O in solution. When the data is fit (either a parabola can be used or higher D$_2$O solvents can be arbitrarily made negative for a linear fit), the point where the line intersects the x-axis is called the match point for the sample. The percentage of D$_2$O in the solvent that correlates with the match point of the macromolecule is the concentration of D$_2$O that will be used in subsequent multi-component measurements.

For all SAS experiments, but especially for contrast variation experiments, it is

![Figure 1.8 – Schematic of contrast variation series. As the ratio of H:D is varied (as illustrated by a gradient from blue to green), different particles become “visible”.](image-url)
essential to have an accurate determination of the concentration of the sample. Using equation 3 to look at scattering angle at zero with the form factor normalized:

\[ I(0) = N(\Delta \rho \bar{V})^2 = C\Delta \rho \bar{V}^2 M \]  

Equation 4

where C, \( \bar{V} \), and M are the concentration, the partial specific volume of the particle, and mass of the particle, respectively. (Whitten & Trewhella, 2009) Equation 4 shows that the scattering at zero-angle is directly proportional to the concentration of the sample.

If a contrast matching experiment is being undertaken, care should be taken to ensure matching out of the particle. An expansion of equation 3 for a two-component system allows for the total scattered intensity to be approximated:

\[ I(q) = N[\Delta \rho_1^2 \bar{V}_1^2 P_{11}(q) + \Delta \rho_2^2 \bar{V}_2^2 P_{22}(q) + 2\Delta \rho_1 \bar{V}_1 \Delta \rho_2 \bar{V}_2 P_{12}(q)] \]

\[ = \Delta \rho_1^2 I_{11}(q) + \Delta \rho_2^2 I_{22}(q) + \Delta \rho_1 \Delta \rho_2 I_{12}(q) \]  

Equation 5

Equation 5 shows that the scattering profiles of the individual components are influenced by the contrast. An additional term arises, \( I_{12}(q) \), that describes the interference between the scattering elements; this term is referred to as the cross-term. Each of these terms provides enough information to build models of the system being investigated. MULCh (ModULes for the analysis of small-angle neutron contrast variation data form biomolecular assemblies) is a program that can determine the contrast of each component within a system and then breakdown the total scattering profile into the individual scattering profiles for each component and the respective cross-term. (Whitten et al., 2008) The output from MULCh can then be modeled with programs such as SASREF or MONSA. (Petoukhov & Svergun, 2005, Svergun, 1999) For additional information on data analysis, see the annex on small angle scattering.
1.3 Research Outline

The goal of this dissertation was to gain mechanistic insight through a structural perspective on the cellulases Cel7A and CDHIIA. *Trichoderma reesei* Cel7A is an extensively studied CBH whose mechanism for processing along cellulose is still not fully understood. Through crystallography and molecular dynamics simulations, a processivity mechanism is proposed (chapter 2). The CDHIIAs from *Myriococcum thermophilum* and *Neurosopra crassa* were studied using SAS to structurally explain the activity difference of these CDHs under various pH and cation conditions (chapter 3). A subsequent SANS study on *N. crassa* CDHIIA and *N. crassa* LPMO was performed to gain insight on this complex (chapter 4). These studies have shed light on new aspects of the structure-function-dynamics relationship of two essential cellulose-degrading enzymes. Thoughts arising from these findings and potential experiments to tackle them are presented in chapter 5.
1.4 References


CHAPTER 2

Crystal Structures of Wild-type *Trichoderma reesei* Cel7A Catalytic Domain in Open and Closed States

Due to the similarity and nature of these papers, this chapter will be divided into two parts. The first section contains the crystallization note published in Acta Crystallographica F describing our crystallization conditions of full-length Cel7A in two new crystal forms. The second part is comprised of the article in press in FEBS Letters discussing the crystal structures and accompanying analysis of molecular dynamics simulation.
PART I

Crystallization and preliminary X-ray diffraction analysis of *Hypocrea jecorina* Cel7A in two new crystal forms


2.1a Abstract

Cel7A (previously known as celllobiohydrolase I) from *Hypocrea jecorina* was crystallized in two crystalline forms, neither of which have been previously reported. Both forms co-crystallize under the same crystallization conditions. The first crystal form belonged to space group C2, with unit-cell parameters $a = 152.5$, $b = 44.9$, $c = 57.6$ Å, $\beta = 101.2^\circ$, and diffracted X-rays to 1.5 Å resolution. The second crystal form belonged to space group P6$_3$22, with unit-cell parameters $a = b \equiv 155$, $c \equiv 138$ Å, and diffracted X-rays to 2.5 Å resolution. The crystals were obtained using full-length Cel7A, which consists of a large 434-residue N-terminal catalytic domain capable of cleaving cellulose, a 27-residue flexible linker and a small 36-residue C-terminal carbohydrate-binding module (CBM). However, a preliminary analysis of the electron-density maps suggests that the linker and CBM are disordered in both crystal forms. Complete refinement and structure analysis are currently in progress.
2.2a Introduction

Improving the enzymatic degradation of cellulose remains one of the key challenges in the industrial production of biofuel from lignocellulosic biomass. A significant problem is that cellulose is composed of crystalline microfibrils, which are insoluble in the native form. In order to digest crystalline cellulose, its surface must be disrupted to allow enzymes to access individual glucan chains. Understanding how cellulases from wood-degrading organisms are able to efficiently break down cellulose could greatly benefit industrial production, both in the design of more effective pre-treatment processes and by guiding the engineering of more efficient enzymes.

The wood-degrading filamentous fungus *Hypocrea jecorina* (an anamorph of *Trichoderma reesei*) secretes a complex mixture of cellulases. Although the secretome of *H. jecorina* is composed of several enzymes including endoglucanases, exoglucanases, polysaccharide monooxygenases and β-glucosidases, the exoglucanase Cel7A comprises about 60–80% of the secreted mixture. Cel7A, previously known as cellobiohydrolase I (CBHI), is a well-studied exoglucanase, which hydrolyzes cellulose to cellobiose and is one of the few cellulases that are capable of degrading both amorphous and crystalline cellulose. Its ability to bind to and disrupt the surface of crystalline cellulose enables the endoglucanases to access and rapidly hydrolyze the individual glucan chains separated from the cellulose microfibrils by its action, an effect known as synergy. The efficiency with which Cel7A is able to bind, disrupt and hydrolyze crystalline cellulose makes this enzyme a key target for converting biomass into biofuels.

Cel7A consists of a large 434-residue N-terminal catalytic domain (CD) capable of
cleaving amorphous cellulose, a flexible linker and a small 36-residue C-terminal carbohydrate-binding module (CBM) that is required for the efficient degradation of crystalline cellulose. Studies suggest that the reducing end of a cellulose chain is threaded through the long catalytic tunnel of the CD, where it is hydrolyzed to cellobiose, which is released at the exit of the tunnel. While the flexibility of the linker region (Beckham et al., 2010) makes the full-length protein structure very challenging to determine using high-resolution techniques, structures of the CD and CBM structures have been determined separately by X-ray crystallography (Divne et al., 1994, 1998; Ståhlberg et al., 1996) and NMR (Kraulis et al., 1989; Creagh et al., 1996), respectively. However, low-resolution solution-scattering information on full-length Cel7A has been obtained using both small-angle X-ray scattering (SAXS; Abuja et al., 1989) and small-angle neutron scattering (SANS; Pingali et al., 2011).

Here, we sought to determine the structure of full-length *H. jecorina* Cel7A (*HjCel7A*) in two new crystal forms in order to probe potential interactions between the different domains.

### 2.3 Materials and Methods

#### 2.3.1a *HjCel7A* purification

Full-length *HjCel7A* was purified as previously described (Evans et al., 1994) from a commercial culture filtrate of *T. reessei* ATCC 26921 (Sigma–Aldrich, St Louis, Missouri, USA). Briefly, buffer exchange of the filtrate was performed on a Sephadex G-75 column
Purification was then carried out on a DEAE Sepharose column (GE Healthcare) equilibrated in 50 mM sodium acetate buffer pH 5.0 and eluted with 1 M NaCl. Chromatofocusing on a Mono-P column (Pharmacia) was carried out with a 25 mM N-methylpiperazine buffer pH 5.4 using a linear gradient of 1:10 diluted Polybuffer 74 pH 3.0 (GE Healthcare). The buffer was then exchanged to 50 mM sodium acetate pH 5.0. SDS-PAGE was used to check the purity of the full-length protein (Figure 2.1).

2.3.2a Crystal growth

*Hj*Cel7A was concentrated to 10 mg ml⁻¹. Crystals were grown by vapor diffusion. Hanging drops consisting of 2 µl protein solution and 0.5 µl reservoir solution were equilibrated against a 1 ml well and incubated at 35 °C. Crystals belonging to space groups C2 and P6₃22 co-crystallized from 25–31% PEG 3350, PEG 4000, or PEG 6000 in 100 mM HEPES pH 6.4 (Figure 2.2). Crystal growth was improved by the addition of 0.5 µl Silver
Bullet Screen D4 (Hampton Research, Aliso Viejo, California, USA), which consists of 5 mM gadolinium(III) chloride hexahydrate, 5 mM samarium(III) chloride hexahydrate, 50 mM benzanidine hydrochloride, 0.25%(w/v) salicin, 20 mM HEPES sodium pH 6.8. Ligand-bound crystals were prepared by addition of 10 mM cellohexaose (Megazyme, Bray, Ireland). Crystals were cryoprotected in reservoir solution containing 40% PEG prior to data collection.

2.3.3a Data collection and refinement

X-ray diffraction data were collected on the SER-CAT and GM/ CA beamlines at the Advanced Photon Source (APS), Argonne National Laboratory. Data were collected at 173 °C. HKL-2000 (Otwinowski & Minor, 1997) was used to index, integrate and scale the data. Data statistics are summarized in Table 2.1. Structure-factor amplitudes were obtained from intensities using TRUNCATE (French & Wilson, 1978) from the CCP4 suite (Winn et al., 2011). The data from the C2 crystal and from the P6322 crystal grown with cellohexaose were phased using Phaser (McCoy et al., 2007) using PDB entry 2v3i (A. Fagerstrom, T.
Liljefors, J. Stahlberg, M. Sandgren, U. Berg & R. Isaksson, unpublished work) as the search model. The molecular-replacement solution will be subjected to rounds of refinement using REFMAC5 (Murshudov et al., 2011; Winn et al., 2011) followed by iterative cycles of manual rebuilding in Coot (Emsley & Cowtan, 2004) and subsequent refinement with REFMAC5. The complete models will be reported elsewhere.

Fluorescence analysis of a P6$_{3}$22 crystal on the GM/CA beamline at APS identified a peak at 6720 eV corresponding to the Sm$^{3+}$ absorption edge. A complete single-wavelength anomalous dispersion (SAD) data set was collected at this peak wavelength for experimental phasing. The Sm$^{3+}$ substructure was determined using the SHELXC/D/E pipeline (Sheldrick,

Table 2.1 – X-ray data collection statistics

<table>
<thead>
<tr>
<th></th>
<th>C2</th>
<th>P6$_{3}$22 SERCAT</th>
<th>P6$_{3}$22 GM/CA†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.85</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>35.20-1.50</td>
<td>37.65-2.70</td>
<td>47.87-2.50</td>
</tr>
<tr>
<td></td>
<td>(1.56-1.50)</td>
<td>(2.80-2.70)</td>
<td>(2.59-2.50)</td>
</tr>
<tr>
<td>Space group</td>
<td>C2</td>
<td>P6$_{3}$22</td>
<td>P6$_{3}$22</td>
</tr>
<tr>
<td>Unit cell parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a (Å)</td>
<td>152.46</td>
<td>155.516</td>
<td>153.58</td>
</tr>
<tr>
<td>b (Å)</td>
<td>44.90</td>
<td>155.51</td>
<td>153.68</td>
</tr>
<tr>
<td>c (Å)</td>
<td>57.57</td>
<td>137.96</td>
<td>137.57</td>
</tr>
<tr>
<td>α (°)</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>β (°)</td>
<td>101.25</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>γ (°)</td>
<td>90</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>R$_{merge}$ ‡ (%)</td>
<td>7.0 (32.8)</td>
<td>12.4 (49.1)</td>
<td>18.8 (144.1)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>59492 (5164)</td>
<td>26987 (2643)</td>
<td>33616 (3293)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>3.3 (2.3)</td>
<td>5.6 (5.2)</td>
<td>11.3 (3.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.1 (84.7)</td>
<td>99.3 (99.4)</td>
<td>99.9 (99.8)</td>
</tr>
<tr>
<td>$〈I/〈σ(I)〉$</td>
<td>19.9 (3.5)</td>
<td>18.8 (4.2)</td>
<td>14.7 (1.6)</td>
</tr>
<tr>
<td>Wilson B Factor (Å$^2$)</td>
<td>15.7</td>
<td>29.1</td>
<td>42.8</td>
</tr>
</tbody>
</table>

† The percentage of correlation between intensities from random half-data sets, CC$_{1/2}$, is 99.6% (63.4%) (Karplus & Diederichs, 2012). ‡ $R_{merge} = \Sigma_{hkl} \Sigma_{i} |I_{i}(hkl) - 〈I(hkl)〉|/ \Sigma_{hkl} \Sigma_{i} I_{i}(hkl)$.
Output from the SHELX program was then input into ARP/wARP (Langer et al., 2008) to build the model. REFMAC5 will be used for subsequent refinement steps.

2.4a Results and Discussion

Previous crystallographic studies of HjCel7A were conducted using the CD prepared from full-length HjCel7A submitted to papain cleavage. These CD crystals belonged to space group I222. Here, the crystallization of full-length HjCel7A produced crystals in two new space groups, C2 and P6\(_3\)22, that diffracted X-rays to 1.5 and 2.5 Å resolution, respectively. Preliminary analysis of the electron-density maps shows clear density for the CD in both crystal forms, but electron density for the linker or CBM is weak or lacking. Crystallographic disorder is expected to some degree as the high flexibility of the linker in HjCel7A is known (Tomme et al., 1988; Beckham et al., 2010; Payne et al., 2013). The Matthews coefficient (Matthews, 1968) calculated for the C2 crystal with one CD alone in the asymmetric unit is 1.98 Å\(^3\) Da\(^{-1}\) and the solvent content is 37.89%. The Matthews coefficient calculated for the full-length protein is 1.73 Å\(^3\) Da\(^{-1}\) with a solvent content of 28.88% (Winn et al., 2011). For the P6\(_3\)22 crystal, the Matthews coefficient calculated with just the CD is 4.80 Å\(^3\) Da\(^{-1}\) and the solvent content is 74.40%, while the Matthews coefficient for the full-length protein was 4.19 Å\(^3\) Da\(^{-1}\) with a solvent content of 70.69%. While we cannot exclude proteolysis of the full-length protein during crystallization, SDS–PAGE analysis of melted crystals does not indicate that the protein in the crystals is cleaved (Fig. 2.1), and the crystal packing observed in both space groups provides ample room for the linker and CBM to be present in the crystal but to be disordered (Fig. 2.3).
While the Cel7A linker and CBM are required for the degradation of crystalline cellulose, the CD alone is equally efficient as the full-length protein on amorphous cellulose (Cruys-Bagger et al., 2013). This suggests that the CD possesses the required machinery not only to hydrolyze a cellodextrin chain but also to process along a cellulose chain as cellobiose is being cleaved and released. Molecular-dynamics simulations have suggested that the exo-loop and exit loop that span residues 247–252 and 370–394, respectively, adopt different conformations depending on the binding state of the tunnel (Colussi et al., 2012; Bu et al., 2013). However, experimentally only the exo-loop is observed to move in the available structures of the CD alone (Ossowski et al., 2003; Textor et al., 2013). Analysis of these loop regions in our new C2 and P6322 crystal forms may shed new light on the processive motion of the protein during its catalytic cycle. Model building and refinement are now in progress.

Figure 2.3 – Crystal packing diagram of the HjCel7A CD. Left, C2 crystal packing (ac face). Right, P6322 crystal packing (ab face).
Acknowledgements

The authors would like to thank Ruslan (Nukri) Sanishvili and Tim Grüne for assistance with fluorescence analysis and single-wavelength anomalous dispersion during the 2013 APS data-collection workshop and CCP4 school. Funding for this work was provided in part by the NSF (award 1069091). Research at Oak Ridge National Laboratory was supported by the US Department of Energy, Office of Biological and Environmental Research, through the Genomic Science Program, under SFA Biofuels Contract FWP ERKP752, and through support for the Center for Structural Molecular Biology (CSMB) under Contract FWP ERKP291, using facilities supported by the Office of Basic Energy Sciences of the US Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle LLC for the US Department of Energy under Contract DE-AC05-00OR22725.

2.5a References


PART II

Crystal structures of wild-type *Trichoderma reesei* Cel7A catalytic domain in open and closed states


2.1b Abstract

*Trichoderma reesei* Cel7A efficiently hydrolyzes cellulose. We report here the crystallographic structures of the wild-type *Tr*Cel7A catalytic domain (CD) in an open state and, for the first time, in a closed state. Molecular dynamics (MD) simulations indicate that the loops along the CD tunnel move in concerted motions. Together the crystallographic and MD data suggest that the CD cycles between the tense and relaxed forms that are characteristic of work producing enzymes. Analysis of the interactions formed by R251 provides a structural rationale for the concurrent decrease in product inhibition and catalytic efficiency measured for product binding site mutants.

