Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases

**Kai Deng[1,2], Taichi E. Takasuka[3†], Christopher M. Bianchetti[3,4], Lai F. Bergeman[3],
Paul D. Adams[1,5,6], Trent R. Northen[1,5] and Brian G. Fox[3,7]***

[1] US Department of Energy Joint BioEnergy Institute, Emeryville, CA, USA, [2] Sandia National Laboratories, Livermore, CA, USA, [3] US Department of Energy Great Lakes Bioenergy Research Center, Madison, WI, USA, [4] Department of Chemistry, University of Wisconsin-Oshkosh, Oshkosh, WI, USA, [5] Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [6] Department of Bioengineering, University of California Berkeley, Berkeley, CA, USA, [7] Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

Chemically synthesized nanostructure-initiator mass spectrometry (NIMS) probes derivatized with tetrasaccharides were used to study the reactivity of representative *Clostridium thermocellum* β-glucosidase, endoglucanases, and cellobiohydrolase. Diagnostic patterns for reactions of these different classes of enzymes were observed. Results show sequential removal of glucose by the β-glucosidase and a progressive increase in specificity of reaction from endoglucanases to cellobiohydrolase. Time-dependent reactions of these polysaccharide-selective enzymes were modeled by numerical integration, which provides a quantitative basis to make functional distinctions among a continuum of naturally evolved catalytic properties. Consequently, our method, which combines automated protein translation with high-sensitivity and time-dependent detection of multiple products, provides a new approach to annotate glycoside hydrolase phylogenetic trees with functional measurements.

Keywords: cellulase, assay, kinetics, Nimzyme, mass spectrometry, protein engineering, biofuels

INTRODUCTION

The enzymatic hydrolysis of plant cell wall material is a formidable task because of the complexity of the plant cell wall (Himmel et al., 2007). In most currently deployed cellulosic ethanol plants, enzyme cocktails containing multiple classes of polysaccharide-degrading enzymes are used to hydrolyze plant biomass into fermentable sugars. Understanding the function, synergy, and stability of enzymes is thus of paramount importance in biofuels production.

Polysaccharide-degrading enzymes are classified into families in the carbohydrate active enzyme (CAZy) database (Henrissat and Davies, 1997; Cantarel et al., 2009; Levasseur et al., 2013), including glycoside hydrolases (GHs), pectic lyases (PLs), carbohydrate esterases (CEs), and others. Only a small fraction of the enzymes included in CAZy have a function assigned by biochemical analyses. One root of this limitation arises from difficulties in succeeding with heterologous expression of enzymes after selection from phylogenetic trees (Watson et al., 2007; Fox et al., 2008; Markley

Abbreviations: AFEX-SG, ammonia fiber expansion pretreated switchgrass; CBM, carbohydrate-binding module; CelE, broad specificity GH family 5 (GH5) domain from *C. thermocellum* Cthe_0797; GH, glycoside hydrolase; IL-SG, ionic liquid pretreated switchgrass; NIMS, nanostructure-initiator mass spectrometry.

et al., 2009; Nair et al., 2009; Pieper et al., 2013). As an option to address this limitation, we (Takasuka et al., 2014; Bianchetti et al., 2015) and others (Beebe et al., 2011, 2014; Madono et al., 2011; Hirano et al., 2013, 2015; Makino et al., 2014) have found that wheat germ cell-free protein translation can be used as an effective expression platform to make functional assignments of enzyme function.

Another limitation arises from experimental complications of carrying out high-throughput multisubstrate assays to screen for enzyme function (Gerlt et al., 2011). A breadth of assay methods have been developed for GHs, including use of soluble and insoluble chromogenic and/or fluorogenic substrates, HPLC, and others (Sharrock, 1988; Decker et al., 2003; Chundawat et al., 2008; Bansal et al., 2009; Dowe, 2009; Dashtban et al., 2010; Selig et al., 2011; Eklof et al., 2012; Horn et al., 2012; Kosik et al., 2012; McCleary et al., 2012; Pena et al., 2012; Whitehead et al., 2012; Wischmann et al., 2012). Each of these approaches has intrinsic advantages, but can suffer in sensitivity, complexity of analysis, throughput time, and volumes of reagents and enzyme needed. In comparison, nanostructure-initiator mass spectrometry (NIMS) offers high sensitivity, simplicity of detection of products derived from biomass hydrolysis, microliters or smaller volumes for reaction, and options for automation (Northen et al., 2008; Deng et al., 2012; de Rond et al., 2013; Heins et al., 2014). Recently, we used oxime-NIMS and numerical integration methods to provide time-dependent, quantitative characterization of reducing sugars released by individual enzymes in reactions with pretreated biomass (Deng et al., 2014).

