

Use of CelR as a model to understand glycoside hydrolase activity

By

Nathaniel Jacob Kuch

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2022

Date of final oral examination: 10/05/2022

The dissertation is approved by the following members of the Final Oral Committee:

Brian G. Fox, Professor, Biochemistry

James A. Dumesic, Professor Emeritus, Chemical and Biological Engineering

Srivatsan Raman, Associate Professor, Biochemistry

Alessandro Senes, Professor, Biochemistry

Ophelia Venturelli, Assistant Professor, Chemical and Biological Engineering

## **Acknowledgments:**

There are a number of people who I want to thank for helping me be successful in graduate school, whether directly or indirectly. First, I would like to thank Brian Fox for his mentorship throughout my time in his lab. He has helped me develop scientific, critical thinking, and general work skills; along with giving me the opportunity and space to explore other facets of biotechnology like my marketing internship or my mentorship of students.

I would like to thank the other members of the Fox lab, both past and present. Evan Glasgow and Kirk Vander Meulen were invaluable as sounding boards to help me refine experimental designs and keep me focusing on tackling problems one day at a time. David Rivera-Kohr has helped with working through the frustrations and stress that come with research, and I will miss our One Piece conversations once I've graduated. My students, Mark Kutschke and Alex Parker have been a joy to work with, and it astounds me how much they've grown and how helpful they have been, taking over entire projects from experimental design to data analysis. They've both been inquisitive students who've pushed me to learn the material better to keep up with their questions as their own understanding of our work has grown.

Next, I would like to thank Curtis Knox and everyone on the marketing team at Promega for making my internship both fun and educational. I learned a lot about how the biotechnology industry works in general, and about the role that marketing plays in particular. On that note, I would like to thank the Biotechnology Training Program for enabling me to go on my internship to Promega, and also for encouraging me to think outside of my corner of science. I believe this encouragement played a large part in my successful interdisciplinary collaborations with the Dumesic lab.

The Dumesic lab, and everyone else involved with the Great Lakes Bioenergy Research Center have also been a great help in my development as a scientist. They have provided the resources and opportunity to learn more about fields of science that I would have never thought about otherwise. I would also like to thank the Integrated Program in Biochemistry for the opportunity to pursue my PhD here at UW Madison, and for all the support over the years that I've been here, especially Kate Ryan, who has been a treasure trove of knowledge and extremely helpful in answering all of my questions in my time working toward this dissertation. Additionally, I would like to thank the members of my thesis committee for their help over the years.

Finally, I would like to thank my friends and family for their support over these past six years. Their support has made this journey much lighter, from daily calls to coming out to stay with me for a month when I broke my leg. I would especially like to thank my Dungeons and Dragons group; Sam Gowland, Gillian Kalmey, Emily Tran, and Jennie Rhiner. You've all been a bright point throughout not just the day-to-day stress of graduate school, but also the past two years living through the pandemic. How you have managed to put up with my shenanigans this long I don't know, but I'm thankful for your presence every day.

**Abstract:**

Efficient conversion of lignocellulosic biomass into chemically and biologically upgradeable components is a cornerstone of biofuel and bioproducts research. As such, learning more about how the thermochemical pretreatments and enzymes used for lignocellulosic biomass deconstruction interact with one another and the biomass used is a key part of advancing the field. This work begins with characterization of a family 9 glycoside hydrolase, CelR, determining the changes to kinetics, binding, and thermostability brought about by the presence or absence of carbohydrate binding domains. The structure of CelR is also described for the first time, and the potential mechanistic role of an active site adjacent calcium ion in coordinating residue motion to promote either substrate binding or product release is explored. Then, CelR is used as a model enzyme in the investigation of changes in lignocellulosic biomass from  $\gamma$ -valerolactone pretreatment through solid-state nuclear magnetic resonance (ssNMR) experiments. This work shows how the ratio of peaks corresponding to the 4-carbon of cellulose in lignocellulosic biomass strongly correlates with soluble sugar yields after hydrolysis with CelR. Spectral fitting is then used to deconvolute the 4-carbon peaks to identify changes in specific subdomains of the cellulose throughout pretreatment and enzyme hydrolysis. The  $I_{\beta}$  subdomain is shown to form in the course of pretreatment; and be resistant to not only hydrolysis by CelR but also by the commercial cellulase cocktail CTec3. This thesis concludes with proposing a novel system for classifying and thinking about multifunctional cellulase activity, providing a framework that can be used to support and direct future research endeavors. This thesis represents advances in the understanding of family 9 glycoside hydrolases, which are attractive targets for study and enzyme engineering efforts for their ability to interact synergistically with both endo- and exo-cellulase classes. It also presents strong examples of

using ssNMR to monitor chemical and physical changes to lignocellulosic biomass throughout the deconstruction process. The ratio of 4-carbon peaks that was found to correlate to hydrolysis yields could prove to be a potent tool for evaluating thermochemical pretreatments, while subdomain tracking was showcased as a powerful technique for identifying specific molecular changes and bottlenecks to enzymatic hydrolysis of lignocellulosic biomass. The tier system presented in the review of multifunctional cellulases advances dialogue in the field and expands what enzymes could be considered multifunctional and highlights under-explored research avenues that arise from the goal of fleshing out understanding of these tiers and how they integrate with one another.

**Table of contents:**

	<b>Page</b>
<b>Acknowledgements</b> .....	i
<b>Abstract</b> .....	iii
<b>Table of Contents</b> .....	v
<b>List of figures and tables</b> .....	vi
<b>List of frequently used abbreviations</b> .....	x
<b>Chapter 1</b> .....	1
Introduction.....	
<b>Chapter 2</b> .....	25
Contributions of carbohydrate binding modules and calcium to the activity of glycoside hydrolase family 9 cellulase CelR.....	
<b>Chapter 3</b> .....	68
Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pretreatment for enzymatic conversion of residual cellulose.....	
<b>Chapter 4</b> .....	123
Evolution of the cellulose microfibril through gamma-valerolactone assisted co-solvent and enzymatic hydrolysis.....	
<b>Chapter 5</b> .....	169
Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy.....	
<b>Chapter 6</b> .....	185
Summary for the general public.....	
<b>Chapter 7</b> .....	187
Significance & future directions.....	
<b>Appendix A</b> .....	194
Drop-NIMS: A combinatorial droplet microfluidic device integrated with NIMS mass spectrometry analysis.....	

## List of figures and tables:

	<b>Description</b>	<b>Page</b>
<b>Chapter 1</b>		
<b>Figure 1.1</b>	Evidence for historical fermentation.....	2
<b>Figure 1.2</b>	Relationships between biofuel classifications.....	3
<b>Figure 1.3</b>	Cartoon representation of major components in lignocellulosic biomass.....	5
<b>Figure 1.4</b>	Cartoon depicting the goal of thermochemical pretreatment and the specific workflow for $\gamma$ -valerolactone pretreatment.....	7
<b>Figure 1.5</b>	Cartoon and scanning electron micrograph of cellulose microfibril packing in plant cell walls.....	9
<b>Figure 1.6</b>	General mechanisms for inverting or retaining GHs.....	11
<b>Figure 1.7</b>	Cartoon showing endo and exo cellulase activity.....	12
<b>Figure 1.8</b>	Cartoon representation of processive hydrolysis of a cellulose chain	15
<b>Figure 1.9</b>	Cartoon structures of five GH9 enzymes.....	16
<b>Chapter 2</b>		
<b>Figure 2.1</b>	Overview of the CelR constructs used in this work.....	29
<b>Figure 2.2</b>	First derivative data from differential scanning fluorimetry protein melt experiments.....	31
<b>Figure 2.3</b>	Lines of best fit for binding and hydrolysis of PASC and Avicel.....	33
<b>Table 2.1</b>	Kinetic and binding parameters for hydrolysis of either PASC or Avicel.....	34
<b>Figure 2.4</b>	Cartoon and surface representations of CelR with and without CBM3c.....	36
<b>Figure 2.5</b>	View of the CelR active site cleft.....	37
<b>Figure 2.6</b>	Interactions and residue movement around active site $\text{Ca}^{2+}$ ...	39
<b>Figure 2.7</b>	Differences in ligation of $\text{Ca}^{2+}$ and positions of adjacent residues....	41
<b>Figure S2.1</b>	Relative activity for CelR constructs on PASC.....	55
<b>Figure S2.2</b>	Relative activity for CelR constructs on Avicel.....	56
<b>Figure S2.3</b>	Kinetic and binding data for CelR constructs on PASC.....	57
<b>Figure S2.4</b>	Kinetic and binding data for CelR constructs on Avicel.....	58
<b>Table S2.1</b>	Data collection and refinement statistics.....	59
<b>Figure S2.5</b>	Hydrolysis yields of CelR constructs after EGTA treatment.....	61
<b>Chapter 3</b>		
<b>Figure 3.1</b>	Schematic representation of plant cell wall.....	70
<b>Figure 3.2</b>	Example of run of the GVL pretreatment process.....	75
<b>Figure 3.3</b>	P39 poplar $^{13}\text{C}$ MAS-NMR spectra.....	77
<b>Figure 3.4</b>	P39 poplar CP/MAS-NMR spectra.....	78
<b>Figure 3.5</b>	P39 poplar Proton spin-lock edited (PSRE) CP/MAS NMR spectra..	80

<b>Table 3.1</b>	Reaction conditions, $X_{NMR}$ , and enzymatic sugar yields for various pretreated biomasses.....	81
<b>Figure 3.6</b>	P39 poplar HR/MAS NMR spectra.....	83
<b>Figure 3.7</b>	Enzymatic sugar yields of variably pretreated P39 poplar.....	84
<b>Figure 3.8</b>	Enzymatic sugar yields versus $X_{NMR}$ for variably pretreated P39 or variable biomasses pretreated with GVL.....	85
<b>Figure 3.9</b>	Enzymatic sugar yields versus $X_{NMR}$ for never-dried samples.....	86
<b>Figure 3.10</b>	Normalized, integrated areas of the main hydration resonance.....	87
<b>Figure 3.11</b>	$T_2$ relaxation time constants for water in the main pore structure of cellulose.....	88
<b>Figure 3.12</b>	Schematic of lignocellulosic biomass structural features at various stages of pretreatment.....	92
<b>Figure S3.1</b>	Normal CP/MAS spectrum for non-pretreated P39 poplar biomass...	102
<b>Figure S3.2</b>	Integrated signal for the regions denoted in Figure S3.1 as a function of spin-lock delay.....	103
<b>Figure S3.3</b>	P39 poplar normal and PSRE CP/MAS spectra.....	106
<b>Figure S3.4</b>	P39 poplar PSRE CP/MAS sub spectra corresponding to crystalline amorphous cellulose.....	107
<b>Figure S3.5</b>	HR/MAS NMR spectra with a varying Hahn-echo delays applied...	108
<b>Figure S3.6</b>	Integrated are for the regions denoted in Figure S3.5 as a function of the Hahn-echo delay.....	109
<b>Figure S3.7</b>	P39 poplar normal CP/MAS spectra throughout GVL pretreatment..	110
<b>Table S3.1</b>	CP/MAS integrated signals for variably pretreated P39 poplar.....	112
<b>Figure S3.8</b>	ATR-FTIR spectra for pretreated and unpretreated P39 poplar.....	113
<b>Figure S3.9</b>	Representative X-ray diffractogram for P39 poplar and XRD-derived estimates cellulose microfibril diameter throughout pretreatment.....	114
<b>Table S3.2</b>	Representative mass balance for GVL pretreatment of P39 poplar...	116
<b>Chapter 4</b>		
<b>Figure 4.1</b>	Schematic diagram of GVL-assisted mild acidolysis pretreatment and subsequent enzymatic hydrolysis with CelR.....	127
<b>Figure 4.2</b>	Analysis of switchgrass, sorghum, P39 and NM6 poplar composition over different GVL pretreatment conditions.....	129
<b>Figure 4.3</b>	Microfibril structure proposed in literature along with comparison of spectral fitting between literature and DS140 treated P39 poplar..	132
<b>Table 4.1</b>	Chemical shifts (ppm) and full width half maxima (FWHM) for C4 peak assignments of P39 and NM6 poplar.....	133
<b>Figure 4.4</b>	$X_{fit}$ vs. enzymatic sugar yield upon hydrolysis by CelR or CTec3.....	135
<b>Figure 4.5</b>	Changes of fitted cellulose domains resulting from GVL pretreatments of P39 and NM6.....	137
<b>Figure 4.6</b>	Changes in fitted cellulose domains in GVL pretreated P39 poplar after enzymatic hydrolysis by CelR.....	139



<b>Table 4.2</b>	$X_{fit}$ , Lateral Fibril Dimension (LFD) and Lateral Fibril Aggregate Dimension (LFAD) estimations for P39 and NM6 poplar based on NMR populations.....	140
<b>Table 4.3</b>	Segal Crystallinity Index (CrI), coherence length (L), and lattice spacing (d) for the (200) plane of cellulose microfibril and a schematic of cellulose microfibril cross section.....	142
<b>Table S4.1</b>	Mass Closure analysis on a dry basis of feedstocks for P39 Poplar, NM6 Poplar, Switchgrass, and Sorghum.....	157
<b>Table S4.2</b>	Composition analysis of GVL treated P39 Poplar, NM6 Poplar, Switchgrass, and Sorghum on a dry basis.....	157
<b>Table S4.3</b>	Hydrolysis of commercial substrates PASC and Avicel by CelR.....	158
<b>Figure S4.1</b>	Comparisons of $^{13}\text{C}$ CP/MAS NMR spectra of P39 Poplar treated at 140°C (DS 140) using PSRE or spin lock delays.....	160
<b>Figure S4.2</b>	Effect of removing para-crystalline and inaccessible surface area peaks from the spectral fitting procedure on P39 poplar DS 140 treated $\text{C}_4$ spectrum.....	161
<b>Table S4.4</b>	Remaining weight percentage of P39 cellulose throughout GVL-assisted and enzymatic hydrolysis steps.....	162
<b>Figure S4.3</b>	Changes of fitted cellulose regions in the $\text{C}_4$ peak resulting from GVL pretreatments for P39 and NM6 poplar.....	162
<b>Figure S4.4</b>	Changes of fitted cellulose regions in the $\text{C}_4$ peak resulting from enzymatic hydrolysis with CelR.....	163
<b>Figure S4.5</b>	Changes of fitted cellulose regions in the $\text{C}_4$ peak resulting from enzymatic hydrolysis with CTec3.....	163
<b>Figure S4.6</b>	WAXS plots for DL 100, DS 120, and DS 140 GVL-treated P39 poplar.....	164
 <b>Chapter 5</b>		
<b>Figure 5.1</b>	Simplified structure of LCB reacting with different classes of enzymes.....	170
<b>Figure 5.2</b>	Tiers of function in MFCs.....	172
<b>Figure 5.3</b>	Specificity ratios for enzymes assayed on substrates with distinct backbone compositions.....	174
<b>Figure 5.4</b>	Chord plot illustration of the connections between GH and other domains in Tier 2 and 3 enzymes.....	176
 <b>Appendix A</b>		
<b>Figure A.1</b>	Assembly and operation of microfluidics device.....	197
<b>Figure A.2</b>	Droplet enzyme reaction with mass spectrometry barcode.....	205
<b>Figure A.3</b>	Rapid screening of glycosyl hydrolases with model hemicellulose compounds.....	208
<b>Figure SA.1.1</b>	PDMS molding.....	214
<b>Figure SA.1.2</b>	Align glass layer on PDMS mold using markers.....	216
<b>Figure SA.1.3</b>	Secure the glass slide with 2 Neodymium magnets.....	216

<b>Figure SA.1.4</b>	Separate the droplet chip (Glass/PDMS) from the master mold.....	217
<b>Figure SA.1.5</b>	Assemble NIMS and microfluidic chip with a chip holder.....	218
<b>Figure SA.1.6</b>	Generate droplets using BIO-RAD's QX200 Droplet Generator.....	220
<b>Figure SA.1.7</b>	Loading droplet cocktail into the microfluidic system.....	221
<b>Figure SA.1.8</b>	Seal the droplet chip against the NIMS surface.....	221
<b>Figure SA.1.9</b>	Merging droplets using an AC voltage treater.....	222
<b>Figure SA.1.10</b>	Separate the NIMS chip for MS imaging.....	223
<b>Figure SA.2</b>	Droplet Array Printing and MS Imaging.....	224
<b>Figure SA.3</b>	A demonstration of droplet pairing and merging.....	225
<b>Figure SA.4</b>	Alternative versions of the microfluidics chip with multiple sub-wells.....	226
<b>Figure SA.5</b>	Droplet random pairing and merging on ITO-coated glass.....	226
<b>Figure SA.6</b>	Specific products of reactions shown in Figure A.3c-e.....	228
<b>Figure SA.7</b>	Verapamil and NIMS-tagged substrates used in this study.....	230
<b>Figure SA.8</b>	Structures of mass spectrometry barcodes used in this study.....	231
<b>Figure SA.9</b>	Replicate counts for chip experiments shown in Fig. A.3.....	231

## List of frequently used abbreviations:

AS	Accessible surface
CAZy	Carbohydrate Active Enzyme Database
CBM(x)	Carbohydrate binding module (from family x)
CP/MAS	Cross-polarization magic angle spinning
CrI	Segal crystallinity index
DL	Delignification
DMSO	Dimethyl sulfoxide
DS	Desaccharification
DSF	Differential scanning fluorimetry
ESI	Electrospray ionization
FTIR	Fourier transform infrared
FWHM	Full width half maxima
G1	Glucose
G2	Cellobiose
GH(x)	Glycoside hydrolase (from family x)
GLBRC	Great Lakes Bioenergy Research Center
GVL	$\gamma$ -valerolactone
HR/MAS	High-resolution magic angle spinning
IAS	Inaccessible surface
ITO	Indium tin oxide
LCB	Lignocellulosic biomass
LDI	Laser desorption ionization
LFAD	Lateral fibril aggregate dimension
LFD	Lateral fibril dimension
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MeCN	Acetonitrile
MFC	Multifunctional cellulase
MSI	Mass spectrometry imaging
NIMS	Nanostructure-initiator mass spectrometry
NMP	N-methyl-pyrrolidinone
NMR	Nuclear magnetic resonance
OMAAT	OpenMSI Arrayed Analysis Toolkit
PASC	Phosphoric acid swollen cellulose
PDMS	Polydimethylsiloxane
PSRE	Proton spin-relaxation edited
ssNMR	Solid-state nuclear magnetic resonance
THF	Tetrahydrofuran
WAXS	Wide-angle X-ray scattering
X1	Xylose
X2	Xylobiose
X3	Xylotriose
X4	Xylotetraose
XRD	X-ray diffraction

## **Chapter 1:**

### **Introduction**

#### **Fermentation:**

Having a basic understanding of fermentation is important for understanding much of the research done on biofuels and bioproducts because it is one of the most commonly used technologies is their production (1-5). Pertinent to the work presented here, a large body of research in the field focuses on generating sugars (such as glucose) that are then fermented by any of a number of microorganisms into either biofuels/bioproducts or their precursors (6-8). As such, this thesis will begin with a brief discussion of fermentation and its history.

Fermentation is a metabolic process where organic molecules are converted into various alcohols, acids, and carbon dioxide in the absence of oxygen (9). The traditional fermentation molecular pathway converts sugars into pyruvate, which is then converted to ethanol as the end waste product to maintain redox balance (10-12). Fermentation is attractive for biofuels and bioproducts research because, in addition to flexibility of inputs leading to sugar, the output of the reaction can also be changed to any of a number of desirable products (10-12).

Humanity has used fermentation since antiquity as both a preservation and production method (13). The ubiquity of fermentation across the world is evidenced by representations in art and in residue from pottery used to store fermented goods; as shown in Figure 1.1A and 1.1B (adapted from Motsch, et al.) respectively (14). Scientific study of fermentation began to take off in the mid-1800s with the proposal by Schwann and Kützing that yeast was responsible for conversion of sugar to ethanol, made possible by Leeuwenhoek's invention of the microscope in the late-1600s (9).



Figure 1.1: Illustration from a German illuminated manuscript (Hans Franck, Biermesser, Mendel Band I (1506), pg 125v) depicting a beer measurer (A) and a pottery fragment from Vix-Mont Lassois, France (B) analyzed for evidence of fermentation residue, picture adapted from Motsch, et al. (14).

Production of compounds other than ethanol took off in the early-1900s, and included products such as glycerol, acetone, and butanol (9). As fermentation technology advanced, it began to be used as a production platform for a variety of bioproducts like antibiotics, and biofuels. Currently, although attention has shifted from ethanol-based biofuels to other compounds, fermentation remains an integral part of the bioprocessing pipeline (9,15,16).

### **Biofuels & bioproducts:**

Biofuels have been in use since the beginning of the automotive industry, with a number of bioethanol blends, such as Agrol, being widely used until petroleum based fuels became cheaper in the aftermath of World War 2 (17). Interest in biofuels was renewed in the 1970s due to the oil crisis, and since then the field of biofuels research has expanded to include the

production of non-fuel bioproducts as potential replacements for non-fuel chemicals derived from petroleum (17-20).

As biofuel and bioproduct technologies have been developed, multiple classification systems have been used to organize the field. Two of the more popular systems are the generation and processing/production method systems. As shown in Figure 1.2, the generation system breaks biofuels/bioproducts into groups by feedstock with first generation biofuels using food or starchy feedstocks such as corn and sugar cane, the second generation using lignocellulosic or waste biomass such as switchgrass and manure, and the third generation using oils derived from microalgae (17). The processing or production method system classifies biofuels by the technologies involved in their production, with conventional biofuels relying on well-established and commercial technology, while advanced biofuels use technologies that are the subject of ongoing research and development (17).

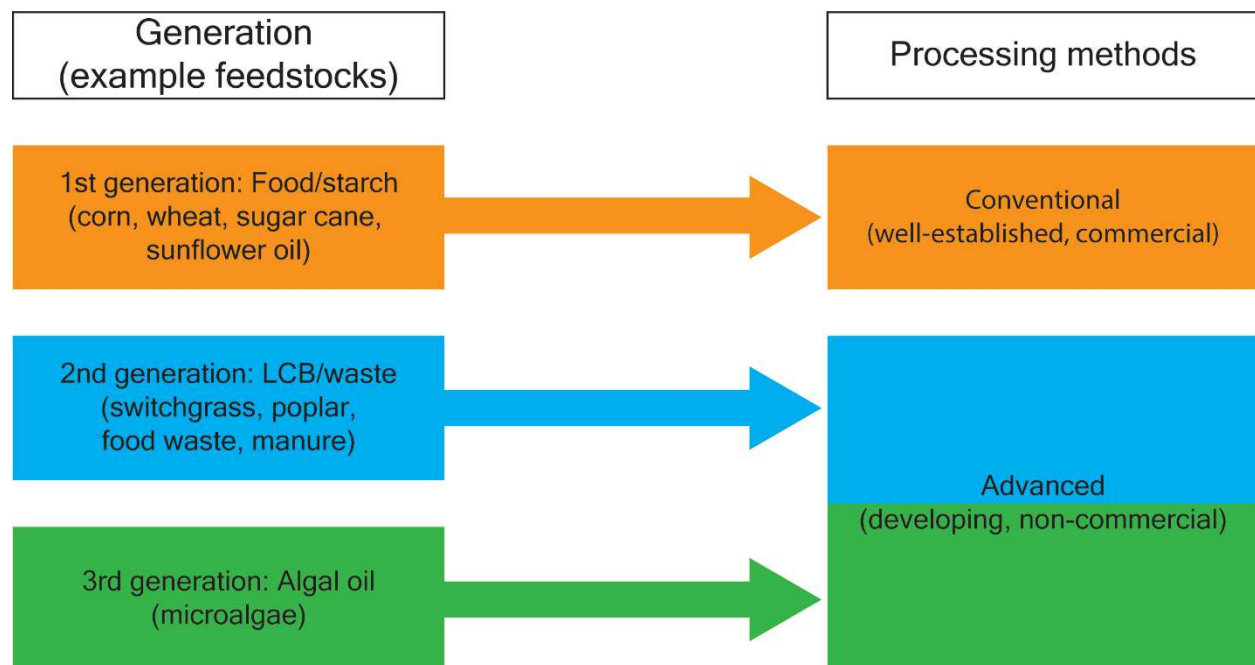


Figure 1.2: Diagram showing the relationship between two common frameworks when discussing biofuels: generation and processing method. Classification of biofuel generation is made by feedstock, while processing methods are classified by the level of technological development and ongoing research on a given method.

The transition from first to second generation biofuels was, in part, motivated by the food vs. fuel debate, which raised questions about the economic and environmental impacts of using edible material to produce fuel instead of feeding people and livestock (1,21). To side-step this issue, researchers began to transition to other feedstocks that could produce fermentable sugars, namely lignocellulosic biomass (LCB), which refers to the complex collection of multiple sugars and aromatic polymers of plant primary and secondary cell walls that serve a structural role; as opposed to the energy storage role of the starch material and already soluble sugars used in first generation biofuels (8).

### **Lignocellulosic biomass composition:**

LCB is formed from three broad classes of polymers arranged in a complex matrix that has evolved to resist degradation (8,22). The first is cellulose, a glucose polymer consisting of  $\beta$ -1,4 glycoside bonds which aggregate to form crystalline fibrils. The second is hemicellulose, which refers to a number of different polysaccharides, such as xylan, mannan, and xyloglucan (22,23). Because hemicellulose can be any of a number of polysaccharides, there is no one common structure that is adopted by hemicelluloses. The final component of LCB is lignin, which is a branching, covalently linked polymer composed of aromatic compounds (24,25).

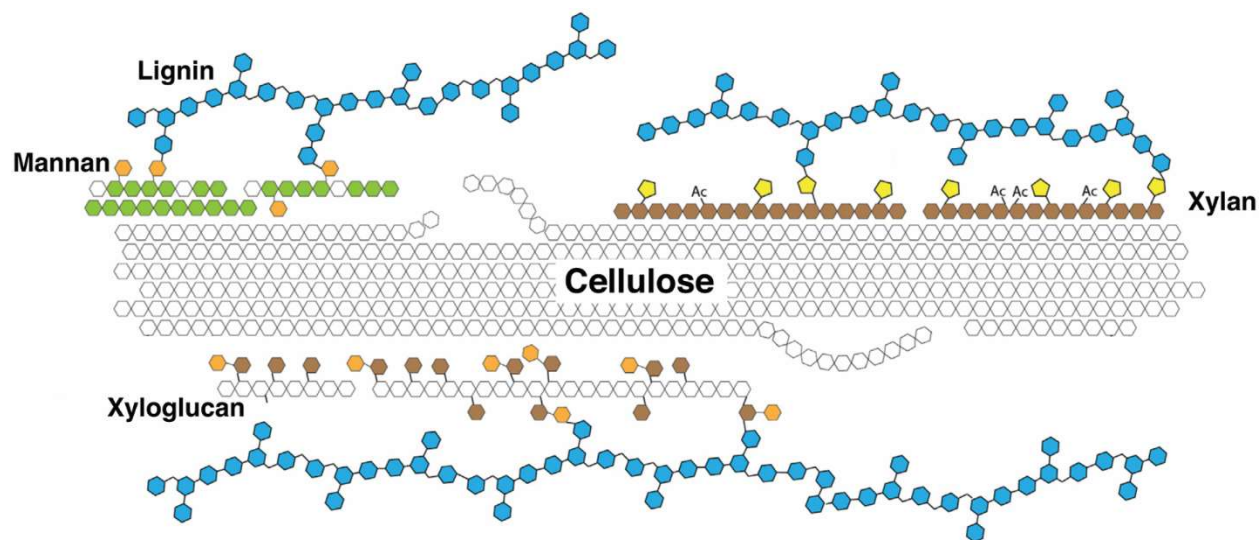


Figure 1.3: Adapted from Glasgow, et al. (26). Cartoon representation of major components in lignocellulosic biomass: cellulose; hemicellulose (xylan, mannan, and xyloglucan given as examples); and lignin.

Figure 1.3, adapted from Glasgow, et al., depicts a cartoon representation of LCB (26). The organization of components in the figure aligns with current models of the cell wall, with cellulose forming a core crystalline fibril which associates with hemicellulose at the surface, and through the hemicellulose with lignin (22,23,25,26). The exact ratios of LCB components varies from species to species, and within a plant based on tissue and cell wall type. Generic ranges for primary cell walls (differentiating/actively growing cells) are 20-30% cellulose, 20-45% hemicellulose and 5-10% pectins (similar to hemicellulose but often composed of heteropolymers of galacturonan) while secondary cell walls (terminally differentiated cells) range from 40-80% cellulose, 10-40% hemicellulose, and 5-25% lignin (24,25). In addition to these major components, there are also a number of proteins and other molecules associated with the cell wall for synthesis, modification, and maintenance purposes (24).



**Thermochemical pretreatments:**

As a key structural component in plants, evolutionary pressure has favored LCB that provides resistance to degradation as a defense against pathogens. However, this resistance complicates the use of LCB as a biofuels/bioproduct feedstock (5,8,27,28). Because efficient deconstruction of LCB is a key requirement of biofuel/bioproduct production, numerous degradation strategies have been explored (3,4,8,29). Generally, these strategies fall into two broad categories: thermochemical processes, or enzymatic hydrolysis. Both categories have advantages; thermochemical processes typically have much faster reaction times, while enzymatic hydrolysis has high product specificity and can operate under mild conditions (6,7,30). The downsides for these categories are that thermochemical treatments often require harsh chemicals and high temperatures, while enzymes are expensive and require long reaction times to produce high soluble sugar yields (5,29,30).

The combination of these factors has led to a paradigm where both strategies are employed, with an initial thermochemical pretreatment being done to remove lignin, hemicellulose, and the other minor components from LCB, followed by enzymatic hydrolysis of the then more reactive cellulose fraction (1,5,28,30). The general premise of pretreatment is shown in Figure 1.5A, adapted from Walker, et al., which is Chapter 3 of this thesis (31).

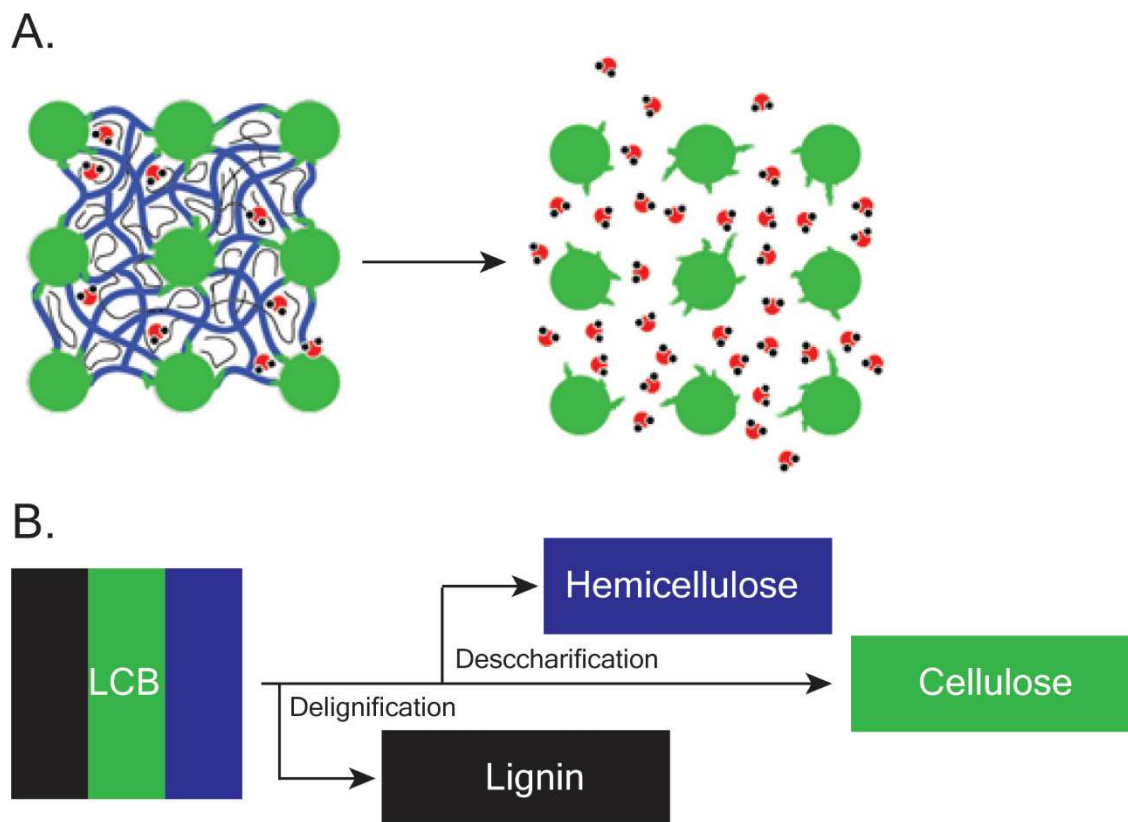


Figure 1.4: Generic cartoon depicting the goal of thermochemical pretreatment (A) adapted from Walker, et al., and the specific workflow for  $\gamma$ -valerolactone (GVL) pretreatment (B) (31). In panel A cellulose (green) is isolated from hemicellulose (blue) and lignin (black) through the pretreatment process. In panel B LCB (black/green/blue stripes) is separated through two sequential steps into pure lignin (black), hemicellulose (blue), and cellulose (green) fractions.

While there are many methods for thermochemically pretreating LCB, for example: dilute acid, ammonia fiber expansion, ionic liquids, etc.; the method most relevant to this work is  $\gamma$ -valerolactone (GVL) pretreatment (29,31). GVL pretreatment is a versatile technique that uses dilute acid in a GVL/water cosolvent system to hydrolyze lignin, hemicellulose, and cellulose. Hydrolysis can be done at elevated temperatures to produce products like 5-hydroxymethylfurfural, furfural, or levulinic acid, or in successive steps at increasing temperatures to fractionate the LCB into first lignin ( $\sim 100$  °C) and then hemicellulose ( $\sim 140$  °C), leaving solids containing almost entirely cellulose (2,29,31). In the case of LCB fractionation, liquid lignin and hemicellulose fractions can be biologically or chemically upgraded to a variety

of valuable bioproducts, while the cellulose solids can be enzymatically hydrolyzed to produce fermentable sugars, which is covered in greater detail in chapters 3 and 4 of this work (2,31).

### **Cellulose structure:**

To enable more efficient degradation, a great deal of work has been done to better understand the structure of cellulose in LCB and how it changes throughout various pretreatment processes (22,23,31,32). On a primary level, the  $\beta$ -1,4 glycoside bonds in cellulose result in linear chains that form crystals through hydrogen bonding between parallel chains to form sheets and hydrophobic interactions between faces of stacking cellulose sheets (23). The crystalized cellulose chains are called microfibrils, which have been shown to form aggregates called macrofibrils (22,32). Crystalline cellulose has been observed to fall into one of four different types. Type I cellulose, the naturally occurring form, consists of parallel chains within the microfibril and is the relevant form for the work presented in this thesis, but briefly, type II has antiparallel chains and forms after alkali treatment of type I cellulose, type III is formed through treating either type I or II with amines, and type IV is formed from treating type III with glycerol at high temperature (33).

Within type I cellulose, microfibrils are typically 3-4 nm in diameter, with macrofibrils varying in size from 14-23 nm in diameter depending on other factors such as lignin and hemicellulose content, with aggregates of macrofibrils being observed with diameters up to 60 nm (32). Perhaps unsurprisingly, the different chemical environments within and surrounding microfibrils give way to a multitude of subdomains on the surfaces and within the crystal, as shown in Figure 1.5A. Chapter 4 goes into further detail about detection of subdomains via solid-

state nuclear magnetic resonance (ssNMR), and how subdomain distributions change over the course of enzymatic hydrolysis (34-36).

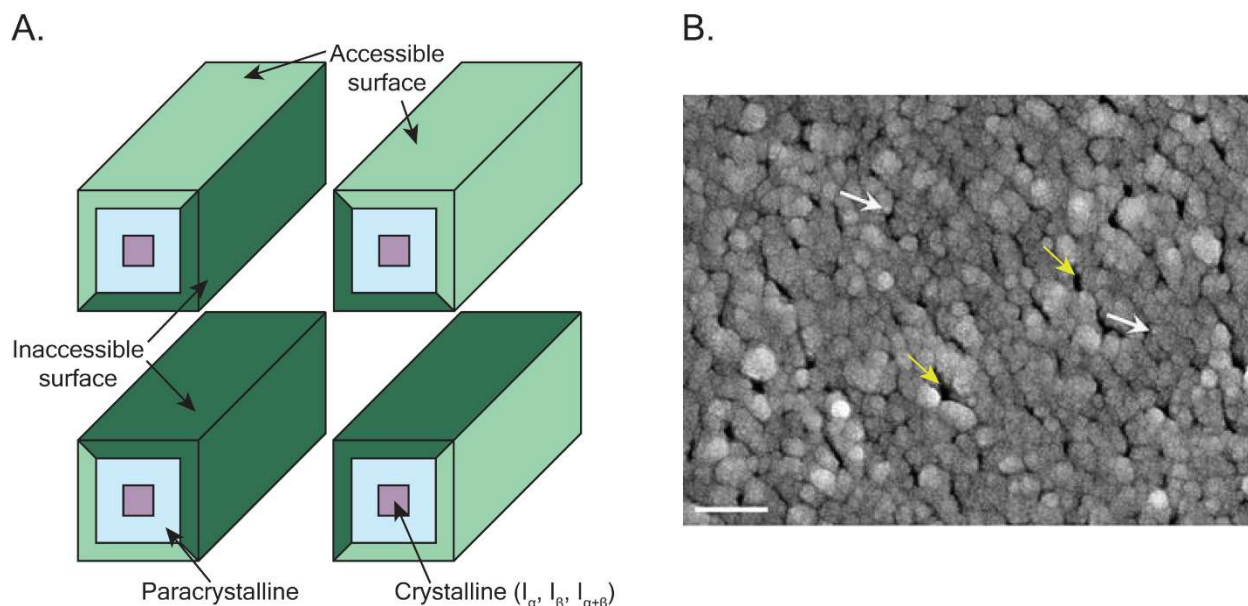


Figure 1.5: Cartoon (A) and scanning electron micrograph (B) of cellulose microfibril packing in plant cell walls. Microfibrils contain an interior crystalline region (purple) surrounded by paracrystalline cellulose (cyan). The microfibril surface is divided into two distinct populations, accessible (light green) and inaccessible (dark green) based on placement within the macrofibril. SEM image adapted from Donaldson et al. shows cellulose microfibrils (white arrows) aggregating in tight bundles with open regions between the microfibrils (yellow arrows); scale bar = 100 nm (32).

Figure 1.5 presents a cartoon model of four cellulose microfibrils arranged into a macrofibril, colored by subdomain. The purple, crystalline core of the microfibrils contains a combination of  $I_{\alpha}$  and  $I_{\beta}$  cellulose, the cyan region just outside the core represents paracrystalline material, and the light green and dark green regions represent the accessible and inaccessible surfaces respectively. The difference between the surfaces comes from the packing of microfibrils into macrofibrils; as shown in Figure 1.5B, adapted from Donaldson et al., inaccessible surfaces occur when microfibrils pack against each other, and accessible surfaces occur when microfibrils are in contact with hemicellulose, lignin, or the surrounding

environment (32,37,38).  $I_{\alpha}$  and  $I_{\beta}$  cellulose in the microfibril core differ in the arrangement of the 6-carbon and associated hydroxyl which results in  $I_{\alpha}$  having an alternating pattern (A-A'-A-A'...) forming a triclinic crystal structure, whereas  $I_{\beta}$  has alternating sheets (B-B-B... or B'-B'-B'...) forming a monoclinic crystal structure (34). Of the two forms of cellulose,  $I_{\beta}$  is the more thermodynamically stable, with current thoughts in the field being that rearrangement of  $I_{\alpha}$  to  $I_{\beta}$  occurs as a sliding of the sheets across one another, coupled with a movement of the 6-carbon and hydroxyl group (39,40).

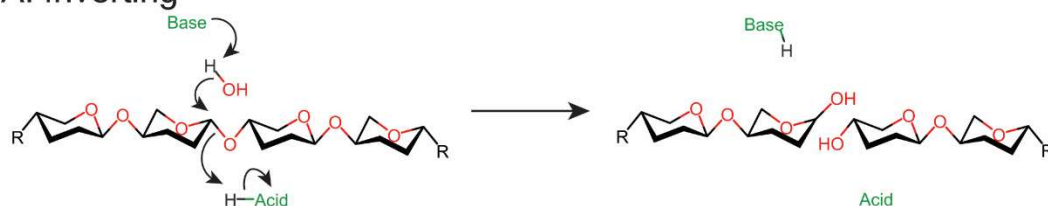
### **Carbohydrate active enzymes:**

The enzymes used to hydrolyze pretreated LCB, regardless of the specific pretreatment method, are classified as carbohydrate active enzymes; a classification encompassing a number of activities that revolve around the synthesis, modification, and degradation of carbohydrates (27,41,42). The Carbohydrate Active Enzyme (CAZy) database (<http://www.cazy.org>) classifies half of these enzymes, based on sequence alignment, as glycoside hydrolases (GHs) which degrade polysaccharides through hydrolysis of the glycosidic bonds (43-46). CAZy also uses sequence similarity to group GHs into numbered families that often have similar mechanisms, substrates, and structural features.

Glycoside bond hydrolysis is almost always done through either an inverting or retaining mechanism, as shown in Figure 1.6. Inverting GHs (Figure 1.6A) activate water with a catalytic base, and protonate the leaving group with a catalytic acid, resulting in inversion of stereochemistry at the anomeric carbon. In contrast, retaining GHs (Figure 1.6B) use the catalytic base to make a direct nucleophilic attack on the glycoside bond and form a covalently linked enzyme-substrate intermediate that is then released by water cleaving the enzyme-substrate bond,

resulting in retention of stereochemistry at the anomeric carbon (47-50). Typically, the amino acids used as the catalytic base/acid are a combination of aspartate and/or glutamate (47-49). While GH is used as a generic label, further classification is often done by the polysaccharide that they target, for example this work focuses on the cellulase CelR, which hydrolyzes cellulose specifically (26,51,52). Although usually a handy descriptor, this nomenclature is sometimes incomplete as some GHs are capable of hydrolyzing multiple substrates (26,52). This is referred to commonly as multifunctionality, and occasionally an enzyme with high substrate promiscuity is classified based on only one of its potential substrates (26,52).

### A. Inverting



### B. Retaining

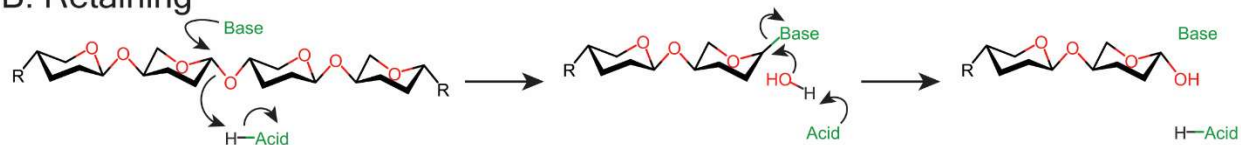


Figure 1.6: General mechanisms for inverting (A) or retaining (B) GHs (47,48). For clarity, hydroxyls, hydrogens, and 6-carbons are not shown. Inverting GHs cleave the glycoside bond in one step through an activated water, while retaining GHs form a temporary covalent bond between the enzyme and substrate. Both mechanisms often use aspartate and/or glutamate as catalytic residues.

Beyond classifying GHs by their target substrate, another typology is to classify them based on the location of glycoside bonds they can target relative to the ends of polysaccharide chains (26,53-55). As shown in Figure 1.7 using cellulases as an example, the two classifications are endo- (purple) and exo- (teal and yellow) GHs, with exo-GHs targeting either the reducing (yellow) or non-reducing (teal) end of the polysaccharide chain and endo-GHs not having a

restriction on hydrolyzing bonds in relation to either end of the polysaccharide chain (55-57). For long polysaccharides, endo-GHs are more likely to hydrolyze bonds in the interior of the chain and generate novel ends that can be targeted by exo-GHs; which results in synergy between these three classes where a combination of these activities produces soluble reducing sugar faster than equimolar reactions containing only one of these activities (26,55).

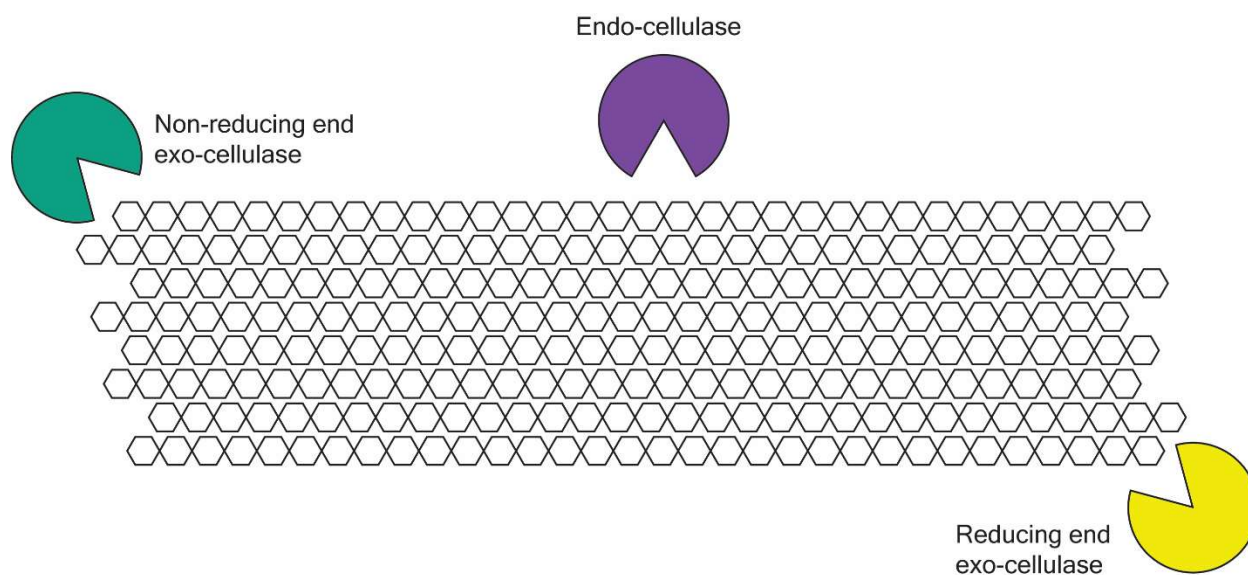


Figure 1.7: Cartoon showing the three broad types of cellulase activity. These types are endo-cellulases (purple), which can hydrolyze glycosidic bonds at any position in the cellulose, and exo-cellulases, which can only hydrolyze glycosidic bonds at cellulose chain ends. Depending on which end of the cellulose chain is targeted, exo-cellulases are further classified as either non-reducing end (teal) or reducing end (yellow) targeting.

Relevant to this work, and also discussed in chapter 2, there exists some ambiguity in the exact definition of exo-cellulases. Most seem to agree on classifying enzymes that exclusively produce glucose or cellobiose from hydrolyzing a chain end as an exo-cellulase, but members of GH family 9 (GH9) have been shown to produce cellotetraose from hydrolyzing chain ends and consequently have been frequently referred to as endo-cellulases (53,55,58-60). For the remainder of this work, the term endo-cellulase or endo-GH will refer to GHs whose activity are

not tied or referenced to a chain end and exo-cellulase or exo-GH will refer to GHs who always catalyze reactions at some specific distance from a chain end. So GH9s that hydrolyze a polysaccharide chain end to produce cellotetraose will be referred to as exo-cellulases instead of endo-cellulases.

Carbohydrate binding modules (CBMs) are common accessory domains for carbohydrate active enzymes, including many GHs. CBMs primarily play a role in binding to target polysaccharides, with different CBM families having affinities for different polysaccharides, much like the different GH families (61-63).

In addition to polysaccharide binding, CBMs can also impact GH thermostability and substrate accessibility by, for example, disrupting hydrogen bonding between cellulose chains (64-66). Numerous studies have found that CBMs play a role in targeting appended domains to specific regions of LCB, and may serve in modulating substrate reactivity (62,67-69). Because of their roles in substrate binding and thermostability, and the often insoluble or partially insoluble nature of polysaccharides, especially cellulose and LCB, the interactions between CBMs and GHs has been an important aspect of understanding and improving LCB degradation for biofuel and bioproduct purposes (61,62,68).

### **Glycoside hydrolase kinetics & processivity:**

The insolubility of LCB has made research into CBMs important to the field because they can serve to anchor GHs to the substrate, thereby increasing the local concentration of substrate for the GH to act on (66,70). The insolubility of LCB also causes issues when measuring kinetic parameters of GHs, as classical Michaelis-Menten (MM) kinetics, and quasi-steady state approximations both assume free diffusion of both substrate and enzyme (71,72). This issue is



commonly addressed in one of two ways; using a soluble substrate such as carboxymethyl cellulose as a stand in for an insoluble polysaccharide, such as cellulose; or making modifications to the MM model to account for insoluble substrates (70,73-75). The matter is also complicated further on cellulose in particular with its structural heterogeneity, with some models including parameters that cover cellulose characteristics ranging from crystallinity to more abstract variables, like cellulose digestibility decreasing as it is hydrolyzed (76-78). Interestingly, a number of papers have found that, in practice, MM kinetics can fit cellulase kinetic data very well (70,74,75). One reason for this could be that amorphous cellulose or loose chain ends could have sufficiently free movement to produce results similar to if the substrate was able to freely diffuse (76).

An important topic when discussing cellulase kinetics, is processive activity. As illustrated in Figure 1.8, processivity refers to the enzyme-substrate complex persisting after release of the product to undergo another round of catalysis, rather than dissociating as occurs in MM kinetics (79,80). Processive hydrolysis has been well established for GH6 and GH7 exo-cellulases, as well as being suggested for some enzymes in GH5 and, pertinent to this work, GH9 (60,79-82). A number of reviews have covered the myriad methods used to experimentally determine cellulase processivity, but in brief, they have ranged from atomic force microscopy to comparison of the fraction of soluble vs. insoluble sugars produced, indicating the position of hydrolysis (53,60,79-83). Efforts to quantify processivity fall broadly into one of two categories: apparent processivity, and intrinsic processivity (79,84). Apparent processivity is the measure of catalytic events per processive run, which can be difficult to measure and change dramatically depending on the substrate and other assay conditions (79,84). Intrinsic processivity, on the other hand, describes the theoretical potential for processivity and has been applied to other processive

enzymes such as nucleic acid polymerases and helicases in addition to cellulases (79,84).

Mathematically, intrinsic processivity is defined as the catalytic rate coefficient ( $k_{cat}$ ) divided by the dissociation rate coefficient ( $k_{off}$ ) (79,84). Processivity, as it relates to this work, is discussed in further detail in chapter 2.

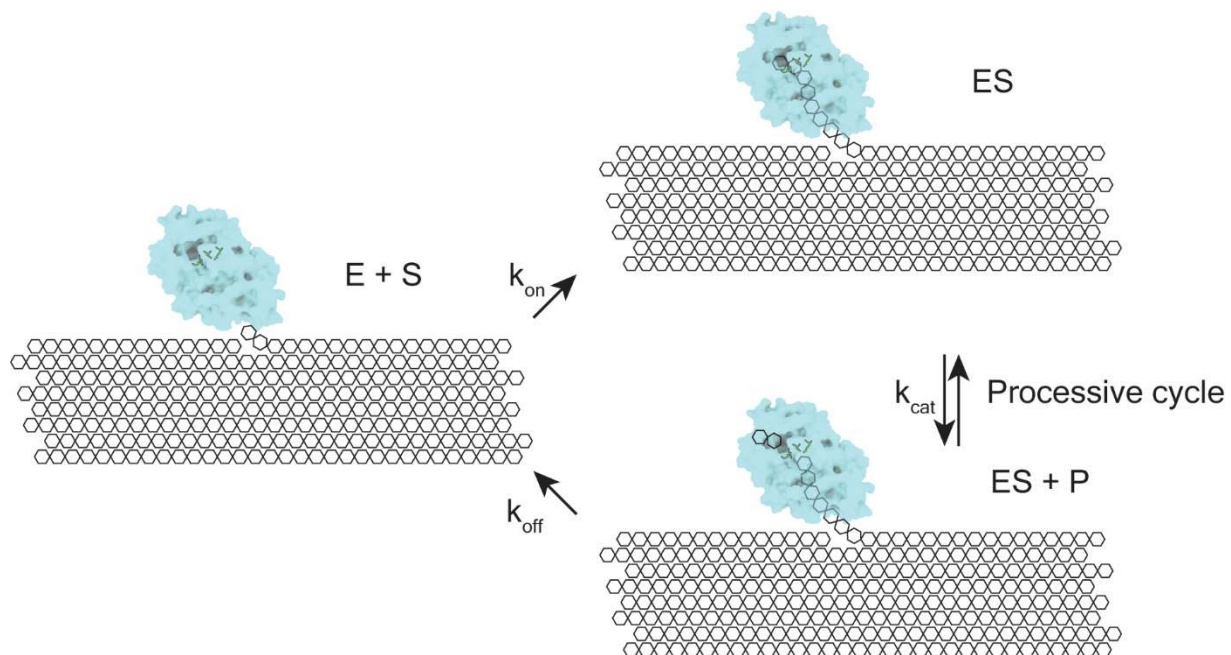


Figure 1.8: Cartoon representation of processive hydrolysis of a cellulose chain by 4C4C (50). The processive cycle occurs when the enzyme and substrate complex (ES) fully reforms after hydrolysis and product release rather than dissociating into separate enzyme and substrate ( $E + S$ ). This has been modeled as the ratio of  $k_{cat}/k_{off}$ .

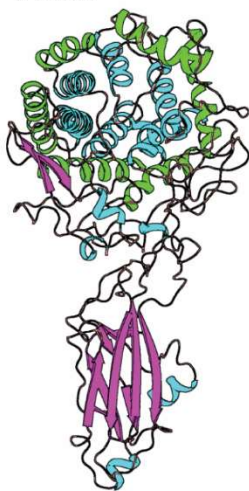
### Glycoside hydrolase family 9:

The cellulase that this work revolves around, CelR, is a member of GH9, with a single catalytic domain, family 3c CBM (CBM3c), and dockerin domain (85,86). As shown in Figure 1.9A, CelR's (7UNP's) CBM3c domain has a  $\beta$ -sandwich fold which packs tightly up against the  $\alpha/\alpha_6$ -barrel fold of the catalytic domain. This is a well-conserved domain arrangement, as shown in Figure 1.9B-C, where the  $\beta$ -sandwich fold of the CBM is packed tightly to the catalytic

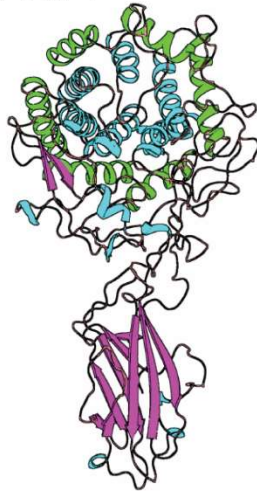
domain. The  $\alpha/\alpha_6$ -barrel fold of the catalytic domain is also well-conserved in GH9, even in the members without CBMs, as shown in Figure 1.9D-E.

GH9s are classified in the CAZy database as being cellulases, with the examples in Figure 1.9 covering a range of cellulase activities, with 7UNP, 4TF4, and 1G87 (Figure 1.9A-C) being annotated as processive endo-cellulases, 1UT9 (Figure 1.9D) a cellobiose producing exo-cellulase, and 1KS8 (Figure 1.9E) being an endo-cellulase (55-57,87). As noted in the earlier discussion of endo- and exo-cellulases, Figure 1.9A-C release predominantly cellotetraose from hydrolysis of the non-reducing end of cellulose chains, and so will be referred to as non-reducing end exo-cellulases moving forward (55,59).

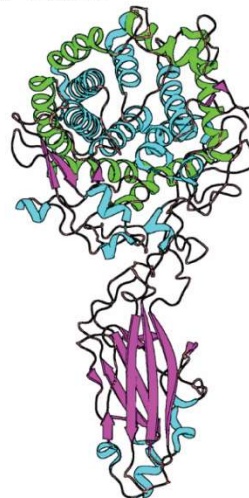
A. 7UNP



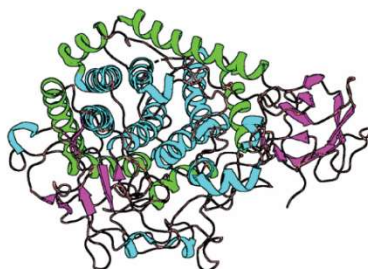
B. 4TF4



C. 1G87



D. 1UT9



E. 1KS8

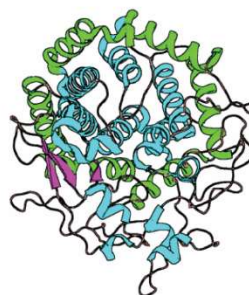


Figure 1.9: Cartoon representations of five GH9 enzymes, 7UNP, 4TF4 (55), 1G87 (87), 1UT9 (57), and 1KS8 (56). The conserved  $\alpha/\alpha_6$ -barrel fold is highlighted by coloring the exterior helices of the barrel green to contrast the rest of the helices colored blue. When CBMs are present, they adopt a  $\beta$ -sandwich fold and are tightly packed against the catalytic domain instead of being connected via a long, flexible linker.

It has been noted in literature that non-reducing end exo-cellulases like CelR or Cel9A (4TF4) have the capacity to act synergistically with both endo- and other exo-cellulases (55,60). This synergy has been attributed to these enzymes targeting chain ends, which allow synergistic interaction with endo-cellulases, and these enzymes producing cellotetraose, which is a substrate for other exo-cellulases (55,60). Structural features of GH9s, and CelR in particular, are further discussed in chapter 2, but briefly, there is evidence that the role of the CBM is to align the substrate and the enzyme active site cleft, to disrupt hydrogen bonding between the target cellulose chain and neighboring chains in the sheet, and to improve thermostability (54,83,88) (chapter 2). Additionally, processivity of these enzymes is thought to be thermodynamically driven by increasingly favorable binding energies as the active site cleft is progressively filled, coupled with some degree of motion by aromatic side chains along the active site cleft (58,89).

### **Thesis overview:**

Prior to the work presented in this thesis, there was no structure of CelR, and interactions between cellulases and GVL pretreatment had not been studied. Chapter 2 presents the first structures of CelR, including structures of the catalytic domain before and after hydrolysis of cellohexaose and identifies a group of calcium-coordinating residues that move in response to different substrates bound. Chapter 3 describes the relationship between resonances in the ssNMR spectra and reducing sugar yields after hydrolysis of pretreated LCB by CelR, covering a range of co-solvents used in a dilute acid pretreatment protocol. The ssNMR resonance identified

in Chapter 3 are explored further in Chapter 4, which follows the changes in cellulose microfibril subdomains throughout GVL pretreatment and enzymatic hydrolysis and identifies I $\beta$  cellulose as being unhydrolyzed even at yields exceeding 30% of the theoretical maximum. Chapter 5 reviews literature on multifunctional cellulases, and presents a three-tier approach to multifunctionality, including functional changes brought about by addition of CBMs or other domains. The work presented in this thesis is summarized for the general public in Chapter 6, while Chapter 7 reiterates the key findings from this work along with presenting a path for future research. Appendix A goes on to detail the development of an integrated microfluidics and mass spectrometry device for high-throughput enzyme characterization using GH8s and GH43s as example test cases.

#### Citations:

1. Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Lee, T. S., Tullman-Ercek, D., Voigt, C. A., Simmons, B. A., and Keasling, J. D. (2011) Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci U S A* **108**, 19949-19954
2. Alonso, D. M., Hakim, S. H., Zhou, S., Won, W., Hosseinaei, O., Tao, J., Garcia-Negron, V., Motagamwala, A. H., Mellmer, M. A., Huang, K., Houtman, C. J., Labbe, N., Harper, D. P., Maravelias, C., Runge, T., and Dumesic, J. A. (2017) Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Sci Adv* **3**, e1603301
3. Bhalla, A., Fasahati, P., Particka, C. A., Assad, A. E., Stoklosa, R. J., Bansal, N., Semaan, R., Saffron, C. M., Hodge, D. B., and Hegg, E. L. (2018) Integrated experimental and techno-economic evaluation of two-stage Cu-catalyzed alkaline-oxidative pretreatment of hybrid poplar. *Biotechnol Biofuels* **11**, 143
4. Cai, C. M., Zhang, T. Y., Kumar, R., and Wyman, C. E. (2013) THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chem* **15**, 3140-3145
5. Kucharska, K., Rybarczyk, P., Holowacz, I., Lukajtis, R., Glinka, M., and Kaminski, M. (2018) Pretreatment of Lignocellulosic Materials as Substrates for Fermentation Processes. *Molecules* **23**
6. Roman-Leshkov, Y., Barrett, C. J., Liu, Z. Y., and Dumesic, J. A. (2007) Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature* **447**, 982-985
7. Oke, M. A., Annuar, M. S. M., and Simarani, K. (2016) Mixed Feedstock Approach to Lignocellulosic Ethanol Production-Prospects and Limitations. *Bioenerg Res* **9**, 1189-1203

8. Ning, P., Yang, G., Hu, L., Sun, J., Shi, L., Zhou, Y., Wang, Z., and Yang, J. (2021) Recent advances in the valorization of plant biomass. *Biotechnol Biofuels* **14**, 102
9. Zeng, A.-P. (2021) Fermentation and Bioprocess Development. in *Kirk-Othmer Encyclopedia of Chemical Technology*. pp 1-71
10. Zheng, J., Tashiro, Y., Wang, Q., and Sonomoto, K. (2015) Recent advances to improve fermentative butanol production: genetic engineering and fermentation technology. *J Biosci Bioeng* **119**, 1-9
11. Li, T., Chen, X. B., Chen, J. C., Wu, Q., and Chen, G. Q. (2014) Open and continuous fermentation: products, conditions and bioprocess economy. *Biotechnol J* **9**, 1503-1511
12. Dos Santos Vieira, C. F., Mauger Filho, F., Maciel Filho, R., and Pinto Mariano, A. (2019) Acetone-free biobutanol production: Past and recent advances in the Isopropanol-Butanol-Ethanol (IBE) fermentation. *Bioresour Technol* **287**, 121425
13. Craig, O. E. (2021) Prehistoric Fermentation, Delayed-Return Economies, and the Adoption of Pottery Technology. *Current Anthropology* **62**, 9p
14. Motsch, A., Schorer, B., Bardel, D., Winkler, A., Sacchetti, F., Chaume, B., Della Casa, P., Buckley, S., Cafisso, S., Fries-Knoblach, J., Krausse, D., Hoppe, T., Stockhammer, P., and Spiteri, C. (2019) New insights into Early Celtic consumption practices: Organic residue analyses of local and imported pottery from Vix-Mont Lassois. *PLoS ONE* **14**, 19p
15. Humphrey, A. (1998) Shake flask to fermentor: What have we learned? *Biotechnol Progr* **14**, 3-7
16. Demain, A. L. (1998) Microbial natural products: Alive and well in 1998. *Nat Biotechnol* **16**, 3-4
17. Jeswani, H. K., Chilvers, A., and Azapagic, A. (2020) Environmental sustainability of biofuels: a review. *Proc Math Phys Eng Sci* **476**, 20200351
18. Ramos, J. L., and Duque, E. (2019) Twenty-first-century chemical odyssey: fuels versus commodities and cell factories versus chemical plants. *Microb Biotechnol* **12**, 200-209
19. Clomburg, J. M., Crumbley, A. M., and Gonzalez, R. (2017) Industrial biomanufacturing: The future of chemical production. *Science* **355**
20. Steen, E. J., Kang, Y. S., Bokinsky, G., Hu, Z. H., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-U182
21. Robertson, G. P., Hamilton, S. K., Barham, B. L., Dale, B. E., Izaurralde, R. C., Jackson, R. D., Landis, D. A., Swinton, S. M., Thelen, K. D., and Tiedje, J. M. (2017) Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes. *Science* **356**
22. Cosgrove, D. J. (2022) Building an extensible cell wall. *Plant Physiol* **189**, 1246-1277
23. Xu, H., Giannetti, A., Sugiyama, Y., Zheng, W., Schneider, R., Watanabe, Y., Oda, Y., and Persson, S. (2022) Secondary cell wall patterning-connecting the dots, pits and helices. *Open Biol* **12**, 210208
24. Gigli-Bisceglia, N., Engelsdorf, T., and Hamann, T. (2020) Plant cell wall integrity maintenance in model plants and crop species-relevant cell wall components and underlying guiding principles. *Cell Mol Life Sci* **77**, 2049-2077
25. Meents, M. J., Watanabe, Y., and Samuels, A. L. (2018) The cell biology of secondary cell wall biosynthesis. *Ann Bot* **121**, 1107-1125
26. Glasgow, E., Vander Meulen, K., Kuch, N., and Fox, B. G. (2021) Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy. *Curr Opin Biotechnol* **67**, 141-148

27. Rafiei, V., Velez, H., and Tzelepis, G. (2021) The Role of Glycoside Hydrolases in Phytopathogenic Fungi and Oomycetes Virulence. *Int J Mol Sci* **22**
28. Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F., and da Silva, S. S. (2018) The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization. *Bioresour Technol* **264**, 370-381
29. Luterbacher, J. S., Rand, J. M., Alonso, D. M., Han, J., Youngquist, J. T., Maravelias, C. T., Pfleger, B. F., and Dumesic, J. A. (2014) Nonenzymatic Sugar Production from Biomass Using Biomass-Derived gamma-Valerolactone. *Science* **343**, 277-280
30. Johnson, E. (2016) Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuel Bioprod Bior* **10**, 164-174
31. Walker, T. W., Kuch, N., Vander Meulen, K. A., Clewett, C. F. M., Huber, G. W., Fox, B. G., and Dumesic, J. A. (2020) Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pre-treatment for enzymatic conversion of residual cellulose. *ACS Sustain Chem Eng* **8**, 6551-6563
32. Donaldson, L. (2007) Cellulose microfibril aggregates and their size variation with cell wall type. *Wood Sci Technol* **41**, 443-460
33. Nunes, R. C. R. (2017) Rubber nanocomposites with nanocellulose. *Woodh Pub S Compos S* **75**, 463-494
34. Kono, H., and Numata, Y. (2006) Structural investigation of cellulose I-alpha and I-beta by 2D RFDR NMR spectroscopy: determination of sequence of magnetically inequivalent D-glucose units along cellulose chain. *Cellulose* **13**, 317-326
35. Wang, T., Yang, H., Kubicki, J. D., and Hong, M. (2016) Cellulose Structural Polymorphism in Plant Primary Cell Walls Investigated by High-Field 2D Solid-State NMR Spectroscopy and Density Functional Theory Calculations. *Biomacromolecules* **17**, 2210-2222
36. Atalla, R. H., and Vanderhart, D. L. (1984) Native cellulose: a composite of two distinct crystalline forms. *Science* **223**, 283-285
37. Hult, E. L., Larsson, P. T., and Iversen, T. (2000) A comparative CP/MAS C-13-NMR study of cellulose structure in spruce wood and kraft pulp. *Cellulose* **7**, 35-55
38. Larsson, P. T., Hult, E. L., Wickholm, K., Pettersson, E., and Iversen, T. (1999) CP/MAS 13C-NMR spectroscopy applied to structure and interaction studies on cellulose I. *Solid State Nucl Magn Reson* **15**, 31-40
39. Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose Ibeta from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **124**, 9074-9082
40. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose I(alpha) from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* **125**, 14300-14306
41. Amin, K., Tranchimand, S., Benvegnu, T., Abdel-Razzak, Z., and Chamieh, H. (2021) Glycoside Hydrolases and Glycosyltransferases from Hyperthermophilic Archaea: Insights on Their Characteristics and Applications in Biotechnology. *Biomolecules* **11**
42. Higgins, M. A., Whitworth, G. E., El Warry, N., Randriantsoa, M., Samain, E., Burke, R. D., Vocadlo, D. J., and Boraston, A. B. (2009) Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. *J Biol Chem* **284**, 26161-26173

43. Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022) The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* **50**, D571-D577
44. Baker, P., Hill, P. J., Snarr, B. D., Alnabelseya, N., Pestrak, M. J., Lee, M. J., Jennings, L. K., Tam, J., Melnyk, R. A., Parsek, M. R., Sheppard, D. C., Wozniak, D. J., and Howell, P. L. (2016) Exopolysaccharide biosynthetic glycoside hydrolases can be utilized to disrupt and prevent *Pseudomonas aeruginosa* biofilms. *Sci Adv* **2**, e1501632
45. Snarr, B. D., Baker, P., Bamford, N. C., Sato, Y., Liu, H., Lehoux, M., Gravelat, F. N., Ostapska, H., Baistrocchi, S. R., Cerone, R. P., Filler, E. E., Parsek, M. R., Filler, S. G., Howell, P. L., and Sheppard, D. C. (2017) Microbial glycoside hydrolases as antibiofilm agents with cross-kingdom activity. *Proc Natl Acad Sci U S A* **114**, 7124-7129
46. Fleming, D., Chahin, L., and Rumbaugh, K. (2017) Glycoside Hydrolases Degrade Polymicrobial Bacterial Biofilms in Wounds. *Antimicrob Agents Chemother* **61**
47. KOSHLAND Jr., D. E. (1953) STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS. *Biological Reviews* **28**, 416-436
48. McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol* **4**, 885-892
49. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853-859
50. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Gotz, A. W., Sandgren, M., Withers, S. G., Stahlberg, J., and Beckham, G. T. (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J Am Chem Soc* **136**, 321-329
51. Diez-Municio, M., Herrero, M., Olano, A., and Moreno, F. J. (2014) Synthesis of novel bioactive lactose-derived oligosaccharides by microbial glycoside hydrolases. *Microb Biotechnol* **7**, 315-331
52. Glasgow, E. M., Vander Meulen, K. A., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Deutsch, S., and Fox, B. G. (2019) Extent and Origins of Functional Diversity in a Subfamily of Glycoside Hydrolases. *J Mol Biol* **431**, 1217-1233
53. Chiriac, A. I., Cadena, E. M., Vidal, T., Torres, A. L., Diaz, P., and Pastor, F. I. (2010) Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* displaying a novel architecture. *Appl Microbiol Biotechnol* **86**, 1125-1134
54. Kundu, S. (2019) Insights into the mechanism(s) of digestion of crystalline cellulose by plant class C GH9 endoglucanases. *J Mol Model* **25**, 240
55. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat Struct Biol* **4**, 810-818
56. Khademi, S., Guarino, L. A., Watanabe, H., Tokuda, G., and Meyer, E. F. (2002) Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Crystallogr D Biol Crystallogr* **58**, 653-659
57. Schubot, F. D., Kataeva, I. A., Chang, J., Shah, A. K., Ljungdahl, L. G., Rose, J. P., and Wang, B. C. (2004) Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*. *Biochemistry* **43**, 1163-1170
58. Kumar, K., Singh, S., Sharma, K., and Goyal, A. (2021) Computational modeling and small-angle X-ray scattering based structure analysis and identifying ligand cleavage mechanism by processive endocellulase of family 9 glycoside hydrolase (HtGH9) from *Hungateiclostridium thermocellum* ATCC 27405. *J Mol Graph Model* **103**, 107808



59. Ravachol, J., Borne, R., Tardif, C., de Philip, P., and Fierobe, H. P. (2014) Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium *Clostridium cellulolyticum*. *J Biol Chem* **289**, 7335-7348
60. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* **180**, 1709-1714
61. Grondin, J. M., Duan, D., Kirilin, A. C., Abe, K. T., Chitayat, S., Spencer, H. L., Spencer, C., Campigotto, A., Houliston, S., Arrowsmith, C. H., Allingham, J. S., Boraston, A. B., and Smith, S. P. (2017) Diverse modes of galacto-specific carbohydrate recognition by a family 31 glycoside hydrolase from *Clostridium perfringens*. *PLoS One* **12**, e0171606
62. Walker, J. A., Takasuka, T. E., Deng, K., Bianchetti, C. M., Udell, H. S., Prom, B. M., Kim, H., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol Biofuels* **8**, 220
63. Pasari, N., Adlakha, N., Gupta, M., Bashir, Z., Rajacharya, G. H., Verma, G., Munde, M., Bhatnagar, R., and Yazdani, S. S. (2017) Impact of Module-X2 and Carbohydrate Binding Module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass. *Sci Rep* **7**, 3700
64. Meng, D. D., Ying, Y., Chen, X. H., Lu, M., Ning, K., Wang, L. S., and Li, F. L. (2015) Distinct roles for carbohydrate-binding modules of glycoside hydrolase 10 (GH10) and GH11 xylanases from *Caldicellulosiruptor* sp. strain F32 in thermostability and catalytic efficiency. *Appl Environ Microbiol* **81**, 2006-2014
65. Foumani, M., Vuong, T. V., MacCormick, B., and Master, E. R. (2015) Enhanced Polysaccharide Binding and Activity on Linear beta-Glucans through Addition of Carbohydrate-Binding Modules to Either Terminus of a Glucooligosaccharide Oxidase. *PLoS One* **10**, e0125398
66. Herve, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci U S A* **107**, 15293-15298
67. Hall, J., Black, G. W., Ferreira, L. M., Millward-Sadler, S. J., Ali, B. R., Hazlewood, G. P., and Gilbert, H. J. (1995) The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochem J* **309** ( Pt 3), 749-756
68. Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Basle, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J., and Gilbert, H. J. (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc Natl Acad Sci U S A* **109**, 20889-20894
69. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781
70. Zhang, Y. H., and Lynd, L. R. (2004) Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol Bioeng* **88**, 797-824
71. Michaelis, L., and Menten, M. L. (1913) Die kinetik der invertinwirkung. *Biochemische Zeitschrift* **49**, 333-369
72. Briggs, G. E., and Haldane, J. B. (1925) A Note on the Kinetics of Enzyme Action. *Biochem J* **19**, 338-339

73. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* **66**, 506-577, table of contents
74. Bansal, P., Hall, M., Realff, M. J., Lee, J. H., and Bommarius, A. S. (2009) Modeling cellulase kinetics on lignocellulosic substrates. *Biotechnol Adv* **27**, 833-848
75. Brunecky, R., Donohoe, B. S., Yarbrough, J. M., Mittal, A., Scott, B. R., Ding, H. S., Taylor, L. E., Russell, J. F., Chung, D. W., Westpheling, J., Teter, S. A., Himmel, M. E., and Bomble, Y. J. (2017) The Multi Domain Caldicellulosiruptor bescii CelA Cellulase Excels at the Hydrolysis of Crystalline Cellulose. *Sci Rep-Uk* **7**
76. Hall, M., Bansal, P., Lee, J. H., Realff, M. J., and Bommarius, A. S. (2010) Cellulose crystallinity--a key predictor of the enzymatic hydrolysis rate. *FEBS J* **277**, 1571-1582
77. Drissen, R. E. T., Maas, R. H. W., Van Der Maarel, M. J. E. C., Kabel, M. A., Schols, H. A., Tramper, J., and Beftink, H. H. (2007) A generic model for glucose production from various cellulose sources by a commercial cellulase complex. *Biocatal Biotransfor* **25**, 419-429
78. Desai, S. G., and Converse, A. O. (1997) Substrate reactivity as a function of the extent of reaction in the enzymatic hydrolysis of lignocellulose. *Biotechnol Bioeng* **56**, 650-655
79. Beckham, G. T., Stahlberg, J., Knott, B. C., Himmel, M. E., Crowley, M. F., Sandgren, M., Sorlie, M., and Payne, C. M. (2014) Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Curr Opin Biotech* **27**, 96-106
80. Wilson, D. B., and Kostylev, M. (2012) Cellulase processivity. *Methods Mol Biol* **908**, 93-99
81. Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside Hydrolase Processivity Is Directly Related to Oligosaccharide Binding Free Energy. *Journal of the American Chemical Society* **135**, 18831-18839
82. Wu, S., and Wu, S. (2020) Processivity and the Mechanisms of Processive Endoglucanases. *Appl Biochem Biotechnol* **190**, 448-463
83. Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Appl Environ Microbiol* **73**, 3165-3172
84. Kurasin, M., and Valjamae, P. (2011) Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* **286**, 169-177
85. Zverlov, V. V., Schantz, N., and Schwarz, W. H. (2005) A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-beta-1,4-glucanase producing cellotetraose. *Fems Microbiol Lett* **249**, 353-358
86. Deng, K., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases. *Front Bioeng Biotech* **3**
87. Mandelman, D., Belaich, A., Belaich, J. P., Aghajari, N., Driguez, H., and Haser, R. (2003) X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J Bacteriol* **185**, 4127-4135
88. Li, Y., Irwin, D. C., and Wilson, D. B. (2010) Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of *Thermobifida fusca* Cel9A. *Appl Environ Microbiol* **76**, 2582-2588

89. Jeng, W. Y., Liu, C. I., Lu, T. J., Lin, H. J., Wang, N. C., and Wang, A. H. (2019) Crystal Structures of the C-Terminally Truncated Endoglucanase Cel9Q from *Clostridium thermocellum* Complexed with Cellodextrins and Tris. *ChemBiochem* **20**, 295-307

## Chapter 2:

### **Contributions of carbohydrate binding modules and calcium to the activity of glycoside hydrolase family 9 cellulase CelR**

Nathaniel J. Kuch, Mark E. Kutschke, Alex Parker, Craig A. Bingman, Brian G. Fox

This work has been submitted to the Journal of Biological Chemistry

Nathaniel Kuch made the following contributions: experimental design and execution, data analysis, writing and revising of manuscript

#### **Abstract:**

Enzymatic deconstruction of lignocellulosic biomass is crucial for production of biofuels and bioproducts. Better understanding interactions between catalytic and binding domains, and other structural features offer potential avenues for improvement. Glycoside hydrolase family 9 (GH9) enzymes, which show both exo- and endo-cellulase activity, are attractive targets for improvement. This study examines a GH9 from *Acetovibrio thermocellus*, including variants of the catalytic domain with the native family 3c carbohydrate binding module (CBM3c), a non-native CBM3a, both CBMs, and neither. Binding to amorphous phosphoric acid swollen cellulose (PASC) was equally strong for all constructs, but catalysis was severely impaired in the absence of CBM3c. CBM3c increased thermostability by 25°C and catalytic efficiency on PASC and crystalline cellulose (Avicel) by 10×. CBM3a gave a modest increase in thermostability in the absence of CBM3c (+5°C), and improved binding by 5× on Avicel and catalytic efficiency on PASC and Avicel by 2-5×. EGTA treatment reduced activity of CBM3a constructs on Avicel, but otherwise had minimal effects. Comparison of structures for the catalytic domain bound to either cellohexaose or cellobiose (result of slow turnover of cellohexaose by active site variant E439Q in the crystal), and catalytic domain with CBM3c without substrate revealed multiple

residues near a calcium in the catalytic domain whose ligands and adjacent residues rearrange to alternatively promote substrate binding or product release. This work provides new insight into structural contributions of the extended active site cleft to the reactivity of CelR and perhaps other GH9 enzymes.