2.2b Introduction

Plant-degrading fungi utilize cellulolytic enzymes that synergistically work to breakdown cellulose into monomeric glucose through hydrolytic and oxidative processes. (Yang *et al.* 2011; Horn *et al.* 2012; Dimarogona *et al.* 2013) Commonly secreted oxidative enzymes include cellobiose dehydrogenases (CDH) and polysaccharide monooxygenases (PMO).(Dimarogona *et al.* 2013; Wu *et al.* 2013; Beeson *et al.* 2015) Hydrolytic enzymes,
including cellobiohydrolases (CBH), endoglucanases (EG), and β-glucosidases, have been under extensive investigation since the 1950’s because of their industrial importance. (Reese et al. 1950) EGs are characterized by a binding cleft that randomly binds and cleaves amorphous cellulose, while CBHs have long active site tunnels that bind cellulose end chains and processively hydrolyze both amorphous and crystalline cellulose into cellobiose. (Payne et al. 2015)

Cel7A, a secreted CBHI, from the ascomycete Trichoderma reesei is a two-domain protein comprised of an N-terminal family 7 glycoside hydrolase catalytic domain (CD) connected to a C-terminal family 1 carbohydrate-binding module (CBM). The CBM binds to the cellulose and is essential for efficient crystalline cellulose degradation. (Van Tilbeurgh et al. 1986; Cruys-Bagger et al. 2013) The CD features an active site tunnel that loads onto the reducing-end of a cellulose strand and threads it through for hydrolysis to occur. (Divne et al. 1994; Divne et al. 1998; Payne et al. 2015) In the tunnel, the glucose monomers of the threaded cellulose strand interact with residues along nine subsites (Figure 2.4). The chain enters at the -7 subsite with the +2 subsite positioned at the rear. (Divne et al. 1994) The strand is threaded through the tunnel during the “pre-slide” mode. (Beckham et al. 2014) In the subsequent “slide mode” state, each glucose moiety is in the chair conformation at all nine subsites. For the Michaelis complex to form, the cellulose chain must be rotated to induce a distorted non-chair conformation of the -1 sugar. The glycosidic bond between the -1 and +1 glucose moieties is hydrolyzed leading to a cellobiose product; a state that is referred to as the “unprimed” glycosyl-enzyme intermediate (GEI) mode. (Beckham et al.
Cellobiose is released from the exit side of the tunnel and the newly created reducing-end is rethreaded to the +2 subsite.

Eight loops form the active site tunnel (Figure 2.4). Loops A1 and B1 interact with the cellulose strand at the tunnel entrance. Loop A1 interacts directly with the polysaccharide chain, while the backbone of residues within B1 forms water-mediated interactions. Loop B1 hydrogen bonds with loop B2, which forms the -4 and -3 subsites. Loop B3 (formerly referred to as the exo-loop), interacts with the -2 and +1 glucosyl residues, and exhibits conformational variability in previously reported crystal structures. Loop B4 binds to the product at the +4 subsite during product release. On the other side of the tunnel, loops A2, A3, and A4 are connected as a large, well-structured loop. Loop A2 does not interact directly with the cellulose strand, but reorients when the CD complexes with cellulose. Residues in loop A3 are part of the -3 subsite. Although adjacent to the product-binding site, loop A4 does not appear to interact directly with the
cellulose strand. However, computational work has noted high flexibility in loop A4 suggesting a potential role in product binding and/or expulsion. (Momeni et al. 2013)

The tunnel loops have been shown to be important for both substrate stabilization and product expulsion in TrCel7A. Notably, a salt bridge between R251 and D259 within loop B3 restricts this loop conformational freedom and allows for maximal activity along crystalline substrates. (Silveira & Skaf 2015; Atreya et al. 2016) In addition, Silveira and Skaf demonstrated that in silico mutations of the product-binding site affect areas throughout the tunnel, most notably the entrance site. (Silveira & Skaf 2015) Loop motions are better characterized in other CDHs. MD simulations on the CBHII from Thermobifida fusca, Cel6B, have shown that opening and closing of the active site loops are required for product release and processive action to occur. (Wu et al. 2013) In Melanocarpus albomyces Cel7B, a CBHI, loops corresponding to A4, B2, and B3 orient themselves to stabilize the cellobextrin chain with loops B2 and B3 forming a more enclosed tunnel during the simulation. (Granum et al. 2014)

While product inhibition limits all enzyme efficiency, the heterogeneous environment that exists during lignocellulose hydrolysis increases the likelihood for product inhibition of cellulases. MD simulations of empty, substrate bound, and product bound CD revealed that Y381, on loop A4, and R251, on loop B3, drive the formation of a more enclosed tunnel conformation. (Silveira & Skaf 2015) The simulations further identified a number of inter-residue and product-enzyme interactions at the exit side of the tunnel essential for enzyme activity. A biochemical analysis of simulated mutants demonstrated that a decrease in product inhibition by mutations in the product-binding site is almost always accompanied by
a decrease in catalytic efficiency. (Atreya et al. 2016) Other than these investigations on product inhibition, a direct role in cellulose threading for the TrCel7A loops has not been demonstrated. Whereas glycoside hydrolase 6 CBHs display clear loop reorganization in reported crystal structures, this has not been observed for family 7 CBHs. (Varrot et al. 1999) However, in a small angle neutron scattering investigation of the pH dependence of TrCel7A conformation in solution, Pingali et al. discovered that the CD flexibility increased at the catalytically optimal pH. (Pingali et al. 2011) In subsequent computational simulations, the effect of pH on overall CD flexibility was found to affect loop motions. (Bu et al. 2013) These studies suggest that the protonation state of charged residues govern loop conformation, explaining in part why cellulases may have such a narrow optimal pH activity range.

Despite extensive biochemical and structural studies, a long-standing question remains on how TrCel7A extracts a single cellodextrin chain from a cellulose microfibril. An extensive network of intra- and intermolecular hydrogen bonds confers high stability to crystalline cellulose, which must be disrupted to release a single chain from the crystal. The CBM is known to play an important role, promoting adsorption on the substrate through π-CH interactions between three tyrosine residues and the hydrophobic plane of the cellulose (Din et al. 1994; Boraston et al. 2004; Liu et al. 2011), and perhaps inserting under a free chain end. (Mulakala & Reilly 2005b) However, limited proteolysis studies have also shown that the CD alone is capable of processing crystalline cellulose, though with severely decreased activity. (Tomme et al. 1988) In addition, the CD alone was shown to be capable of sliding on crystalline cellulose by AFM studies (Igarashi et al. 2009), and shown to be
capable of sliding independently of the CBM using an optical tweezer based single molecule mobility assay. (Brady et al. 2015) Mutation studies have shown that W40 of the CD is essential for processivity, as the full-length W40A mutant did not slide on crystalline cellulose. (Igarashi et al. 2009) Similarly, an E212Q inactive mutant did not slide on crystalline cellulose, suggesting that the hydrolysis and loading steps are coupled. Taken together, these results suggest that the CD possesses the complete machinery required to extract a single cellulose chain, but that the CBM substantially assists this process.

Structural analyses of transient conformations along reaction coordinates are challenging. While an X-ray crystallographic structure typically provides a snapshot of an enzyme trapped in a specific state, analysis of several structures determined under differing conditions or crystal forms, or an ensemble of crystallographic models can help gain insights into flexibility. (Echols et al. 2003) In this work, we examine loop variability and implications for TrCel7A processive mechanism by comparing structures from several different crystal forms, and find that the loops adopt conformations not previously reported in other crystallographic studies. Specifically, the closed TrCel7A structure provides evidence for loop variability and supports a role for the TrCel7A tunnel loops in cellodextrin chain abstraction. To better understand our crystallographic observations, a molecular dynamics simulation of the CD with a cellononaose molecule loaded in the tunnel was analyzed.
2.3b Materials and Methods

2.3.1b Crystal structures

Full-length, wild-type, Cel7A purification, crystallization, and X-ray data collection have been previously reported. (Evans et al. 1994; Bodenheimer et al. 2014) C2 crystals were initially grown from benzamidine containing solutions. (Bodenheimer et al. 2014) C2 crystals were reproduced from identical conditions without benzamidine. The data reduction statistics for the benzamidine free crystals are reported in Supplementary Information Table 2.2. The refinement statistics for the structures discussed here are presented in Table 2.2.

The structures were solved by molecular replacement using PhaserMR with 2V3I as the search model. (McCoy et al. 2007) Refinement with TLS restraints was performed by REFMAC5 and followed by manual rebuilding of the structure in Coot. (Emsley et al. 2010; Murshudov et al. 2011; Winn et al. 2011)

2.3.2b Molecular dynamics analysis

The structure used for MD simulations was the Cel7A CD with cellononaose spanning the active site tunnel. (Knott et al. 2014a) Trajectory files generated from previous studies were generously shared by Drs. Brandon Knott and Gregg Beckham. (Knott et al. 2014a) MD analysis and principal component analysis (PCA) was done in VMD and bio3d. (Humphrey et al. 1996; Grant et al. 2006)
2.4b Results

Full-length, wild-type \( TrCel7A \) was crystallized in space groups C2 and P6\(_3\)22 (PDB ID 5TC9 and 4P1J, respectively). As previously reported, electron density for the linker and CBM is lacking. (Bodenheimer et al. 2014)

<table>
<thead>
<tr>
<th>Table 2.2 – Refinement statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>PDB ID</td>
</tr>
<tr>
<td>LIGAND</td>
</tr>
<tr>
<td>Space group</td>
</tr>
<tr>
<td>R-work</td>
</tr>
<tr>
<td>R-free</td>
</tr>
<tr>
<td>Number of non-hydrogen atoms</td>
</tr>
<tr>
<td>Macromolecules</td>
</tr>
<tr>
<td>Ligands</td>
</tr>
<tr>
<td>Water</td>
</tr>
<tr>
<td>Protein residues</td>
</tr>
<tr>
<td>RMS (bonds)</td>
</tr>
<tr>
<td>RMS (angles)</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
</tr>
<tr>
<td>Ramachandran favored (%)</td>
</tr>
<tr>
<td>Average ( B )-factor (Å(^2))</td>
</tr>
</tbody>
</table>

2.4.1b Product interactions

The 4P1J (P6\(_3\)22 space group) structure has cellobiose bound to the product-binding site of the tunnel (Figure 2.5). Overall, the cellobiose-protein interactions in our structure are similar to those observed in the truncated E212Q inactive \( TrCel7A \) cellobiose complex (PDB ID 3CEL). (Ståhlberg 1996) We focus here on the interactions known to be key to product inhibition. (Atreya et al. 2016) O5 and O6 from the +2 glucosyl moiety form hydrogen bonds with R394 NH1 (3.3 and 3.2 Å, respectively), while O1 on the same moiety is hydrogen bonded to R394 NH2. The extensive hydrogen bond network between loop A4 R394 and the
product +2 moiety supports its role in product inhibition. Mutation of R394 to alanine suppresses this network alleviating product inhibition. (Silveira & Skaf 2015; Atreya et al. 2016)

In contrast to previous observations, the product +2 glucosyl moiety is not hydrogen bonded to D259. (Knott et al. 2014a) Instead, D259 OD1 forms a hydrogen bond with R251 NE (3.0 Å) and D259 OD2 is hydrogen bonded to Y252 OH (2.7 Å). The main chain D259 carbonyl group, however, forms a hydrogen bond with the O2 of the +1 glucosyl moiety (3.0 Å) as previously reported. (Ståhlberg 1996) D259A has been shown to mitigate product inhibition to a lesser extent than R394A or R251K. (Atreya et al. 2016) The variability observed across structures in the hydrogen bond interactions between D259 and the product provides structural insight as to why this is the case. Water-mediated interactions between O1 of the +2 glucosyl moiety and R267 NH2 and F338 O (loop B4) and between O6 of the +2 glucosyl moiety and D262 OD2 complete the hydrogen bond network that firmly positions
the product in the +1 and +2 sites. Silveira and Skaf identified the salt bridge between D262 and R267 as essential for the correct positioning of R394. (Silveira & Skaf 2015)

Computational disruption of the R267 and D262 salt bridge by mutating R267 to alanine resulted in the formation of a new salt bridge between R394 and D262. Biochemical characterization confirmed that the new salt bridge lead to decreased product binding and lower activity on cellulose, as observed for R394A, R251K, and D259A. (Silveira & Skaf 2015; Atreya et al. 2016) In our structure, D262 and R267 both form water mediated interactions with the product and therefore it is not unexpected that mutations of either of these two residues do not affect product inhibition significantly. This structural observation does confirm however that the larger decrease in catalytic efficiency of R267A compared to D262A is due to the mispositioning of R394.

The 5TC9 structure (C2 space group) reveals an empty tunnel and loops organized in novel conformations. In the empty tunnel structure, Y381 side chain position is shifted

Figure 2.6 – Hydrogen bond network formed by R251 within the product-binding site. The orange dashed lines schematized the long range interaction between R251 and Y381. Top, open tunnel structure (4P1J). Cellobiose is not shown to simplify the figure. Bottom, closed tunnel structure (5TC9).
inwards by 2.4 Å when compared to the cellobiose complex (Figure 2.6). Noticeably, the electron density allows for R251 to be modeled in a dual-position (Figure 2.6). In the alternate R251 conformation, the guanidinium group flips and points towards D259 while it points toward T246 in all reported TrCel7A structures, including our cellobiose bound structure. This reorientation alters the electrostatics in the vicinity of R251. While the canonical salt bridge between R251 and D259 is maintained, R251 now forms a hydrogen bond with T252 OH (3.3 Å) instead of forming the usual hydrogen bond with T246 OG1. The release of the interaction between T246 and R251 would contribute to a loosening of the interaction between T246 and a product occupied +1 site. R251 dual-position and higher B-factors despite extensive interactions throughout the product-binding site support that the mobility of R251 is functional. We propose that the R251 electrostatic switch is critical for the open-close transition (detailed below) and essential for catalysis.

2.4.2b Loop analysis

MD simulations have shown that a shortening of the long-distance interactions between Y381 and R251 lead to a reorganization of the tunnel from an open to a closed tunnel conformation. (Silveira & Skaf 2015) To date, only the open tunnel conformation has been observed in crystallographic structures. In our wild-type TrCel7A empty tunnel structure, the distance between Y381 OH and R251 guanidinium group is shorter by approximately 2 Å compared to the distances for the wild-type TrCel7A cellobiose complex. As predicted by the simulation, we observe the entire tunnel in an enclosed conformation. This is the first time the closed state of TrCel7A is experimentally modeled.
To compare the loop conformations of TrCel7A closed and open tunnel, the structures were superimposed and analyzed for main chain Cα root mean square deviation (RMSD) using the program LSQKAB (Kabsch 1976). For reference, RMSD plots calculated for previously reported TrCel7A structures are presented in Supplementary Information Figure 2.11. Extended loop variability is observed in the closed tunnel structure (PDB ID 5TC9). The greatest deviations for the tunnel forming loops are up to 1.4 Å on loop A1 (A100 Cα), 2.2 Å on loop B3 (D249 Cα) and 4.2 Å on loop A4 (N384 Cα) (Figure 2.7). A more subtle displacement is observed at loop B2, which twists away from the upstream loop B1, by 0.6 Å (N197 Cα). Loop A3 contains residue Y371, which flips in the close tunnel structure; this flip requires Y247 to turn inwards pulling loop B3 by 2.1 Å towards the center of the tunnel (Figure 2.8). With all of these loop conformational changes along the tunnel lead to a tunnel volume decrease from 3,464 Å³ in the open tunnel structure to 2,803 Å³ in the closed tunnel structure (Dundas et al. 2006).

Figure 2.7 – Close up of loops from superposed 4P1H (dark shade) and 4P1J (light shade). Top, left to right: A1, A2, A3, A4. Bottom left to right: B1, B2, B3, B4.
It is important to note that while the A1, B3, and A4 loops’ conformations observed in the closed state structure deviates markedly from previously reported TrCel7A structures, these loops display similar temperature factors than the rest of the protein (Figure 2.9). This is in marked contrast with the higher temperature factors of these loops previously observed (Fig. 2.9). This observation supports the hypothesis that these loop conformations forms a stable state and are visited during catalysis.

Superposition of the closed state structure with 6CEL – which has cellobiose bound at the +3 and +4 subsites, shows that loop A4 adopts a conformation that would sterically clash with the binding of cellobiose at these subsites due to the inward motion of Y381 and D385 side chains. Superposition of the closed state structure with 4C4C shows that, at the tunnel entrance, the inward motion of loop A1 leads to stronger hydrogen bond interactions between Q101 OE1 and O4 of the -6 glucosyl moiety (from 3.3 Å to 2.9 Å), V104 N and O4 of the -4 glucolysl moiety (from 3.2 Å to 2.5 Å) (Knott et al. 2014b). In addition W40 is shifted towards the -7 glucosyl moiety by 0.5 Å.
The role of salt bridges has been previously investigated for \(Tr\)Cel7A catalytic cycle, but has not been examined for a potential role in loop dynamics (Knott \textit{et al.} 2014a; Silveira \& Skaf 2015). Along the A-side of the tunnel, a conserved salt bridge exists between K102 and E408. The salt bridge between K102 and E408 links loops A1 and A2. With loop A2 existing as part of larger loop containing loops A3 and A4 as well, this salt bridge may allow for united movement along the A-side of the tunnel. To determine if the closed loop state that is modeled in the 4P1H structure is biologically relevant and if this state is visited during cellulose strand complexation, an MD simulation with a cellononaose molecule loaded into the active site tunnel was analyzed followed by PCA. Analysis of the RMSD for the cellononaose strand and tunnel twisting and compressing through the simulation with the entire CD moving in correlated movements. Loops B4 and A4 have the most significant

Figure 2.9 – Analysis of \(Tr\)Cel7A conformational flexibility.