Here, we report a new use of NIMS to provide quantitative analysis of time-dependent reactions of cellulases. The enzymes selected for this study were from *Clostridium thermocellum*, a Gram-positive anaerobe with high cellulolytic capacity (Ding et al., 2008; Fontes and Gilbert, 2010; Smith and Bayer, 2013). The *C. thermocellum* genome encodes ~130 CAZyme domains and ~90 carbohydrate-binding module (CBM) domains (Feinberg et al., 2011). The majority of CAZyme domains also possess dockerin domains, which serve to recruit these enzymes into the cellulosome via dockerin–cohesin interactions (Ding et al., 2008; Smith and Bayer, 2013). The specific gene regulatory and protein secretory patterns of this model consolidated bioprocessing organism have also been well described (Brown et al., 2007; Gold and Martin, 2007; Roberts et al., 2010; Feinberg et al., 2011; Raman et al., 2011; Riederer et al., 2011), and many of the enzymes have been characterized. Given this state of knowledge, individual enzymes from *C. thermocellum* have proven useful for the development and testing of new approaches for assignment of GH function.

In this work, we have used chemically synthesized tetrasaccharide-NIMS probes to study the reactivity of some cellulases from *C. thermocellum*. Patterns of reactivity identified by using the tetrasaccharide-NIMS probes provide a diagnostic approach to assess reaction specificity and also provide comparative apparent rate information. Our results show diagnostic patterns for reactions of a β-glucosidase, relaxed but varied specificity of several endoglucanases and high specificity of a cellobiohydrolase with the model substrate. Time-dependent reactions of these polysaccharide-selective enzymes were modeled by numerical integration, which provides a quantitative basis to make functional distinctions among a continuum of naturally evolved reactive properties. Consequently, this method, which combines high-sensitivity detection of multiple products with quantitative numerical analysis of their time-dependent formation, provides a new approach to enhance the annotation of GH phylogenetic trees with functional measurements.

MATERIALS AND METHODS

Enzyme Preparation

Methods for cloning, cell-free translation, and purification of the enzymes studied have been reported elsewhere (Takasuka et al., 2014). Briefly, enzymes were cloned by PCR amplification of catalytic domains as indicated by the first and last codons indicated in **Table 1**. Cloned genes were transferred into an optimized wheat germ cell-free translation plasmid pEU-HSCB (Beebe et al., 2011; Takasuka et al., 2014), which is also available from the NIH Protein Structure Initiative Materials Repository (http://psimr.asu.edu/). Enzymes were prepared by cell-free translation using either bilayer or dialysis methods (Beebe et al., 2011, 2014; Makino et al., 2014), and active enzymes were identified (Takasuka et al., 2014). The enzymes listed in **Table 1** were also cloned by PCR into the *Escherichia coli* expression vector pEC_CBM3a to create enzyme_CBM3a fusion proteins, CelAcc_CBM3a. The vector pEC_CBM3a is a derivative of pEU_HSBC_CBM3a (Takasuka et al., 2014) that yields fusion proteins having an N-terminal enzyme catalytic domain fused by an ~40 aa linker sequence to the CBM3a domain from Cthe_3077. A stop codon was added to the PCR primer used to amplify the 3′ end of the BglA gene so that no fusion to CBM3a was produced from pEU_HSBC_CBM3a. As needed, protein coding sequences were transferred between pEU and pEC vectors by use of FlexiVector cloning (Blommel et al., 2009). Methods for PCR amplification, capture and sequence verification of protein coding sequences, and transformation into *E. coli* 10G competent cells (Lucigen, Middleton, WI, USA) for DNA manipulations and *E. coli* B834 for protein expression were as previously reported (Takasuka et al., 2014). Additional details of the properties and methods for the use of pEU and pVP are described elsewhere (Aceti et al., 2015).