### **Introduction:**

Efficient and economical deconstruction of complex cellulosic biomass into simpler, easily fermentable sugars and aromatic compounds is an integral part of the developing biofuel and bioproduct technology landscape (1-6). Strategies for deconstruction often include combination of thermochemical treatment and enzymatic hydrolysis (1,5,7-9). Thermochemical treatments can involve high temperatures and harsh chemicals and produce inhibitory or non-fermentable byproducts (1,5,7,8). In contrast, enzymatic hydrolysis is often slower and complete hydrolysis often requires complex mixtures of enzymes secreted by fungi, with the potential inclusion of enzymes that are not suitable for hydrolyzing the various forms of pretreated biomass (3,9-12). One example is biomass pretreated by *g*-valerolactone, which effectively removes the hemicellulose and lignin fractions, and leaves a highly enriched cellulosic fraction consisting of both amorphous and crystalline domains of polysaccharide (8,13). The action of both endo- (cleaving within a cellulose chain) and exo- (acting at the end of the chain) glycoside hydrolases is typically required to hydrolyze insoluble polysaccharides. Glycoside hydrolase family 9 (GH9) includes members that have both activities (14,15), and so has attracted interest as targets for basic research and technological application.

Enzymes are a major operational expense in the deconstruction of cellulosic biomass, which has prompted efforts to find or engineer more efficient and economical options (12,16-

18). Developing a better understanding of glycoside hydrolase function and the impact of various non-catalytic domains, such as carbohydrate binding modules (CBMs), is needed to advance the goal of improved biomass deconstruction. Multiple studies have shown that incorporation (or removal) of non-catalytic domains such as CBMs can modulate thermostability, binding specificity, and catalytic efficiency (19-23).

Cel9A from *Thermobifida fusca* XY is one of the best characterized examples from glycoside hydrolase family 9 (GH9) (24-27). The complete Cel9A-90 (Tfu\_2176, UniProt P26221, Q47MW0) consists of an (a/a)<sub>6</sub> catalytic domain, a CBM family 3c (CBM3c) domain, a fibronectin domain, and a CBM family 2 (CBM2) domain. This enzyme has high activity on crystalline cellulose and, notably, exhibits both endo- and exocellulase activity (24-27). A truncated form, Cel9A-68, consisting of the catalytic and CBM3c domains, has also been studied and several crystal structures of this enzyme and its mutated variations have been solved (25-27). These studies, and others (28,29), have identified the catalytic residues and others along the extended active site cleft and on the surface of the CBM3c domain that interact with the oligosaccharide. Cel9A has also become important model system for computational evaluation of the mechanism of processivity (30-32), meaning that the enzyme carries out multiple cycles of catalysis by advancing along the cellulose chain (33,34).

CelR (Cthe\_0578, UniProt A3DCY5), from *Acetovibrio thermocellus* ATCC 27405<sup>1</sup> is a lesser-studied GH9 that consists of catalytic, CBM3c, and dockerin domains (35). Dockerin domains are responsible for recruitment of numerous proteins and enzymes into the cellulosome. Like other members of GH9, CelR has been proposed to have both exo- and endo- reactivity

---

<sup>1</sup> Over time, this organism has previously been named *Clostridium thermocellum* ATCC 27405, *Clostridium thermocellum* DSM 1237, *Hungateiclostridium thermocellum* ATCC 27405, and *Ruminiclostridium thermocellum* ATCC 27405.

(15), and to also have processive reactivity (35). Previous work showed that CelR and similar GH9 enzymes produce cellotetraose as an initial product from hydrolysis of cellulose (35) with cellotetraose hydrolyzed to either cellotriose and glucose or two cellobiose, and cellotriose hydrolysed to cellobiose and glucose (35,36). Cellobiose is not hydrolyzed. The CBM3c is required for processive activity (27,37-40). Along with their established catalytic capabilities, Cel9A and CelR are also of interest because of their thermostability and reactivity at high temperature (~50-70°C) with crystalline cellulose, which is considered difficult to enzymatically digest (37). Recently, we have focused on the reaction of CelR with GVL-treated plant biomass (8,13), and identified the I<sub>b</sub> crystal form of cellulose as a bottleneck to hydrolysis by CelR and also commercial enzyme preparations (41).

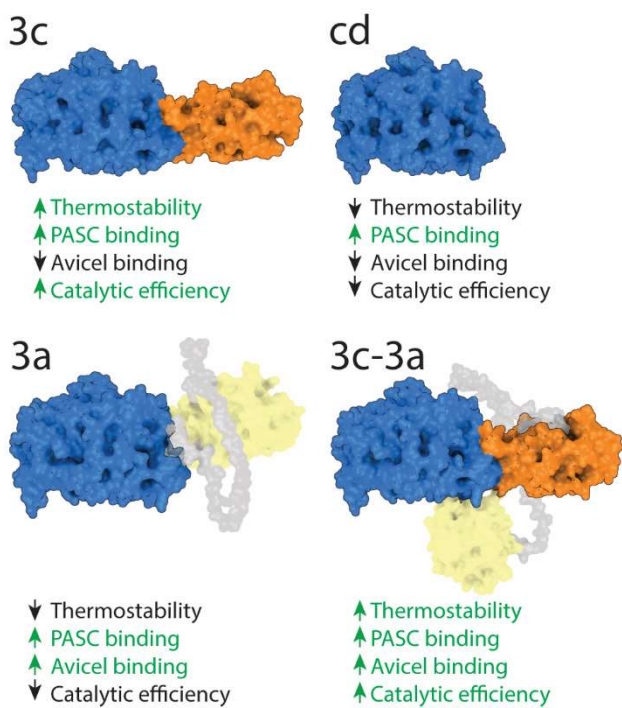
In this work, the thermal melting, binding, and kinetic parameters of four different assemblies of CelR with its native CBM3c and a non-native CBM3a are presented along with crystal structures of mutated enzyme bound to either substrate- (cellohexaose) or product- (cellobiose, arising from slow in-crystal hydrolysis of cellohexaose by active site variant E439Q). In biology, Ca<sup>2+</sup> is frequently associated with cellular signaling or energy transduction, however, it is also well suited to support protein stability and catalysis (42,43), so it is found in many enzymes including GHs and CBMs (27,44). The CelR constructs studied here contain three distinct Ca<sup>2+</sup> binding sites. Biochemical results suggest that Ca<sup>2+</sup> present in engineered CBM3a is more labile while Ca<sup>2+</sup> bound near the active site cleft and in CBM3c are tightly bound. Variations in the ligands of Ca<sup>2+</sup> and adjacent residues in the catalytic domain influence both substrate binding at the -4 subsite of the active site cleft and product release at the -3 subsite. In the presence of cellohexaose, residues D278 and D279 provide hydrogen bonding interactions with the glucosyl group in the -4 subsite, while after hydrolysis of cellohexaose in

the crystal, H226, another closely associated residue, moves away from the –3 subsite of the active site cleft to possibly facilitate release of the preferred product cellotetraose.

Concomitantly, cellobiose is retained in the clamp-like region of the active site cleft, supporting the concept that differential affinity for glucosyl units along the active site cleft may contribute to processivity in GH9 (45,46). Although the inclusion of both CBMs provides modest improvements, the presently engineered addition of CBM3a did not increase specific activity commensurate with the increase in molecular mass of the engineered catalyst. Moreover, based on the ratio of  $k_{cat}/k_{off}$ , CelR is a modestly processive enzyme when compared to structurally distinct, highly processive enzymes such as fungal GH7 (47,48). These findings inform potential use of CelR and similar enzymes in biomass deconstruction.

## Results:

### *CelR constructs and summary of results*





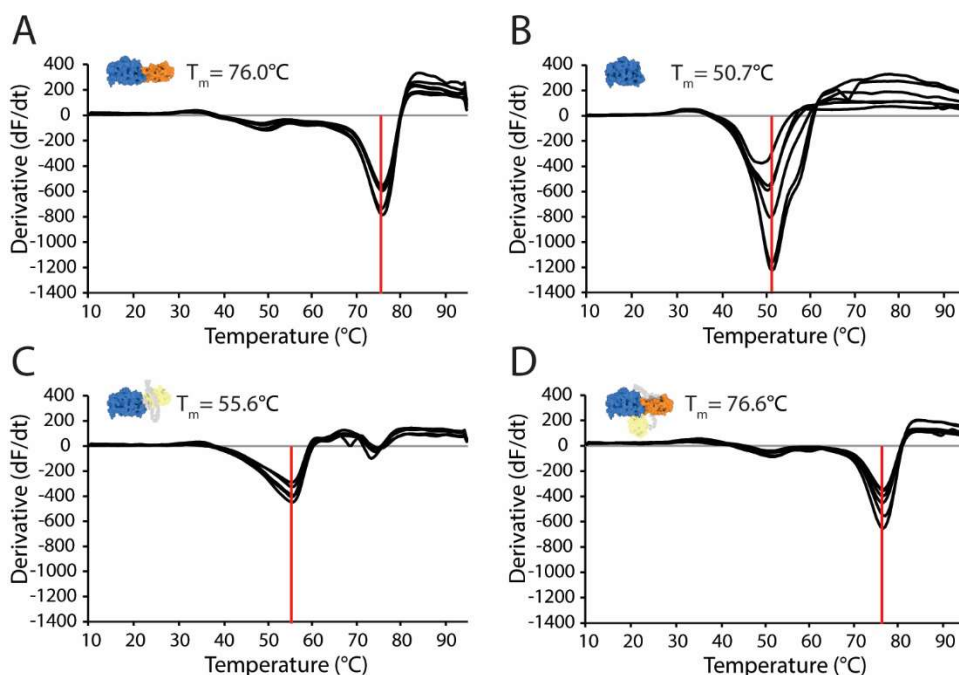
**Figure 2.1:** Overview of the X-ray (cd and 3c) or AlphaFold predicted (3a and 3c-3a) structures of the CelR constructs used in this work alongside a qualitative summary of changes in experimentally determined parameters.

In this work, the CelR catalytic domain was engineered to contain different combinations of CBMs (Figure 2.1): the catalytic domain and native CBM3c (3c, *blue* and *orange* surfaces); the catalytic domain alone (cd, *blue* surface); addition of a non-native flexible linker (*gray* transparent surface) and CBM3a from *A. thermocellus* ATCC 27405 to the catalytic domain (3a, *yellow* transparent surface); and addition of CBM3a to 3c (3c-3a). The linker used to attach CBM3a comes from the CipA gene of *A. thermocellus* ATCC 27405, and thus mimics a natural connection between cellulase domains in the *A. thermocellus* cellulosome. The 3c-3a construct also loosely mimics the native form of Cel9A from *T. fusca*, which includes the GH9 catalytic domain, closely associated CBM3c, and a CBM2 attached by a long, flexible linker (25). CBM2 is known to bind to crystalline cellulose (49), although the linker connecting CBM2 is apparently susceptible to proteolysis (27,29,30). Consequently, most Cel9A research has been on the truncated form containing only the catalytic domain and CBM3c.

Figure 2.1 provides a relative assessment of thermostabilities, capacity to bind to either amorphous or crystalline cellulose, and catalytic efficiencies ( $k_{\text{cat}}/K_M$ ). The structures of cd and 3c were determined by crystallography while AlphaFold was used to generate iconic representations of 3a and 3c-3a. Representations containing CBM3a are for illustrative purposes only and do not imply a preferred position of the CBM3a domain or specific interactions with either cd or CBM3c; indeed, results indicate these do not likely exist. As summarized in Figure 2.1, the catalytic domain alone (cd) is capable of binding to amorphous cellulose but has low catalytic efficiency, the presence of CBM3c (3c) promotes thermostability and increased

catalytic efficiency with amorphous cellulose, while the addition of CBM3a to either cd or 3c promotes binding to crystalline cellulose but does not uniquely contribute to catalytic efficiency.

### Thermostability of *CelR* constructs



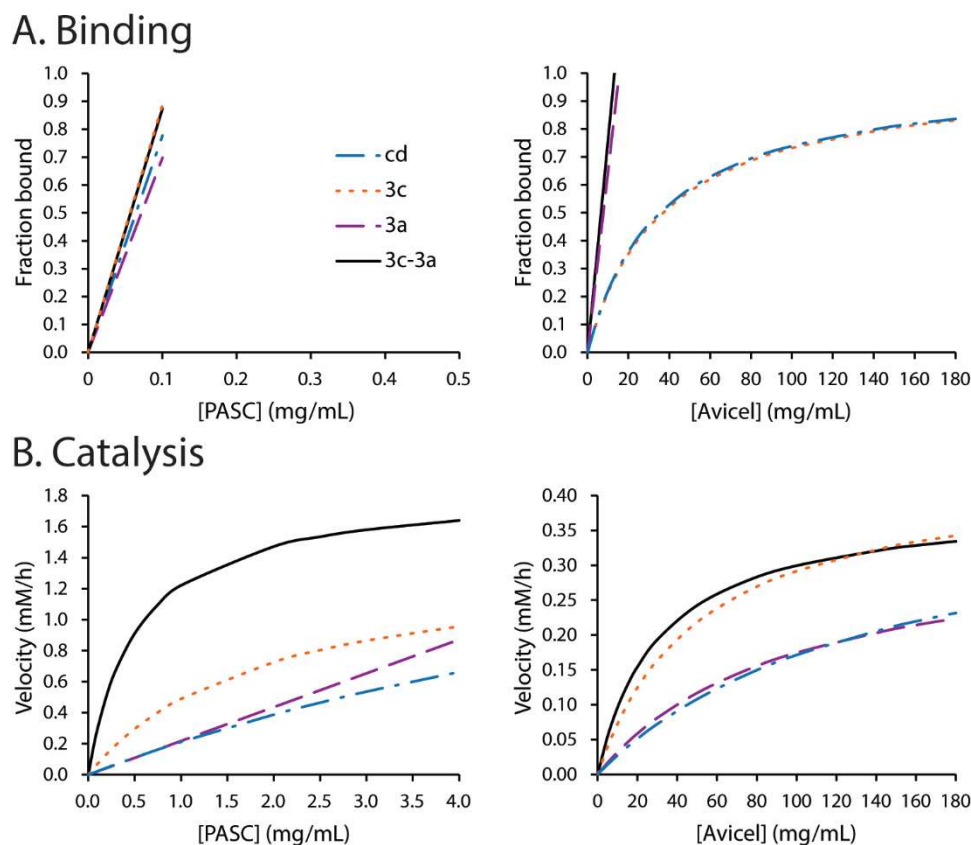
**Figure 2.2:** First derivative (*black, solid*) data from differential scanning fluorimetry (DSF) protein melt experiments for: 3c (A); cd (B); 3a (C); and 3c-3a (D).  $T_m$  (*red, vertical*) for each construct was determined from the minimum of the first derivative curves:  $76.0 \pm 0.0^\circ\text{C}$  for 3c;  $50.7 \pm 0.9^\circ\text{C}$  for cd;  $55.6 \pm 0.2^\circ\text{C}$  for 3a; and  $76.6 \pm 0.2^\circ\text{C}$  for 3c-3a. Measurements were made in triplicate from two experimental replicates.

Differential scanning fluorimetry (DSF) was used to track the thermal denaturation of the *CelR* constructs. Figure 2.2 shows that all constructs were denatured in the DSF experiments, with the first derivative of the fluorescence (*solid black lines*) decreasing to a minimum at the melting temperature of the construct, before increasing as the SYPRO dye dissociated from the melted polypeptide. The  $T_m$  is shown as a vertical *red* line. The four constructs exhibit different behaviors. The most thermostable constructs were 3c-3a ( $T_m = 76.6 \pm 0.2^\circ\text{C}$ ) and 3c ( $T_m = 76.0 \pm 0.0^\circ\text{C}$ ), while the least thermostable construct was cd ( $T_m = 50.7 \pm 0.9^\circ\text{C}$ ). Addition of CBM3a

provided a small improvement in thermostability relative to cd alone ( $T_m = 55.6 \pm 0.2^\circ\text{C}$ ) and showed a second melting feature at  $\sim 75^\circ\text{C}$ , likely corresponding to the melting of CBM3a. Other small, secondary features in the derivative curves are not indicative of major unfolding events.

*Binding affinity of CelR constructs at optimal pH and temperature for hydrolysis*

A further assessment of how different CBMs impact CelR was made through comparing the binding affinity of the four constructs containing the E439Q inactivating mutation with phosphoric acid swollen cellulose (PASC) and Avicel, which are either amorphous or crystalline forms of cellulose, respectively. For hydrolysis reactions with PASC, the optimal pH and temperature were determined to be pH 7.0 and  $40^\circ\text{C}$  for cd, pH 6.0 and  $65^\circ\text{C}$  for 3c, pH 6.5 and  $45^\circ\text{C}$  for 3a, and pH 6.0 and  $75^\circ\text{C}$  for 3c-3a (Figure S2.1), while for reactions with Avicel the optima were determined to be pH 7.0 and  $45^\circ\text{C}$  for cd, pH 7.0 and  $65^\circ\text{C}$  for 3c, pH 6.5 and  $45^\circ\text{C}$  for 3a, and pH 5.5 and  $65^\circ\text{C}$  for 3c-3a (Figure S2.2). Thus, addition of CBMs gave a moderate decrease in optimal pH (7.0 for cd versus 5.5-6.0 for 3c-3a) and an increase in the optimal temperature for catalysis in a manner similar to increases in melting temperature observed as CBMs were added.



**Figure 2.3:** Lines of best fit for (A) binding and (B) hydrolysis for PASC and Avicel. Individual experimental results from triplicate experiments overlaid with the line of best fit can be found in Figures S2.3 and S2.4. Kinetic and binding behaviors for the constructs indicate that the different CBMs have distinct impacts on both binding and catalysis. Additional CBMs do not change the affinity for PASC, while CBM3a provides the strongest binding for Avicel. CBM3c provides maximum velocity for hydrolysis of Avicel, while CBM3c and CBM3a appear to interact synergistically for hydrolysis of PASC.

Figure 2.3 shows the lines of best fit from kinetics and binding experiments to highlight the different patterns between binding and hydrolysis of PASC and Avicel. Graphs containing the results from triplicate experiments as well as the line of best fit for each combination of construct and substrate can be found in Figures S2.3 and S2.4. Neither cd nor 3a show saturation kinetics behavior on PASC, and while 3c does show saturating behavior, 3c-3a has a much higher maximum velocity. This is contrasted by the binding data, which indicates that neither CBM significantly changes the affinity of cd for PASC, and binding experiments produce curves

that are linear until saturation, indicating tight binding. On PASC, 3c-3a has only a marginal improvement in  $K_M$  compared to 3c, indicating that CBM3a provides little benefit to catalysis of crystalline cellulose beyond the benefits provided by CBM3c. On Avicel, all constructs show saturating kinetics behavior, and 3c and 3c-3a have both higher maximum velocities and lower  $K_M$ -values than cd and 3a. The binding data for Avicel shows that CBM3a confers tight binding, while CBM3c only provides only marginal benefits for this crystalline substrate. Taken together, Figure 2.3 indicates that CBM3c provides the most impact to catalytic capabilities, while CBM3a improves binding to crystalline cellulose, and the combination of both CBMs provides synergistic improvement to the hydrolysis of amorphous cellulose.

**Table 2.1:** Kinetic and binding parameters for hydrolysis of either PASC or Avicel

Substrate	Enzyme	$K_D$ (mg/mL)	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (mg/mL)	$k_{cat}/K_M$	Specific activity ( $U/mg$ ) <sup>c</sup>
PASC	cd	$0.07 \pm 0.02^b$	$\geq 0.7$	$\geq 4$	nd <sup>a</sup>	$0.1 \pm 0.01$
	3c	$0.06 \pm 0.01^b$	$3.9 \pm 0.2$	$1.9 \pm 0.1$	$2.1 \pm 0.1$	$2.9 \pm 0.1$
	3a	$0.07 \pm 0.00^b$	$\geq 0.9$	$\geq 4$	nd <sup>a</sup>	$0.1 \pm 0.003$
	3c-3a	$0.06 \pm 0.01^b$	$5.1 \pm 0.2$	$0.5 \pm 0.03$	$10 \pm 0.3$	$2.5 \pm 0.2$
Avicel	cd	nd <sup>a</sup>	$0.11 \pm 0.02$	$142 \pm 47$	$0.001 \pm 0.0001$	$0.003 \pm 0.0003$
	3c	$37 \pm 5.8$	$1.2 \pm 0.1$	$50 \pm 7.0$	$0.02 \pm 0.001$	$0.14 \pm 0.002$
	3a	$7.8 \pm 0.1^b$	$0.10 \pm 0.001$	$98 \pm 2.2$	$0.001 \pm 0.00002$	$0.004 \pm 0.001$
	3c-3a	$6.7 \pm 0.7^b$	$1.1 \pm 0.1$	$31 \pm 4.5$	$0.04 \pm 0.002$	$0.11 \pm 0.002$

a: nd, not determined; saturation kinetics behavior not observed

b: Linear response observed until until saturation, value calculated from a linear fit of the experimental data

c:  $U$ , unit, is defined as production of 1  $\mu$ mol of reducing sugar per minute. The specific activity calculation includes the mass of the different constructs: cd, 53.3 kDa; 3c, 70.7 kDa; 3a, 73.4 kDa; 3c-3a, 93.0 kDa.

Table 2.1 contains the kinetic and binding parameters determined from the lines of best fit shown in Figure 2.3. All four constructs share similar, high affinity for PASC, with  $K_D$  in the

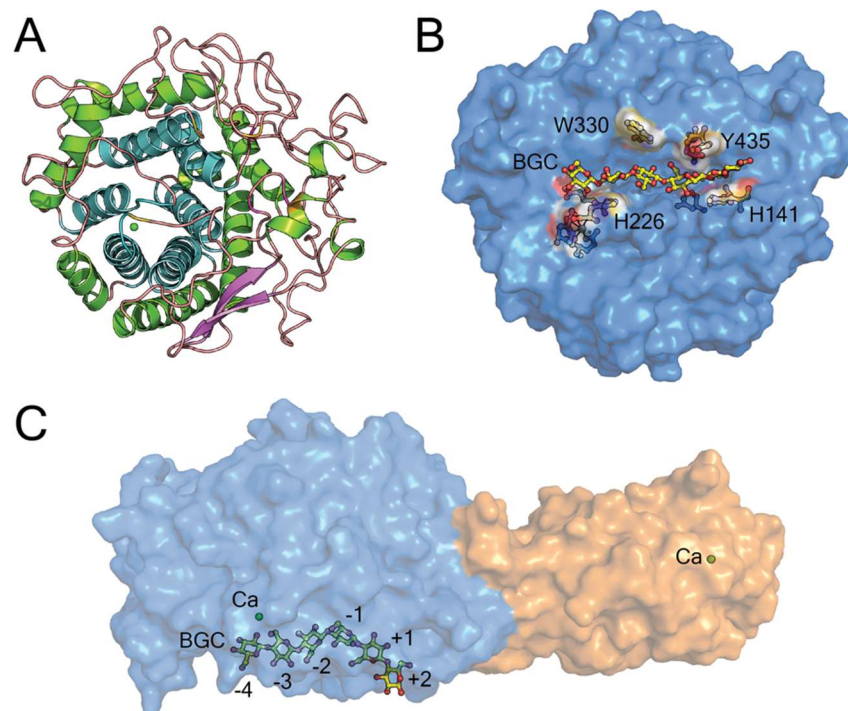
range of 0.06 mg/mL. CBM3c appears to be required for effective hydrolysis of PASC, as both cd and 3a fail to reach saturation in the substrate range tested. Comparing construct 3c-3a to 3c, the addition of CBM3a results in a small improvement in  $k_{\text{cat}}$ , and a 4× decrease in  $K_M$ , resulting in an overall improvement in  $k_{\text{cat}}/K_M$  of approximately 5×. On Avicel, CBM3c is again required for effective hydrolysis, corresponding to an approximately 10× increase in  $k_{\text{cat}}$  (comparing cd and 3a to 3c and 3c-3a respectively). As with PASC, 3c-3a has a higher catalytic efficiency than 3c, although the improvement is due to the lower  $K_M$ , resulting in an approximate doubling of  $k_{\text{cat}}/K_M$ . Binding to Avicel is improved by the addition of both CBM3c and CBM3a, although the most significant improvement comes from CBM3a, with a 6× improvement in  $K_D$  from 3c to 3c-3a and an approximately 6× lower  $K_D$  comparing 3a to 3c.

Specific activity follows a broadly similar trend as catalytic efficiency on both PASC and Avicel, with 3c and 3c-3a showing an order of magnitude higher specific activity than either cd or 3a. However, 3c has higher specific activity compared to 3c-3a for both substrates, indicating that 3c is more active per mass of enzyme ( $U/mg$ ). While 3c-3a is more active per enzyme molecule ( $k_{\text{cat}}/K_M$ ) for reaction with PASC, tighter binding given by CBM3a did not improve catalytic efficiency with Avicel.

### *Structure of CelR*

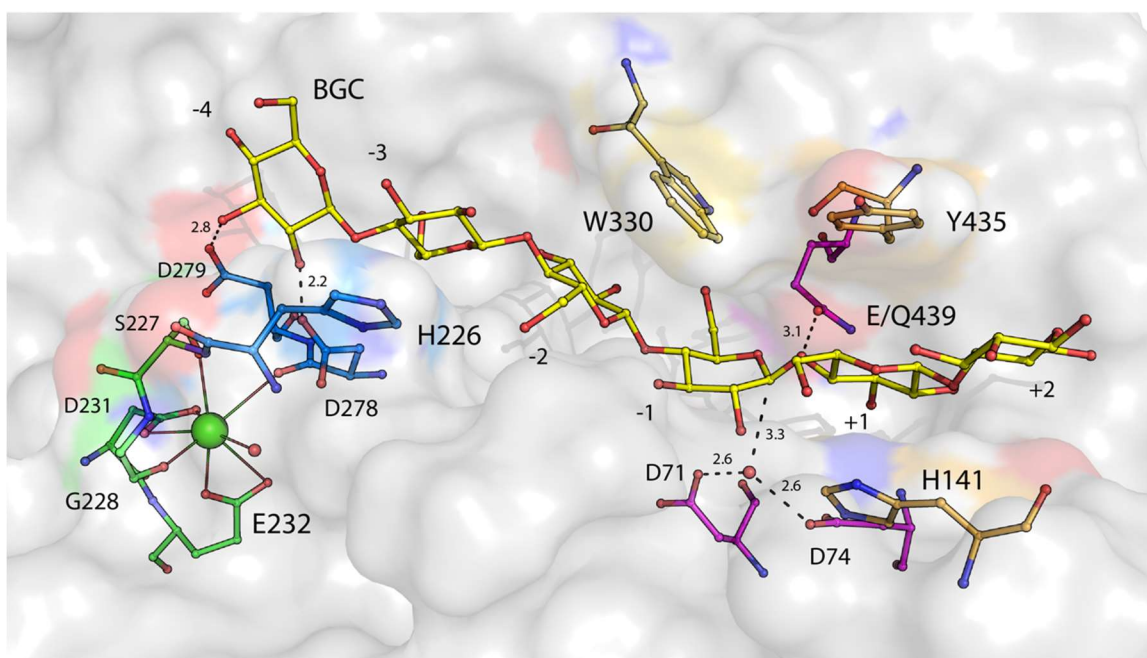
Table S2.1 provides the data collection and refinement statistics for three structures: cd E439Q bound to a substrate (PDB 7v0i, i.e., BGC, cellohexaose); a product (PDB 7v0j, i.e., CBI, cellobiose, resulting from slow hydrolysis of cellohexaose by the mutated enzyme during crystallization); and 3c with no substrate bound (PDB 7unp). All three structures have similar completeness,  $R$ -values, and resolution. Figure 2.4A and 2.4B show cartoon and surface

representations of the catalytic domain, respectively. Figure 2.4C shows a transparent side view of the intact catalytic and CBM3c domains (Figure 2.4C), the positions of  $\text{Ca}^{2+}$  atoms in these domains and a placement of cellobiose (BGC, *yellow sticks*) from alignment with 7v0i. The catalytic domain has an  $(\alpha/\alpha)_6$ -barrel fold (Figure 2.4A), with the *cyan* helices forming the base of the active site cleft and providing the ligands to  $\text{Ca}^{2+}$ . Loops (*salmon*) provide the surface of the active site. Figure 2.4B shows that the active site lies along a cleft in the catalytic domain along with the positions of the catalytic residues D71, D74 and E439 (mutated to Q439 in 7v0i and 7v0j). Figure 2.4C shows that the catalytic domain and CBM3c provide a continuous surface for interaction with cellulose, including possible binding interactions that extend beyond the active site cleft by seven or so glucosyl units (30,35). The CBM3c has a  $\beta$ -sandwich fold and others have described how the CBM3c might assist in guiding an oligosaccharide into the active site cleft, which extends from the -4 to +2 positions for the bound glucosyl units (25-27).



**Figure 2.4:** (A) Cartoon representation of 7v0j with the outer helices of the  $(\alpha/\alpha)_6$ -barrel fold colored green and inner helices colored cyan. (B) Transparent surface representation of 7v0i showing bound cellohexaose (BGC, *yellow sticks*) and important residues along the active site cleft. (C) 90° rotation of the catalytic domain relative to (B) and transparent view of 7unp with cellohexaose superimposed from alignment with 7v0i. Sugar binding sites in the active site cleft are labeled from -4 to +2 (non-reducing to reducing end).  $\text{Ca}^{2+}$  near the active site and in the CBM3c domain are shown as *green spheres*.

### Overview of the active site cleft



**Figure 2.5:** View of the CelR active site cleft in 7v0i encompassing -4 to +2 sugar sites with clamp residues (H141, Y435, *orange sticks*), mobile residues (H226, *marine sticks*; W330, *gold sticks*), catalytic residues (D71, D74, E/Q439, *magenta sticks*),  $\text{Ca}^{2+}$  (*green sphere*),  $\text{Ca}^{2+}$ -coordinating residues (S227, G228, D231, E232), cellohexaose (BGC *yellow sticks*), and residues hydrogen bonding to the glucosyl unit in the -4 position (D278, D279, *marine sticks*) from 7v0i. A water hydrogen bonded by D71 and D74 is well positioned to serve as the nucleophile for the hydrolysis reaction.

Figure 2.5 shows a closer view of the active site cleft and three regions contributing to catalysis. These include a ‘clamp’ provided by H141 and Y435 (*orange sticks*) that contacts the glucosyl unit bound in the +1 position, and is immediately adjacent to the catalytic residues D71, D74 and E439 (*magenta sticks*, Q439 in 7v0i). A water molecule hydrogen bonded by D71 and D74 lies  $\sim 3$  Å from the C1 of the glucosyl unit in the -1 position, which is site of hydrolysis.

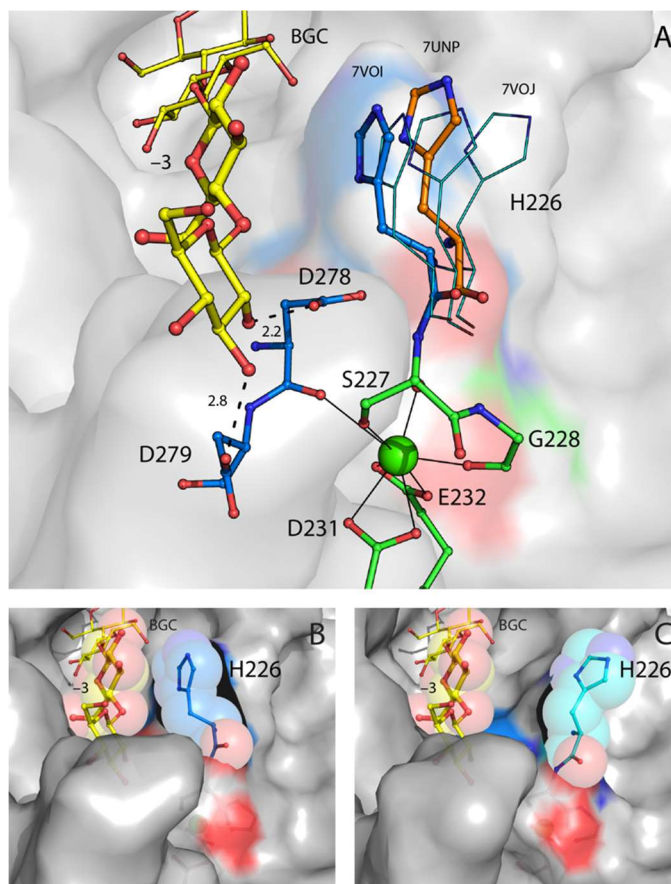


Among the three solved structures of CelR, the position of Q439 is variable depending on occupancy of the active site cleft, but regardless the closest sidechain atom lies within  $\sim 3$  Å of O4 of the glucosyl unit in the +1 position so is well positioned to serve as the general acid for the hydrolysis reaction. The identification of the catalytic residues (27,29), active site nucleophile (27), and the role of clamp residues (H141 and Y435) in positioning the oligosaccharide for hydrolysis between the +1 and -1 subsites (27,38) have been well described by others.

Residues W330 and Y435, both conserved in the GH9 family, lie along the active site cleft and adopt different conformations whether substrate is present or not. These rearrangements are proposed to facilitate productive substrate binding and have been implicated in processivity (27,46). Although the movement of CelR Y435 in response to substrate binding is modest, CelR W330 moves  $\sim 6$  Å or more in response to substrate binding and makes  $\pi$ -stacking interactions with a glucosyl unit bound in the -2 position as seen in other GH9 members (27,46).

#### *Interactions of H226 at the -3 subsite*

In CelR, H226 (Figure 2.5, *gold sticks*) also lies along the active site cleft. In other GH9 showing processivity (27,50), a Trp residue is most often observed at this position, although in Cel9G no residue is positioned similarly to H226 (38). In Cel9A, W209 does not change position dependent on whether substrate is present or not, and its pyrrole ring superimposes with the position of the imidazole ring of CelR H226 in the cellohexaose-bound form (Figure 2.5).

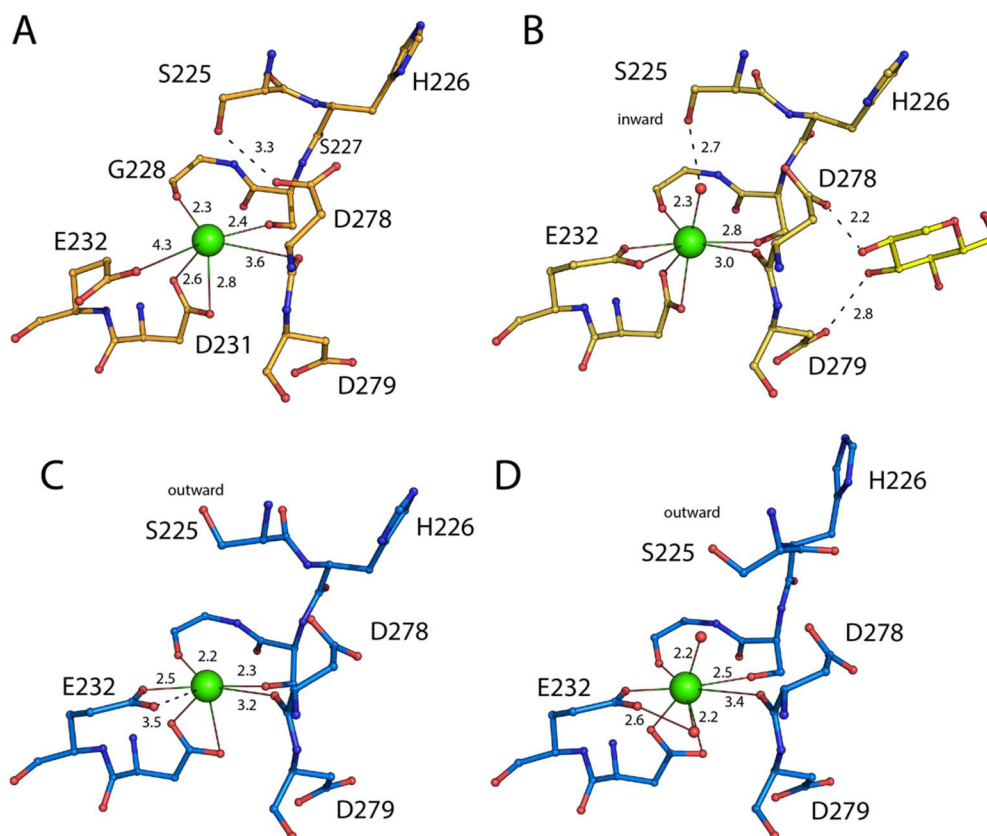


**Figure 2.6:** A, Interaction of substrate (BGC, *yellow* sticks and lines) with H226 (*gold* sticks at the  $-3$  subsite), D278, and D279 (*marine* sticks at the  $-4$  subsite) in the C chain from 7v0i. The position of H226 in the other two copies of the protein in the asymmetric unit of 7v0i are indistinguishable: not shown for clarity. Other ligands to  $\text{Ca}^{2+}$  are shown as *green* sticks. The position of H226 in the unbound structure (7unp, *orange* sticks), and in the three copies of the protein in product bound structure 7v0j (*cyan* lines), the latter of which are variable.

Figure 2.6A shows the region near to H226 from the unbound (*orange* sticks, 7unp), substrate bound (BGC in *yellow* sticks and lines extending down the active site cleft toward the site of hydrolysis; 7v0i), and product bound structures (CBI in the clamp region; 7v0j, not visible in this image). In the substrate bound structure 7v0i, H226 (*marine*, also Figure 2.6B) is within Van der Waals contact of the glucosyl unit in the  $-3$  site (*yellow* lines), and so helps place the substrate into the active site as reported for Trp in other GH9 at this position (27,50). In CelR, D278 and D279 hydrogen bond to the O2 and O3 hydroxyl groups of the glucosyl unit in the  $-4$

site ( $\sim 2.2$  and  $\sim 2.8$  Å, respectively), providing additional stabilization of bound oligosaccharide. These interactions are also seen in Cel9A bound to cellotriose or cellotetraose (3tf4, 4tf4), albeit with longer bond lengths than seen in CelR.

In the product bound structure 7v0j, H226 (*cyan* lines in Figure 2.6A) is observed in different positions in the three polypeptide chains found in the asymmetric unit. The outermost of these positions releases the steric constraint on the glucosyl unit at the  $-3$  site of the active site cleft (Figure 2.6C), and so may facilitate release of cellotetraose even as cellobiose (or a longer oligosaccharide interacting with the CBM3c) is retained in the distal portion of the active site cleft. Indeed, the 7v0j structure arose from slow hydrolysis of cellohexaose soaked into crystals of the E439Q mutated enzyme, which evidently retains low activity, but nevertheless retains cellobiose after release of cellotetraose. In the unbound structure 7unp, which has only one polypeptide in the asymmetric unit, H226 occupies an intermediate position (*orange* sticks, Figure 2.6A).



**Figure 2.7:** Differences in ligation of  $\text{Ca}^{2+}$  and positions of adjacent residues in structures of CelR. Key residues and bond lengths are identified. A, unbound structure from 7unp. E232 provides a monodentate, long weak interaction with  $\text{Ca}^{2+}$  and D278 is rotated toward S225 OG. B, substrate bound structure from chain C of 7v0i; chain B is indistinguishable. E232 has rotated to providing bidentate coordination to  $\text{Ca}^{2+}$  along with D231; S225 OG has an ‘inward’ conformation and makes a hydrogen bonding interaction with water bound to  $\text{Ca}^{2+}$  while D278 has rotated toward the glucosyl unit in the  $-4$  subsite and makes a hydrogen bond. C, product bound structure from chain C of 7v0j. E232 has a bidentate coordination with  $\text{Ca}^{2+}$  and S227 OG has an ‘outward’ conformation away from  $\text{Ca}^{2+}$ . D, product bound structure from chain B of 7v0j. Two waters are bound to  $\text{Ca}^{2+}$  and S227 OG has an outward projection away from  $\text{Ca}^{2+}$ . Chains A and B from 7v0i and chains A and C from 7v0j have similar positions of protein ligands and an additional water between E232 and  $\text{Ca}^{2+}$ .

### *Ligation of $\text{Ca}^{2+}$ is influenced by substrate/product binding*

Calcium is a versatile metal ion. It accepts variation in number of ligands, coordination geometries, and bond lengths (43,51). It can also exhibit rapid exchange of ligand positions, a property called fluxional binding (52,53). There are three  $\text{Ca}^{2+}$  binding environments in the CelR constructs examined here: 1) at the end of the catalytic cleft near the  $-4$  subsite (Figure 2.7); 2)

in CBM3c (38,46,50); and 3) in the engineered CBM3a (44). While the  $\text{Ca}^{2+}$  sites in CBM3c and CMB3a have been previously described (38,44,46,50), less attention has been given to  $\text{Ca}^{2+}$  in the GH9 catalytic domain. In the CelR structures,  $\text{Ca}^{2+}$  was modeled in full occupancy in all subunits, and addition of  $\text{Ca}^{2+}$  to catalytic assays gave no enhancement of activity. Treatment of enzyme with EGTA impacted catalysis only when CBM3a was present (Figure S2.5), likely owing to the removal of the surface-bound  $\text{Ca}^{2+}$  shown by others to be required for stability and/or cellulose binding (38,44,50,54-56). The catalytic properties of cd and 3c were not significantly impacted, indicating the well buried positions of  $\text{Ca}^{2+}$  in these enzymes were not accessible to EGTA. Figure 2.7 show the breadth of ligation environments present in the CelR cd, which differ when the active site cleft is empty, or when product (cellobiose, CBI) or substrate (cellohexaose, BGC) are bound.

In the unbound structure (7unp, Figure 2.7A), E232 provides a weak monodentate interaction ( $\sim 4.3$  Å) with  $\text{Ca}^{2+}$  (Figure 2.7A), but in the substrate bound form 7v0i (Figure 2.7B), E232 undergoes a dramatic rearrangement as it changes rotamer orientation to assume a full bidentate coordination of  $\text{Ca}^{2+}$  with bond lengths of 2.5 Å and 3.2 Å.

In all structures, D231 provides a symmetric bidentate coordination to  $\text{Ca}^{2+}$  with bond lengths ranging from 2.4 Å to 2.9 Å, typical of oxygen bound to  $\text{Ca}^{2+}$ . G228 O, S227 OG, and D278 O also provide bonds to  $\text{Ca}^{2+}$ ; the bonds from these residues change their length and geometry depending on the disposition of E232 and whether additional waters are bound to  $\text{Ca}^{2+}$ . For example, while the configuration of E232 in chain B of product bound 7v0j (Figure 2.7C) is similar to that shown for chain C of the substrate bound 7v0i (Figure 2.7B), the electron density of chains A and C from 7v0j and chains A and B from 7v0i are better modeled by including a water bound to  $\text{Ca}^{2+}$  and hydrogen bonded by E232 (e.g., Figure 2.7D for chain B from 7v0j).

Rearrangements of the bonding interactions of S227 OG influence the positioning of adjacent H226 (Figure 2.6) and S225 (Figure 2.7). When substrate is bound (Figure 2.7B), S225 OG assumes an ‘inward’ configuration and makes an  $\sim 2.5$  Å hydrogen bond with water bound to  $\text{Ca}^{2+}$ . This apparently stabilizes the interaction of H226 with the glucosyl unit at the  $-3$  subsite. When product is bound in the distal clamp site, S225 OG is observed in both inward (like that shown in Figure 2.7B) and ‘outward’ configurations among the three polypeptides in the asymmetric unit, with corresponding breadth of configurations observed for H226 (Figure 2.6). Changes in the bond length of D278 O to  $\text{Ca}^{2+}$  are also observed between the unbound and substrate/product bound structures, and these changes influence the orientation and distance of the carboxylate groups of D278 and D279 to the  $-4$  subsite. Thus, as the bond length of D278 O to  $\text{Ca}^{2+}$  decreases from 3.6 Å to 3.0 Å (Figure 2.7A versus 2.7B), the carboxyl group of D278 rotates  $\sim 90^\circ$  and moves along with the carboxyl group of D279  $\sim 1.5$  Å to provide hydrogen bonds to the glucosyl unit in the  $-4$  subsite (2.2-2.4 Å for D278 and 2.8-3.0 Å for D279). These positions of the carboxylate group are retained with only minor distances in the product bound structure even as there is no glucosyl unit occupying the  $-4$  subsite and S225 and H226 show multiple conformations.

## **Discussion:**

### *Thermostability*

This study shows that CelR’s thermostability is substantially modulated by the presence or absence of CBM3c and CBM3a. Thermostability is most impacted by CBM3c, which gives a 20-25°C increase in  $T_m$  to the catalytic domain. The thermostability benefits are likely due to the extensive protein interface between these two domains, agreeing with previous findings on

CBM3c in other enzymes (57-59) and the stabilizing contribution of tightly bound  $\text{Ca}^{2+}$ . In contrast, CBM3a has only a small impact on thermostability when linked to either cd or 3c as done here, and the different domains appear to melt independently at widely separated temperatures ( $\sim 50^\circ\text{C}$  versus  $\sim 70^\circ\text{C}$ ). The linker used in this study was intended to provide flexible access to substrate instead of stabilizing protein-protein interactions with cd such as provided in 3c.

### *Binding*

The structure of CelR (Figure 2.4) shows cd has a recessed cleft that can only be accessed by a single polysaccharide chain, such as found in amorphous cellulose. Correspondingly, binding to amorphous cellulose was not improved by either CBM3c or CBM3a (Figure 2.3A and Table 2.1,  $K_D$  values for PASC). The enclosed active site cleft provides several interactions with oligosaccharides (shown here for cellohexaose and cellobiose) characteristic of processive GH9; particularly steric interactions at the +2 and +1 binding sites, the His/Tyr clamp at the +1 site, three hydrogen bonds coordinating the +2 sugar (15,25-27,29,30,38,46), and in addition the interactions with the -3 and -4 subsites modulated by  $\text{Ca}^{2+}$  ligation shown here (Figures 2.6 and 2.7). The recessed cleft and the loop at the end of the active site cleft (Figure 2.6B and 2.6C) presumably hinders access of the planar surface of crystalline cellulose, Avicel, to the active site, but would still allow interactions at less ordered and ragged ends of crystalline fibers.

The binding interactions of family 3 CBMs arise from a small number of hydrogen bonds and a track of aromatic residues that provide weak, hydrophobic stacking interactions with the bound cellulose chain (27,60-62). In GH9, CBM3c has a flat binding surface that extends towards the active site cleft. For CelR, CBM3c provides no improvement in binding to

amorphous cellulose when compared to cd alone, and only a weak binding interaction with crystalline cellulose.

CBM3a provides an  $\sim 5\times$  improvement in binding to Avicel when added to either cd or 3c (Figure 2.3A and Table 2.1). This is in accord with earlier studies that show CBM3a binds tightly to crystalline cellulose (23,44), as the planar arrangement of binding interactions are not limited to a single strand of cellulose; rather, the binding interaction includes neighboring cellulose strands in the crystalline plane in addition to the strand being actively hydrolyzed (23,44,60).

#### *Catalytic properties of CelR constructs*

As with binding, hydrolysis is modulated by the inclusion of either CBM3c or CBM3a. For Avicel, CelR constructs containing CBM3c (either 3c or 3c-3a) possessed both  $\sim 10\times$  higher  $k_{\text{cat}}$  and 3 times lower  $K_M$  values when compared to the constructs lacking CBM3c (either cd or 3a). In comparison, while CBM3a provided an  $\sim 5\times$  improvement to binding ( $K_D$ ), it only provided  $<2\times$  improvement to  $K_M$  values when added to either cd or 3c. This indicates that the binding affinity provided by CBM3a (and the presumed flexibility of the linker) do not necessarily lead to catalytically productive binding to Avicel.

The pattern of CBM3c improving both  $k_{\text{cat}}$  and  $K_M$  is also apparent for hydrolysis of PASC, although the magnitude of improvement is unclear as the velocity of neither cd nor 3a saturated in the tested PASC concentrations. Because of this, it is not possible to conclude whether the addition of CBM3a to cd provides an improvement in  $K_M$  as seen on Avicel, but comparison of results for 3c and 3c-3a revealed only modest  $<2\times$  improvement in  $k_{\text{cat}}$  and  $4\times$  improvement in  $K_M$ . The kinetic improvements gained from the addition of CBM3a most likely



arise from increase in proximity (local concentration effect) or residence time (altered  $k_{\text{off}}$ ) of the enzyme in close contact with cellulose, allowing opportunities for the catalytically relevant binding interactions of CBM3c and cd to manifest. Furthermore, with CBM3c, others have noted that alignment of its productive binding sites with the active site cleft seems to support better positioning the substrate for successful binding and hydrolysis (25-27,38,63).

#### *Catalytic efficiency ( $k_{\text{cat}}/K_M$ ) versus specific activity (U/mg)*

A motivating factor for research on cellulases is the need to improve efficiency of the industrial deconstruction of plant biomass (3,16,18). In the context of the use of enzymes in production of biofuels and other renewable products (6,17,64), there are several ways to consider enzyme performance (10,64), including the catalytic efficiency ( $k_{\text{cat}}/K_M$ , i.e., mol of product per mol of enzyme per unit time), specific activity (U/mg, i.e., mol of product per mg of enzyme per unit time), turnover number ( $k_{\text{cat}}$ , i.e., mol product per mol of enzyme) and process mass intensity (grams of input materials, including the mass of enzyme, needed to produce a gram of product). Table 2.1 and Figure 2.4 show that while 3c-3a is superior to 3c for catalytic efficiency and turnover number, 3c has better specific activity and better process mass intensity because its comparable turnover number arises from a smaller molecular mass. Thus, although the addition of linker and CBM3a made the engineered enzyme a more efficient catalyst, the additional protein mass did not give a proportional improvement in specific activity or process mass intensity. In effect, the engineered CBM3a provided no useful improvement in application.

#### *Processivity and exo- or endo- hydrolysis mode*

The reactivity of GH9 has generated considerable interest (24-27), and assessments of processivity and mode of hydrolysis (exo- and/or endo-) are typically based on a determination of the ratio of soluble and insoluble reducing sugar produced from hydrolysis of an insoluble substrate (25,26,35,65,66). Alternatively, the ratio  $k_{cat}/k_{off}$  has also been used as a measure of processivity to evaluate other known processive and non-processive GHs (47,48). This ratio can be calculated from the results of Table 2.1 ( $K_D$ ,  $K_M$ ,  $k_{cat}$ ) and eq 1 to provide  $k_{off}$ :

$$K_M = K_D + \frac{k_{cat}}{k_{on}} \quad (1)$$

Using this approach, earlier studies of processive GH7 gave values  $>100$ , while non-processive GHs gave ratios ranging from  $\sim 5$ -20 (47,48). In this work, we used parameters obtained from normalized assays applied to a broad specificity GH5 subfamily 4 clade (67), and found processivity ratios of  $>150$  for xyloglucan (the kinetically preferred substrate in this clade (67-70)) and glucomannan, indicating an unexpected level of processivity. For CelR, the processivity ratio was lower for reactions with cellulose, ranging from 0.4 (3c on Avicel) to 31 (3c on PASC). Based on this evaluation of kinetic and binding parameters, and comparison to other cellulases, CelR appears to be weakly processive. Since CelR is naturally found bound to a megadalton cellulosome complex by a short linker, the imperative for processivity may be quite different than for a freely diffusing enzyme like Cel9A.

The structure presented here, along with structures and biochemical results for other GH9 (15,35,46), show how CelR produces cellotetraose as its primary hydrolysis product from long oligosaccharides, as the extended substrate receives stabilizing interactions from the  $-4$  to  $+2$  sites prior to hydrolysis (15,35,36). In this sense, CelR is an endolytic cellulase that releases a short, soluble reducing oligosaccharide. However, endolytic hydrolysis, such as suggested for an enzyme like EngD having an active site cleft open on both ends (67,71), would require

considerable conformational distortion of a longer polysaccharide to avoid unfavorable steric interactions with the loop blocking the end of the active site channel in CelR (Figure 2.6B and 2.6C); this loop is seen in many other GH9 (63), but not all (63,72). Binding of an extended polysaccharide would also disrupt the favorable interactions at the  $-4$  subsite provided by  $\text{Ca}^{2+}$  ligands D278 and D279.

### *Structural features of CelR*

The CelR active site is contained in a narrow cleft recessed in the catalytic domain (Figure 2.4B). As is common in many GHs, the walls of this cleft contain residues capable of interacting through Van der Waals and  $\pi$ -stacking interactions between the glucosyl units and residues such as Trp, Tyr, and His (27,38,45,73-75). These non-specific interactions allow for greater substrate mobility than the hydrogen bonding interactions lining the bottom of the active site cleft and at the  $-4$  subsite, which hold the substrate more firmly in place (45,50,74,75).

Although its role in substrate binding has been described elsewhere (27,38,46), the clamp region at the  $+1$  subsite in CelR retains cellobiose after hydrolysis of cellohexaose and release of cellotetraose. Thus, the interactions of H141 and Y435 may also play a role in the retention of longer oligosaccharides interacting with CBM3c, and thus stabilize the enzyme-cellulose complex to potentially enable processive hydrolysis. The closed conformation of Trp residues with the  $-2$  subsite and their role in promoting a catalytically competent configuration between the  $+1$  and  $-1$  position has also been previously reported (27). W330 adopts a similar closed conformation in substrate bound 7v0i. Moreover, the open or closed position of W330 overall correlates with the variation in position of H226 at the  $-3$  subsite (Figure 2.6). H226 is positioned between S225 and S227, which both interact with the  $\text{Ca}^{2+}$  bound near the  $-4$  subsite

and change bond lengths and geometry depending on the presence of substrate or product. These changes propagate to changes in the positioning of H226 and control of access of substrate and product. Moreover, movements of D278, bonded to  $\text{Ca}^{2+}$  by its O atom, are influenced by changes in bonding of E232 on the opposite side of  $\text{Ca}^{2+}$  (and also adjacent D279). This flexibility allows the formation of stabilizing hydrogen bonds with the glucosyl unit in the -4 subsite that help hold the substrate in position for effective catalysis. These potentially coordinated motions may also contribute to catalytic efficiency and perhaps processivity of reaction.

## **Experimental procedures:**

### *Protein expression and purification*

Rosetta cells (Millipore-Sigma) cells transformed with sequence-confirmed plasmids (pVP67K, N-terminal His<sub>8</sub> TEV-cleavable tag) and plated on selective media. Protein expression was performed using autoinduction (76,77). Starter cultures were grown in minimal medium containing glucose overnight and used to inoculate autoinduction medium. Cells were grown at 25°C for 24 h. Cells were harvested by centrifugation and lysed using lysozyme and sonication. Lysate was centrifuged at 20,000 rpm for 45 min, and the supernatant was clarified with a 0.4 μm PES filter. The constructs were separated by nickel affinity chromatography on an AKTA Start FPLC. Purified protein was desalted and concentrated to 10-20 mg/mL.

### *Protein melting experiments*

Protein melt curves were acquired and analyzed using a BioRad CFX Real Connect System. Purified protein solutions were created in differential scanning fluorimetry buffer (25

mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 7.2). Protein denaturation was measured by the change in fluorescence of 128× SYPRO Orange dye (Invitrogen, Carlsbad, CA, USA). The dye solutions were excited at 530 ± 20 nm. Samples contained 20 μM protein and were heated in a 96 well plate from 10 °C to 95 °C at a rate of +1.7 °C per min, with fluorescence measured every 0.5 °C. Chicken lysozyme (Sigma Aldrich, St. Louis, MO, USA) was run with 16× SYPRO Orange as a control. Melt curves were measured for triplicate samples over two identical experiments.

### *Temperature and pH optima*

Optimal temperature for reaction of each construct with either phosphoric acid swollen cellulose (PASC, prepared in house from Avicel) or Avicel (Sigma Aldrich, St. Louis, MO, USA) was determined in triplicate by incubating 250 μL of 1 μM (cd, 3a) or 0.1 μM (3c, 3c-3a) enzyme with either 5 mg/mL PASC or 50 mg/mL Avicel in 100 mM phosphate buffer, pH 6 for 1 h (PASC) or 24 h (Avicel) from 35-85 °C in 5 °C increments with shaking at 1050 rpm on an Eppendorf thermomixer R (catalog #5355). The pH optimum for these reaction conditions was determined at the optimum temperature, using 100 mM acetate buffer (pH 4-5.5), or 100 mM phosphate buffer (pH 6-8) in 0.5 pH unit increments. Enzyme only and substrate only samples were used for controls

Samples were centrifuged for 2 min at 15,000 rpm (21,130 × g) to pellet residual solids, and the concentration of soluble reducing sugar was determined using the Pierce BCA Protein Assay Kit from Thermo Fisher Scientific (Waltham, MA). Briefly, 100 μL of BCA working solution and 5 μL of reaction supernatant were heated at 80 °C for 30 min and absorbance was measured at 562 nm. A glucose standard curve was used to convert absorbance units to reducing sugar concentrations.

### *Enzyme kinetics*

Steady-state kinetic measurements for each construct were made in triplicate on PASC by incubating 250  $\mu\text{L}$  of 0.05  $\mu\text{M}$  enzyme with up to 4 mg/mL PASC for 10 min with shaking at 1050 rpm on an Eppendorf thermomixer R (catalog #5355) and either 100 mM acetate or phosphate buffer at pH and temperature determined by the optima screening. Samples were centrifuged for 30 s at 15,000 rpm to pellet residual solids, and the concentration of soluble reducing sugar was determined using a modification of the BCA method described above, by mixing 50  $\mu\text{L}$  of working solution and 50  $\mu\text{L}$  of sample supernatant instead of 100  $\mu\text{L}$  and 5  $\mu\text{L}$  respectively.

Kinetic parameters for the constructs were determined in triplicate on Avicel using the same methodology as above, except 1  $\mu\text{M}$  enzyme was incubated with up to 180 mg/mL Avicel for 30 min, and the samples were centrifuged for 3 min. Reducing sugar concentrations were measured as described previously, with 100  $\mu\text{L}$  of working solution and 5  $\mu\text{L}$  of sample supernatant. Enzyme only and substrate only samples were used as controls for both sets of kinetic experiments.

### *Enzyme binding to insoluble substrates*

Binding to either PASC or Avicel was determined by pull-down of eGFP fluorescence. Assays were conducted in triplicate with 250  $\mu\text{L}$  of 1  $\mu\text{M}$  of E439Q mutated enzyme having an N-terminal eGFP-tag. Binding was to either PASC concentrations up to 0.5 mg/mL or Avicel concentrations up to 180 mg/mL for 10 min with shaking at 1050 rpm on an Eppendorf thermomixer R (catalog #5355) at the optimal temperature and pH for hydrolysis reactions.

Samples were centrifuged at 15,000 rpm ( $21,130 \times g$ ) for 1 min to pellet enzyme bound to substrate and the fluorescence of 100  $\mu$ L of supernatant was measured at 485 nm excitation and 535 nm emission. Enzyme only and substrate only samples were used as controls.

### *Crystallization, X-ray diffraction, and refinement*

Crystallization experiments and structure determination were conducted in the Collaborative Crystallography Core in the Department of Biochemistry at the University of Wisconsin (Madison, WI, USA). A STP Labtech Mosquito crystallization robot set up Hampton IndexHT and Molecular Dimensions JCSG+ general screens and optimizations in MRC SD-2 crystallization plates. If needed, crystals were cryoprotected, then harvested in MiTeGen MicroMounts and immersed in liquid nitrogen.

Diffraction screening and data collection was conducted at LS-CAT beamlines 21ID-F (0.97872Å) and 21ID-G (0.97856Å) at the Advanced Photon Source, Argonne National Labs. Data was collected on a Rayonix MX3000HE. Diffraction data was reduced and scaled using XDS (78) and XSCALE (79). The Phenix suite (80) of crystallography programs provided the framework for structure solution and refinement. After molecular replacement in Phaser (81), models were improved using phenix.autosol (82). Afterward, models were subjected to alternating rounds of phenix.refine (83) and rebuilding in Coot (84). Active site carbohydrate was placed in the difference electron density maps with a combination of interactive building and automatic docking in phenix.ligandfit (85). Stereochemical restraints for cellobiose and cellohexaose were improved using the GRADE server (86) which uses the Cambridge Crystallographic Data Center program MOGUL (87). Stereochemical restraints were more

successful when the carbohydrates were treated as ligand monomers in refinement, rather than polymers. The structures were validated with MOLPROBITY (88).

Diffraction data extending to 2.0Å for CelR-3c (7unp) came from a crystal grown by combining 200 nL of protein at 29.5mg/mL, 200 nL reservoir composed of 20% w/v polyethylene glycol (PEG) 6000, and 0.1M sodium citrate buffer, pH 5. The crystal was cryopreserved in reservoir solution supplemented with 10% ethylene glycol. Data was collected using 21ID-F on 2020-07-25. The structure was solved by molecular replacement by models derived from PDB ID 1k72. Phaser first located the larger catalytic domain (1k72:19-459) and then the smaller cellulose binding domain (1k72:478-626).

Data extending to 1.9Å for CelR-cc with bound carbohydrate was collected on 21ID-G on 2021-09-19. The structures were solved using the catalytic domain of 7unp. A crystal grown from 0.2 M magnesium chloride, 0.1 M HEPES (2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid) buffer, pH 7.5, 30% v/v polyethylene glycol 400, 2mM cellohexaose (Megazyme) provided data for CelR-cc hexameric cellulose (7v0i). Strong electron density for cellohexaose was present in the active site cleft of each of the three crystallographically independent monomers in the asymmetric unit.

A second, distinctive crystal morphology grew from a similar condition, 0.05 M magnesium chloride, 0.1 M HEPES buffer, pH 7.5, 30% v/v polyethylene glycol monomethyl ether 550, 2 mM cellohexaose. Although these ellipsoidal-shaped crystals were also birefringent, they completely lacked the distinct faces exhibited by the crystals leading to 7v0i. Despite their unusual appearance, their unit cell constants were very similar to 7v0i crystals and gave 2.4 Å data. Surprisingly, electron density in the active site cleft could only accommodate a bound cellobiose (7v0j).



**Data availability:**

The atomic coordinates and structure factors (codes 7v0i, 7v0j, 7unp) have been deposited in the Protein Data Bank (<http://wwpdb.org/>) Diffraction frame data has been deposited in <https://www.proteindiffraction.org>

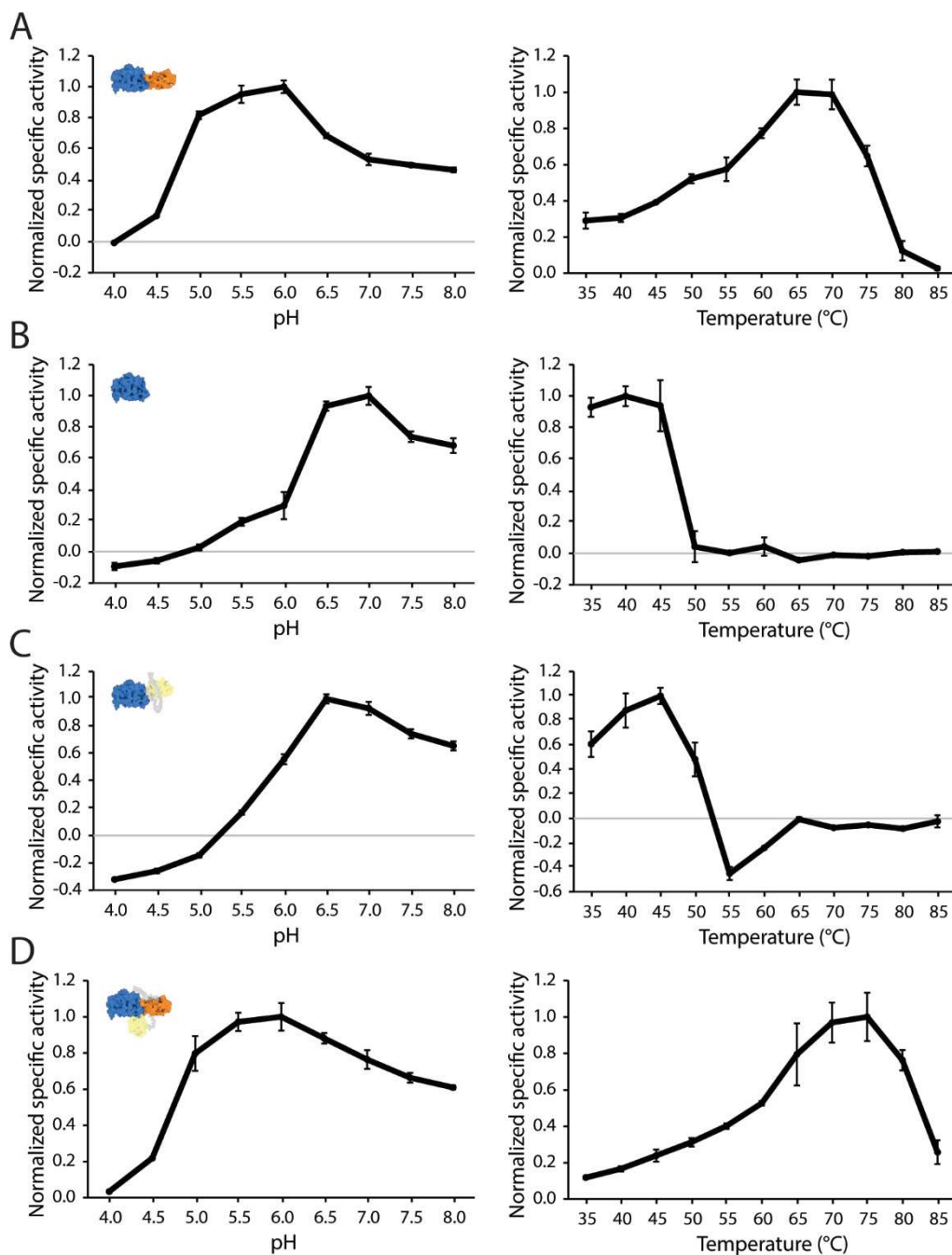
**Acknowledgments:**

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported under Contract No. DE-AC02-05CH11231. The use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). The use of Life Science Collaborative Access Team at the Advanced Photon Source was supported by the College of Agricultural and Life Sciences, Department of Biochemistry, and the Office of the Vice Chancellor for Research and Graduate Education of the University of Wisconsin–Madison. Crystallization, data collection, and structure solution and refinement were conducted by the Department of Biochemistry Collaborative Crystallography Core at the University of Wisconsin–Madison, which is supported by user fees and the department. NJK was supported by the UW-Madison Biotechnology Training Program (NIH 5 T32 GM008349). We thank Craig Ogata, Michael Becker, Nagarajan Venugopalan, Spencer Anderson, Joe Brunzelle, Elena Kondrashkina, and Zdislaw Wawrzak for beamline support and Justin Acheson for assistance in the AlphaFold predictions of 3a and 3c-3a.