Normalized B factor = \(\frac{B - \langle B \rangle}{\sigma(B)}\)
Figure 2.10 – Molecular dynamics simulation RMSD analysis for cellononaose (top, black), loop B4 (middle, cyan), and loop A4 (bottom, green).
movements with shifts going towards the center of the tunnel. Interestingly, the PCA revealed a straightening of the cellononaose strand reminiscent of a transition from the Michaelis complex to the slide mode.

2.5b Discussion

While the open to closed tunnel transition of the TrCel7A CD had been suggested by MD simulation, we report here the first experimental evidence. In the closed state, loops A1, B3, and A4 are much closer to the center of the tunnel than in previously reported structures (Divne et al. 1994; Ståhlberg 1996; Divne et al. 1998; Ståhlberg et al. 2001; von Ossowski et al. 2003; Knott et al. 2014b). The closing of the tunnel appear to be driven by a long-range interaction between Y381 and R251. Long-range interactions are known to play an essential role in protein biochemistry, ranging from protein recognition to folding (Go & Taketomi 1978; Kannan et al. 2016). Analysis of the hydrogen bonding and electrostatic interactions between R394, R251, surrounding residues, and the product in both the open state and closed state provides a structural rationale for the largest decrease in product inhibition and catalytic activity measured for the industrially patented R251K/R394A double mutant (Hanson et al. 2013; Atreya et al. 2016). Our results suggest that the R251 electrostatic switch may drive the open to closed transition.

Electrostatics have previously been shown to be important during TrCel7A catalytic cycle, but have not been investigated as to their role in loop dynamics (Knott et al. 2014a; Silveira & Skaf 2015). The results presented here suggest that the K102-E408 salt bridge
promotes connectedness along the A-side of the tunnel in the inward shift observed along this entire side of the tunnel.

The largest shifts on the B-side were seen at loop B3, which display a continuum of conformations when all TrCel7A structures are superposed. Loop B3 contains Y247 that flips with Y371 on loop A3. While the Y247-Y371 flip has previously been described, no functional implications have been made (Divne et al. 1998; Borisova et al. 2015). We propose that the Y247-Y371 flip helps coordinate movements between the two sides of the tunnel.

Results from the MD simulations support the conclusions drawn from the X-ray crystal structures. PCA highlighted the correlated movements present throughout the CD. Along the A-side of the tunnel, a coordinated movement involves an approach to the cellononaose molecule with the greatest amount of movement occurring at loop A4. While the B-side constricts with side A, the B loops move disjointedly. The MD simulation showed that the tunnel active site loops twist and stretch the cellononaose molecule, which supports the twisting and constricting motions seen in the crystal structures.

It is enticing to draw parallels between the -7 to +4 series of individual subsites along the tunnel and a group of processive molecular motors such as kinesins or between the tunnel as a whole and the monomeric KIF1A kinesin postulated to move by biased diffusion (Okada & Hirokawa 1999, 2000). Within the TrCel7A tunnel, alternating attractive interactions at the front of the tunnel and repulsive interaction at the back of the tunnel could favor unidirectional motion. The monomeric KIF1A kinesin slides on microtubule after its tight interaction with a binding site is released upon a conformational change of the K-loop. The
conformational changes observed upon transition from the open to the closed state suggest that, along the reaction coordinates, TrCel7A may adopt a closed conformation to simultaneously destabilize the binding of cellobiose at the back of the tunnel while increasing the binding of the glucosyl moiety towards the entrance. We propose that during the repeated catalytic cycles of the processive TrCel7A enzyme, the CD tunnel cycles between closed (tense) and open (relaxed) conformations and this cycle promotes unidirectional motion along the cellodextrin chain. The closed or tense form is likely visited after glycosidic bond cleavage, promoting unidirectional motion by creating a pattern of interactions and repulsions with the cellulose strand that differ from the pattern required for the hydrolysis step. After the CD has slid by two glucosyl units, the tunnel loosens allowing for cellulose strand positioning and hydrolysis. Transition from the tense to relaxed forms would provide part of the energy required to separate the strand from the solid substrate.

The energy released by the tense to relaxed transition induced by the substrate could contribute to releasing the energy required to pull a single cellodextrin chain. Our structural observations therefore support the conclusion drawn from recent AFM studies, but also very early work by Ståhlberg et al (Ståhlberg et al. 1991), that the CD alone possesses the necessary machinery to process along crystalline cellulose. Both the energy released by the cleavage of the glycosidic bond (approximately -12.5 kJ/mol; ~-32.5kJ/mol if cellobiose concentration is kept low (Mulakala & Reilly 2005a)) and the energy released upon reorganization from a tense to a relaxed form drive cellulase processivity. Based on the data reported here we propose that the tunnel forming loops are actively engaged in extracting a single cellulose chain and that this process is linked to the hydrolysis step. Finally, the
significant movement of loop A4 may promote cellobiose expulsion and solvation. This is an important step as cellobiose solvation provides 5.73 kJ/mol that can contribute toward the activation barrier of 50-60 kJ/mol (Knott et al. 2014a; Brady et al. 2015).

Acknowledgements

We thank Gregg T. Beckham and Brandon C. Knott for providing the molecular dynamics simulations. We thank Dean Myles for his input to improve the manuscript.

This work was supported by the USDA National Institute of Food and Agriculture, Hatch project 211001, the National Science Foundation IGERT award #1069091 and the Graduate Opportunity (GO!) Program at Oak Ridge National Laboratory. Part of this work was conducted in the Center for Structural Molecular Biology supported by the Office of Biological and Environmental Research Project ERKP291, using facilities supported by the U.S. Department of Energy, managed by UT-Battelle, LLC under contract No. DE-AC05-00OR22725. Data were collected at Southeast Regional Collaborative Access Team (SER-CAT) beamlines at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.
2.6b References


## 2.7b Supplementary Information

Table 2.3 – Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Space group</strong></td>
<td>C2</td>
</tr>
<tr>
<td><strong>Cell dimensions</strong></td>
<td></td>
</tr>
<tr>
<td>$a, b, c$ (Å)</td>
<td>152.45, 44.83, 57.52</td>
</tr>
<tr>
<td>$\alpha, \beta, \gamma$ (°)</td>
<td>90.00, 101.11, 90.00</td>
</tr>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>42.94 - 1.60 (1.64 - 1.60)</td>
</tr>
<tr>
<td>$R_{merge}$</td>
<td>0.085 (0.181)</td>
</tr>
<tr>
<td>$I / \sigma I$</td>
<td>9.1 (3.9)</td>
</tr>
<tr>
<td><strong>Completeness (%)</strong></td>
<td>96.79 (93.17)</td>
</tr>
<tr>
<td><strong>Redundancy</strong></td>
<td>2.9 (2.9)</td>
</tr>
<tr>
<td><strong>Wilson B-factor (Å²)</strong></td>
<td>10.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resolution (Å)</strong></td>
<td>1.60</td>
</tr>
<tr>
<td><strong>Unique reflections</strong></td>
<td>142366</td>
</tr>
<tr>
<td>$R_{work} / R_{free}$</td>
<td>0.154 / 0.193</td>
</tr>
<tr>
<td><strong>No. atoms</strong></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>3254</td>
</tr>
<tr>
<td>Ligand/ion</td>
<td>46</td>
</tr>
<tr>
<td>Water</td>
<td>458</td>
</tr>
<tr>
<td><strong>R.m.s. deviations</strong></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td></td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td></td>
</tr>
<tr>
<td><strong>Ramachandran favored (%)</strong></td>
<td>98</td>
</tr>
<tr>
<td><strong>Ramachandran outliers (%)</strong></td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 2.11 – RMSD of different *Tr*Cel7A crystal structures.
CHAPTER 3

Small Angle Scattering Reveals pH and Cation Induced Conformational Changes for Neurospora crassa and Myriococcum thermophilum Cellobiose Dehydrogenase IIA

3.1 Abstract

Cellobiose dehydrogenases (CDH) have gained a renewed interest due to its potentially versatile applications in sectors from biofuel production to biomedical devices. The CDHIIA variant is comprised of a cytochrome domain (CYT), a dehydrogenase domain (DH), and a family 1 carbohydrate-binding module that are connected by flexible linkers. Cellobiose is oxidized at the DH domain with the electrons being transferred to the CYT domain. In the fungal cellulase system, CDHIIA subsequently donates electrons to a lytic polysaccharide monooxygenase (LPMO). Previous studies have shown a dramatic increase in interdomain electron transfer (IET) rate in the presence of cations. Interestingly, Neurospora crassa (Nc) CDHIIA does not have an increase in IET, while Myriococcum thermophilum (Mt) CDHIIA has a dramatic increase in activity when divalent cations are present even when at less than optimal pH conditions. Small angle scattering experiments on both MtCDHIIA and NcCDHIIA have been performed to characterize the structural rearrangements that occur in the presence of cations at different pH’s. The results show that NcCDHIIA undergoes conformational changes as the pH is changed, but is not significantly affected by cations. In contrast, MtCDHIIA goes through extensive structural rearrangements in the presence of cations, especially Ca$^{2+}$. 

83
3.2 Introduction

Cellobiose dehydrogenases (CDH) are co-secreted with a consortium of other cellulose degrading enzymes, by most wood-degrading fungi in the presence of cellulose. (Henriksson et al. 1995; Hallberg et al. 2000) CDH comprises greater than 2% by mole of the secreted cellulases in the fungal secretome indicating a natural importance of these enzymes to fungal degradation of cellulose. (Harreither et al. 2011; Phillips et al. 2011) When CDH genes have been deleted in *N. crassa*, a significant decrease in cellulase activity was observed. (Phillips et al. 2011) Upon addition of CDH from *M. thermophilum* to the *N. crassa* Δcdh-1 strain, cellulase activity was recovered. (Phillips et al. 2011) These results show the essential role of CDHs in cellulase cocktails and the “mix-and-match” optimization capabilities and possibilities for CDHs and LPMOs from various organisms.

All CDHs are comprised of an N-terminal cytochrome domain with a haem b

Figure 3.1 – Crystal structure of *Neurospora crassa* CDHIIA. Cytochrome domain, red; dehydrogenase domain, yellow; CBM, orange. (PDB ID 4Q17)
cofactor, a dehydrogenase domain containing a flavin adenine dinucleotide (FAD) cofactor. (Figure 3.1) The CDHIIA variant also has a C-terminal family 1 carbohydrate-binding module (CBM). Flexible linkers connect each of these domains. The dehydrogenase domain carries out a two-electron oxidation of cellobiose, or other small sugars, converting the cellobiose into cellubiose-1,5-lactone. Electrons are stored within the dehydrogenase domain FAD cofactor. These electrons are then passed one at a time to the haem group within the cytochrome domain for electron transfer to an electron acceptor, such as PMO. To complete this transfer of electrons, the domains must come together so that the redox centers are within 14 Å of each other. (Page et al. 1999)

Current applications of CDH range from biofuel production, sensor technology, to biomedical devices. (Thallinger et al. 2014; Cipri et al. 2016; Tegl et al. 2016; Wang & Lu 2016) Each of these applications requires CDH to maintain activity at various pH’s and under less than favorable enzymatic conditions, thus requiring a robust, tunable enzyme to operate under a spectrum of environments. Previous work grouped CDHs from ascomycetes into acidic, neutral, and alkaline based on their activity pH optimum. (Harreither et al. 2011) The alkaline category of CDHs has activity pH optima within a neutral to alkaline region. Acidic CDHs have a pH optimum around pH 5.0 and have a narrow pH range at which they are capable of functioning. (Harreither et al. 2011) MtCDH is characterized as an acidic CDH. (Harreither et al. 2007) The intermediate category of CDHs, in which NcCDH falls, have a pH optima between 5.0 and 6.0, but are capable of maintaining their activity at a broader pH range than acidic CDHs. (Harreither et al. 2011) This has been demonstrated by CDH maintaining its activity at pH 7.5 within 50-70% of its maximum activity.
Electrostatic mapping at the domain interface of the dehydrogenase and cytochrome domains using the recently solved crystal structures show a distinct difference between \textit{N. crassa} and \textit{M. thermophilum} CDH (Figure 3.2). For \textit{N. crassa} CDHIIA, the interface transitions from postitive (along the dehydrogenase domain, Fig 3.2E) at pH 5.5 to neutral at pH 7.5 (Fig 3.2F). While \textit{M. thermophilum} has an electronegative interface at the cytochrome domain that increases in electronegativity as the pH increases (Fig 3.2C-D) and electropositive and electronegative patches along the dehydrogenase domain. This distinction at the interface likely plays a role in variations in interdomain electron transfer (IET) rates and pH optima preferences between CDHs from different organisms.\cite{Harreither2011}

Recent experiments have shown that IET can be increased or regained in CDHIIA at pH conditions away from its optimal activity pH.\cite{Schulz2012, Kielb2015, Kracher2015} Initial work investigated immobilized class I CDH from \textit{Phanerochaete chrysosporium}, and class II CDHs from both \textit{Humicola insolens} and \textit{M. thermophilum}.\cite{Schulz2012} There was a two-fold increase in catalytic current for \textit{MtCDH} in the presence of monovalent cation salt, KCl, whereas a five-fold increase was observed when the divalent cation salt, \textit{CaCl}_2, was added to the buffer. Kracher \textit{et al} further investigated this phenomenon by varying cations based on their atomic radii and electronegativity and
measured changes in IET of MtCDH (MgCl$_2$, CaCl$_2$, SrCl$_2$, and BaCl$_2$ were tested). (Kracher et al. 2015) It was found that changing the atomic radius and electronegativity did not affect the IET rate. All CDHs had increased IET rates when away from optimal pH conditions in the presence of CaCl$_2$ except for the Sclerotium rolfsii and N. crassa CDHs. Based on these activity studies and docking models, the authors hypothesized that the cations undergo complexation with CDHIIA at the interface of the cytochrome and dehydrogenase domains, allowing the domains to more easily come together and electrons to be transferred at a more efficient rate. (Schulz et al. 2012; Kracher et al. 2015)

Recent work determined the structure of full-length MtCDHIIA and NcCDHIIA. (Tan et al. 2015) MtCDH was captured in a closed conformation, while NcCDHIIA maintained an open conformation. Tan et al noted the high flexibility of NcCDHIIA and difficulty of obtaining electron density for the cytochrome domain. This observation suggests a continuum of conformations are visited. Small-angle X-ray scattering (SAXS) work was subsequently performed to understand the conformational landscape of these two enzymes. Based on ensemble optimization modeling (EOM) from the SAXS data, the data confirmed that the CDHs existed in an ensemble of confirmations. EOM analysis also concluded that the presence of the inhibitor cellobiose-1,5-lactam reduced conformational variability. Interestingly, this indicates that the presence of an inhibitor molecule in the binding pocket of the dehydrogenase domain limits the flexibility of the entire enzyme.

In this work, we aim to structurally characterize the effect of pH, cellobiose, and cations on the CDHs from M. thermophilium and N. crassa. Through small-angle scattering, one can elucidate the conformational rearrangements that are occurring in solution due to
changes in the environment. By using the recently solved structures of MtCDHIIA and
NcCDHIIA for SAS modeling, a multi-state model, when appropriate, in each condition is
proposed.