Synthesis of Cellotetraose-NIMS Substrate

The cellotetraose-NIMS substrate (**Figure 1A**) is an amphiphilic molecule that has a sugar head group coupled to a perfluorinated (F17) tag. The detailed synthetic procedure has been reported previously (Deng et al., 2012).

Enzyme Reactions

An enzyme reaction consisted of 10 µL of 50 mM phosphate, pH 6.0, supplemented with 1 µL of 1 mM cellotetraose-F17 dissolved in water. An aliquot of each enzyme preparation (containing 1–10 ng of enzyme) was added to initiate the reaction and the resulting mixture was incubated at 37°C. At times of 5, 10, 20, 40, 80, and 120 min, 0.2 µL of the reaction mixture was withdrawn for analysis.

TABLE 1 | *Clostridium thermocellum* enzymes studied in this work.

Gene locus	Name	CAZy family	First codon[a]	Last codon[b]	Function[c]	CMX class[d]	Reference
Cthe_0212	BglA	GH1	1	448	Exo-β-glucosidase		Grabnitz et al. (1991)
Cthe_0269	CelA	GH8	34	368	Endo-β-1,4-glucanase	CX	1CEM; Alzari et al. (1996)
Cthe_0040	CelI	GH9	28	887	Endo-β-1,4-glucanase	C	2XFG
Cthe_0797	CelE	GH5	36	388	Cellulase, xylanase, mannanase	CMX	Deng et al. (2014, Takasuka)
Cthe_0578	CelR	GH9	27	640	Endo-β-1,4-glucanase; cellotetraohydrolase	C	Zverlov et al. (2005)
Cthe_0405	CelL	GH5	31	430	–	CX	Deng et al. (2014)
Cthe_0412	CelK	GH9	28	809	Cellobiohydrolase	C	Kataeva et al. (1999)
Cthe_3077[e]	CipA	CBM3a	323	523	Cellulose-binding module		Kataeva et al. (1999)

[a]First codon of the indicated gene locus that was included in the PCR primer design (Takasuka et al., 2014).
[b]Last codon of the indicated gene locus that was included in the PCR primer design.
[c]Function assigned from annotation as defined in CAZy (Cantarel et al., 2009), from experimental evidence cited in the table, or a combination of both.
[d]Representation of the breadth of substrate specificity for each enzyme (Deng et al., 2014). The CMX classification indicates that CelE can hydrolyze cellulose, xylan, or mannan; CX indicates that CelA and CelL can hydrolyze cellulose and xylan, while CelI, CelR, and CelK can only hydrolyze cellulose. This classification derives from reactions with pure polysaccharides and pretreated biomass (Deng et al., 2014; Takasuka et al., 2014).
[e]CBM3a was subcloned from the scaffoldin CipA gene.

Mass Spectrometry

In each case, 0.2 µL per reaction sample was spotted onto the NIMS surface and removed after an incubation of ~30 s. A grid drawn manually on the NIMS chip using a diamond-tip scribe helped with spotting and identification of sample spots in the spectrometer. Chips were loaded using a modified standard MALDI plate. NIMS was performed on a 4800 MALDI TOF/TOF mass analyzer from AB Sciex (Foster City, CA, USA). In each case, signal intensities were identified for the ions of the cellotetraose substrate and, when present, each product shown in **Figure 1**. For each assay, ~1000 laser shots were collected. Enzyme activities were determined by measuring the intensity ratios of each product over the intensity total of ions of for the cellotetraose-, cellotriose-, cellobiose-, glucose-, and aglycone-NIMS (**Figure 2**).

Kinetic Analyses

The time dependence of hydrolysis of the tetrasaccharide-NIMS was analyzed by non-linear global optimization of differential equations accounting for the appearance and decay of products (Deng et al., 2014) using Mathematica routine NDSolve and the Nelder-Mead simplex method for constrained minimization (Nelder and Mead, 1965). The differential equations corresponding to the kinetic scheme of **Figure 3** are as follows:

$$y[1] = \text{cellotetraose-NIMS} \tag{1}$$

$$y[2] = \text{cellotriose-NIMS} \tag{2}$$

$$y[3] = \text{cellobiose-NIMS} \tag{3}$$

$$y[4] = \text{glucose-NIMS} \tag{4}$$

$$y[5] = \text{aglycone-NIMS} \tag{5}$$

$$dy[1]/d[t] = -(k1 + k9 + k11 + k13)\, y[1][t] \tag{6}$$

$$dy[2]/d[t] = (k1)\, y[1][t] - (k3 + k15 + k17)\, y[2][t] \tag{7}$$

$$dy[3]/d[t] = (k9)\, y[1][t] + (k3)\, y[2][t] - (k5 + k19)\, y[3][t] \tag{8}$$

$$dy[4]/d[t] = (k11)\, y[1][t] + (k15)\, y[2][t] + (k5)\, y[3][t] - (k7)\, y[4][t] \tag{9}$$

$$dy[5]/d[t] = (k13)\, y[1][t] + (k17)\, y[2][t] + (k19)\, y[3][t] + (k7)\, y[4][t] \tag{10}$$

Initial guesses for apparent rate constants were made by visual inspection of the match between the results of single NDSolve calculations and the experimental data. This process was continued in an iterative way until a set of initial apparent rates that adequately matched the experimental data was obtained. Successive rounds of least squares parameter optimization with adjustment of parameter constraints were carried out until the sum of the squares difference between calculated and experimental values reached a minimum and no parameter was artificially constrained.

RESULTS AND DISCUSSION

Enzymes Chosen for Study

Clostridium thermocellum enzymes were chosen for this study based on previous transcriptomic and proteomic results (Gold and Martin, 2007; Raman et al., 2011; Riederer et al., 2011) and other biochemical and structural results (**Table 1**). Genes encoding these enzymes were expressed using wheat germ cell-free protein synthesis and the translated proteins were assayed using fluorogenic substrates (Takasuka et al., 2014); among the synthesized enzymes, 13 reacted with MUG or MUC, 11 reacted with MUX or MUX2, and 5 reacted with other diagnostic fluorogenic substrates. Reactions of these enzymes with ionic liquid pretreated switchgrass (IL-SG) have been published (Deng et al., 2014). Enzymes from cell-free translation reactions that showed promising characteristics were produced by expression in *E. coli* and purified for use in the studies described here.
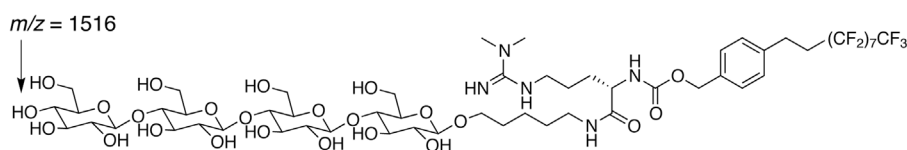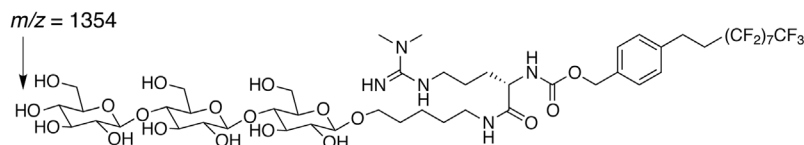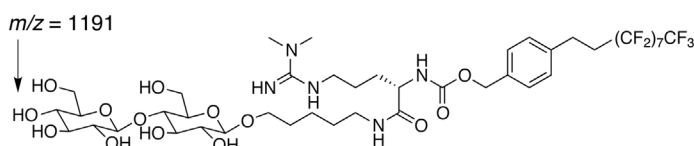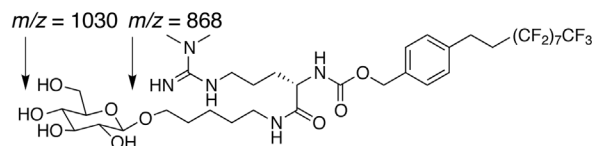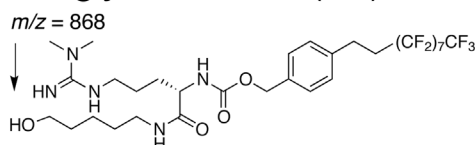
FIGURE 1 | Structure of cellotetraose-NIMS and *m/z* values for products obtained from hydrolysis at the indicated anomeric position. **(A)** cellotetraose-NIMS; **(B)** cellotriose-NIMS; **(C)** cellobiose-NIMS; **(D)** glucose-NIMS; **(E)** aglycone-NIMS.
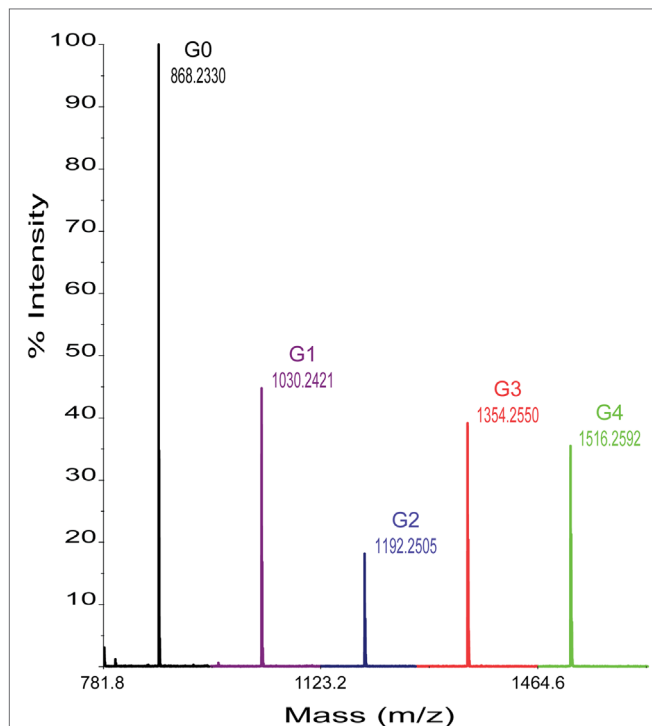
Cellotetraose-NIMS Substrate