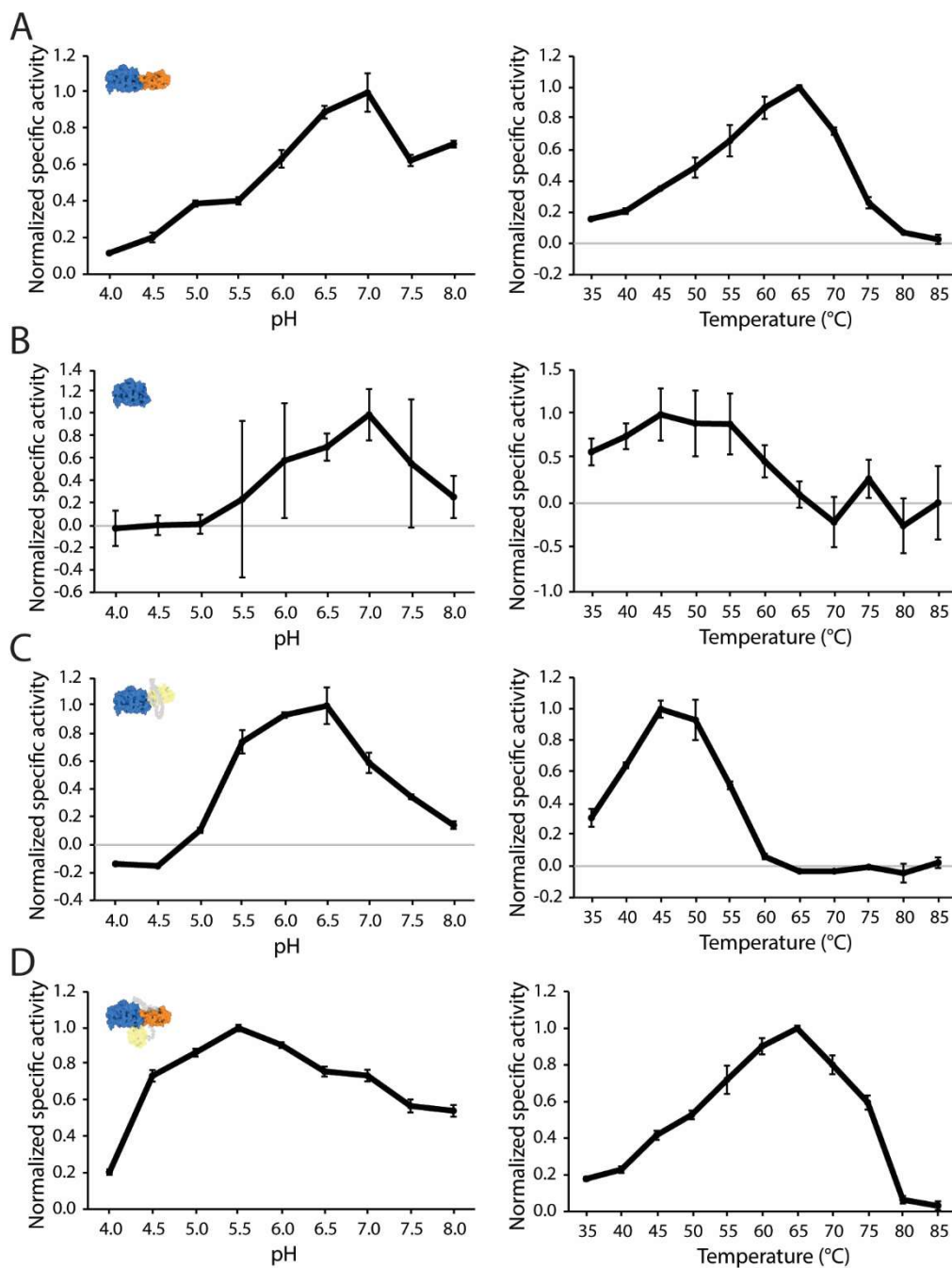
**Conflict of interest:**

The authors declare that they have no conflicts of interest with the contents of this article.

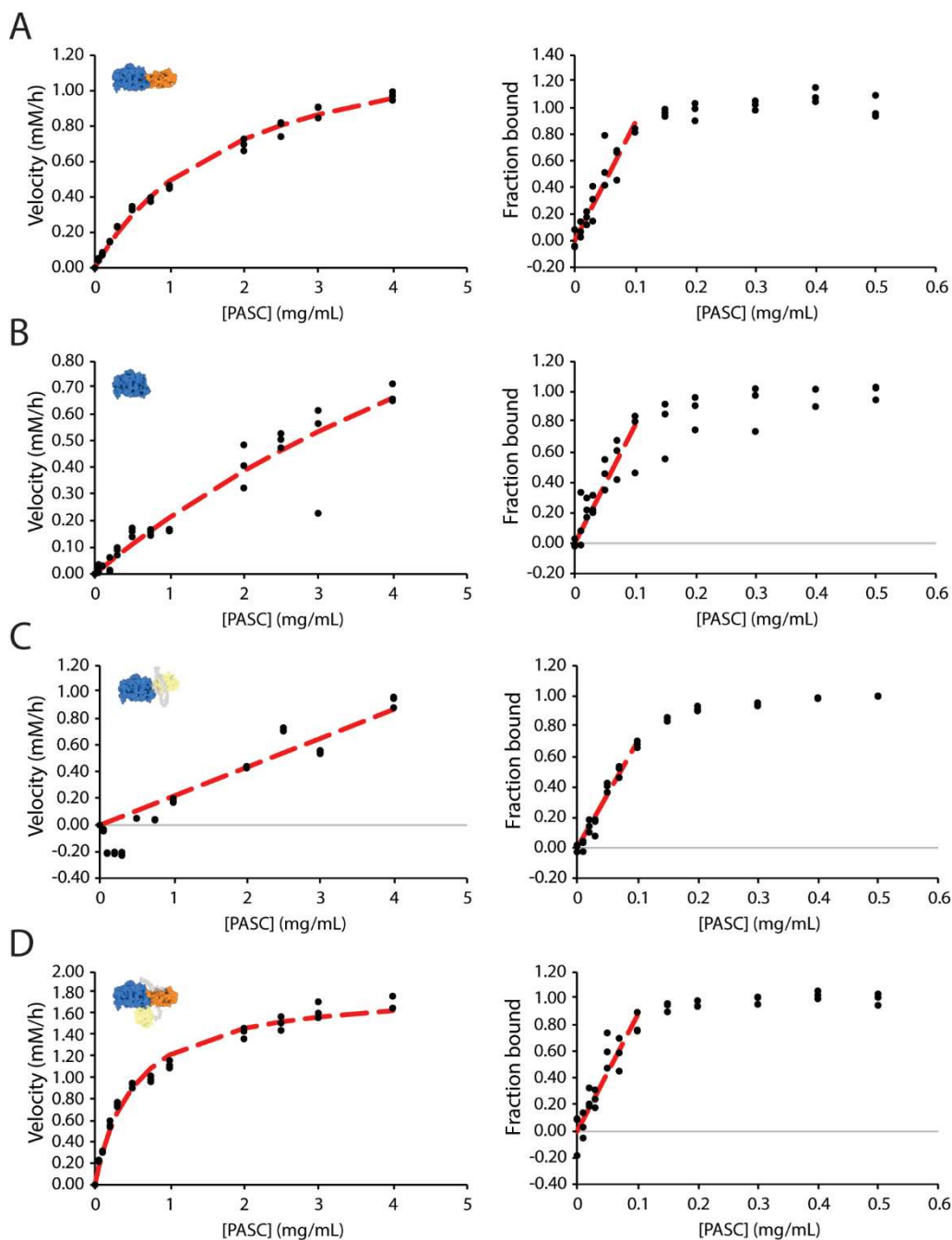
**Supporting information:**



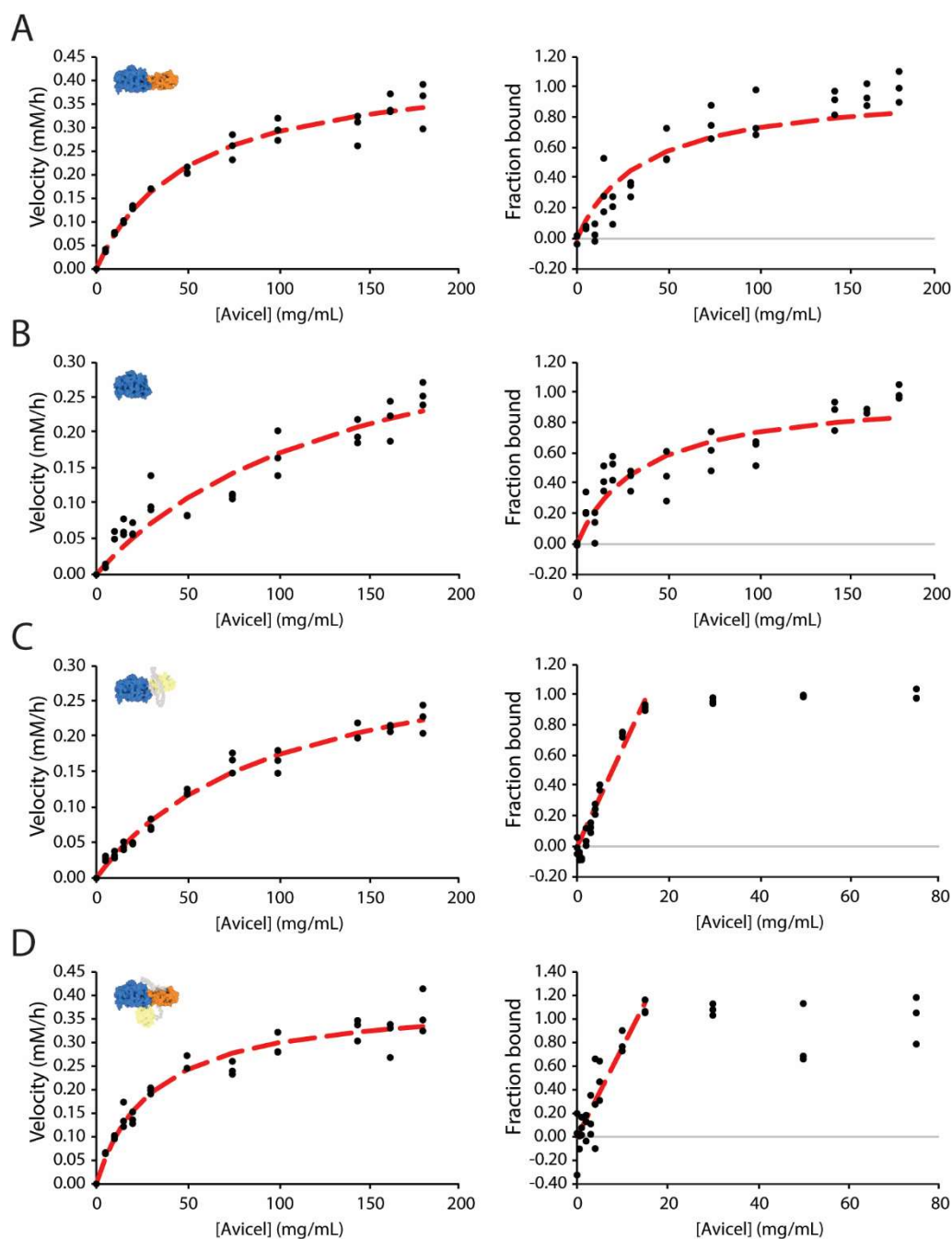
**Figure S2.1:** Relative activity for constructs 3c (A), cd (B), 3a (C), and 3c-3a (D) on PASC from pH 4.0-8.0 and 35-85 °C. Activities were measured in triplicate. The optimal pHs were 6.0, 7.0, 6.5, and 6.0 for 3c, cd, 3a, and 3c-3a respectively. The optimal temperatures were 65, 40, 45, and 75 °C for 3c, cd, 3a, and 3c-3a respectively.



**Figure S2.2:** Relative activity for constructs 3c (A), cd (B), 3a (C), and 3c-3a (D) on Avicel from pH 4.0-8.0 and 35-85 °C. Activities were measured in triplicate. The optimal pHs were 7.0, 7.0, 6.5, and 5.5 for 3c, cd, 3a, and 3c-3a respectively. The optimal temperatures were 65, 45, 45, and 65 °C for 3c, cd, 3a, and 3c-3a respectively.



**Figure S2.3:** Kinetic and binding data for 3c, cd, 3a, and 3c-3a on PASC along with the line of best fit. Data was fit using a traditional Michaelis-Menten equation for kinetics, although neither cd nor 3a saturated activity. The tight binding of all four constructs resulted in a straight line until reaching saturation. Because of this, a linear equation was used for fitting the linear portion of their binding curves.

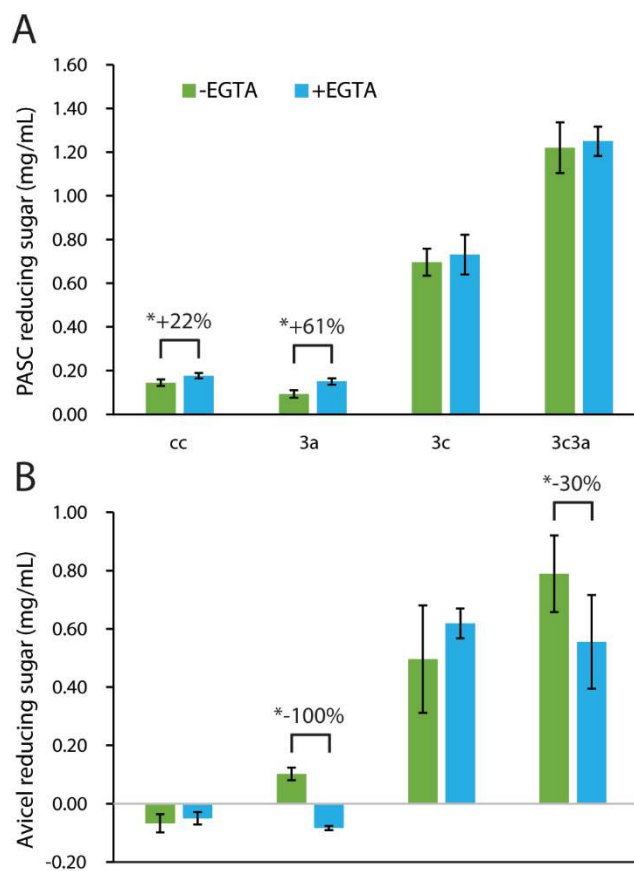


**Figure S2.4:** Kinetic and binding data for 3c, cd, 3a, and 3c-3a on Avicel along with the line of best fit. Data was fit using a traditional Michaelis-Menten equation for kinetics. Binding for 3c was fit using a single binding site equation, while the tight binding of 3a and 3c-3a resulted in a straight line until reaching saturation. Because of this, 3a and 3c-3a were both fit with only a linear equation for the linear portion of their binding curves.

**Table S2.1:** Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	<b>7V0I CelR-catalytic domain bound to cellulose hexamer</b>	<b>7V0J CelR-catalytic domain bound to cellulose dimer</b>	<b>7UNP CelR-3c</b>
<b>Wavelength</b>	0.9786	0.9786	0.9787
<b>Resolution range</b>	39.87 - 1.9 (1.968 - 1.9)	40.17 - 2.4 (2.486 - 2.4)	36.8 - 2.0 (2.071 - 2.0)
<b>Space group</b>	C 2 2 21	C 2 2 21	P 1 21 1
<b>Unit cell</b>	93.302 162.419 210.522 90 90 90	92.585 161.637 210.439 90 90 90	53.751 91.076 66.909 90 110.986 90
<b>Total reflections</b>	937513 (93401)	460077 (45613)	152710 (15355)
<b>Unique reflections</b>	125449 (12423)	61908 (6071)	40523 (4058)
<b>Multiplicity</b>	7.5 (7.5)	7.4 (7.5)	3.8 (3.8)
<b>Completeness (%)</b>	99.92 (99.95)	99.86 (99.80)	99.55 (99.58)
<b>Mean I/sigma(I)</b>	17.41 (1.43)	14.33 (1.40)	8.45 (1.86)
<b>Wilson B-factor</b>	34.18	50.13	19.98
<b>R-merge</b>	0.07696 (1.275)	0.1097 (1.436)	0.1437 (0.7598)
<b>R-meas</b>	0.0827 (1.368)	0.1181 (1.542)	0.1681 (0.8877)
<b>R-pim</b>	0.03009 (0.4941)	0.04323 (0.5591)	0.08648 (0.4556)
<b>CC1/2</b>	0.999 (0.637)	0.999 (0.654)	0.991 (0.683)
<b>CC*</b>	1 (0.882)	1 (0.889)	0.998 (0.901)
<b>Reflections used in refinement</b>	125393 (12422)	61876 (6071)	40513 (4059)
<b>Reflections used for R- free</b>	3037 (307)	3061 (311)	2006 (206)
<b>R-work</b>	0.1870 (0.3125)	0.1957 (0.3544)	0.1617 (0.2340)
<b>R-free</b>	0.2317 (0.3257)	0.2277 (0.3775)	0.2080 (0.2939)
<b>CC(work)</b>	0.968 (0.772)	0.969 (0.791)	0.968 (0.873)

<b>CC(free)</b>	0.950 (0.797)	0.956 (0.733)	0.945 (0.719)
<b>Number of non-hydrogen atoms</b>	11697	11066	5443
<b>macromolecules</b>	10830	10816	4895
<b>ligands</b>	375	138	2
<b>solvent</b>	663	178	546
<b>Protein residues</b>	1357	1357	608
<b>RMS (bonds)</b>	0.003	0.004	0.003
<b>RMS (angles)</b>	0.63	0.98	0.54
<b>Ramachandran favored (%)</b>	96.23	95.63	96.37
<b>Ramachandran allowed (%)</b>	3.63	4.07	3.63
<b>Ramachandran outliers (%)</b>	0.15	0.30	0.00
<b>Rotamer outliers (%)</b>	0.82	0.82	0.00
<b>Clashscore</b>	1.31	1.28	0.53
<b>Average B-factor</b>	44.22	63.18	23.71
<b>macromolecules</b>	43.98	63.15	22.95
<b>ligands</b>	59.52	81.52	36.33
<b>solvent</b>	43.44	57.45	30.47
<b>Number of TLS groups</b>	20	19	4



**Figure S2.5:** Hydrolysis yields of CelR constructs on PASC (A) or Avicel (B) after 24 h incubation at room temperature with (blue) or without (green) EGTA. Hydrolysis was done at the optimal pH and temperature for the construct/substrate pair for 2 h (PASC) or 24 h (Avicel) without shaking. Results were analyzed via the modified BCA detailed in the methods; differences in yields that had  $p \leq 0.05$  in a two-tailed t-test are noted with \*, and the percent change in yield caused by addition of EGTA.

## References:

1. Roman-Leshkov, Y., Barrett, C. J., Liu, Z. Y., and Dumesic, J. A. (2007) Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature* **447**, 982-985
2. West, R. M., Liu, Z. Y., Peter, M., and Dumesic, J. A. (2008) Liquid alkanes with targeted molecular weights from biomass-derived carbohydrates. *ChemSusChem* **1**, 417-424
3. Yang, B., and Wyman, C. E. (2008) Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuel Bioprod Bior* **2**, 26-40
4. Moody, J. W., McGinty, C. M., and Quinn, J. C. (2014) Global evaluation of biofuel potential from microalgae. *Proc Natl Acad Sci U S A* **111**, 8691-8696
5. Bhalla, A., Fasahati, P., Particka, C. A., Assad, A. E., Stoklosa, R. J., Bansal, N., Semaan, R., Saffron, C. M., Hodge, D. B., and Hegg, E. L. (2018) Integrated experimental and techno-economic evaluation of two-stage Cu-catalyzed alkaline-oxidative pretreatment of hybrid poplar. *Biotechnol Biofuels* **11**, 143



6. Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Lee, T. S., Tullman-Ercek, D., Voigt, C. A., Simmons, B. A., and Keasling, J. D. (2011) Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci U S A* **108**, 19949-19954
7. Cai, C. M., Zhang, T. Y., Kumar, R., and Wyman, C. E. (2013) THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chem* **15**, 3140-3145
8. Luterbacher, J. S., Rand, J. M., Alonso, D. M., Han, J., Youngquist, J. T., Maravelias, C. T., Pfleger, B. F., and Dumesic, J. A. (2014) Nonenzymatic sugar production from biomass using biomass-derived gamma-valerolactone. *Science* **343**, 277-280
9. Lorenci Woiciechowski, A., Dalmas Neto, C. J., Porto de Souza Vandenberghe, L., de Carvalho Neto, D. P., Novak Sydney, A. C., Letti, L. A. J., Karp, S. G., Zevallos Torres, L. A., and Soccol, C. R. (2020) Lignocellulosic biomass: acid and alkaline pretreatments and their effects on biomass recalcitrance - conventional processing and recent advances. *Bioresour Technol* **304**, 122848
10. Nguyen, Q. A., and Saddler, J. N. (1991) An integrated model for the technical and economic-evaluation of an enzymatic biomass conversion process. *Bioresource Technol* **35**, 275-282
11. Jorgensen, H., Kristensen, J. B., and Felby, C. (2007) Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuel Bioprod Bior* **1**, 119-134
12. Zhao, X., Cheng, K., and Liu, D. (2009) Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Appl Microbiol Biotechnol* **82**, 815-827
13. Walker, T. W., Kuch, N., Vander Meulen, K. A., Clewett, C. F. M., Huber, G. W., Fox, B. G., and Dumesic, J. A. (2020) Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pre-treatment for enzymatic conversion of residual cellulose. *ACS Sustain Chem Eng* **8**, 6551-6563
14. Ravachol, J., Borne, R., Tardif, C., de Philip, P., and Fierobe, H. P. (2014) Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium *Clostridium cellulolyticum*. *J Biol Chem* **289**, 7335-7348
15. Leis, B., Held, C., Bergkemper, F., Dennemarck, K., Steinbauer, R., Reiter, A., Mechelke, M., Moerch, M., Graubner, S., Liebl, W., Schwarz, W. H., and Zverlov, V. V. (2017) Comparative characterization of all cellulosomal cellulases from *Clostridium thermocellum* reveals high diversity in endoglucanase product formation essential for complex activity. *Biotechnol Biofuels* **10**
16. Hendriks, A. T., and Zeeman, G. (2009) Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresour Technol* **100**, 10-18
17. Oke, M. A., Annuar, M. S. M., and Simarani, K. (2016) Mixed feedstock approach to lignocellulosic ethanol production-prospects and limitations. *Bioenerg Res* **9**, 1189-1203
18. Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F., and da Silva, S. S. (2018) The path forward for lignocellulose biorefineries: bottlenecks, solutions, and perspective on commercialization. *Bioresour Technol* **264**, 370-381
19. Hettle, A., Fillo, A., Abe, K., Massel, P., Pluvinage, B., Langelaan, D. N., Smith, S. P., and Boraston, A. B. (2017) Properties of a family 56 carbohydrate-binding module and its role in the recognition and hydrolysis of beta-1,3-glucan. *J Biol Chem* **292**, 16955-16968

20. Nakamura, A., Tasaki, T., Ishiwata, D., Yamamoto, M., Okuni, Y., Visootsat, A., Maximilien, M., Noji, H., Uchiyama, T., Samejima, M., Igarashi, K., and Iino, R. (2016) Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *Trichoderma reesei* Cel6A and its domains on crystalline cellulose. *J Biol Chem* **291**, 22404-22413
21. Sammond, D. W., Payne, C. M., Brunecky, R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2012) Cellulase linkers are optimized based on domain type and function: insights from sequence analysis, biophysical measurements, and molecular simulation. *PLoS One* **7**, e48615
22. Tang, C. D., Li, J. F., Wei, X. H., Min, R., Gao, S. J., Wang, J. Q., Yin, X., and Wu, M. C. (2013) Fusing a carbohydrate-binding module into the *Aspergillus usamii* beta-mannanase to improve its thermostability and cellulose-binding capacity by in silico design. *PLoS One* **8**, e64766
23. Walker, J. A., Takasuka, T. E., Deng, K., Bianchetti, C. M., Udell, H. S., Prom, B. M., Kim, H., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol Biofuels* **8**, 220
24. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J Bacteriol* **180**, 1709-1714
25. Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Appl Environ Microbiol* **73**, 3165-3172
26. Li, Y., Irwin, D. C., and Wilson, D. B. (2010) Increased crystalline cellulose activity via combinations of amino acid changes in the family 9 catalytic domain and family 3c cellulose binding module of *Thermobifida fusca* Cel9A. *Appl Environ Microbiol* **76**, 2582-2588
27. Sakon, J., Irwin, D., Wilson, D. B., and Karplus, P. A. (1997) Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. *Nat Struct Biol* **4**, 810-818
28. Kostylev, M., Moran-Mirabal, J. M., Walker, L. P., and Wilson, D. B. (2012) Determination of the molecular states of the processive endocellulase *Thermobifida fusca* Cel9A during crystalline cellulose depolymerization. *Biotechnol Bioeng* **109**, 295-299
29. Zhou, W., Irwin, D. C., Escovar-Kousen, J., and Wilson, D. B. (2004) Kinetic studies of *Thermobifida fusca* Cel9A active site mutant enzymes. *Biochemistry* **43**, 9655-9663
30. Escovar-Kousen, J. M., Wilson, D., and Irwin, D. (2004) Integration of computer modeling and initial studies of site-directed mutagenesis to improve cellulase activity on Cel9A from *Thermobifida fusca*. *Appl Biochem Biotechnol* **113-116**, 287-297
31. Costa, M. G. S., Silva, Y. F., and Batista, P. R. (2018) Computational engineering of cellulase Cel9A-68 functional motions through mutations in its linker region. *Phys Chem Chem Phys* **20**, 7643-7652
32. Oliveira, O. V., Freitas, L. C., Straatsma, T. P., and Lins, R. D. (2009) Interaction between the CBM of Cel9A from *Thermobifida fusca* and cellulose fibers. *J Mol Recognit* **22**, 38-45
33. Breyer, W. A., and Matthews, B. W. (2001) A structural basis for processivity. *Protein Sci* **10**, 1699-1711

34. Wilson, D. B., and Kostylev, M. (2012) Cellulase processivity. *Methods Mol Biol* **908**, 93-99
35. Zverlov, V. V., Schantz, N., and Schwarz, W. H. (2005) A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-beta-1,4-glucanase producing cellotetraose. *FEMS Microbiol Lett* **249**, 353-358
36. Deng, K., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases. *Front Bioeng Biotechnol* **3**, 165
37. Gal, L., Gaudin, C., Belaich, A., Pages, S., Tardif, C., and Belaich, J. P. (1997) CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. *J Bacteriol* **179**, 6595-6601
38. Mandelman, D., Belaich, A., Belaich, J. P., Aghajari, N., Driguez, H., and Haser, R. (2003) X-Ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides. *J Bacteriol* **185**, 4127-4135
39. Hirano, K., Nihei, S., Hasegawa, H., Haruki, M., and Hirano, N. (2015) Stoichiometric assembly of the cellulosome generates maximum synergy for the degradation of crystalline cellulose, as revealed by *in vitro* reconstitution of the *Clostridium thermocellum* cellulosome. *Appl Environ Microbiol* **81**, 4756-4766
40. Gilmore, S. P., Lillington, S. P., Haitjema, C. H., de Groot, R., and O'Malley, M. A. (2020) Designing chimeric enzymes inspired by fungal cellulosomes. *Synth Syst Biotechnol* **5**, 23-32
41. Gilcher, E., Kuch, N., Walker, T. W., Vander Meulen, K. A., Clewett, C. F. M., Huber, G. W., Fox, B. G., and Dumesic, J. A. (2022) unpublished data.
42. Liao, S. M., Liang, G., Zhu, J., Lu, B., Peng, L. X., Wang, Q. Y., Wei, Y. T., Zhou, G. P., and Huang, R. B. (2019) Influence of calcium ions on the thermal characteristics of alpha-amylase from thermophilic *Anoxybacillus sp. GXS-BL*. *Protein Pept Lett* **26**, 148-157
43. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008) Data mining of metal ion environments present in protein structures. *J Inorg Biochem* **102**, 1765-1776
44. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J* **15**, 5739-5751
45. Knott, B. C., Crowley, M. F., Himmel, M. E., Stahlberg, J., and Beckham, G. T. (2014) Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity. *J Am Chem Soc* **136**, 8810-8819
46. Jeng, W. Y., Liu, C. I., Lu, T. J., Lin, H. J., Wang, N. C., and Wang, A. H. (2019) Crystal structures of the C-terminally truncated endoglucanase Cel9Q from *Clostridium thermocellum* complexed with cello-dextrins and Tris. *Chembiochem* **20**, 295-307
47. Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K., and Westh, P. (2013) Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: effects of substrate structure and carbohydrate binding domain. *Biochemistry* **52**, 8938-8948
48. Kurasin, M., and Valjamae, P. (2011) Processivity of cellobiohydrolases is limited by the substrate. *J Biol Chem* **286**, 169-177

49. Simpson, P. J., Xie, H., Bolam, D. N., Gilbert, H. J., and Williamson, M. P. (2000) The structural basis for the ligand specificity of family 2 carbohydrate-binding modules. *J Biol Chem* **275**, 41137-41142
50. Petkun, S., Rozman Grinberg, I., Lamed, R., Jindou, S., Burstein, T., Yaniv, O., Shoham, Y., Shimon, L. J., Bayer, E. A., and Frolow, F. (2015) Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker. *PeerJ* **3**, e1126
51. Chai, K. P., Othman, N. F., Teh, A. H., Ho, K. L., Chan, K. G., Shamsir, M. S., Goh, K. M., and Ng, C. L. (2016) Crystal structure of *Anoxybacillus* alpha-amylase provides insights into maltose binding of a new glycosyl hydrolase subclass. *Sci Rep* **6**, 23126
52. Sanan, T. T., Muthukrishnan, S., Beck, J. M., Tao, P., Hayes, C. J., Otto, T. C., Cerasoli, D. M., Lenz, D. E., and Hadad, C. M. (2010) Computational modeling of human paraoxonase 1: preparation of protein models, binding studies, and mechanistic insights. *J Phys Org Chem* **23**, 357-369
53. Aggarwal, G., Prajapati, R., Tripathy, R. K., Bajaj, P., Iyengar, A. R., Sangamwar, A. T., and Pande, A. H. (2016) Toward understanding the catalytic mechanism of human paraoxonase 1: site-specific mutagenesis at position 192. *PLoS One* **11**, e0147999
54. Hershko Rimon, A., Livnah, O., Rozman Grinberg, I., Ortiz de Ora, L., Yaniv, O., Lamed, R., Bayer, E. A., Frolow, F., and Voronov-Goldman, M. (2021) Novel clostridial cell-surface hemicellulose-binding CBM3 proteins. *Acta Crystallogr F Struct Biol Commun* **77**, 95-104
55. Yaniv, O., Halfon, Y., Shimon, L. J., Bayer, E. A., Lamed, R., and Frolow, F. (2012) Structure of CBM3b of the major cellulosomal scaffoldin subunit ScaA from *Acetivibrio cellulolyticus*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 8-13
56. Santos, C. R., Paiva, J. H., Sforca, M. L., Neves, J. L., Navarro, R. Z., Cota, J., Akao, P. K., Hoffmam, Z. B., Meza, A. N., Smetana, J. H., Nogueira, M. L., Polikarpov, I., Xavier-Neto, J., Squina, F. M., Ward, R. J., Ruller, R., Zeri, A. C., and Murakami, M. T. (2012) Dissecting structure-function-stability relationships of a thermostable GH5-CBM3 cellulase from *Bacillus subtilis* 168. *Biochem J* **441**, 95-104
57. Chiriac, A. I., Cadena, E. M., Vidal, T., Torres, A. L., Diaz, P., and Pastor, F. I. (2010) Engineering a family 9 processive endoglucanase from *Paenibacillus barcinonensis* displaying a novel architecture. *Appl Microbiol Biotechnol* **86**, 1125-1134
58. Su, X., Mackie, R. I., and Cann, I. K. (2012) Biochemical and mutational analyses of a multidomain cellulase/mannanase from *Caldicellulosiruptor bescii*. *Appl Environ Microbiol* **78**, 2230-2240
59. Yi, Z., Su, X., Revindran, V., Mackie, R. I., and Cann, I. (2013) Molecular and biochemical analyses of CbCel9A/Cel48A, a highly secreted multi-modular cellulase by *Caldicellulosiruptor bescii* during growth on crystalline cellulose. *PLoS One* **8**, e84172
60. Chundawat, S. P. S., Nemmaru, B., Hackl, M., Brady, S. K., Hilton, M. A., Johnson, M. M., Chang, S., Lang, M. J., Huh, H., Lee, S. H., Yarbrough, J. M., Lopez, C. A., and Gnanakaran, S. (2021) Molecular origins of reduced activity and binding commitment of processive cellulases and associated carbohydrate-binding proteins to cellulose III. *J Biol Chem* **296**, 100431
61. Linder, M., Mattinen, M. L., Kontteli, M., Lindeberg, G., Stahlberg, J., Drakenberg, T., Reinikainen, T., Pettersson, G., and Annala, A. (1995) Identification of functionally

- important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci* **4**, 1056-1064
62. Varrot, A., Schulein, M., and Davies, G. J. (1999) Structural changes of the active site tunnel of *Humicola insolens* cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry* **38**, 8884-8891
  63. Kumar, K., Singh, S., Sharma, K., and Goyal, A. (2021) Computational modeling and small-angle X-ray scattering based structure analysis and identifying ligand cleavage mechanism by processive endocellulase of family 9 glycoside hydrolase (HtGH9) from *Hungateiclostridium thermocellum* ATCC 27405. *J Mol Graph Model* **103**, 107808
  64. Johnson, E. (2016) Integrated enzyme production lowers the cost of cellulosic ethanol. *Biofuel Bioprod Bior* **10**, 164-174
  65. Beckham, G. T., Stahlberg, J., Knott, B. C., Himmel, M. E., Crowley, M. F., Sandgren, M., Sorlie, M., and Payne, C. M. (2014) Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. *Curr Opin Biotechnol* **27**, 96-106
  66. Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J Am Chem Soc* **135**, 18831-18839
  67. Glasgow, E. M., Kemna, E. I., Bingman, C. A., Ing, N., Deng, K., Bianchetti, C. M., Takasuka, T. E., Northen, T. R., and Fox, B. G. (2020) A structural and kinetic survey of GH5\_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *J Biol Chem* **295**, 17752-17769
  68. Glasgow, E. M., Vander Meulen, K. A., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Deutsch, S., and Fox, B. G. (2019) Extent and origins of functional diversity in a subfamily of glycoside hydrolases. *J Mol Biol* **431**, 1217-1233
  69. Attia, M. A., Nelson, C. E., Offen, W. A., Jain, N., Davies, G. J., Gardner, J. G., and Brumer, H. (2018) *In vitro* and *in vivo* characterization of three *Cellvibrio japonicus* glycoside hydrolase family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnol Biofuels* **11**, 45
  70. Xi, L., Sheng, D., Ye-Fei, W., Xiao-Qing, M., Haixia, Z., Xinquan, W., Li-Shan, Y., Yingang, F., and Fu-Li, L. (2017) Structural insights into the substrate specificity of a glycoside hydrolase family 5 lichenase from *Caldicellulosiruptor* sp. F32. *Biochemical Journal* **474**, 17p
  71. Foong, F. C., and Doi, R. H. (1992) Characterization and comparison of *Clostridium cellulovorans* endoglucanases-xylanases EngB and EngD hyperexpressed in *Escherichia coli*. *J Bacteriol* **174**, 1403-1409
  72. Khademi, S., Guarino, L. A., Watanabe, H., Tokuda, G., and Meyer, E. F. (2002) Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*. *Acta Crystallogr D Biol Crystallogr* **58**, 653-659
  73. Igarashi, K., Koivula, A., Wada, M., Kimura, S., Penttila, M., and Samejima, M. (2009) High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *J Biol Chem* **284**, 36186-36190
  74. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Gotz, A. W., Sandgren, M., Withers, S. G., Stahlberg, J., and Beckham, G. T. (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J Am Chem Soc* **136**, 321-329

75. Nakamura, A., Watanabe, H., Ishida, T., Uchihashi, T., Wada, M., Ando, T., Igarashi, K., and Samejima, M. (2014) Trade-off between processivity and hydrolytic velocity of cellobiohydrolases at the surface of crystalline cellulose. *J Am Chem Soc* **136**, 4584-4592
76. Blommel, P. G., Becker, K. J., Duvnjak, P., and Fox, B. G. (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol Prog* **23**, 585-598
77. Fox, B. G., and Blommel, P. G. (2009) Autoinduction of protein expression. *Curr Protoc Protein Sci* **Chapter 5**, Unit 5 23
78. Kabsch, W. (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132
79. Diederichs, K. (2006) Some aspects of quantitative analysis and correction of radiation damage. *Acta Crystallographica: Section D (Wiley-Blackwell)* **62**, 6p
80. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., and Stockwell, D. H. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica: Section D, Structural Biology* **75**, 17p
81. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674
82. Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Li-Wei, H. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica: Section D (Wiley-Blackwell)* **65**, 20p
83. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352-367
84. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
85. Terwilliger, T. C., Klei, H., Adams, P. D., Moriarty, N. W., and Cohn, J. D. (2006) Automated ligand fitting by core-fragment fitting and extension into density. *Acta Crystallographica: Section D (Wiley-Blackwell)* **62**, 8p
86. Smart, O. S., Womack, T. O., Sharff, A., Flensburg, C., Keller, P., Paciorek, W., Vonrhein, C., and Bricogne, G. (2011) Grade, version 1.2.20 Cambridge, United Kingdom, Global Phasing Ltd., <https://www.globalphasing.com>.
87. Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E., and Orpen, A. G. (2004) Retrieval of crystallographically-derived molecular geometry information. *J Chem Inf Comput Sci* **44**, 2133-2144
88. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21

## Chapter 3:

### **Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pre-treatment for enzymatic conversion of residual cellulose**

Theodore W. Walker, Nathaniel Kuch, Kirk A. Vander Meulen, Catherine F. M. Clewett, George W. Huber, Brian G. Fox and James A. Dumesic

This work was originally published as:

Walker T.W., Kuch N., Vander Meulen K.A., Clewett C.F.M., Huber G.W., Fox B.G., Dumesic J.A. 2020. Solid-state NMR studies of solvent-mediated, acid-catalyzed woody biomass pretreatment for enzymatic conversion of residual cellulose. *ACS Sustainable Chemistry & Engineering*. 10.1021/acssuschemeng.0c01538

Nathaniel Kuch made the following contributions: planning and execution of enzyme assays, data analysis, writing and revising of manuscript

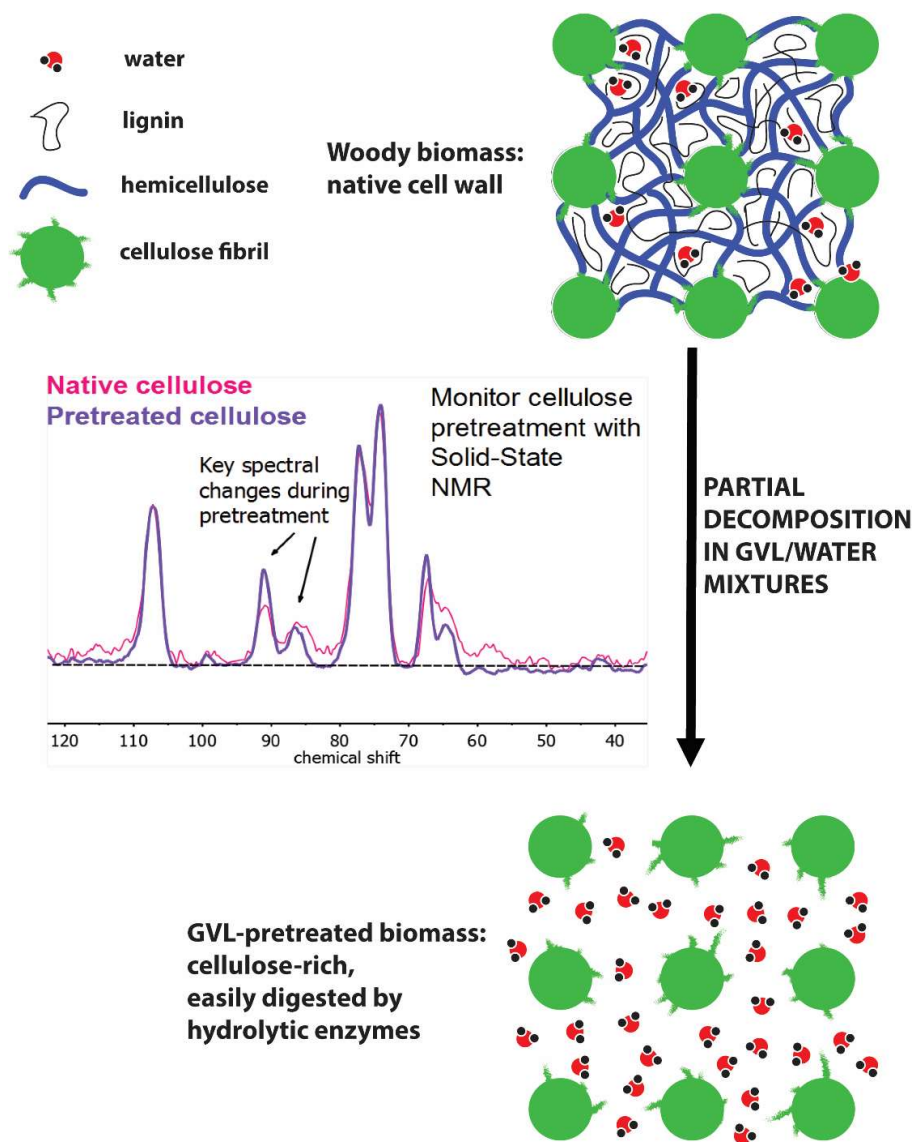
#### **Abstract:**

Enzymes selectively hydrolyze the carbohydrate fractions of lignocellulosic biomass into corresponding sugars, but these processes are limited by low yields and slow catalytic turnovers. Under certain conditions, the rates and yields of enzymatic sugar production can be increased by pretreating biomass using solvents, heat and dilute acid catalysts. However, the mechanistic details underlying this behavior are not fully elucidated, and designing effective pretreatment strategies remains an empirical challenge. Herein, using a combination of solid-state and high-resolution magic-angle-spinning NMR, infrared spectroscopy and X-ray diffractometry, we show that the extent to which cellulase enzymes are able to hydrolyze solvent-pretreated biomass can be understood in terms of the ability of the solvent to break the chemical linkages between cellulose and non-cellulosic materials in the cell wall. This finding is of general significance to enzymatic biomass conversion research, and implications for designing improved biomass conversion strategies are discussed. These findings demonstrate the utility of solid-state NMR as

a tool to elucidate the key chemical and physical changes that occur during the liquid-phase conversion of real biomass.

**Key words: biomass, pretreatment, solid-state NMR, cellulose, enzymatic conversion**

**Graphical Abstract:**



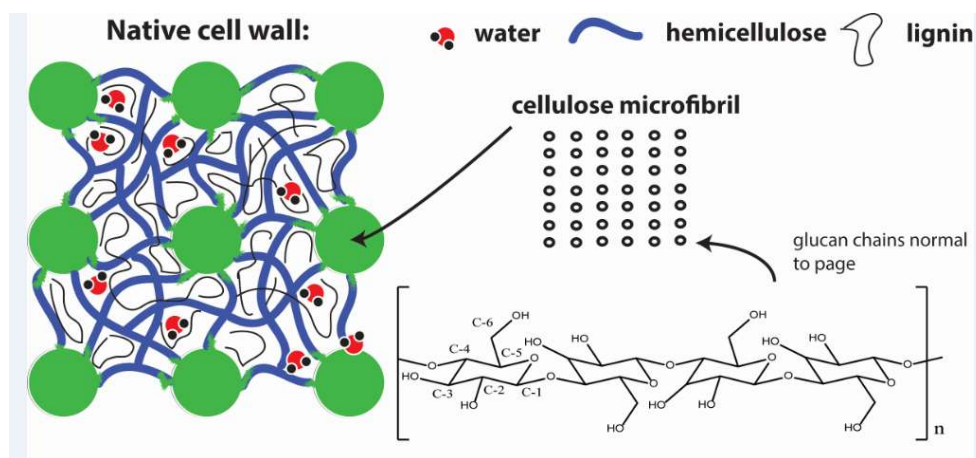
**Introduction:**

Lignocellulosic plant matter (*biomass*) is a heterogeneous, polymeric material composed mainly of xylan (hemicellulose), glucan (cellulose) and lignin, along with entrained water, minerals, proteins and extractives.<sup>1-3</sup> Of these constituent fractions, cellulose is present in



amorphous phases and relatively crystalline phases<sup>4</sup> with characteristic lateral dimensions (microfibrils).<sup>5-6</sup> The microfibrils are bound to the less-ordered (amorphous) cellulose, hemicellulose and lignin via covalent and non-covalent interactions to form the cell wall.<sup>7-8</sup> These structures are represented schematically in Figure 3.1.

As an abundant and naturally occurring source of carbon, deconstruction of biomass to yield platform chemicals is a strategy to partially displace global demand for fossil resources.<sup>9-11</sup> To this end, several technologies have been proposed, including enzymatic conversion<sup>12-13</sup> and solvent-based methods.<sup>14-15</sup> Of these two examples, solvent-assisted conversion of biomass over acid catalysts is characterized by rapid catalytic turnovers, but suffers from poor selectivity and often requires elevated temperatures ( $\geq 160^\circ\text{C}$ ).<sup>9, 16</sup> Hydrolytic enzymes selectively hydrolyze cellulose and hemicellulose into corresponding sugars at mild temperatures,<sup>17</sup> but these processes are limited by high operating costs and slow catalytic turnovers.<sup>18-19</sup>



**Figure 3.1.** Schematic representation of the cellulose-, hemicellulose- and lignin-containing structures present in the cell wall of woody biomass. Cellulose microfibrils are semi-crystalline in nature, and are constituted by parallel chains of glucan arrayed in ordered planes (represented here by an arbitrary  $m \times n$  matrix, with ellipses denoting single chains whose long axes are oriented normal to the page). The six carbon centers present in the repeating units of cellulose are labeled according to IUPAC convention.

One strategy to address these challenges is to combine enzymatic hydrolysis of biomass with a solvent-mediated pretreatment step,<sup>20</sup> whereby the solvent (usually containing a dilute acid catalyst) first dissolves the hemicellulose and lignin fractions, leaving behind a cellulose-rich solid that is more easily hydrolyzed by enzymes than the native cell wall.<sup>21</sup> However, the details connecting the physical and chemical changes effected by the solvent to the improved yield of the enzymatic hydrolysis step are not fully elucidated. For example, it has been proposed that the solvent alters the cellulose pore structure,<sup>22</sup> and/or the crystallinity of cellulose;<sup>23</sup> these changes hypothetically improve the accessibility of the glycosidic linkages to enzymes, or the affinity of enzymes for the surface accessible facets of the cellulose, respectively. However, attempts to verify these hypotheses have often proven to be inconclusive,<sup>24</sup> or yielded contradictory results.<sup>25-26</sup> As such, few fundamental approaches exist to understand the effectiveness of biomass pretreatment strategies, as determined by improved sugar yields in the enzymatic hydrolysis step.

Herein, we present a strategy to probe the physical and chemical properties of crystalline cellulose present in raw and solvent-pretreated biomass using a combination of high resolution magic-angle-spinning (HR/MAS) and solid-state cross-polarization (CP)/MAS nuclear magnetic resonance spectroscopy (NMR), along with infrared spectroscopy (FTIR) and X-ray diffraction (XRD) techniques. Proton spin-relaxation-edited (PSRE) CP/MAS NMR is used to differentiate <sup>13</sup>C signals originating from the crystalline and amorphous portions of cellulose, which overlap in non-PSRE NMR spectra. We show that as hemicellulose and lignin are removed by the solvent (observable by NMR and FTIR), the characteristic dimensions of the remaining crystalline cellulose microfibrils (observable by XRD) are unchanged. In contrast, as the lignin and hemicellulose are removed, the magnetic environment at the exterior of the *crystalline*

cellulose microfibrils (observable by  $^{13}\text{C}$  PSRE CP/MAS NMR) is quantifiably altered. The latter result indicates that the cohesive forces in the interior of the crystalline cellulose microfibrils are comparable to those bonding the microfibril exteriors to hemicellulose and lignin in the plant cell wall. Accordingly, the presence of hemicellulose and lignin likely prevents cellulose-specific enzymes from binding to the surface-accessible glycosidic linkages on the microfibril exterior, inhibiting cellulose hydrolysis.  $^1\text{H}$  HR/MAS NMR reveals that as the residual cellulose from pretreated-biomass is dried, distinct hydration environments form within the pore structure, and water binds more strongly with the surface accessible facets of the pore walls. This behavior likely slows the diffusion of cellulases within the cellulose pore structure, further hindering enzymatic conversion.

We combine these characterization results with enzymatic reactivity studies to show that the NMR-observable reductions in cellulose-xylan interactions and increased water binding strength in solvent-pretreated biomass can be correlated with the yields of sugars afforded by hydrolysis of the residual cellulose by engineered cellulases. To our knowledge, this study represents the first use of solid-state and HR/MAS NMR to establish this relationship. This finding is of general significance, as the correlation between reduced cellulose-xylan interactions and increased enzymatic sugar production from cellulose is of general importance between different biomass types (corn stover, hard woods, etc.) and solvents (tetrahydrofuran, dimethyl sulfoxide, etc.). Therefore, any strategy to reduce cellulose-xylan interactions in the cell wall while preserving the native cellulose pore structure should produce a corresponding increase in the enzymatic digestibility of the resulting cellulose. Furthermore, we believe that this study demonstrates the use of solid-state NMR to study the liquid-phase conversion of real biomass,

affording insights that complement techniques such as solution-phase NMR, X-ray scattering and FTIR.

## Results:

### *Solvent-mediated pretreatment of biomass over sulfuric acid*

Figure 3.2 displays a three-step pretreatment process based on prior work by our own group and others.<sup>10, 14, 27-29</sup> Generally, this process consists of:

- (1) A low temperature (60-120°C) ***lignin removal step***, wherein raw biomass is treated in a solvent system consisting of mostly of an organic solvent, with some water and a dilute mineral acid catalyst.
- (2) A higher temperature (100-160°C) ***sugar removal step***, wherein the residual solids from Step (1) are treated in a solvent system, again consisting of an organic solvent and dilute acid catalyst, but with more water than Step (1), and;
- (3) A ***bleaching step***, wherein trace impurities are removed from the remaining solids to yield a pristine cellulose product.

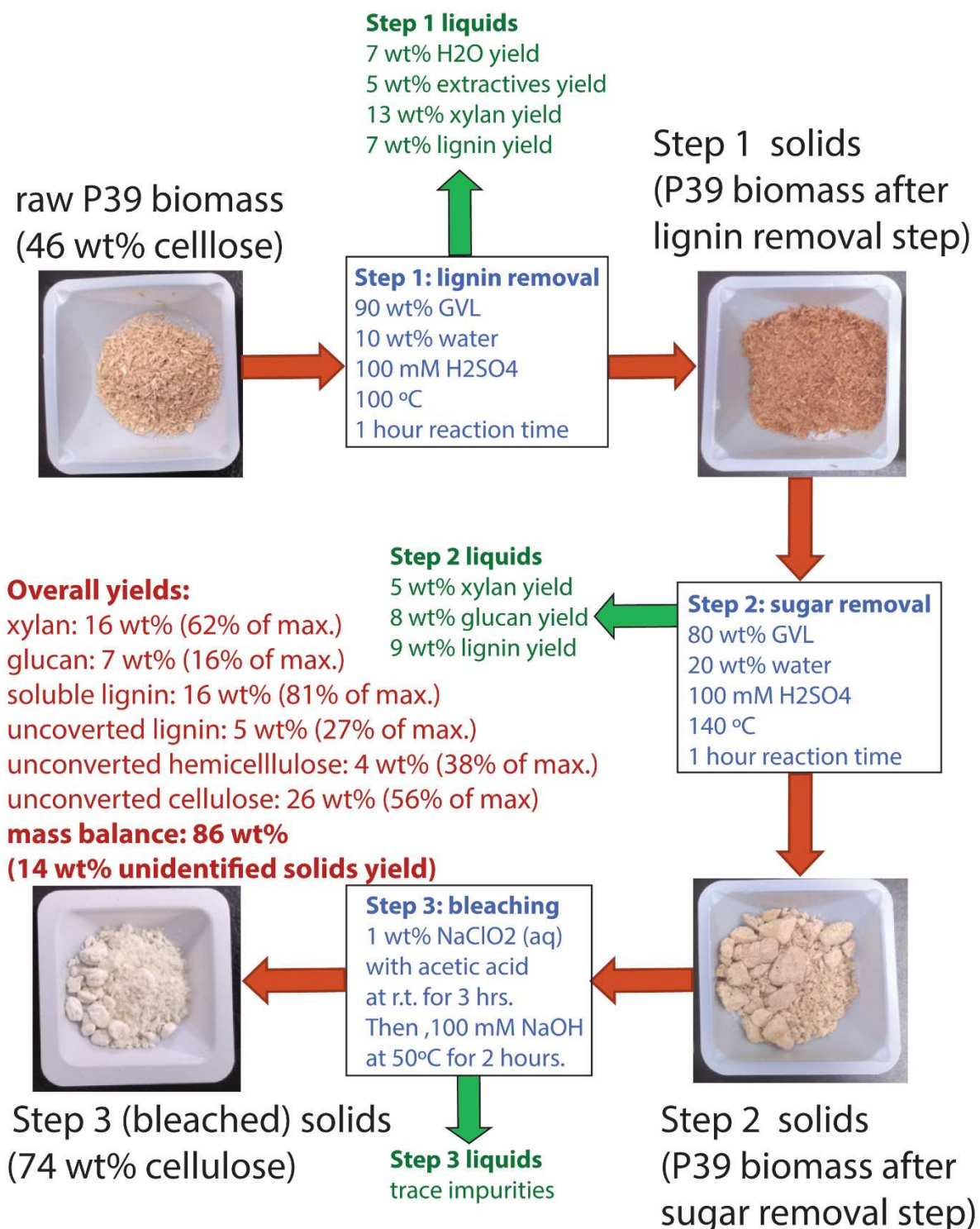
The purpose of the first step is to remove as large a portion of the lignin as possible in a form that closely resembles the native lignin in the cell wall; higher temperatures and solvents systems containing more water alter the chemical nature of the extracted lignin.<sup>30-32</sup> The first step also removes part of the hemicellulose in the form of soluble sugar monomers and oligomers (xylan), along with the entrained water and extractives in the biomass. The purpose of the second step is to remove additional amounts of hemicellulose and lignin, but a portion of the cellulose may also be removed in this step in the form of soluble six-carbon sugar monomers and oligomers (glucan). The bleaching step, if used, removes trace lignin and degraded sugars (humins),<sup>33</sup> which have been shown to inhibit the activity of some hydrolytic enzymes.<sup>34</sup>

In the example displayed in Figure 3.2, the organic solvent used is  $\gamma$ -valerolactone (GVL), a non-toxic and bio-renewable solvent<sup>35</sup> that has the demonstrated abilities to solubilize

all fractions of woody biomass,<sup>14</sup> and to facilitate selective liquid-phase biomass conversion reactions over acid catalysts.<sup>27, 36</sup> Other organic solvents may be used, so long as they are miscible with water (which is required to solubilize the carbohydrate fractions of biomass),<sup>37</sup> and are stable at the desired pretreatment temperatures. Other solvents used in this study were: tetrahydrofuran (THF), acetonitrile (MeCN); *N*-methyl-pyrrolidinone (NMP); and dimethyl sulfoxide (DMSO).

The biomass type shown in Figure 3.2 is P39 poplar, an engineered deciduous energy crop optimized for growth on marginal lands (*i.e.* lands that are not useable for growing food).<sup>38</sup> Three other high yield but non-food-competitive biomass feedstocks were studied in this work: corn stover, switchgrass, sorghum and an additional engineered poplar (NM6).<sup>39</sup> The catalyst (H<sub>2</sub>SO<sub>4</sub>) used in Figure 3.2 is a standard reagent used in demonstrated biomass pretreatment/conversion strategies.<sup>40</sup> The bleaching step is based on industry-standard protocols, using water as a solvent.<sup>41</sup>

In Figure 3.2, representative reaction conditions are shown, with corresponding yields of soluble products in each liquid fraction. Note that the mass balance for the overall process is less than 100%, which results from the formation of unaccountable soluble and insoluble polymeric species (humins), which are ubiquitous in biomass conversion processes.<sup>42</sup> The yields shown in Figure 3.2 are typical for those reported under similar conditions.<sup>14, 27</sup> Table 3.1, below, summarizes all pretreatment conditions investigated in this study.

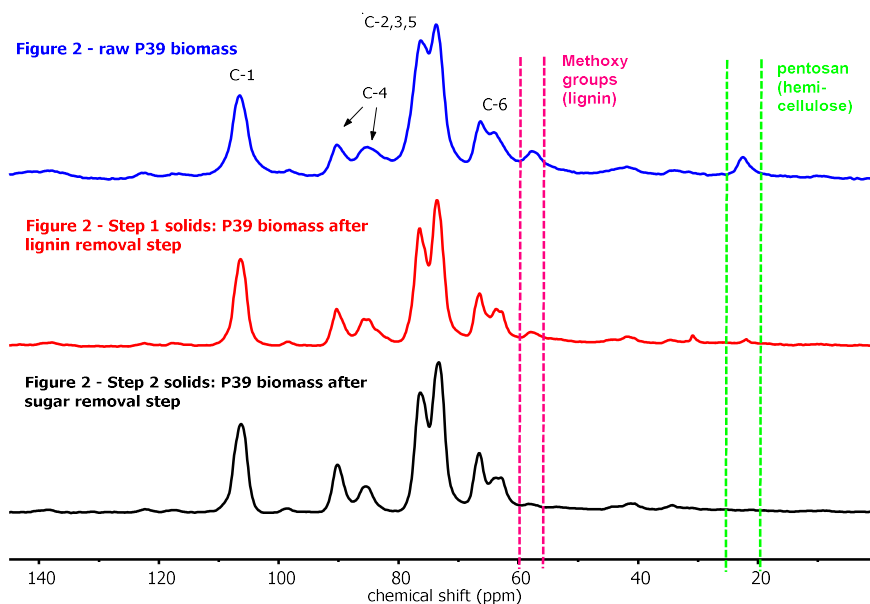


**Figure 3.2.** An exemplary solvent-assisted biomass pretreatment sequence based on the general two-step lignin- and then sugar-removal process described above, followed by an additional bleaching step. Reaction conditions:  $\gamma$ -valerolactone (GVL) solvent with HPLC-grade water as a cosolvent. Sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) is used as catalyst. Each step was carried out at 9/1 wt/wt liquids to solids in pressure-sealed glass reactors, with mixing facilitated by a magnetic stir bar agitated at 500 rpm. See ESI for further details.

### *<sup>13</sup>C solid-state cross-polarization magic-angle-spinning (CP/MAS) NMR*

<sup>13</sup>C solid-state CP/MAS NMR is a technique whereby magnetization is transferred from the protons to their attached carbon nuclei in a solid sample for the purpose of enhancing the <sup>13</sup>C signal and improving resolution for this low natural-abundance nucleus. The sample is spun at high frequency (~1-10 kHz) to mimic the solution-phase behaviors that average out chemical shift anisotropies in liquid samples and produce sharp, well-separated features in the NMR spectrum. This technique has seen increasing use in biomass conversion research as late, as it is a non-destructive technique that allows for the *in-situ* characterization of whole biomass.<sup>43-45</sup> See methods and ESI for details.

Figure 3.3 displays <sup>13</sup>C CP/MAS NMR spectra for three of the representative biomass samples shown in Figure 3.2: raw P39 biomass, the same biomass pretreated in GVL-water mixtures to remove lignin (Figure 3.2, Step 1 solids as described in Entry 1 of Table 3.1), and then hemicellulose (Figure 3.2, Step 2 solids as described in Entry 9 of Table 3.1). Key resonances are delineated in Figure 3.3, corresponding to the six carbon centers in cellulose,<sup>46</sup> as well as the acetate groups in hemicellulose, and the methoxy groups present in lignin.<sup>47</sup> These latter two resonances decrease in intensity by about 74 and 65%, respectively, in the pretreated samples as compared to the raw P39, indicating a corresponding reduction in the lignin and hemicellulose content, as expected. See ESI for quantitative, NMR-enabled assays of lignin, hemicellulose cellulose contents in raw and solvent-pretreated biomass samples. This behavior is confirmed by Fourier Transform Infrared Spectroscopy (FTIR), also shown the ESI.



**Figure 3.3.**  $^{13}\text{C}$  MAS-NMR spectra corresponding to native P39 biomass, and P39 biomass pretreated to remove lignin and hemicellulose. Conditions for the pretreatment steps associated with each spectrum are noted in the Figure 3.2.

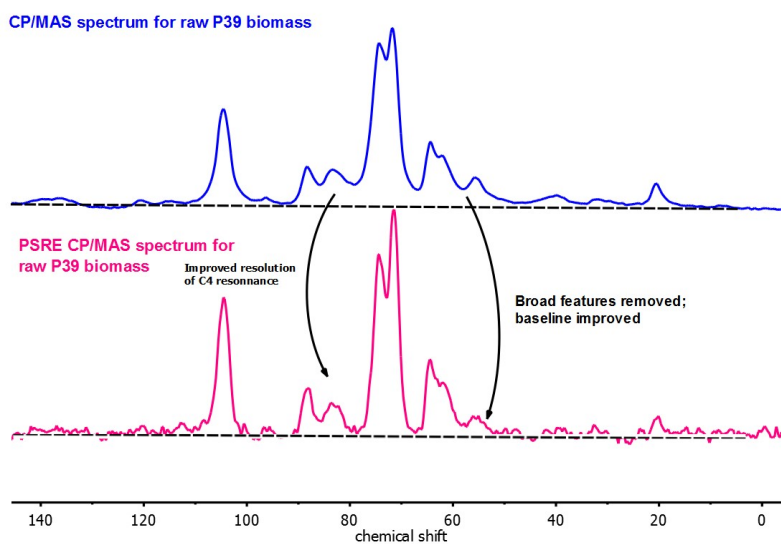
#### *Proton spin-relaxation edited (PSRE) CP/MAS NMR*

Proton spin-relaxation editing (PSRE) is a solid-state NMR technique whereby differences in the rotating-frame  $^1\text{H}$  relaxation behavior for crystalline materials (such as cellulose) and amorphous materials (such as hemicellulose or lignin) are exploited to differentiate their corresponding NMR signals when they have similar chemical shifts. Briefly, a key time delay in the CP/MAS pulse sequence (spin-lock time) is modulated; increasing spin-lock time results in less magnetization being transferred to the carbon nuclei, and therefore decreased signal in the CP/MAS spectrum. Importantly, the extent to which the NMR signal for different carbon nuclei in a solid sample is reduced at a fixed spin-lock time is a function of the chemical environment, allowing for signals for different materials or solid phases to be differentiated, even if their NMR signals overlap in the chemical shift domain. Herein, we followed the method developed by Newman *et al.*<sup>48</sup> to filter out the solid-state NMR signals



corresponding to amorphous cellulose, hemicellulose and lignin in biomass, allowing for the NMR signal corresponding to crystalline cellulose to be isolated and analyzed. See the methods section for more details, and the ESI for a more complete description of this technique.

Figure 3.4 displays the total  $^{13}\text{C}$  CP/MAS NMR spectra for raw P39 biomass, along with the PSRE sub-spectrum corresponding only to the crystalline cellulose in the same sample. Note that several broad spectral features (corresponding to amorphous cellulose, lignin and hemicellulose) are mostly removed in the PSRE spectrum, resulting in better-separated peaks and a flat baseline as compared to the normal CP/MAS spectrum.<sup>48</sup> The resonances corresponding to the C-4 and C-6 carbon centers in the PSRE and the unedited CP/MAS NMR spectra are split into two non-equivalent features. This behavior has been attributed to differences in the chemical environment between cellulose chains located on the outside *vs.* the inside of the microfibrils.<sup>49</sup> In particular, the C-4 carbon resonances that correspond to these two environments are well separated in the PSRE spectrum.

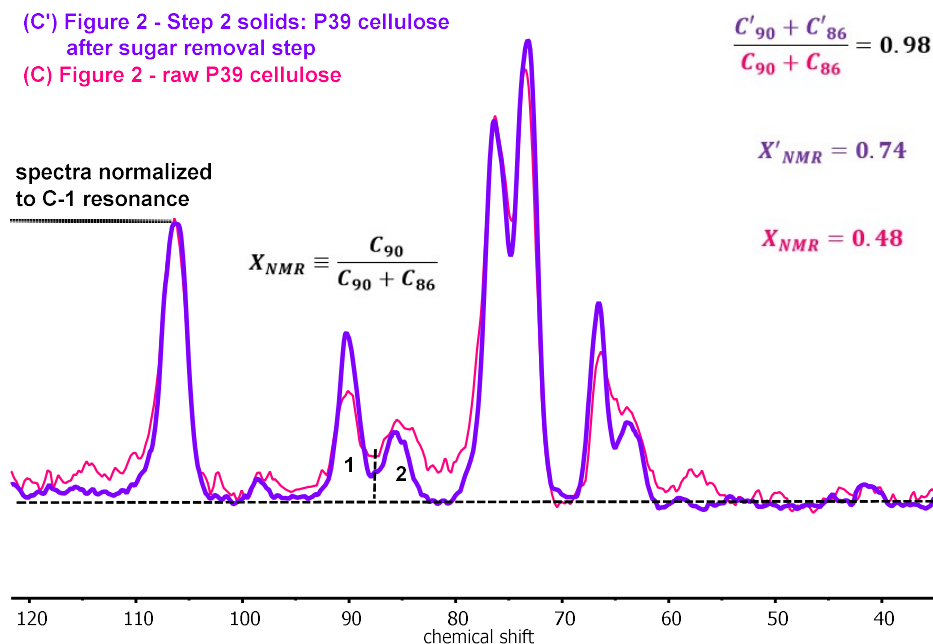


**Figure 3.4.** CP/MAS-NMR spectra for native P39 biomass. The top display is the unedited spectrum produced by averaging 2056 scans with a 10 sec recycle delay time and 0.03 ms spin-lock mixing time. The bottom display is the proton spin-relaxation edited (PSRE) spectrum corresponding to the crystalline cellulose fraction of native P39.

Following these insights, PSRE cellulose spectra corresponding to the native and pretreated biomass investigated in this study were collected, and compared on the basis of the relative peak areas for the downfield C-4 resonance at ~90 ppm and the upfield C-4 resonance at ~86 ppm, denoted  $C_{90}$  and  $C_{86}$ , respectively. Figure 3.5 displays this procedure. When comparing the PSRE cellulose spectra corresponding to raw and pretreated P39 biomass (with peak areas normalized to the C-1 resonance), note that the combined area for the upfield and downfield C-4 peaks remains constant. In contrast, the *relative* C-4 peak areas, are expressed by the relationship:

$$X_{NMR} = \frac{C_{90}}{C_{90} + C_{86}} \quad (1)$$

reveals an enrichment of the downfield C-4 peak area going from the raw to the pretreated P39 cellulose. Together, these results indicate a *redistribution* of the C-4 carbons from one chemical environment to the other, rather than a selective depletion of one phase during the GVL pretreatment process. This behavior has been attributed to a reduction in xylan-cellulose interactions at the surface of the cellulose microfibrils.<sup>50-52</sup> Furthermore, the average size of the crystalline cellulose microfibrils (observable by XRD; see ESI) did not decrease during pretreatment, indicating that changes in  $X_{NMR}$  during biomass pretreatment could not be attributed to an increase in the fraction of glucan chains on the exterior of the microfibrils, as reported elsewhere.<sup>48</sup> Accordingly, we attribute increases in  $X_{NMR}$  resulting from solvent-mediated pretreatment of biomass to the destruction of cellulose-hemicellulose linkages at the surface of the cellulose microfibrils.



**Figure 3.5.** Proton spin-lock edited (PSRE) CP/MAS NMR spectra corresponding to the crystalline cellulose fraction of native P39 biomass (pink display), and P39 biomass pretreated in GVL/water systems to remove lignin and hemicellulose (purple display).

### *Enzymatic hydrolysis of residual cellulose*

The GVL-pretreated solids derived from each step in Figure 3.2 were collected, washed with water, filtered to produce a solid product of about 70 wt% moisture content, and hydrolyzed in the presence of engineered cellulases (see ESI for details). Similar experiments were also conducted for the different biomass types (corn stover, switchgrass, sorghum and NM6-poplar), and solvent systems (THF, DMSO, NMP and MeCN) used in this study. Bleaching steps were applied to some of the GVL-pretreated samples prior to enzymatic hydrolysis to investigate the effects of this procedure on the digestibility of the residual cellulose. Finally, some of the bleached and unbleached GVL-pretreated celluloses were dried in a vacuum oven at 85°C overnight, and then subjected to enzymatic hydrolysis as well. The yields of soluble sugars derived from this enzymatic hydrolysis step were then compared to assess the effect of each pretreatment step on the enzymatic reactivity of the residual cellulose, whereby higher yields of

sugars from the enzymatic hydrolysis step indicate a more effective pretreatment process. These results are displayed in Table 3.1, along with the values of  $X_{NMR}$ .

**Table 3.1.** Reaction conditions for solvent-assisted pretreatment of various biomass, and corresponding enzymatic sugar yields and the NMR-derived observable  $X_{NMR}$ . The temperature in the lignin removal step was 100°C for each entry. The catalyst was 100 mM H<sub>2</sub>SO<sub>4</sub> for all lignin- and sugar-removal steps. Bleaching was carried out as described in the methods above. See ESI for full details and corresponding liquid-phase product assays. Enzymatic hydrolysis was performed with 10 wt% CelR cellulase (see methods) and 0.5 mg/mL pretreated material in a total volume of 1 mL for 24 hours with shaking.