3.3 Materials and Methods

3.3.1 Protein expression and purification

NcCDHIIA and MtCDHIIA were both expressed in KM71H Pichia pastoris cells
(Invitrogen). The Invitrogen EasySelect Pichia protein expression protocol was followed for
protein expression. (Invitrogen 2010) Since KM71H is a Mut$^S$ strain of P. pastoris, the Mut$^S$
protocol was followed. Both CDHIIAs were expressed and purified following the same
protocol previously established by Invitrogen and Sygmund et al, respectively. (Invitrogen
2010; Sygmund et al. 2012) Cells were grown in buffered complex glycerol (BMGY) media
overnight at 30 °C to an OD$_{600}$ between 4 and 6. Cells were spun down at 5,000 x g for 5
minutes at room temperature to remove the BMGY. The cells were resuspended in buffered
complex methanol (BMMY) media between one-tenth to one-fifth the original
culture/BMGY volume to induce expression. 100% methanol was added to a final
concentration of 0.5% 24 hrs after induction. After 48 hrs of induction, the culture was spun
down at 5,000 x g for 10 minutes at room temperature. The supernatant was decanted and
ammonium sulfate was slowly added to a final concentration of 20% saturated solution. After
addition of the ammonium sulfate, the supernatant was spun at 5,000 x g for 15 minutes at
room temperature. The supernatant from the ammonium sulfate cut was steri-filtered with a
0.2 µm PES filter (Nalgene).
The filtered supernatant was applied to a Phenyl Sepharose HP column (GE Healthcare). CDHIIA was eluted using a linear gradient of 20% (NH$_4$)$_2$SO$_4$ to 0% (NH$_4$)$_2$SO$_4$. The buffer was exchanged to 50 mM sodium acetate, pH 5.5 then applied to a Mono Q HR column (Pharmacia). CDHIIA eluted with a linear gradient from 0 M NaCl to 1 M NaCl. A Superdex 200 column (GE Healthcare) was used as a final polishing and buffer exchange step. Molar absorption coefficients for $\text{NcCDH}$ and $\text{MtCDH}$ are $\varepsilon_{280} = 174$ mM$^{-1}$ cm$^{-1}$ and $\varepsilon_{280} = 159$ mM$^{-1}$ cm$^{-1}$, respectively.

| Table 3.1 – Buffer conditions tested for $\text{MtCDH}$ (M) and $\text{NcCDH}$ (N). |
|------------------|------------------|------------------|------------------|
| **50 mM NaOAc, pH 5.5** | **50 mM PIPES, pH 7.5** |
| 20 mM KCl (M / N) | 20 mM KCl (M / N) |
| 40 mM KCl (M / N) | 40 mM KCl (M / N) |
| 60 mM KCl (M / N) | 60 mM KCl (M / N) |
| 20 mM CaCl$_2$ (M / N) | 20 mM CaCl$_2$ (M / N) |
| 40 mM CaCl$_2$ (M / N) | 40 mM CaCl$_2$ (M / N) |
| 60 mM CaCl$_2$ (M / N) | 60 mM CaCl$_2$ (M / N) |
| 10 mM cellobiose (N) | – |
| 25 mM cellobiose (N) | – |
| 50 mM cellobiose (N) | – |
| 100 mM cellobiose (N) | – |
3.3.2 Small angle neutron scattering (SANS) sample preparation, data collection and reduction

NcCDHIIA was measured using the Bio-SANS instrument at the High Flux Isotope Reactor of Oak Ridge National Laboratory. (Heller et al. 2014) Buffers were exchanged using a 30 kD molecular weight cutoff (MWCO) concentrator (Vivaspin) followed by spin-filtering with 0.2 µm nylon membrane filter (VWR) the protein (Table 3.1 for buffer conditions tested). 50 mM sodium acetate (NaOAc), pH 5.5, and 50 mM piperazine-N,N’-bis(2-ethanesulfonic acid) (PIPES), pH 7.5, were used. All samples were measured in 2 mm quartz banjo cells at ambient temperature. Measurements were performed with protein concentrations of 2 mg mL⁻¹ and 100% D₂O buffers to minimize incoherent background and decrease measurement times. Detector distances of 0.3 and 6 m with a wavelength of 6 Å provided an adequate q range of 0.008 to 0.789 Å⁻¹. Samples were measured for 20 min and 90 min at the 0.3 and 6 m detector configurations, respectively. Data were corrected for transmission, detector sensitivity, dark current, and sample background. Data were reduced using MantidPlot. (Arnold et al. 2014)

3.3.3 Small angle X-ray scattering sample preparation, data collection and reduction

Buffers were exchanged with a 30 kD MWCO concentrator (Vivaspin) followed by spin-filtering with 0.2 µm nylon membrane filter (VWR) prior to measurements (Table 3.1 for conditions measured). 50 mM sodium acetate (NaOAc), pH 5.5, and 50 mM piperazine-N,N’-bis(2-ethanesulfonic acid) (PIPES), pH 7.5, were used. MtCDHIIA was measured on a Rigaku Bio-SAXS 2000 at the Center for Structural Molecular Biology in Oak Ridge National Laboratory. The instrument uses Cu Kα radiation (λ = 1.54 Å) and was collimated to
provide a sufficient q range of 0.01 Å to 0.67 Å. Measurements were done at ambient temperature. Samples were measured for 40 min with the detector refreshing every 5 min. Data were corrected for transmission and sample background. Reduction, averaging, and buffer subtraction were done with Rigaku SAXS Lab 3.1.0b14.

3.3.4 – Data evaluation and modeling

The Guinier analysis and P(r) were analyzed using PRIMUS and GNOM, respectively. (Svergun 1992; Konarev et al. 2003) The missing linker region was built into MtCDHIIA using MODELLER. (Yang et al. 2012) This step was not required for NcCDHIIA. Glycans were added onto both CDHs with the glycam server. (Woods 2005-2016) For the N-linked glycans, 2 N-acetylglucosamines with 8 mannoses were modeled at glycosylation sites identified from the crystal structure. (Brethauer & Castellino 1999) On O-linked glycosylation sites, 4 mannoses were attached at glycosylation sites based on the crystal structure. Glycan patterns were chosen based on core structures produced by P. pastoris. (Brethauer & Castellino 1999) Modeling of the small angle scattering data was done on the MultiFoXS server. (Schneidman-Duhovny et al. 2016) 10,000 conformations were sampled with an additional rigid body connection file used to maintain heme coordination at the cytochrome domain. MultiFoXS follows the rapidly exploring random tree search with calculating SAXS profiles for each generated model. Multi-state model enumeration and scoring based on the SAXS profile was calculated by minimizing the chi score:

$$
\chi = \sqrt{\frac{1}{S} \sum_{i=1}^{s} \left( \frac{I_{\text{exp}}(q_i) - c \sum_{n} w_n I_n(q_i,c_1,c_2)}{\sigma(q_i)} \right)^2}
$$
where $S$, $I_{\text{exp}}(q)$, $c$, $w_n$, $I_n(q, c_1, c_2)$, and $\sigma(q)$ are the number of points in the profile, the experimental profile, scale factor, weight given to the calculated profile, calculated profile, and experimental error of the measured profile. The multi-state model in each condition was chosen based on the lowest chi value and deviation.

3.3.5 – Sedimentation velocity

Analytical ultracentrifugation (AUC) was performed on a Beckman XL-I. Sedimentation velocity (SV) data were measured at wavelengths at 250 nm, 280 nm, and 420 nm. CDH sample concentrations were measured at 6 µM and spun at 50,000 rpm at 18 °C. SEDFIT was used to fit data using the c(S) distribution model of the Lamm equation.

3.4 Results

To ensure the purity of the samples and that they were free of aggregates, SDS-

Figure 3.3 – CDHIIA samples tested for purity and aggregates before SAS experiments. SDS-PAGE gels of NcCDHIIA (left) and MtCDHIIA (middle). SV showed NcCDH sample (right).
PAGE, UV spectroscopy, and analytical ultracentrifugation (AUC) were conducted. (Figure 3.3) RZ values, the absorbance ratio of $A_{420}/A_{280}$ (absorbance peak for the oxidized heme/absorbance peak of the protein in the sample) for both CDHIIA samples were 0.60, indicating sufficient purity in accordance with previously reported work.(Harreither et al. 2011; Sygmund et al. 2012)

3.4.1 – pH Effect on MtCDH and NcCDH

Earlier in this study, NcCDHIIA was measured on the SIBYLS SAXS beam line at the Advanced Light Source, Lawrence Berkeley National Laboratory. Analysis of the data found significant increases in the radius of gyration ($R_g$) and maximum dimension ($d_{max}$) compared to preliminary measurements on a home-source SAXS instrument and the Bio-SANS beam line (performed for initial characterization and determination of potential D$_2$O-induced aggregation). This caused concerns of photoreduction of the heme and/or FAD cofactors, which would cause structural rearrangements in the enzyme. This phenomenon has been observed in cytochrome P450 reductase.(Huang et al. 2013) The other potential reason for the increases in the dimensions of the particle are X-ray induced aggregation due to

<table>
<thead>
<tr>
<th></th>
<th>R$_g$ (Å)</th>
<th>d$_{max}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nc pH 5.5</td>
<td>39.15</td>
<td>147</td>
</tr>
<tr>
<td>Nc pH 7.5</td>
<td>40.56</td>
<td>145</td>
</tr>
<tr>
<td>Mt pH 5.5</td>
<td>37.25</td>
<td>117</td>
</tr>
<tr>
<td>Mt pH 7.4</td>
<td>36.95</td>
<td>111</td>
</tr>
</tbody>
</table>
breaking of bonds and formation of radicals that has been described by Kuwamoto et al. (Kuwamoto et al. 2004) This led to cautious initial measurements of NcCDHIIA on the BioSANS instrument to ensure no X-ray radiation damage to the cofactors and protein.

The SANS and SAXS data from NcCDHIIA and MtCDHIIA, respectively, are plotted on a log(I) vs log(q) scale in Figure 3.4. Both CDHs were measured at 2 mg mL\(^{-1}\). Proteins were investigated at pH 5.5, a pH where both enzymes are optimally active, and pH 7.5, a condition where both enzymes are less active. Dissimilarities between the two proteins can be seen by the presence of a shoulder in the MtCDHIIA at q = 0.12 Å that is not observed in the NcCDHIIA data. Guinier plots used all data possible and showed linear trends, indicating that each sample was free of aggregates and interparticle interactions (Figure 3.5). The R\(_g\)'s determined by the Guinier analysis for NcCDH at pH 5.5 and 7.5 were 37.57 ± 0.25 Å and 38.70 ± 0.26 Å, respectively. For MtCDH, the R\(_g\)'s at pH 5.5 and 7.5 was 36.65 ± 0.21 Å and 34.19 ± 0.20 Å, respectively.

Figure 3.4 – Guinier plot of Mt (pH 5.5, red; pH 7.4, orange) and NcCDH (pH 5.5, green; pH 7.5, blue). Curves are offset for clarity.
Figure 3.5 – Scattering curves of *MtCDH* (pH 5.5, red; pH 7.4, orange) and *NcCDH* (pH 5.5, green; blue) accompanied with the P(r) curves).
P(r) profiles obtained from the data were calculated by GNOM (Figure 3.4, summary of dimensions in Table 3.2). MtCDHIIA shows a very slight compaction going from the optimal activity pH 5.5 to the less favorable pH 7.5 as indicated by a decrease in the \(d_{\text{max}}\) from 117 Å to 111 Å and a negligible decrease in \(R_g\) from \(37.25 \pm 0.13\) Å to \(36.94 \pm 0.08\) Å. An increase in pH led to minimal changes in \(NcCDH\) \(d_{\text{max}}\) (147 Å to 145 Å) and \(R_g\) (39.15 ± 0.45 Å to 40.56 ± 0.39 Å). A prominent shoulder located around 75 Å is observed in the \(MtCDH\) P(r) curve at both pH conditions. The most interesting features arise in the \(NcCDH\) transition from pH 5.5 to 7.5. A shoulder in the P(r) curve arises at approximately 75 Å that is not seen in the pH 5.5 P(r) curve, indicating structural rearrangements are occurring when the pH is shifted to a less favorable pH.

To better understand the changes in the different pH samples with \(NcCDH\), modeling of the SAS data was performed. MultiFoXS generates the desired number of conformations of the input pdb. The generated conformers have their scattering curves calculated. The calculated curves are weighted and summed together to produce a curve as close to the experimental curve as possible. At pH 5.5, a three state model fits to the data best indicating that there were three predominant conformations visited (Figure 3.6). The predominant state had an \(R_g\) of 36.67 Å that was visited 60% of the time. The two other states were an extended and compact conformer with \(R_g\)’s of 46.36 Å and 31.85 Å that comprised 30.3% and 9.8%, respectively, of the other conformers. At pH 7.5 there was a disappearance of the compact conformer with the two states’ \(R_g\)’s at 38.48 Å and 47.20 Å existing 75.9 % and 24.1% of the time. The chi values were high on these due to a poor fit at the high q region (chi values for pH 5.5 and 7.5 were 1.77 and 2.44, respectively). This is potentially due to the difference in
scattering of the hydration shell that exists between SANS and SAXS for which the SAXS modeling software could not account. However, these models aid in explaining that the formation of the shoulder, which indicates there is an increase in the distance between the cytochrome and dehydrogenase domains at a less favorable pH.

Figure 3.6 – Multistate modeling of NcCDHIIA. Top, pH 5.5 represented as a three state model with the most prevalent conformer in black ($R_g = 36.67 \text{ Å}$, $w = 60\%$) lightening to the less visited states (gray: $R_g = 46.36 \text{ Å}$, $w = 30.3\%$; light gray: $R_g = 31.85 \text{ Å}$, $w = 9.8\%$). Bottom, pH 7.5 represented as a two state model with black being the most visited conformer ($R_g = 38.48 \text{ Å}$, $w = 75.9\%$) and light gray being the less visited state ($R_g = 37.20 \text{ Å}$, $w = 24.1\%$). To the right of both models are the accompanying calculated profiles overlayed with the experimental profiles. Glycans and cofactors not shown for ease of visualization.
3.4.2 – Effect of monovalent cations on CDH

The monovalent cation, K\(^+\), was measured in increasing concentrations up to 60 mM. Previous studies observed a two-fold activity increase with KCl concentrations up to 100 mM at pH 5.5.\(^{(Schulz et al. 2012)}\) To understand if pH played a role in KCl structural rearrangements, pH at 5.5 and 7.5 were measured with KCl concentrations increasing to 60 mM.

SANS curves of the KCl series for \(NcCDHIIA\) are plotted on a log(I) vs log(q) scale in figure 3.7. For \(NcCDHIIA\), the SANS profiles at both pH 5.5 and 7.5, the SANS profiles have a similar trend with very little change in the curve. Guinier plots utilized all possible data points and showed linear trends for each sample, except for the 20 mM KCl at pH 7.5 sample. This sample had an upturn at low q, indicating that aggregates were present. This measurement was excluded from further analysis. The P(r) profiles show a lack of conformational change as indicated by a lack of change in the overall curvature of the P(r) profile (Figure 3.7 and Table 3.3). This is not surprising since the electrostatic mapping indicates a positively charged domain interface at pH 5.5 that becomes neutral at pH 7.5, so

<table>
<thead>
<tr>
<th>Table 3.3 – Effect of K(^+) on (NcCDH) dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_g) (Å)</td>
</tr>
<tr>
<td>pH 5.5</td>
</tr>
<tr>
<td>0 mM K(^+)</td>
</tr>
<tr>
<td>20 mM K(^+)</td>
</tr>
<tr>
<td>40 mM K(^+)</td>
</tr>
<tr>
<td>60 mM K(^+)</td>
</tr>
</tbody>
</table>
Figure 3.7 – NcCDH SANS data on a log-log plot next to the respective P(r) curves. Red, 0 mM K⁺; orange, 20 mM K⁺; green, 40 mM K⁺; blue, 60 mM K⁺.
the addition of cations would not interact with residues at the interface.

SAXS curves, plotted on a log(I) vs log(q) scale, for the MtCDHIIA KCl series are shown in Figure 3.8. The Guinier analysis for all samples showed linear trends, which signify that the samples were free of aggregates and interparticle interactions. At pH 5.5, the $R_g$ determined from the Guinier region slightly increased as KCl increased to 60 mM (from $36.65 \pm 0.21$ Å at 0 mM K$^+$ to $38.18 \pm 0.35$ Å at 60 mM K$^+$). A similar trend was observed at pH 7.5, the $R_g$ increased from $34.19 \pm 0.20$ Å to $37.43 \pm 0.29$ Å as the K$^+$ concentration increased from 0 mM to 60 mM. The P(r) curves at pH 5.5 show a gradual and subtle increase in the $d_{max}$ (from 117 Å at 0 mM KCl to 129 Å at 60 mM KCl) as the KCl concentration increases (Figure 3.8 and summarized in Table 3.4). The P(r) profiles also indicate that very slight structural rearrangements are occurring as indicated by a slight decrease in the prominence of the shoulder present around 75 Å. However, at pH 7.5 the change is a little less gradual. A change in $d_{max}$ is not seen until 60 mM KCl is added where the $d_{max}$ changes from 111 Å to 124 Å (Figure 3.7). The conformational changes are subtle at pH 7.5 as well. Very slight changes in the shoulder are seen at 20 mM and 40 mM KCl, but

<table>
<thead>
<tr>
<th>Table 3.4 – Effect of K$^+$ on MtCDH dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_g$ (Å)</td>
</tr>
<tr>
<td>pH 5.5</td>
</tr>
<tr>
<td>0 mM K$^+$</td>
</tr>
<tr>
<td>20 mM K$^+$</td>
</tr>
<tr>
<td>40 mM K$^+$</td>
</tr>
<tr>
<td>60 mM K$^+$</td>
</tr>
</tbody>
</table>
Figure 3.8 – *MtCDH* SAXS data plotted on a log-log plot next to the respective P(r) curves. Red, 0 mM K⁺; orange, 20 mM K⁺; green, 40 mM K⁺; blue, 60 mM K⁺.
the most noticeable alteration occurs at 60 mM KCl where the shoulder becomes slightly less prominent. Indicating greater amounts of KCl are required to elicit a structural conformational change at less favorable pH condition.

Modeling of *Mt*CDH shows that for both pH conditions there was a departure from the compact conformation (an R\textsubscript{g} of approximately 33.6 Å) that comprised 31.9% and 40.3% for pH 5.5 and pH 7.4, respectively, of the present conformers upon addition of 20 mM K\textsuperscript{+}. At pH 5.5, the predominant conformer had an R\textsubscript{g} of 39.22 Å at 20 mM K\textsuperscript{+} that decreased to 36.03 Å at 60 mM K\textsuperscript{−}(Figure 3.9). At pH 7.4, there was a disappearance of the compact conformer, but the presence of the extended conformation (R\textsubscript{g} > 38 Å) significantly decreased by 32%. The results of the modeling indicate that higher K\textsuperscript{+} concentrations will aid in bringing the two domains closer together and enable increased conformational flexibility.

3.4.3 – *Effect of divalent cations on CDH*

The divalent salt, CaCl\textsubscript{2}, was added in 20 mM increments to a final concentration of 60 mM. To understand if there was a pH dependence on CaCl\textsubscript{2} structural response, the Ca\textsuperscript{2+} series was performed at pH 5.5 and pH 7.5.