Figure 1 shows the structure of cellotetraose-NIMS and the products that can be formed by various GH reactions. In the synthesized probe, the tetra-saccharide is linked to the NIMS probe by a potentially hydrolyzable anomeric linkage. Synthesis of the NIMS probe and the tetra-saccharide derivatives are summarized in Materials and Methods (Deng et al., 2012; de Rond et al., 2013). The guanidium group on the NIMS probe provides improved ionization properties in the mass spectrometry experiment, while the perfluorinated portion of the NIMS probe provides hydrophobic anchoring of the molecule into the NIMS surface. Enzyme-catalyzed hydrolysis of the anomeric linkages give rise to a cascade of potential products retained on the NIMS surface. Reactions of GHs can progressively remove single glucose units or carry out other reactions that remove cellobiose, cellotriose, or cellotetraose.

Kinetic Scheme

Figure 2 shows a representative mass spectrum obtained after partial reaction with BglA (Cthe_0212), a β-glucosidase. At the selected time point in the reaction (120 min), the cellotetraose-NIMS probe (G4, green) has been partially converted into a mixture of cellotriose (G3, red), cellobiose (G2, blue), glucose (G1, purple), and aglycone (G0, black) derivatives of the NIMS probe. **Figure 3** shows a kinetic scheme that accounts for the potential products shown in **Figure 1**. The scheme accounts for release of one or more glucose units from the cellotetraose-NIMS probe (G4) and its successive products. Time course profiles provide the fundamental data used in this work for numerical analysis of enzyme hydrolysis reactions.

β-Glucosidase BglA Reaction

The nucleotide sequence of BglA (Grabnitz et al., 1991) was published before the genome sequence and annotated to be