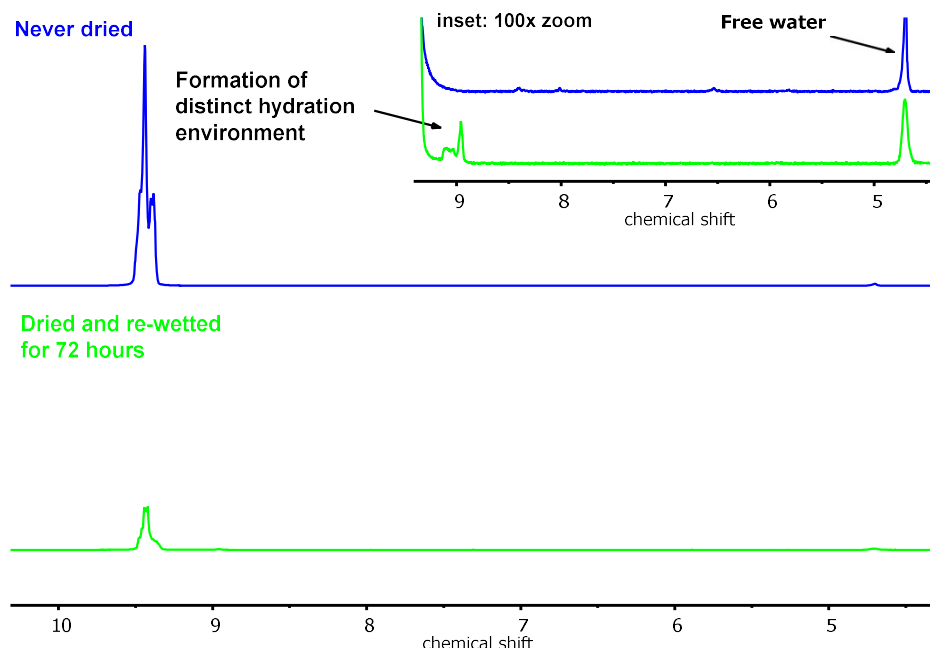
Entry	Biomass type	Organic solvent (wt%) lignin removal step	Organic solvent (wt%) sugar removal step	Temp. – sugar removal step / °C	Bleached ? (Y/N)	Dried? (Y/N)	$X_{NMR}$	Enzymatic sugar yield from residual cellulose (wt% of theor. maximum)
1	P39	GVL (90)	--	--	N	N	0.504	0.0
2	P39	GVL (90)	--	--	N	Y	0.485	0.0
3	P39	GVL (90)	--	--	Y	N	0.530	0.0
4	P39	GVL (90)	--	--	Y	Y	0.507	0.0
5	P39	GVL (90)	GVL (80)	120	N	N	0.545	0.3
6	P39	GVL (90)	GVL (80)	120	N	Y	0.522	0.0
7	P39	GVL (90)	GVL (80)	120	Y	N	0.608	3.7
8	P39	GVL (90)	GVL (80)	120	Y	Y	0.580	1.8
9	P39	GVL (90)	GVL (80)	140	N	N	0.745	31.8
10	P39	GVL (90)	GVL (80)	140	N	Y	0.710	7.4
11	P39	GVL (90)	GVL (80)	140	Y	N	0.775	38.0
12	P39	GVL (90)	GVL (80)	140	Y	Y	0.747	11.7
13	P39	THF (90)	THF (80)	140	N	N	0.779	32.3
14	P39	MeCN (90)	MeCN (80)	140	N	N	0.684	14.1
15	P39	NMP (90)	NMP (80)	140	N	N	0.609	6.4
16	P39	DMSO (90)	DMSO (80)	140	N	N	0.633	6.9
17	NM6	GVL (90)	GVL (80)	140	N	N	0.711	28.1
18	Sorghum	GVL (90)	GVL (80)	140	N	N	0.582	7.9
19	switchgrass	GVL (90)	GVL (80)	140	N	N	0.545	10.4
20	Corn stover	GVL (90)	GVL (80)	140	N	N	0.698	20.1

### <sup>1</sup>H High-Resolution (HR)/MAS NMR spectroscopy

<sup>1</sup>H High-Resolution (HR)/MAS NMR is a technique whereby solution-phase protons in liquid/solid mixtures can be characterized. Radiofrequency pulse strengths in HR/MAS NMR experiments are set to probe the liquid-phase protons in the same manner as a solution NMR experiment, but high frequency sample spinning is still required to average out chemical shift

anisotropies resulting from intermolecular interactions at the liquid-solid interface. See ESI for details.

Figure 3.6 displays representative HR/MAS spectra corresponding to P39 biomass pretreated in GVL-water solvents to remove hemicellulose and lignin for never-dried and dried materials (Entries 9 and 10 in Table 3.1). The dried material has been re-wetted by soaking in water (7/3 water to dry solids) for 72 hours (see methods and ESI). HR/MAS spectra were collected with 100 single-rotor-cycle interpulse delays in a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence for a total of 50 ms transverse relaxation time. Note that each of the spectra within Figure 3.6 display resonances corresponding to the solution-phase protons in the wetted cellulose; *i.e.*, to liquid water only. Without any magnification, only one resonance is visible: a large multiplet at about 9.5 ppm corresponding to water wetting the surface-accessible facets of the cellulose. The inset of Figure 3.6 shows the same spectra magnified 100 times. In these magnified images, we see the singlet at 4.7 ppm corresponding to bulk water. Notably, there is a small feature present in the spectra corresponding to the re-wetted material that is not present in the never-dried material. The appearance of this feature indicates the formation of distinct hydration environment upon drying and re-wetting, which we attribute to a change in the cellulose pore structure upon drying. This behavior has been attributed to water becoming trapped in segregated hydration environments upon rapid changes in the pore structure of semi-crystalline polymeric materials.<sup>53</sup> Accordingly, assign the minor feature at 8.9 ppm to a partially “collapsed” cellulose pore structure that forms upon drying.



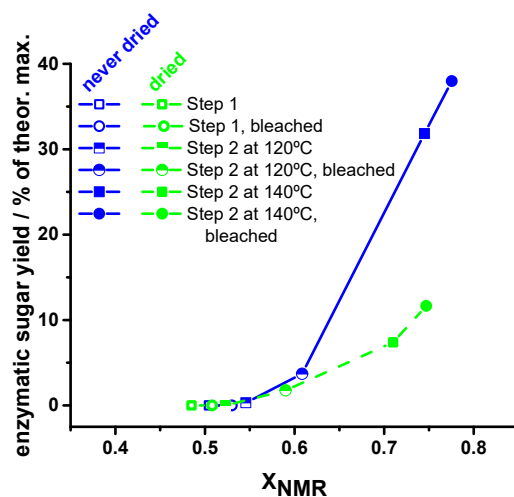
**Figure 3.6.** HR/MAS NMR spectra with a 50 ms Hahn-echo delay applied (100 single-rotor-cycle interpulse delays in a Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence) for cellulose from P39 biomass pretreated in 90 wt% GVL at 100°C for one hour, then 80 wt% GVL at 140°C for one hour (Table 3.1, Entries 9 and 10). Top spectrum corresponds to cellulose that was never dried, and analyzed directly after collection from the pretreatment steps. Bottom spectrum corresponds to the same material after drying in a vacuum oven overnight, then re-wetting by soaking in water for 72 hours.

### Discussion:

To understand the effects of different solvent-mediated biomass pretreatment strategies on the enzymatic digestibility of the residual cellulose, we first analyze the GVL-pretreated P39 biomass samples and corresponding enzymatic sugar yields to probe the effects of the lignin removal step, the sugar removal step, and the bleaching and drying steps. In doing so, we establish that the enzymatic sugar yields from GVL-pretreated P39 cellulose can be correlated with respect to  $X_{NMR}$ , and we propose a physical basis for this behavior. We then establish that this relationship can be extended across all pretreated cellulose samples investigated in this study, regardless of the native biomass type or pretreatment solvent used.

*Effects of GVL-pretreatment, bleaching and drying on the enzymatic digestibility of cellulose derived from P39 biomass*

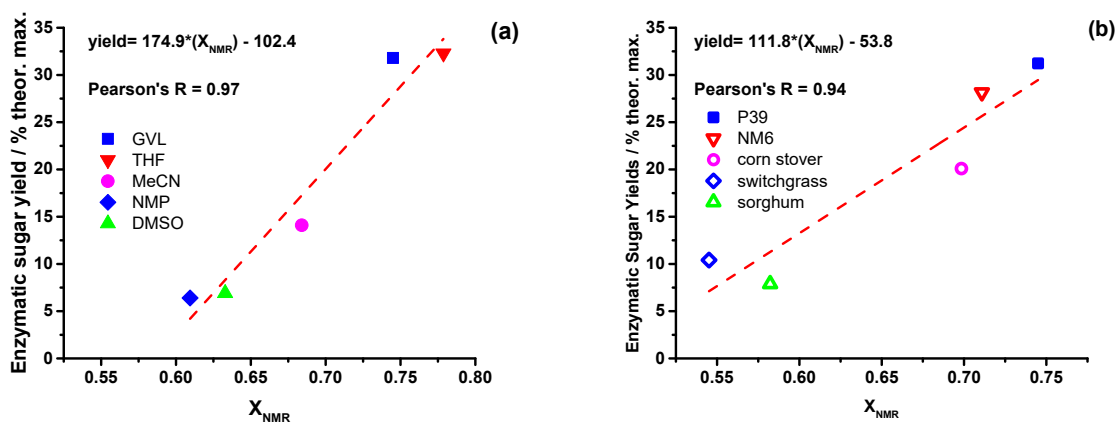
If increases in  $X_{NMR}$  correspond to a reduction in cellulosic material that is bonded to hemicellulose and lignin, and if these extracellular materials occlude enzymes' access to the cellulose microfibrils<sup>25</sup>, then the value of  $X_{NMR}$  might serve as a predictor of the enzymatic digestibility of the residual cellulose derived from GVL-pretreated biomass. Following this hypothesis, Figure 3.7 displays the yields of enzymatic sugars from the hydrolysis of GVL-pretreated P39 (both with and without additional bleaching) as function of  $X_{NMR}$ . Two curves are shown: one corresponding to the never-dried samples, and the same samples after drying in a vacuum oven at -15 in. Hg for 24 hours (open symbols). In general, for both the never-dried and dried GVL-pretreated P39 celluloses, the yield of enzymatic sugars after a 24 reaction time increases monotonically with  $X_{NMR}$ .



**Figure 3.7.** Enzymatic sugar yields from P39 biomass pretreated with GVL-water solvent systems under various conditions, both with and without bleaching steps (entries 1 through 12 in Table 3.1).

We next compare enzymatic sugar yields as a function of  $X_{NMR}$  for the different biomass types and solvent systems studied, where there were no bleaching steps or drying, and the sugar

removal step was fixed at 140°C (Figure 3.8). As demonstrated by the linear regression analyses displayed in Figure 3.8, enzymatic reactivity generalizes universally with  $X_{NMR}$  across all solvent systems and biomass types at a fixed set of reaction times and temperatures for the lignin removal and sugar removal steps.



**Figure 3.8.** Enzymatic sugar yields versus  $X_{NMR}$  for: (a) P39 biomass pretreated in various water/organic-solvent mixtures to remove lignin and hemicellulose (entries 11, 15-18 in Table 3.1), and; (b) different biomass types pretreated in GVL-water solvent mixtures to remove lignin and hemicellulose (entries 11, 19-22 in Table 3.1). Least squares linear fit to the data are shown, demonstrating a quantitative correlation between  $X_{NMR}$  and enzymatic sugar yields across various biomass types and pretreatment solvents.

Finally, Figure 3.9 displays enzymatic sugar yields as a function of  $X_{NMR}$  for all entries in Table 3.1 corresponding to *never-dried* celluloses. As demonstrated by the linear regression in Figure 3.9, there is a quantitative, linear and *universal correlation* between  $X_{NMR}$  and enzymatic reactivity (as expressed by sugar yields) for all biomass types, solvent systems, and pretreatment temperatures tested in this study. Therefore, the results displayed in Figure 3.9 indicate that for a fixed set of enzymatic hydrolysis conditions,  $X_{NMR}$  serves as a *universal predictor* of the enzymatic digestibility of cellulose following pretreatment in mixtures of water with sulfuric acid and polar aprotic cosolvents. That said, it is possible that different catalysts (such as bases) or cosolvents (such as proton donors or ionic liquids) would effect different



changes, which may or may not improve the enzymatic digestibility of the residual cellulose; only the effects of Brønsted-acid catalysts and polar aprotic cosolvents are studied herein.

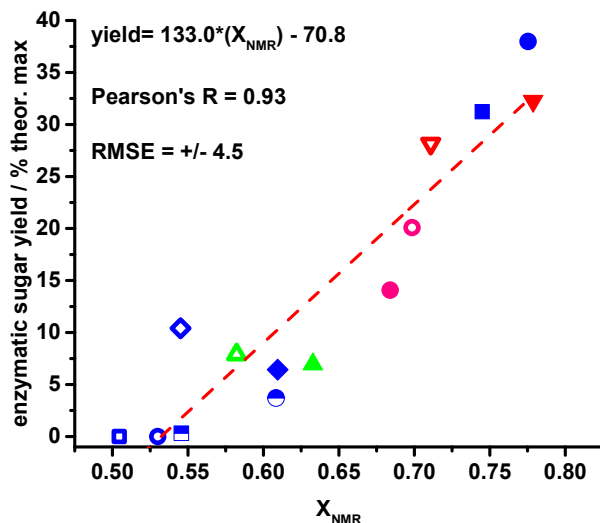
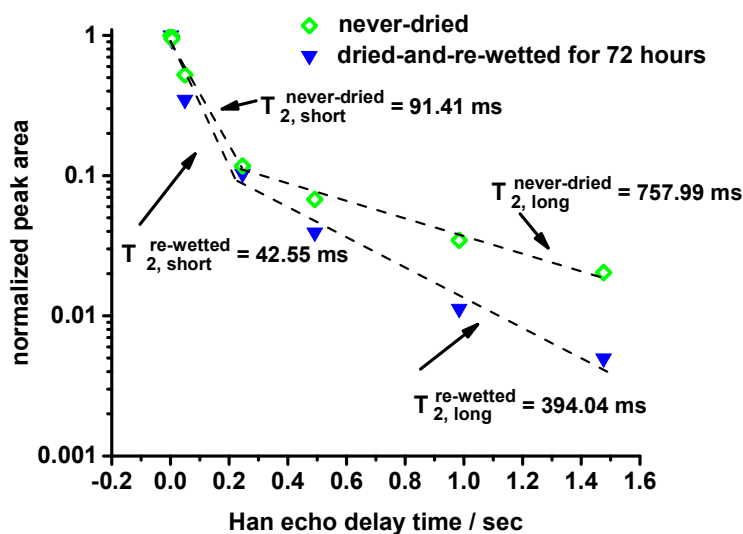


Figure 3.9. enzymatic sugar yields versus  $X_{NMR}$  for all never-dried entries in Table 3.1. Symbols correspond to those datapoints indicated in the legends in Figures 3.7 and 3.8.

#### *The effects of drying on the hydration behavior of GVL-pretreated P39 cellulose*

In contrast to the never-dried celluloses, the dried celluloses in Figure 3.7 exhibit a slight decrease in  $X_{NMR}$ , but systematically lower yields of enzymatic sugars at similar values of  $X_{NMR}$  as compared to the never dried samples. This result indicates that decreasing the moisture content of pretreated cellulose alters its physical or chemical properties, *other* than the fraction of microfibril surface chains liberated from the lignin-hemicellulose matrix, and that these changes correspond to a decrease in enzymatic digestibility. Furthermore, the results displayed in Figure 3.6 indicate that drying GVL-pretreated celluloses alters the pore structure, creating distinct hydration environments as compared to the never-dried material. It has been reported that polymeric materials composed of identical monomer units but synthesized using different methods exhibit similar behavior, whereby different preparation methods produce distinct pore

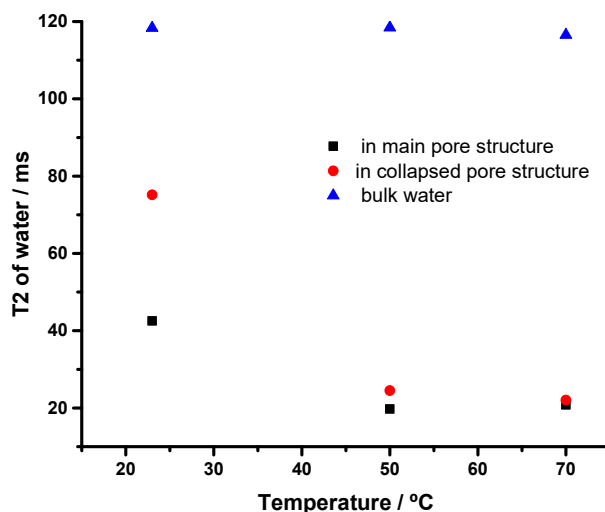
structures, and therefore different hydration environments between materials.<sup>53</sup> These differences in hydration environments can be characterized in terms of the binding strengths of water with the surface accessible facets of the pore structure, the diffusivity of water within the pores, and the extent to which water exchanges readily between the pore structure and the bulk fluid. Changes in these characteristics would suggest changes in the ease with which macromolecules, such as cellulases, diffuse from the bulk solution to the interior pores of the cellulose, explaining differences in reactivity between dried and never-dried cellulose materials, and these changes can be probed using  $^1\text{H}$  HR/MAS NMR.<sup>54-55</sup>



**Figure 3.10.** Normalized, integrated areas of the main hydration resonance at  $\sim 9.45$  ppm in the  $^1\text{H}$  HR-MAS spectra as a function of the number of CPMG rotor cycle delays or, equivalently, the Hahn echo delay time. Integrals shown are for never-dried, and dried-then-re-wetted P39 biomass that was treated with GVL-water solvents to remove hemicellulose and lignin (entries 9 and 10 in Table 3.1, respectively).

Figure 3.10 displays the total integrated area of the  $^1\text{H}$  HR/MAS resonance at  $\sim 9.5$  ppm (which we attribute to the main pore structure of wetted cellulose) as a function of the Hahn Echo delay time in a rotor-synchronized CPMG pulse sequence for GVL-pretreated P39 cellulose. Closed symbols correspond to samples that were never dried, while open symbols

correspond to the same samples after being dried in a vacuum oven overnight, and then re-wetted by soaking in DI water at room temperature for 72 hours. Transverse spin-spin relaxation times ( $T_2$ ) were estimated by fitting the data in Figure 3.8 to a bi-exponential curve (shown in dashed lines). Accordingly, two time constants were assigned to the  $T_2$  relaxation process for the never-dried and the dried-and-re-wetted GVL-pretreated cellulose samples: one long relaxation time on the order of 100 ms ( $T_{2, long}$ ), and one short relaxation time on the order of 10 ms ( $T_{2, short}$ ). These time constants are also displayed in Figure 3.10. Going from the never-dried to the dried-and-re-wetted samples, the time constants associated with the spin-spin-relaxation processes for water within the cellulose pore structure roughly double. Short spin-spin relaxation times have been attributed to water that binds more strongly within a polymer matrix.<sup>56</sup>



**Figure 3.11.**  $T_2$  relaxation time constants for water in the main pore structure of cellulose (9.4 ppm in Figure 3.11) and the “collapsed” pore structure of cellulose that forms upon drying (8.9 ppm in Figure 3.11) as a function of temperature. Spectra were collected with a CPMG filter (Hahn echo delay time) of 2 to 250 ms, and the absolute integrals of the corresponding resonances were analyzed as a function of Hahn echo delay time using a single exponential decay to estimate spin-spin relaxation time constants ( $T_2$  values). The standard error for these six  $T_2$  measurements was  $\pm 4.3\%$ .

We next examined the temperature-dependent spin-spin relaxation behavior of the  $^1\text{H}$  HR/MAS resonances in Figure 3.6 that we attribute bulk water (4.7 ppm), water in the main pore structure (9.4 ppm) and “collapsed” pore structure (8.9 ppm) of dried-and-re-wetted cellulose. These data are displayed in Figure 3.11. Briefly,  $^1\text{H}$  HR-MAS NMR spectra were collected using a CPMG pulse sequence with a Hahn-echo delay between 4 and 250 ms, with the temperature varied between 23 and 70°C. The absolute integrals of the aforementioned resonances were analyzed as a function of Hahn echo delay time using a single exponential decay to estimate spin-spin relaxation time constants ( $T_2$  values; see ESI).

At room temperature, the spin-spin relaxation times ( $T_2$  values) of bulk water, and water in the main and collapsed pore structure of previously dried cellulose are clearly differentiated. If water were able to exchange readily between the cellulose pore structure and the bulk water, then it would be expected that the  $T_2$  values of the corresponding resonances would approach one another with increasing temperature, as the rate of exchange increases.<sup>53</sup> However, we find that the  $T_2$  values for bulk water is constant with temperature. In contrast, the  $T_2$  values for water in the main and collapsed cellulose pore structures are initially differentiated at 23°C, but decrease and approach one another with increasing temperature, becoming indistinguishable within the error (+/- 4.3% or about 1 ms). This behavior indicates that while water may exchange (diffuse) between the main or secondary hydration environments (main pore structure and collapsed pore structures) at these temperatures, water does not exchange readily between the cellulose pore structures and the bulk.

All things considered, we interpret the results displayed in Figures 3.6, 3.7, 3.10 and 3.11 to indicate that upon drying GVL-pretreated cellulose, the cellulose pore structure partially collapses, trapping water in a number of segregated hydration environments wherein water binds

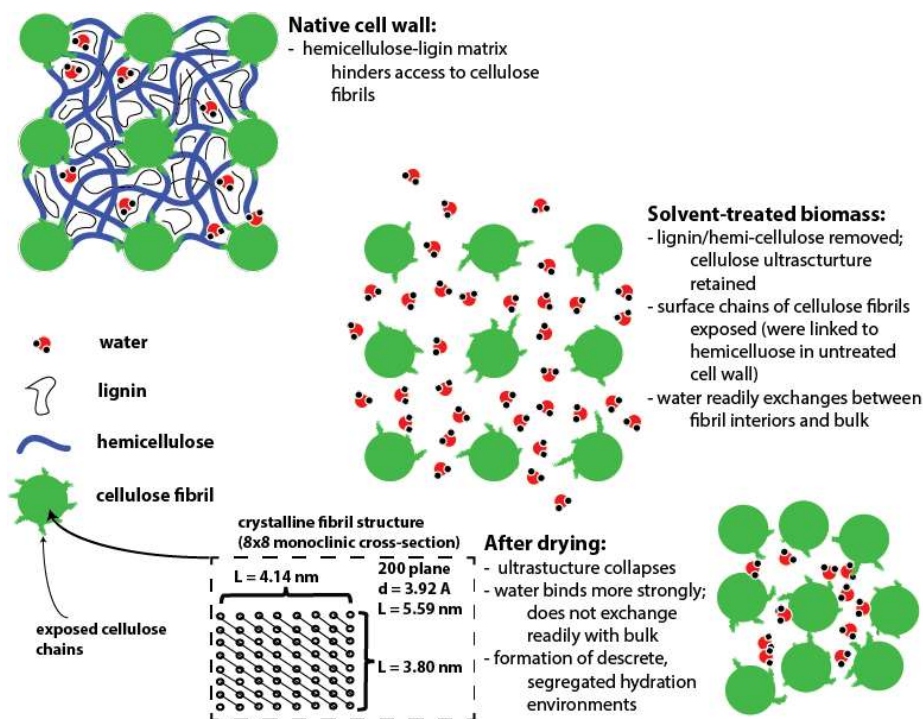
more strongly with the surface-accessible facets of cellulose, and does not exchange readily with the bulk water outside the cellulose pores. We expect that these changes in hydration behavior, as compared to the never-dried cellulose, inhibit the diffusion of cellulases into the pore structure, as well as the binding of the cellulases to the surface accessible facets of the cellulose, limiting enzymatic sugar yields. Therefore, it follows that drying biomass may decrease the enzymatic digestibility of the entrained cellulose in general. This result is of general significance to biomass conversion research, as laboratory samples may be dried prior to enzymatic hydrolysis to determine moisture content, or to increase shelf life.

*Physical interpretations and predicting the effectiveness of solvent-assisted biomass pretreatment strategies using NMR*

We propose in Figure 3.12 a physical interpretation for the key changes effected by solvent-mediated pretreatment of woody biomass over acid catalysts. Note that the proposed microfibril dimensions shown in Figure 3.11 are derived from XRD measurements of P39 poplar, and are in agreement with those proposed elsewhere,<sup>48, 57</sup> see ESI for details. As native lignocellulosic biomass is pretreated in semi-aqueous organic solvent systems (for example GVL-water mixtures) over acid catalysts and/or bleached, lignin, hemicellulose and other extracellular materials are removed from the cell wall. At sufficiently low temperatures and short reaction times, however, the cellulose microfibrils are largely untouched, so that their characteristic dimensions do not change. Furthermore, without a drying step, water occupies the interstitial areas between the remaining microfibrils, so that the pore structure formed by these features in the native cell wall is preserved. The result is a mostly cellulose-containing solid

through which water can easily diffuse and exchange with the bulk liquid, so that hydrolytic enzymes (cellulases) can readily migrate to the pore structure interior.

To facilitate cellulose hydrolysis, enzymes must also bind to the glycosidic linkages exposed on the surface-accessible facets of the cellulose microfibrils, and so the proportion of these linkages that are exposed to the cellulases ultimately determines the enzymatic reactivity of the cellulose. This fraction of exposed linkages is determined by the extent to which the solvent is able to remove not just the bulk extracellular material, but to destroy the linkages between the cellulose and the hemicellulose/lignin matrices. As discussed above, this characteristic feature can be probed using CP/MAS NMR, though the observable  $X_{NMR}$ . Upon drying, the cellulose pore structure collapses, trapping remaining water into a number of segregated hydration environments. The water in these hydration environments binds more strongly than the water present in the never-dried material, so that it does not diffuse as easily, nor exchange as readily with the bulk liquid. This behavior hinders the ability of macromolecules such as enzymes to diffuse into the cellulose pore structure interior, reducing the reactivity of hydrolytic enzymes, and limiting the yields of enzymatic sugars.



**Figure 3.12.** Schematic representation of the structural features of lignocellulosic biomass at various stages of the pretreatment process. Characteristic dimensions of the cellulose microfibrils are for native P39 biomass; values are estimated based on the combined NMR/XRD analysis as described in the ESI.

## Conclusions:

At short reaction times and lower temperatures, pretreatment of lignocellulosic biomass in aqueous mixtures of organic solvents with dilute acid catalysts removes lignin, hemicellulose and extracellular material from the bulk interstitial areas between crystalline cellulose microfibrils in the cell wall. At higher temperatures and longer reaction times, the characteristic dimensions of the crystalline cellulose domains remain unchanged, but the linkages between the exterior of the cellulose microfibrils and the lignin/hemicellulose matrix are broken, exposing surface-accessible cellulose chains on the outsides of the microfibrils. The extent to which these surface-accessible cellulose chains are liberated from the lignin and hemicellulose is observable by solid state NMR, and herein we express these changes in terms of a single, quantitative metric. Changes in this NMR-observable metric as function of biomass pretreatment conditions

(reaction time and temperature) universally predict the yields of enzymatic sugars achieved by hydrolysis of the residual cellulose, regardless of biomass type or the organic pretreatment solvent used. Drying the residual cellulose collected from solvent-pretreated biomass alters the pore structure constituted by the cellulose microfibrils, so that water binds more strongly within the pores, and does not exchange readily with the bulk. This behavior limits the diffusion of enzymes into the cellulose pore structure, limiting enzymatic reactivity of the cellulose accordingly.

The results of this study suggest that solvent-mediated pretreatments (and biomass pretreatment methods in general) that are able to liberate the surface-accessible facets of the cellulose microfibrils from the surrounding lignin and hemicellulose render biomass more readily hydrolysable by enzymes, and likewise improve the yields of enzymatic sugars from biomass. The factors which connect the composition and properties of specific solvent systems to their ability to effect these key changes is the subject of future work. Finally, the moisture content of the pretreated cellulose should be carefully controlled prior to enzymatic hydrolysis, as changes in moisture content alter the structure of the residual cellulose and limit enzymatic reactivity.

Together, these results demonstrate the utility of solid-state and HR/MAS NMR as tools to understand the key chemical and physical changes that occur during liquid-phase conversion of real biomass. The most important corollary is that solid-state NMR can resolve details regarding not just the bulk composition of different biomass types, but also the chemical linkages between distinct phases like lignin, hemicellulose and cellulose. As demonstrated herein, these details can complement those afforded by more standard biomass characterization techniques, and lead to new insights. Moreover, details beyond those generated in this study might be



obtained by bringing more advanced solid-state NMR techniques to bear, such as 2D-dimensional solid-state NMR. We hope that such efforts will represent a greater fraction of published efforts in biomass conversion research, going forward.

## **Experimental:**

### *Materials*

Corn stover, switchgrass, sorghum, and poplars (NM6<sup>58</sup> and P39<sup>59</sup>) were obtained from the Great Lakes Bioenergy Research Center (GLBRC) in the form of debarked, 5 mm chips, and used as received. Water (Fisher, HPLC grade), acetone (Sigma, 98%), ethanol (Fisher, HPLC grade, 200 proof),  $\gamma$ -valerolactone (GVL; Sigma, 95%), tetrahydrofuran (THF; Acros, 99%+ anhydrous with 200 ppm BHT inhibitor), acetonitrile (MeCN; Sigma, 98%), dimethyl sulfoxide (DMSO; Sigma, 98%), and *N*-methyl pyrrolidinone (NMP; Sigma, 99%) were obtained from vendors and used as received. Sulfuric acid (98 wt%), sodium hydroxide, sodium chlorite and acetic acid were obtained from Sigma and used as received.  $\beta$ -*D*-glucose, xylose, furfural, and 5-hydroxymethylfurfural (Sigma-Aldrich ACS reagent grade) were obtained from vendors, and used as calibration standards.

CelR\_cbm3a was created through fusion of the Hungateiclostridium thermocellum CelR enzyme domain (GenBank accession code CAE51308.1 residues 31 – 642, i.e., GH9 and CBM3c modules only) with an additional 153-residue CBM3a from the scaffolding protein CipA, connected via the 48-residue CipA-native linker sequence. The lac operon-controlled expression plasmid was shipped to a third-party vendor, where it was expressed and purified at scale.

### *Sample prep and solvent-mediated pretreatment of biomass over sulfuric acid*

Briefly, solid biomass samples were charged to sealed glass reactors containing a described amount of an appropriate solvent system (e.g., pure water or 90/10 GVL/water on a mass basis) along with a sulfuric acid catalyst or appropriate bleaching agent (e.g., NaClO<sub>2</sub>). Reactors were placed in an oil bath at the desired temperature (e.g., 100 °C) and aggerated with a magnetic stir bar. After an appropriate period of time, reactors were quenched in an ice bath, and the liquid and solid contents were separated with filtration, washed (if needed), and characterized using the techniques described elsewhere in this report (e.g., HPLC, NMR, etc.). See ESI for details.

*<sup>1</sup>H/<sup>13</sup>C Proton spin-relaxation (T<sub>1ρ</sub>) edited cross-polarization magic-angle-spinning (PSRE CP/MAS) NMR*

About 200 mg of solid sample was packed into 4-mm thin-walled silicon nitride rotors and sealed with glass-filled Torlon caps. <sup>1</sup>H/<sup>13</sup>C PSRE CP/MAS NMR spectra were acquired on a Bruker Avance III 500 MHz spectrometer with a proton radio frequency of 500.22 MHz and a <sup>13</sup>C radio frequency of 125.76 MHz. Each rotor was spun at 4 kHz in a 4 mm Doty Scientific MAS probe. <sup>1</sup>H and <sup>13</sup>C 90° pulse lengths of 2.30 and 4.55 μs were used, respectively. Spectra shown represent 2056 signal averages with a 0.3 second acquisition time and an 8 second recycle delay between scans. The spin-lock radiofrequency strength was  $\gamma B_H / (2\pi) = 46$  kHz during spin-locking with a spin-lock time of 3-10000 μs, as described below. Cross-polarization radiofrequency strength was 99.53 kHz (maximum) during cross-polarization with a 70-100% ramp and a contact time of 2000 μs. The decoupler radiofrequency was 100 kHz during acquisition. <sup>13</sup>C spectra were secondary referenced to the up field adamantane peak at 28.7 ppm referenced to TMS. See ESI for full details and a more complete description of this technique.

### *Attenuated total reflectance infrared (ATR-FTIR) spectroscopy*

Attenuated total reflection–Fourier transform infrared spectroscopy (ATR–FTIR) was conducted using a Bruker Optics Vertex system with a diamond-germanium ATR single reflection crystal. Untreated and solvent-pretreated biomass samples were dried in a vacuum oven overnight to remove water content prior to analysis, and were pressed uniformly against the diamond surface using a spring-loaded anvil. Sample spectra were obtained in triplicates using an average of 128 scans over the range between  $400\text{ cm}^{-1}$  and  $4000\text{ cm}^{-1}$  with a spectral resolution of  $2\text{ cm}^{-1}$ . Air was used as background.

### *X-ray diffraction characterizations of native and pretreated cellulose*

Cellulose microfibrils scatter incident X-rays to produce a diffraction pattern consistent with a monoclinic unit cell.<sup>60</sup> Accordingly, X-ray diffractograms were collected for the native P39 and solvent-pretreated biomass samples, using a Bruker D8 Discovery diffractometer with a  $\text{Cu } K_{\alpha}$  X-ray source operating at 1000 kV and 100 mA, with a 5-mm aperture and 600 sec exposure time. Biomass samples were analyzed without additional grinding to reduce particle size. A separate sample of boron nitride was analyzed to assess the inherent line broadening of the instrument.

### *Enzymatic hydrolysis of residual cellulose*

Hydrolysis reactions were prepared with 0.5 mg/mL pretreated cellulose and 0.05 mg/mL CelR in 0.1M phosphate buffer, pH 6.0 to a final volume of 1 mL. Reactions were incubated in a Heidolph Titramax 1000, with a Heidolph Inkubator 1000 used to control temperature. Reactions

were run for 24 hours at 50 °C with 1050 rpm (0.0185 x g) shaking. After incubation, reactions were centrifuged for 5 minutes at 21,130 x g. The concentration of soluble sugar produced was determined via the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Briefly, 100 µL of working solution and 5 µL of supernatant were heated for 15 minutes at 80 °C. Control experiments used were CelR only, cellulose only, and buffer only samples, and a glucose standard curve was used for converting absorbance units to mg/mL glucose in solution. Four reactions were performed for each condition, with supernatant being sampled in triplicate.

### *<sup>1</sup>H High-Resolution (HR)/MAS NMR spectroscopy*

<sup>1</sup>H HR/MAS NMR experiments were performed on a Bruker Avance 500 MHz spectrometer with a proton radiofrequency of 500.22 MHz equipped with a 4 mm Doty Scientific MAS NMR probe. Proton 90° pulses were applied with a pulse length of 2.48 µs with a 0.5 second acquisition time and a 10 second recycle delay. Solvent-pretreated biomass samples of about 70 wt% moisture content were sealed into 30 µL Kel-F HR/MAS rotor inserts (Bruker Biospin Inc.), with the MAS experiments performed at 4 kHz with a ±0.1 K temperature regulation between 296 and 343 K. Dried, solvent-pretreated biomass samples were re-wetted by soaking in an amount of water corresponding to the moisture content of the never-dried samples (typically ~70 wt% water to 30 wt% solids) for 72 hours, and then loaded into 30 µL Kel-F HR/MAS rotor inserts. The sample temperatures under HR/MAS conditions ( $T_{\text{MAS}}$ ) were calibrated using a neat ethylene glycol thermometer. The one-dimensional <sup>1</sup>H NMR spectra were obtained using a single pulse Bloch decay sequence with 16 scan averages and 10 seconds recycle delays. Transverse spin–spin relaxation times ( $T_2$ ) were found using a rotor-synchronized CPMG (Carr–Purcell–Meiboom–Gill) pulse sequence with 16 signal averages, and inter-pulse

delays varied between 10 and 20000 rotor cycles (corresponding to 0.03 to 8 seconds transverse spin-spin relaxation times). The  $^1\text{H}$  NMR chemical shifts were referenced to the internal standard of neat water,  $\delta = +4.7$  ppm at 298 K with respect to the chemical shift of TMS,  $\delta = 0$  ppm.

### **Acknowledgements:**

This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494.

### **Supporting Information:**

#### **Methods and Analytical details:**

##### *Sample Preparation*

All reaction conditions are summarized in Table 3.1 of the main text; as an example, we describe the sample preparation using Entry 9. For a typical experiment, 6 grams of as-received 5-mm biomass chips (P39 poplar) were loaded into a 60- mL pressure-sealed glass reactor equipped with a magnetic stir bar. 60 grams of an appropriate solvent system (80 wt% GVL with 20 wt% water) containing 100 mM sulfuric acid were then added, and the glass reactor was sealed with a screw-top cap equipped with a Teflon O-ring. The glass reactor was next submerged in a silicon oil bath at the desired temperature (140°C). The reactor contents were agitated by rotating the stir bar at 500 rpm. The reactor was held at the desired temperature for a period of time (one hour), and then quenched in an ice bath at 0°C for 15 minutes. The reactor contents were then removed, and the liquid and solid fractions were separated by vacuum filtration with a fine-mesh filter paper and a Buchner funnel connected to a house vacuum system at 0.5 atm.

The solid fraction from the reactor was washed with 3 volumes of HPLC grade water, and dried with vacuum filtration to produce a solid filter cake of about 70 wt% moisture content.

The bleaching step, when applied, first consisted of stirring the solids at room temperature for three hours in a solvent system consisting of water with 1 wt% dissolved  $\text{NaClO}_2$  and 0.33 mL acetic acid per gram of  $\text{NaClO}_2$ . The resulting solids were then separated with vacuum filtration, and stirred for an additional 4 hours at  $50^\circ\text{C}$  in a solvent system consisting of water with pH adjusted to 11 using sodium hydroxide (1.0 mM NaOH). The solids/liquids ratio in both bleaching steps was fixed at 3/97 by weight. The resulting solids from the second bleaching step were then washed with three volumes of HPLC grade water, and dried using vacuum filtration for one hour to produce a solid filter cake of about 60 wt% moisture content. When applied, a drying step consisted of placing the washed, filtered solids from the solvent pretreatment or bleaching steps in a vacuum oven at  $85^\circ\text{C}$  and 0.5 atm overnight, and taking the difference in mass between the dried and never-dried material as the moisture content of the sample.

The liquid fraction from the reactor was collected, diluted with water ten times by weight to precipitate the GVL-soluble lignin, and then vacuum filtered again to remove the solid lignin. The lignin was vacuum treated overnight, and weighed to determine the lignin yield. The diluted, lignin-free liquid fraction was next filtered with a PTFE 0.2-micron syringe filter, and analyzed for soluble sugar and furan content using a Shimadzu-1020 HPLC equipped with a proton-based ion exclusion column (BioRad HPX-87H), a refractive index detector and a UV-vis scanning photo diode array. The mobile phase used was 5 mM sulfuric acid, with a flow rate of 0.6 mL/min at  $30^\circ\text{C}$ . The glucose, xylose, furfural and 5-hydroxymethyl furfural content of the sub-micron filtered, lignin-free liquids was then determined using calibration curves and external standards. The glucan and xylan oligomer content of the filtered, lignin-free liquids was then determined by adding 3 wt%  $\text{H}_2\text{SO}_4$ , stirring at  $120^\circ\text{C}$  for one hour, and then re-analyzing the samples using HPLC. The difference in glucose and xylose content between the hydrolyzed and

non-hydrolyzed samples was taken to determine the soluble glucan and xylan oligomer content, respectively. Note that the xylan and glucan oligomer yields, plus the small quantity of sugar degradation products (furfural and 5-hydroxymethylfurfural) detected were added to the corresponding sugar yields (xylose and glucose, respectively) to determine total soluble carbohydrate yield.

The moisture content of the untreated biomass was determined by drying the as-received samples in a vacuum oven overnight, as described above. The extractives content (waxes, proteins, minerals, etc.) of the untreated biomasses was determined by drying the as-received biomasses overnight, and then processing the dried samples under reflux in a Soxhlet extractor using water, then ethanol, and then acetone as extraction solvents, with each step being carried out at the boiling point of the corresponding extraction solvent. After correcting for the original moisture content, the wt% of total extractives was determined by taking the mass difference between the dried and extracted biomass samples.

#### *Proton spin-relaxation-edited $^1\text{H}/^{13}\text{C}$ cross-polarization magic-angle-spinning (PSRE CP/MAS)*

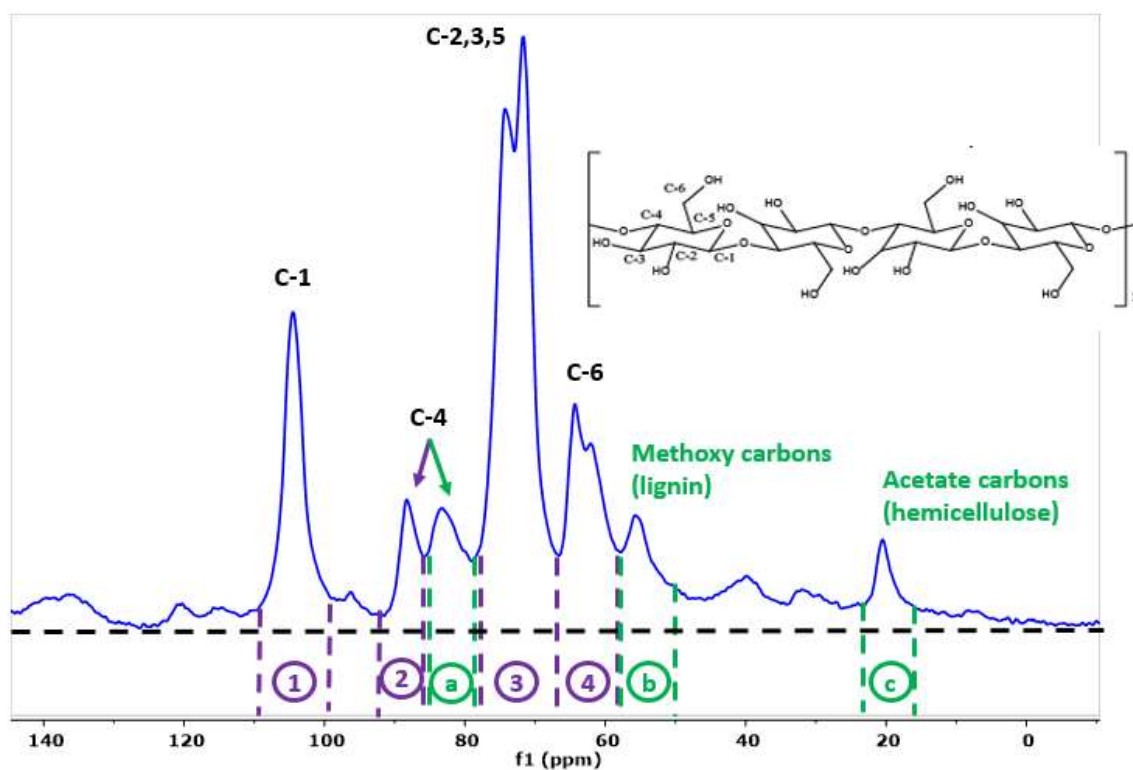
##### *NMR of cellulose*

Briefly,  $^{13}\text{C}$  CP/MAS NMR spectra are collected with a spin-lock delay applied prior to cross polarization. During spin-locking, the  $^1\text{H}$  net magnetic vector is first oriented along the transverse plane by application of a standard  $90^\circ$  pulse, and is then held in the transverse plane by the sustained application of a rotating magnetic field. As longer spin-lock delays are applied, magnetization decays in the  $^1\text{H}$  rotating frame, so that less magnetization is transferred to the carbon nuclei during cross-polarization. The result is collection of various CP/MAS spectra which, in order of increasing spin-lock delays, are attenuated to an increasing extent.

Importantly, however, the proton radio frequency during spin-locking is optimized so as to obtain the greatest contrast in rotating-frame relaxation time constants ( $T_{1\rho}$ 's) between crystalline cellulose and the amorphous materials in the sample. That is, protons associated with crystalline materials lose magnetization in the transverse plane either faster or slower than amorphous materials during spin-locking. Newman and co-workers have shown that for cellulosic biomass, this contrast is optimized at a spin-lock frequency of 46 kHz, which is used herein, so that crystalline cellulose generally has about a 2-3 times longer  $T_{1\rho}$  value than amorphous cellulose, hemicellulose or lignin.<sup>48</sup>

Figure S3.1 displays a normal CP/MAS spectrum (no spin-lock delay) for native P30 poplar biomass. The main resonances corresponding to the carbon nuclei in crystalline cellulose are labeled.<sup>46</sup> A glucan tetramer unit is displayed in the inset for reference. Also featured in the CP/MAS spectrum in Figure S3.1 are resonances corresponding to methoxy moieties in the lignin fraction of the biomass, and acetate moieties in hemicellulose. The proton spin-relaxation-editing (PSRE) method developed by Newman and co-workers<sup>48</sup> is calibrated by first measuring the transverse relaxation rates ( $T_{1\rho}$ 's) for the crystalline cellulose and non-crystalline fractions of the biomass by varying the spin-lock delay time in a CP/MAS pulse sequence, as described above.

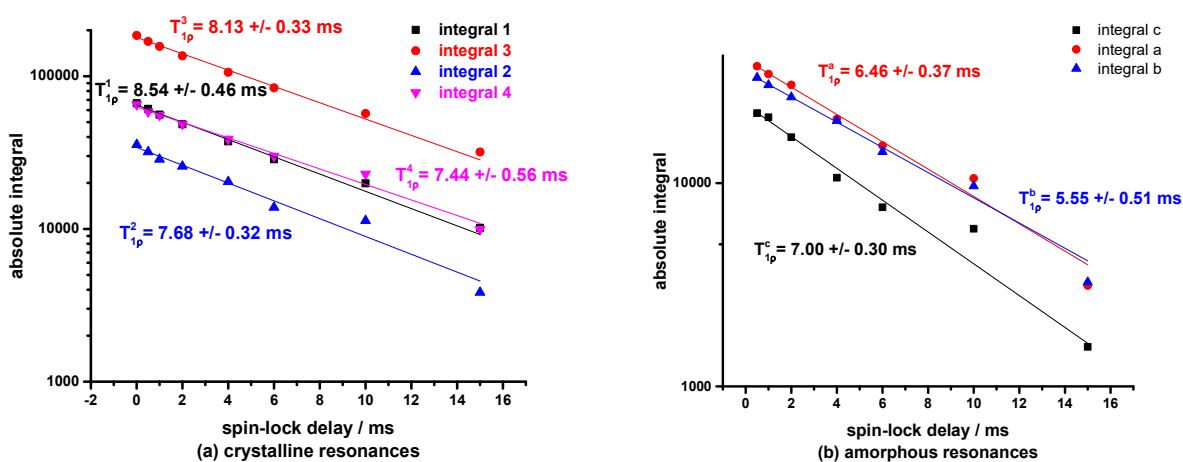




**Figure S3.1.** Normal CP/MAS spectrum for non-pretreated P39 poplar biomass. Key resonances and integral regions are denoted.  $^1\text{H}$  and  $^{13}\text{C}$   $90^\circ$  pulse lengths of 2.30 and 4.55  $\mu\text{s}$  were used, respectively. 2056 signal averages were used with a cross polarization contact time of 2000  $\mu\text{s}$ , using a 70-100% ramped cross polarization pulse with the proton and carbon radio frequencies matched at 99.53 kHz (max.). MAS speed was 4 kHz. The decoupler radiofrequency was 100 kHz during acquisition.  $^{13}\text{C}$  spectra were secondary referenced to the up field adamantane peak at 28.7 ppm referenced to TMS.

To this end, as shown in Figure S3.1, four integral regions are delineated corresponding to **crystalline cellulose resonances (regions 1-4)**, and three to the **non-crystalline materials (regions a, b and c)**. Region (a) in Figure S3.1 corresponds to amorphous cellulose, although crystalline cellulose signal overlaps the amorphous signal in this region. Region (b) corresponds to amorphous lignin, although signal corresponding to lignin-cellulose linkages with longer  $T_{1,\rho}$ 's likely overlap the amorphous lignin in this region. Region (c) corresponds to amorphous hemicellulose, and is reasonable well separated from overlapping signals. Note that the entire spectrum displays broad features corresponding to ill-defined amorphous materials in the biomass, which appears to distort the base line, though the spectrum has been appropriately

baseline-corrected and phased. Finally, note that these aforementioned broad, amorphous features and the general multiplicity of resonances makes peak deconvolution or line fitting problematic for these real biomass samples. Accordingly, following the method of Newman and others, we integrate the total signal in the regions defined in Figure S3.1, without fitting peaks individually.



**Figure S3.2.** Integrated signal for the regions denoted in Figure S3.1 as a function of spin-lock delay. Spin locking was performed at 46 kHz with a spin-lock time of 3-15000  $\mu$ s. Otherwise, acquisition parameters are the same as those denoted in Figure S3.1.

Figure S3.2 displays the integrated signal for the regions denoted in Figure S3.1 as a function of the spin-lock delay at a spin-locking radiofrequency of 46 kHz. Individually, each integrated region as a function of spin-lock time fits a first-order exponential decay. For the crystalline resonances (Figure S3.2(a)), the average characteristic time constant that describes this process ( $T_{1,\rho}$ ) is  $7.94 \pm 0.42$  ms. For the amorphous materials (Figure S3.2(b)),  $T_{1,\rho}$ 's vary between 5.5 and 7.00 ms with an average of  $6.34 \pm 0.60$  ms. In general, the  $T_{1,\rho}$ 's corresponding to the amorphous materials are shorter than those corresponding to crystalline cellulose, as expected. However, the variance between  $T_{1,\rho}$ 's for distinct amorphous resonances

(+/- 10%) is greater than that between  $T_{1,\rho}$ 's for distinct crystalline cellulose resonances (+/- 5%).

This behavior is attributable to the fact that the distinct amorphous resonances are associated with different amorphous materials, with likewise different  $T_{1,\rho}$ 's, whereas the crystalline resonances are associated with a single phase (crystalline cellulose microfibrils).

Finally, we produce a proton spin-relaxation-edited (PSRE) CP/MAS subspectra that displays only  $^{13}\text{C}$  signal corresponding to crystalline or amorphous materials in the biomass being analyzed. We do this by assuming that the total spectrum ( $\mathbf{S}$ ) obtained with a short (0.003 ms) spin-lock delay is a linear combination of signal from the crystalline ( $\mathbf{C}$ ) and amorphous ( $\mathbf{A}$ ) materials in the biomass:

$$S = C + A \quad (\text{S1})$$

Similarly, we assume that the attenuated spectrum ( $\mathbf{S}'$ ) obtained with a longer (e.g. 10 ms) spin-lock delay is a combination of signal from the crystalline and amorphous fractions in the biomass sample, but that signal from each phase has been reduced by a fraction ( $f_C$  or  $f_A$ ) by the application of the spin-lock delay:

$$S' = f_C C + f_A A \quad (\text{S2})$$

$f_C$  and  $f_A$  can be calculated as a function of spin-lock time ( $t_{SL}$ ) using Figure S3.2, where:

$$f_C = e^{-\frac{t_{SL}}{T_{1\rho}^{crystalline}}} \quad (\text{S3})$$

And

$$f_A = e^{-\frac{t_{SL}}{T_{1\rho}^{amorphous}}} \quad (\text{S4})$$

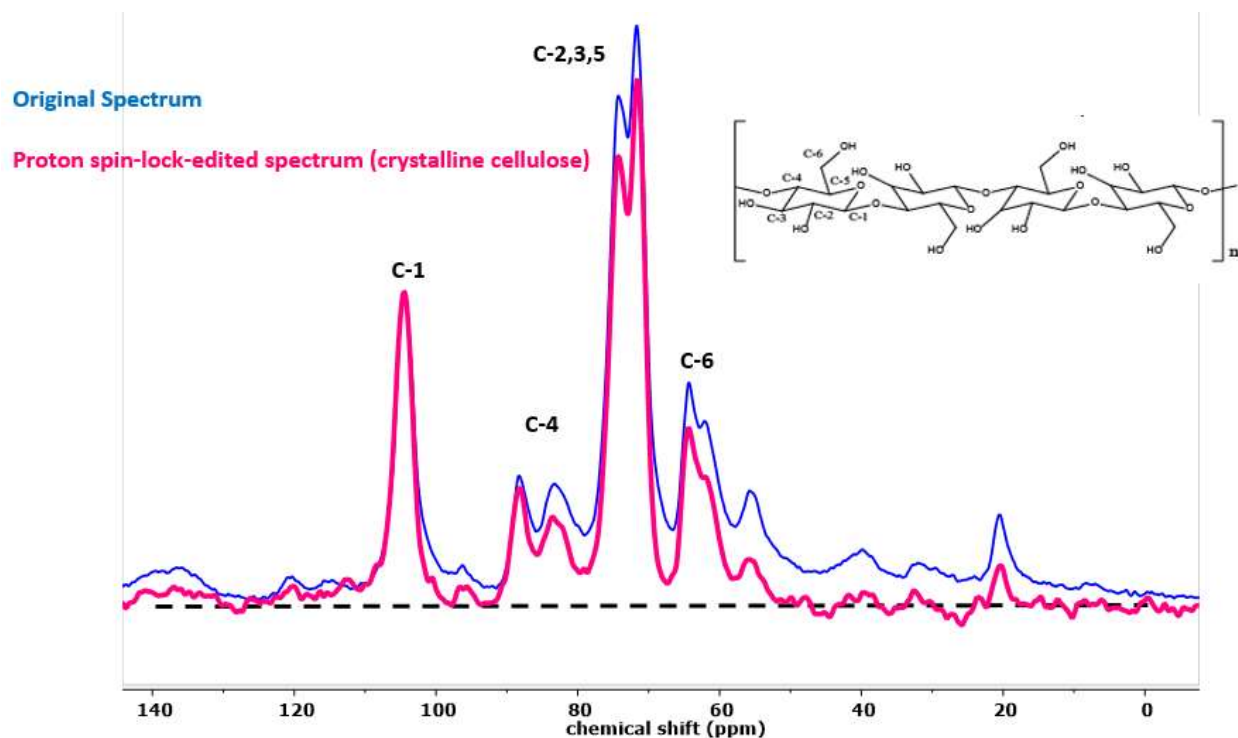
These relationships can be expressed as a matrix operation:

$$\begin{bmatrix} 1 & 1 \\ f_C & f_A \end{bmatrix} \begin{bmatrix} C \\ A \end{bmatrix} = \begin{bmatrix} S \\ S' \end{bmatrix} \quad (\text{S5})$$

And the vector (C, A) is solved for to reproduce the sub spectra corresponding to crystalline cellulose and the amorphous materials in the solid biomass sample.

Importantly, to facilitate the PSR-editing procedure described above, a representative timescale ( $T_{1,\rho}$ ) must be chosen for crystalline cellulose and for the amorphous “phase” as a whole, even though the latter consists of multiple distinct materials with a range of corresponding  $T_{1,\rho}$ 's. The PSRE method best differentiates the crystalline cellulose phase from other components in the biomass when the two materials have the greatest contrast in  $T_{1,\rho}$ 's. Following this insight, and in accordance with other reports,<sup>48</sup> we select the methoxy lignin peak at 55.8 ppm to represent the amorphous, non-cellulosic materials in the biomass, and calculate  $f_C$  in Equation S3 using the  $T_{1,\rho}$  associated with this resonance.

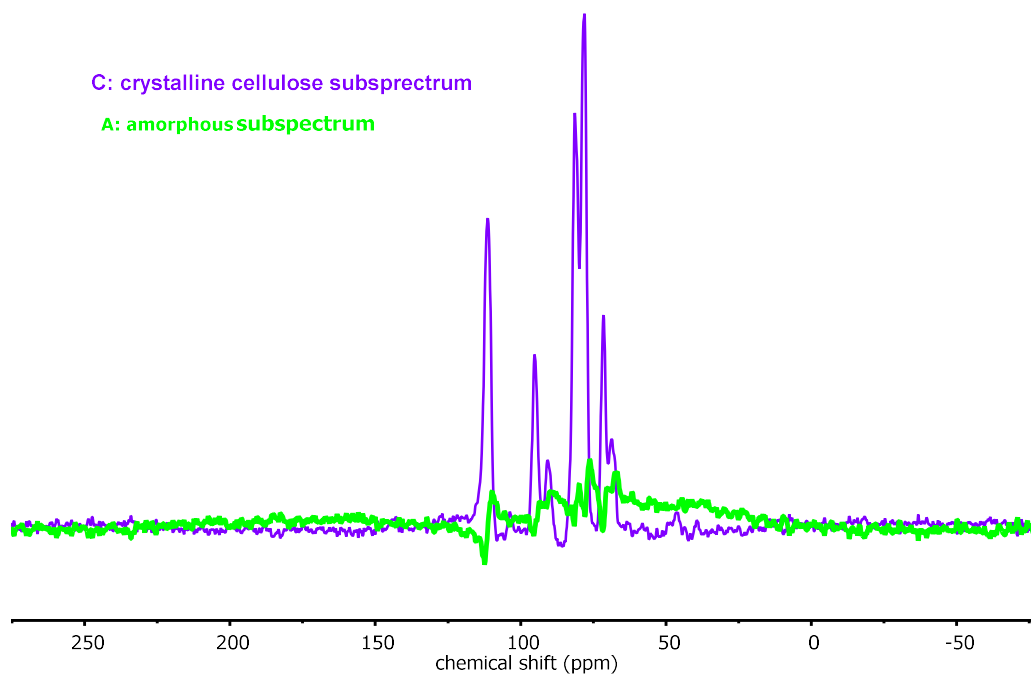
Figure S3.3 displays the normal and PSRE-CP/MAS spectra for native P39 biomass. The PSRE spectrum corresponds to the crystalline cellulose signal only. Note how the heights and areas of the C-1 and downfield C-4 peak (crystalline cellulose) are conserved between the normal and PSRE CP/MAS spectra, whereas a portion of the signal has been subtracted from other regions, due to the signal from amorphous materials being filtered out. Note that the S/N ratio has been reduced comparing the PSRE spectrum to the normal CP/MAS spectrum, but the baseline in the PSRE spectrum is flat due to the lack of signal from broad, amorphous features.



**Figure S3.3.** Normal (blue) and PSRE CP/MAS spectra (pink) corresponding to native (non-pretreated) P39 biomass and the crystalline cellulose therein, respectively. The PSRE spectrum was estimated based on the procedure described above and in the main text.  $T_{1,\rho}$ 's for crystalline cellulose and the amorphous phase were set at 7.94 and 5.55 ms, respectively.

Finally, Figure S3.4 displays representative sub-spectra for the crystalline and amorphous materials in pretreated cellulose. The pretreatment procedure is described in Entry 9 and Table 3.1 in the main text. In Figure S3.4, a chemical shift range from -75 to 275 ppm to show that, after the PSRE procedure described above, the resulting sub-spectra each retain phasing and a baseline that goes to zero in the limits of the observed sweep width. Comparing the sub-spectra for the crystalline cellulose (purple) and amorphous biomass material (green), it is apparent that the crystalline phase is characterized by sharp spectral features and a flat baseline. In contrast, the amorphous phase is characterized by broad features that are slightly offset compared to their crystalline counterparts. This behavior is characteristic of semi-crystalline polymeric materials. In

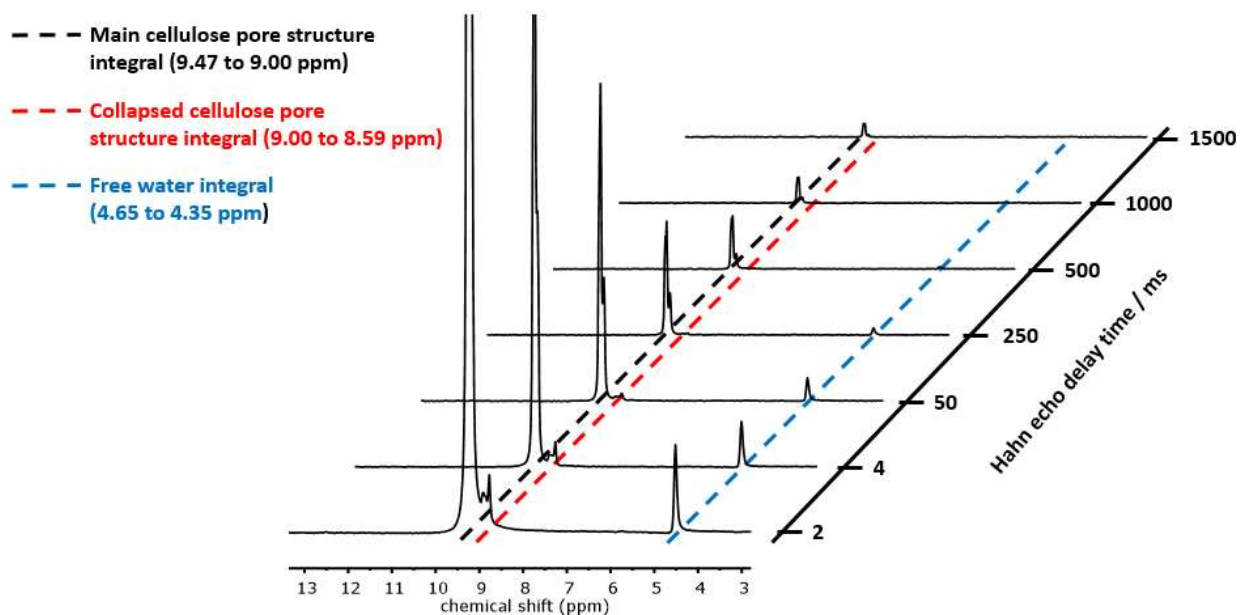
particular, note the large amorphous signal that overlaps the C-4 region of the crystalline cellulose sub spectrum.



**Figure S3.4.** Proton spin-relaxation-edited (PSRE) CP/MAS sub spectra corresponding to crystalline cellulose (purple) and amorphous material (green) in the cellulose-rich material obtained from P39 biomass as described in Entry 9 of Table 3.1 of the main text.

### *<sup>1</sup>H high-resolution (HR)/MAS NMR of cellulose*

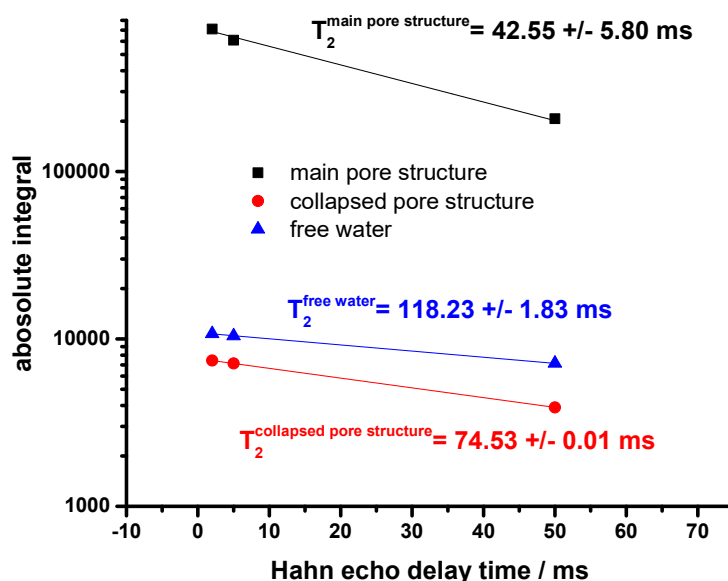
Figure S3.5 displays the HR/MAS spectra for the “Step 2 solids” from Figure 3.2 (or Table 3.1, Entry 9) of the main text, collected after being dried overnight, then rewetted in 30/70 wt/wt solids/water for 72 hours. Three integral regions are denoted, corresponding to the main pore structure, the “Collapsed” pore structure, and the “free water” in the cellulose, as described in the main text. Each spectra in Figure S3.5 is collected at room temperature and at a different Hahn-echo delay time in a rotor-synchronized CPMG pulse sequence, with signal reduction as a function of increasing Hahn-echo delay time resulting from loss of transverse plane magnetization due to spin-spin relaxation.



**Figure S3.5.** HR/MAS NMR spectra with a varying Hahn-echo delays applied (or equivalently varying single-rotor-cycle interpulse delays) in a rotor-synchronized Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence. Spectra correspond to cellulose from P39 biomass pretreated in 90 wt% GVL at 100°C for one hour, then 80 wt% GVL at 140°C for one hour, then dried overnight and re-wetted in 30/70 solids/water by weight for 72 hours (Table 3.1, Entry 10 of the main text). The one-dimensional  $^1\text{H}$  NMR spectra were obtained using a single pulse Bloch decay sequence with 16 scan averages and 10 seconds recycle delays. Proton  $90^\circ$  pulses were applied with a pulse length of 2.48  $\mu\text{s}$  with a 0.5 second acquisition time and a 10 second recycle delay. The spinning rate was 4 kHz.

As described in the main text, the integrated area of the three resonances shown in Figure S3.4 as a function of the Hahn-echo delay time fit a bi-exponential curve with two corresponding  $T_2$  time constants for each resonance. For each of the three integral regions in Figure S3.4, the shorter of these two time constants corresponds to water more tightly bound within the pore structure of the cellulose. Herein, we are interested in how changes in the cellulose pore structure upon drying affect the mobility of water and therefore, presumably, the ease with which macromolecules such as enzymes can diffuse from the bulk water to the cellulose pore interiors. Accordingly, we fit the integrated regions in Figure S3.5 to a single exponential decay within the short range of Hahn-echo delay times (2-50 ms) to capture the initial loss of signal corresponding to these fast-relaxing protons. Figure S3.6 displays these results. The  $T_2$  time constants denoted

for each region (free water, water in the main pore structure and water within the “collapsed” pore structure) are the T<sub>2</sub> time constants shown in Figure 3.15 of the main text. This procedure (collecting HR/MAS at and recording integrals as a function of the Hahn-echo delay time within the range of 2-50 ms) is repeated at 50 and 70°C, and the resulting T<sub>2</sub>'s for each region are calculated to construct Figure 3.16 in the main text.



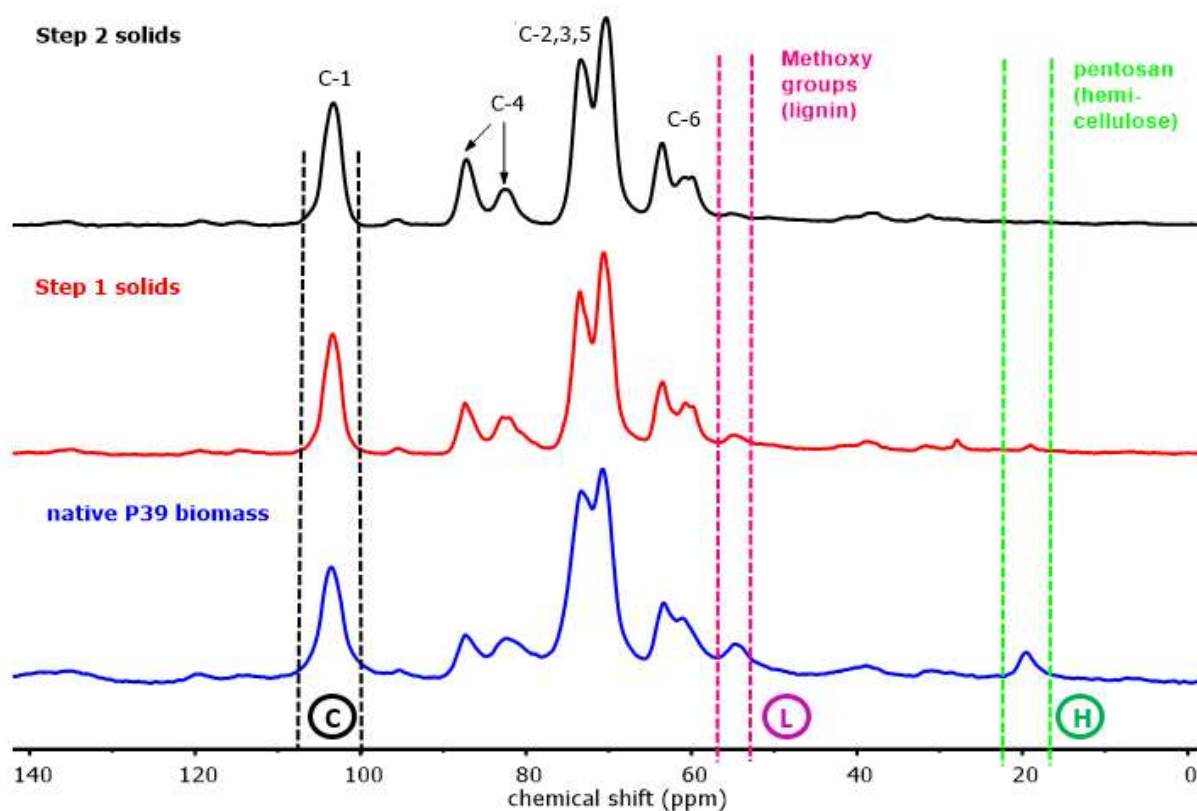
**Figure S3.6.** Integrated are for the regions denoted in Figure S3.5 as a function of the Hahn-echo delay time in a rotor-synchronized CPMAS pulse sequence at room temperature. The one-dimensional <sup>1</sup>H NMR spectra were obtained using a single pulse Bloch decay sequence with 16 scan averages and 10 seconds recycle delays. Proton 90° pulses were applied with a pulse length of 2.48 μs with a 0.5 second acquisition time and a 10 second recycle delay. The spinning rate was 4 kHz.

#### *Native biomass assay by CP/MAS NMR*

To calculate yields and mass balances of solid, non-soluble materials (*e.g.*, those shown in Figure 3.2 of the main text) after GVL-pretreatment, normal CP/MAS spectra of native and GVL-pretreated biomass were analyzed as described below. First, a normal CP/MAS spectrum of the dry, extractive free biomass (prepared as per methods described in the main text) was



collected. The regions corresponding to lignin (**L**), hemicellulose (**H**), and the C-1 resonance of crystalline cellulose (**C**) (denoted in Figure S3.7 below) were integrated, and these integrals were recorded (Table S3.1).



**Figure S3.7.** Normal CP/MAS spectra for native P39 poplar and the same biomass pretreated in GVL-solvent mixtures to remove lignin (Step 1 solids) and sugars (Step 2 solids) as referred to in Figure 3.2 of the main text.  $^1\text{H}$  and  $^{13}\text{C}$   $90^\circ$  pulse lengths of 2.30 and 4.55  $\mu\text{s}$  were used, respectively 2056 signal averages were used with a cross polarization contact time of 2000  $\mu\text{s}$ , using a 70-100% ramped cross polarization pulse with the proton and carbon radio frequencies matched at 99.53 kHz (max.). MAS speed was 4 kHz. Proton decoupler radio frequency was 100 kHz during acquisition.  $^{13}\text{C}$  spectra were referenced to the up field adamantane peak at 28.7 ppm in a separate standard.

Next, published reports were referenced to determine the actual wt% composition of lignin, hemicellulose and cellulose of the corresponding dry, extractive-free biomass type. In Figure S3.7, the example biomass type is P39 poplar, which is assumed to be identical to NM6 poplar.<sup>61</sup> Next, we assumed that the integrated signals from each of the three regions denoted in

Figure S3.7 are representative of the actual amount of total cellulose, hemicellulose and lignin in the native (non-GVL-pretreated) biomass. Under this assumption, the integrated signals ( $I_i$ ) from each of the ( $i$ ) regions corresponding to C-1 cellulose, lignin and hemicellulose can be scaled by a factor ( $f_i$ ), so that the wt% composition of lignin, hemicellulose and cellulose in the dry, extractive free biomass is expressible by the relationships denoted in Equations S6-S8:

$$\frac{f_{C-1}I_{C-1}}{f_{C-1}I_{C-1}+f_L I_L+f_H I_H} = \text{wt\% cellulose} \quad (\text{S6})$$

$$\frac{f_L I_L}{f_{C-1}I_{C-1}+f_L I_L+f_H I_H} = \text{wt\% lignin} \quad (\text{S7})$$

$$\frac{f_H I_H}{f_{C-1}I_{C-1}+f_L I_L+f_H I_H} = \text{wt\% hemicellulose} \quad (\text{S8})$$

The scaling factors  $f_{C-1}$ ,  $f_L$ , and  $f_H$  are found by setting the wt% of lignin, cellulose and hemicellulose equal to known values for each biomass type, and using the integrated CP/MAS signals from the regions  $I_{C-1}$ ,  $I_L$ , and  $I_H$  to solve Equations S6-S8 simultaneously. Once the factors  $f_{C-1}$ ,  $f_L$ , and  $f_H$  are known, the integrated signals  $I_{C-1}$ ,  $I_L$ , and  $I_H$  derived from CP/MAS spectra of GVL-pretreated biomass can be used to estimate the wt% composition of cellulose, hemicellulose and lignin in each sample. Exemplary results following from this procedure are also displayed for P39 biomass in Table S3.1. A similar procedure was performed for all entries in Table 3.1 in the main text, and the wt% cellulose of each sample was used to calculate the yield of enzymatic sugars ( $Y_{enzymatic\ sugar}$ ) using Equation S9:

$$Y_{enzymatic\ sugar} = \frac{C_{reducing\ end} * V * MM_{glucose}}{M_{biomass} * m_{cellulose}} \quad (\text{S9})$$

Where  $C_{reducing\ end}$  is the molar concentration of soluble reducing end sugars (as estimated by methods described in main text, and assumed to be equal to the concentration of monomeric glucose in solution) after 24 hours incubation with enzyme CelR,  $V$  is the total reaction volume,  $MM_{glucose}$  is the molar mass of glucose,  $M_{biomass}$  is the initial mass of biomass charged to the

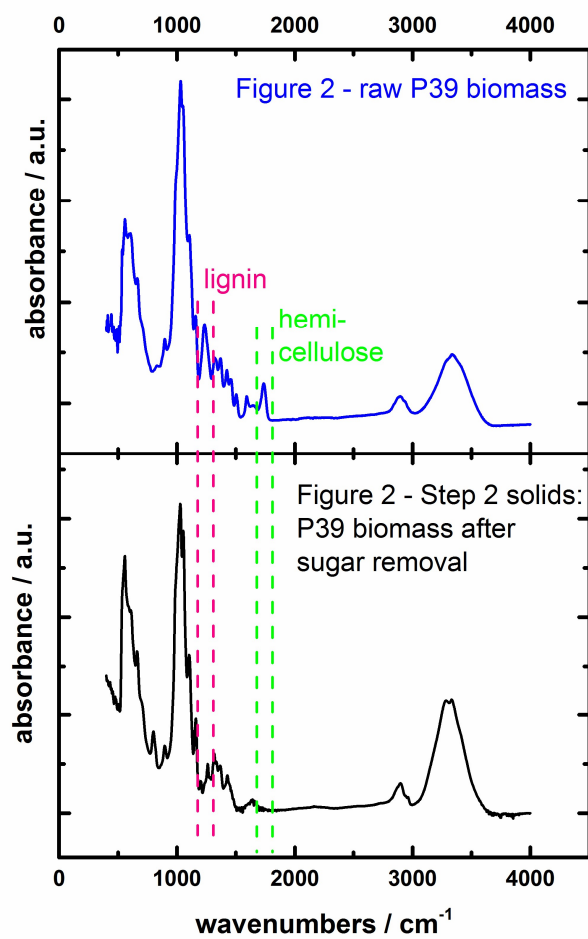
reactor, and  $m_{cellulose}$  is the mass fraction of cellulose (wt% cellulose in Table S3.1 divided by 100%).