SANS data curves for *Nc*CDHIIA are plotted on log(I) vs log(q) scale in Figure 3.10. The Guinier analysis of the pH 5.5 and pH 7.5 CaCl\textsubscript{2} series showed linear trends at low q indicating that each sample was free of aggregation and interparticle interactions. The
Figure 3.9 – Best modeling solutions of MtCDHIIA resulted in 2 state solutions for pH 5.5. The most prevalent state is shown in black with the less prevalent state shown in gray. Top shows modeling results for 0 mM K$^+$ (black: $R_g = 37.88$ Å, $w = 68.1$%; light gray: $R_g = 33.76$ Å, $w = 31.9$%). Middle shows modeling results for 20 mM K$^+$ (black: $R_g = 39.22$ Å, $w = 70.8$%; light gray: $R_g = 35.07$ Å, $w = 29.2$%). Bottom, modeling results for 60 mM K$^+$ (black: $R_g = 36.03$ Å, $w = 94.2$%; light gray: $R_g = 50.76$ Å, $w = 5.8$%). To the right of the models are the accompanying calculated profiles overlaid with the experimental profiles. Glycans and cofactors not shown for ease of visualization.
Figure 3.10 – $N_c$CDH Ca$_{2+}$ series SANS data on a log-log plot with the respective P(r) curves. Red, 0 mM Ca$^{2+}$; orange, 20 mM Ca$^{2+}$; green, 40 mM Ca$^{2+}$; blue, 60 mM Ca$^{2+}$.
addition of Ca\(^{2+}\) did not affect the overall conformation of NcCDH whether it was at pH 5.5 or 7.5 (Figure 3.10 and summary in Table 3.5). Like the K\(^+\) series, this makes biochemical sense, due to the electrostatics of the NcCDH domain interface.

| Table 3.5 – Effect of Ca\(^{2+}\) on NcCDH dimensions |
|-----------------------------------------------|---------------|---------------|---------------|---------------|
| pH 5.5 | pH 7.5 | pH 5.5 | pH 7.5 |
| R\(_g\) (Å) | d\(_{\text{max}}\) (Å) | R\(_g\) (Å) | d\(_{\text{max}}\) (Å) |
| 0 mM Ca\(^{2+}\) | 39.15 | 147 | 40.56 | 145 |
| 20 mM Ca\(^{2+}\) | 40.08 | 148 | 40.54 | 144 |
| 40 mM Ca\(^{2+}\) | 39.77 | 146 | 40.57 | 145 |
| 60 mM Ca\(^{2+}\) | 39.98 | 154 | 40.86 | 148 |

The SAXS data for the MtCDHIIA are plotted on a log(I) vs log(q) scale in Figure 3.11. Guinier analysis for both the pH 5.5 and pH 7.4 CaCl\(_2\) series had linear trends indicating all samples were free of aggregates and interparticle interactions. Based on the P(r) profiles for pH 5.5, the initial addition of 20 mM Ca\(^{2+}\) led to a dramatic increase of the d\(_{\text{max}}\) from 117 Å to 130 Å (Figure 3.11 with results summarized in Table 3.6). Interestingly, the R\(_g\) only increased from 37.25 Å to 37.56 Å. At 40 mM Ca, there is a slight increase in R\(_g\) (38.86 Å) with an additional increase in d\(_{\text{max}}\) (146 Å). At the final Ca\(^{2+}\) concentration, there was a slight decrease in the R\(_g\) (38.24 Å) and d\(_{\text{max}}\) (138 Å). Based on the pH 5.5 data from the divalent cation series, the presence of 20 mM Ca\(^{2+}\) causes significant structural rearrangements shown by a significant increase in d\(_{\text{max}}\), while a decrease in the dimensions are observed at 60 mM CaCl\(_2\).
The P(r) profiles for the Ca\(^{2+}\) series at pH 7.4 have a similar trend as the Ca\(^{2+}\) series at pH 5.5 (Figure 3.11 and Table 3.5). Upon addition of Ca\(^{2+}\), the R\(_g\) and d\(_{\text{max}}\) increased from 36.94 Å and 111 Å to 38.09 Å and 145 Å. As Ca\(^{2+}\) increases from 40 mM to 60 mM, a similar decrease in d\(_{\text{max}}\) is observed (147 Å to 132 Å). The R\(_g\) appears to very slightly increase, but both of the R\(_g\) values are within the error of the other (38.06 ± 0.34 Å and 38.35 ± 0.31 Å for 40 mM and 60 mM Ca\(^{2+}\), respectively). These results indicate that Ca\(^{2+}\) induces significant conformational changes in MtCDHIIA at pH 7.4.

To gain additional insight on the structural changes caused by Ca\(^{2+}\), multi-state modeling was performed on the MtCDH SAXS data. At pH 5.5, the initial addition of 20 mM Ca\(^{2+}\), leads to the appearance of an extended conformer with a R\(_g\) of 47.33 Å and disappearance of the more compact conformer present at no salt (R\(_g\) = 33.76 Å) (Figure 3.12). However, this open state at 20 mM Ca\(^{2+}\) is present in 9.1% of the conformers. At 60 mM Ca\(^{2+}\) the largest R\(_g\) is 38.88 Å (dominating 14.2% of the overall structures), while the compact conformer at 33.19 Å reappears (occurring in 17.2% of the time). At pH 7.4, the

| Table 3.6 – Effect of Ca\(^{2+}\) on MtCDH dimensions |
|----------------|----------------|----------------|----------------|
|                | R\(_g\) (Å) | d\(_{\text{max}}\) (Å) | R\(_g\) (Å) | d\(_{\text{max}}\) (Å) |
| pH 5.5         | pH 7.4       | pH 5.5         | pH 7.4       |
| 0 mM Ca\(^{2+}\) | 37.25        | 117            | 36.95        | 111            |
| 20 mM Ca\(^{2+}\) | 37.56        | 130            | 38.09        | 145            |
| 40 mM Ca\(^{2+}\) | 38.68        | 146            | 38.06        | 147            |
| 60 mM Ca\(^{2+}\) | 38.24        | 138            | 38.35        | 132            |
Figure 3.11 – *MtCDH* Ca$^{2+}$ series SAXS data on a log-log plot next to the respective $P(r)$ curves. Red, 0 mM Ca$^{2+}$; orange, 20 mM Ca$^{2+}$; green, 40 mM Ca$^{2+}$; blue, 60 mM Ca$^{2+}$.
predominant states have R_g’s of 38.83 Å and 33.61 Å (occurring 48.8% and 40.3 % of the time), with an intermediate state of 36.17 Å comprising 10.9% of the conformers in the sample. There is no significant change in the R_g distribution upon the initial addition of Ca^{2+}. At 60 mM Ca^{2+}, the most prevalent conformer has a R_g of 34.98 Å, but the open state R_g massively increases to 49.88 Å. The trends show that as the Ca^{2+} concentration increases, the most prevalent conformers show a decrease in R_g; however, as Ca^{2+} is added, the presence of massively extended conformations appear. This would suggest that the presence of Ca^{2+} predominately acts to bring the domains together. Ca^{2+} also increases the enzyme’s conformational variability.

3.4.4 – Effect of cellobiose on CDH

Increasing amounts of cellobiose were added to NcCDHIIA in 50 mM sodium acetate, pH 5.5, to determine if substrate induced conformational changes within NcCDHIIA. SANS data is plotted on a log(I)vs log(q) scale in Figure 3.13. The Guinier analysis of the cellobiose series shows linear trends indicating there are no signs of interparticle interactions or aggregation. The P(r) profiles indicate that there are no significant changes in the dimensions when cellobiose is added in concentrations up to 100 mM (Figure 3.13). While the P(r) did not show signs of structural changes, the Kratky plot shed some light on what may be occurring (Figure 3.13). As the concentration of the cellobiose increases, the Kratky plot decays at slower rates and converge to the x-axis at concentrations greater than 25 mM cellobiose. This implies that higher concentrations of cellobiose induce conformational flexibility within NcCDHIIA.
Discussion

Previous work on both *M. thermophilum* and *N. crassa* CDHIIA showed strong pH dependence for both interdomain and direct electron transfer pathways. (Harreither *et al.* 2007; Harreither *et al.* 2012) Harreither *et al.* initial characterization of CDHs into groups
Figure 3.13 – *NcCDH* cellobiose SANS data plotted on log-log axis (top left). P(r) curve indicates little change in dimensions (top right). Kratky plot shows increase in flexibility with increase in substrate (bottom).

Based on their optimal activity pHs provided some preliminary insight on potential differences between *NcCDH*, an intermediate CDH, and *MtCDH*, an acidic CDH. (Harreither et al. 2011) Electrostatic mapping of the interface between the dehydrogenase and cytochrome domains shed structural light on why *NcCDH* and *MtCDH* would be in different pH optima groups. *NcCDH* dehydrogenase domain shifts from a positive to neutral interface when the pH increases from 5.5 to 7.5. Compared to *MtCDH* that has a negatively charged cytochrome domain that becomes slightly more negative when the pH is increased. These studies were performed to gain a better understanding on how various conditions, such as pH, substrate, and cations, affect these proteins differently.
The \textit{NcCDH} SAS data indicated that as the pH increased from the optimal activity pH of 5.5 to the less favorable activity pH 7.5, \textit{NcCDH} undergoes conformational rearrangements. This was observed by the formation of a shoulder in the P(r) profile as the pH increased from 5.5 to 7.5, which can indicate a shift from a globular, compact state to the separation of two domains. Modeling showed that the pH optimum favors a more closed state, while the less optimal pH leads to a more open CDH. These results indicate that as the pH increases, the attractive charges decrease between the cytochrome and dehydrogenase domains causing a natural opening of the enzyme, rather than repulsive charges driving the two domains apart. This would explain how \textit{NcCDH} is capable of maintaining activity at a higher pH.

The \textit{MtCDH} SAXS data showed that the pH did not appreciably affect the conformation of the enzyme. There was a slight decrease in the $d_{\text{max}}$, but the overall P(r) profile indicated no significant changes in the domain organization. Previous biochemical studies showed \textit{MtCDH} to have a dramatic drop in activity when the pH deviated away from the pH optima of 5.5. The electrostatic calculations show that cytochrome domain is negatively charged at pH 5.5, but becomes extremely negatively charged at pH 7.5. However, this change in charge does not significantly affect the distance between the dehydrogenase and cytochrome domains. This could indicate that there is a more defined interface required for IET to occur that changes in response to pH. Together these results help explain why these two seemingly very similar enzymes respond differently to changes in the pH.

From here we proceeded to investigate how monovalent and cations affected the conformational landscape of both \textit{Nc-} and \textit{MtCDH}. Spectroscopic-based structural
observations concluded that cations cause domain reorganization within MtCDH, but no
definitive structural work has shown how cations affect CDH. (Kielb et al. 2015) For NcCDH
at both pH 5.5 and 7.5, no structural changes were observed when K\(^+\) and Ca\(^{2+}\) were added.
Based on the fairly neutral charge of the \textit{N. crassa} domain interface with the dehydrogenase
domain being positively charged at pH 5.5, it is obvious as to why the addition of cations
would not affect the structure of the protein. These results correlate with biochemical assays
that found \textit{NcCDHIIA} to be unresponsive to cations. (Kracher et al. 2015)

However, \textit{MtCDH} had the largest structural rearrangements in response to cations.
Upon addition of K\(^+\) slight conformational changes occurred that directly corresponded with
previous observations where only moderate changes in IET were noted. (Schulz et al. 2012)
An increase in pH required higher concentrations of K\(^+\) to cause a change in the
conformation. The more dramatic change was observed when Ca\(^{2+}\) was added to \textit{MtCDH}. A
substantial increase in the \(d_{\text{max}}\) with negligible changes in the \(R_g\) indicated that Ca\(^{2+}\) induced
significant structural reorganization. Multistate modeling showed that the predominant
conformer was a more compact shape as cations were added, but an extended state was
present as well. This indicated that cations increased the conformational freedom of \textit{MtCDH},
but primarily promoted compact conformers.

Finally, the effect of substrate interactions was investigated. The \textit{NcCDH} SANS data
shows that as the concentration of cellobiose increases there is an increase in flexibility.
Igarashi \textit{et al} found that cellobiose bound to the active site inhibited electron transfer from
the flavin to the heme in \textit{Phanerochaete chrysosporium}. (Igarashi \textit{et al.} 2002) This would
indicate that the release of the cellobionolactone product from the dehydrogenase domain is
required for IET to occur. Our SANS results suggest that in the presence of a substrate, CDH will become slightly more flexible to allow the substrate to easily enter the dehydrogenase domain’s binding pocket. Tan et al suggested the presence of an inhibitor at the dehydrogenase domain would decrease conformational variability based on EOM results, which aligns with the Kratky analysis of our NcCDH on the substrate.(Tan et al., 2015)

3.6 Conclusions

From these studies a clearer understanding on how the environment affects these CDHs has been gained. Previous work found that Nc- and MtCDHIIA were both affected by pH with MtCDHIIA being significantly more sensitive to changes in the pH.(Harreither et al, 2011) The data showed conformational rearrangements occurring for NcCDHIIA when transitioning from the optimal pH to the less catalytically active pH, while MtCDHIIA did not undergo structural changes. This finding may indicate that there is a more defined electron transfer interface for MtCDHIIA due to the highly charged cytochrome domain, while the dispersed positively charged dehydrogenase domain at NcCDHIIA interface may allow for a less specific interaction for IET to occur. The presence of cellobiose increases the flexibility of NcCDHIIA, which may allow for ease of substrate entry to and/or product release from the binding pocket of the dehydrogenase domain. Definitive structural evidence now shows that cations, especially Ca^{2+}, allow MtCDHIIA to undergo significant conformational rearrangements, which would allow for increased IET rates. By understanding the structural implications of changing the environment, a more targeted approach on optimizing CDHs and their interface can be pursued.
3.7 References


CHAPTER 4

Investigation of the *Neurospora crassa* Cellobiose Dehydrogenase IIA and *Neurospora crassa* Lytic Polysaccharide Monooxygenase-2 Complex

4.1 Abstract

One of the mechanisms fungi utilize to degrade cellulose is oxidative cleavage of glycosidic bonds. Lytic polysaccharide monooxygenase (LPMO) and cellobiose dehydrogenase (CDH) are two secreted cellulases that work together to degrade cellulose through oxidation. The CDHIIA variant is comprised of a cytochrome domain, dehydrogenase domain, and a family 1 carbohydrate-binding module (CBM). CDH oxidizes short, soluble sugars, specifically cellobiose, at the dehydrogenase domain. Electrons are transferred from the dehydrogenase domain to the cytochrome domain. The cytochrome domain then shuttles electrons to LPMO. The reduced copper ion in LPMO is then capable of oxidizing cellulose. Due to the transient interaction of CDH and LPMO, structural studies of the complex are challenging. Here, the complex was studied from the perspective of how *Neurospora crassa* CDHIIA structurally responds to interactions with *Neurospora crassa* LPMO-2 in solution. Small angle neutron scattering (SANS) enables components of a system to be masked by varying the concentration of D$_2$O in the buffer. This capability of SANS was utilized to explore CDHIIA conformational reorganization in the presence of LPMO. The match point of partially deuterated NcLPMO was determined to be 69% D$_2$O. When the CDH-LPMO complex was measured, significant decreases in the radius of gyration ($R_g$) and maximum dimension ($d_{\text{max}}$) were observed. These results indicate that upon binding of LPMO, CDH undergoes
conformational rearrangements, specifically a compaction. This research provides new information on the binding interactions in the transient CDH-LPMO complex.

4.2 Introduction

The cellulose and other polysaccharides present in biomass are appealing avenues for renewable materials and biofuels. Currently, biorefineries utilize cellulose-degrading enzymes to breakdown the cellulose into monomeric glucose for downstream production into biofuels. However, one of the biggest hurdles with cellulose degradation is overcoming its crystallinity, which makes it recalcitrant to degradation and enzymatic attack.

Fungal cellulases frequently utilized for cellulose breakdown employ both hydrolytic and oxidative processes. (Payne et al. 2015) Hydrolytic enzymes include: cellobiohydrolases (CBH) that processively hydrolyze cellulose strands into cellobiose; endoglucanases (EG) cleave amorphous cellulose; and β-glucosidases that hydrolyzes the glycosidic bond in celllobiose. Enzymes that oxidize cellulose and its byproduct include lytic polysaccharide monooxygenases (LPMO) and celllobiose dehydrogenases (CDH). With the recent discovery of these oxidative enzymes, research has shifted gears to better understand the oxidative mechanisms fungal cellulases utilize to breakdown cellulose.

CDH is a flavocytochrome enzyme that catalyzes the oxidation of celllobiose into celllobiono-1,5-lactone. These enzymes contain a dehydrogenase domain with a FAD cofactor and cytochrome domain with a Met-His coordinated heme b cofactor. (Hallberg et al. 2000) The CDHIIA variant contains a small family 1 carbohydrate-binding module (CBM). A flexible linker approximately 20 residues long connects the cytochrome and
dehydrogenase domains. A second linker five residues in length is expected between the
dehydrogenase domain and CBM, but the crystal structure of CDH has shown the CBM
tucked into the dehydrogenase domain. (Tan et al. 2015) The two-electron oxidation of
cellobiose leads to the reduction of FAD to FADH$_2$. Electrons are then shuttled to the heme
cofactor one at a time. The cytochrome domain transfers the electrons to either a LPMO or
another electron acceptor.