**Table S3.1.** CP/MAS integrated signals for the C-1 cellulose, methoxy-lignin and acetate-hemicellulose resonances in Figure S3.6, along with scaling factors ( $f_i$ 's) and wt% compositions of cellulose, hemicellulose and lignin for native and GVL-pretreated P39 poplar biomass as estimated using the CP/MAS-enabled methods described above.

Native P39 biomass									
Table 3.1 Entry # (from main text)	$I_{C-1}$	$I_L$	$I_H$	$f_{C-1}$ (calculated)	$f_L$ (calculated)	$f_H$ (calculated)	Wt% cellulose (known)	Wt% lignin (known)	Wt% hemicellulose (known)
--	694227	224597	172643	0.998	1.460	1.776	52.3	24.7	23.1
GVL-pretreated P39 biomass									
Table 3.1 Entry # (from main text)	$I_{C-1}$	$I_L$	$I_H$	$f_{C-1}$ (fixed)	$f_L$ (fixed)	$f_H$ (fixed)	Wt% cellulose (calc.)	Wt% lignin (calc.)	Wt% hemicellulose (calc.)
5	521373	113521	65318	0.998	1.460	1.776	64.9	20.7	14.5
9	535620	7666	45198	0.998	1.460	1.776	73.6	15.3	11.0

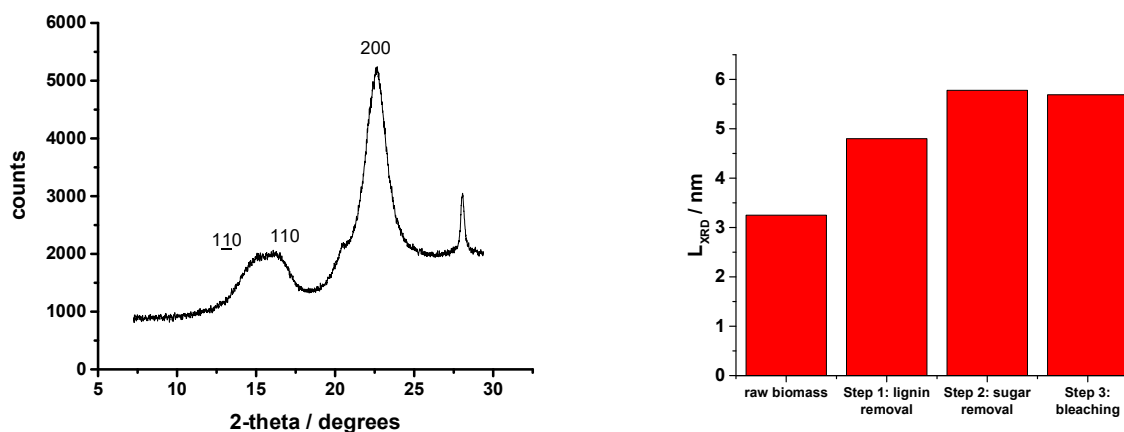
#### Attenuated total reflectance (ATR)-FTIR spectroscopy

The removal of lignin and hemicellulose from P39 biomass achieved by the solvent-mediated pretreatment is confirmed by attenuated-total-reflectance infrared spectroscopy (ATR-FTIR). These results are shown in Figure S3.8, wherein the spectral features corresponding to lignin and hemicellulose vanish going from native P39 biomass to the pretreated material.<sup>62</sup> Note that the spectral bands corresponding to hemicellulose and cellulose at about 1000  $\text{cm}^{-1}$  overlap, so that the relative intensities of the bands at 500 and 1000  $\text{cm}^{-1}$  change as hemicellulose is removed from the surface of the cellulose microfibrils.



**Figure S3.8.** ATR-FTIR spectra for native P39 biomass (top) and P39 pretreated in GVL-water solvent systems to remove hemicellulose and lignin (bottom). Note that signals corresponding to cellulose and hemicellulose overlap at  $\sim 1000$  wavenumbers.

*X-ray diffraction characterizations of native and pretreated cellulose*



**Figure S3.9.** Representative X-ray diffractogram for native P39 biomass (left) and XRD-derived estimates of the diameter of the cellulose microfibrils in native and solvent-pretreated P39 biomasses.

Figure S3.9 displays a representative diffractogram corresponding to cellulose in raw P39 biomass. The diffraction peaks associated with the  $1\bar{1}0$ ,  $110$  and  $200$  facets of the monoclinic unit cell are indicated; their positions on the  $2\theta$  scale correspond to inter-planar spacings of 5.91, 5.43 and 3.92 Angstroms, respectively. After the appropriate baseline corrections and accounting for the inherent line broadening of the instrument, the full width at half-maximum (FWHM) intensity for the  $200$  peak was collected for each sample. These data are used to estimate the lateral dimensions of the ordered cellulose domains (microfibrils), by invoking the standard relationship:<sup>63</sup>

$$L_{XRD} = \frac{K\lambda}{\beta \cos\theta} \quad (\text{S10})$$

The term  $L_{XRD}$  is the average cross-sectional length of a cellulose microfibril,  $\beta$  is the full width at half maximum intensity of the  $200$  diffraction peak (in radians),  $\theta$  is the Bragg angle (again in radians),  $\lambda$  is the wavelength of the scattered X-ray beam, and  $K$  is a factor on the order of 0.9 (which is the value used here). Based on the line broadening of the  $200$  diffraction peak, the apparent size of the microfibrils ranges from 3.25 for native P39 biomass, to about 5.5 nm for

GVL-pretreated P39. These values are within the range of microfibril dimensions reported for similar cellulosic materials.<sup>48, 57</sup> The greatest variance in the apparent value of  $L_{XRD}$  corresponds to the native and minimally processed P39 samples, while  $L_{XRD}$  approaches a near-constant value as lignin and hemicellulose are removed. It should be noted that phenomena other than changes in the characteristic dimensions of an ordered phase can result in broadening of X-ray diffraction features. These additional phenomena include epitaxial strain, lattice defects, and the presence of amorphous material in the crystal matrix.<sup>64-65</sup>

The aforementioned XRD-enabled measurements allow for a microfibril structure to be proposed for P39 biomass by considering the following insights:

1. Assuming that the removal of amorphous material by the solvent is responsible for the reduced line broadening going from native to pretreated P39 biomass, we expect that the true cross-sectional dimensions of P39 cellulose microfibrils falls within the range represented by the pretreated samples: 4.9-5.8 nm.
2. Furthermore, the positions of the 110,  $1\bar{1}0$  and 200 peaks displayed in Figure S3.9(a) indicate a monoclinic unit cell with lattice parameters (interplanar spacings) of 5.43, 5.91 and 3.92 Angstroms, respectively.
3. The crystal structure of cellulose microfibrils in hardwood has been reported as a monoclinic unit cell with a characteristic angle near 90 degrees.<sup>48, 57</sup>

A microfibril geometry constituted by an 8x8 rectangular array of glucan chains with side lengths of 4.24 and 3.80 nm, and a diagonal length of 5.59 nm would be consistent with these observations. This is the structure proposed in Figure 3.11 of the main text.

### *Exemplary mass balances*

Table S3.2 displays representative mass balances for the 2-step lignin- and sugar-removal process outlined in Figure 3.5 of the main text, using two different temperatures for the sugar removal step (120 and 140°C). the composition of the GVL-pretreated solids reflect those shown

in Table S3.1. The yields of water, extractives, and soluble glucan, xylan and lignin were estimated using methods described in Section S1.1. The yield of humins was back-calculated so as to close the mass balance, and the resulting yield of humins in each step reflects those typically reported in the literature for similar processes. The carbon balances likewise reflect those typically reported for similar processes.<sup>14, 27</sup>

**Table S3.2.**

<b>Key Process Step</b>	<b>Step 1 (lignin removal)</b>	<b>Step 2 (sugar removal)</b>	<b>Step 1 (lignin removal)</b>	<b>Step 2 (sugar removal)</b>
<b>biomass in</b>	<b>P39</b>	<b>delignified P39 100C</b>	<b>P39</b>	<b>delignified P39 100C</b>
water (wt%)	7%	0%	7%	0%
extractives (wt%)	5%	0%	5%	0%
cellulose (wt%)	46%	65%	46%	65%
xylan (wt%)	23%	14%	23%	14%
lignin (wt%)	20%	21%	20%	21%
<b>total biomass (g)</b>	<b>5.9</b>	<b>4.0660</b>	<b>6.04</b>	<b>4.1446</b>
<b>biomass loading (solvent/solids ratio)</b>	<b>10</b>	<b>14.5106</b>	<b>10</b>	<b>14.5732</b>
<b>organic cosolvent</b>	<b>GVL</b>	<b>GVL</b>	<b>GVL</b>	<b>GVL</b>
FORMULA	C5H8O2	C5H8O2	C5H8O2	C5H8O2
density	1.05	1.05	1.05	1.05
molar mass	100.12	100.12	100.12	100.12
<b>cosolvent/water ratio</b>	<b>9</b>	<b>4</b>	<b>9</b>	<b>4</b>
<b>catalyst</b>	<b>sulfuric acid</b>	<b>sulfuric acid</b>	<b>sulfuric acid</b>	<b>sulfuric acid</b>
FORMULA	H2SO4	H2SO4	H2SO4	H2SO4
density	1.84	1.84	1.84	1.84
molar mass	98.079	98.079	98.079	98.079
<b>catalyst concentration (M)</b>	<b>0.1</b>	<b>0.1</b>	<b>0.1</b>	<b>0.1</b>
<b>Batch time (hr)</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
<b>Batch temperature (°C)</b>	<b>100</b>	<b>120</b>	<b>100</b>	<b>140</b>
catalyst loading (g catalyst/g biomass)	0.0938	0.1368	0.0938	0.1374
<b>biomass out</b>	<b>delignified P39 100C</b>	<b>desach P39 120C</b>	<b>delignified P39 100C</b>	<b>desach P39 140C</b>
water (wt%)	0%	0%	0%	0%
extractives (wt%)	0%	0%	0%	0%
cellulose (wt%)	65%	69%	65%	74%
xylan (wt%)	14%	13%	14%	11%
lignin (wt%)	21%	18%	21%	15%
<b>total solids out (g)</b>	<b>4.066</b>	<b>2.5671</b>	<b>4.1446</b>	<b>2.102</b>
<b>Feed (g)</b>	<b>65.4537</b>	<b>63.6232</b>	<b>67.0069</b>	<b>65.1150</b>

Extractives (solid)	C6H12O6	0.2950	0.0000	0.3020	0.0000
Cellulose (solid)	C6H10O5	2.7022	2.6396	2.7663	2.6907
Xylan (solid)	C5H8O4	1.3334	0.5855	1.3650	0.5968
Lignin (solid)	C8H8O3	1.1564	0.8417	1.1838	0.8579
H2O	H2O	6.3130	11.8000	6.4628	12.0800
cosolvent	--	53.1000	47.2000	54.3600	48.3200
catalyst	--	0.5537	0.5564	0.5669	0.5696
Furfural	C5H4O2	0.0000	0.0000	0.0000	0.0000
Lignin (soluable)	C8H8O3	0.0000	0.0000	0.0000	0.0000
Extractives (soluable)	C6H12O6	0.0000	0.0000	0.0000	0.0000
Glucose	C6H12O6	0.0000	0.0000	0.0000	0.0000
Xylose	C5H10O5	0.0000	0.0000	0.0000	0.0000
humins		0	0	0	0
<b>Product (g or kg/hr)</b>		<b>65.4537</b>	<b>63.6232</b>	<b>67.0078</b>	<b>65.1150</b>
Extractives (solid)	C6H12O6	0.0000	0	0.0000	0
Cellulose (solid)	C6H10O5	2.6396	1.7790	2.6907	1.5487
Xylan (solid)	C5H8O4	0.5855	0.3266	0.5968	0.2322
Lignin (solid)	C8H8O3	0.8417	0.4619	0.8579	0.3213
H2O	H2O	6.2438	11.7624	6.3939	12.0091
cosolvent	C5H8O2	53.1000	47.2000	54.3600	48.3200
catalyst	catalyst	0.5537	0.5564	0.5669	0.5696
Furfural	C5H4O2	0.0488	0	0.0577	0
Lignin (soluable)	C8H8O3	0.4130	0.531	0.4228	0.5436
Extractives (soluable)	C6H12O6	0.2950	0	0.3020	0
Glucose	C6H12O6	0.0216	0.0542	0.0270	0.5006
Xylose	C5H10O5	0.7109	0.2678	0.7321	0.174
humins	C8H8O3	0.0000	0.6838	0.0000	0.8957

CARBON FLOWS		Step 1	Step 2 (120°C)	Step 1	Step 2 (140°C)
component	carbon wt%	lignin removal	sugar removal	lignin removal	sugar removal
Extractives in	40%	0.1180	0.0000	0.1208	0.0000
cellulose in	44%	1.2009	1.1731	1.2294	1.1958
xylan in	45%	0.6061	0.2661	0.6205	0.2713
solid lignin in	63%	0.7299	0.5313	0.7473	0.5415
extractives out	40%	0.1180	0.0000	0.1208	0.0000
cellulose out	44%	1.1731	0.7906	1.1958	0.6883
xylan out	45%	0.2661	0.1485	0.2713	0.1056
solid lignin out	63%	0.5313	0.2916	0.5415	0.2029
furfural out	62%	0.0305	0.0000	0.0360	0.0000
soluble lignin out	63%	0.2607	0.3352	0.2669	0.3431
glucose out	40%	0.0086	0.0217	0.0108	0.2002
xylose out	40%	0.2841	0.1070	0.2926	0.0695
carbon balance:	--	<b>100.66%</b>	<b>86.00%</b>	<b>100.65%</b>	<b>80.14%</b>



PERFORMANCE METRICS:	Entry -->	Step 1	Step 2 (120°C)	Step 1	Step 2 (140°C)
metric	Step -->	lignin removal	sugar removal	lignin removal	sugar removal
xylan conversion (mol%)		56.09%	44.21%	56.28%	61.08%
glucan conversion (mol%)		2.31%	32.60%	2.73%	42.44%
lignin conversion (mol%)		27.22%	45.11%	27.53%	62.54%
xylose yield (mol%)		46.88%	40.22%	47.16%	25.63%
fufural yield (mol%)		5.03%	0.00%	5.81%	0.00%
glucose yield (mol%)		0.72%	1.85%	0.88%	16.74%

## References:

1. Bobleter, O., Hydrothermal degradation of polymers derived from plants. *Progress in Polymer Science* **1994**, *19* (5), 797-841.
2. Mosier, N.; Wyman, C.; Dale, B.; Elander, R.; Lee, Y.; Holtzapple, M.; Ladisch, M., Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource technology* **2005**, *96* (6), 673-686.
3. Jenkins, B.; Baxter, L.; Miles Jr, T.; Miles, T., Combustion properties of biomass. *Fuel processing technology* **1998**, *54* (1-3), 17-46.
4. O'SULLIVAN, A. C., Cellulose: the structure slowly unravels. *Cellulose* **1997**, *4* (3), 173-207.
5. Nishiyama, Y., Structure and properties of the cellulose microfibril. *Journal of Wood Science* **2009**, *55* (4), 241-249.
6. Frey-Wyssling, A., The Fine Structure of Cellulose Microfibrils. *Science* **1954**, *119* (3081), 80-82.
7. Salmén, L., Micromechanical understanding of the cell-wall structure. *Comptes Rendus Biologies* **2004**, *327* (9), 873-880.
8. Harris, P. J.; Stone, B. A., Chemistry and molecular organization of plant cell walls. *Biomass recalcitrance: deconstructing the plant cell wall for bioenergy* **2009**, 61-93.
9. Chheda, J. N.; Huber, G. W.; Dumesic, J. A., Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition* **2007**, *46* (38), 7164-7183.
10. Wettstein, S. G.; Alonso, D. M.; Gürbüz, E. I.; Dumesic, J. A., A roadmap for conversion of lignocellulosic biomass to chemicals and fuels. *Current Opinion in Chemical Engineering* **2012**, *1* (3), 218-224.
11. Gaddy, B. E.; Sivaram, V.; Jones, T. B.; Wayman, L., Venture capital and cleantech: The wrong model for energy innovation. *Energy Policy* **2017**, *102*, 385-395.

12. Jørgensen, H.; Kristensen, J. B.; Felby, C., Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels, Bioproducts and Biorefining* **2007**, *1* (2), 119-134.
13. Nguyen, Q.; Saddler, J., An integrated model for the technical and economic evaluation of an enzymatic biomass conversion process. *Bioresource Technology* **1991**, *35* (3), 275-282.
14. Luterbacher, J. S.; Rand, J. M.; Alonso, D. M.; Han, J.; Youngquist, J. T.; Maravelias, C. T.; Pfleger, B. F.; Dumesic, J. A., Nonenzymatic sugar production from biomass using biomass-derived  $\gamma$ -valerolactone. *Science* **2014**, *343* (6168), 277-280.
15. Cai, C. M.; Zhang, T.; Kumar, R.; Wyman, C. E., THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chemistry* **2013**, *15* (11), 3140-3145.
16. Liu, Z.; Zhang, F.-S., Effects of various solvents on the liquefaction of biomass to produce fuels and chemical feedstocks. *Energy conversion and management* **2008**, *49* (12), 3498-3504.
17. Kumar, R.; Singh, S.; Singh, O. V., Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of industrial microbiology & biotechnology* **2008**, *35* (5), 377-391.
18. Klein-Marcuschamer, D.; Oleskowicz-Popiel, P.; Simmons, B. A.; Blanch, H. W., The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and bioengineering* **2012**, *109* (4), 1083-1087.
19. Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M. *Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover*; National Renewable Energy Lab.(NREL), Golden, CO (United States): 2011.
20. Zhao, X.; Cheng, K.; Liu, D., Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied Microbiology and Biotechnology* **2009**, *82* (5), 815-827.
21. Hendriks, A.; Zeeman, G., Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource technology* **2009**, *100* (1), 10-18.
22. Mooney, C. A.; Mansfield, S. D.; Beatson, R. P.; Saddler, J. N., The effect of fiber characteristics on hydrolysis and cellulase accessibility to softwood substrates. *Enzyme and Microbial Technology* **1999**, *25* (8-9), 644-650.
23. Chundawat, S. P.; Bellesia, G.; Uppugundla, N.; da Costa Sousa, L.; Gao, D.; Cheh, A. M.; Agarwal, U. P.; Bianchetti, C. M.; Phillips Jr, G. N.; Langan, P., Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *Journal of the American Chemical Society* **2011**, *133* (29), 11163-11174.
24. Chang, V. S.; Burr, B.; Holtzapfle, M. T., Lime pretreatment of switchgrass. In *Biotechnology for fuels and chemicals*, Springer: 1997; pp 3-19.
25. Fan, L.; Lee, Y. H.; Beardmore, D., The influence of major structural features of cellulose on rate of enzymatic hydrolysis. *Biotechnology and Bioengineering* **1981**, *23* (2), 419-424.
26. Grethlein, H. E., The Effect of Pore Size Distribution on the Rate of Enzymatic Hydrolysis of Cellulosic Substrates. *Bio/Technology* **1985**, *3* (2), 155-160.
27. Alonso, D. M.; Hakim, S. H.; Zhou, S.; Won, W.; Hosseinaei, O.; Tao, J.; Garcia-Negron, V.; Motagamwala, A. H.; Mellmer, M. A.; Huang, K., Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science advances* **2017**, *3* (5), e1603301.
28. Shuai, L.; Questell-Santiago, Y. M.; Luterbacher, J. S., A mild biomass pretreatment using  $\gamma$ -valerolactone for concentrated sugar production. *Green Chemistry* **2016**, *18* (4), 937-943.

29. Motagamwala, A. H.; Won, W.; Maravelias, C. T.; Dumesic, J. A., An engineered solvent system for sugar production from lignocellulosic biomass using biomass derived  $\gamma$ -valerolactone. *Green Chemistry* **2016**, *18* (21), 5756-5763.
30. Luterbacher, J. S.; Azarpira, A.; Motagamwala, A. H.; Lu, F.; Ralph, J.; Dumesic, J. A., Lignin monomer production integrated into the  $\gamma$ -valerolactone sugar platform. *Energy & Environmental Science* **2015**, *8* (9), 2657-2663.
31. Yokoyama, T., Revisiting the Mechanism of  $\beta$ -O-4 Bond Cleavage During Acidolysis of Lignin. Part 6: A Review. *Journal of Wood Chemistry and Technology* **2015**, *35* (1), 27-42.
32. Deuss, P. J.; Scott, M.; Tran, F.; Westwood, N. J.; de Vries, J. G.; Barta, K., Aromatic monomers by in situ conversion of reactive intermediates in the acid-catalyzed depolymerization of lignin. *Journal of the American Chemical Society* **2015**, *137* (23), 7456-7467.
33. Johar, N.; Ahmad, I.; Dufresne, A., Extraction, preparation and characterization of cellulose fibres and nanocrystals from rice husk. *Industrial Crops and Products* **2012**, *37* (1), 93-99.
34. Yang, Q.; Pan, X., Correlation between lignin physicochemical properties and inhibition to enzymatic hydrolysis of cellulose. *Biotechnology and Bioengineering* **2016**, *113* (6), 1213-1224.
35. Horváth, I. T.; Mehdi, H.; Fábos, V.; Boda, L.; Mika, L. T.,  $\gamma$ -Valerolactone—a sustainable liquid for energy and carbon-based chemicals. *Green Chemistry* **2008**, *10* (2), 238-242.
36. Mellmer, M. A.; Sener, C.; Gallo, J. M. R.; Luterbacher, J. S.; Alonso, D. M.; Dumesic, J. A., Solvent effects in acid-catalyzed biomass conversion reactions. *Angewandte chemie international edition* **2014**, *53* (44), 11872-11875.
37. Mellmer, M. A.; Alonso, D. M.; Luterbacher, J. S.; Gallo, J. M. R.; Dumesic, J. A., Effects of  $\gamma$ -valerolactone in hydrolysis of lignocellulosic biomass to monosaccharides. *Green Chemistry* **2014**, *16* (11), 4659-4662.
38. Unda, F.; Kim, H.; Hefer, C.; Ralph, J.; Mansfield, S. D., Altering carbon allocation in hybrid poplar (*Populus alba*  $\times$  *grandidentata*) impacts cell wall growth and development. *Plant biotechnology journal* **2017**, *15* (7), 865-878.
39. Hamilton, S.; Hussain, M.; Bhardwaj, A.; Basso, B.; Robertson, G., Comparative water use by maize, perennial crops, restored prairie, and poplar trees in the US Midwest. *Environmental Research Letters* **2015**, *10* (6), 064015.
40. Lloyd, T. A.; Wyman, C. E., Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Bioresource technology* **2005**, *96* (18), 1967-1977.
41. Morán, J. I.; Alvarez, V. A.; Cyras, V. P.; Vázquez, A., Extraction of cellulose and preparation of nanocellulose from sisal fibers. *Cellulose* **2008**, *15* (1), 149-159.
42. Patil, S. K.; Lund, C. R., Formation and growth of humins via aldol addition and condensation during acid-catalyzed conversion of 5-hydroxymethylfurfural. *Energy & Fuels* **2011**, *25* (10), 4745-4755.
43. Falco, C.; Perez Caballero, F.; Babonneau, F.; Gervais, C.; Laurent, G.; Titirici, M.-M.; Baccile, N., Hydrothermal carbon from biomass: structural differences between hydrothermal and pyrolyzed carbons via  $^{13}\text{C}$  solid state NMR. *Langmuir* **2011**, *27* (23), 14460-14471.
44. Fu, L.; McCallum, S. A.; Miao, J.; Hart, C.; Tudryn, G. J.; Zhang, F.; Linhardt, R. J., Rapid and accurate determination of the lignin content of lignocellulosic biomass by solid-state NMR. *Fuel* **2015**, *141*, 39-45.
45. Foston, M., Advances in solid-state NMR of cellulose. *Current opinion in biotechnology* **2014**, *27*, 176-184.

46. Atalla, R. H.; Gast, J.; Sindorf, D.; Bartuska, V.; Maciel, G., Carbon-13 NMR spectra of cellulose polymorphs. *Journal of the American Chemical Society* **1980**, *102* (9), 3249-3251.
47. Gilardi, G.; Abis, L.; Cass, A. E., Carbon-13 CP/MAS solid-state NMR and FT-IR spectroscopy of wood cell wall biodegradation. *Enzyme and Microbial Technology* **1995**, *17* (3), 268-275.
48. Newman, R. H., Estimation of the lateral dimensions of cellulose crystallites using <sup>13</sup>C NMR signal strengths. *Solid State Nuclear Magnetic Resonance* **1999**, *15* (1), 21-29.
49. Earl, W. L.; VanderHart, D., Observations by high-resolution carbon-13 nuclear magnetic resonance of cellulose I related to morphology and crystal structure. *Macromolecules* **1981**, *14* (3), 570-574.
50. Larsson, P. T.; Hult, E.-L.; Wickholm, K.; Pettersson, E.; Iversen, T., CP/MAS <sup>13</sup>C-NMR spectroscopy applied to structure and interaction studies on cellulose I. *Solid State Nuclear Magnetic Resonance* **1999**, *15* (1), 31-40.
51. Newman, R.; Hemmingson, J.; Suckling, I., Carbon-13 nuclear magnetic resonance studies of kraft pulping. *Holzforschung-International Journal of the Biology, Chemistry, Physics and Technology of Wood* **1993**, *47* (3), 234-238.
52. Hult, E.-L.; Larsson, P. T.; Iversen, T., A comparative CP/MAS <sup>13</sup>C-NMR study of the supermolecular structure of polysaccharides in sulphite and kraft pulps. *Holzforschung* **2002**, *56* (2), 179-184.
53. Alam, T. M.; Childress, K. K.; Pastoor, K.; Rice, C. V., Characterization of free, restricted, and entrapped water environments in poly (N-isopropyl acrylamide) hydrogels via <sup>1</sup>H HRMAS PFG NMR spectroscopy. *Journal of Polymer Science Part B: Polymer Physics* **2014**, *52* (23), 1521-1527.
54. Zhang, C.; Li, P.; Zhang, Y.; Lu, F.; Li, W.; Kang, H.; Xiang, J.-f.; Huang, Y.; Liu, R., Hierarchical porous structures in cellulose: NMR relaxometry approach. *Polymer* **2016**, *98*, 237-243.
55. Corsaro, C.; Mallamace, D.; Vasi, S.; Pietronero, L.; Mallamace, F.; Missori, M., The role of water in the degradation process of paper using <sup>1</sup>H HR-MAS NMR spectroscopy. *Physical Chemistry Chemical Physics* **2016**, *18* (48), 33335-33343.
56. Selig, M. J.; Thygesen, L. G.; Felby, C., Correlating the ability of lignocellulosic polymers to constrain water with the potential to inhibit cellulose saccharification. *Biotechnology for biofuels* **2014**, *7* (1), 159.
57. Leppänen, K.; Andersson, S.; Torkkeli, M.; Knaapila, M.; Kotelnikova, N.; Serimaa, R., Structure of cellulose and microcrystalline cellulose from various wood species, cotton and flax studied by X-ray scattering. *Cellulose* **2009**, *16* (6), 999-1015.
58. Zalesny Jr, R. S.; Wiese, A. H.; Bauer, E. O.; Riemenschneider, D. E., Sapflow of hybrid poplar (*Populus nigra* L. × *P. maximowiczii* A. Henry 'NM6') during phytoremediation of landfill leachate. *Biomass and bioenergy* **2006**, *30* (8-9), 784-793.
59. Wilkerson, C.; Mansfield, S.; Lu, F.; Withers, S.; Park, J.-Y.; Karlen, S.; Gonzales-Vigil, E.; Padmakshan, D.; Unda, F.; Rencoret, J., Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science* **2014**, *344* (6179), 90-93.
60. Sugiyama, J.; Vuong, R.; Chanzy, H., Electron diffraction study on the two crystalline phases occurring in native cellulose from an algal cell wall. *Macromolecules* **1991**, *24* (14), 4168-4175.

61. Zamora, D. S.; Wyatt, G. J.; Apostol, K. G.; Tschirner, U., Biomass yield, energy values, and chemical composition of hybrid poplars in short rotation woody crop production and native perennial grasses in Minnesota, USA. *Biomass and Bioenergy* **2013**, *49*, 222-230.
62. Zhou, C.; Jiang, W.; Via, B. K.; Fasina, O.; Han, G., Prediction of mixed hardwood lignin and carbohydrate content using ATR-FTIR and FT-NIR. *Carbohydrate polymers* **2015**, *121*, 336-341.
63. Edwards, A., HP Klug and LE Alexander, x-ray diffraction procedures for polycrystalline and amorphous materials: Wiley-Interscience, New York, 2nd edn., 1974, xxv+ 966 pp. price£ 18.55. Elsevier: 1975.
64. Ungár, T., Microstructural parameters from X-ray diffraction peak broadening. *Scripta Materialia* **2004**, *51* (8), 777-781.
65. Hindeleh, A.; Johnson, D., Peak resolution and X-ray crystallinity determination in heat-treated cellulose triacetate. *Polymer* **1972**, *13* (1), 27-32.

## Chapter 4:

### Evolution of the cellulose microfibril through gamma-valerolactone assisted co-solvent and enzymatic hydrolysis

Elise B. Gilcher, Nathaniel Kuch, Joshua T. Del Mundo, Leida M. Vázquez Ramos, Samantha F. Ausman, Catherine F. M. Clewett, Esther W. Gomez, Enrique D. Gomez, Brian G. Fox, Thatcher W. Root and James A. Dumesic

This work was submitted to ACS Sustainable Chemistry & Engineering

Nathaniel Kuch made the following contributions: planning and execution of enzyme assays, data analysis, writing and revising of manuscript

#### Abstract:

Biomass recalcitrance during deconstruction remains a key bottleneck to affordable biomass processing technologies. A clear connection between the cell wall structure and biomass deconstruction is necessary to understand how lignocellulosic material is broken down to valuable monomeric components. Here, we monitor changes in the cellulose microfibril domains in poplar, sorghum and switchgrass throughout gamma-valerolactone (GVL)-water co-solvent treatment and enzymatic hydrolysis using solid-state  $^{13}\text{C}$  cross-polarization magic angle spinning (CP/MAS-NMR) spectroscopy and wide-angle X-ray scattering (WAXS). Spectral fitting of NMR peaks corresponding to different cellulose microenvironments at the  $\text{C}_4$  carbon center suggest that a mildly acidic GVL-water co-solvent pretreatment of poplar leads to nearly full removal of xylan-cellulose linkages, which primes the cellulose for enzymatic attack. The spectral fitting also suggests that the pretreatment causes significant depletion of the inaccessible fibril surface domains with an increase in more thermally stable crystalline resonances ( $I_\beta$ ). WAXS confirmed a decrease in lattice spacing between (200) crystalline planes with increasing co-solvent pretreatment severity. These results are interpreted as an opening of bound microfibril surfaces previously inaccessible to the co-solvent, which leaves behind a more thermally stable,

crystalline domain that is potentially prone to relaxation and recrystallization. Full conversion of residual GVL treated biomass was achieved after the GVL co-solvent pretreatment at 140°C using a commercial enzyme cocktail, CTec3, which contains different cellulases and other enzymes. Spectral fitting of enzymatically hydrolyzed samples by a single engineered cellulase, CelR, suggests that the residual cellulose recalcitrance is mainly due to an inability of CelR to digest the I<sub>β</sub> crystalline domain present in pretreated samples. This work helps bridge the knowledge gap between cell wall structure throughout GVL-assisted and CelR enzymatic biomass deconstruction by tracking the evolution of structural domains within the cellulose microfibril, and directs recommendations for future studies. Our findings inform inquiry into larger questions of cellulose recalcitrance through GVL and CelR enzymatic hydrolysis, and give insight to subsequent required steps for full cellulose conversion with attention to the most recalcitrant cellulose structures.

### **Introduction:**

Plant cell walls are comprised of three main biopolymers within an intricate heterogeneous matrix: cellulose (C<sub>6</sub> sugars); hemicellulose (C<sub>5</sub> and C<sub>6</sub> sugars); and lignin (polyaromatics). As such, plant biomass is a renewable, abundant and carbon neutral<sup>1</sup> energy source that can be used to produce biofuels<sup>2,3</sup> and biomass-derived platform chemicals<sup>4-6</sup>. Continued interest in cellulosic biofuels and biomass-derived products has motivated efforts to understand the molecular determinants of biomass recalcitrance and thereby increase total sugar yields in biomass fractionation and deconstruction processes. However, biomass recalcitrance remains a key bottleneck to development of effective biomass processing technologies.<sup>7-9</sup>

Biomass recalcitrance is attributed to multiple factors, including the high degree of organization and binding within the cellulose ultrastructure<sup>10</sup> as well as the presence of lignin and hemicellulose within the matrix<sup>11</sup>. Cellulose consists of many tightly bound, parallel polymer chains of the C<sub>6</sub> sugar  $\beta$ -glucose that interact strongly with neighboring hydroxyl groups through inter- and intra-molecular hydrogen bonding. This behavior stiffens the linear chains and forms cellulose microfibrils, imparting rigidity to the plant cell wall. The strong binding also makes cellulose highly resistant to enzymatic, microbial, and chemical deconstruction to solubilized sugars<sup>2,10</sup>. Furthermore, hemicellulose and lignin adhere to cellulose microfibrils, limiting accessibility to the surrounding solvent or enzymatic active sites.

Mild acid hydrolysis<sup>12</sup>, alkaline treatment<sup>13,14</sup>, and organosolv<sup>15,16</sup> can chemically fractionate lignocellulosic biomass to depolymerized sugars and solubilized lignin compounds. Our previous work demonstrated that biomass-derived  $\gamma$ -valerolactone (GVL)<sup>17</sup> is effective at deconstructing biomass in GVL-water co-solvent systems, with up to 70% carbohydrate yields for C<sub>5</sub> and C<sub>6</sub> sugars.<sup>18</sup> While purely solvent-based systems feature rapid catalytic turn-over, they require high operating temperatures and often yield lower purity of monomeric sugars compared to other deconstruction methods.<sup>19,20</sup> In contrast, enzymatic systems boast higher purity<sup>21</sup> of monomeric sugars with milder reaction temperatures and ambient pressures,<sup>15,22</sup> but suffer from low turnover numbers and long reaction times while showing little activity on untreated biomass.<sup>23-25</sup>

The relationship between cell wall structure and cellulose recalcitrance remains a key challenge in the advancement of lignocellulosic deconstruction for biofuels and biomass-derived products. <sup>13</sup>C-cross-polarization magic-angle-spinning nuclear magnetic resonance (CP/MAS-NMR) has been used to identify different domains within the cellulose microfibril via the C<sub>4</sub> resonance, which is spread over a wide chemical shift range and has little overlap with other



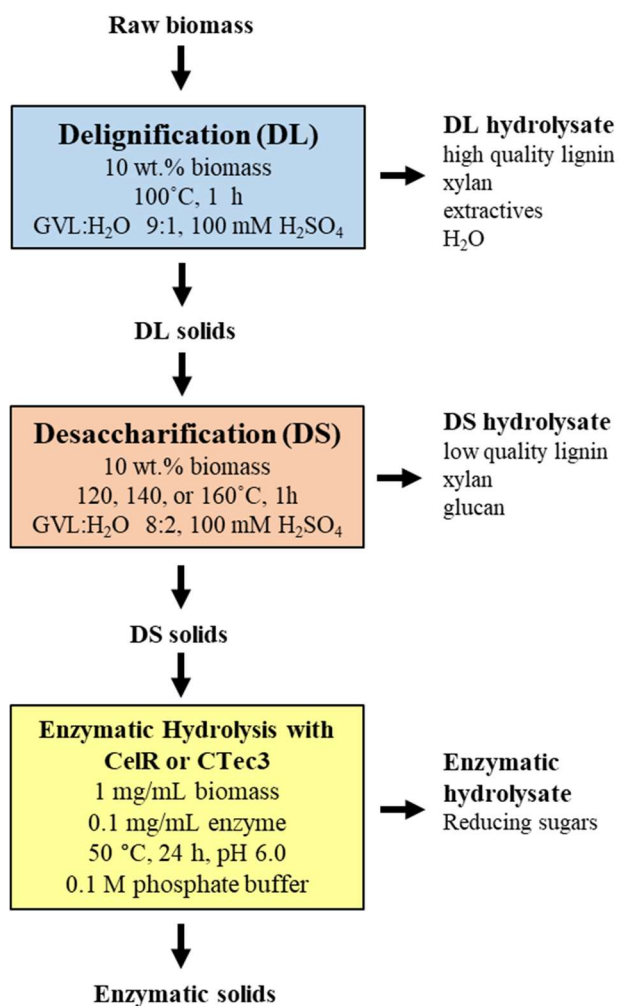
cellulose resonances, enabling non-linear, least squares spectral-line fitting of the C<sub>4</sub> resonance. A model was developed to identify individual magnetic micro-environments of cellulose I contributing to the C<sub>4</sub> resonance<sup>26,27</sup> which has been applied to monitor the structure and interactions of various cellulose I biomass types throughout pulping<sup>22,28,29</sup> and various hydrothermal<sup>30</sup>, chemical<sup>12,31</sup> and enzymatic<sup>15,32</sup> hydrolysis treatments.

Recently, we used CP/MAS NMR to study the combination of GVL co-solvent pretreatment with enzymatic hydrolysis. Our initial examination showed a correlation between chemical breakage of cellulosic and non-cellulosic linkages and increased sugar yields from hydrolysis with an engineered cellulase.<sup>20</sup> A single observable metric from the CP/MAS NMR ( $X_{\text{NMR}}$ ) was found to predict enzymatic sugar yields independent of co-solvent system or biomass type used, and suggested an increase in crystallinity of cellulose remaining after GVL-pretreatments of increasing severity.<sup>20</sup> However, physical changes within the cellulose microfibril structure throughout the GVL-assisted pretreatments were not well understood, and the recalcitrance of GVL pretreated biomass samples to enzymatic hydrolysis was largely unexplained.

A clear understanding of the structure and chemical properties of lignocellulose throughout deconstruction is required to increase biomass utilization through reduced recalcitrance and lead to commercialization of biofuels and bioproducts.<sup>7,23,33</sup> In this work, changes in the cellulosic microfibril domains caused by GVL co-solvent assisted chemical pretreatment and subsequent enzymatic hydrolysis are examined using CP/MAS NMR and wide angle X-ray scattering. We provide a molecular level understanding of the interplay between GVL treatment and enzyme hydrolysis with implications for selection of the minimal number of enzymes and process steps needed to achieve full biomass deconstruction.

## Results:

### *Pretreatment Process*



**Figure 4.1:** Schematic diagram showing the GVL-assisted mild acidolysis pretreatment and subsequent enzymatic hydrolysis with CelR, adapted from reference.<sup>20</sup>

Figure 4.1 shows a schematic representation of a “two-step” GVL-assisted mild acidolysis of biomass explored in conjunction with enzymatic hydrolysis of the remaining cellulosic fraction. We substituted the small 60 mL glass reactors used in previous work<sup>20</sup> with a 450 mL stirred stainless steel batch reactor to provide higher throughput of biomass, more effective mixing of the biomass slurry, higher yield of residual solids to support more detailed characterizations and

post-treatment reactions. All raw feedstocks were debarked (if applicable) and ground to 5 mm mesh size and charged into a batch, stirred tank reactor without any prior treatment for delignification.

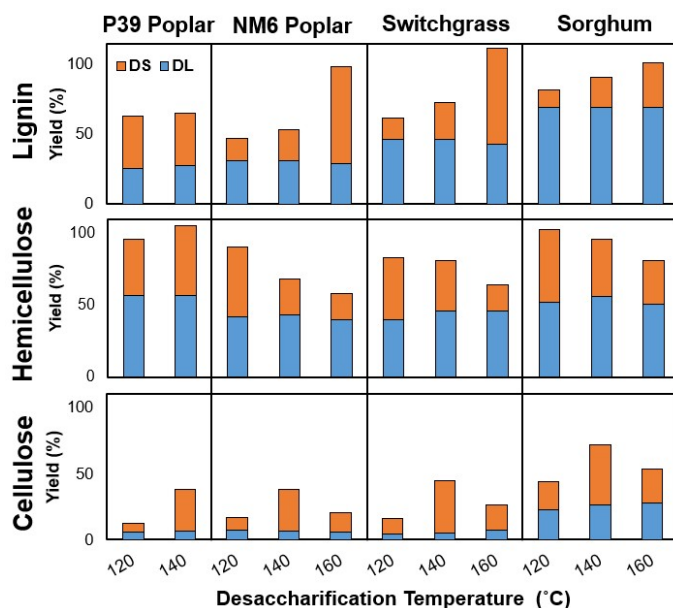
The delignification step (Figure 4.1, DL) consists of a solvent system containing ~90 wt% GVL, and 10 wt% water, and dilute 100 mM sulfuric acid catalyst added to the reactor containing 10 wt% biomass. This mild treatment condition provided a soluble lignin that more closely resembled native lignin in the cell wall, and is more suitable for subsequent processing by hydrogenolysis, than condensed lignin formed with more severe treatment conditions. Hemicellulose is also extracted as sugar monomers into the aqueous fraction (hydrolysate) along with small amounts of cellulose under these conditions. Residual solids consist primarily of cellulose and lesser amounts of hemicellulose and lignin.

The liquid hydrolysate and residual solids are separated through filtration and the solids are charged into a second batch, stirred tank reactor for a more severe GVL-assisted desaccharification step (Figure 4.1, DS). In this second step of treatment, temperature was varied to explore the effect of different GVL-assisted hydrolysis conditions on the insoluble cellulosic material. Lower desaccharification temperatures yielded a cellulose-enriched, solid product and a desaccharification hydrolysate contained the remaining soluble lignin, notably at a lower quality than obtained in the first step, as well as the remaining hemicellulose sugar and increasing amounts of C<sub>6</sub> sugar monomers as the temperature was increased. Desaccharification at 160°C yielded only a liquid product, where all constituents of the biomass feed were completely hydrolyzed. Cellulose-enriched solids were subjected to enzymatic digestion to encompass an overall biomass deconstruction system.

### Analysis of Hydrolysates

Four biomass feedstocks were selected to examine the effects of GVL hydrolysis process over different starting materials and treatment conditions: wildtype P39 poplar; NM6 poplar; switchgrass; and sorghum. Raw wildtype feedstocks were treated under delignification (DL) conditions at 100°C and subsequently under desaccharification (DS) at either 120, 140, or 160°C. All treatment conditions yielded a solid residual biomass fraction except for desaccharification at 160°C (DS 160), which completely solubilized all biomass types.

Figure 4.2 depicts the yields in the hydrolysate solution for lignin (determined by weight after water-induced precipitation, washing and drying), hemicellulose (defined as xylose, arabinose, galactose, mannose, and furfural content determined by HPLC), and cellulose (defined as glucose, cellobiose, and 5-hydroxymethylfurfural content determined by HPLC) after GVL-assisted acidolysis pretreatment over each biomass type.



**Figure 4.2:** Analysis of hydrolysate composition obtained from P39 poplar, NM6 poplar, switchgrass, and sorghum over different GVL pretreatment conditions. Standard deviations are under 2% for all

measurements. DL, delignification step; DS, desaccharification step. Mass composition analyses for the feedstocks is shown in Table S4.1.

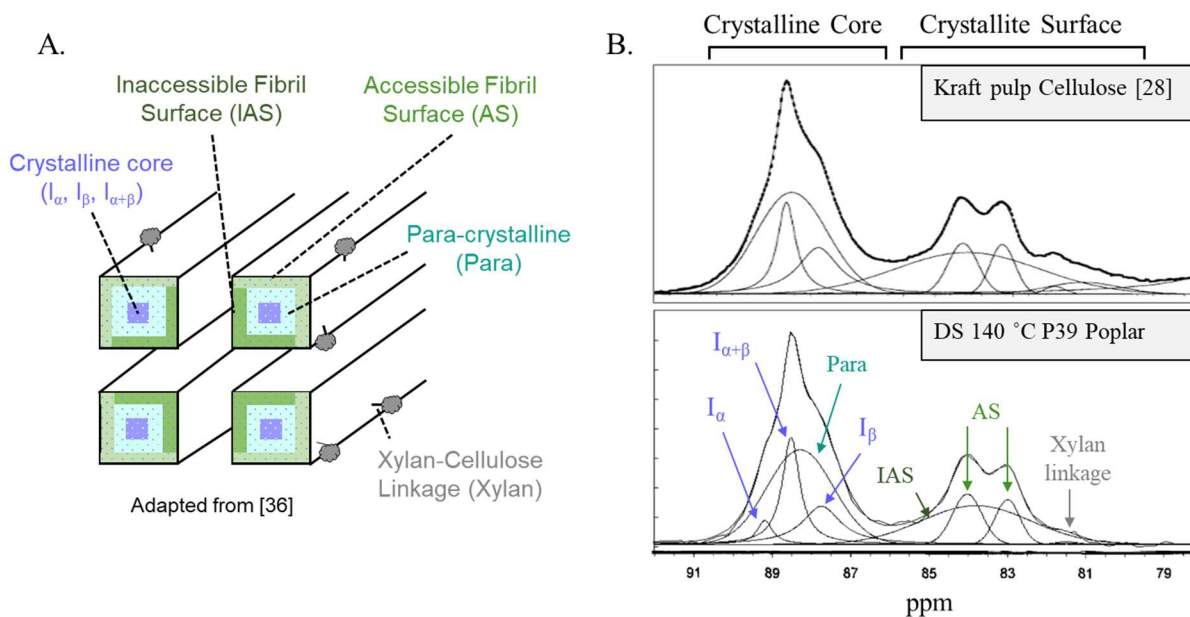
The amount of solubilized lignin in the hydrolysate solution increases with increasing overall treatment severity for all feedstocks. Poplar samples show lower solubilized lignin from the delignification step alone than for the grass samples. Interestingly, sorghum shows over twice the amount of lignin released as compared to poplar. Significant amounts of solubilized lignin are also yielded from desaccharification at elevated temperature; however, this lignin quality is expected to be lower (more condensed) than that released in the delignification step (100°C), as is reported with other higher temperature or more severe solvent systems.<sup>10,13,18,34,35</sup> When pretreated at 160°C, nearly all lignin is released across all feedstocks. This behavior is expected since no residual solid biomass is present after this high temperature treatment.

A substantial amount (~50%) of hemicellulose was hydrolyzed to soluble sugars (xylose, arabinose, galactose, mannose, and furfural) at the delignification temperature (DL, 100°C) across all feedstocks. The desaccharification step yielded additional hydrolysis of hemicellulose content across all biomass types, with best yields found at 120°C. In most cases (P39 poplar being the exception), the hemicellulose yields decreased at higher temperature, due to degradation of hemicellulose to undesirable byproducts and humins. Small amounts of cellulose are hydrolyzed to soluble C<sub>6</sub> sugars in the delignification step, with the yields increasing with temperature. The decrease in yield of soluble C<sub>6</sub> sugars at 160°C is expected due to conversion to various undesired products including levulinic acid, formic acid, and further degradation to humins.<sup>30</sup> The amount of C<sub>6</sub> sugars observed from the different feedstocks are roughly equivalent, except for sorghum, which shows nearly 3× concentration of solubilized C-6 sugars after the DL step and 2× the total C<sub>6</sub> sugars after both pretreatment steps. Notably, the sorghum

feedstock is consistently yielded the highest solubilization of lignin and hydrolysis of hemicellulose and cellulose at almost every condition examined. The behavior of the grass feedstocks is considered in the Discussion section so that additional results can be focused on behavior of the poplar feedstocks.

### *CP/MAS NMR of Residual Solids*

Figure 4.3A shows a schematic cross section of a cellulose I fibril containing 4 microfibrils. Each microfibril is composed of many parallel glucose chains, shown from an end-on view as blue dots. Within the crystalline core of a microfibril, depicted as the purple region at the center of each microfibril,  $I_{\alpha}$  and  $I_{\beta}$  represent triclinic and monoclinic unit cell packings of crystalline cellulose, and  $I_{\alpha+\beta}$  a mixture of these. Among the crystalline packings, the  $I_{\beta}$  monoclinic structure is thermodynamically most stable.<sup>11,22</sup> The crystalline core is surrounded by a para-crystalline domain (Para, teal), which is less ordered and possesses higher molecular mobility than the more crystalline regions.<sup>26</sup> The fibril also has accessible surface (AS, light green), inaccessible surface (IAS, green) buried by packing of the microfibrils, and linkages between xylan and cellulose (Xylan-Cellulose, gray spheres). Each of these substructures can be associated with  $C_4$  resonances in the CP/MAS NMR spectra (Figure 4.3B), and as shown below, can be differentially altered by GVL treatment, temperature, and enzyme hydrolysis.



**Figure 4.3:** A. A proposed structure of cellulose within the microfibril adapted from literature.<sup>36</sup> B. Comparison of peak identification from literature depiction spectral fitting of Kraft pulp cellulose from Hult et al.<sup>28</sup> and labelled spectral fitting of the C4 carbon center of DS 140 treated P39 poplar from this work.

Figure 4.3B, top panel shows literature results from spectral fitting of C<sub>4</sub> resonance peaks in the CP/MAS NMR spectra of cellulose microfibrils in Kraft pulp<sup>36</sup>, while Figure 4.3B, bottom panel shows comparable fitting for a GVL-treated P39 poplar sample from this work (comparable spectral fitting for GVL-treated NM6 poplar are shown in SI). Based on previous spectral fitting of the overall C<sub>4</sub> resonance of the NMR spectra<sup>27,29,36,37</sup>, we were able to identify several unique structural elements in GVL poplar. Table 4.1 lists the chemical shifts and full width half maximum values obtained from peak fitting of the cellulose fibril domains in P39 and NM6 poplar subjected to GVL delignification (DL, 100°C) and desaccharification (DS, 140°C). The spectral fitting procedure is described in greater detail in the Supporting Information section, Figure S4.1, and Figure S4.2. Changes in fractional areas of these resonances resulting from GVL treatment and enzyme hydrolysis are reported next.

**Table 4.1:** Chemical shifts (ppm) and full width half maxima (FWHM) for C4 peak assignments of P39 and NM6 poplar residual celluloses after chemical pretreatment with GVL-water co-solvent mild acidolysis.

Cellulose Fibril Domains	P39 Assignments				NM6 Assignments			
	ppm	+/- <sup>a</sup>	FWHM	+/-	ppm	+/-	FWHM	+/-
I <sub>α</sub>	89.12	0.01	0.46	0.03	89.16	0.09	0.47	0.00
I <sub>α+β</sub>	88.44	0.01	0.53	0.03	88.44	0.02	0.53	0.00
Para-crystalline	88.25	0.03	2.00	0.00	88.30	0.00	2.00	0.00
I <sub>β</sub>	87.67	0.02	1.18	0.00	87.60	0.00	1.18	0.00
Accessible fibril surface	83.93	0.02	0.88	0.00	83.93	0.01	0.87	0.00
Inaccessible fibril surface	83.66	0.03	3.54	0.54	83.58	0.02	3.80	0.36
Accessible fibril surface	82.94	0.02	0.74	0.00	82.93	0.03	0.74	0.00
xylan-cellulose	81.40	0.12	0.87	0.23	81.47	0.12	1.03	0.39

<sup>a</sup> Errors shown are standard error.

Compared to the reference spectrum of Kraft pulp, GVL-treated P39 poplar also contains I<sub>β</sub> and I<sub>α+β</sub> fractions and an additional I<sub>α</sub> fraction in the crystalline core (Figure 4.3B). These peaks were fit with a Lorentzian distribution to reflect the high degree of organization that results in a narrow distribution of isotropic shift and a signal dominated by life-time line broadening.<sup>29</sup> The surrounding para-crystalline fraction, centered at ~88 ppm, was fit with a Gaussian distribution reflecting the higher degree of disorder in this cellulose form. The broad resonance between 80 and 86 ppm consists of two sharp, Gaussian peaks correlated to two distinct, inequivalent accessible fibril surfaces (Figure 4.3, AS).<sup>29</sup> A broad, Gaussian resonance overlapping the accessible surface peaks represents cellulose from inaccessible fibril surfaces (Figure 4.3, IAS), which has been shown to be largely unaffected by various solvent-based degradation



methods.<sup>27,36</sup> Compared to the Kraft pulp reference spectrum, the GVL P39 poplar sample contains a substantially depleted xylan-cellulose fraction centered at 81 ppm, representing the hydrogen bonded interaction between xylan and cellulose microfibrils.<sup>27,38</sup>

#### *Enzymatic hydrolysis of residual cellulose and $X_{fit}$*

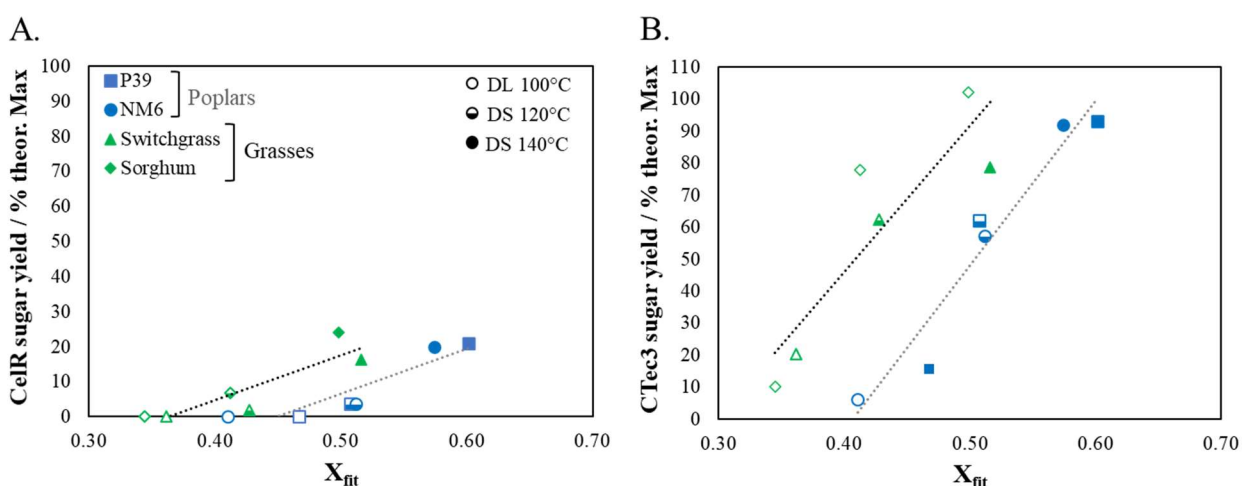
Our previous work established a linear relationship<sup>20</sup> between increasing  $X_{NMR}$ , defined as the ratio of the integrated area between ~90 - 86 ppm to the integrated area of the total C<sub>4</sub> resonance, and increasing sugar yields from enzyme hydrolysis. Physically,  $X_{NMR}$  represents a loosely defined relationship between the core domains (crystalline and para-crystalline) to the total composition of cellulose sample. This linear relationship held for P39 poplar samples undergoing GVL treatment with different co-solvents, as well as for different biomass feedstock undergoing delignification at 100°C followed by desaccharification at 140°C.

Here we improve on the predictive power of  $X_{NMR}$  by using spectral fitting and literature-defined parameters to determine the fractional areas of crystalline, para-crystalline, and surface domains contributing to structure-informed ratio named  $X_{fit}$ ,

$$X_{fit} = \frac{Area(I_{\alpha} + I_{\beta} + I_{\alpha+\beta} + para)}{Area(I_{\alpha} + I_{\beta} + I_{\alpha+\beta} + para + IAS + AS + xylan)}$$

For these enzyme hydrolysis studies, P39 and NM6 poplar, sorghum, and switchgrass were subjected to GVL-assisted delignification at 100°C and then desaccharification at either 120 or 140°C. The solid remaining from the treated samples were washed, filtered, and subjected to enzymatic hydrolysis using a single enzyme, CelR from *H. thermocellum*, or CTec3, an industrially optimized multi-enzyme mixture (e.g. cellulases, hemicellulases, glucosidases, etc).

CelR is a member of glycoside hydrolase family GH9 from *H. thermocellum*, where the natural enzyme is part of the cellulosomal complex. For this work, an engineered enzyme consisting of the naturally occurring GH9 catalytic and carbohydrate binding module CBM3c was fused to a C-terminal CBM3a. CelR preferentially converts hexosaccharides into cellobiose.<sup>39</sup> CelR produces various soluble oligosaccharides from reaction with amorphous cellulose, and lesser reactivity with Avicel (a processed form of cellulose).<sup>39</sup>



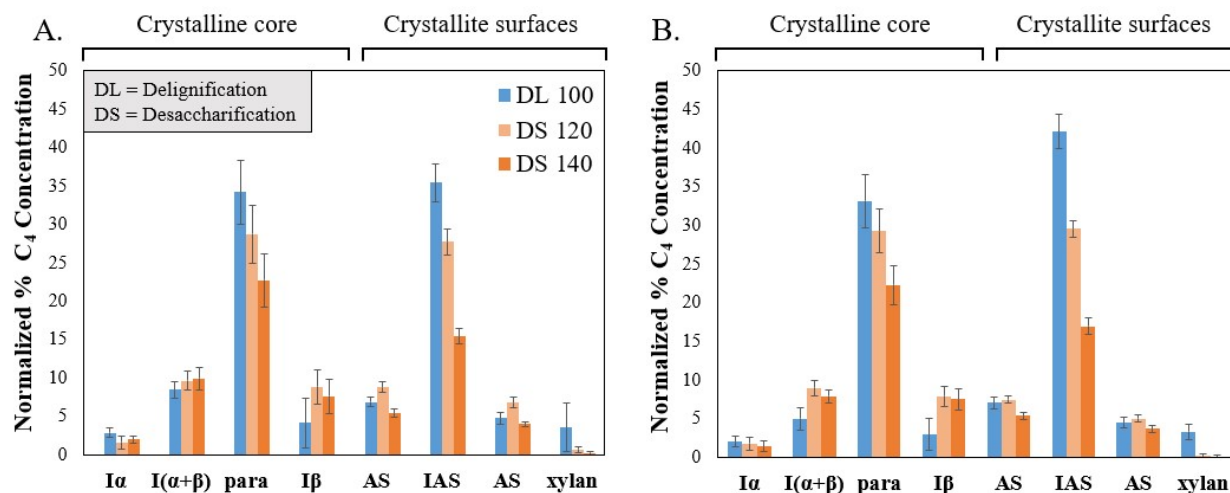
**Figure 4.4:**  $X_{fit}$  vs. theoretical maximum of enzymatic sugar yield using A. CelR and B. CTec3 out of the percent of theoretical maximum cellulose for P39 poplar, NM6 poplar, sorghum, and switchgrass after various GVL hydrolysis pretreatments. The theoretical maxima for cellulose sugar yields are calculated out of the available cellulose fraction measured by composition analysis for each GVL-treated sample (see Table S4.2). Linear fit parameters for A. poplar: slope = 127.3,  $R^2=0.82$ , grasses: slope = 138.0,  $R^2=0.76$  and B. poplar: slope = 514.04,  $R^2=0.93$ , grasses: slope = 458.6,  $R^2=0.78$ . Enzymatic hydrolysis conditions: 1 mg/mL GVL-treated biomass and 0.1 mg/mL CelR or CTec3 (10 wt.% cellulase) in 0.1 M phosphate buffer, pH 6.0, 1 mL total volume, 24 h reaction time at 50 °C with shaking. Unfilled, half-filled, and filled symbols represent the DL 100°C, DS 120°C, and DS 140°C treatment conditions, respectively.

Figure 4.4 shows linear correlations between  $X_{fit}$  and improved enzyme hydrolysis for P39 and NM6 poplar, switchgrass, and sorghum. All samples that underwent only delignification give a  $X_{fit}$  value less than 0.5. P39 and NM6 poplars show close agreement in  $X_{fit}$  as the treatment

severity is increased, yielding up to 21% of the theoretical maximum sugar yield by CelR (Figure 4.4A) after desaccharification at 140°C. While both poplar samples lay on a similar trendline, both grasses examined in this work, 2019 crops of sorghum and switchgrass, show considerable offset from the poplar samples. For grass samples, the sugar yields show a consistent shift to lower  $X_{\text{fit}}$  values. Across all samples, no enzymatic activity was observed from CelR after the DL treatment at 100 °C, and less than 10% soluble sugar was obtained from CTec3 (Figure 4.4B). The multi-enzyme CTec3 blend was able to provide complete hydrolysis of poplar samples at  $X_{\text{fit}}$  value of ~0.6 (achieved at 140°C) and of grass samples at  $X_{\text{fit}}$  ~0.5 (achieved at 120°C.). For comparison, the single enzyme CelR achieve 20-27% hydrolysis at these same  $X_{\text{fit}}$  values. Linear, overlapping trends of P39 and NM6 poplar are still observed for enzymatic conversion over CTec3, as well as for switchgrass and sorghum.

The application of the  $X_{\text{fit}}$  analysis to the yields from CelR hydrolysis expands upon previously reported enzymatic reactivity with GVL-treated substrates<sup>20</sup> and presents the additional potential to identify specific cellulose substructures contributing to recalcitrance by monitoring each domain throughout the pretreatment process.

#### *Spectral Fitting of CP/MAS NMR Spectra from GVL-Poplar Treatment*



**Figure 4.5:** Changes of fitted cellulose domains resulting from GVL pretreatments of A. P39 and B. NM6 poplar. The percent of C<sub>4</sub> concentration in each domain is out of the total carbon (area, 100% total) present in the C<sub>4</sub> resonance. Concentrations of each domain are then normalized to the cellulose content remaining compared to the total cellulose content of the DL 100 sample (determined by the HPLC analysis, see Table S4.4 for scaling factors) and reflect the conversion of cellulose from hydrolysis reactions and/or structurally rearrangement during the GVL treatment. See Figure 4.1 for GVL-assisted hydrolysis conditions.

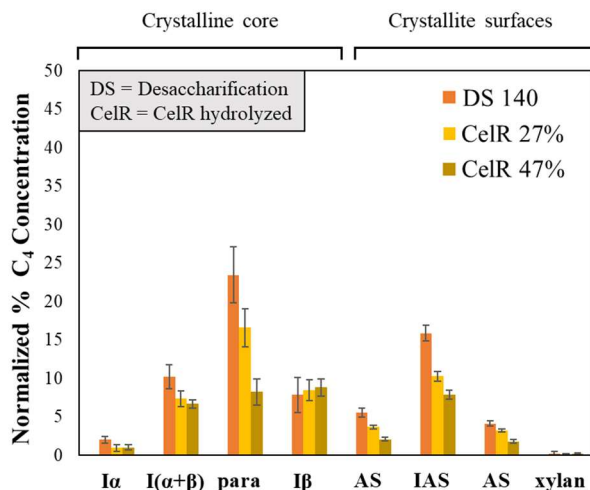
Figure 4.5 shows changes in relative proportions of the cellulose fibril domains from Table 4.1 after GVL treatment of P39 and NM6 poplar by peak fitting following procedures outlined in the Supporting Information section. Figure 4.5 demonstrates that changes in the cellulose structure throughout hydrolysis with the GVL pretreatment are detectable as changes in the proportions of microenvironments within the cellulosic microfibrils. A decrease in the normalized % C<sub>4</sub> concentration indicates hydrolysis of a domain by the GVL co-solvent system, whereas an increase suggests the formation of more cellulose within a domain. We note that the normalization adjusts the total peak area to reflect the mass balance as the samples undergo cellulose conversion. The sum of all domains for each treatment condition sum to the total cellulosic carbon remaining compared to the total carbon in the DL 100 sample (e.g. all domains

for the P39 DL 100 sample sum to 100%, while P39 DS 120 domains sum to 93%, see Table S4.4)

Large changes were observed within some of the microfibril domains during pretreatment. Xylan units disappear after the 120°C desaccharification, suggesting that xylan bound to the surface of the cellulose microfibril is solubilized and removed by the GVL solution at this pretreatment condition. The resonance peak assigned to inaccessible surface area also decreases throughout each GVL hydrolysis step. The para-crystalline region consistently decreases throughout GVL-assisted hydrolysis. Although little change in the normalized proportion of the  $I_{\alpha}$  and  $I_{\alpha+\beta}$  core populations are observed during the pretreatment, the  $I_{\beta}$  peak shows a significant increase after DS 120°C. Hydrolysis with the mildly acidic GVL-water co-solvent completely hydrolyzes all cellulose domains, including  $I_{\beta}$ , at 160°C for 1 h, as no residual solid fraction remain after this treatment, and unwanted degradation products are produced.

These trends were further confirmed for the GVL hydrolysis of NM6 poplar. Figure 4.5B depicts how the GVL hydrolysis of NM6 poplar mirrors the decreasing inaccessible and para-crystalline populations throughout increasing treatment severity seen in P39 poplar. This analysis was not possible on grass feeds, as the NMR spectra of the crystalline regions overlapped too significantly to differentiate  $I_{\alpha}$  and  $I_{\beta}$  core regions to a reasonable degree of certainty.

*Spectral Fitting of CP/MAS NMR Spectra from CelR Hydrolysis of GVL-Poplar*



**Figure 4.6:** Changes in fitted cellulose domains in GVL pretreated P39 poplar after enzymatic hydrolysis by CelR (0, 27, and 47% conversion). The percent of  $C_4$  concentration in each domain is out of the total carbon (area) present in the  $C_4$  resonance. Concentrations of each domain are then normalized to the cellulose content remaining compared to the total cellulose content of the DL 100 sample (determined by the reducing sugar colorimetric assay, see Table S4.4 for scaling factors) and changes in peak area reflect the conversion of cellulose from hydrolysis reactions and/or structural rearrangement during the GVL treatment. Enzymatic hydrolysis conditions: 20 mg/mL GVL-treated biomass and 2 mg/mL CelR (10 wt.% CelR) in 0.1 M phosphate buffer, pH 6.0, 30 mL total volume, 24 h or 72 h reaction time at 70 °C with no shaking.

$C_4$  carbon center spectral fitting was performed on GVL-treated P39 biomass samples after desaccharification at 140°C and hydrolysis by CelR. This GVL pretreatment condition was chosen because it was most readily hydrolyzed by CelR, depicted in Figure 4.4. DS 140°C samples before and after enzymatic digestion were analyzed to understand how CelR attacks and proceeds to convert GVL pretreated solids, as well as why CelR did not achieve 100% conversion of the pretreated cellulose in previous investigations. As described previously, the total  $C_4$  spectral area was normalized to the cellulosic carbon content of the DL 100°C biomass sample, and a loss of concentration reflects conversion of cellulose from that domain by CelR.

Upon 27 and 47% hydrolysis of DS 140°C solids with CelR, all regions within the cellulose microfibril show evidence of conversion except the I $\beta$  region. The para-crystalline region within

the microfibril cores and the inaccessible fibril surfaces appear to undergo the greatest conversion. Interestingly, the “least ordered” crystalline domain, para-crystalline cellulose, shows the most significant hydrolysis, followed by inaccessible surface cellulose. A significant decrease in the accessible surface domain was observed at the higher percentage of enzymatic conversion. All other crystalline regions showed moderate digestion.

**Table 4.2:**  $X_{fit}$ , Lateral Fibril Dimension (LFD) and Lateral Fibril Aggregate Dimension (LFAD) estimations for P39 and NM6 poplar based on NMR populations.

Treatment	P39 Poplar			NM6 Poplar		
	$X_{fit}^a$	LFD <sup>b</sup> (nm)	LFAD <sup>c</sup> (nm)	$X_{fit}^a$	LFD <sup>b</sup> (nm)	LFAD <sup>c</sup> (nm)
DL 100	0.47	4.0	18.5 <sup>d</sup>	0.43	3.4	18.5 <sup>f</sup>
DS 120	0.51	4.2	12.9	0.53	4.2	15.8 <sup>g</sup>
DS 140	0.60	5.5	15.8 <sup>e</sup>	0.62	5.1	15.8
CelR 27%	0.64	6.1	16.1	--	--	--
CelR 47%	0.68	6.3	18.6	--	--	--

<sup>a</sup> Standard errors are less than 0.02

<sup>b</sup> Standard errors are less than 0.12 nm

<sup>c</sup> Standard errors are typically less than 1.0 nm unless noted otherwise

<sup>d-g</sup> Standard error of d: 3.8 nm, e: 2.1 nm. f: 4.2 nm, g: 1.6 nm

In Table 4.2,  $X_{fit}$  values are compared after all GVL-assisted treatment conditions and after enzymatic hydrolysis conditions examined in Figure 4.5-Figure 4.6.  $X_{fit}$  parameters increase with increased GVL pretreatment severity for both poplar samples, in agreement with literature trends of increasing cellulose crystallinity after enzymatic hydrolysis of organosolv pretreated Loblolly pine.<sup>15</sup> Both poplar samples reach a maximum  $X_{fit}$  parameter of ~0.60 for the 140°C

desaccharification. Upon reaction with CelR, crystallinity estimates through  $X_{\text{fit}}$  increases further for P39 poplar to 0.68.

Estimations regarding the fibril geometry and dimensions can be calculated with the fitted  $C_4$  domain compositions by assuming a square cross section of the cellulose fibril aggregate, described in detail in the literature.<sup>27,40</sup> In brief, concentrations of domains can be correlated to cellulose microfibril and fibril aggregate dimension by assuming a standard microfibril geometry, in this case assuming a square fibril cross section as applied in Wickholm et al.<sup>27</sup> For a square cross section, the fraction of signal intensity of cellulose at the fibril surface,  $q$ , is

represented by

$$q = 4n - 4n^2$$

(Eq. 1:

$$q = \frac{4n-4}{n^2} \quad (\text{Eq. 1})$$

In this equation,  $q$  can represent the fraction of accessible surface of cellulose or the fraction of all surfaces (accessible and inaccessible) to the bulk. When  $q$  represents the former, solving for  $n$  constitutes the estimated number of cellulose chains along the full width of the fibril aggregate (shown in Figure 4.3A), and when  $q$  represents all accessible and inaccessible surfaces,  $n$  estimates the number of cellulose chains along the length of a microfibril within the aggregate (one quarter of the cross section of Figure 4.3A). The width of a cellulose strand was assumed to be 0.57 nm.<sup>40</sup>

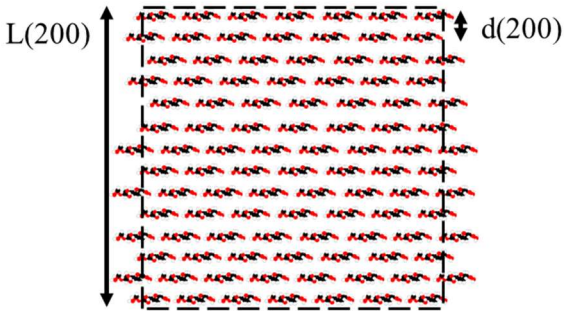
Fibril dimensions estimated for P39 and NM6 poplar range between 3.4 and 6.5 nm. These values fall within ranges reported in the literature for cellulose type I, which is common to higher plants.<sup>32</sup> Lateral fibril dimension of both poplar samples increases as the GVL hydrolysis becomes more severe, and increases further after enzymatic hydrolysis by CelR.



After desaccharification is performed on the delignified sample, the lateral fibril aggregate size decreases despite an increase in lateral fibril dimension. The P39 poplar shows a larger decrease in the aggregate dimension at DS 120°C, while both P39 and NM6 samples show the same aggregate dimension of 15.8 nm after treatment at DS 140°C conditions. After the DS 140°C sample is hydrolyzed by CelR, the estimated fibril aggregate dimension increases slightly.

### Wide Angle X-ray Scattering Analysis

**Table 4.3:** Segal Crystallinity Index (CrI), coherence length (L), and lattice spacing (d) for the (200) plane of cellulose microfibril and a schematic of cellulose microfibril cross section, denoting lattice spacing and coherence length for the (200) plane, adapted from [41,42].

P39	CrI <sup>a</sup>	L(200) <sup>b</sup>	d(200) <sup>c</sup>	L(200)	d(200)
		(nm)	(Å)		
DL 100	0.42	4.95	4.05		
DS 120	0.53	5.18	4.03		
DS 140	0.63	5.39	4.01		

<sup>a</sup> Errors are less than 0.006

<sup>b</sup> Errors are less than 0.04 nm

<sup>c</sup> Errors are less than 0.003 Å

While solid-state NMR measures populations of carbon spins within different magnetic environments across the bulk residual cellulose sample, wide angle X-ray scattering (WAXS) measures changes in lattice spacing between planes within a specific crystalline cellulose matrix. Segal crystallinity index (CrI) measurements from WAXS analysis on GVL-treated, P39 poplar samples confirmed trends of increasing crystallinity observed from solid-state NMR results. CrI values show strong agreement with the values determined from  $X_{fit}$ .

Table 4.3 depicts WAXS measurements of the lattice spacing,  $d$ , and coherence length,  $L$ , in the (200) plane of crystalline cellulose. The corresponding equations to calculate these measurements are described in the SI. The results show statistically significant growth of the lateral coherence length. Additionally, measurement of the lattice spacing in the (200) plane shows a statistically significant decrease from 4.05 to 4.01 Å in the space between parallel sheets of crystalline cellulose.

### **Discussion:**

This work extends our previous findings from solid-state nuclear magnetic resonance (NMR) spectroscopy to identify defining features of residual cellulose in various solvent-pretreated biomasses and correlate them with enzymatic sugar yields. Additional quantitative and structural information has been obtained by definition of a parameter  $X_{\text{fit}}$ , which incorporates changes in spectral properties of different allomorphs of cellulose, and permits tracking of changes in their relative proportions during GVL treatment.

Both poplar and grasses were effectively deconstructed by GVL treatment, albeit with some difference in properties and offset in optimal temperature for GVL treatment needed. The trends displayed in Figure 4.2 demonstrate that extraction of the three major fractions within biomass is optimized at different temperature of the GVL pretreatment. Hemicellulose is the easiest major fraction to remove but succumbs to degradation at moderate desaccharification temperatures (140°C). Furthermore, as evidenced by the low amount of solubilized C<sub>6</sub> sugar, the desired cellulose yields are maximized at moderate desaccharification temperatures (140°C). As with hemicellulose, cellulose can be fully solubilized with 160°C treatment, albeit with mounting consequences of thermal degradation on the yields of soluble sugars. Clearly, there is no one

temperature for optimized extraction of all solubilized fractions for the GVL pretreatment conditions examined in this work, and optimal conditions must be selected based on the desired fraction for solubilization.

Given this limitation and the significant degradation at full solubilization (DS 160°C), a partial solubilization of GVL treated biomass combined with enzymatic hydrolysis of remaining solids was determined the most promising method to increase overall sugar yields. To explore the combined process of GVL and enzymatic hydrolysis in more detail residual solids from DL 100, DS 120, and DS 140°C treatments were analyzed using solid-state NMR spectroscopy and underwent enzymatic hydrolysis in hopes of attaining full biomass fractionation and understanding the mechanism of cellulose degradation.

Comparison of increasing  $X_{\text{fit}}$  values to increasing enzymatic yield of enzymatic sugars in Figure 4.4 show separate trendlines describing the behavior of poplar and grass samples. As the temperature of GVL treatment increased,  $X_{\text{fit}}$  also increased, corresponding to an increase in crystallinity of the residual cellulose samples. Similar trends of increasing crystallinity with increasing treatment severity were noted with dilute acid hydrothermal pretreatment of *Populus* and switchgrass<sup>31</sup> as well as pine sawdust,<sup>12</sup> due to preferential hydrolysis of non-crystalline material and/or recrystallization to different cellulose domains or allomorphs. The composition analysis of the sorghum feedstock used here shows unusually high extractives content (around 30 wt.%) for sorghum (Table S4.1), which likely influenced the high hydrolysis yield with CelR at high temperatures. These results highlight that, despite the previous hypothesis that all biomass samples lie along the same trendline, we cannot solely rely on a single linear equation to predict the enzymatic yields from  $X_{\text{fit}}$  for all biomass material as we expand to a wider array of treatment conditions over different biomass feedstocks. However, the slopes of the trendlines are similar

for different biomass feeds using CelR samples (127 for poplar and 138 for grasses) as well as for Ctec3 (514 for poplar and 459 for grasses).

A considerable recalcitrance appears to exist for GVL-pretreated biomass samples when undergoing enzymatic conversion by CelR (Figure 4.4A). To examine if this recalcitrance was observed for other cellulase systems, a commercial enzyme cocktail, CTec3, was examined under the same reaction conditions (Figure 4.4B). CTec3 contains a mixture of cellulases that allowed for higher conversions of residual GVL-pretreated materials, likely through multiple deconstruction mechanisms exhibited by various cellulases in the mixture. Importantly, the near quantitative sugar yields after DS 140 treatment show GVL-pretreated biomass samples can be readily hydrolyzed by the stronger enzymatic cocktail.

While identical enzymatic hydrolysis conditions with CelR yielded significantly less sugar than with CTec3 (roughly 25% of DS 140°C samples were hydrolyzed by CelR), CelR still converted roughly 25% of sugars in the DS 140°C sample as a single, genetically engineered exocellulase, without the aid of any additional enzymes. This result suggests that CelR could be implemented in a minimal enzymatic cocktail system, where a small number of enzymes are employed to achieve similar activity of more complex enzymatic cocktail systems like CTec3 that can contain significantly more cellulases. However, investigation into the mechanism of deconstruction for CelR is needed to understand how the engineered cellulase hydrolyzes cellulosic matter within heterogeneous cellulose sources. This insight could determine limitations of CelR in hydrolyzing complex biomass sources that are candidates for industrial biomass deconstruction, like those examined here.

To probe this question, further analysis of the GVL and enzymatic residual cellulose samples were carried out to explore structure property relationships of cellulose recalcitrance throughout

both the GVL pretreatment and CelR hydrolysis steps. We note that switchgrass and sorghum samples produced NMR spectra that were too broad and lacked defining features for further spectral fitting analysis due to significant overlap of the para-crystalline and inaccessible peaks with the more crystalline resonances. This challenge has been reported elsewhere in the literature.<sup>43</sup>

In Figure 4.5, relative reductions in the proportion of the xylan-cellulose linkage region after desaccharification at 120°C agrees with the hydrolysate analysis in Figure 4.2, which shows near complete yields of xylan-derived sugars in solution after the same treatment condition. The loss of xylan content at DS 120°C also aligns with when the pulp first becomes enzymatically digestible to CelR, visible in Figure 4.2A. The relative depletion of inaccessible cellulose shows that the GVL co-solvent system readily accesses domains within the cellulose microfibril which are reported to be inaccessible to other solvents systems like methanol and acid hydrolysis with 2.5 M HCl at 100°C.<sup>27</sup> The increase of the relative accessible surface area after the lowest desaccharification treatment, followed by a subsequent decrease of relative accessible surface is not fully understood from this work but may reflect dissolution of fibrils and subsequent re-aggregation or recrystallization behavior.

The increase in the most ordered domain,  $I_{\beta}$  cellulose, despite the decrease in other crystalline domains (i.e. para-crystalline) suggests that the stronger Van Der Waals forces and hydrogen bonds that hold  $I_{\beta}$  together likely contribute to this region being the last domain to be hydrolyzed by the co-solvent system. Indeed, the  $I_{\beta}$  population remains relatively unconverted even after treatment at 140°C. Similar increases in  $I_{\beta}$  crystalline cellulose have been reported with dilute-acid pretreatments in water and water-ethanol organosolv.<sup>12,15</sup> Furthermore, recrystallization is evidenced throughout GVL pretreatment by the increase in the  $I_{\beta}$  domain after

normalization of the cellulose content in each sample. The increase in the  $I_{\beta}$  region suggests that more of this phase is being formed throughout treatment than was initially present in the DL 100°C sample. Therefore, the shift to higher crystallinity with increasing treatment severity is not solely due to the removal of more labile domains. Interestingly, the increase in sample crystallinity via the formation of new  $I_{\beta}$  cellulose at DS 120°C is also observed for residual NM6 poplar solids.

Regarding the mechanism of the GVL pretreatment on residual biomass samples, we note that if the degradation mechanism were radial, one could expect a continuous increase in the accessible surface population relative to the total  $C_4$  peak as the hydrolysis severity is increased. This trend was not observed in our investigation, as seen in Figure S4.3. Our results suggest that inaccessible domains “opened” during hydrolysis do not become magnetically equivalent to accessible surface regions, as evidenced by the lack of increase in accessible domains after the DS 140°C treatment despite significant loss of inaccessible peak populations.

The spectral fitting of CelR hydrolyzed solids (Figure 4.6) showed that CelR most readily digests lesser bound, less crystalline cellulose within a fibril, which is consistent with previously reported reactivity with amorphous cellulose and crystalline Avicel (Table S4.3). Pu et al. similarly reported lower susceptibility of  $I_{\beta}$  cellulose to enzymatic hydrolysis when compared to  $I_{\alpha}$ , para-crystalline, and non-crystalline regions by a similar, commercial cellulase.<sup>22</sup> When the same cellulose domains are plotted along an axis comparing relative  $C_4$  proportions that are not normalized by cellulose content, the relative accessible fibril surface concentrations increase slightly, but still remain relatively constant (Figure S4.4). This behavior suggests that enzymatic conversion of the cellulose fibrils by CelR is not limited to the outermost regions of the fibril but is likely occurring at both inaccessible and para-crystalline domains within a microfibril

aggregate. Additionally, the constant ratio of accessible surface to total C<sub>4</sub> cellulosic carbon suggests end-on hydrolysis, where the width of the CelR digested fibrils does not significantly change. If CelR digested the residual cellulosic fibril aggregates via radial degradation, we should expect an increase in relative accessible surface as the ratio of surface area to total volume of the fibril grows.

A further examination of DS 140 solids was completed under similar degrees of enzymatic conversion using CTec3 to determine if a preferential conversion of the I<sub>β</sub> region was observed at similar degrees of conversion (Figure 4.4B). Lower loadings of CTec3 were used than in earlier CTec3 hydrolyses (Figure 4.4) to ensure that sufficient solids remained for analysis after the reaction. At 24 and 36% conversion of cellulose by CTec3, few differences were noted between the spectral fitting of CTec3 converted samples, indicating that the I<sub>β</sub> region is not preferentially digested by CTec3 at these conditions (i.e. loadings and reaction times), and that conversion of this domain likely continues after removal of other, more labile domains (Figure S4.5).

From these results, it appears that the GVL pretreatment “clears” the surface of cellulose from xylan and extra-cellulosic material that hinder digestion by CelR. However, GVL pretreatment also appears to yield a more recalcitrant cellulose packing within the crystalline core, I<sub>β</sub>, which is apparently recalcitrant to CelR. This process prevents full conversion of cellulose to smaller sugars by CelR alone. However, enzymatic cocktails, such as CTec3, which contain diverse mixtures of enzymes, can readily digest all residual cellulosic domains, including I<sub>β</sub> at high loadings and sufficient reaction times, that appear otherwise recalcitrant to CelR.

The growth of the lateral fibril dimension (LFD) in Table 4.2 suggests a possible restructuring via coalescence of cellulose chains, where fibril widths grow as hemicellulose and lignin material are removed from the cellulosic matrix. Van der Waals and hydrogen bonds along

the cellulose chain are unoccupied and available to interact with the -OH bonds along the now accessible cellulose fibril surfaces. This change can allow for partial relaxation of fibril distortions, as well as more continuous ordering between fibrils. However, our results cannot rule out a deconstruction mechanism that begins on fibrils with a smaller lateral dimension, thereby increasing the remaining bulk average lateral length as conversion proceeds. Similar increases in lateral fibril dimensions throughout hydrothermal<sup>30</sup>, dilute acid<sup>31</sup>, and enzymatic<sup>32</sup> hydrolysis treatments have been reported elsewhere, along with an increase in fibril aggregation.

A less clear degradation behavior is observed in the lateral fibril aggregate dimension (LFAD) throughout hydrolysis when the aggregate dimension increases as the LFD decreases. Foston et al.<sup>31</sup> also reported increase in the lateral fibril dimension of dilute acid treated biomass, after a similar decrease in LFAD during initial dilute acid treatment. This behavior provides evidence for the microfibril aggregates breaking apart as non-cellulosic material is fractionated by the co-solvent-assisted pretreatment. The final increase in LFAD after desaccharification at 140 °C could be due to the structure containing less amorphous material in the crystalline lattice (as described regarding Figure 4.5), likely further encouraging microfibril aggregation.

Comparing NMR analysis with wide angle X-ray scattering (WAXS) data (Table 4.3) the crystallinity index (CrI) gave crystallinity estimates that were in agreement with  $X_{fit}$ . The growth in lateral coherence length,  $L$ , provide further support for a growing average lateral fibril dimension, similarly suggested from NMR. Furthermore, the decrease in the lattice spacing ( $d$ ) of the (200) plane is consistent with a tightening of the crystalline lattice structure within the core cellulose as the GVL treatment severity increases, causing the crystalline cores to become more tightly packed. The WAXS measurements confirmed that the average spacing within crystalline domains became more densely packed while the lateral fibril dimension in the (200) plane



increased throughout the GVL pretreatment. These results agree with the changes in proportion of decomposed peaks from solid-state NMR showing a transition in crystalline domains to the  $I_{\beta}$  population, which is a tighter core packing of crystalline cellulose. This behavior points toward a restructuring of the crystalline cellulose during chemical treatment at moderate temperatures examined in the proposed pretreatment method, as well as aggregation of cellulose strands onto previously existing microfibrils as non-cellulosic material is solubilized and less ordered cellulose is removed from the matrix.

### **Conclusions:**

In this study, we examined a mildly acidic GVL-H<sub>2</sub>O co-solvent pretreatment of two poplar and two grass feedstocks using hydrolysate analysis, solid-state NMR, and WAXS characterization techniques. A solid-state NMR crystallinity estimate,  $X_{\text{fit}}$ , was used to quantify the impacts of GVL treatment and enzymatic hydrolysis on crystalline, surface and other structural domains in cellulose.  $X_{\text{fit}}$  gave predictive, linear trendlines depending on biomass feedstock type when examining different biomass samples over a wider range of GVL pretreatment conditions. Hemicellulose sugar solubilization throughout the GVL pretreatment was favored at the lowest desaccharification temperature examined, 120°C, after a lignin removal step and cellulose sugar solubilization was favored at a higher desaccharification temperature, 140°C. At higher temperatures, these sugars were degraded to undesired products. Tracking the evolution of cellulose microfibril domains in P39 and NM6 poplar samples throughout GVL-water mild acidolysis showed the removal of xylan-cellulose linkages after delignification at 100°C. With increasing treatment severity, there was a progressive depletion of para-crystalline and inaccessible surface domains. These changes were accompanied by an increase in the core  $I_{\beta}$

domain, suggesting a restructuring to a more ordered, crystalline cellulose. WAXS analysis showed a shrinking of the lattice spacing between the (200) crystalline plane with increasing GVL pretreatment severity, indicating a tightening of the crystalline cellulose matrix. Studies of P39 poplar after enzymatic hydrolysis revealed that the most crystalline domain of the core cellulose, I<sub>β</sub>, was resistant to hydrolysis by the single enzyme CelR. In contrast, the multi-enzyme blend CTec3 was able to fully hydrolyze GVL cellulose. This work provides mechanistic understanding of how microfibrils change throughout GVL-treatment and enzymatic hydrolysis. Changes in the spectral signatures of various cellulose substructure provide insight into the origins of enzymatic cellulose recalcitrance and can guide future GVL assisted and chemical hydrolysis methods in pursuit of efficient, solvent-assisted biomass conversion and utilization.

## **Experimental Section:**

### *Materials*

5 mm debarked poplar (P39 and NM6), switchgrass, and sorghum chips were received from the Great Lakes Bioenergy Research Center (GLBRC) and underwent no additional pretreatments. 18 MΩ Milli-Q water, γ-valerolactone (GVL; Sigma, 95%), and sulfuric acid (Sigma, 98 wt.%) were obtained and used as received. CelR-CBM3a consists of the *Hungateiclostridium thermocellum* cellulase CelR (GenBank: CAE51308.1, residues 31-642, the GH9 and CBM3c domains) with an additional 48 residue CipA-native linker connecting to the 153 residue CBM3a from the CipA scaffolding protein. The expression plasmid, under lac operon control, was expressed and purified at scale by a third-party vendor.

### *Sample Preparation and GVL-water Co-solvent Acidolysis*

30 g of untreated 5 mm chips of the desired feedstock were loaded into a 450 mL Parr reactor with 300 mL of the desired ratio of GVL to water (90:10 or 80:20 wt%) and 100 mM sulfuric acid catalyst. Reactors were stirred with overhead mechanical stirring and placed in a furnace set to the desired reaction temperature. Once the system reached the set temperature, it was held for 1 h before quenching with an ice bath. Liquid and solid contents were separated with vacuum filtration through a fine mesh glass filter and washed (if needed) with excess Milli-Q water while stirring the slurry. Separated, washed solids and liquid hydrolysate solutions were analyzed with different techniques described below (NMR, HPLC, etc). For samples undergoing desaccharification, solids separated from the delignification reactor were filtered, roughly six grams of solids were removed and washed for analysis, and the remainder was loaded into the reactor without additional washing or treatment for the second step: desaccharification at 120, 140, or 160°C.

### *Hydrolysis of Solubilized Sugars, HPLC Analysis, and Lignin Quantification*

Liquid hydrolysate samples were analyzed with methods described in detail elsewhere.<sup>20</sup> Briefly, samples were diluted 10 times by weight in Milli-Q water, filtered with a 0.2 µm PTFE syringe filter, and analyzed with a Shimadzu-1020 HPLC equipped with a BioRad HPX-87H proton-based ion exclusion column and refractive index and UV-vis scanning photo diode array detectors. All diluted samples underwent an additional hydrolysis step for 1 h at 120°C in 3 wt% H<sub>2</sub>SO<sub>4</sub> to reduce depolymerized glucose and xylose oligomers. Oligomer content was added to the glucose and xylose yields to determine total soluble carbohydrate content. Lignin was precipitated from the hydrolysate solution by ten times dilution in water and followed by

centrifugation. Solids were washed with centrifugation and flash frozen before weighing to determine lignin yield.

#### *Enzymatic Hydrolysis of Residual Solids*

Reactions were prepared with 1 mg/mL GVL-treated biomass and 0.1 mg/mL of either CelR or CTec3 in 0.1 M phosphate buffer at pH 6.0 in a 1 mL total volume. Reactions were incubated for 24 h at 50°C with 1050 rpm (0.0185 g) shaking in a Heidolph Titramax 1000 using a Heidolph Inkubator 1000 to control temperature. Reactions were centrifuged for 5 min at 21130 g after incubation to pellet any residual solids and the concentration of soluble sugar was determined via the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Briefly, 100 µL of BCA working solution and 5 µL of reaction supernatant were heated at 80°C for 15 min. Control samples contained either CelR only, GVL-treated biomass only, or buffer only; and a glucose standard curve was used to convert absorbance units to mg/mL reducing sugar in solution. Reactions were prepared in quadruplet for each condition, with reaction supernatant being sampled in triplicate. High conversion CelR samples (27 and 47% conversion CelR samples) underwent reaction conditions of: 2 mg CelR/mL, 20 mg residual solids/mL, 30 mL final volume, 0.1 M phosphate buffer with pH of 6.0, 70°C, no shaking, and 24 or 72 hours incubation for 27 and 47% conversion, respectively. High conversion CTec3 samples (24 and 36% conversion CTec3 samples) underwent reaction conditions of: 0.4 mg CTec3/mL, 20 mg residual solids/mL, 25 mL final volume, pH 5.0, 0.1 M acetate buffer, 52.5°C, no shaking, and 24 or 96 hours incubation for 24 and 36% conversion, respectively.

#### *Native Biomass Analyses via NREL Mass Closure Method*

Native biomass was prepared by milling to a 60-mesh particle size using a Wiley mill. The moisture and ash content of the milled biomass was analyzed via thermogravimetric analysis (Mettler Toledo TGA/DSC 3+). The biomass was extracted with hot water followed by hot ethanol to collect and analyze-soluble components. Both extracted fractions were dried and gravimetrically weighed for total extractive composition. The water extracted fraction was further assayed to quantify the water soluble sugar composition (glucose, fructose, and sucrose) and water soluble oligosaccharide composition. Freely soluble sugars were directly assayed using HPLC analysis (Agilent Infinity 1260 equipped with an RID detector using the BioRad Aminex HPX-87P column). Oligosaccharide composition analysis was performed on the water extracted fraction using 2M TFA hydrolysis followed by the alditol acetate procedure for GC/MS analysis (Agilent 7890 / 5975C). The extractives-free biomass was oven dried at 50°C overnight and subsequently assayed for moisture and ash. The lignin content and carbohydrate composition were assayed using the two-stage sulfuric acid hydrolysis described by the NREL procedure.<sup>44</sup> Monosaccharides from the hydrolysate fraction were quantitated via HPLC analysis (Agilent Infinity 1260 equipped with an RID detector using the BioRad Aminex HPX-87P column). The HPLC vials used for the monosaccharide composition were re-assayed on the HPLC system using a BioRad Aminex HPX-87H column to determine the acetyl content. The hydrolysate fraction was assayed for the acid soluble lignin content via UV spectrometry (Molecular Devices SpectraMax Plus 384). The acid insoluble residue remaining after the second stage acid hydrolysis was collected via filtration, dried, then gravimetrically assayed. After the gravimetric analysis, the residue was further analyzed using thermogravimetric analysis (Mettler Toledo TGA/DSC 3+) to determine the acid insoluble lignin content.

*<sup>13</sup>C Cross-Polarization Magic-Angle-Spinning (CP/MAS) NMR and Proton Spin-Relaxation Editing (PSRE)*

Washed, never dried residual solid samples were analyzed with solid-state NMR techniques described in detail elsewhere.<sup>20</sup> Roughly 200 mg of solvent-pretreated sample was packed into a 4mm thin-walled silicon nitride rotor with Kel-F caps. <sup>13</sup>C spectra were acquired on a 500 MHz Bruker Avance III spectrometer with a proton radio frequency of 500.22 MHz and a <sup>13</sup>C frequency of 125.76 MHz. Rotors were spun in a 4mm Doty Scientific DI4 CPMAS probe at 4 kHz and 90° pulse lengths of 2.30 and 4.55 μs were used for <sup>1</sup>H and <sup>13</sup>C, respectively. During spin-locking, the spin-lock radiofrequency strength was  $\gamma B_H / (2\pi) = 46$  kHz for a time of 3–10000 μs, A 0.3 s acquisition time and an 8 s recycle delay between scans were implemented with a cross-polarization contact time of 2000 μs with a 70-100% ramp at a 99.53 kHz radiofrequency strength. A spinal64 decoupler<sup>45</sup> with a radiofrequency of 100 kHz was applied during acquisition. A secondary reference to the upfield adamantane peak at 28.7 ppm was used to indirectly reference the <sup>13</sup>C spectra to TMS.

*Spectral fitting of C<sub>4</sub> NMR Peak*

A non-linear, least squares solver was used in DMFit software<sup>46</sup> to fit C<sub>4</sub> cellulose subpeaks following methods described by Larsson et al.<sup>26</sup> Initial guesses for chemical shifts and peak widths were taken from fits reported in literature on similar kraft pulp samples.<sup>28</sup> DMFit applies a mismatch criterion and minimization algorithm using the quadratic distance between the experimental and modelled spectrum and iterates to a local minimum, without making additional assumptions about the dataset. This process is described in more detail in the Supporting Information section, Figure S4.1, and Figure S4.2.

### *Wide Angle X-Ray Scattering (WAXS) sample preparation, measurement, and analysis*

GVL treated poplar samples were air dried. 2.0 g of dried sample were milled with a Wiley mill at 100 mesh, then in a RETSCH CryoMill at 25 Hz for 30 s. The powders were pressed into pellets of 3 mm diameter and 0.3 - 0.5 mm thickness. Transmission WAXS was performed using a Xenocs Xeuss 2.0 instrument under vacuum. WAXS images were reduced to 1D data using Nika software.<sup>47</sup> WAXS plots were normalized by pellet thickness. Background subtraction prior to crystallinity analyses was performed using a two-point linear background. Crystalline and amorphous peaks were fitted to Gaussians. Crystallinity index was calculated by deconvoluting peaks or reflections, and is defined as the ratio of the area of crystalline cellulose peaks to the area of the entire diffraction curve.<sup>48</sup> Further description of WAXS calculations and GVL-treated sample raw data can be found in the Supporting Information (Figure S4.6).

### **Acknowledgments:**

This material is based upon work supported by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018409. The Bruker Avance III 500 NMR spectrometer was supported by the UW-Madison Bender Fund (2012) and UW2020 (2017). WAXS studies were supported by The Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by DOE, Office of Science, BES under Award # DE-SC0001090.

### **Supporting Information:**

**Table S4.1.** Mass Closure analysis on a dry basis of feedstocks for P39 Poplar, NM6 Poplar, Switchgrass, and Sorghum. Moisture analysis for feedstocks averaged 3.5 wt%  $\pm$  0.1 wt%. Mass closure analysis reflects this moisture content. Extractives content includes water and ethanol soluble extractives. Cellulose content consists of glucose, and hemicellulose content consists of xylose, arabinose, mannose, and galactose, and acetals. Lignin content is composed of acid soluble and acid insoluble lignin.

<b>biomass</b>	<b>wt%</b> <b>extractives</b>	<b>wt%</b> <b>cellulose</b>	<b>wt%</b> <b>xylan</b>	<b>wt%</b> <b>lignin</b>	<b>wt%</b> <b>Ash</b>	<b>wt%</b> <b>proteins</b>	<b>wt%</b> <b>Total</b>
<b>2018 P39 Poplar, native</b>	4.4	50.9	24.0	20.0	0.17	2.3	101.8
<b>2018 NM6 Poplar, native</b>	5.5	45.1	23.7	24.2	0.33	1.97	100.8
<b>2019 Sorghum, native</b>	29.7	28.9	23.5	12.0	4.94	5.30	104.3
<b>2019 Switchgrass, native</b>	11.4	34.1	29.8	19.0	3.31	4.48	102.2

**Table S4.2.** Composition analysis of GVL treated P39 Poplar, NM6 Poplar, Switchgrass, and Sorghum on a dry basis. All poplar values were based off the values quantified for NM6 Poplar. The cellulose fraction consists of glucose, the hemicellulose fraction consists of xylose, arabinose, mannose, and galactose, and the lignin content consists of acid soluble and acid insoluble lignin.

<b>biomass</b>	<b>wt%</b> <b>extractives</b>	<b>wt%</b> <b>cellulose</b>	<b>wt%</b> <b>xylan</b>	<b>wt%</b> <b>lignin</b>	<b>wt%</b> <b>ash</b>	<b>wt.%</b> <b>Total</b>
<b>NM6 DL100</b>	0	67.6	15.7	18.2	0.0	101.5
<b>NM6 DS120</b>	0	90.2	4.7	8.4	0.2	103.4
<b>NM6 DS140</b>	0	95.6	0.0	3.6	0.6	99.8
<b>Sorghum DL100</b>	0	67.6	14.3	9.2	3.2	94.3
<b>Sorghum DS120</b>	0	88.7	1.7	3.1	3.2	96.7
<b>Sorghum DS140</b>	0	88.8	1.0	2.1	6.4	98.3
<b>Switchgrass DL100</b>	0	60.9	14.4	15.6	2.0	92.9
<b>Switchgrass DS120</b>	0	86.5	1.8	6.8	3.0	98.1



<b>Switchgrass DS140</b>	0	86.8	1.1	5.6	5.1	98.6
--------------------------	---	------	-----	-----	-----	------

**Table S4.3:** Enzymatic reactivity of commercial substrates PASC and Avicel, representing amorphous and crystalline cellulose samples by CelR. Enzymatic hydrolysis conditions: 1 mg/mL GVL-treated biomass and 0.1 mg/mL CelR or CTec3 in 0.1 M phosphate buffer, pH 6.0, 1 mL total volume, 24 h reaction time at 50 °C with shaking. digestion with 10 wt.% CelR cellulase on 1 mg/mL GVL pretreated solids in 1mL volume for 24 h at 50 °C with shaking.

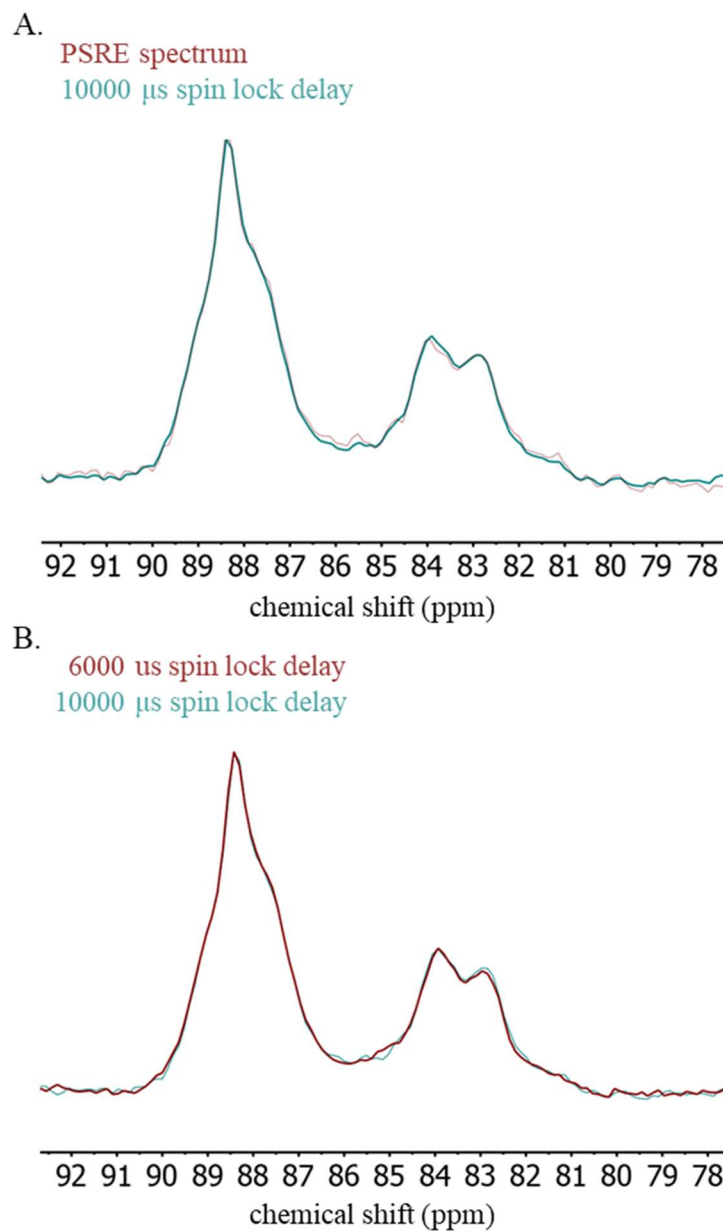
<b>Substrate</b>	<b>CelR sugar yield / % theor. Max.</b>
<b>PASC (Phosphoric Acid Swollen Cellulose)</b>	63
<b>Avicel</b>	3

*Procedure for spectral fitting of C<sub>4</sub> resonance in the <sup>13</sup>C CP/MAS NMR spectra*

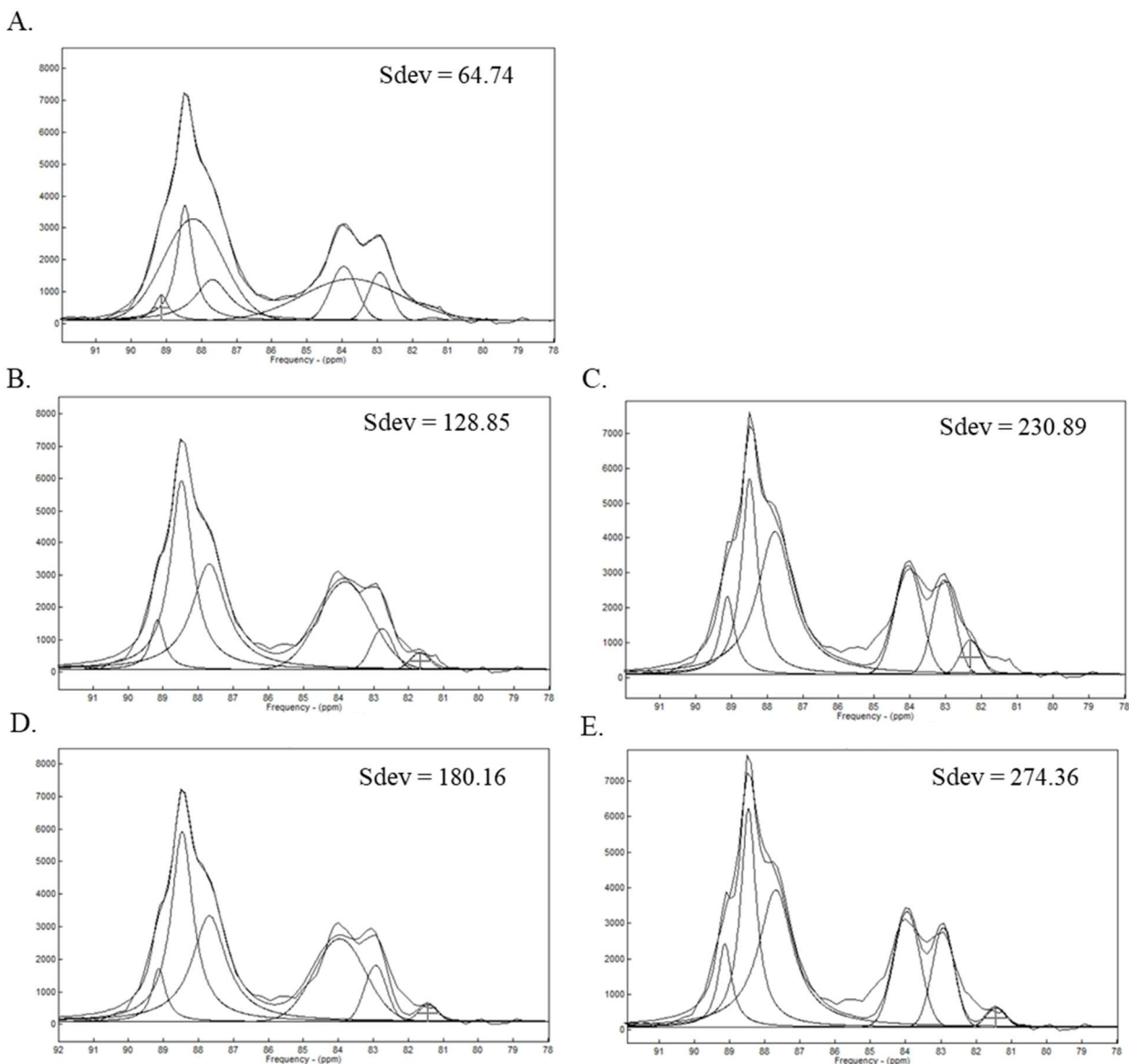
Spectral fitting was completed on the C<sub>4</sub> resonance of spectra that underwent a spin lock of either 6,000 or 10,000 μs. The PSRE method was not used for the spectral fitting analysis, since comparison of the C<sub>4</sub> resonance of the PSRE resulting spectrum to the spectra resulting from a spin lock time >6,000 μs showed line shapes that closely aligned. Furthermore, analyzing the spectra after only the delayed spin lock circumvented the decrease in signal to noise that occurred from using the PSRE subtraction method, enabling the spectral fitting. Additionally, both 6,000 and 10,000 μs spin lock times produced nearly identical line shapes for all samples (when scaled to matching amplitudes, as shown in Figure S4.1B). Peak assignments for chemical shifts, full width half maxima (FWHM), and line types from Hult et al.<sup>28</sup> were used as initial guesses. First, the crystalline peaks, accessible surface peaks, and xylan-cellulose linkage peaks were added based on similar kraft pulp sample spectra. After accounting for shifts due to different calibration methods used to assign the zero point of the chemical shift, FWHM were kept near the ranges suggested in literature (± 0.05 ppm). If a parameter was not stable and would solve to an unreasonable value or chemical shift away from the range established in

literature, this value was locked and not allowed to fluctuate. If this parameter was the chemical shift, the shift was assigned directly under the visible peak that corresponds to it in the spectrum.

Finally, all broad peaks (i.e. para-crystalline and the inaccessible surface peak) were assigned based off literature chemical shifts. Full width half maxima were assigned based on literature but also accounted for remaining residual space in the spectrum, as is seen with large fluctuations in the accessible peak in the literature.<sup>26, 29, 43, 49</sup> The decrease in FWHM of the IAS peak is represented by the higher error for the IAS peak shown in Table 4.1. Upon removal of the two broader peaks (IAS and paracrystalline), the standard deviation (as generated by DMFit Fitting Software) doubles at minimum. Even when chemical shifts and FWHM are allowed to fluctuate to unreasonable positions or values that far outlie from literature reported ranges (i.e. Figure S4.2 B and D), the quality of fit is decreased significantly. The most crystalline GVL-treated spectra was fit first (DS140), since this spectra had the clearest crystalline peaks from which to start the assignment process, and then the set of chemical shifts and full width half maxima generated from the more crystalline samples was used to fit all other samples.



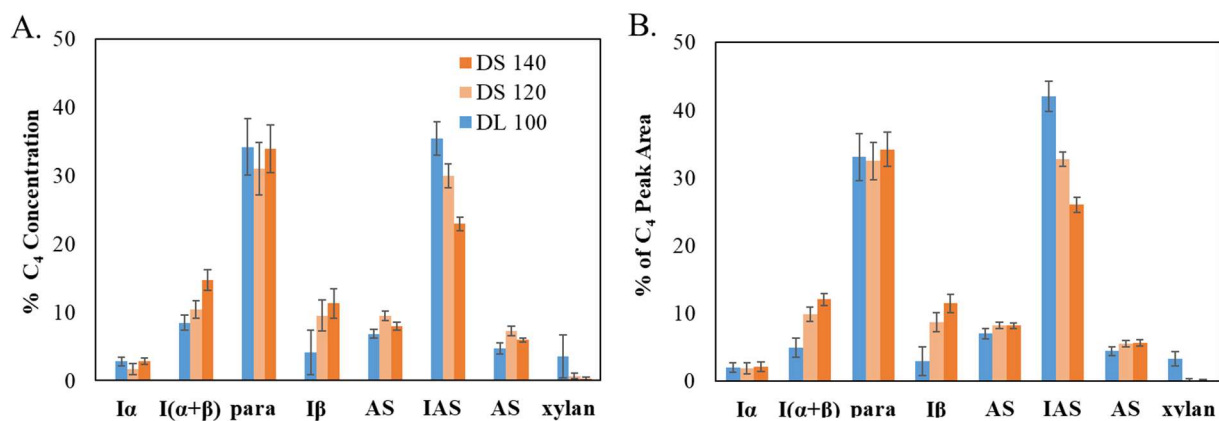
**Figure S4.1.** Comparisons of  $^{13}\text{C}$  CP/MAS NMR spectra of P39 Poplar treated until desaccharification at  $140^\circ\text{C}$  (DS 140) using different analysis methods. A. Overlay of the  $\text{C}_4$  resonances using the PSRE subtraction method outlined in previous work<sup>20</sup> and the 10,000  $\mu\text{s}$  spin lock delay spectrum which has not undergone a PSRE subtraction of the amorphous spectrum. B. Overlay of two spin lock delay times (6,000 and 10,000  $\mu\text{s}$ )



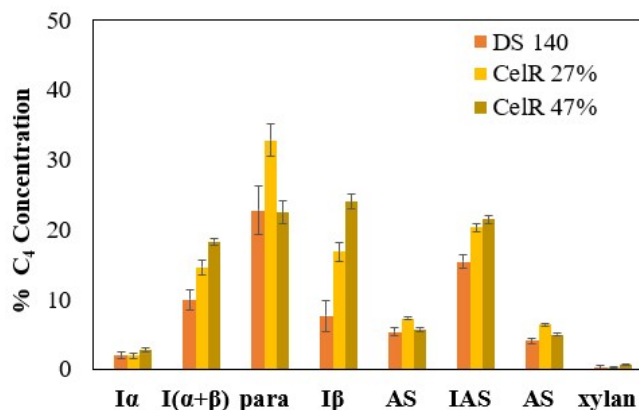
**Figure S4.2:** The effect of removing para-crystalline (para) and inaccessible surface area (IAS) peaks from the spectral fitting procedure on the P39 poplar DS 140 treated C<sub>4</sub> spectrum. A. Spectral fitting with all peaks described in the literature and outlined in Figure 4.3. B. Spectral fitting upon removal of both para and IAS peaks. No new restrictions were placed on the fitting parameters. As such, the left accessible surface peak width was allowed to vary, and the  $I_{\alpha+\beta}$  peak were allowed to vary along full width half maxima (FWHM), as was the case in A. Additionally, all chemical shifts were allowed to vary, as was also the case in A. C. Spectral fitting upon removal of both para and IAS peaks, with all peak chemical shifts locked. D. Spectral fitting upon removal of both para and IAS peaks, with all peak chemical shifts locked. E. Spectral fitting upon removal of both para and IAS peaks, with all peak chemical shifts and FWHM locked.

**Table S4.4:** Remaining weight percentage of P39 cellulose throughout GVL-assisted and enzymatic hydrolysis steps, determined by high pressure liquid chromatography and reducing sugar colorimetric assays, respectively.

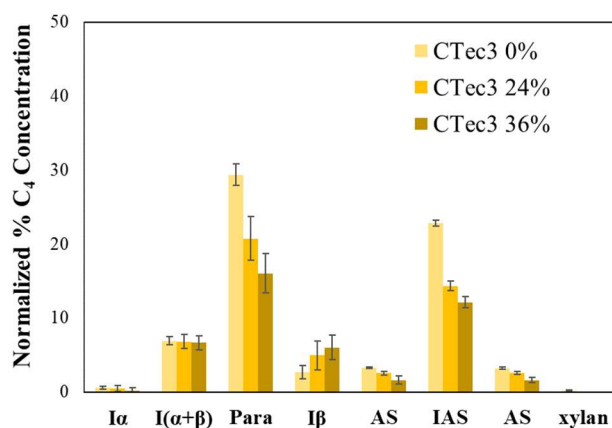
	Treatment Condition	Conversion of cellulose (%)	Cellulose remaining After DL 100 step (%)	Normalization factor applied to fitting
P39	Raw biomass	0	--	--
	DL 100	6.1	100	1.00
	DS 120	6.9	93	0.93
	DS 140	31	69	0.69
	CelR 27%	27	51	0.51
	CelR 47%	47	37	0.37
	CTec3 24%	24	52	0.52
	CTec3 36%	50	44	0.44
NM6	Raw biomass	0	--	--
	DL 100	6.8	100	1.00
	DS 120	9.5	91	0.91
	DS 140	32	68	0.68



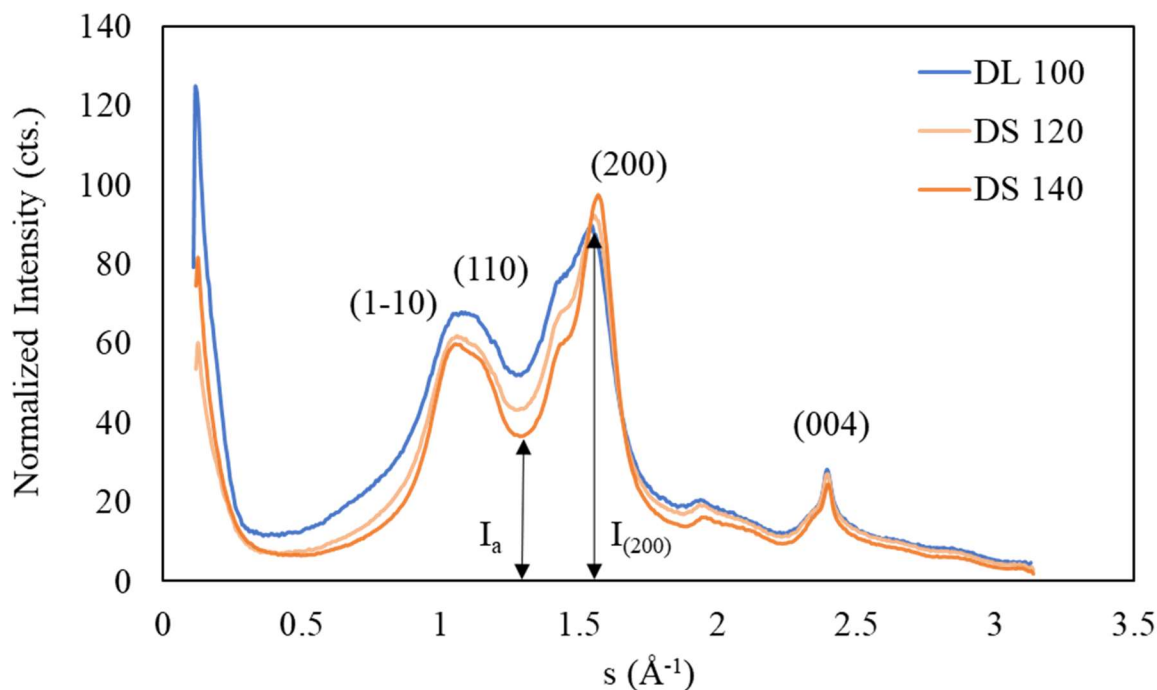
**Figure S4.3:** Changes of fitted cellulose regions in the C<sub>4</sub> peak resulting from GVL pretreatments. A, P39 and B, NM6 poplar. Concentrations of each domain are out of the total C<sub>4</sub> population (i.e. the total C<sub>4</sub> peak area) and sum to 100% for each condition. See Figure 4.1 in main text for GVL-assisted hydrolysis reaction conditions.



**Figure S4.4:** Changes of fitted cellulose regions in the C<sub>4</sub> peak resulting from enzymatic hydrolysis with CelR. Concentrations of each domain are out of the total C<sub>4</sub> population (i.e. the total C<sub>4</sub> peak area) and sum to 100% for each condition. Enzymatic hydrolysis conditions: 20 mg/mL GVL-treated biomass and 2 mg/mL CelR (10 wt.% CelR) in 0.1 M phosphate buffer, pH 6.0, 30 mL total volume, 24 h or 72 h reaction time at 70 °C with no shaking.



**Figure S4.5:** Changes of fitted cellulose regions in the C<sub>4</sub> peak resulting from enzymatic hydrolysis with CTec3 of the P39 poplar biomass treated with DL 100 and DS 140 conditions. The percent of C<sub>4</sub> concentration in each domain is out of the total carbon (area) present in the C<sub>4</sub> resonance. Concentrations of each domain are then normalized to the cellulose content remaining compared to the total cellulose content of the DL 100 sample (determined by the reducing sugar colorimetric assay, see Table S4.4 for scaling factors) and changes in peak area reflect the conversion of cellulose from hydrolysis reactions and/or structural rearrangement during the GVL treatment. Enzymatic hydrolysis conditions for CTec3 0%: 20 mg/mL GVL-treated biomass and 0.4 mg/mL CTec3 (10 wt.% CTec3) in 0.1 M phosphate buffer, pH 5.0, 30 mL total volume. 24 h or 72 h reaction time at 52.5 °C with no shaking. CTec3 0% sample underwent aforementioned reaction conditions, but with 0 mg/mL CTec3 for 24 h reaction time.



**Figure S4.6:** WAXS plots for DL 100, DS 120, and DS 140 GVL-treated P39 poplar wood.  $s$  represents the scattering vector and spectra are background-subtracted.

The scattering vector,  $s$ , along the horizontal axis is defined by

$$s = 4\pi \sin \theta / \lambda$$

where  $\theta$  is the scattering angle and  $\lambda$  is the X-ray wavelength. The lattice spacing,  $d$ , along the (200) plane was calculated by

$$d(200) = s(200)/2\pi$$

where  $s(200)$  is the scattering angle along the (200) plane.  $L(200)$  is defined by the Scherrer equation:

$$L(200) = K\lambda/\beta\cos\theta$$

where  $\beta$  is the full-width at half maximum of the (200) peak and  $K$  is a dimensionless shape factor taken to be 0.9<sup>50</sup>. The formula to determine the Segal Crystallinity Index<sup>51</sup> is as follows:

$$CrI = (I(200) - I_a) / I(200)$$

where  $I_a$  represents the amorphous content of the sample and  $I(200)$  represents the crystalline content.

## References:

1. Srivastava RK, Shetti NP, Reddy KR, Kwon EE, Nadagouda MN, Aminabhavi TM. Biomass utilization and production of biofuels from carbon neutral materials. *Environmental Pollution*. 2021;276:116731.
2. Yang B, Wyman CE. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining: Innovation for a sustainable economy*. 2008;2(1):26-40.
3. West RM, Liu ZY, Peter M, Dumesic JA. Liquid alkanes with targeted molecular weights from biomass-derived carbohydrates. *ChemSusChem: Chemistry & Sustainability Energy & Materials*. 2008;1(5):417-424.
4. Alonso DM, Wettstein SG, Mellmer MA, Gurbuz EI, Dumesic JA. Integrated conversion of hemicellulose and cellulose from lignocellulosic biomass. *Energy & Environmental Science*. 2013;6(1):76-80.
5. Román-Leshkov Y, Barrett CJ, Liu ZY, Dumesic JA. Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature*. 2007;447(7147):982.
6. Chang H, Motagamwala AH, Huber GW, Dumesic JA. Synthesis of biomass-derived feedstocks for the polymers and fuels industries from 5-(hydroxymethyl) furfural (HMF) and acetone. *Green Chemistry*. 2019;21(20):5532-5540.
7. Gilna P, Lynd LR, Mohnen D, Davis MF, Davison BH. Progress in understanding and overcoming biomass recalcitrance: a BioEnergy Science Center (BESC) perspective. *Biotechnology for Biofuels*. 2017;10(1):285.
8. Patil SK, Lund CR. Formation and growth of humins via aldol addition and condensation during acid-catalyzed conversion of 5-hydroxymethylfurfural. *Energy & Fuels*. 2011;25(10):4745-4755.
9. Lynd LR, Laser MS, Bransby D, et al. How biotech can transform biofuels. *Nature biotechnology*. 2008;26(2):169-172.
10. Beckham GT, Matthews JF, Peters B, Bomble YJ, Himmel ME, Crowley MF. Molecular-level origins of biomass recalcitrance: decrystallization free energies for four common cellulose polymorphs. *The Journal of Physical Chemistry B*. 2011;115(14):4118-4127.
11. Pu Y, Hallac B, Ragauskas AJ. Plant Biomass Characterization: Application of Solution- and Solid-State NMR Spectroscopy. In: *Aqueous Pretreatment of Plant Biomass for Biological and Chemical Conversion to Fuels and Chemicals*. 2013:369-390.
12. Sannigrahi P, Ragauskas AJ, Miller SJ. Effects of two-stage dilute acid pretreatment on the structure and composition of lignin and cellulose in loblolly pine. *BioEnergy Research*. 2008;1(3-4):205-214.
13. Rinaldi R, Jastrzebski R, Clough MT, et al. Paving the way for lignin valorisation: recent advances in bioengineering, biorefining and catalysis. *Angewandte Chemie International Edition*. 2016;55(29):8164-8215.



14. Woiciechowski AL, Neto CJD, de Souza Vandenberghe LP, et al. Lignocellulosic biomass: Acid and alkaline pretreatments and their effects on biomass recalcitrance—Conventional processing and recent advances. *Bioresource technology*. 2020;304:122848.
15. Sannigrahi P, Miller SJ, Ragauskas AJ. Effects of organosolv pretreatment and enzymatic hydrolysis on cellulose structure and crystallinity in Loblolly pine. *Carbohydrate Research*. 2010;345(7):965-970.
16. Zhao X, Cheng K, Liu D. Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. *Applied microbiology and biotechnology*. 2009;82(5):815-827.
17. Wettstein SG, Alonso DM, Chong Y, Dumesic JA. Production of levulinic acid and gamma-valerolactone (GVL) from cellulose using GVL as a solvent in biphasic systems. *Energy & Environmental Science*. 2012;5(8):8199-8203.
18. Luterbacher JS, Azarpira A, Motagamwala AH, Lu F, Ralph J, Dumesic JA. Lignin monomer production integrated into the  $\gamma$ -valerolactone sugar platform. *Energy & Environmental Science*. 2015;8(9):2657-2663.
19. Chheda JN, Huber GW, Dumesic JA. Liquid-phase catalytic processing of biomass-derived oxygenated hydrocarbons to fuels and chemicals. *Angewandte Chemie International Edition*. 2007;46(38):7164-7183.
20. Walker TW, Kuch N, Vander Meulen KA, et al. Solid-State NMR Studies of Solvent-Mediated, Acid-Catalyzed Woody Biomass Pretreatment for Enzymatic Conversion of Residual Cellulose. *ACS Sustainable Chemistry & Engineering*. 2020;8(16):6551-6563.
21. Klass DL. *Biomass for renewable energy, fuels, and chemicals*. Elsevier; 1998.
22. Pu Y, Ziemer C, Ragauskas AJ. CP/MAS <sup>13</sup>C NMR analysis of cellulase treated bleached softwood kraft pulp. *Carbohydrate Research*. 2006;341(5):591-597.
23. Klein-Marcuschamer D, Oleskiewicz-Popiel P, Simmons BA, Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and bioengineering*. 2012;109(4):1083-1087.
24. Kumar R, Singh S, Singh OV. Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. *Journal of industrial microbiology and biotechnology*. 2008;35(5):377-391.
25. Humbird D, Davis R, Tao L, et al. *Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover*. National Renewable Energy Lab.(NREL), Golden, CO (United States);2011.
26. Larsson PT, Wickholm K, Iversen T. A CP/MAS<sup>13</sup>C NMR investigation of molecular ordering in celluloses. *Carbohydrate Research*. 1997;302(1-2):19-25.
27. Wickholm K, Larsson PT, Iversen T. Assignment of non-crystalline forms in cellulose I by CP/MAS <sup>13</sup>C NMR spectroscopy. *Carbohydrate Research*. 1998;312(3):123-129.
28. Hult E-L, Larsson PT, Iversen T. A comparative CP/MAS <sup>13</sup>C-NMR study of cellulose structure in spruce wood and kraft pulp. *Cellulose*. 2000;7(1):35-55.
29. Larsson PT, Hult E-L, Wickholm K, Pettersson E, Iversen T. CP/MAS <sup>13</sup>C-NMR spectroscopy applied to structure and interaction studies on cellulose I. *Solid State Nuclear Magnetic Resonance*. 1999;15(1):31-40.
30. Borrega M, Larsson PT, Ahvenainen P, et al. Birch wood pre-hydrolysis vs pulp post-hydrolysis for the production of xylan-based compounds and cellulose for viscose application. *Carbohydrate Polymers*. 2018;190:212-221.

31. Foston M, Ragauskas AJ. Changes in lignocellulosic supramolecular and ultrastructure during dilute acid pretreatment of Populus and switchgrass. *Biomass and Bioenergy*. 2010;34(12):1885-1895.
32. Peciulyte A, Karlström K, Larsson PT, Olsson L. Impact of the supramolecular structure of cellulose on the efficiency of enzymatic hydrolysis. *Biotechnology for biofuels*. 2015;8(1):56.
33. Wyman CE. What is (and is not) vital to advancing cellulosic ethanol. *TRENDS in Biotechnology*. 2007;25(4):153-157.
34. Li Y, Shuai L, Kim H, et al. An “ideal lignin” facilitates full biomass utilization. *Science advances*. 2018;4(9):eaau2968.
35. Abu-Omar MM, Barta K, Beckham GT, et al. Guidelines for performing lignin-first biorefining. *Energy & Environmental Science*. 2021;14(1):262-292.
36. Zuckerstätter G, Terinte N, Sixta H, Schuster KC. Novel insight into cellulose supramolecular structure through <sup>13</sup>C CP-MAS NMR spectroscopy and paramagnetic relaxation enhancement. *Carbohydrate polymers*. 2013;93(1):122-128.
37. Hult E-L, Larsson PT, Iversen T. A CP/MAS <sup>13</sup>C-NMR study of supermolecular changes in the cellulose and hemicellulose structure during kraft pulping. *Nordic Pulp & Paper Research Journal*. 2001;16(1):33-39.
38. Simmons TJ, Mortimer JC, Bernardinelli OD, et al. Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nature Communications*. 2016;7:13902.
39. Zverlov VV, Schantz N, Schwarz WH. A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo- $\beta$ -1, 4-glucanase producing cellotetraose. *FEMS microbiology letters*. 2005;249(2):353-358.
40. Newman RH. Estimation of the lateral dimensions of cellulose crystallites using <sup>13</sup>C NMR signal strengths. *Solid State Nuclear Magnetic Resonance*. 1999;15(1):21-29.
41. Kubicki JD, Yang H, Sawada D, O'Neill H, Oehme D, Cosgrove D. The Shape of Native Plant Cellulose Microfibrils. *Scientific Reports*. 2018;8(1):13983.
42. Jarvis MC. Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2018;376(2112):20170045.
43. Wickholm K, Hult E-L, Larsson PT, Iversen T, Lennholm H. Quantification of cellulose forms in complex cellulose materials: a chemometric model. *Cellulose*. 2001;8(2):139-148.
44. Sluiter J, Sluiter A. Summative mass closure. *NREL, NREL/TP-510-48087*. 2010:1-10.
45. Fung B, Khitrin A, Ermolaev K. An improved broadband decoupling sequence for liquid crystals and solids. *Journal of magnetic resonance*. 2000;142(1):97-101.
46. Massiot D, Fayon F, Capron M, et al. Modelling one-and two-dimensional solid-state NMR spectra. *Magnetic resonance in chemistry*. 2002;40(1):70-76.
47. Ilavsky J. Nika: software for two-dimensional data reduction. *Journal of Applied Crystallography*. 2012;45(2):324-328.
48. Park S, Baker JO, Himmel ME, Parilla PA, Johnson DK. Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnology for biofuels*. 2010;3(1):1-10.
49. Larsson PT, Westlund P-O. Line shapes in CP/MAS <sup>13</sup>C NMR spectra of cellulose I. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 2005;62(1):539-546.
50. Murdock CC. The form of the X-ray diffraction bands for regular crystals of colloidal size. *Physical review*. 1930;35(1):8.

51. Segal L, Creely JJ, Martin Jr A, Conrad C. An empirical method for estimating the degree of crystallinity of native cellulose using the X-ray diffractometer. *Textile research journal*. 1959;29(10):786-794.

## Chapter 5:

### **Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy**

Evan Glasgow, Kirk Vander Meulen, Nate Kuch and Brian G Fox

This work was originally published as:

Glasgow, E., Vander Meulen, K., Kuch, N., Fox, B. G. 2021. Multifunctional cellulases are potent, versatile tools for a renewable bioeconomy. *Current opinion in biotechnology*, 67, 141–148. <https://doi.org/10.1016/j.copbio.2020.12.020>

Nathaniel Kuch made the following contributions: conceptualization of review content, investigation of literature sources, manuscript revision and editing

#### **Abstract:**

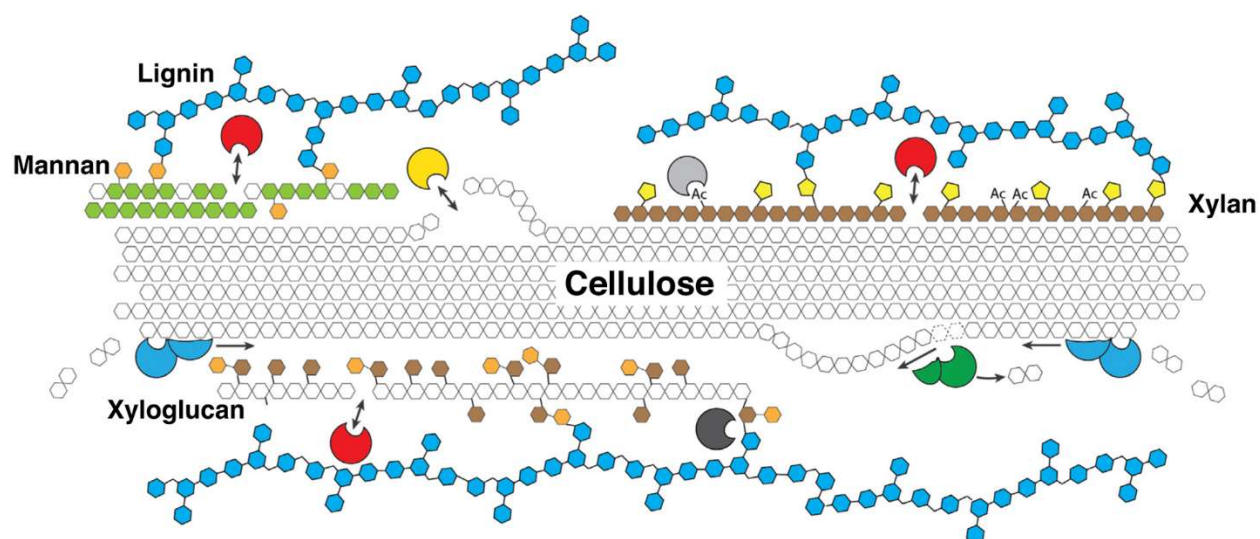
Enzyme performance is critical to the future bioeconomy based on renewable plant materials. Plant biomass can be efficiently hydrolyzed by multifunctional cellulases (MFCs) into sugars suitable for conversion into fuels and chemicals, and MFCs fall into three functional categories. Recent work revealed MFCs with broad substrate specificity, dual exo-activity/endo-activity on cellulose, and intramolecular synergy, among other novel characteristics. Binding modules and accessory catalytic domains amplify MFC and xylanase activity in a wide variety of ways, and processive endoglucanases achieve autosynergy on cellulose. Multidomain MFCs from *Caldicellulosiruptor* are heat-tolerant, adaptable to variable cellulose crystallinity, and may provide interchangeable scaffolds for recombinant design. Further studies of MFC properties and their reactivity with plant biomass are recommended for increasing biorefinery yields.

#### **Introduction:**

Global efforts to reduce carbon emissions are spurring technological innovations worldwide. As part of an integrated bioeconomy, the synthesis of fuels and bioproducts from

renewable plant biomass may supplant petroleum-derived products, reduce net carbon emissions and promote energy security. Inedible lignocellulosic biomass (LCB) contains vast reserves of convertible carbohydrates and aromatics for next-generation fuel and product synthesis but recovering these molecules from the polymeric matrix of a plant cell wall remains a significant challenge.

Physical and chemical pretreatments render LCB susceptible to cellulases, which hydrolyze the exposed polysaccharides into small, fermentable sugars. The complex polymeric composition of LCB (Figure 5.1) requires several distinct enzyme activities to liberate as much sugar as possible. While enzymatic deconstruction has several advantages over chemical hydrolysis, enzyme cocktails are a major (up to 30%) operational expense for biorefineries [1•]. Reducing the cost of LCB deconstruction is imperative for a viable bioeconomy to produce renewable fuels and bioproducts at competitive prices, and enzymes are a critical cost-reduction target.



**Figure 5.1:** Simplified structure of LCB reacting with different classes of enzymes. Processive exocellulases (blue) depolymerize cellulose from reducing or non-reducing ends. Endocellulases (yellow) create new chain ends in amorphous regions. Hemicellulases (red) hydrolyze non-cellulose

polysaccharides. Processive endocellulases (green) create new ends and continue to depolymerize. Carbohydrate-targeting and lignin-targeting esterases (light gray and dark gray, resp.) remove acetyl side chains and cleave linkages to lignin. Glucose (white hexagon), mannose (green hexagon), xylose (brown hexagon), galactose (orange hexagon), arabinose (yellow pentagon), lignin (blue hexagon).

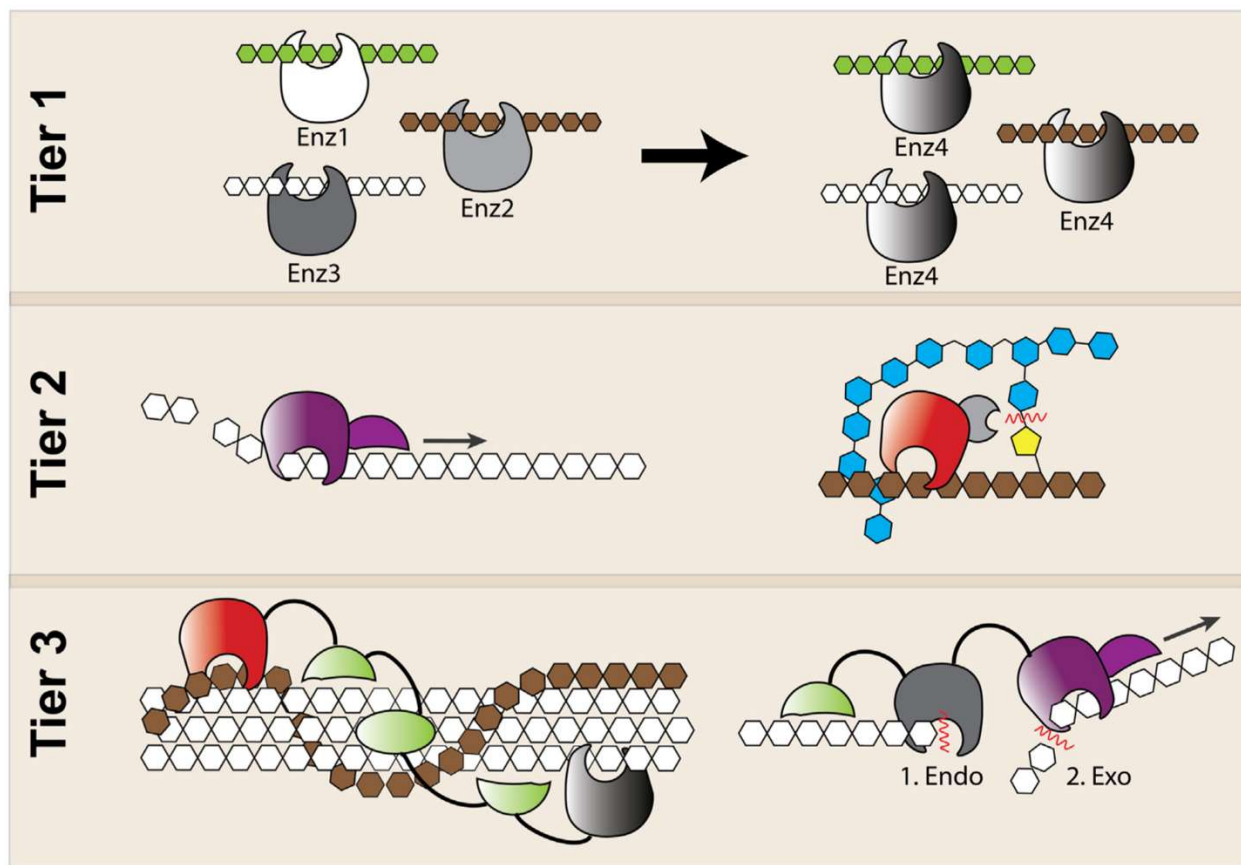
Recent investigations identified glycoside hydrolases (GH) with multiple catalytic functions. Here, we define multifunctional cellulases (MFCs) as a subset of GHs that possess, at minimum,  $\beta$ -(1,4)-endoglucanase or cellobiohydrolase activity with two or more polysaccharide types. Multifunctional cellulases (MFCs) may provide a route to lower enzyme costs by reducing the complexity of enzyme cocktails while increasing efficiency of enzyme hydrolysis. This review highlights recent discoveries, mechanistic investigations, and laboratory-scale applications of MFCs and recommends several considerations for future technology development. We also note that many trends in MFC research are occurring in parallel on xylanases, spurred by the abundance of xylan in LCB and the special utilities of pentose sugars in biosynthetic pathways [2].

Better understanding of MFCs may simplify the composition of enzyme cocktails, lower the amount of enzyme needed, and, along with organic co-solvents, may enable the as-yet unmet goal of biomass-agnostic deconstruction, in which a single chemoenzymatic process efficiently deconstructs a variety of LCB feedstocks. Catalytic versatility can thus enable a more robust bioeconomy operable across a variety of agricultural regions and climatic zones [3].

### **A multi-tiered view of MFCs:**

We propose a novel, three-tier functional classification system (Figure 5.2) to help group MFCs by complexity and to distinguish among the different factors contributing to multifunctionality. Tier 1 MFCs consolidate breadth of substrate specificities into a single GH active site, while Tier 2 MFCs incorporate additional non-GH domains or reaction mechanisms

that further amplify the native activity. A Tier 3 MFC brings two or more GH domains that catalyze closely associated steps in LCB deconstruction and may synergize in reaction on the same substrate. Tier 3 MFCs may be built from Tier 1 and 2 components, just as the GH domains in Tier 2 MFCs may have Tier 1 broad specificity.

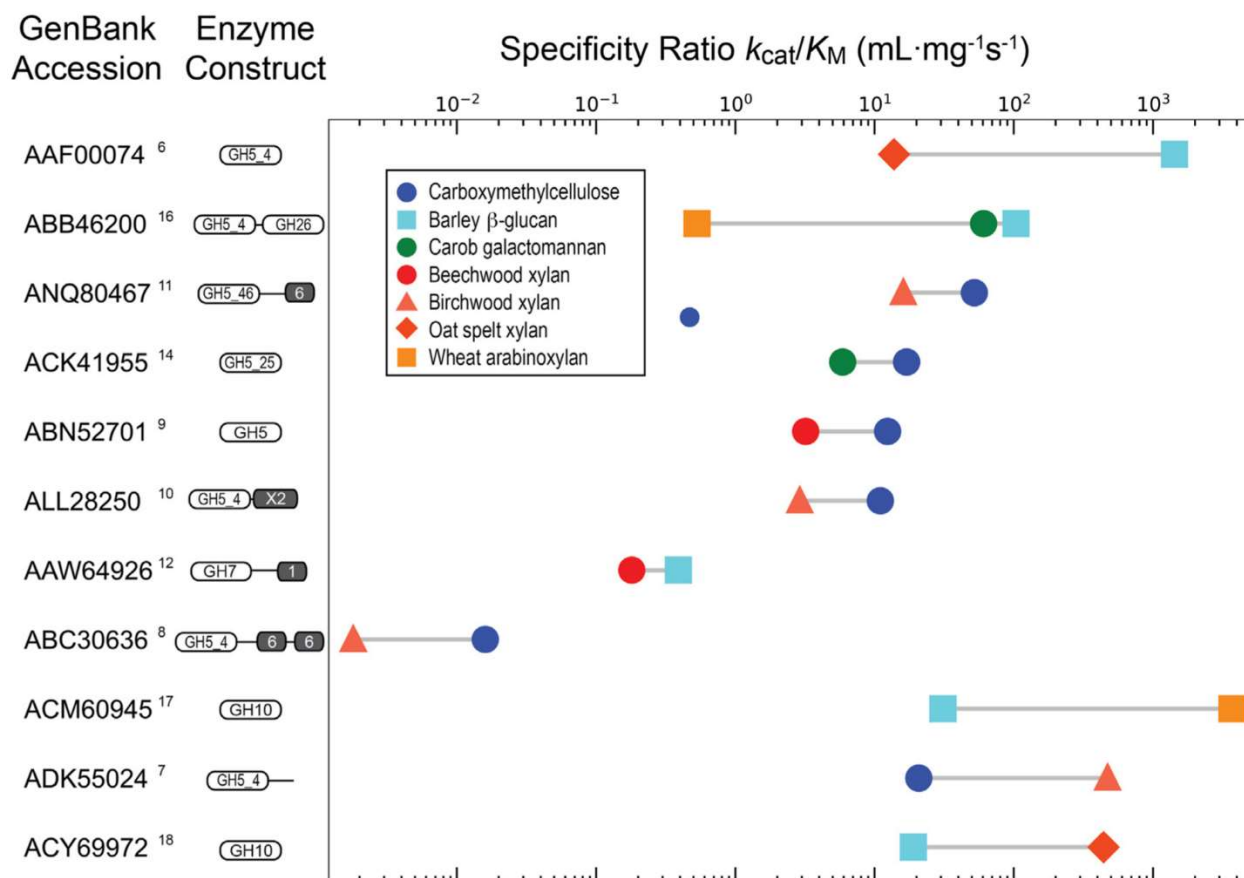


**Figure 5.2:** Tiers of function in MFCs. Several monospecific enzymes (left) can be replaced by a single Tier 1 MFC (right) possessing broad substrate specificity. Tier 2 MFCs possess additional domains that confer increased activity on native substrates by promoting processivity (left), cleaving crosslinks or branches (right), and so on. Tier 3 enzymes combine several GH and binding domains to achieve intramolecular synergy, for example, by xylanase-cellulase fusion (left) or endoglucanase-exoglucanase fusion (right). Polysaccharide colors and shapes are as in Figure 5.1.