LPMOs are copper metalloenzymes. The copper ion is oriented at the active site by a
histidine brace. These enzymes have been recategorized from the glycoside hydrolase 61
family to auxiliary activity 9 cellulases, due to their low endoglucanase activity. (Levasseur et
al. 2013; Lombard et al. 2014) These enzymes are found within fungi and bacteria. There are
three classes of LPMOs that are categorized on where they oxidize the cellulose strand. Type
1 oxidizes at the C1 position creating an aldonic acic upon glycosidic bond oxidation. (Vaaje-
Kolstad et al. 2010; Phillips et al. 2011) Type 2 LPMOs oxidize the cellulose strand at the
C4 creating position ultimately generating a gemdiol species. (Phillips et al. 2011; Beeson et
al. 2012; Isaksen et al. 2014) The third class of LPMOs has been shown to oxidize the C1
and C4 position of glucan chains. (Li et al. 2012) Previous studies have shown that various
small molecules, such as gallic acid and ascorbic acid, can reduce LPMO in place of
CDH. (Kittl et al. 2012; Bulakhov et al. 2016; Frandsen et al. 2016)

Structural structures of different classes of LPMOs from various organisms have been
solved aiding in mechanistic insights. (Karkehabadi et al. 2008; Harris et al. 2010; Quinlan et
al. 2011a; Li et al. 2012; Wu et al. 2013; Borisova et al. 2015; Tan et al. 2015; Westereng et
al. 2015; Frandsen et al. 2016) These studies have shown that LPMO is capable of oxidizing
not only the cellulose surface, but soluble cellooligosaccharides as well. (Langston et al. 2011; Frandsen et al. 2016) NMR studies have suggested that the cytochrome domain of CDH binds directly to the copper site of the LPMO. (Courtade et al. 2016)

The first proposed catalytic mechanism involved the LPMO resting state copper at the Cu(II) state (Figure 4.1). (Phillips et al. 2011) Reduction of the copper atom through the CDH or small molecule electron transfer occurs, which promotes binding of molecular oxygen. A superoxo intermediate is formed with a transition of Cu(I) back to Cu(II). Depending upon the site of oxidation, a H is abstracted from the polysaccharide at the C1 or C4 position, creating a hydroperoxide at the Cu site. A second electron from CDH, or a small molecule, and a proton leads to the formation of a copper oxo species, which hydroxylates the radical on the polysaccharide. By adding the oxygen atom, LPMO destabilizes the glycosidic bond.

Recent density functional theory calculations report that Cu(I) binds molecular oxygen creating a Cu(II) superoxo intermediate (Figure 4.2). (Kim et al. 2014) The transfer of two electrons and protons creates a reactive oxygen species at the copper site that binds the substrate. H abstraction creates a radical on the substrate and the resulting hydroxyl at the PMO active site is added to the substrate bringing the copper back to the Cu(I) state. Despite differences in the proposed mechanisms intermediates, both of the mechanisms state that two electrons are required per LPMO cleavage reaction.

Cellulase optimization studies have found that adding CDH and LPMO significantly improves cellulose degradation. (Igarashi et al. 1998; Igarashi et al. 2002; Langston et al. 2011; Phillips et al. 2011; Bulakhov et al. 2016; Wang & Lu 2016) The ability for LPMO to create new reducing or non-reducing ends within crystalline cellulose strands is of major
importance, since this allows increased loading sites for CBHs. (Quinlan et al. 2011b) While CDH is the enzymatic electron donor for LPMO, several studies have found that small molecules, such as ascorbic acid and gallic acid, are capable of reducing the active site copper ion as well. (Vaaje-Kolstad et al. 2010; Quinlan et al. 2011a; Dimarogona et al. 2012; Kittl et al. 2012; Bulakhov et al. 2016) Gallic acid, a phenol, has been demonstrated to inhibit cellulases. (Ximenes et al. 2011) While it seems more convenient and cost-effective to replace an enzyme with a small molecule in a cellulase cocktail, it has been argued that the presence of CDH also prevents product inhibition of cellulbiohydrolases through the oxidation of cellulose degradation products. (Igarashi et al. 1998; Wang & Lu 2016)

While the synergism of CDH and LPMO have been reported, little is known about how these enzymes interact with respect to electron transfer. Attempts to characterize the binding of the full-length CDH and PMO have proved difficult due to the transient nature of their interaction. (Wu 2013; Courtade et al. 2016) Tan et al confirmed through rapid-kinetics experiments that the cytochrome domain, rather than the dehydrogenase domain, is responsible for transferring electrons to the LPMO. (Tan et al. 2015) Docking studies suggested that the CDH heme propionate arm was near the copper ion in LPMO during electron transfer. (Tan et al. 2015) In the crystal structure of Myriococcus thermophilum, the propionate group of the heme is facing towards the FAD. These docking studies were done of just the cytochrome domain of CDH and did not take into account the LPMO substrate-binding site, which is the site of the copper ion. With these structural observations combined, the current sequence of interactions required for electron transfer is: 1) dehydrogenase domain shuttles electrons to cytochrome domain; 2) cytochrome domain dissociates from
dehydrogenase domain to interact with copper site of LPMO, which has either not bound to
the substrate yet or must dissociate from the substrate.

In this study, we aim to characterize the structural effects NcLPMO has on its
electron donor, NcCDH. Due to the transient nature of the CDH and LPMO interaction,
increasing concentrations of LPMO were added to observe concentration-dependent
structural changes. SANS is ideally suited to study this weakly interacting system through
contrast matching. Through this work we identify how CDH responds to LPMO and propose
a binding mechanism utilized for electron transfer between the two proteins.

4.3 Materials and Methods

4.3.1 – Neurospora crassa Cellobiose Dehydrogenase IIA expression and purification

NcCDHIIA was expressed in KM71H Pichia pastoris (Invitrogen). The Invitrogen
EasySelect Pichia protein expression protocol was followed for protein expression. Cells
were grown in 4 L buffered complex glycerol (BMGY) media overnight at 30 °C. Cells were
spun down at 5,000 x g for 5 minutes at room temperature to remove the BMGY. The cells
were resuspended in buffered complex methanol (BMMY) media at one-fifth the original
culture/BMGY volume (800 mL) to induce expression. 100% methanol was added to a final
concentration of 0.5% 24 hrs after induction. After 48 hrs of induction, the culture was spun
down at 5,000 x g for 10 minutes at room temperature. The supernatant was decanted and
ammonium sulfate was slowly added to a final concentration of 20% saturated solution. After
addition of the ammonium sulfate, the supernatant was centrifuged at 5,000 x g for 15
minutes at room temperature. The supernatant from the ammonium sulfate cut was steri-
filtered with a 0.2 µm PES filter (Nalgene).

The filtered supernatant was applied to a Phenyl Sepharose HP column (GE
Healthcare) equilibrated with 50 mM sodium acetate, pH 5.5, containing 20% (NH₄)₂SO₄
saturated solution. CDHIIA was eluted using a linear gradient of 20% (NH₄)₂SO₄ to 0%
(NH₄)₂SO₄. The buffer was exchanged to 50 mM sodium acetate, pH 5.5 then applied to a
Mono Q HR column (Pharmacia). CDHIIA eluted with a linear gradient from 0 M NaCl to 1
M NaCl. A Superdex 200 column (GE Healthcare) was used as a final polishing and buffer
exchange step into 50 mM sodium acetate, pH 5.5.

4.3.2 – Partially deuterated Neurospora crassa lytic polysaccharide monooxygenase-2
expression and purification

NcLPMO was expressed in SuperMan5 Pichia pastoris cells (BioGrammatics). Cells
were adapted to D₂O over several weeks. The cells were grown in a 3 L bioreactor with D₂O
and hydrogenated carbon source and reagents allowing for partial deuteration. Once enough
cell mass accumulated and glycerol in the media was consumed, methanol was added to
induce NcLPMO expression. Fermentation media was centrifuged to separate cells from
supernatant. Ammonium sulfate was slowly added to the supernatant to a final concentration
of 30% saturated solution. Precipitants were removed through centrifugation. The resulting
supernatant was steri-filtered with a 0.2 µm PES filter (Nalgene).

The supernatant was applied to a Phenyl Sepharose FF (GE Healthcare) column that
has been equilibrated with 25 mM sodium acetate, pH 5.5 containing 30% (NH₄)₂SO₄
saturated solution. LPMO is eluted with a linear gradient of 30% (NH₄)₂SO₄ to 0%
(NH₄)₂SO₄. The buffer was exchanged to 20 mM sodium acetate, pH 5.5 and then injected onto a Mono Q HR column (Pharmacia). LPMO eluted with a linear gradient from 0 M to 1 M NaCl. Gel filtration was used as a final polishing step and for buffer exchange into 20 mM sodium acetate, pH 5.5.

4.3.3 – Lytic polysaccharide monooxygenase contrast variation series

To determine the match point of the partially deuterated LPMO, samples were exchanged into 50 mM sodium acetate, pH 5.5, buffer containing 20 %, 40 %, 60 %, 80 %, or 100 % D₂O. LPMO was measured at a concentration of 6 mg mL⁻¹ for each point. SANS measurements were performed on the extended Q-range SANS (EQ-SANS) instrument at the Spallation Neutron Source, Oak Ridge National Laboratory. (Zhao et al. 2010) EQ-SANS was configured in the 60 Hz mode with a wavelength band of 2.5 Å to 6.1 Å. The sample-to-detector distance was set to 4 m. Samples were loaded into 1 mm banjo quartz cuvettes (Hellma USA, Plainville, NY) and measured at 18 °C. The data were reduced and corrected for transmission, detector sensitivity, dark current, and sample background using MantidPlot. (Arnold et al. 2014) The match point was determined by plotting the square root of the forward scattering intensity against the percentage of D₂O in the buffer.

4.3.4 – CDH and LPMO contrast matching series

Measurements of the CDH-LPMO complex were done in a 70% D₂O 50 mM sodium acetate, pH 5.5, buffer. CDH was kept at a constant concentration of 17 µM, while LPMO was increased from 17 µM to 170 µM. The sample-to-detector distance was set to 4 m. EQ-SANS was configured in the 30 Hz mode, or frame-skipping mode, with wavelength bands 2.5 Å to 6.1 Å and 9.4 Å to 13.4 Å. Frame skipping provided a q-range of 0.005 Å⁻¹ to
0.2 Å⁻¹. Samples were loaded into 1 mm banjo quartz cuvettes (Hellma USA, Plainville, NY) and measured at 18 °C. The data was reduced and corrected for transmission, detector sensitivity, dark current, and sample background using MantidPlot. (Arnold et al. 2014) Primus and GNOM were used for data analysis. (Svergun 1992; Konarev et al. 2003)

4.3.5 – Cellobiose dehydrogenase crowding experiment

To ensure that observations made from the CDH-LPMO SANS experiment were due to LPMO interaction rather than protein crowding effects, a crowding experiment was performed. NcCDHIIA was measured at 2 mg mL⁻¹ in the presence of 0.5 – 4 molal betaine and 0.5 – 3 molal PEG 200. Buffers were exchanged with a centrifugal concentrator (Vivaspin) followed by centrifugal filtration (VWR) before measurements. SAXS measurements were performed on the Rigaku Bio-SAXS 2000 at the Center for Structural Molecular Biology in Oak Ridge National Laboratory. Measurements were done at ambient temperature. Samples were measured for 60 min. Data were corrected for transmission and sample background then reduced and buffer subtracted using Rigaku SAXSLab 3.1.0b14. Primus and GNOM were used for data analysis. (Svergun 1992; Konarev et al. 2003) Osmolality of buffers were measured on a Wescor Vapro 5520 in triplicate then averaged.

4.3.6 – Determination of NcCDH-NcLPMO binding constants

Sedimentation velocity and sedimentation equilibrium experiments were performed on a Beckman XL-I. NcLPMO and NcCDH were exchanged into 50 mM sodium acetate, pH 5.5. To track changes in CDH mobility, wavelengths at 250 nm, 280 nm and 420 nm were measured. For the sedimentation velocity experiments, samples were spun at 50,000 rpm at 18 °C. The sedimentation equilibrium experiments were spun at 10,000 rpm at 4 °C.
Surface plasmon resonance (SPR) experiments were completed on a Biacore 3000 (GE Healthcare). LPMO was exchanged into 10 mM sodium acetate, pH 4.5, for optimal LPMO immobilization. LPMO was crosslinked onto a Biacore CM5 (GE Healthcare) sensor chip with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide•HCl (EDC) and N-hydroxysuccinimide (NHS). The protein was flowed over the CM5 chip for 20 min at 10 µL/min. Ethanolamine flowed over the chip to remove weakly bound LPMO and cap free NHS esters. CDH was exchanged into either a 10 mM HEPES, pH 7.4, 150 mM NaCl, 0.005% Tween 20 buffer or 50 mM sodium acetate, pH 5.5, 150 mM NaCl, 0.005% Tween 20 buffer. To ensure complete removal of CDH, 10 mM glycine, pH 10.2, was flowed over the CM5 chip after each injection.

4.4 Results

4.4.1 – CDH-LPMO SANS Measurements

NC-LPMO was measured at 6 mg mL\(^{-1}\) for the match point experiments. The data points for 20%, 40%, 60%, 80%, and 100% D\(_2\)O buffers were reduced, substracted, and analyzed at the beamline to determine the match point for LPMO (Figure 4.1). The Guinier analysis for each of these points showed linear trends indicating that these samples were free of aggregates and interparticle interactions. Using the intensity of radiation scattered at zero angle, I(0), calculated from the Guinier analysis, the square-root of I(0) was plotted against the percent D\(_2\)O in the sample (Figure 4.2). Data points for 80% and 100% were plotted as negative values for ease of match point determination. The match point for the partially deuterated LPMO was determined to be 68.7% D\(_2\)O. For dilute systems, the protein being
matched should have negligible scattering contributions within ± 5% D₂O of the match point, so measurements for the CDH-LPMO complex were done in a 70% D₂O 50 mM sodium acetate, pH 5.5, buffer. (Stanley, personal communication)

For the complex measurements, EQ-SANS was configured in the frame-skipping mode. This mode generates two scattering curves for a sample (one for low q and another for
Figure 4.3 – CDH-LPMO complex series SANS data plotted on a log-log scale.

Figure 4.4 – Normalized P(r) profiles for CDH-LPMO series. Ratio shown as molar ratio of CDH : deuterated LPMO. CDH alone, red; 1:1, orange; 1:5, green; 1:10, blue.

Figure 4.5 – Ratio of deuterated LPMO plotted as a function of deuterated LPMO. Error bars are from the P(r) analysis.
high q) that are merged together. Upon analysis of the data, only the first frame was used for each sample, which provided a sufficient q-range of 0.01 Å\(^{-1}\) to 0.21 Å\(^{-1}\) (Figure 4.3) The CDH-LPMO experiments were performed with increasing concentrations up to 10-fold molar excess concentrations of LPMO to increase the amount of CDH complexed with LPMO. Guinier analysis of each sample showed a linear trend at low q, signifying that all of the complex samples were free of aggregation and interparticle interactions.

P(r) profiles showed that each increase in LPMO led to a decrease in the \(R_g\) and \(d_{\text{max}}\) (Figure 4.4). The 1:1 LPMO:CDH ratio showed a decrease in \(R_g\) and \(d_{\text{max}}\) from 39.61 ± 1.80 Å and 148 Å to 34.47 ± 0.99 Å and 128 Å. The 1:5 sample had a slightly less dramatic decrease in \(R_g\) and \(d_{\text{max}}\) at 32.84 ± 0.64 Å and 115 Å. At 1:10, another significant drop in \(R_g\) and \(d_{\text{max}}\) was measured (28.46 ± 0.41 Å and 97 Å, respectively). Change in \(R_g\) as a function of increasing LPMO concentration is easily seen in Figure 4.5. The P(r) curves also show significant changes with each increase in LPMO. A shoulder in the P(r) curve begins to form around 75 Å. This trend was observed in Chapter 3, when the cytochrome and dehydrogenase domains came closer together to form a more compact structure. These changes in curvature are additional indications of significant structural rearrangements within CDH as the CDH-LPMO complex increases.

4.4.2 – CDH Crowding Experiment

To understand how CDHIIA structurally responded to crowding conditions and determine if our LPMO-CDH complex series results were results of crowding or not, CDH was measured in the presence of crowders, betaine and PEG200.(Bhattacharya et al. 2014; Sarkar & Pielak 2014) The samples shown are only for 0.5 m and 1 m betaine and 0.5 m,
1 m, and 2 m PEG200. These samples had linear trends within the Guinier region, which indicated that they were free of aggregates and interparticle interactions. The higher crowder concentrations showed clear signs of aggregation. In Figure 4.6, both betaine and PEG200 show very small decreases in $d_{\text{max}}$, but significantly increase as the osmolality increases. This increase in $R_g$ and $d_{\text{max}}$ could be signs of conformational changes within CDH or subtle signs of aggregation that were not observed with the Guinier analysis. Initially, PEG200 modestly decreases the $R_g$, but it does not appear to significantly vary as osmolality increases. The addition of 1 m betaine led to a nearly 2.5 Å increase in $R_g$. The disparity in structural responses to the crowders may be due to their chemistry. Betaine is a charged molecule with a cationic quaternary ammonium and an anionic carboxylate group, while PEG200 is a

![Figure 4.6](image_url)

**Figure 4.6** – Effect of crowders on NcCDHIIA. The $R_g$ (circle) and $d_{\text{max}}$ (triangle) of CDHIIA are plotted as a function of osmolality. Betaine series, gray; PEG 200 series, blue.
neutral molecule. Back calculated osmolalities (assuming osmolality = molality at dilute conditions) for LPMO were 0.0167 mmol/kg, 0.0835 mmol/kg, and 0.167 mmol/kg for the 1:1, 1:5, and 1:10 CDH:LPMO conditions, respectively. Addition of LPMO to CDH at these concentrations should not have induced conformational changes due to crowding. The results of this crowding experiment show that significant increases in osmotic pressure lead to increases in both \( R_g \) and \( d_{\text{max}} \). These findings are counter to what was observed in the CDH-LPMO experiment, supporting our findings that the structural rearrangements were due to binding rather than crowding-induced compaction.