**Tier 1 MFCs: broad substrate specificity in single domains:**

The simplest MFCs are single-domain GHs with broad substrate specificity. According to the Carbohydrate-Active Enzymes Database (CAZy, <http://www.cazy.org/>), families enriched with MFCs include GH5, 26, 44, 45, and 74 [4], with GH5 being most enriched in broad-specificity endoglucanases. GH5 is divided into subfamilies to help map its many specificities into sequence-similar groups [5]. GH5 endoglucanase activity often co-occurs with endoxylanase [6–12] or endomannanase [13••,14•] activity, in some cases with both [15••,16], as in subfamily 4 (GH5\_4). The glucanase efficiency is typically highest (Figure 5.3), although xylanase activity sometimes exceeds it [7], mimicking the specificity of GH10 xylanases (see ACM60945 [17] and ACY69972 [18] in Figure 5.3). The identification of gluco-mannanase [19],  $\beta$ -(1,3)(1,4)-mixed-linkage glucanase [20,21] and xyloglucanase [22] activities demonstrates that MFCs accept variations in sugar monomers and linkages, and also tolerate branching. Interestingly, Tier 1 MFCs from GH9 have been found in herbivorous insect genomes, with high expression in the digestive tract and cross-reactivity on cellulose and either xylan or xyloglucan [23,24]. Critically, some Tier 1 MFCs are as efficient as ‘specialized’ monospecific GHs on each of their multiple substrates [25], illustrating that broad specificity and high reactivity are not mutually exclusive. The comparable reactivity of MFCs and monospecific GHs supports the promise of MFCs in designed enzyme cocktails.





**Figure 5.3:** Specificity ratios ( $k_{cat}/K_M$ ) for enzymes assayed on substrates with distinct backbone compositions.  $\beta$ -1,4-glucan substrates (carboxymethylcellulose and barley  $\beta$ -1,3-glucan, blue shades);  $\beta$ -1,4 mannose backbone (carob galactomannan, green);  $\beta$ -1,4 xylose backbone substrates (xylans from beechwood, birchwood and oat spelt, red shades); arabinoxylan (orange). GenBank accession code (with reference as superscript) and enzyme domain structure are shown on the left. Enzyme and CBM domains are denoted as ovals labeled with black and white text, respectively, and linker regions are also displayed to approximate proportional scale.

Tier 1 MFCs with more open binding clefts can tolerate a diversity of polysaccharides, including some with non- $\beta$ -(1,4) linkages or charged sugar residues [18]. For example, a dual cellulase/xylanase from GH26 [with an ( $\alpha\beta$ )8-barrel fold like that observed in GH5] [26] also hydrolyzed anionic  $\beta$ -(1,4)-polyglucuronic acid in a Ca<sup>2+</sup>-independent fashion distinct from pectate lyases. In another example, GH131 MFCs (with a  $\beta$ -jelly roll fold) have a widely accessible cleft that can hydrolyze  $\beta$ -(1,4) and  $\beta$ -(1,3) bonds equally well [27]. Activity on chitosan (the cationic, deacetylated form of chitin) is also observed in cellulases encompassing

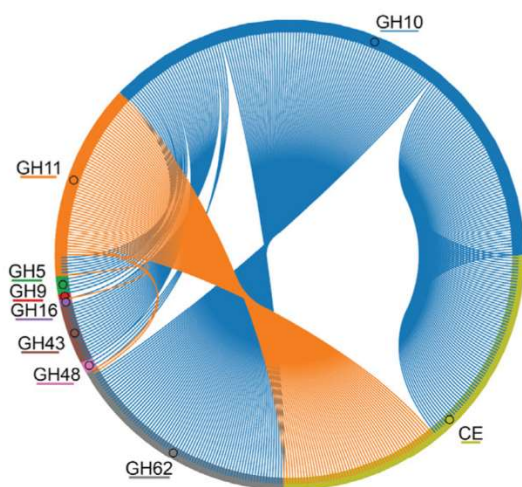
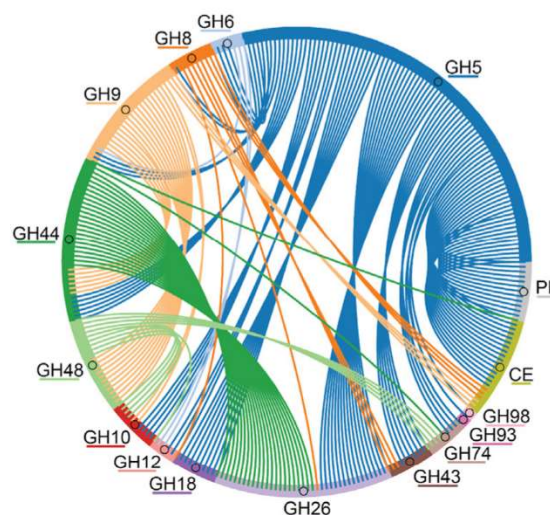
several distinct folds and mechanisms [28]. Thus, MFCs are widely distributed among GH families, offering many options for design of customized enzyme cocktails.

### **Tier 2 MFCs: multi-domain amplification of reactivity:**

The combination of a GH and a carbohydrate-binding module (CBM) is the most abundant example of a Tier 2 MFC and provides a mechanism to direct the reactivity of an MFC onto different polysaccharides [29]. Various combinations of MFC and CBM can give  $\sim 10^5$  distribution of the specificity ratio (kcat/KM) across a breadth of hexose and pentose polysaccharides (Figure 5.3).

CelE, a cellulosome-associated MFC from *Hungateiclostridium thermocellum*, is a multidomain enzyme consisting of a GH5 endoglucanase domain, a dockerin domain, and a CE2 domain. The GH5 domain is a Tier 1 MFC that degrades cellulose, mannan, xylan, and several heteropolysaccharides [15•,29,30], while the CE domain deacetylates xylan [31] and further exposes the hemicellulose to the GH5 active site, demonstrating a Tier 2 functionality.

GH amplification can also be achieved by fusion of carbohydrate esterase (CE) domains to xylanases from GH10 and GH11 (Figure 5.4a). When added in excess as separate proteins, CEs can potentiate xylanases up to 20-fold on some plant biomass [32•]. Non-covalent ‘xylanosome’ complexes can confer over threefold synergy by keeping xylanase and acetylxylan esterase in close proximity [33]. Although some of these fusions bind to amorphous cellulose [34], their cellulase activities have only been sparsely explored.

**(a) Xylanase Families****(b) Cellulase Families**

**Figure 5.4:** Chord plot illustration of the connections between GH and other domains in Tier 2 and 3 enzymes. Arcs connect modules present in a single gene according to separate listings in the CAZy database. (a) Tier 2 and 3 endoxylanase (GH10, GH11) connections to CE and other GH domains. (b) Tier 2 and 3 endoglucanase (GH5, GH6, GH8, GH9, GH44, or GH48) connections to CE, PL, other GH domains. For simplicity, the multiple carbohydrate esterase (CE) and polysaccharide lyase (PL) families have been merged into single composite entries.

Some xylanases, such as XynY and XynZ from *H. thermocellum* and Xyn10A/Fae1A from a soil consortium [35], harbor a p-feruloyl esterase domain (CE1), which cleaves linkages to lignin and releases ferulic acid. To our knowledge, this lignin-targeting module has not been found in cellulases, suggesting selective evolution of the most appropriate pairs of catalytic domains in Tier 2 MFCs.

Disruption of substrate superstructure is another Tier 2 mechanism, exemplified by the action of swollenins. Swollenins can increase the activity of both xylanases and cellulases [36,37] up to ~3-fold by weakening the non-covalent interactions between polysaccharide chains. Swollenin fusions to CEs [38] and xylanases [39] have shown promise, but fusions to cellulases have been limited to expansins, the bacterial counterpart [40], and show less impact on activity than swollenins.

Processivity is yet another Tier 2 function that amplifies the native ability of exocellulases to hydrolyze  $\beta$ -(1,4)-glucosidic linkages by threading the substrate polymer into the active site without dissociation [41•]. The effect is limited to crystalline cellulose and impacts  $k_{cat}$  rather than  $K_M$  [42]. Processivity is often assisted by a tunnel-shaped active site and an intimately associated CBM. Different exocellulase families exhibit strict specificity for reducing (GH7 and 48) or non-reducing (GH6) ends.

Processivity is not limited to exocellulases. Processive endoglucanases (PEGs) are less understood, but they have been found in microbes lacking a canonical exocellulase [43••]. PEGs are fascinating because they first create new chain ends by endocellulase action, and then processively depolymerize the polysaccharide by exocellulase action (Figure 5.1, green). PEGs found in GH5 may possess CBMs from families 1, 3, or 6, although their precise function in processivity is less clear and may even be inhibitory [44]. PEGs found in GH9 show close association between the GH domain and a CBM3. One case was found in GH48, demonstrating PEGs may evolve in both endoglucanase and exoglucanase families.

### **Tier 3 MFCs: multi-domain intramolecular synergy:**

Two or more GH domains of differing (or similar) reactivity can be encoded by a single gene (Figure 5.4b). GH5, GH26, and GH44 are three families often found fused to additional GH domains. Fusion of endocellulase and exocellulase (or  $\beta$ -glucosidase) motifs may create intramolecular synergy [45•,46], but this is not a general rule: additional CBMs are often required [47]. Synergy in these multidomain enzymes is a complex property, and the position of CBM relative to GH is critical [48].

Bacteria in the genus *Caldicellulosiruptor* are the top sources of Tier 3 MFCs [49•]. Their MFCs are composed typically of two GH domains, interspersed with one to four CBMs, mostly from family 3 [49•,50•]. While they do not assemble into cellulosomes (which are complex assemblies of enzymes noncovalently bound to scaffoldin proteins by dockerin/cohesion pairs), the arrangements of GHs and CBMs in *Caldicellulosiruptor* MFCs recapitulate the function of cellulosomes with superior stability and specific activity without the noncovalent assembly domains. These MFCs also deconstruct LCB more effectively than mixtures of free fungal enzymes, and some, like CelA from *Caldicellulosiruptor bescii* [51••], are agnostic to the crystallinity index of cellulose [52]. CelA connects a GH9 PEG to a GH48 exocellulase, and the enzyme is optimally active at 75°C. The modular assembly, thermostability, and advances in native host expression [53] suggest considerable technological promise for *Caldicellulosiruptor* MFCs.

MFC reactivity is not confined to LCB polysaccharides. Recently, a cellulase was discovered in hot spring bacteria [54] possessing activity on agar and carrageenan (found in red algae) and amylose (from starch). The enzyme is composed of three GH70 domains, and its ability to degrade polycyclic, heterogeneous, and sulfated polysaccharides (in addition to both  $\alpha$ -glycosidic (1,4)-glycosidic bonds and  $\beta$ -(1,4)-glycosidic bonds) offers promise for processing marine biomass and enzymatic synthesis of bioactive compounds [55].

### **Conclusions:**

Tier 1 broad specificity enzymes add to a growing parts list for cellulase design, yet their ability to degrade LCB is typically limited, despite their broad reactivity. Although improvements in stability are common [56], it has been difficult to surpass modest improvements

in catalytic efficiency or achieve broad specificity by sequence optimization of single domains [57], highlighting the importance of bioprospecting for novel Tier 1 MFCs [58,59] and the need for additional screening methods specifically for endo-acting MFCs.

Engineered modularity offers a different approach and greater promise for generating new MFCs by taking advantage of Tier 2 and Tier 3 functionality. Tier 2 phenomena require deeper mechanistic insights, including structural dynamics and other constraints on the interacting domains. In Tier 2, fusion to CBMs and other non-GH domains can expand potential by adding new substrate binding and kinetic mechanisms.

Tier 1 and Tier 2 MFCs are building blocks for multidomain, potentially synergistic Tier 3 enzymes. Combinatorial assembly of multi-domain Tier 3 MFCs can leverage Golden Gate assembly [60] and other rapid domain-swapping techniques [61]. The most compatible domains may come from families with high natural frequency of multidomain fusions, such as GH5, 9, 26, 44, and 48 (Figure 5.4b). Substituting these domains into the scaffolds of existing Tier 3 cellulases may be a promising strategy, since inter-domain interactions are sensitive to linker length [62], and linker glycosylation can be important [63•]. Focusing on MFCs with thermal stability and salt tolerance will facilitate future integration with thermochemical biomass pretreatments, which are critical to obtaining maximum sugar yields in the biorefinery.

Finally, LCB structure itself deserves equal attention if yields of monomer sugars are to be maximized, since LCB composition varies greatly between plant species, growth, and harvest conditions. Enzymatic deconstruction changes the properties of LCB, as do the myriad physical and chemical treatments preceding enzyme addition. These changes impact the amount and quality of lignin residue, the valorization of which is another key component of a sustainable bioeconomy [64•]. Experiments in LCB imaging, simulation, and engineering can also reveal

bottlenecks in enzymatic hydrolysis and inform the best use of MFCs in an innovative deconstruction pipeline.

### **Acknowledgements:**

This work was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494. EG and NK were supported by the UW-Madison Biotechnology Training Program (NIH 5 T32 GM008349).

### **Footnotes:**

Conflict of interest statement  
Nothing declared.

### **References and recommended reading:**

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. • Chandel AK, Garlapati VK, Singh AK, Antunes FAF, da Silva SS: The path forward for lignocellulose biorefineries: bottlenecks, solutions, and perspective on commercialization. *Bioresour Technol* 2018, 264:370–381
2. Naidu DS, Hlangothi SP, John MJ: Bio-based products from xylan: a review. *Carbohydr Polym* 2018, 179:28–41.
3. Oke MA, Annuar MSM, Simarani K: Mixed feedstock approach to lignocellulosic ethanol production-prospects and limitations. *BioEnergy Res* 2016, 9:1189–1203.
4. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014, 42:D490–D495.
5. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B: Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Bio* 2012, 12.
6. Liu J, Tsai C, Liu J, Cheng K, Cheng C: The catalytic domain of a *Piromyces rhizinflata* cellulase expressed in *Escherichia coli* was stabilized by the linker peptide of the enzyme. *Enzyme Microb Technol* 2001, 28:582–589.
7. Chang L, Ding M, Bao L, Chen Y, Zhou J, Lu H: Characterization of a bifunctional xylanase/endoglucanase from yak rumen microorganisms. *Appl Microbiol Biotechnol* 2011, 90:1933–1942.
8. Ghatge SS, Telke AA, Kang SH, Arulalapperumal V, Lee KW, Govindwar SP, Um Y, Oh DB, Shin HD, Kim SW: Characterization of modular bifunctional processive endoglucanase Cel5 from *Hahella chejuensis* KCTC 2396. *Appl Microbiol Biotechnol* 2014, 98:4421–4435.

9. Yuan SF, Wu TH, Lee HL, Hsieh HY, Lin WL, Yang B, Chang CK, Li Q, Gao J, Huang CH et al.: Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *J Biol Chem* 2015, 290:5739–5748. [PMC free article]
10. Rattu G, Joshi S, Satyanarayana T: Bifunctional recombinant cellulase-xylanase (rBhcell-xyl) from the polyextremophilic bacterium *Bacillus halodurans* TSLV1 and its utility in valorization of renewable agro-residues. *Extremophiles* 2016, 20:831–842.
11. Tan H, Miao R, Liu T, Yang L, Yang Y, Chen C, Lei J, Li Y, He J, Sun Q et al.: A bifunctional cellulase-xylanase of a new *Chryseobacterium* strain isolated from the dung of a straw-fed cattle. *Microb Biotechnol* 2018, 11:381–398.
12. Han C, Yang R, Sun Y, Liu M, Zhou L, Li D: Identification and characterization of a novel hyperthermostable bifunctional cellobiohydrolase-xylanase enzyme for synergistic effect with commercial cellulase on pretreated wheat straw degradation. *Front Bioeng Biotechnol* 2020, 8:296.
13. ••. Liang PH, Lin WL, Hsieh HY, Lin TY, Chen CH, Tewary SK, Lee HL, Yuan SF, Yang B, Yao JY et al.: A flexible loop for mannan recognition and activity enhancement in a bifunctional glycoside hydrolase family 5. *Biochim Biophys Acta Gen Subj* 2018, 1862:513–521
14. •. Chen ZW, Friedland GD, Pereira JH, Reveco SA, Chan R, Park JI, Thelen MP, Adams PD, Arkin AP, Keasling JD et al.: Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J Biol Chem* 2012, 287:25335–
15. ••. Glasgow EM, Vander Meulen KA, Takasuka TE, Bianchetti CM, Bergeman LF, Deutsch S, Fox BG: Extent and origins of functional diversity in a subfamily of glycoside hydrolases. *J Mol Biol* 2019, 431:1217–1233
16. Palackal N, Lyon CS, Zaidi S, Luginbuhl P, Dupree P, Goubet F, Macomber JL, Short JM, Hazlewood GP, Robertson DE et al.: A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. *Appl Microbiol Biotechnol* 2007, 74:113–124.
17. Chu YD, Tu T, Penttinen L, Xue XL, Wang XY, Yi ZL, Gong L, Rouvinen J, Luo HY, Hakulinen N et al.: Insights into the roles of non-catalytic residues in the active site of a GH10 xylanase with activity on cellulose. *J Biol Chem* 2017, 292:19315–19327.
18. Shi P, Tian J, Yuan T, Liu X, Huang H, Bai Y, Yang P, Chen X, Wu N, Yao B: *Paenibacillus* sp. strain E18 bifunctional xylanase-glucanase with a single catalytic domain. *Appl Environ Microbiol* 2010, 76:3620–3624.
19. Wang YW, Bai Y, Shu T, Fan P, Zhang HS, Turunen O, Xiong HR, Yu LJ: Characterization of a versatile glycoside hydrolase Cel5M from *Pectobacterium carotovorum* HG-49 for ramie degumming. *Text Res J* 2020, 90:1602–1615.
20. Dorival J, Ruppert S, Gunnoo M, Orłowski A, Chapelais-Baron M, Dabin J, Labourel A, Thompson D, Michel G, Czjzek M et al.: The laterally acquired GH5 ZgEngAGH5\_4 from the marine bacterium *Zobellia galactanivorans* is dedicated to hemicellulose hydrolysis. *Biochem J* 2018, 475:3609–3628.
21. Meng DD, Liu X, Dong S, Wang YF, Ma XQ, Zhou HX, Wang XQ, Yao LS, Feng YG, Li FL: Structural insights into the substrate specificity of a glycoside hydrolase family 5 lichenase from *Caldicellulosiruptor* sp. F32. *Biochem J* 2017, 474:3373–3389.
22. Attia MA, Nelson CE, Offen WA, Jain N, Davies GJ, Gardner JG, Brumer H: In vitro and in vivo characterization of three *Cellvibrio japonicus* glycoside hydrolase family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnol Biofuels* 2018, 11.



23. Shelomi M, Heckel DG, Pauchet Y: Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea). *Insect Biochem Mol Biol* 2016, 71:1–11.
24. Shelomi M, Wipfler B, Zhou X, Pauchet Y: Multifunctional cellulase enzymes are ancestral in Polyneoptera. *Insect Mol Biol* 2020, 29:124–135.
25. Glasgow EM, Kemna EI, Bingman CA, Ing NL, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG: A structural and kinetic survey of GH5\_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *J Biol Chem* 202010.1074/jbc.RA120.015328. Online ahead of print.
26. Patel AB, Patel AK, Shah MP, Parikh IK, Joshi CG: Isolation and characterization of novel multifunctional recombinant family 26 glycoside hydrolase from Mehsani buffalo rumen metagenome. *Biotechnol Appl Biochem* 2016, 63:257–265.
27. Anasontzis GE, Lebrun MH, Haon M, Champion C, Kohler A, Lenfant N, Martin F, O’Connell RJ, Riley R, Grigoriev IV et al.: Broad-specificity GH131 beta-glucanases are a hallmark of fungi and oomycetes that colonize plants. *Environ Microbiol* 2019, 21:2724–2739.
28. Xia W, Liu P, Liu J: Advance in chitosan hydrolysis by non-specific cellulases. *Bioresour Technol* 2008, 99:6751–6762.
29. Walker JA, Takasuka TE, Deng K, Bianchetti CM, Udell HS, Prom BM, Kim H, Adams PD, Northen TR, Fox BG: Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol Biofuels* 2015, 8:220.
30. Walker JA, Pattathil S, Bergeman LF, Beebe ET, Deng K, Mirzai M, Northen TR, Hahn MG, Fox BG: Determination of glycoside hydrolase specificities during hydrolysis of plant cell walls using glycome profiling. *Biotechnol Biofuels* 2017, 10:31.
31. Montanier C, Money VA, Pires VM, Flint JE, Pinheiro BA, Goyal A, Prates JA, Izumi A, Stalbrand H, Morland C et al.: The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol* 2009, 7:e71.
32. • Kmezik C, Bonzom C, Olsson L, Mazurkewich S, Larsbrink J: Multimodular fused acetylferuloyl esterases from soil and gut Bacteroidetes improve xylanase depolymerization of recalcitrant biomass. *Biotechnol Biofuels* 2020, 13
33. Srikrishnan S, Chen W, Da Silva NA: Functional assembly and characterization of a modular xylanosome for hemicellulose hydrolysis in yeast. *Biotechnol Bioeng* 2013, 110:275–285.
34. Bolam DN, Xie HF, White P, Simpson PJ, Hancock SM, Williamson MP, Gilbert HJ: Evidence for synergy between family 2b carbohydrate binding modules in *Cellulomonas fimi* xylanase 11A. *Biochemistry* 2001, 40:2468–2477.
35. Wang RN, Yang JS, Jang JM, Liu JW, Zhang Y, Liu L, Yuan HL: Efficient ferulic acid and xylo-oligosaccharides production by a novel multi-modular bifunctional xylanase/feruloyl esterase using agricultural residues as substrates. *Bioresour Technol* 2020, 297.
36. Santos CA, Ferreira-Filho JA, O’Donovan A, Gupta VK, Tuohy MG, Souza AP: Production of a recombinant swollenin from *Trichoderma harzianum* in *Escherichia coli* and its potential synergistic role in biomass degradation. *Microb Cell Fact* 2017, 16:83.
37. Nakatani Y, Yamada R, Ogino C, Kondo A: Synergetic effect of yeast cell-surface expression of cellulase and expansin-like protein on direct ethanol production from cellulose. *Microb Cell Fact* 2013, 12:66.
38. Levasseur A, Saloheimo M, Navarro D, Andberg M, Monot F, Nakari-Setälä T, Asther M, Record E: Production of a chimeric enzyme tool associating the *Trichoderma reesei* swollenin

- with the *Aspergillus niger* feruloyl esterase A for release of ferulic acid. *Appl Microbiol Biotechnol* 2006, 73:872–880.
39. Zhang W, Liu C, Qu M, Pan K, OuYang K, Song X, Zhao X: Construction and characterization of a chimeric enzyme of swollenin and xylanase to improve soybean straw hydrolysis. *Int J Biol Macromol* 2020, 156:558–564.
40. Nakashima K, Endo K, Shibasaki-Kitakawa N, Yonemoto T: A fusion enzyme consisting of bacterial expansin and endoglucanase for the degradation of highly crystalline cellulose. *RSC Adv* 2014, 4:43815–43820.
41. • Vermaas JV, Kont R, Beckham GT, Crowley MF, Gudmundsson M, Sandgren M, Stahlberg J, Valjamae P, Knott BC: The dissociation mechanism of the processive cellulase TrCel7A. *Biophys J* 2020, 118:531a–532a
42. Cruys-Bagger N, Tatsumi H, Ren GR, Borch K, Westh P: Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: effects of substrate structure and carbohydrate binding domain. *Biochemistry* 2013, 52:8938–8948.
43. •• Wu S, Wu S: Processivity and the mechanisms of processive endoglucanases. *Appl Biochem Biotechnol* 2020, 190:448–463
44. Wu B, Zheng S, Pedroso MM, Guddat LW, Chang S, He B, Schenk G: Processivity and enzymatic mechanism of a multifunctional family 5 endoglucanase from *Bacillus subtilis* BS-5 with potential applications in the saccharification of cellulosic substrates. *Biotechnol Biofuels* 2018, 11:20.
45. • Li XQ, Xia JL, Zhu XY, Bilal M, Tan ZB, Shi H: Construction and characterization of bifunctional cellulases: Caldicellulosiruptor-sourced endoglucanase, CBM, and exoglucanase for efficient degradation of lignocellulose. *Biochem Eng J* 2019, 151
46. Nath P, Dhillon A, Kumar K, Sharma K, Jamaldeen SB, Moholkar VS, Goyal A: Development of bi-functional chimeric enzyme (CtGH1-L1-CtGH5-F194A) from endoglucanase (CtGH5) mutant F194A and beta-1,4-glucosidase (CtGH1) from *Clostridium thermocellum* with enhanced activity and structural integrity. *Bioresour Technol* 2019, 282:494–501.
47. Brunecky R, Subramanian V, Yarbrough JM, Donohoe BS, Vinzant TB, Vanderwall TA, Knott BC, Chaudhari YB, Bomble YJ, Himmel ME et al.: Synthetic fungal multifunctional cellulases for enhanced biomass conversion. *Green Chem* 2020, 22:478–489.
48. Sajjad M, Khan IM, Zafar R, Ahmad S, Niazi UHK, Akhtar MW: Influence of positioning of carbohydrate binding module on the activity of endoglucanase CelA of *Clostridium thermocellum*. *J Biotechnol* 2012, 161:206–212.
49. • Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM: The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'. *Extremophiles* 2020, 24:1–15
50. • Conway JM, Crosby JR, Hren AP, Southerland RT, Lee LL, Lunin VV, Alahuhta P, Himmel ME, Bomble YJ, Adams MWW et al.: Novel multidomain, multifunctional glycoside hydrolases from highly lignocellulolytic *Caldicellulosiruptor* species. *AIChE J* 2018, 64:4218–4228
51. •• Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MWW, Lunin VV et al.: Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 2013, 342:1513–1516
52. Brunecky R, Donohoe BS, Yarbrough JM, Mittal A, Scott BR, Ding HS, Taylor LE, Russell JF, Chung DW, Westpheling J et al.: The multi domain *Caldicellulosiruptor bescii* CelA cellulase excels at the hydrolysis of crystalline cellulose. *Sci Rep* 2017, 7.

53. Chung D, Young J, Bomble YJ, Vander Wall TA, Groom J, Himmel ME, Westpheling J: Homologous expression of the *Caldicellulosiruptor bescii* CelA reveals that the extracellular protein is glycosylated. *PLoS One* 2015, 10.
54. Li J, Gu X, Pan A: Multifunctional alpha-amylase Amy19 possesses agarase, carrageenase, and cellulase activities. *Int J Biol Macromol* 2019, 126:585–594.
55. Park SH, Lee CR, Hong SK: Implications of agar and agarase in industrial applications of sustainable marine biomass. *Appl Microbiol Biotechnol* 2020.
56. Komor RS, Romero PA, Xie CB, Arnold FH: Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng Des Sel* 2012, 25:827–833.
57. Zheng F, Tu T, Wang XY, Wang Y, Ma R, Su XY, Xie XM, Yao B, Luo HY: Enhancing the catalytic activity of a novel GH5 cellulase GtCel5 from *Gloeophyllum trabeum* CBS 900.73 by site-directed mutagenesis on loop 6. *Biotechnol Biofuels* 2018, 11.
58. Ferrer M, Ghazi A, Belouqui A, Vieites JM, Lopez-Cortes N, Marin-Navarro J, Nechitaylo TY, Guazzaroni ME, Polaina J, Waliczek A et al.: Functional metagenomics unveils a multifunctional glycosyl hydrolase from the family 43 catalysing the breakdown of plant polymers in the calf rumen. *PLoS One* 2012, 7:e38134.
59. Maruthamuthu M, Jimenez DJ, Stevens P, van Elsas JD: A multi-substrate approach for functional metagenomics-based screening for (hemi)cellulases in two wheat straw-degrading microbial consortia unveils novel thermoalkaliphilic enzymes. *BMC Genomics* 2016, 17.
60. Geddes BA, Mendoza-Suarez MA, Poole PS: A Bacterial Expression Vector Archive (BEVA) for flexible modular assembly of golden gate-compatible vectors. *Front Microbiol* 2018, 9:3345.
61. Maervoet VET, Briers Y: Synthetic biology of modular proteins. *Bioengineered* 2017, 8:196–202.
62. Rizk M, Antranikian G, Elleuche S: Influence of linker length variations on the biomass-degrading performance of heat-active enzyme chimeras. *Mol Biotechnol* 2016, 58:268–279.
63. • Chung D, Sarai NS, Knott BC, Hengge N, Russell JF, Yarbrough JM, Brunecky R, Young J, Supekar N, Wall TV et al.: Glycosylation is vital for industrial performance of hyperactive cellulases. *ACS Sustain Chem Eng* 2019, 7:4792–4800
64. • Liu ZH, Le RK, Kosa M, Yang B, Yuan JS, Ragauskas AJ: Identifying and creating pathways to improve biological lignin valorization. *Renew Sustain Energy Rev* 2019, 105:349–362

## Chapter 6:

### Summary for the general public

Fossil fuel emissions are the leading cause of climate change. To combat this issue, a large amount of research is being done to find alternatives to fuels and products that are made with petroleum and oil. One of these alternatives is to make biofuels from plants, which is what my work has focused on.

Turning plants into biofuels reduces carbon emissions because the carbon dioxide made when using the fuel is the same amount of carbon dioxide that the plant consumes as it grows. This balance of carbon dioxide used by the plants and produced by the fuels is an example of a carbon neutral process because the total carbon dioxide in the atmosphere has not changed. Biofuels are also renewable because the material needed to make them can be regrown. Fossil fuels, on the other hand, are non-renewable since they cannot be replenished after using them. The renewability of biofuels makes them a more reliable energy source for posterity.

To make biofuels from plants, you need to release sugars in the plant, like the starch in corn kernels, into an edible form for a microbe like yeast that can ferment the edible sugar into a biofuel like ethanol. Because corn is an important food source for both people and animals, using corn for biofuels is not economically or environmentally viable. Therefore, my research has focused on the sugars found in non-food plants, such as poplar trees or switchgrass. However, the sugars in non-food plants are much more difficult to process into an edible form for microbes than sugars like starch.

To achieve this difficult task, my work given in Chapter 2 first looks at an enzyme, a small cellular machine, to figure out how swapping out different parts of the enzyme change how well it works at breaking down the inedible sugar into an edible form. Then, in Chapters 3 and 4,

I have collaborated with members of the Dumesic lab on deepening our understanding of their method for adding chemicals to the inedible sugar to make it easier for my enzyme to break down. From this work we learned how well four different versions of the enzyme, with different parts, worked on two types of inedible sugar. Having one specific part made the enzyme nearly ten times better at converting both of the inedible sugars. The collaboration work found a method that was able to predict how well my enzyme converted various inedible sugars after the Dumesic lab's chemical treatment, and then further identified a specific part of the inedible sugar that my enzyme was not able to convert at all.

Our results have advanced the field's understanding of my enzymes and similar enzymes; and produced potentially important tools for grading different chemical treatments and enzymes in the future by identifying a part of the inedible sugar that is difficult to digest and needs further study to understand how to effectively convert it. This thesis provides a foundation for further research into improving enzymes like mine, as well developing tools that could be useful when examining the commercial potential of new chemical treatments. Taken all together, this work represents an advancement in our understanding of how to convert inedible sugars into edible sugars that can be used to produce renewable, carbon neutral biofuels and bolster sustainability and energy security.

## **Chapter 7:**

### **Significance & future directions**

#### **Significance:**

Many of the biofuels and bioproducts currently being developed rely on the deconstruction of lignocellulosic biomass (LCB). This deconstruction is often accomplished through an initial thermochemical pretreatment step which increases the reactivity of LCB for subsequent processing by glycoside hydrolases (GHs). Both of these steps are critical, because pretreatments are either unable to fully deconstruct the LCB or they produce undesirable side products; and GHs have little to no activity on untreated LCB. Because pretreatment and enzymatic hydrolysis are central parts of the biofuels and bioproducts field, better understanding how GHs function, and how they interact with both LCB and pretreatment methods is a key part of advancing the field.

In this chapter I will summarize the work detailed in Chapters 2-5 and highlight their major findings. I will also comment briefly on how the work presented in the chapter accomplishes the goal of better understanding GH functions, and the interactions between LCB, pretreatments, and GHs before discussing possible directions for furthering this research in future work.

Chapter 2 presents work done to better understand the behavior of GHs from family 9 (GH9s) through study of the model cellulase CelR. GH9s are an intriguing family of enzymes because some members (including homologs of CelR) have been shown to interact synergistically with both endo-cellulases and other exo-cellulases. To begin, four constructs consisting of the catalytic domain with the native family 3c carbohydrate binding module

(CBM3c), non-native CBM3a, both CBMs or neither CBM were made and their thermostability, kinetics, and binding were tested. Thermostability was greatest in constructs containing CBM3c, with a  $T_m$  25°C higher than the catalytic domain alone. CBM3a modestly improved thermostability (+5°C) in the absence of CBM3c. CBM3c is also critical for CelR to properly function, as shown by the fact that the only constructs for which activity showed saturation kinetic behavior on phosphoric acid swollen cellulose (PASC), an amorphous cellulose, were those containing CBM3c, while addition of CBM3c increased catalytic efficiency on Avicel, a crystalline cellulose, 20×. CBM3a provided a 2-5× increase in catalytic efficiency on cellulose if CBM3c was also present but had negligible effect if CBM3c was absent. Binding to PASC was equally tight for all constructs, regardless of whether a CBM was present or not. Binding to Avicel was improved by CBM3c but was improved much more by the presence of CBM3a.

Chapter 2 also reports on three novel structures of CelR, one of the catalytic domain and CBM3c without substrate, and two of the catalytic domain alone with an E439Q active site mutation bound to either substrate cellohexaose or product cellobiose (after a slow turnover of cellohexaose in the crystal). These structures provide new evidence that multiple residues, especially H226 and D278, in the active site channel rearrange to alternatively facilitate substrate binding or product release. The motion of these two residues appears to be facilitated and coordinated by a nearby calcium ion that is highly conserved in GH9.

Chapter 2 advances the biofuel and bioproduct field by thoroughly characterizing CelR as a model GH9 and examining how it interacts with multiple arrangements of CBM domains. This work also provides the first structures of CelR and identifies a potential mechanistic role of the well-conserved calcium ion near the active site cleft, which has not been focused before my thesis work. In addition to providing a starting point for further examination of the role of this

calcium ion in GH9s, perhaps being relevant to processivity, the detailed kinetic measurements made in this work allowed for the calculation of  $k_{\text{off}}$  and the theoretical processivity of CelR, defined as the ratio of  $k_{\text{cat}}/k_{\text{off}}$ . CelR's theoretical processivity fell well short of values for other well-known processive cellulases from GH6 and GH7 as well as values from putatively processive cellulases from GH5, indicating that CelR is at best only conducting short processive runs, and may not be processive at all. This is the first time that the theoretical processivity for a GH9 has been calculated, and again provides impetus for others in the field to adopt this metric as a potent tool for evaluating processivity in this family.

Chapter 3 presents collaborative work done with the Dumesic lab to better understand the interactions between CelR as a model cellulase and  $\gamma$ -valerolactone (GVL) pretreated LCB. Solid-state nuclear magnetic resonance (ssNMR) spectra were collected, and the ratio between peaks in the split 4-carbon resonance ( $\chi_{\text{NMR}}$ ) was found to represent the portion of cellulose at the surface of a microfibril that does not interact with lignin or hemicellulose; and correlated with the yield of reducing sugar after hydrolysis by CelR. This correlation held across all of the LCBs and organic-aqueous cosolvent systems tested. The ssNMR data also indicated that drying, commonly done to extend sample shelf-life for storage, resulted in a collapse of the cellulose pore structure; and the lower accessibility of the cellulose to CelR corresponded to lower hydrolysis yields.

Chapter 3 provides a clear example of the utility of ssNMR as a technique for probing the chemical and physical changes that LCB undergoes throughout the deconstruction process. The  $\chi_{\text{NMR}}$  metric is a potentially valuable tool for developing or refining dilute acid pretreatment techniques, given the correlation with yields from enzymatic hydrolysis. Chapter 3 also highlights the importance of sample storage and handling on hydrolysis yields, which has



implications for the design of biofuel refineries, as well as experimental design for biofuels and bioproducts research involving enzymatic LCB hydrolysis.

Chapter 4 presents collaborative work done with the Dumesic lab to further investigate the split 4-carbon resonance discussed in Chapter 3. Spectral deconvolution was used to identify chemical environments within the cellulose microfibrils and track changes throughout GVL pretreatment and enzymatic hydrolysis. This data showed the complete removal of cellulose-xylan interactions, and a rearrangement of cellulose in the microfibrils into the more crystalline subdomains, especially  $I_{\beta}$ , and a corresponding decrease in paracrystalline and inaccessible surface material. This rearrangement is particularly noteworthy, as the  $I_{\beta}$  subdomain was consistently resistant to hydrolysis by not only the model cellulase CelR, but the commercial cellulase cocktail CTec3. Rearrangement of cellulose to the  $I_{\beta}$  crystal form is understandable, as it is more thermostable than the other forms of cellulose and has been previously shown to form from heating cellulose.

ssNMR data was also used to calculate the lateral fibril dimension (LFD) and lateral fibril aggregate dimension (LFAD) through the ratios of accessible and inaccessible surface signal intensities. LFD corresponds to the microfibril dimensions, while LFAD corresponds to the macrofibril dimensions. LFD increases throughout both GVL pretreatment and CelR hydrolysis, while LFAD decreases during pretreatment but increases again after enzyme hydrolysis; this indicates that microfibrils are either expanding or growing throughout conversion and that macrofibrils are disrupted by GVL pretreatment but reaggregate during the course of CelR hydrolysis. Wide-angle X-ray scattering (WAXS) data showed the spacing between cellulose chains in the microfibril actually slightly decreased, which means that the LFD increase is likely

due to microfibril growth. This growth could be explained by partially digested microfibrils coalescing to form larger microfibrils throughout the course of deconstruction.

Chapter 4 further fills the knowledge gap of how cellulose structures change in response to thermochemical and enzymatic deconstruction by applying spectral deconvolution techniques to ssNMR spectra. In addition to the value of the methods and protocols developed in this work, the identification of  $I_{\beta}$  as a cellulose subdomain that is particularly hydrolysis resistant has wide implications for the field as a whole, highlighting the importance of considering cellulose rearrangements when developing pretreatment strategies or identifying enzyme candidates for use in cellulase cocktails. Chapter 4 also provides molecular insight that could help explain why cellulose reactivity decreases as it is hydrolyzed through aggregation and condensation of partially hydrolyzed microfibrils into larger micro- and macrofibril structures.

Chapter 5 presents a proposal for a novel framework for characterizing GH multifunctionality consisting of three tiers with distinct features. The first tier of multifunctionality encompasses catalytic domains which possess a broad specificity range, which is a more traditional view of multifunctionality. The second tier represents catalytic domains whose activity is enhanced or adjusted by the addition of non-GH domains. This tier includes CelR, and other processive GH9s, as processivity is thought to sharply decrease with the removal of their CBM. The last tier of multifunctionality describes constructs with multiple catalytic domains and non-GH domains that work together synergistically.

This way of thinking about multifunctionality is beneficial to the field because it provides specificity to discussions around the subject, while also broadening what is considered multifunctionality. This model also helps direct discussion and research in the field towards considering the effects of mixing the various tiers of multifunctionality, for example replacing a

highly specific catalytic domain from a tier three enzyme with a more broadly specific catalytic domain from tier one. Using this framework could inspire a great deal of novel work in the field not only through mixing tiers of multifunctionality in recombinant designs, but also through highlighting aspects of the tiers that are ripe for further study and engineering efforts.

**Future directions:**

Further structural and mechanistic study of CelR is one of the most interesting directions for future studies to head. Collaborating with the NMR Facility at Madison (NMRFAM) to monitor LCB hydrolysis could provide direct molecular evidence for or against the processivity of CelR. Additionally, the roles of residues like the highly mobile H226, W330 and Y435 residues along the active site cleft, as well as motion around the active site calcium could be interrogated through labeling specific amino acids. Another method for interrogating the mechanism of CelR and other potentially processive cellulases would be time-resolved serial X-ray crystallography, which could help probe the degree of movement and positional flexibility of H226 in particular, as well as determining how coordinated movement of residues around the active site calcium are. The added benefit of both these directions for future work is that the techniques developed could be applied to other GHs, especially potentially processive GH5s or other GH9s that have been previously characterized.

More mutational studies could also be rewarding. Of particular interest would be H226W or H226 deletion mutations, as other homologous GH9s have either a tryptophan or no aromatic residue at this position. These mutants, in conjunction with the previously mentioned mechanistic studies, could answer the question of whether the mobility of H226 is retained when

a tryptophan occupies this position, as well as whether removal of H226 altogether significantly alters CelR's function.

The work with hydrolysis of GVL pretreated LCB could be continued by expanding the conditions examined. Testing whether the correlation between  $\chi_{\text{NMR}}$  and enzymatic hydrolysis yields holds for other pretreatment methods (such as Cu-AHP, AFEX, etc.) would be an important piece of follow-up work that would more clearly define the utility of  $\chi_{\text{NMR}}$  as a metric for evaluating pretreatments. If the correlation holds across multiple pretreatments then it becomes even more powerful as a tool than it already is, and if the correlation does not hold then researchers interested in using it have a clearer understanding of when it should be applied.

Expanding the ssNMR techniques to other pretreatment methods would also be beneficial for the subdomain deconvolution described in Chapter 4, as it could identify which pretreatments result in increased  $I_{\beta}$  content. This could be coupled with hydrolysis of the pretreated LCB samples with one or multiple enzymes, alone or in cocktails, to determine whether any enzymes show the ability to react with  $I_{\beta}$  cellulose. In this vein, expanding the conditions and biomasses evaluated in this manner with GVL pretreatment could provide interesting data, especially a more detailed screen of desaccharification temperatures, to test trade-offs between general enzymatic reactivity of the pretreated material and production of  $I_{\beta}$  cellulose during the pretreatment. Additionally, tracking subdomains at even higher hydrolysis levels could provide further insight into whether  $I_{\beta}$  is not hydrolyzed at all or just hydrolyzed much more slowly than other subdomains. It could also be beneficial to leverage cryo-electron microscopy alongside WAXS experiments to verify the cellulose fibril behaviors indicated by the LFD and LFAD results in Chapter 4.

## Appendix A:

### **Drop-NIMS: A combinatorial droplet microfluidic device integrated with NIMS mass spectrometry analysis**

Noel Ha, Jenny Onley, Markus de Raad, Kai Deng, Peter W. Kim, Fangchao Song, La Zhen Han, Peter Andeer, Mai Pham, Kenneth Sale, Nathaniel Kuch, Mark Kutschke, Alex Parker, Brian Fox, (Paul Adams), Trent R. Northen [Draft list of authors, not necessarily complete or in order.]

This work will be submitted to *Lab on a Chip*

Nathaniel Kuch made the following contributions: preliminary enzyme assays, enzyme cloning and expression, writing and review of manuscript

#### **Introduction:**

Coupling microfluidic technologies with high specificity analytical methods is desirable for a wide number of biotechnological and pharmaceutical applications (de Rond et al. 2015; Kempa et al. 2019; Ha et al. 2021). Most mass spectrometry methods coupled with microfluidics have focused on electrospray ionization (ESI) mass spectrometry (Ha et al. 2021). However, there are several reports of coupling microfluidics with laser desorption ionization (LDI) mass spectrometry which has potential for high throughput and small sample volumes. Microfluidics coupled with matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) is a powerful technique (Haidas et al. 2019; Xu et al. 2021) although the required addition of matrix adds complexity. Lapierre et al. successfully coupled digital microfluidics with matrix-free mass spectrometry but this method was low throughput (Lapierre et al. 2011). Alternative methods that enable high throughput and matrix-free mass spectrometry directly from a microfluidics device are desirable. For an in-depth review of coupled microfluidics and mass spectrometry techniques that are relevant to biotechnology, see Ha et al. 2021.

One surface-based matrix-free mass spectrometry technique is nanostructure-initiator mass spectrometry (NIMS), a powerful technology for rapid and sensitive analysis of a variety of samples. Target samples are deposited onto a specialized surface and analyzed by irradiating with a laser. NIMS is amenable to high throughput applications since only small (nanoliter to microliter) sample volumes are required. One useful application of NIMS is the characterization of enzymes (Deng et al. 2018; de Rond et al. 2019; Glasgow et al. 2020). However, high throughput methods for setting up enzymatic reactions and transferring them to the NIMS chip surface are desired. A device that integrated microfluidics with NIMS would enable reactions at nL volumes and further improve scale-up. Heinemann et al. developed a device that excelled at droplet manipulation, but it was difficult to scale due to the complexity of fluid handling. Additionally, samples had to be premixed, limiting the number of combinations that could be tested simultaneously (Heinemann et al. 2017). Recently the Blainey lab reported a droplet microfluidics device for spectroscopic imaging (Kulesa et al. 2018; Kehe et al. 2019). The device randomly combines oil-encapsulated droplets, making it particularly appealing for integrating with mass spectrometry due to the potential for many novel chemical combinations.

This paper describes a new device (Drop-NIMS) where the Blainey lab droplet microfluidic design was adapted and coupled with NIMS to introduce mass spectrometry-based analysis. Here, Drop-NIMS was leveraged to sample the combinatorial space of enzymes and substrates (i.e., enzymatic targets). Barcodes (compounds with unique masses) were used to track the droplets so that multiple substrates and enzymes could be tested simultaneously. After droplets were pairwise combined and merged, the oil encapsulating the droplets evaporated and the samples deposited onto the NIMS chip for mass spectrometry imaging. Drop-NIMS was applied to assay the substrate specificity of glycoside hydrolases (GHs), enzymes relevant to

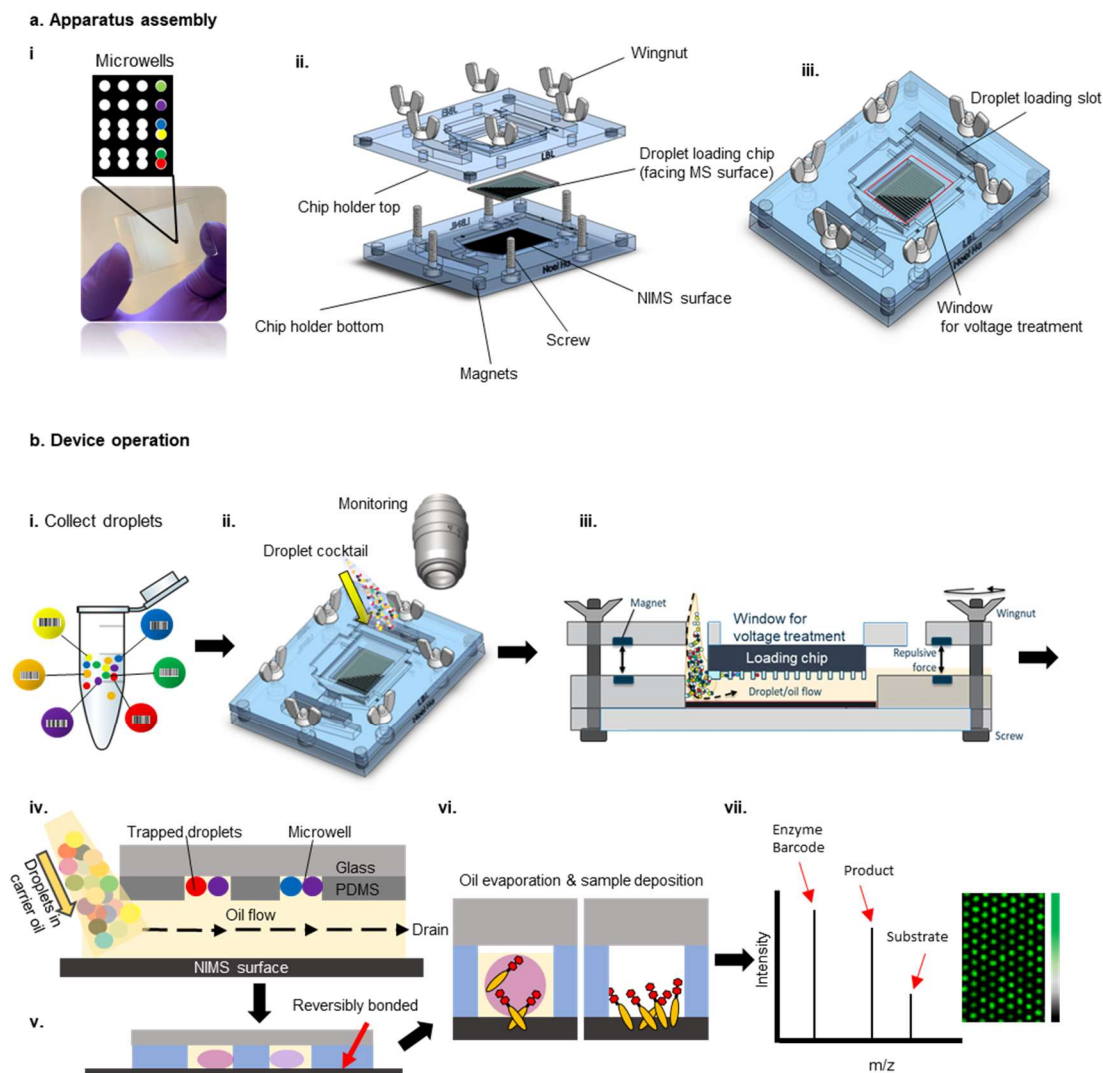
multiple industries including food, household products, and biofuel production. Five GHs and four substrates were screened with effectively 30 unique combinations and up to 90 replicates per combination. This study demonstrated that Drop-NIMS can be used to characterize enzymes in a high throughput and detailed manner, with relative ease of device construction and use.

## **Materials and methods:**

### *Microfluidic droplet assay chip design and fabrication*

The microwell array chip, similar to previously-published designs (Kulesa et al. 2018; Kehe et al. 2019; Fig. A.1a), is made of a polydimethylsiloxane (PDMS) layer bonded on a glass slide. The PDMS layer contains an array of 1,120 microwells (35 columns by 32 rows).

Variations in the microwell array chip were created wherein the wells were in groups of one, two, three, or four (Fig. SA.4) cylindrical sub-wells. The sub-wells are approximately 145  $\mu\text{m}$  in diameter to entrap droplets with a diameter of 110  $\mu\text{m}$  to 130  $\mu\text{m}$ . The microwells are spaced 500  $\mu\text{m}$  apart to prevent cross contamination. The PDMS layer was made using silicon wafer molds (SU-8 2075 from Microchem) fabricated by photolithography per manufacturer's recommended protocol (see Supplemental Methods for more detail). A photo mask was designed for the fabrication of four chips per silicon wafer. The molds were coated with Trichloro (1H,1H,2H,2H-perfluorooctyl) silane by vapor deposition to make the surfaces hydrophobic.



**Figure A.1:** Assembly and operation of microfluidics device. Not to scale. a) Assembly. i. The microwell array chip is composed of a glass layer and a thin PDMS layer containing microwells with 1, 2, 3, or 4 sub-wells. Inset contains example but does not show all wells. ii. The microwell array chip is attached to the top portion of the apparatus and the NIMS chip is attached to the bottom portion of the apparatus. iii. The screws are tightened so that the microwell array chip is close to the NIMS chip. b) Operation of Drop-NIMS. i. Micro-droplets are formed with the QX200 Droplet Generator (BioRad) and pooled. Each enzyme droplet contains a known compound to “barcode” the droplet. ii-iv. Droplet cocktail is transferred to Drop-NIMS (top, side views, and zoomed view). iii. Droplets are buoyant and randomly assemble into the microwells on the PDMS layer as they are washed through with oil. v. Screws are tightened to reversibly seal the PDMS layer against the NIMS surface, and an electric field (not shown) is applied to merge droplets within the microwells. vi. Enzymatic reaction occurs within the merged droplets. After 2 to 3 hours the carrier oil and sample buffer evaporate through the gas permeable PDMS and the droplet contents are deposited onto the NIMS surface. vii. The apparatus is disassembled, and the NIMS chip undergoes mass spectrometry imaging.



To enable microscope imaging, the PDMS layer (approximately 120-150  $\mu\text{m}$  thick) was bonded on the glass substrate. While thickness was not completely uniform, the wells were appropriately sized such that the sub-wells accommodate only one droplet each. PDMS mixed with the curing agent (Dow Corning SYLGARD 184 Silicone Elastomer Kit; 10:1 mixing ratio) was poured on the master wafer and degassed for 30 min. Then, four glass slides (25.4 mm W x 35mm L x 1mm T; precleaned with acetone and isopropyl alcohol) were placed on top of the PDMS-covered mold, covering the respective microwell feature for each slide. The glass slides and the wafer were clamped using 2 neodymium magnets, such that the features on the master came into contact with the glass slide to create through-hole membrane structures. The PDMS was cured for 2 h at 80  $^{\circ}\text{C}$  and the PDMS/glass chip was separated from the mold. A new PDMS chip was used for each oil phase experiment to avoid cross-contamination. For more details on chip design and fabrication, see supplemental methods.

#### *Device assembly and operation*

A custom two-part apparatus was designed and 3D printed to hold the chips together and load the droplets (Fig. A.1a). The microwell array chip was taped to the top part, which had a “window” for viewing and applying an electric field for merging the droplets (Fig. A.1a.ii). A NIMS chip was prepared as previously described (Northen et al. 2007; Woo et al. 2008) and taped to the bottom part (Fig. A.1a). Nylon screws and wingnuts were used to prevent the applied electric field from being concentrated away from the chip while merging the droplets. The screws were tightened so that the microwell array chip was near but not sealed against the NIMS chip. The gap between the two parts was maintained by a balance between the wingnuts and repulsive force from the magnets.

Droplets were produced using the Bio-Rad QX200 Droplet Generator (Pleasanton, CA) according to manufacturer's instructions with a fluorocarbon oil (3M Novec 7500, C<sub>9</sub>H<sub>5</sub>F<sub>15</sub>O) containing 2% 008-FluoroSurfactant (RAN Biotechnologies). Droplets were pooled to prepare a total of 200-400  $\mu$ L (approximately 160,000 droplets) of suspension and gently mixed. The gap between the microwell array chip and the NIMS chip was primed with surfactant-free fluorocarbon oil before the injection of droplets. Droplets were washed with the surfactant-free oil through the apparatus, flowing between the microwell array chip and the NIMS chip (Fig. A.1b iii-iv; Supplemental Video 1). The entire apparatus was gently tilted side to side to allow droplets to randomly fill the microwells, with up to 2,240 droplets (two in each of the 1,120 microwells) in a double-well (2x) droplet design. Excess droplets were washed away with the surfactant-free oil. The wingnuts were tightened until the microwell array chip and the NIMS chip were sealed reversibly against one another (Fig. A.1b.v). The droplets were merged via electrocoalescence as described by Kulesa et al. 2018. Briefly, an electric field was applied to the glass side of the microwell array chip for 10-20 seconds to merge the droplets using a hand-held corona treater (Electro-Technic Products BD-20AC) (Fig. SA.3).

Enzymatic reactions occurred inside the merged droplets after the chip was sealed against the MS surface. The volatile carrier oil and solvent evaporated through the gas-permeable PDMS, inducing sample deposition on the NIMS chip surface and halting the enzymatic reactions. After complete evaporation of oil and solvent, the NIMS chip was separated from the droplet microwell array chip for mass spectrometry imaging (MSI). For more details on device assembly and operation, see supplemental methods.

### *Cell-free protein expression*

Putative xylanases (a subtype of glycoside hydrolase) were identified through the National Center for Biotechnology Information Center (NCBI) Basic Local Alignment Search Tool (BLAST) by searching for sequences similar to known xylanase genes. Identified gene sequences were synthesized at the Joint Genome Institute (Walnut Creek, California) and cloned into pEU (plasmid of Ehime University, Matsuyama) a vector for cell-free expression (Sawasaki et al. 2002; Takasuka et al. 2014). Expression was done as described previously (Takasuka et al. 2014) with sequential transcription-translation using wheat germ extract from Cell-Free Sciences (Yokohama, Japan). Briefly, transcription was done using the SP6 polymerase (Promega, Madison, WI, USA) and translation was done with a bilayered, diffusion-fed translation reaction. Protein production was quantified using band analysis on stain-free sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS PAGE) gels by comparison to a cell-free translation blank.

#### *Preliminary activity screen and cloning*

Successfully translated protein was screened for activity on 1,4-beta-D-mannan and barley beta glucan, low viscosity from Megazyme (Bray, Ireland), and beechwood xylan from Sigma-Aldrich (St. Louis, Missouri); data not shown. Reactions were prepared with 10 mg/mL substrate, 4  $\mu$ L cell-free translation reaction, and 0.1 M acetate buffer at pH 4.5, 5, or 6 to a final volume of 40  $\mu$ L. Reactions were incubated at 32°C for 2 hours before centrifugation for 10 min at 4,000 rpm (3283 xg) to pellet residual solids and the concentration of soluble reducing sugar was determined via the Pierce BCA Protein Assay Kit from Thermo Fisher Scientific (Waltham, Massachusetts). Briefly, 200  $\mu$ L of BCA working solution and 10  $\mu$ L of reaction supernatant were heated at 80°C for 30 min. Control samples contained either cell-free translation only,

substrate only, buffer only, or the cell-free translation blank and substrate. A glucose standard curve was used to convert absorbance units to mg/mL reducing sugar in solution.

### *Protein expression and purification*

Four enzymes with sequence similarity to glycoside hydrolase family 43/subfamily 11 (GH43-11) enzymes were selected for characterization using NIMS analysis. Rosetta cells from Sigma Aldrich were transformed with plasmid containing the selected genes (pVP67K, N-terminal 8His TEV cleavable tag) and expressed via autoinduction as described previously (Studier 2005; Fox and Blommel 2009; Blommel et al. 2007). Centrifuged cells were lysed via lysozyme and sonication; the lysate was centrifuged for 1 hour at 20,000 rpm (31,360 x g) and the supernatant was filtered through a 0.4  $\mu\text{m}$  PES filter. Enzymes were purified from the filtered supernatant via nickel affinity chromatography on an Akta Start FPLC and subsequently desalted and concentrated. Concentrated enzyme solution was frozen in liquid nitrogen and stored at -20°C until later use. The source organisms and accession numbers are as follows: 43-Lact (*Lactobacillus kimchicus*, KRK47366.1), 43-Pedi (*Pediococcus acidilactici*, SJM45318.1), 43-Weis (*Weissella* sp., COI51269.1), and 43-Clos (*Clostridium* sp., SCJ27493.1). The sequence for 43-Weis was extracted from a *Streptococcus pneumoniae* genome, but later bioinformatic analysis showed evidence that the sequence is a contaminating sequence from *Weissella* sp.

### *Droplet enzyme reactions*

Enzymatic reactions were performed at room temperature. A previously-characterized glycoside hydrolase (CelEcc-CBM3) and four previously-uncharacterized glycoside hydrolases were chosen for proof-of-concept. Four NIMS-tagged model substrates were tested (Fig. SA.7),

each with an attached perfluorinated tail: cellobiose (G2, 1102.3 or 1135.2 m/z), xylobiose (X2, 1042.3 m/z), xylotriose (X3, 1189.4 m/z) and xyloetraose (X4, 1321.4 m/z). The drug Verapamil (455.3 m/z) and a previous version of NIMS-tagged cellobiose (1101.8 m/z) were used to test the chip without enzyme. CelEcc\_CBM3a (Walker et al. 2015) or a commercial xylosidase (Megazyme cat. # E-BXSEBP) were used as positive controls, while a commercial glucosidase (catalog #49290-250 mg; Sigma-Aldrich,) or buffer alone were used as negative controls. Pre-tagged NIMS substrates (G2, X2, X3, and X4) were prepared as described previously (Deng et al. 2012; Deng et al. 2014). Mass barcodes were included in enzyme droplets in order to track their locations on the chip. The following were used as mass barcodes from Cambridge Isotopes (Fig. SA.8): L-Carnitine:HCl, O-Dodecanoyl (unlabeled) (cat. # ULM-7199-0.1mg); L-Carnitine:HCl, O-Dodecanoyl (N-Methyl-D3) (cat. # DLM-8162-0.1mg); L-Carnitine:HCl, O-Dodecanoyl (N,N,N-Trimethyl-D9) (cat. # DLM-8215-0.1mg); L-Carnitine:HCl, O-Palmitoyl (unlabeled) (cat. # ULM-7738-PK); and L-Carnitine:HCl, O-Palmitoyl (N-Methyl-D3) (cat. # DLM-1263-0.01). Palmitoyl-<sup>13</sup>C<sub>16</sub>-L-carnitine hydrochloride (cat. # 644323-1.00mg) was purchased from Sigma. Enzyme droplets were generated with a concentration of 733 ug/mL enzyme and 250 μm barcode in 25 mM phosphate buffer, pH 6.0 (366.5 ug/mL enzyme and 125 μm barcode after droplet merging). Substrate droplets contained 1 mM substrate in 25 mM phosphate buffer, pH 6.0 (0.5 mM substrate after droplet merging).

### *Mass spectrometry Imaging (MSI) and Data Analysis*

Buoyancy-based rapid droplet loading followed by sample deposition via evaporation of carrier oil was confirmed using MSI. MSI was performed using an ABI/Sciex 5800 TOF/TOF mass spectrometer with a laser intensity of 2500-3000 over a mass range of 250–1500 Da. Each

position accumulated 20 laser shots. The instrument was controlled using the MALDI-MSI 4800 Imaging Tool using a 50  $\mu\text{m}$  step size between shots. Spectra were recorded in positive reflector mode. The instruments were calibrated using Anaspec Peptide Calibration mixture 1 (catalog #AS-60882; Anaspec, Fremont, CA).

For a quantitative analysis, OpenMSI Arrayed Analysis Toolkit (OMMAT) (De Raad et al. 2017) was used to identify wells displaying peaks for  $m/z$  values of interest. Due to the irregularly-shaped sample spots, the OMMAT script was modified to exclude pixels that did not contain any peaks of interest above a relative signal intensity of 100 arbitrary units (au). The data were then analyzed with a customized script written with pandas (The pandas development team 2021; McKinney 2010) and Pingouin (Vallat 2018).

## **Results & Discussion:**

### *Device Design*

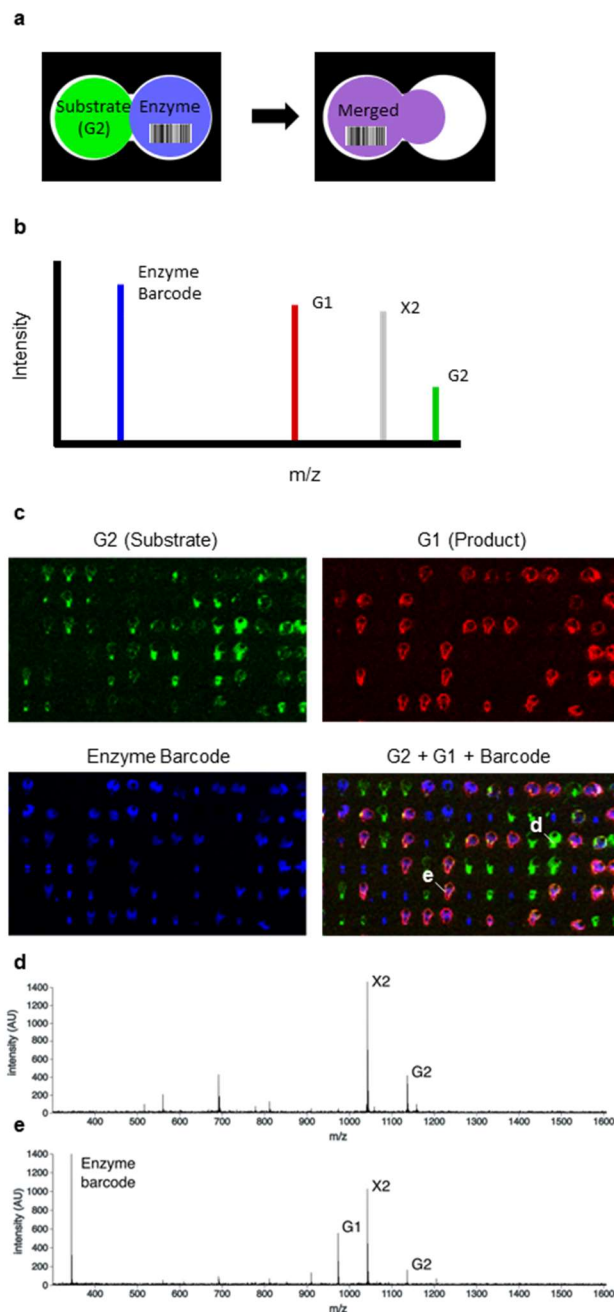
The main objectives of Drop-NIMS are to randomly combine substrate and enzyme droplets and to deposit the mixtures onto the NIMS chip surface. To accomplish random droplet combination, a microwell array chip was created as described earlier. Some wells were observed to not be completely through-hole, which can be explained by the variability in the height of the pillars on the master. Moreover, the glass slide surface is not a perfect plane which can result in this observation as well. However, this was not observed to affect their ability to trap droplets. After droplet loading, samples deposited onto the NIMS chip surface as the oil evaporated, and the apparatus was disassembled for NIMS chip imaging. Droplet loading and merging were tested with droplets containing dye, which were visualized with a stereomicroscope. The microwell array chip was interfaced with the NIMS chip (Fig. SA.3) or with indium tin oxide

(ITO) glass in place of the NIMS chip (Fig. SA.5). In the 2x design (i.e., 2 sub-wells in each microwell), the majority of doublet wells contained two whole droplets and appeared fully merged after electrocoalescence. It was demonstrated that droplet loading and merging works for 2x, 3x, and 4x arrangements as well (Fig. SA.4).

### *Device Operation*

The successful operation of Drop-NIMS was initially demonstrated with two molecules: NIMS-tagged cellobiose (“G2”; 1102 m/z) and verapamil (455 m/z). Microwell array chips with 1x, 2x, 3x, and 4x microwells were used (Fig. SA.2a). Two different chips were used for G2 and for verapamil. Droplets were loaded into the assembled apparatus with the microwell array chip and the NIMS chip as described above. After the oil evaporated, the NIMS chip was removed and imaged using mass spectrometry imaging (MSI). OpenMSI images demonstrated detection of G2 (Fig. SA.2b) and verapamil (Fig. SA.2c) with good signal to noise ratios, indicating that sample deposition was successful.

The device was then applied to enzymatic reactions, using droplets containing a previously-characterized enzyme and droplets containing enzyme substrates. The enzyme was CelEcc\_CBM3a, which is known to break down cellulosic compounds (which are composed of repeating glucose units) (Walker et al. 2015), and cellobiose (two glucose units; “G2”) was used as the enzyme substrate (Fig. A.2a). Droplets containing G2 were tracked by detection of a substrate peak (G2) and/or a product peak (glucose; G1).



**Figure A.2:** Droplet enzyme reaction with mass spectrometry barcode. a) Experimental setup schematic.

Two types of droplets were prepared; one containing cellobiose (G2) and xylobiose (X2) and one containing CelEcc\_CBM3 enzyme with a barcode (DC-1). One possible combination is shown (other combinations are possible). b) Simulation of expected mass spectrometry spectrum with the enzyme substrate (G2), enzyme product (G1), and the barcode for tracking the location of enzyme droplets. c) OpenMSI images of NIMS chip after sample deposition. d-e) Sample spectrum from OpenMSI on spots without (d) and with (e) enzyme. Green indicates signal intensity at 1135  $m/z$  (G2; substrate), gray indicates signal intensity at 1042  $m/z$  (X2; substrate); red indicates signal intensity at 973  $m/z$  (G1; product), and blue indicates signal intensity at 344  $m/z$  (DC-1; barcode).



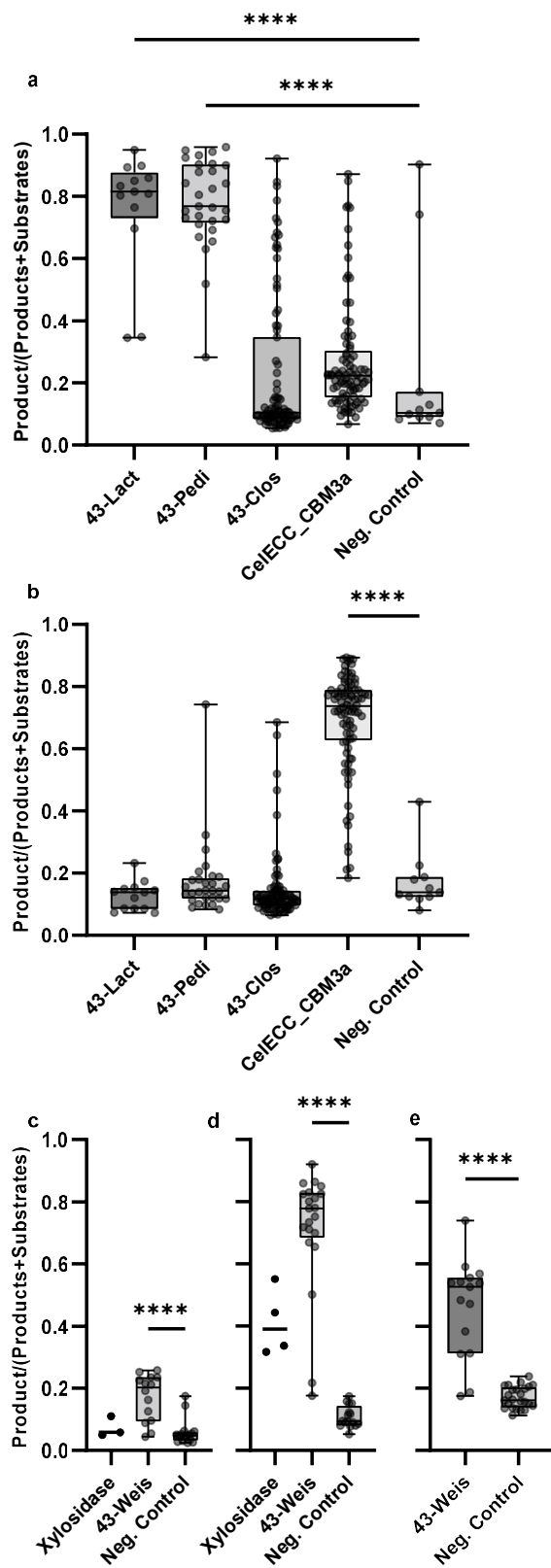
An extra challenge with droplets lacking enzyme substrates (i.e., containing enzymes or buffer only) is that there is no distinct mass spectrometry signal in the selected mass range for tracking the droplets. In order to track enzyme or buffer droplets, compounds of known size (“barcodes”) were added to label the droplets. Thus the location of an enzyme or buffer droplet could be determined by the detection of the mass barcode. Barcodes with the same structures but different isotopes are desirable since they ionize equally in the mass spectrometer (and thus are detected with relatively similar frequencies). Barcodes in this experiment were various isotopes of palmitoyl carnitine or dodecanoyl carnitine (Fig. SA.8), which display high ionization efficiency on NIMS.