4.4.3 – Determination of \( \text{NeCDH-LPMO binding constants} \)

The CDH-LPMO complex is known to be a transient interaction, which makes studying the complex difficult. (Wu 2013; Courtade et al. 2016) Analytical ultracentrifugation (AUC) experiments were performed to investigate the binding of CDH to LPMO. Sedimentation velocity and sedimentation equilibrium experiments showed no signs of complex formation despite up to an eight-fold molar excess of CDH to LPMO (Figure 4.7, up to four-fold excess shown). SPR experiments were performed at the optimal CDH activity pH, 5.5, and less optimal pH, 7.5. The CM5 chip is a matrix of carboxymethylated dextran that prompted initial concerns about non-specific binding of CDH, due to the presence of its CBM. A reference lane that only had NHS, EDC, and ethanolamine flowed over it during the immobilization process was created to check for non-specific binding of CDH. There was no significant non-specific binding observed to the reference lane. Previous SPR studies on \( \text{Phanerochaete chrysosporium} \) CDH noted difficulties removing CDH from the chip, so 10 mM glycine, pH 10.2 was used between runs. (Wu 2013) Our data showed that as soon as
CDH stopped flowing over the chip, CDH dissociation occurred. The 10 mM glycine, pH 10.2, buffer was used between runs as a precaution. Concentration dependent binding was observed, but the chip never reached saturation making $K_d$ determination impossible. The interaction between CDH and LPMO appeared to be too transient to obtain binding constant information with AUC and SPR with the listed conditions.

### 4.5 Discussion

Ideally, mass spectrometry is performed before a contrast variation experiment to determine the amount of deuterium incorporation in the protein. Unfortunately, clean spectra could not be obtained on the in-house mass spectrometer prior to the experiment making deuterium incorporation estimates inaccurate. Based on the experimentally determined match point of 68.7% and calculations from MULCh, the incorporation of non-exchangeable deuterium was approximately 30%.(Whitten et al. 2008) Based on the beam line scientist’s previous experience, being within 5% D$_2$O of the match point would be acceptable for further complex studies since we were in a dilute system.(Stanley, personal communication)
When the LPMO in 70% D₂O was measured, it showed that there was still a slight scattering contribution with an I(0) of 0.00079. To determine if we were “seeing” LPMO in the complex study, the 70% D₂O LPMO sample was scaled proportionally to the concentration of the sample. The scaled 70% D₂O LPMO curve was then subtracted from the complex sample. The LPMO subtracted complex curves were compared against the initial data (Table 4.1). While this method does not account for cross-terms, this is the best method we had to validate that we did not have “bleed through” of LPMO scattering influencing our data. The subtracted curves produced similar Rₚ’s and dₘₚ’s for each concentration point.

<table>
<thead>
<tr>
<th></th>
<th>1:1 (Initial)</th>
<th>1:1 (Sub)</th>
<th>1:5 (Initial)</th>
<th>1:5 (Sub)</th>
<th>1:10 (Initial)</th>
<th>1:10 (Sub)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I(0)</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
<td>0.08</td>
<td>0.09</td>
<td>0.08</td>
</tr>
<tr>
<td>Rₚ (Å)</td>
<td>34.47</td>
<td>34.16</td>
<td>32.84</td>
<td>33.22</td>
<td>28.46</td>
<td>29.10</td>
</tr>
<tr>
<td>dₘₚ (Å)</td>
<td>128.00</td>
<td>124.00</td>
<td>115.00</td>
<td>114.00</td>
<td>97.00</td>
<td>95.00</td>
</tr>
</tbody>
</table>

Initial represents the data analyzed as described in the methods. Sub is the standard curve subtracted by the weighted dLPMO at 70% D₂O curve.

The results from our SANS LPMO-CDH study showed a remarkable reduction in CDH dimensions as the concentration of LPMO was increased. The concentration-dependent reduction that is observed is due to the increased ratio of bound CDH to unbound CDH. Our data suggests that as CDH binds to LPMO, CDH adopts a closed conformation (Figure 4.4). Maintaining a compact conformer while bound to LPMO may allow for a more efficient transfer of electrons between the dehydrogenase and cytochrome domains and LPMO. A compact conformer was not observed when cellobiose was added in the previously discussed SAS studies. This would suggest that when cellobiose is bound in the dehydrogenase domain, CDH will not adopt a closed conformer but will increase in flexibility. From this
finding, we propose that CDH would then be capable of adopting a closed conformation to transfer electrons between the dehydrogenase and cytochrome domains when CDH interacts with LPMO to donate electrons.

The SANS data shows a dramatic decrease in the dimensions of *NcCDH* as the concentration of *NcLPMO* is increased (Figure 4.5 and Table 4.1), which led to initial concerns that these effects were possibly due to crowding instead of interaction-induced conformational changes. With the maximum concentration of *NcLPMO* presented here being less than or equal to 4 mg mL\(^{-1}\), the solution should not be within crowding protein concentrations. To quell this concern, a crowding study on CDH was performed to observe how it responded to increasing osmotic pressure. Our results show that significant increases in osmolality are required for conformational changes to occur in CDH. The conformational changes that were observed were increases in \(R_g\) and \(d_{max}\), which were opposite of the findings in the CDH-LPMO complex study. We conclude that the structural observations made from the CDH-LPMO SANS experiment are due to CDH interacting with LPMO rather than LPMO-inducing crowding conditions.

Recent work to characterize the binding of CDH and LPMO from various organisms has been attempted. Isothermal calorimetry (ITC) has been attempted to study *NcCDHIIA* and *NcLPMO9C* interactions with no success.\(^{(}\text{Courtade et al. 2016}\)) Previous work investigated the interactions between *Phanerochaete chrysosporium* LPMO9D and CDH using surface plasmon resonance (SPR).\(^{(}\text{Wu 2013}\)) The binding never reached equilibrium, so the \(K_d\) value of 80 \(\mu\)M was an approximation rather than a true equilibrium constant. Our attempts to assess the binding of *NcCDHIIA* and *NcLPMO9C* proved to be unfruitful as well.
There were no signs of complex formation in the AUC experiments. For the SPR experiment, concentration-dependent increases in binding were observed. Unfortunately, we were unable to reach saturation to determine the dissociation constant and the CDH off rate was too fast to obtain kinetics information.

Recent nuclear magnetic resonance (NMR) studies confirmed results from a docking study that found CDH, specifically the cytochrome domain, interacted with LPMO at the copper site (Figure 4.8, bottom). (Tan et al. 2015; Courtade et al. 2016) This would require that either: based on the DFT study, CDH pre-loads two electrons to LPMO before polysaccharide binding; or, based on the model from Phillips et al, LPMO would have to

---

Figure 4.8 – Schematic of NcCDH electron transfer cycle. Top, cellobiose bound at the dehydrogenase domain of CDH does not allow IET. Bottom left, CDH bound to LPMO (triangle; Cu, blue circle) electron flow (dashed arrows) for secondary weak interaction site. Bottom right, electron flow from CDH to a strong interaction site on LPMO. Cytochrome domain, red; dehydrogenase domain, yellow; CBM, orange.
break contact with the substrate to transfer the second electron for the formation of the copper oxo species. (Phillips et al. 2011; Kim et al. 2014) It is improbable that LPMO dissociates from the polysaccharide during glycosidic bond oxidation. If the LPMO mechanism proposed by Phillips et al is utilized, then there is likely a second, weaker interaction site for both the LPMO and cytochrome domain. This type of secondary, low-affinity binding site has been reported before with cytochrome c peroxidase and cytochrome c. (Van de Water et al. 2015)

4.6 Conclusion

This work reports the findings from the structural investigations of NcCDHIIA and NcLPMO. The data clearly shows that when CDH interacts with LPMO, CDH adopts a closed conformation. By adopting a compact conformer upon binding with LPMO, the transfer of electrons between the dehydrogenase and cytochrome domains is made possible. Our data cannot discriminate where the CDH and LPMO interact, but it clearly shows that CDH goes from an extended state to a closed state.

Acknowledgements

I would like to thank Wm. Brad O’Dell for providing purified hydrogenated and partially deuterated NcPMO for these experiments. Thank you to Dr. Edward Wright of the Bioanalytical Resource Facility at University of Tennessee for all of the assistance with the CDH-LPMO binding experiments. I would also like to thank Dr. Chris Stanley for his invaluable assistance with the SANS work.
4.7 References


CHAPTER 5

Conclusions and Perspectives

5.1 Cel7A

The structures of an open state and a closed state of Cel7A catalytic domain (CD) were solved. It was concluded that tunnel loops of the CD cycle between these tense and relaxed forms to process along a cellulose strand. While the newly solved crystal structures have provided new insight into the Cel7A processivity mechanism, there are still some lingering questions that could be answered with a series of mutations, activity assays, and crystallographic studies. Along the active site tunnel, Y247 and Y371 from loops B3 and A3, respectively, were observed to flip in response to a shift in loop B3. We have proposed that the pair helps coordinate loop motion on both sides of the tunnel. Early investigations into loop B3 found that mutations (Y247F, truncation, and stabilization through addition of a disulfide bond) led to a decrease in processivity that was more prevalent in the truncation and Y247F mutants. (von Ossowski et al. 2003) The removal of the tyrosine residue would affect ligand interactions and may explain the change in activity. If Y371 was mutated, the role of the proposed tyrosine flip could be further evaluated without disturbing ligand interactions. A salt bridge between K102 and E408 was observed that links loops A1 and A2. These exist as a part of a larger, well-ordered loop containing loops A3 and A4 as well. We proposed that this salt bridge coordinates loop motions along the entire A-side of the tunnel. Disruption of this salt bridge would evaluate if the K102-E408 coordinates these loop motions and, if so, how essential loop coordination is to Cel7A catalytic activity.
Previous work has shown that Cel7A, as well as other cellobiohydrolases, loop flexibility and catalysis is highly pH-dependent. (Pingali et al. 2011; Bu et al. 2013) With the pH optima for Cel7A existing around pH 5.0, the charged residues along the active site tunnel loops and residues responsible for cellulose hydrolysis are susceptible to changes in protonation state.

The cellulose chain is hydrolyzed at the active site of Cel7A by the concerted action of the catalytic residues E212 and E217. (Klarskov et al. 1997) The first hydrolysis reaction step is proposed to be the protonation of the substrate glycosidic oxygen by E217, which would initially be neutral, while charged E212 would act as a nucleophile and attack the substrate anomeric carbon. (Li et al. 2010) In the second step of this mechanism, E217, which would be negatively charged after proton transfer to the substrate, would activate a water molecule. The resulting hydroxyl ion would then replace E212 as the nucleophile, leading to the glycosidic bond cleavage. A third carboxylic residue, D214, is thought to play a role in fine-tuning the pKₐ of E212. (Li et al. 2010) However, the lack of experimental information on the protonation states of these residues prevents definitive confirmation of this mechanism.

Activity assays of H228 located near the catalytic residues have shown its importance in hydrolysis of cellulose. Mutation of H228 in Cel7A leads to a marked decrease in cellobiohydrolase activity. It has been hypothesized that H228 plays a role in maintaining the conformation of the catalytic residues and is in part responsible for the pH dependence observed for Cel7A activity. (Granum et al. 2014) The protonation state of H228 has not been experimentally determined.
Neutron protein crystallography is uniquely suited to determine the protonation state of residues in question. Currently, the structure of *Trichoderma reesei* Cel7A has been solved in space groups $P6_322$ and $C2$ yielding an open and closed state, respectively. (Bodenheimer *et al.* 2014) These space groups can be produced from the same crystallization condition. Since the closed conformation was a new state, optimization of this condition was desirable. Recent work to optimize crystal growth conditions has yielded crystal sizes up to $2.0 \times 1.0 \times 0.3 \text{ mm}^3$, though these crystals did not look single. The largest crystals were grown in a sandwich box (Hampton Research) with drops consisting of 50 $\mu$L 10 mg/mL Cel7A, 10 $\mu$L Silver Bullet Reagent D4 (Hampton Research), and 10 – 20 $\mu$L of reservoir solution. Equal success has been had with replacing the D4 cocktail with a solution of 5 mM Gd(NO$_3$)$_3$$\cdot$$6\text{H}_2\text{O}$ and 20 mM HEPES sodium, pH 6.8. Reservoir solution consists of 25 – 28% PEG 3350 with 100 mM HEPES sodium, pH 6.4. Optimal temperature for crystal growth is 35 ºC. Removal of the linker and CBM with papain was not found to improve crystal growth with these conditions.

Crystals have been tested on the IMAGINE Laue diffractometer at the High Flux Isotope Reactor at Oak Ridge National Laboratory with 1 hour Laue exposure. (Meilleur *et al.* 2013) Diffraction images are from a crystal with dimensions of $1.3 \times 0.8 \times 0.3 \text{ mm}^3$ (Figure 5.1). The diffraction images are of the Cel7A crystal in hydrogenated buffer (top) and D$_2$O-solvent exchanged by vapor diffusion for 14 days (middle) and 100 days (bottom). There was no apparent improvement in signal to noise with extended D$_2$O-solvent exchange time. While we confirmed that we are able to grow single crystals, the resolution of these tests clearly indicates that larger crystals will be required for full data collection.
Figure 5.1 – Neutron diffraction images collected on IMAGINE of *T. reesei* Cel7A in hydrogenated buffer (top), 14 days (middle) and 100 days (bottom) of D$_2$O-solvent exchanged crystal.
Upon successful crystal growth of the CD with an empty tunnel, a pH series will be performed to better understand how pH affects the protonation state of active site residues and tunnel loop residues.

5.2 *Celllobiose Dehydrogenase IIA*

The effects of various chemical environments and LPMO interactions have been investigated with CDIIA from *Neurospora crassa* and *Myriococcus thermophilum*. These studies were performed on enzymes expressed in a standard *Pichia pastoris* cell line at room temperature. *P. pastoris* has extensive N-linked glycosylation at consensus sequences Asn-Xaa-Ser/Thr, with Xaa being any amino acid except for Pro. A typical N-linked glycan structure consists of two N-acetylglucosamines (GlcNac) and between 8 – 14 mannoses (Man) with extensive branching (Figure 5.2). (Bretthauer & Castellino 1999; De Pourcq et al. 2010) These N-linked glycans are significantly smaller than *Saccharomyces cerevisiae* that can have glycans larger than 50 sugar moieties. (Bretthauer & Castellino 1999) These N-

![Figure 5.2 – Typical glycans present in Pichia pastoris. The core N-linked is showed on top with the starred that can have additional mannoses attached. Below is the average O-linked structure that is typically less than 5 mannoses in length. GlcNAc, blue square; Man, green circle.](image-url)
linked sites are very heterogeneous in the amount of linked mannoses. (De Pourcq et al. 2010) O-linked glycans are present at Ser and Thr residues and are typically less than five mannoses in length.

Glycans are known to protect enzymes from proteolysis, help with protein folding, aid in enzymatic activity and stability, and play a role in signaling. (De Pourcq et al. 2010) Extensive work has investigated the role glycosylation plays in cellulase activity. (Klarskov et al. 1997; Harrison et al. 1998; Poon et al. 2007; Adney et al. 2009; Wu et al. 2009; Beckham et al. 2010; Dotsenko et al. 2015) For the Cel7A linker, gradual increase in glycan length allowed for a 16 Å linker extension without altering its inherent flexibility. (Beckham et al. 2010) A linker from Cellulomonas fimi showed that the presence of glycans decreased its flexibility. (Poon et al. 2007) Cel7A expressed in P. pastoris showed a significant decrease in activity compared to the natively expressed protein where glycosylation determined to be the underlying cause for the drop catalytic activity. (Boer et al. 2000) Subsequent mutation studies found that an N-linked site at the entrance of the tunnel led to the overall stability of the enzyme while those at the exit site likely decreased loop flexibility. (Adney et al. 2009)

These studies have led to questions on how glycans, not only affect Cel7A, but also affect CDH flexibility and ultimately its activity. N. crassa and M. thermophilum CDHIIA both have nine glycosylation sites based on crystal structures. (Tan et al. 2015) These sites vary in their location on the enzyme and linkage between different species (Figure 5.3). M. thermophilum predominately has O-linked glycans that are located at the ends of the linker between the dehydrogenase and cytochrome domains. The other glycosylation sites present on MtCDH are N-linked glycans dispersed along the dehydrogenase domain. The majority of
glycosylation sites along NcCDH are N-linked throughout the dehydrogenase domain except for a single site at the cytochrome domain. There are two O-linked glycans found at the dehydrogenase side of the linker between the dehydrogenase and cytochrome domains. The small angle scattering (SAS) results show there are differences in dimensions between these two enzymes that are likely due to the glycosylation patterns. It is tempting to also attribute the dimensional dissimilarities to glycans affecting the rigidity of the linker between the dehydrogenase and cytochrome domains.

Glycoengineering of P. pastoris has allowed proteins to be expressed with homogeneous N-linked glycans with the structure GlcNAc₂Man₅. (Smith et al. 2014) By using these strains of yeast, smaller homogeneous N-linked glycans will be added that will also be more representative the native glycosylation patterns. Sequential removal of the N-linked glycans may shed light on the role they play for these two different CDHs. Mutations of the glycosylated residues along and near the linker between the dehydrogenase and

![Figure 5.3](image-url) – Full-length N. crassa (left) and M. thermophilum (right) with glycans shown in spheres. Glycans added with glycam web server. (Woods Group 2005-2016)
cytochrome domains would allow for a clearer understanding about how glycans affect the mobility of these domains.

Contrast variation SANS measurements on *N. crassa* CDH have been performed to better understand the effect glycans have on the global structure of the protein. The match point for the glycans and protein were calculated with MULCh to be at approximately 24% and 42% D$_2$O, respectively. (Whitten *et al.* 2008) Through these measurements, an initial understanding of how glycans alter domain conformations and glycan orientation along the protein will be gained. Analysis of the data is currently underway.

Previous studies investigating thermophilic cellulases noted that compared to their mesophilic counterparts, there was an increased level of rigidity. (Voutilainen *et al.* 2007; Zheng *et al.* 2012) With *MtCDH* coming from a thermophilic fungi and *NcCDH* from a mesophile, a subsequent study investigating potential correlations between temperature, activity, and flexibility would be of interest. Colussi *et al* have previously shown that Cel7A from the mesophilic ascomycete *Trichoderma harzianum* becomes more flexible at optimal activity temperatures. (Colussi *et al.* 2012) The broad range of applications for these enzymes requires these enzymes to maintain enzymatic activity within a diverse set of environmental conditions. By better understanding these structural parameters, more targeted enzymatic engineering can be performed.
5.3 References


APPENDIX
**Electroporation**

Before beginning the electroporation step, 5-10 µg of linearized DNA needs to be obtained. This amount of DNA can be obtained through a midiprep (QIAGEN) or a similar procedure. The plasmid must be linearized with the restriction enzymes SacI, PmeI, or BstXI (used for pPICZ vectors). Follow the linearization with a restriction digest clean up and resuspend the DNA in sterile water.

To prepare the *Pichia pastoris* cells for electroporation, inoculate 5 mL YPD in a 50 mL flask with the *P. pastoris* strain of choice and grow overnight at 30°C shaking between 250 – 300 rpm. Next, 500 mL of fresh YPD in a 2.8 L baffled flask will be inoculated with 500 µL. Cells need to be grown until they reach an OD$_{600}$ of 1.3 – 1.5. The cells should be centrifuged at 1,500 x g for 5 minutes at 4°C. Discard the YPD and resuspend the cell pellet in 500 mL ice-cold, sterile water. The centrifugation step should be repeated with cell resuspensions of 250 mL of ice-cold water, 20 mL ice-cold 1 M sorbitol, and finally 1 mL ice-cold 1 M sorbitol. The cells resuspended in sorbitol should be kept on ice and used that day.

Combine 80 µL of the sorbitol suspended cells with the prepared 5 – 10 µg of linearized DNA in a pre-chilled 0.2 cm electroporation cuvette. The cells should incubate with the DNA for 5 minutes on ice. Pulse the cells based on the electroporators protocols for *Pichia pastoris*. If there is no setting for *P. pastoris*, use the *Saccharomyces cerevisiae* protocol. Upon completion of electroporation, immediately add 1 mL ice-cold 1 M sorbitol to
the electroporator cuvette. Put the electroporated cells into a sterile 15 mL tube and incubate at 30°C for 1 to 2 hours without shaking. To test for Zeocin selection, 10 – 200 µL of the incubated cells should be plated on YPDS plates with 100 µg/mL Zeocin. An easy way to select for multi-copy recombinants is to plate 100 – 200 µL of cells onto YPDS plates containing 500 – 2,000 µg/mL Zeocin. The plates need to be incubated at 30°C for 3 – 10 days for the 100 µg/mL Zeocin plates and only 2 days for the ≥ 500 µg/mL Zeocin plates.

**Phenotyping**

Depending on the type of strain that is being used, it will determine if phenotyping needs to be done. Knowing the phenotype of the cell is essential to growth and induction strategies. KM71H cells always have the phenotype Mut^S^, which indicates that the AOX1 gene is disrupted and will grow slow on methanol. Transformants in X-33 and GS115 cells typically only have a single crossover recombination where the AOX1 locus occurs. This typically generates the Mut^+^ phenotype, where cells grow on methanol. Sometimes recombination will occur at the 3’AOX1 region of the plasmid, leading to a disruption of the wild-type AOX1 gene. This disruption generates a Mut^S^ phenotype. To test which phenotype has been produced, plate colonies on minimal dextrose and minimal methanol plates. Incubate the plates at 30°C for 2 days. Colonies that only grow on dextrose have the Mut^S^ phenotype, while those than can grow on dextrose and methanol are Mut^S^ (Figure A1). For example, the plates in Figure A1 have Mut^S^ colonies in blocks 1 – 4 and 11 – 15 and Mut^+^ colonies in blocks 5 – 10 and 16.
Growth and Expression

Based on the phenotype of the cells, determine the method to use for cell growth and protein expression. Expression methods are the same for intracellularly expressed or secreted proteins. Typical media used for protein growth/expression is buffered complex glycerol/methanol media (BMGY/BMMY). If proteolytic activity is found during protein expression, minimal glycerol/methanol media (MGY/MMY) can be used. Since the media will not be buffered, check that the protein folding is not affected by the low pH that occurs during growth and expression. To allow for optimal aeration, always use baffled flasks, volumes no greater than 30% of the flask volume, and shake between 250 – 300 rpm. Top the flask with either a silicone sponge closure or a breathable seal. The doubling time for *P. pastoris* is approximately 2 hours.

Figure A1 – Plates for phenotyping *Pichia pastoris*. The left is a minimal dextrose agar plate, where all colonies should grow. The right is a minimal methanol agar plate where only Mut+ strains will grow.
For Mut growth and expression:

1. Inoculate 10 mL BMGY in a 100 mL baffled flask. Incubate between 28 – 30°C for 16 – 18 hours until OD$_{600}$ reaches 2 – 6.

2. Inoculate 1 L BMGY in a 3 L baffled flask with the 10 mL culture. Incubate between 28 – 30°C for 16 – 18 hours until OD$_{600}$ reaches 2 – 6.

3. Centrifuge the cells at 3,000 x g for 5 minutes at room temperature to remove the BMGY. The presence of glycerol will inhibit methanol induction.

4. Cells should be resuspended in 1/10 – 1/5 of the starting volume (100 – 200 mL) of BMMY to induce protein expression. Incubate between 28 – 30°C.

5. Add 0.5% of the induction media volume of 100% methanol every 24 hours. Test for the optimal expression time by removing samples each day for 6 days. Spin the sample down and separate the cells from the supernatant.

6. Centrifuge the cells at 3,000 x g for 5 minutes at room temperature.

For Mut$^+$ growth and expression:

1. Inoculate 10 mL BMGY in a 100 mL baffled flask. Incubate between 28 – 30°C for 16 – 18 hours until OD$_{600}$ reaches 2 – 6.

2. Inoculate 1 L BMGY in a 3 L baffled flask with the 10 mL culture. Incubate between 28 – 30°C for 16 – 18 hours until OD$_{600}$ reaches 2 – 6.

3. Centrifuge the cells at 3,000 x g for 5 minutes at room temperature to remove the BMGY. The presence of glycerol will inhibit methanol induction.

4. Cells should be resuspended with enough BMMY so that the OD$_{600}$ is 1.0. Incubate between 28 – 30°C.
5. Add 0.5% of the induction media volume of 100% methanol every 24 hours. Test for the optimal expression time by removing samples each day for 6 days. Spin the sample down and separate the cells from the supernatant.

6. Centrifuge the cells at 3,000 x g for 5 minutes at room temperature.

**Media Recipes:**

**10X YNB (13.4% Yeast Nitrogen Base with Ammonium Sulfate without amino acids):**

Dissolve either 134 g YNB with (NH₄)₂SO₄ and without amino acids or 34 g YNG without ammonium sulfate and amino acids and 100 g (NH₄)₂SO₄ in 1 L of water and filter sterilize into an autoclaved bottle. The solution is good for about one year when stored at 4°C.

**500X B (0.02% Biotin):**

Dissolve 10 mg biotin in 50 mL of water and filter sterilze. The solution is good for about one year when stored at 4°C.

**10X D (20% Dextrose):**

Dissolve 200 g of D-glucose in 1 L of water then filter sterilize into an autoclaved bottle. The solution is good for about one year at room temperature.

**10X M (5% Methanol):**

Mix 5 mL of methanol into 95 mL of water and filter sterilize into an autoclaved bottle. The solution is good for about two months when stored at 4°C.
10X GY (10% Glycerol):

Mix 100 mL of glycerol into 900 mL of water and filter sterilize into an autoclaved bottle. The solution is good for over a year at room temperature.

YPDS + Zeocin Agar (Yeast Extract Peptone Dextrose Sorbitol – 250 mL)

Dissolve 2.5 g yeast extract, 5 g peptone, 45.6 g sorbitol, and 5 g agar to 225 mL and then autoclave for 20 minutes. Add 25 mL of 10X D and let cool until solution is approximately 60°C to add 250 µL of 100 mg/mL Zeocin. Store plates at 4°C in the dark for 1 – 2 weeks.

MD / MM Agar (Minimal Dextrose / Minimal Methanol – 250 mL)

Add 3.75 g of agar to 200 mL of water and autoclave. Let cool to 60°C then add 25 mL of 10X YNB, 500 µL 500X B, and 25 mL 10X D or 10X M (for MD or MM, respectively). Plates can be stored at 4°C for several months.

BMGY / BMMY (Buffered Glycerol-complex / Buffered Methanol-complex Medium – 1 L)

Add 10 g of yeast extract and 20 g of peptone to 700 mL of water then autoclave for 20 minutes. Once it has cooled to room temperature add 100 mL 1 M potassium phophate buffer, pH 6.0, 100 mL 10X YNB, 2 mL 500X YNB, 100 mL 10X GY or 10X M (for BMGY or BMMY, respectively). Can be stored for about 2 months at 4°C.
**Small Angle Scattering**

*Guinier Analysis*

The Guinier analysis reveals information about the particle at the scattering angle closest to zero. (Guinier & Fournet 1955) This method allows for a quick approximation of the \( R_g \) and \( I(0) \). The Guinier approximation is:

\[
I(q) = I(0)e^{-\frac{q^2R_g^2}{3}}
\]

where \( I(q) \), \( I(0) \), \( R_g \), and \( q \) are intensity as a function of the scattering vector, intensity at zero scattering angle, radius of gyration, and the scattering vector, respectively. A linearization of the approximation yields the following equation:

\[
\ln I(q) = \ln I(0) - \frac{R_g^2}{3}q^2
\]

When performing a Guinier fit, one needs to ensure that the data follows a linear trend. An upward or downward turn at low \( q \) indicates signs of aggregation or interparticle interactions. The approximation is only correct for globular shapes if \( q \cdot R_g < 1.3 \).

*Pair Distribution Function*

The pair distribution function (P(r)) describes the distance between all atoms in a single molecule, assuming particles are monodisperse and not interacting. This is why it is important to establish early on if there are signs of aggregates or interparticle interactions within the sample before continuing data analysis. The P(r) is calculated from the following equation:
$P(r) = \frac{r}{2\pi} \int_0^\infty I(q)q\sin(qr)dq$

where $r$ is radius between scattering points. Small angle scattering programs, such as GNOM, use the inverse Fourier transform of this equation to solve the $P(r)$. (Svergun 1992) $P(r)$ curves give information on the shape of the particle and changes in domain positions. Quantitative values that can be obtained from $P(r)$ are the $R_g$, $I(0)$, and maximum dimension of the particle ($d_{\text{max}}$). The data must be of sufficient q-range for determination of the $d_{\text{max}}$. The q-range should be determined before the experiment based on the dimensions of the particle. To calculate the q-range needed for a given $d_{\text{max}}$ use the equation: $q \leq \pi / d_{\text{max}}$. $d_{\text{max}}$ can be determined from where the curve gradually falls to the x-axis. For an example of $P(r)$ curves where $d_{\text{max}}$ has been properly and improperly determined see Figure A2.

Figure A2 – $P(r)$ curve examples for properly determining $d_{\text{max}}$. The correctly determined $d_{\text{max}}$ is shown in red. The too large $d_{\text{max}}$, green, crosses the x-axes and becomes negative at high $r$ values, while the too small $d_{\text{max}}$, blue, abruptly crosses the x-axis.
**Kratky Plots**

The Kratky plot is useful for determining how folded or unfolded the macromolecule is. It uses the scattering data and plots it as a function of $q^2 \cdot I(q)$ vs $q$. These plots can be made dimensionless by multiplying the $q$-vector by $R_g$ and $I(q)$ by $I(0)^{-1} \cdot (q \cdot R_g)^2$ (Figure A3).

A folded protein would have a peak at low $q$ (magenta), while an unfolded protein (teal) would continue to increase in $q^2 \cdot I(q)$ as $q$ increases. Conformational changes in multi-domain proteins with flexible linkers can be monitored through Kratky plots as well. This is observed through the presence of a broadening of and/or shoulder in the peak at low scattering angles and an increase in $q^2 \cdot I(q)$ at high scattering angles (orange).

Figure A3 – Dimensionless Kratky plot. SAXS curves were generated from crystal structures using the FoXS server to create this plot. Closed conformation of a multi-domain enzyme (PDB ID 4QI6), magenta. Open conformation of a multi-domain enzyme (PDB ID 4QI7), orange. Linker belonging to 4QI7, teal. (Schneidman-Duhovny et al. 2016)
Crystallography

Vapor Diffusion

One of the most commonly used techniques to crystallize macromolecules is through vapor diffusion. Vapor diffusion works by combining a protein solution with a precipitant and allowing them to equilibrate against a larger volume of the precipitant. As the drop equilibrates, water leaves from the drop (from both the protein solution and the added precipitant) and the drop becomes supersaturated. Equilibrium is achieved once the precipitant concentration is nearly equal to what is present in the reservoir. The two methods used for vapor diffusion are hanging drop and sitting drop (Figure A4). These methods have their distinct advantages, but one thing to consider is some molecules will only crystallize with a specific method.

Figure A4 – Schematics of common vapor diffusion methods. Hanging drop (left) involves adding precipitant (blue) to a protein drop (green) on a cover slip that is then sealed over top of a reservoir filled with precipitant. Sitting drop (right) utilizes pedestal to hold the drop that is set up similar to the hanging drop.
Both methods rely on a precipitant, such as a polyethylene glycol or concentrated salt, which is added to a reservoir. To set up a hanging drop tray, a concentrated protein solution is pipetted (typically starting out with 2 µL at a concentration between 5 – 10 mg mL\(^{-1}\)) onto a siliconized glass coverslip. Siliconized glass is essential, since this prevents the crystal from forming on the coverslip. The precipitant from the reservoir is added to the protein drop (for initial screening 2 µL is typically used). Hanging drop allows you to test multiple conditions within a single well, such as various ratios of protein:precipitant, but there is a maximum drop volume that will stay hanging to the cover slip and not fall into the well.

Setting up a sitting drop involves pipetting a concentrated protein solution into a well (these can be built into the tray, microbridges that are added into standard trays, or a large sandwich box). This method typically uses higher volumes, but screening crystallization conditions typically uses the ratio of 2 µL protein solution: 2 µL reservoir solution. Once the drop has been setup, the reservoir needs to be sealed either by tape for a tray, coverslip for
microbridges, or closing the sandwich box with additional vacuum grease to prevent evaporation.

Once an initial crystallization condition is determined, it can be optimized for better quality, bigger size, or any other number of reasons. Typically the easiest method is to observe the drops and see where precipitation is forming and decrease the concentration of the crystallization agent, the protein concentration, or both. When a lot of small crystals have formed within a drop its typically a sign that your drop is within the nucleation region and was not able to transition to the metastable region, where crystal growth occurs. Again, this can be fixed by decreasing the crystallization agent, protein concentration, or both. Another thing to try is varying is the ratio of protein: precipitant in the drop. These are the “easy” things to try first. Further optimization may involve growing in different temperatures, varying the temperature during crystal growth, micro- or macroseeding, and/or testing an additive screen. (Hazemann et al. 2005; Budayova-Spano et al. 2007; Oksanen et al. 2009; Golden et al. 2015)

Molecular Replacement

The phase of each diffraction spot is lost during data collection and needs to be determined to accurately solve a structure. Molecular replacement (MR) is a method utilized to solve the phase of a new data set by using the structure factors from previously solved structures, either solved by crystallography, nuclear magnetic resonance, or electron microscopy, to estimate the phases for the new data. The process involves rotating the model (the known structure) along three rotation axes followed by translation until the model aligns
with the target (unknown structure). These parameters are adjusted until the model’s calculated structure factors, $|F_C|$, are in agreement with the target structure’s observed structure factors, $|F_O|$. From the diffraction experiment, the measured intensities are converted to structure factors, $|F(hkl)|^2$. Once the structure factors are calculated, the Patterson function can be used to generate a Patterson map. The Patterson function is an inverse Fourier transform of the structure factors shown below (Patterson, 1935):

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)|^2 \cos 2\pi (hu + kv + lw)$$

The Patterson function generates a vector map of atomic distances, which gives information on atomic positions within the molecule as well as symmetry related molecules. The need for structural homology is that the model and target structures need to generate similar Patterson maps, so the rotation and translation calculations will find a solution for the target.

Molecular replacement is highly successful with molecules containing ≥ 40% sequence identity, but is possible with sequence identities as low as 30%. Some of the MR programs used are MOLREP and Phaser. (Vagin & Teplyakov 1997; Keegan & Winn 2007; McCoy et al. 2007; Long et al. 2008) The MR program within Phenix is Phaser. (McCoy et al. 2007; Adams et al. 2010) Phaser MR and MOLREP are present in the CCP4 suite. (McCoy et al. 2007; Winn et al. 2011) The resulting score from MOLREP is given as a combinatorial score from the correlation coefficient and the packing function. Ideally, a score above 3.0 is obtained, while > 2.0 is acceptable. Any score ≤ 2.0 should be viewed with caution. Within Phaser a translation function Z-score (reported as TFZ) and log-likelihood gain are give as indicators of how correct the solution is. Ideally the TF Z-score is above 8,
with anything above 7 being acceptable. A TF Z-score below 7 should be viewed with caution. The log-likelihood gain (LLG) should be positive and as high as possible.
References