Drop-NIMS was loaded with droplets containing either CelEcc\_CBM3a and barcode or G2 substrate. Substrate (G2), product (G1), and barcode (for tracking the enzyme) were detected (Fig. A.2c). The G1 product was co-localized with the barcode, indicating that the conversion of G2 to G1 only occurred when enzyme was present. These results demonstrate that the device successfully randomly combined droplets, enabled subsequent enzymatic reactions, and then deposited the reactions onto the NIMS chip surface.

Drop-NIMS has several advantages over our previous NIMS-digital microfluidics device (Heinemann et al.; Ha et al. 2021). Enzymes and substrates can be loaded in separate droplets and randomly combined within the device, rather than premixing solutions. Randomized droplet combinations reduces manual pipetting and creates the potential for many more tested combinations. The simple design of the microwell array means many more samples can be added with minimal effort. Once fabricated, microfluidics expertise is not required to operate the apparatus.

### *Application to Uncharacterized GH Enzymes*

Drop-NIMS was applied to test multiple previously uncharacterized GH enzymes with G2 and xylobiose (X2) on a single chip. GH enzymes with similarity to enzymes in the GH43 family subfamily 11 (enzymes from *Lactobacillus*, *Pediococcus*, and *Clostridium* were labeled 43-Lact, 43-Pedi, and 43-Clos, respectively) were tested. Characterized enzymes in the GH43-11 subfamily are known to target xylooligosaccharides (composed of repeating xylose [X] units) but not G2. Droplets containing 43-Lact, 43-Pedi, 43-Clos, CelEcc\_CBM3a, and buffer alone were loaded along with “substrate droplets” containing a mixture of G2 and X2. Barcodes were included in droplets containing enzymes or buffer to facilitate droplet tracking. Two of the three new enzymes (43-Lact and 43-Pedi) converted X2 substrate to X1 product, while no significant conversion of X2 was observed in droplets containing 43-Clos, CelEcc\_CBM3a, or buffer (Fig. A.3a). As expected, only CelEcc\_CBM3a converted G2 to G1 (Fig. A.3b).



**Figure A.3:** Rapid screening of glycosyl hydrolases with model hemicellulose compounds. Box and whisker plots. a-b) Results of a chip screening multiple enzymes (GH 43-Lact, 43-Pedi, and 43-Clos with CelEcc\_CBM3 as a positive control and buffer as a negative control) with one substrate droplet (containing a mix of G2 and X2). The proportion of products are shown for X2 conversion to X1 (a) and G2 conversion to G1 (b). c-e) Results of a chip screening one enzyme (GH 43-Weis) with commercial xylosidase as a positive control and beta-glucosidase as a negative control) with multiple substrate droplets (G2/X2 mix, X3, and X4). The proportion of products are shown for X2 conversion to X1 (c), X3 conversion to X1 and X2 (d), and X4 conversion to X1, X2, and X3 (e). Boxes show the 25th to 75th percentiles. The line within the box indicates the median. Whiskers indicate maximum and minimum. All samples were compared to the negative control; \*\*\*\*,  $P < 0.0001$ . Xylosidase samples are shown as scatter plots (horizontal line indicates mean) due to low sample size ( $< 5$ ). Y-axes show the sum of the products divided by the sum of products and substrate (e.g., for X3 reactions, y-axis shows  $[X1+X2]/[X1+X2+X3]$ )

An additional GH enzyme from *Weissella*, 43-Weis, was chosen to further characterize its substrate specificity. A second chip was loaded with droplets containing 43-Weis, glycooligosaccharides (G2), and xylooligosaccharides (X2, X3, or X4). X2 and G2 were prepared within the same droplets whereas X3 and X4 were in separate droplets. Droplets containing a commercial xylosidase (which hydrolyzes X2, X3, and X4) served as positive controls and droplets containing glucosidase (which does not hydrolyze X2, X3, or X4) served as negative controls. Like the previously tested enzymes, 43-Weis converted X2 to X1 (Fig. A.3c) but did not convert G2 to G1 (Fig. SA.6a). X3 was converted to X1 (Fig. A.3d; Fig. SA.6b), and X4 was converted to a mixture of X3 and X1 (Fig. A.3e; Fig. SA.6e). As expected, the glucosidase hydrolyzed G2 to G1 but did not hydrolyze X2, X3, or X4 (Fig. A.3c-e). Commercial xylosidase hydrolyzed X3 to X2 and X1, although there were only four replicates (Fig. A.3d). No xylosidase and X4 combinations were observed. Overall this experiment demonstrated that multiple substrates can be tested on one chip. With this multi-substrate chip, barcodes were required for the substrate droplets in addition to the enzyme droplets to distinguish between X3 and X4 droplets (which can be hydrolyzed to the same products).

Drop-NIMS was used to successfully demonstrate that three of the four tested enzymes have xylanase (X4, X3, and/or X2 hydrolysis) activity but not cellulase (e.g., G2 hydrolysis)

activity. Drop-NIMS can rapidly screen glycoside hydrolase enzymes for their activities against oligosaccharides. Drop-NIMS results along with the protein sequence similarity to other GH43 family subfamily 11 proteins support the classification of 43-Lact, 43-Pedi, and 43-Weis as xylanases that target xylooligosaccharides.

### *Loading Efficiency*

The data from two chips are discussed in Figure 3; one contained multiple enzymes with one substrate droplet type (Fig. A.3a and b), and one contained one enzyme with multiple substrate droplet types (Fig. A.3c-e). For these two chips, sufficient replication and minimal droplet breakage and cross-contamination was observed. Figure S9A shows the number of replicates for each chip. In both experiments, the number of substrate droplets was equivalent to the number of non-substrate droplets, to ensure most wells received substrate. The number of replicates of relevant reactions (i.e., ones that contained a substrate droplet and an enzyme droplet) were typically between 10 and 90 per combination type. However, there is some variability; on the chip for Figure 3C-E, there were very few droplets containing xylosidase. Low replication of xylosidase droplets may be due to poor mixing of the droplets prior to droplet loading.

Figure S9B shows the number of droplets per doublet microwell for both chips. The ideal number is two; less than two indicates that one or both sub-wells were empty, and more than two indicates droplet breakage or cross-contamination (since the microwells can only accommodate two full-sized droplets). Most of the wells for both chips had one or two droplets, and wells containing mass spectrometry signals for more than two droplets were excluded from the analyses shown in Figure 3.

### *Future Directions*

Drop-NIMS has potential for further applications in the future. NIMS technology has been applied to enzymes with a variety of different applications including the detection of other biomass-relevant molecules, drugs, and peptides (Northen et al. 2008; Deng et al. 2018; de Rond et al. 2019; Glasgow et al. 2020). Additionally, the PDMS microwell array can be designed to contain three or more sub-wells. Preliminary work demonstrated successful droplet loading and merging with microwells containing three and four sub-wells (Figs. S2 & S4). Some enzymes work synergistically with each other, and testing combinations of three or more droplets (i.e., multiple enzymes and/or multiple substrates) can yield new mechanistic information. The PDMS microwell array can also be enlarged to include more microwells, further increasing the high throughput power of Drop-NIMS.

### **Conclusions:**

In conclusion, Drop-NIMS is a novel combination of NIMS and droplet microfluidics using standard photolithography techniques that can be used by the broader community to characterize enzymatic reactions. The research presented here demonstrates that the new device successfully combines and merges droplets and deposits the resulting contents onto NIMS chips. Drop-NIMS has a higher throughput than traditional NIMS, carrying out more than 1,000 reactions simultaneously while requiring only a small volume (nL) of enzymes and substrates. A multitude of droplets can be randomly combined for many combinations and replicates. Future iterations of the device can be expanded to contain more microwells and to include more than two sub-wells per microwell. The fabrication of the device requires standard photolithography

equipment, and the device can be used with minimal training. While the present study demonstrated the characterization of glycoside hydrolases, Drop-NIMS has potential for broad applications in high throughput characterization of other enzymes important to the food, biofuel, and pharmaceutical industries.

### **Funding acknowledgements:**

This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work was funded in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC0207ER64494). Nate Kuch was supported by the UW-Madison Biotechnology Training Program (NIH 5 T32 GM008349).

### **Conflicts of Interest:**

Noel Ha, Markus de Raad, and Trent Northen are authors of a patent, US20210102954A1, held by the University of California, covering parts of the method described in this manuscript. Trent Northen is an advisor to Brightseed Bio.

### **Supplemental information:**

#### **Part 1. Chip Master and Microwell Array Chip Fabrication:**

##### *Microfluidic Chip Master Fabrication*

1. Clean Si wafer (4", p-type, <100>, 0-100  $\Omega$ -cm, UniversityWafer Inc.) with a solvent wash.
  - Rinse with acetone, isopropanol, and DI water. For better results, ultra-sonicate the wafer in the solvents for 3-5 minutes and use ultrapure (18 M $\Omega$ ) water. For a deep clean, use piranha etch.
  - Remove leftover water by blow-drying the wafer with nitrogen and a dehydration bake at 150°C for 10 min.

2. Spin-coat negative photoresist (SU-8 2075, MicroChem ) on pre-treated wafer. Refer to the MicroChem Nano™ SU-8 2000 manual to customize the recipe.
  - Secure the wafer in an appropriately sized chuck.
  - Ramp up to 500 rpm at 100 rpm/s and dwell for 10 sec while dispensing the photoresist.
  - Ramp up to 1700 rpm at 300 rpm/s and dwell for 30 sec.
  - Ramp down to 0 rpm at 300 rpm/s.
3. Soft bake the coated wafer at 65°C for 5 min and at 95°C for 25 min, followed by 5 min cooling.
4. Exposure
  - Align the photomask (transparency or chrome-gold) with the wafer using a mask aligner (OAI Model 204).
  - Expose the wafer with a near UV (350 - 400 nm) source for 8 sec (at least 260 mJ/cm<sup>2</sup> dosage).
5. Bake the exposed wafer at 65°C for 5 min and at 95°C for 10 min, followed by 5 min cooling.
6. Develop the resist with SU-8 developer for 10 min. Rinse with isopropanol and DI water followed by drying with a gentle stream of nitrogen.
7. Hard bake at 175°C for 10 min. This step is optional but recommended.
8. Silanization of SU-8 master to prevent PDMS adhesion. This step should be conducted in a solvent hood or a glovebox.
  - The wafer was placed (SU-8 features facing up) in a dessicator along with a vial of Trichloro(1H,1H,2H,2H-perfluorooctyl)silane.

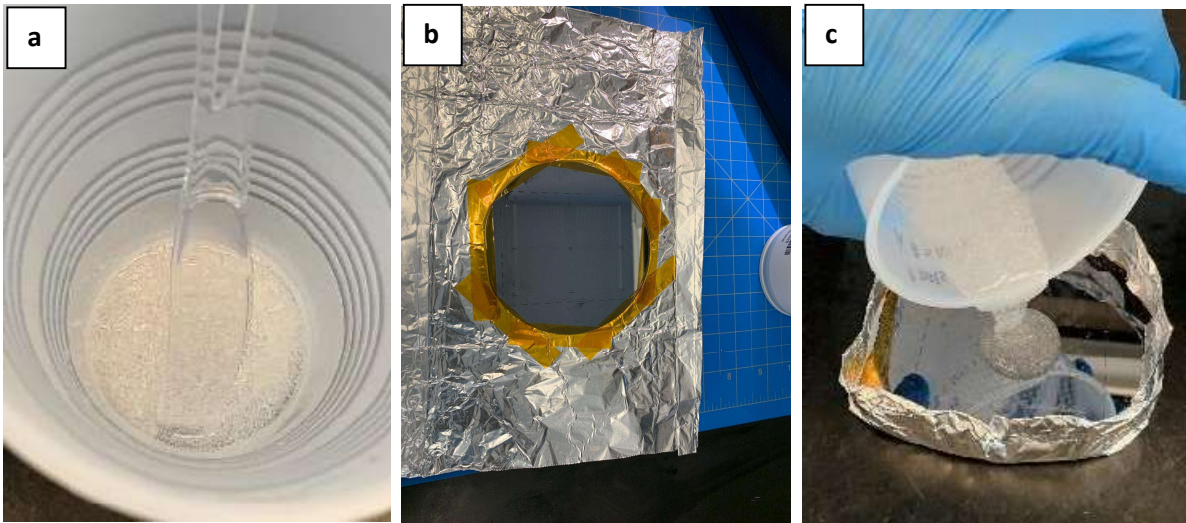


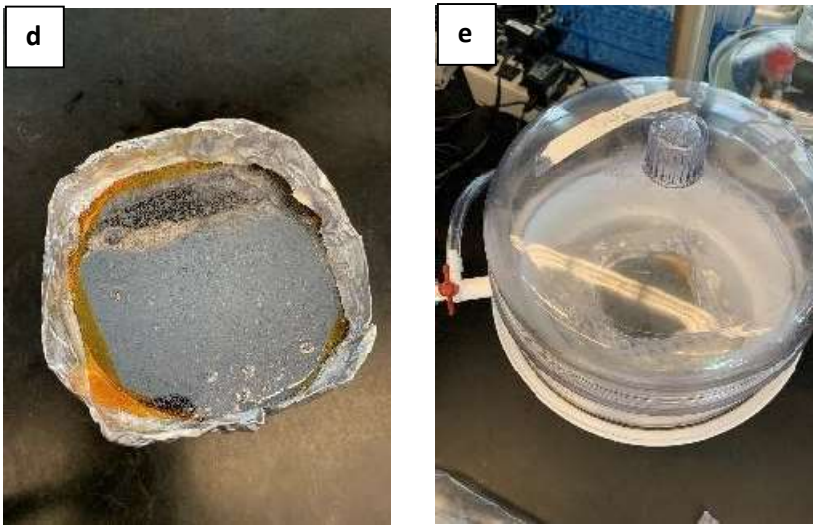
- The dessicator was connected to the vacuum line and left undisturbed for 30 minutes to allow the silanizing agent to evaporate and form a thin layer on the master.

### *Microwell Array Chip Fabrication*

#### 1. PDMS molding:

- Mix PDMS (“Part A”) and curing agent (“Part B”) in 10:1 weight ratio in a clean disposable cup (Fig. SA.1.1a); ~20g of PDMS per mold. PDMS sourced from Dow (SYLGARD™ 184 Silicone Elastomer Kit, material # 4019862).
- Pour mixed liquid PDMS onto a SU-8 master mold (Fig. SA.1.1b-d). SU-8 master mold is silanized (Millipore Sigma catalog #448931) to make the surface hydrophobic to minimize permanent PDMS bonding to the mold.
- Degas for 30 minutes to remove bubbles in a vacuum desiccator (Fig. SA.1.1e).

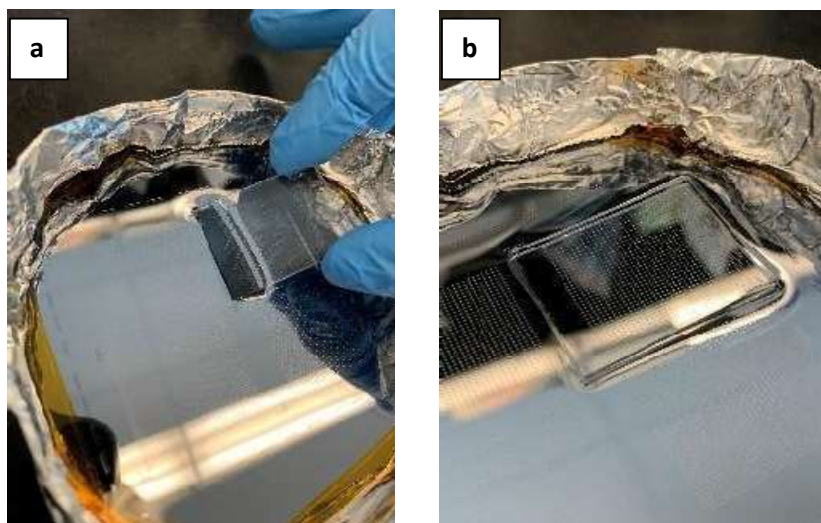




**Figure SA.1.1**

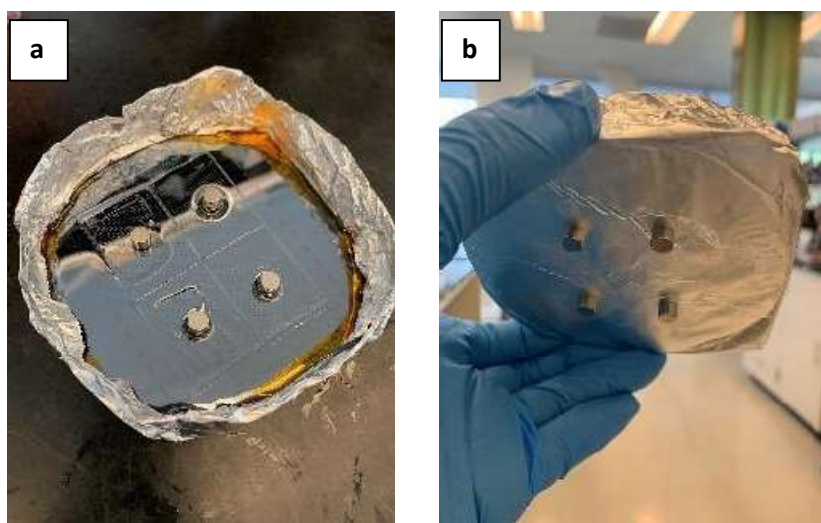
2. Assemble with clean glass slides using magnets

- Cut a glass slide into 25.4 mm W x 35mm L x 1mm T using a diamond scribe.  
(Note: Any commercial glass slides 1" W x 2" L x 1mm T can be used).
- Glass cleaning: Rinse a glass slide with Acetone, isopropyl alcohol, and distilled water. Dry it or wipe with Kimwipe. Blow the dust with an air duster spray.
- Place & align the glass layer on PDMS-filled mold using the alignment marker on the mold (Fig. SA.1.2). Note: 4 droplet chips per mold.



**Figure SA.1.2**

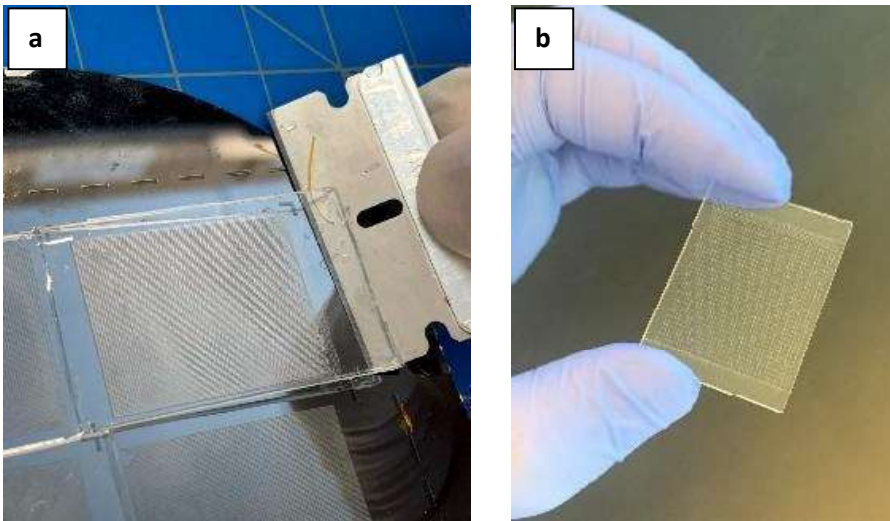
- Secure the glass slide with 2 Neodymium magnets (3/16" Thick, 1/4" OD; McMaster-Carr catalog # 5862K109).



**Figure SA.1.3**

- Repeat the steps above for all 4 positions (Fig. SA.1.3). A total of 8 magnets and 4 glass slides are needed.

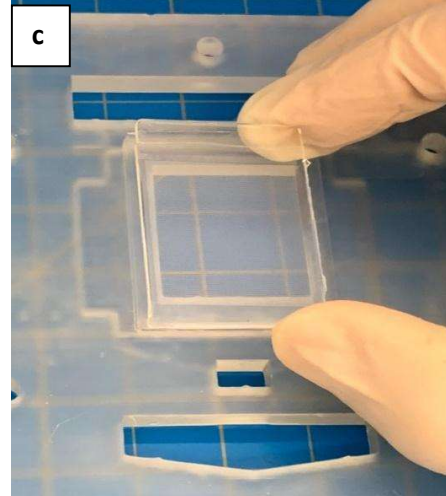
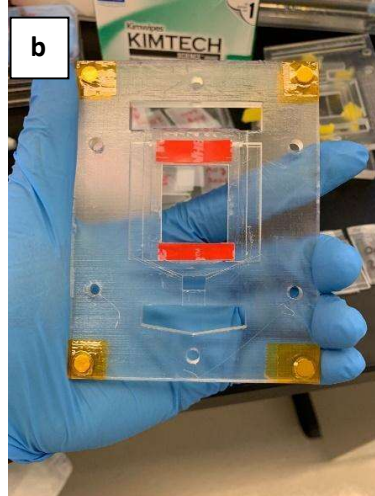
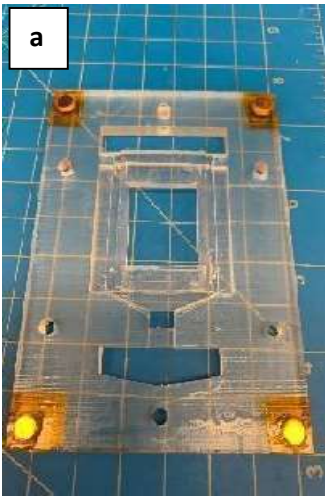
- Degas bubbles overnight in a vacuum desiccator. Note that PDMS will be partially cured at RT overnight.
3. Bake/cure PDMS at 80°C for at least 2 hours (Note: Overnight baking is preferred to fully cure)
  4. After PDMS is cured, separate the droplet chip (Glass/PDMS) from the master mold
    - Remove magnets
    - Remove the surrounding PDMS from the glass surface. Make sure top glass surface is free of PDMS.
    - Gently separate/peel the PDMS/glass chip from the mold using a razor blade (Fig. SA.1.4a). Note: Si master mold is brittle.
    - Trim PDMS (cut ~2mm from both ends to improve sealing against the NIMS chip): The size of PDMS layer should be SMALLER than that of the NIMS chip (Fig. SA.1.4b).

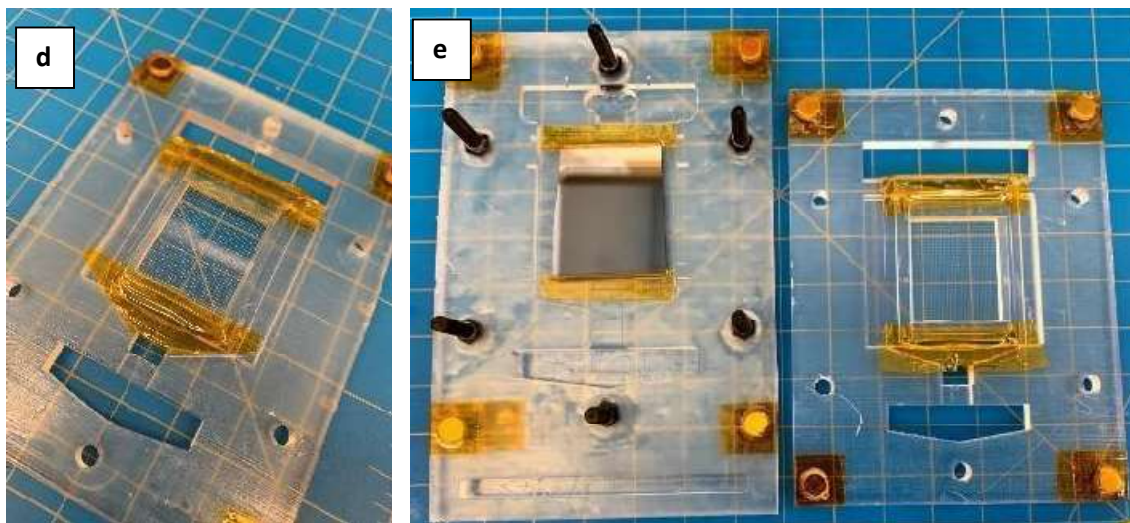


**Figure SA.1.4**

5. Assemble NIMS and microfluidic chip with a chip holder

- Gather 3D print chip holders (top & bottom pieces)
- Tape the microfluidic chip on the chip holder top piece using foam mounting tape (Fig. SA.1.5a-c; McMaster-Carr catalog # 76665A86).
- Secure the chip with [Kapton](#) tape (Fig. SA.1.5d; tape from Digi-Key part # 4393-KPT-1/4-ND). This also helps to prevent droplets from getting into the gap.
- Tape a small NIMS chip (~1" x 1" = 1/4 of a regular NIMS chip) on the chip holder bottom piece using Kapton tape (Fig. SA.1.5e).
- After assembly, should have a top and bottom chip holder with the PDMS chip and the NIMS chip, respectively (Fig. SA.1.5e).





**Figure SA.1.5**

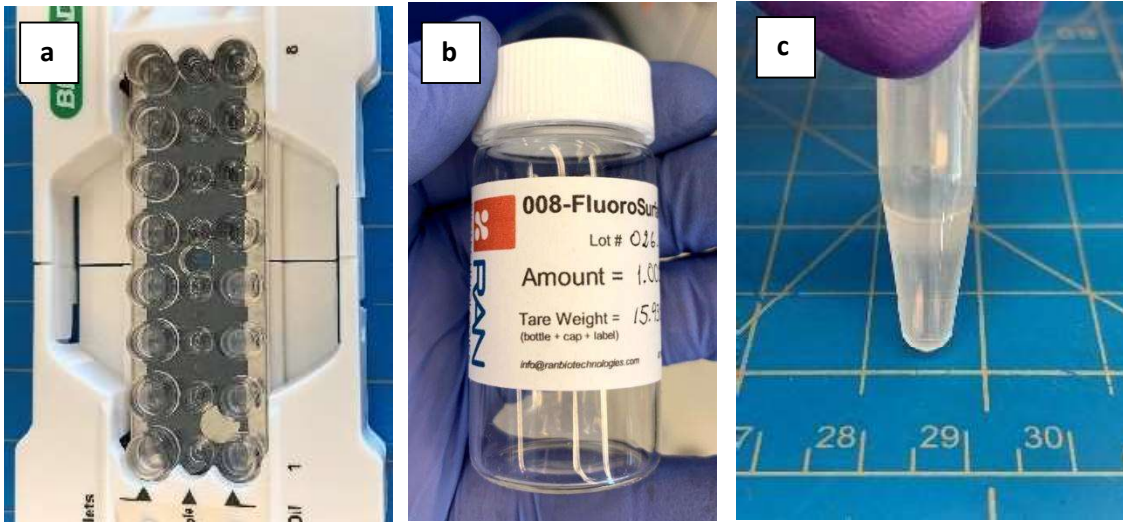
- Assemble two chip holder pieces (Top & Bottom) and adjust the gap using wingnuts (McMaster-Carr catalog # 94924A300). See Fig. A.1Aii. Note: Test the gap by flowing the oil through the gap to ensure sufficient carrier oil is trapped (covering  $\frac{3}{4}$  of the droplet chip surface).

## **Part 2. Running droplet assays:**

See also supplemental videos 1 and 2.

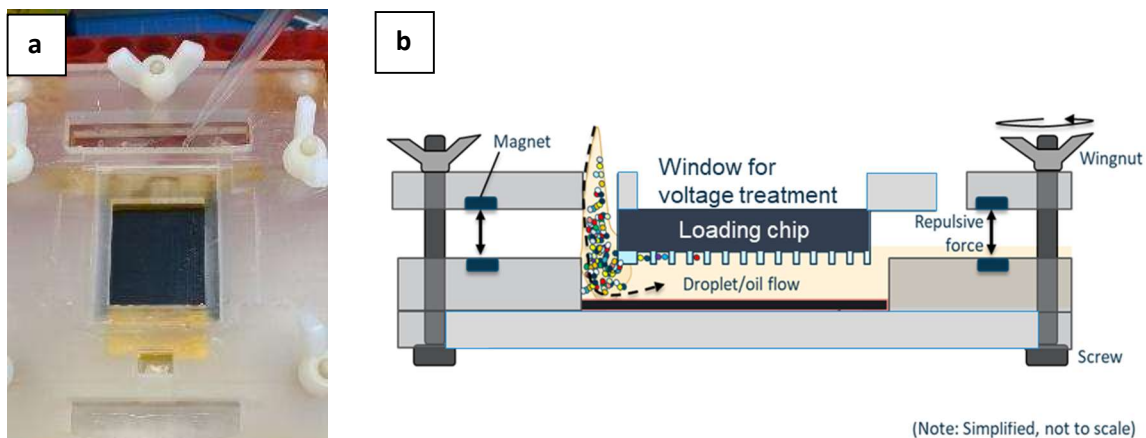
1. Generate droplets using BIO-RAD's QX200 Droplet Generator (catalog # 1864002).
  - Load a new cartridge (Fig. SA.1.6a; BIO-RAD DG8, catalog # 1864008) with sample and carrier oil; 20 $\mu$ l of sample and 70 $\mu$ l of surfactant-added oil (HFE Novec 7500 + 008-Fluorosurfactant from RAN Biotechnologies (Fig. SA.1.6b) in their corresponding wells. It can generate 8 types of droplets each time.
  - After 2 minutes of droplet generation, pool droplets into a new centrifuge tube to prepare droplet cocktail (Fig. SA.1.6c).

- (Optional) visual inspection of generated droplets under the Stereomicroscope:  
Droplet quality check with INCYTO c-chips (hemocytometer chip, item # DHCF012, 20 $\mu$ l chamber).



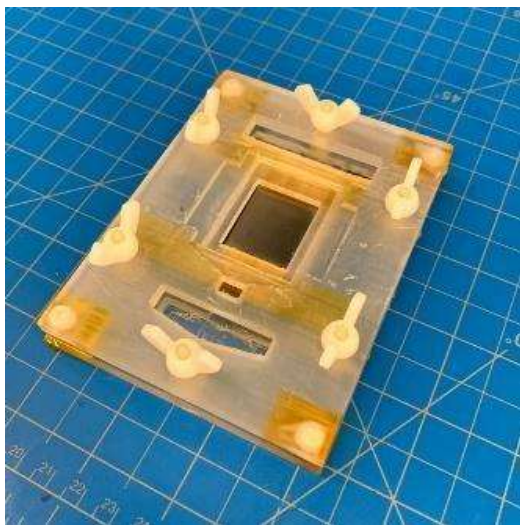
**Figure SA.1.6**

2. Load droplet cocktail (~50 $\mu$ L each) into the microfluidic system using a pipette (use a 200 $\mu$ l pipette tip). Wiggle around the chip holder to rapidly fill microwells with droplets. Pipette extra surfactant-free oil to spread out the droplets if needed.



**Figure SA.1.7**

3. Rinse away unbound droplets with surfactant free oil.
4. Once all microwells are occupied with droplets, seal the droplet chip against the NIMS surface by tightening (hand-tight) the wingnuts. Note: Be careful not to crack the glass.

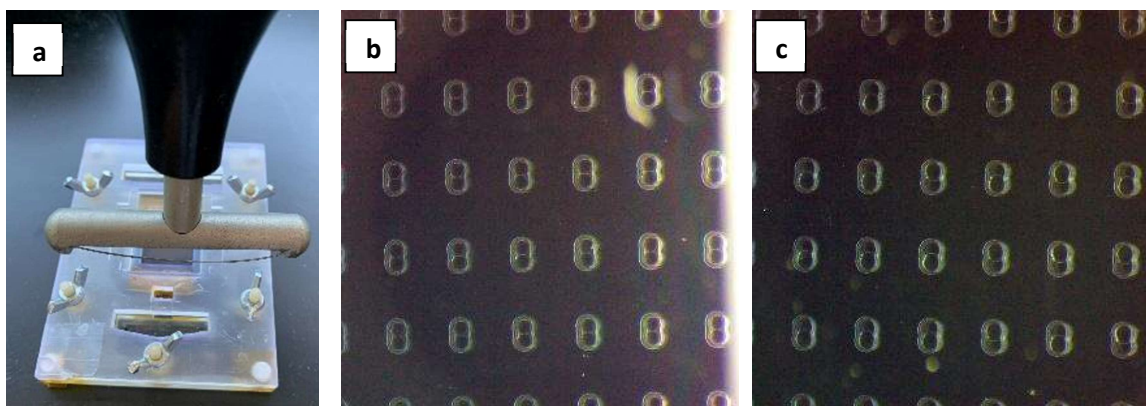


**Figure SA.1.8**

5. Merging droplets using an AC voltage treater (BD-20AC hand-held corona treater from Electro Technic Products, catalog # 12051A):



- Sweeping across the glass (from the top; directly to the glass layer) for 15-20 seconds (Fig. SA.1.9a).
- Confirm complete merging under the stereomicroscope (Fig. SA.1.9b [before] and Fig. SA.1.9c [after]).



**Figure SA.1.9**

6. Droplet enzyme reaction will start immediately upon droplet merging: Incubate the whole system at room temperature (or up to 37°C) overnight. Usually, the reaction will last for approximately two to three hours, then dry out to leave an array of reaction products on the NIMS surface
7. Separate the NIMS chip for MS imaging after the sample & oil have dried out (Fig. SA.1.10a-b).
  - Unscrew wingnuts and carefully separate the chip holder top (containing the PDMS/glass chip) from the chip holder bottom. Remove the NIMS chip from the chip holder bottom.
  - Pipette calibration solution (~0.25uL) onto the NIMS chip in one to two small squares near the edge (e.g., Anaspec peptide 1 solution). Clearly mark the region

of interest with metallic sharpie to quickly locate the region of interest under the MALDI camera (Fig. SA.1.10c).

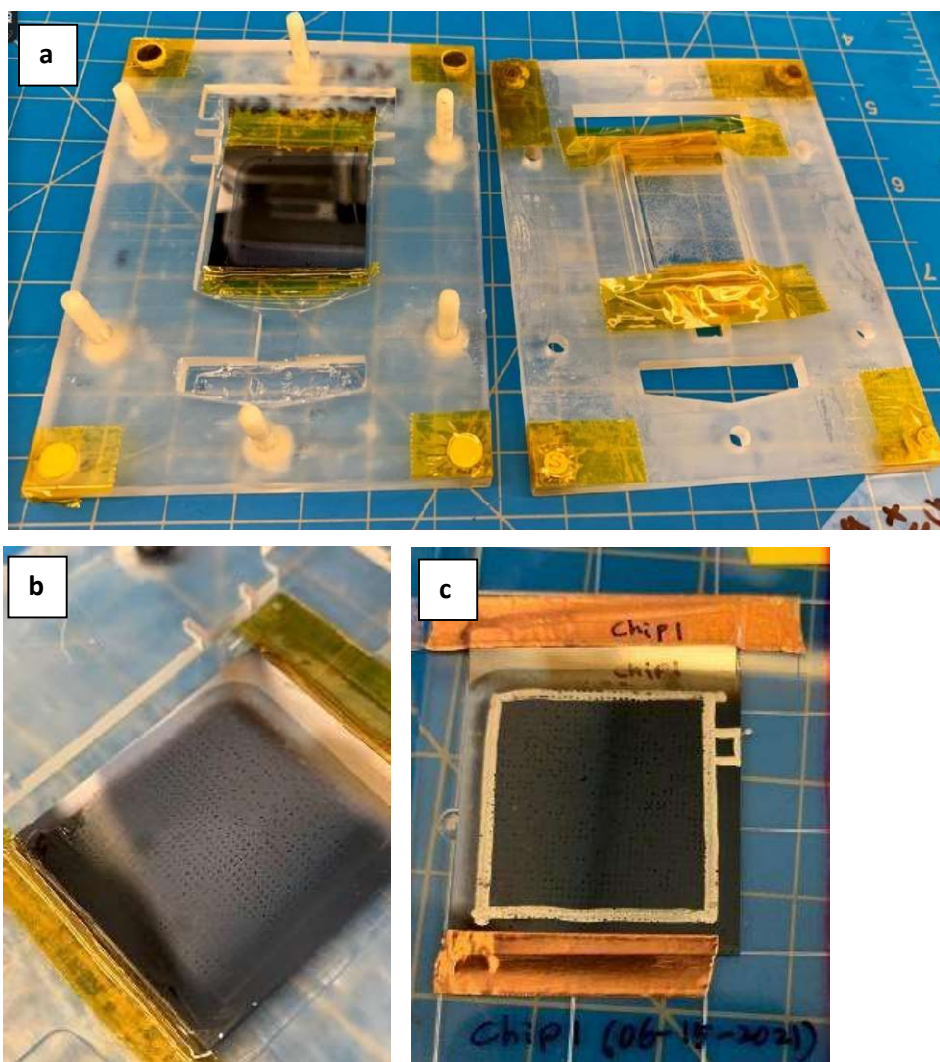
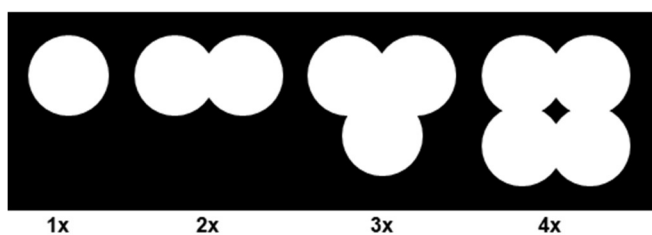
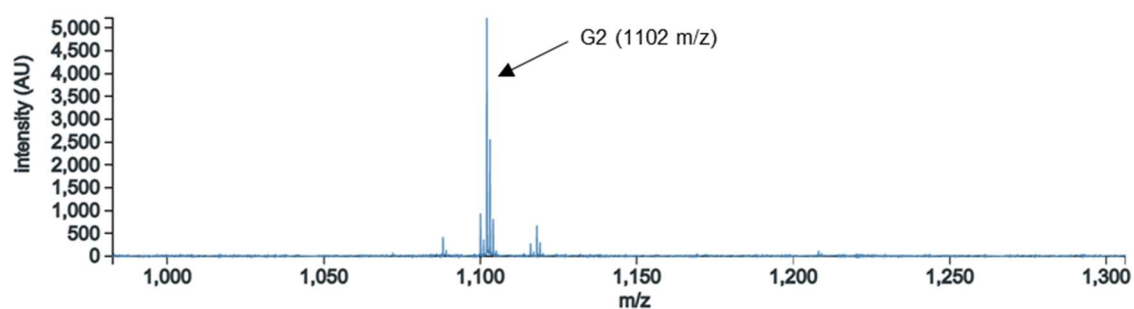


Figure SA.1.10

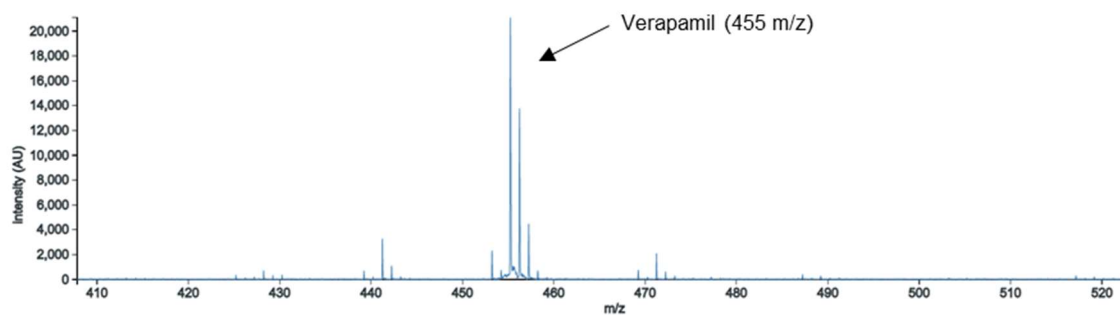
## a. Microwell design



## b



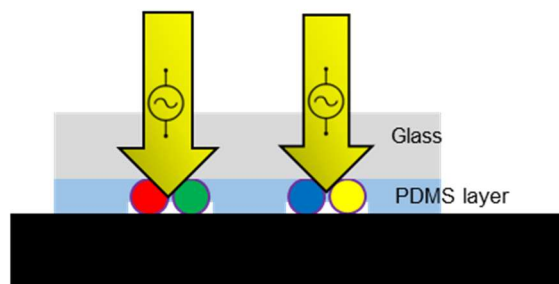
## c



**Figure SA.2:** Droplet Array Printing and MS Imaging. a) Microwell design. Depositing and imaging of substrate droplets b) Verapamil and c) cellobiose with NIMS tag. A prior chip design was used in which microwells were closer together than in final design.

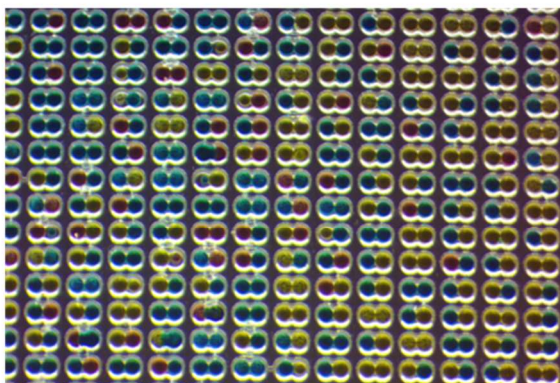
a Droplet Loading

b Electric field applied to merge multiple droplets



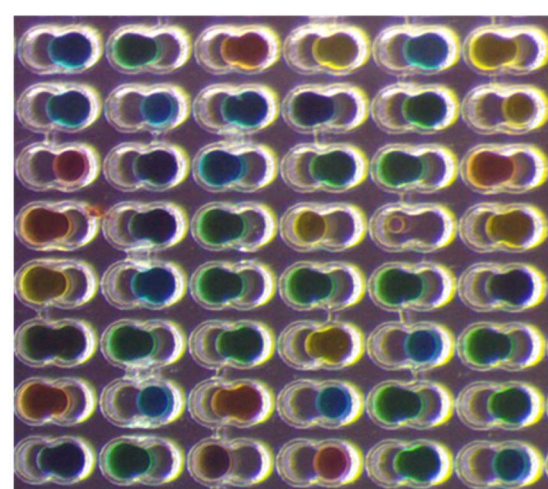
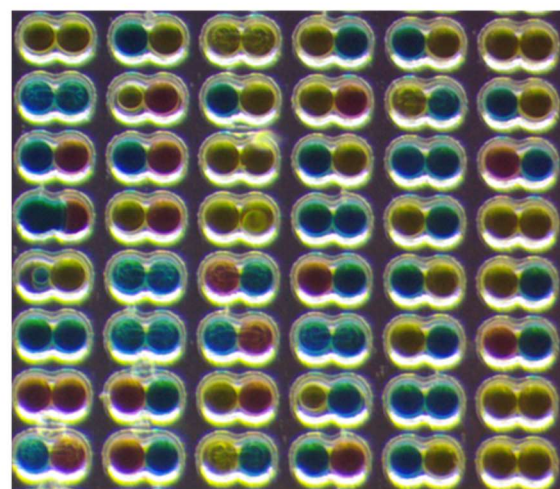
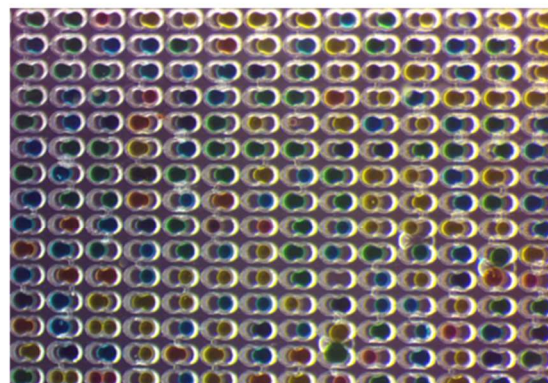
c

Before

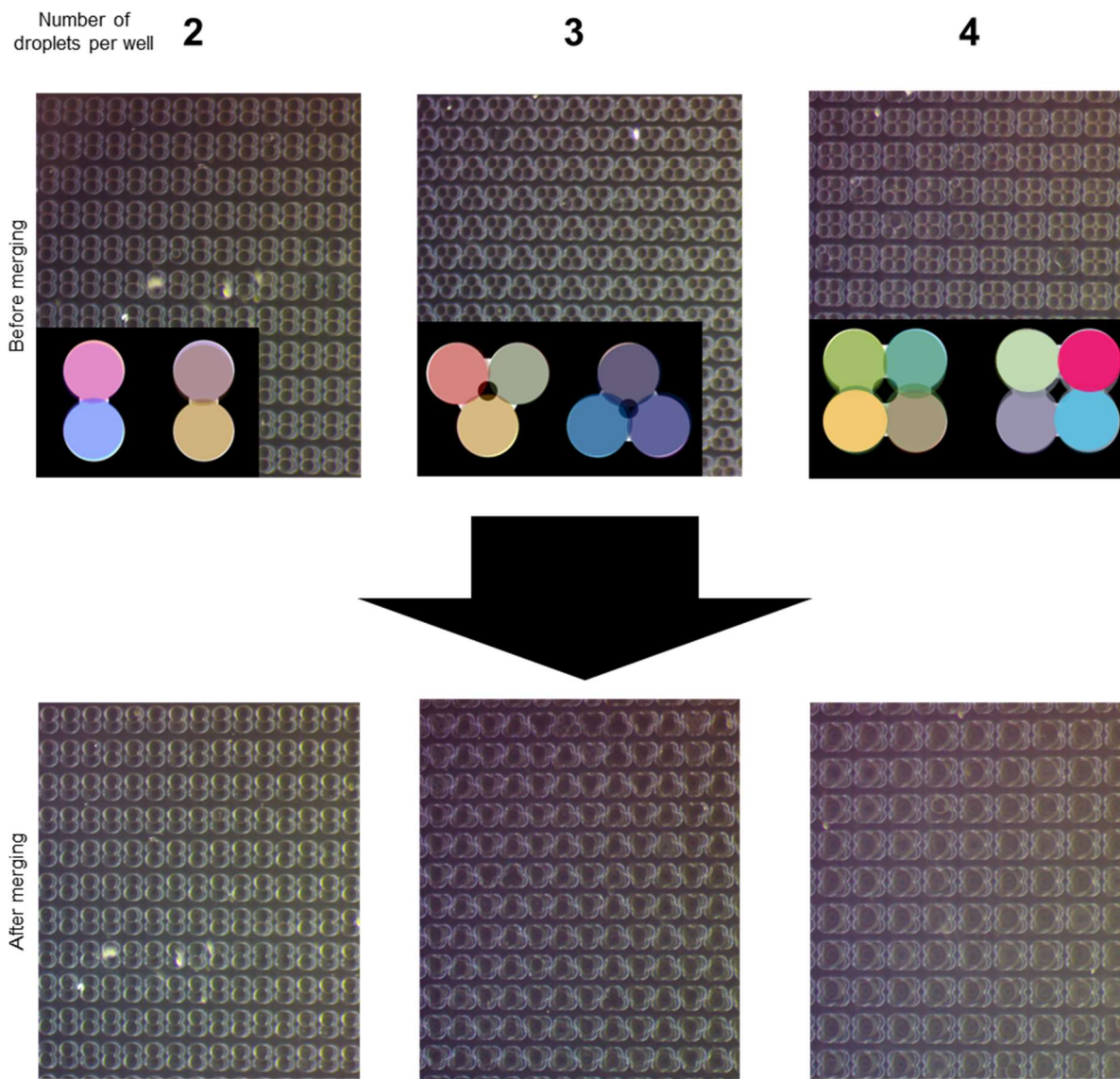


d

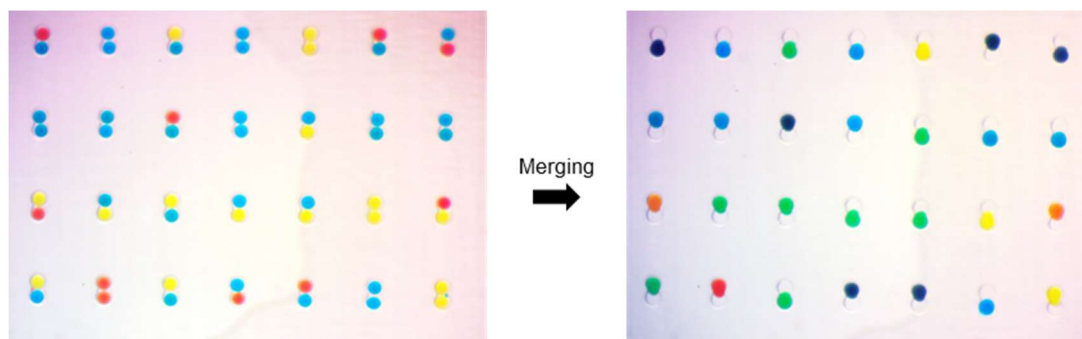
After



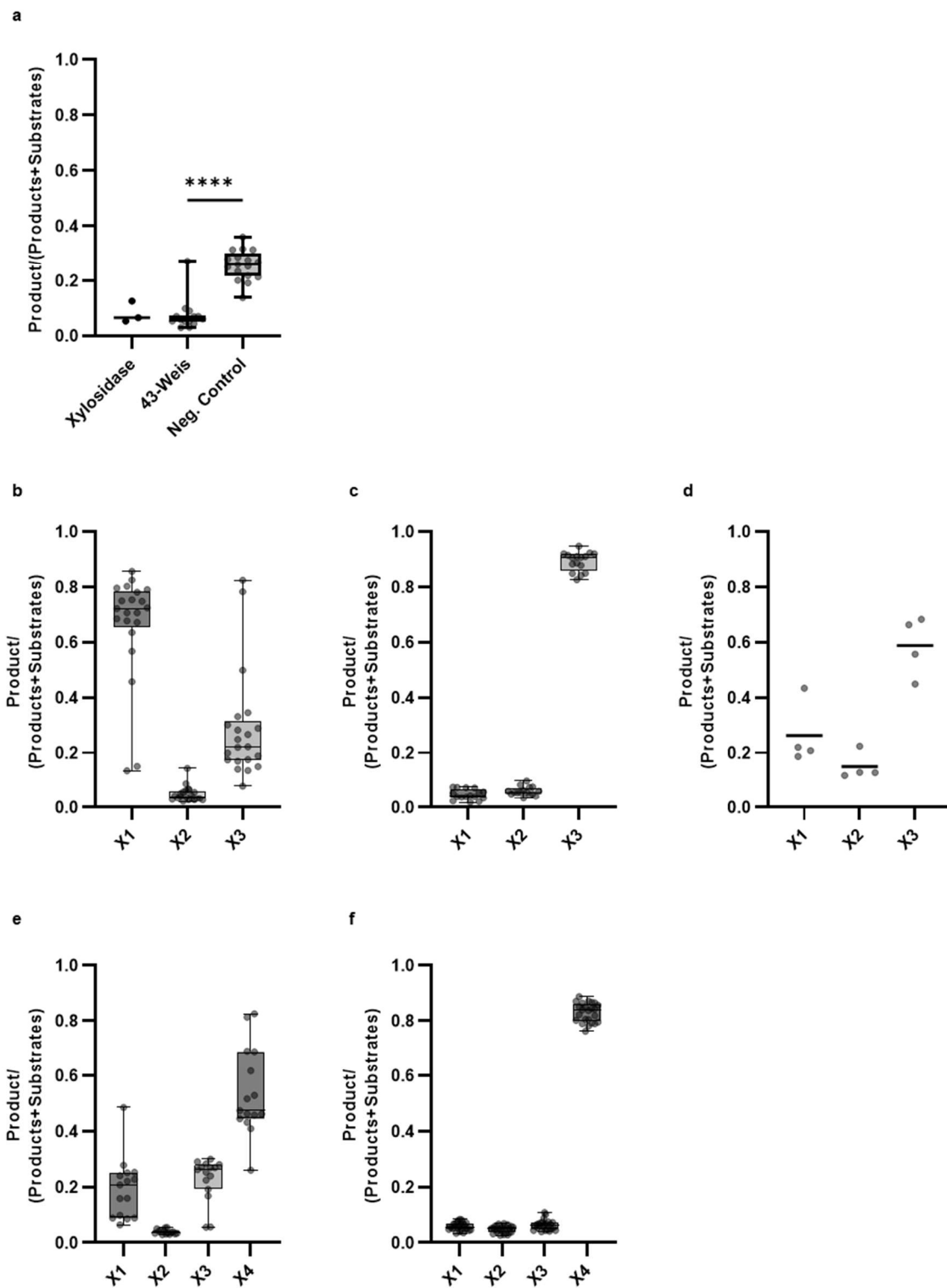
**Figure SA.3:** A demonstration of droplet pairing and merging; before & after merging process. A prior chip design was used in which microwells were closer together than in final design. Colored dyes are shown for visualization. ITO glass was used in place of the NIMS chip. a) Cartoon of loading droplets onto the microfluidics chip. b) An electric field is applied to the glass side of the microfluidics chip. Droplets before (c) and after (d) application of AC voltage. Note: photos are from different fields so wells in before and after pictures do not directly correspond to one another.



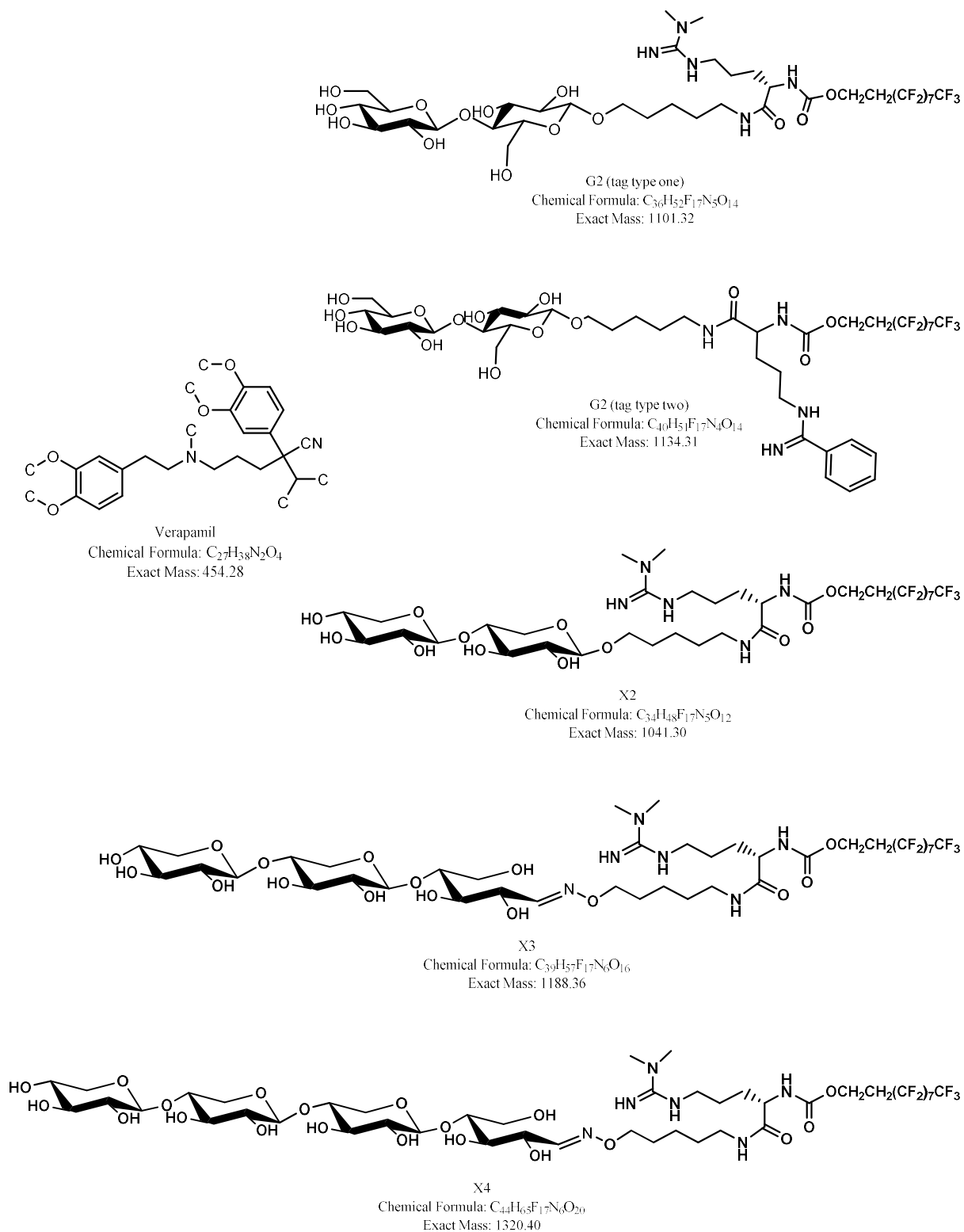
**Figure SA.4:** Alternative versions of the microfluidics chip with multiple sub-wells in each microwell, and subsequent merging. Microwells were closer together than in final design.



**Figure SA.5:** Droplet random pairing and merging on ITO-coated glass. Droplets contained dye-colored buffer.

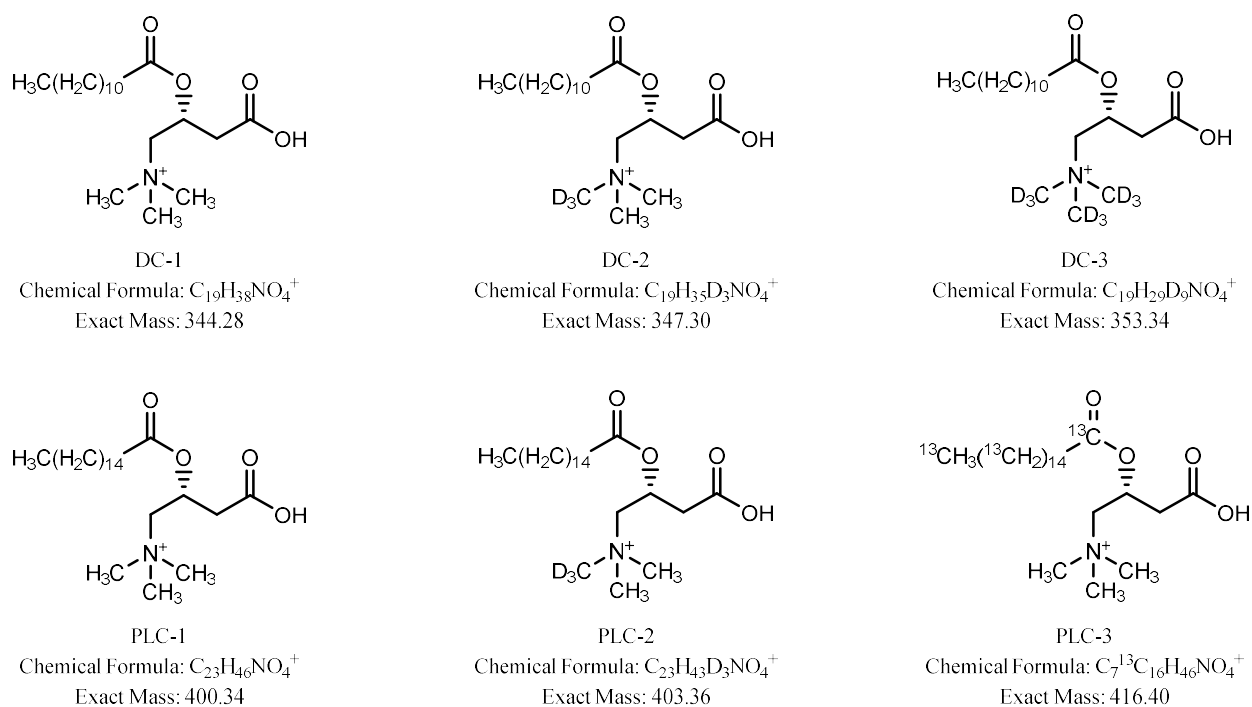


**Figure SA.6:** Specific products of reactions shown in Figure 3c-e. a) G2 reactions, b) X3 reaction products for 43-Weis, c) X3 reaction products for negative control (glucosidase), d) X3 reaction products for commercial xylosidase, e) X4 reaction products for 43-Weis, and f) X4 reaction products for negative control (glucosidase). Boxes show the 25th to 75th percentiles. The line within the box indicates the median. Whiskers indicate maximum and minimum. Statistics only shown for a, in which the negative control is compared to GH43-Weis; \*\*\*\*,  $P < 0.0001$ . Xylosidase samples are shown as scatter plots (horizontal line indicates mean) due to low sample size ( $< 5$ ).

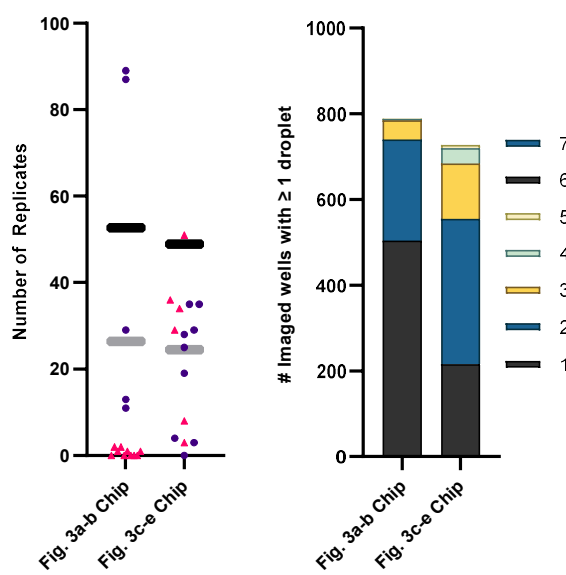


**Figure SA.7:** Verapamil and NIMS-tagged substrates used in this study. G2, cellobiose; X2, xylobiose; X3, xylotriose; X4, xylotetraose.





**Figure SA.8:** Structures of mass spectrometry barcodes used in this study. DC-1, L-Carnitine:HCl O-Dodecanoyl (unlabeled); DC-2, L-Carnitine:HCl, O-Dodecanoyl (N-Methyl-D3); DC-3, L-Carnitine:HCl, O-Dodecanoyl (N,N,N-Trimethyl-D9); PLC-1, L-Carnitine:HCl, O-Palmitoyl (unlabeled); PLC-2, L-Carnitine:HCl, O-Palmitoyl (N-Methyl-D3), and PLC-3, Palmitoyl-13C16-L-carnitine hydrochloride (from Sigma).



**Figure SA.9:** Replicate counts for chip experiments shown in Fig. A.3. a) Numbers of replicates for pairwise combinations. Theoretical number of replicates based on the number of wells imaged, the number of inputs, and either 100% loading efficiency (black bars) or 50% loading efficiency (grey bars). Theoretical numbers do not account for the greater proportion of substrate droplets were used (the number of substrate droplets was equivalent to the number of non-substrate droplets, to ensure most wells received substrate). The number of actual replicate counts for enzyme/substrate or control/substrate combos (purple circles) or non-interesting combinations (e.g., enzyme/enzyme; pink triangles). b) Number of spots with 1 or more droplets.

## References (Formatting not finalized):

- Blommel, Paul G., Katie J. Becker, Petar Duvnjak, and Brian G. Fox. 2007. “Enhanced Bacterial Protein Expression during Auto-Induction Obtained by Alteration of Lac Repressor Dosage and Medium Composition.” *Biotechnology Progress* 23 (3): 585–98. <https://doi.org/10.1021/bp070011x>.
- Deng, K., George, K. W., Reindl, W., Keasling, J. D., Adams, P. D., Lee, T. S., Singh, A. K., & Northen, T. R. (2012). Encoding substrates with mass tags to resolve stereospecific reactions using Nimzyme. *Rapid Communications in Mass Spectrometry*, 26(6), 611–615. <https://doi.org/10.1002/rcm.6134>
- Deng, Kai, Taichi E. Takasuka, Richard Heins, Xiaoliang Cheng, Lai F. Bergeman, Jian Shi, Ryan Aschenbrenner, et al. 2014. “Rapid Kinetic Characterization of Glycosyl Hydrolases Based on Oxime Derivatization and Nanostructure-Initiator Mass Spectrometry (NIMS).” *ACS Chemical Biology* 9 (7): 1470–79. <https://doi.org/10.1021/cb5000289>.
- Deng, Kai, Joel M. Guenther, Jian Gao, Benjamin P. Bowen, Huu Tran, Vimalier Reyes-Ortiz, Xiaoliang Cheng, et al. 2015. “Development of a High Throughput Platform for Screening Glycoside Hydrolases Based on Oxime-NIMS.” *Frontiers in Bioengineering and Biotechnology* 3 (OCT). <https://doi.org/10.3389/fbioe.2015.00153>.
- Deng, K., Zeng, J., Cheng, G., Gao, J., Sale, K. L., Simmons, B. A., Singh, A. K., Adams, P. D., & Northen, T. R. (2018). Rapid characterization of the activities of lignin-modifying enzymes based on nanostructure-initiator mass spectrometry (NIMS). *Biotechnology for Biofuels*, 11(1), 1–9. <https://doi.org/10.1186/s13068-018-1261-2>
- Fox, Brian G., and Paul G. Blommel. 2009. “Autoinduction of Protein Expression.” *Current Protocols in Protein Science* 56 (1): 139–48. <https://doi.org/10.1002/0471140864.ps0523s56>.
- Haidas, D., Bachler, S., Kohler, M., Blank, L. M., Zenobi, R., & Dittrich, P. S. (2019). Micro fluidic Platform for Multimodal Analysis of Enzyme Secretion in Nanoliter Droplet Arrays. *Analytical Chemistry*, 91, 2066–2073. <https://doi.org/10.1021/acs.analchem.8b04506>
- Heinemann, Joshua, Kai Deng, Steve C C Shih, Jian Gao, Paul D Adams, Anup K Singh, and Trent R Northen. 2017. “On-Chip Integration of Droplet Microfluidics and Nanostructure-Initiator Mass Spectrometry for Enzyme Screening.” *Lab on a Chip* 17 (2): 323–31. <https://doi.org/10.1039/C6LC01182A>.
- Hong, Seokho, Myungok Kyung, Inseong Jo, Yong Ro Kim, and Nam Chul Ha. 2018. “Structure-Based Protein Engineering of Bacterial  $\beta$ -Xylosidase to Increase the Production Yield of Xylobiose from Xylose.” *Biochemical and Biophysical Research Communications* 501 (3): 703–10. <https://doi.org/10.1016/j.bbrc.2018.05.051>.
- Kehe, Jared, Anthony Kulesa, Anthony Ortiz, Cheri M. Ackerman, Sri Gowtham Thakku, Daniel Sellers, Sepe Kuehn, Jeff Gore, Jonathan Friedman, and Paul C. Blainey. 2019. “Massively

- Parallel Screening of Synthetic Microbial Communities.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (26): 12804–9. <https://doi.org/10.1073/pnas.1900102116>.
- Klein-Marcuschamer, Daniel, Piotr Oleskowicz-Popiel, Blake A. Simmons, and Harvey W. Blanch. 2012. “The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels.” *Biotechnology and Bioengineering* 109 (4): 1083–87. <https://doi.org/10.1002/bit.24370>.
- Kulesa, Anthony, Jared Kehe, Juan E. Hurtado, Prianca Tawde, and Paul C. Blainey. 2018. “Combinatorial Drug Discovery in Nanoliter Droplets.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (26). <https://doi.org/10.1073/pnas.1802233115>.
- Lapierre, F., Piret, G., Drobecq, H., Melnyk, O., Coffinier, Y., Thomy, V., & Boukherroub, R. (2011). High sensitive matrix-free mass spectrometry analysis of peptides using silicon nanowires-based digital microfluidic device. *Lab on a Chip*, 11(9), 1620–1628. <https://doi.org/10.1039/c0lc00716a>
- McKinney, Wes. 2010. “Data Structures for Statistical Computing in Python.” *Proceedings of the 9th Python in Science Conference* 1 (Scipy): 56–61. <https://doi.org/10.25080/majora-92bf1922-00a>.
- Miller, G. L. (1959). Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, 31(3), 426–428. <https://doi.org/10.1021/ac60147a030>
- Northen, Trent R., Oscar Yanes, Michael T. Northen, Dena Marrinucci, Winnie Uritboonthai, Junefredo Apon, Stephen L. Gollidge, Anders Nordström, and Gary Siuzdak. 2007. “Clathrate Nanostructures for Mass Spectrometry.” *Nature* 449 (7165): 1033–36. <https://doi.org/10.1038/nature06195>.
- The pandas development team. 2021. “Pandas-Dev/Pandas: Pandas 1.2.4.” Zenodo. <https://doi.org/10.5281/ZENODO.4681666>.
- Pollet, Annick, Jan Schoepe, Emmie Dornez, Sergei V. Strelkov, Jan A. Delcour, and Christophe M. Courtin. 2010. “Functional Analysis of Glycoside Hydrolase Family 8 Xylanases Shows Narrow but Distinct Substrate Specificities and Biotechnological Potential.” *Applied Microbiology and Biotechnology* 87 (6): 2125–35. <https://doi.org/10.1007/s00253-010-2659-3>.
- de Raad, Markus, Tristan De Rond, Oliver Rübél, Jay D. Keasling, Trent R. Northen, and Benjamin P. Bowen. 2017. “OpenMSI Arrayed Analysis Toolkit: Analyzing Spatially Defined Samples Using Mass Spectrometry Imaging.” *Analytical Chemistry* 89 (11): 5818–23. <https://doi.org/10.1021/acs.analchem.6b05004>.
- Sawasaki, T., Ogasawara, T., Morishita, R., & Endo, Y. (2002). A cell-free protein synthesis system for high-throughput proteomics. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23), 14652–14657. <https://doi.org/10.1073/pnas.232580399>
- Studier, F. William. 2005. “Protein Production by Auto-Induction in High Density Shaking Cultures.” *Protein Expression and Purification* 41 (1): 207–34. <https://doi.org/10.1016/j.pep.2005.01.016>.
- Takasuka, Taichi E, Johnnie A Walker, Lai F Bergeman, Kirk A Vander Meulen, Shin-ichi Makino, Nathaniel L Elsen, and Brian G Fox. 2014. “Cell-Free Translation of Biofuel Enzymes.” In *Cell-Free Protein Synthesis: Methods and Protocols*, edited by Kirill Alexandrov and Wayne A Johnston, 71–95. Totowa, NJ: Humana Press. [https://doi.org/10.1007/978-1-62703-782-2\\_5](https://doi.org/10.1007/978-1-62703-782-2_5).
- van Tilbeurgh, H., Loontjens, F. G., De Bruyne, C. K., & Claeysens, M. (1988). Fluorogenic and Chromogenic Glycosides as Substrates and Ligands of Carbohydrases. In *Methods in Enzymology* (Vol. 160, pp. 45–59). Academic Press. [https://doi.org/10.1016/0076-6879\(88\)60106-6](https://doi.org/10.1016/0076-6879(88)60106-6)

Vallat, Raphael. 2018. "Pingouin: Statistics in Python." *Journal of Open Source Software* 3 (31): 1026. <https://doi.org/10.21105/joss.01026>.

Walker, Johnnie A., Taichi E. Takasuka, Kai Deng, Christopher M. Bianchetti, Hannah S. Udell, Ben M. Prom, Hyunkee Kim, Paul D. Adams, Trent R. Northen, and Brian G. Fox. 2015.

"Multifunctional Cellulase Catalysis Targeted by Fusion to Different Carbohydrate-Binding Modules." *Biotechnology for Biofuels* 8 (1): 1–20. <https://doi.org/10.1186/s13068-015-0402-0>.

Woo, Hin Koon, Trent R. Northen, Oscar Yanes, and Gary Siuzdak. 2008. "Nanostructure-Initiator Mass Spectrometry: A Protocol for Preparing and Applying NIMS Surfaces for High-Sensitivity Mass Analysis." *Nature Protocols* 3 (8): 1341–49.

<https://doi.org/10.1038/nprot.2008.110>.

Xu, L., Chang, K., Payne, E. M., Modavi, C., Liu, L., Palmer, C. M., Tao, N., Alper, H. S., Kennedy, R. T., Cornett, D. S., & Abate, A. R. (2021). Mapping enzyme catalysis with metabolic biosensing. *Nature Communications*, 12(1), 6803. <https://doi.org/10.1038/s41467-021-27185-9>