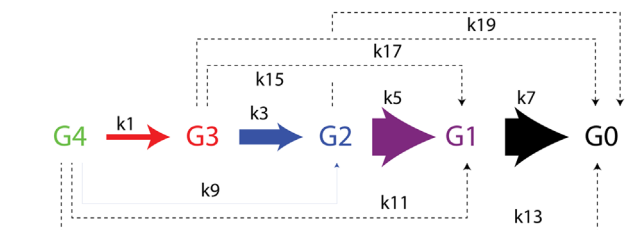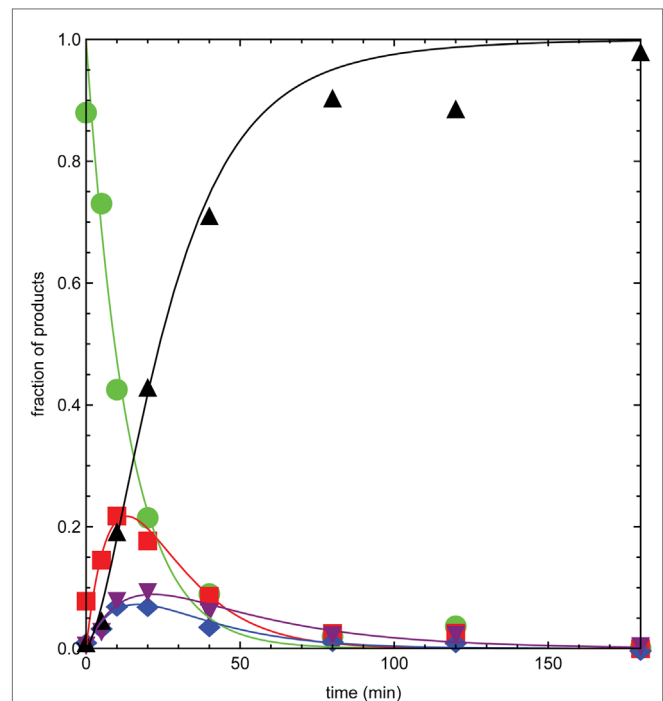**FIGURE 2 | Representative mass spectrum obtained from enzyme hydrolysis of cellotetraose-NIMS.** Mass peaks corresponding to cellotetraose-NIMS (green), cellotriose-NIMS (red), cellobiose-NIMS (blue), glucose-NIMS (purple), and aglycone-NIMS (black) are indicated. The products shown are from reaction of BglA.
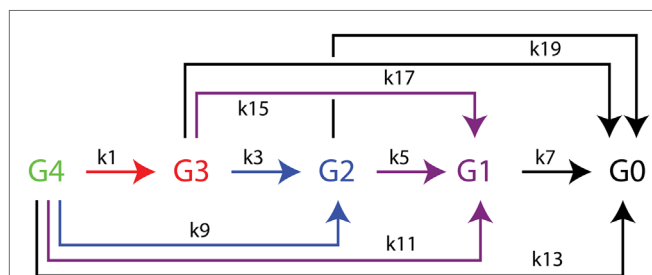
FIGURE 3 | Kinetic scheme for the enzymatic hydrolysis of cellotetraose-NIMS accounting for all products detected. Apparent rate constants determined from numerical simulations of time dependence of enzyme reactions using differential equations 1–10 from the section "Materials and Methods" are found in **Table 2**. Cellotetraose-NIMS, green; cellotriose-NIMS, red; cellobiose-NIMS, blue; glucose, purple; aglycone-NIMS, black.

FIGURE 4 | Numerical analysis of the time course for reaction of BglA with cellotetraose-NIMS. Products are cellotetraose-NIMS (green), cellotriose-NIMS (red), cellobiose-NIMS (blue), glucose- NIMS (purple), and aglycone-NIMS (black). Relative magnitude of the apparent rates shown in **Table 2** indicated by width of arrows in the modified kinetic scheme. A dashed line indicates that the apparent rate was zero.

a β-glucosidase from the GH1 family (Cantarel et al., 2009). The Cthe_02012 gene does not encode a signal peptide, so the entire gene was cloned for the studies described here. Beyond our characterization of the reaction of BglA with IL-SG (Deng et al., 2014), no other functional studies have been reported for this enzyme.

Figure 4 shows the time course for reaction of BglA with cellotetraose-NIMS. The plotted proportions of the different products come from time series of mass spectra like those shown in **Figure 2**. The solid colored lines are results of simulations of the concentration of individual products based on the kinetic scheme of **Figure 3** and the differential equations shown in the section "Materials and Methods." The apparent rate constants provided by the numerical simulation are given in **Table 2**, and a pictorial representation of the relative magnitudes of the apparent rate constants is also given in **Figure 4**. In the time course of the BglA reaction, cellotetraose-NIMS (green circles) was converted to a succession of intermediates by hydrolysis of a single glucose from the position most distal to the NIMS probe. This pattern of reactivity is as expected for the reaction of an exo-β-glucosidase with an oligosaccharide. Thus, cellotriose-NIMS (red squares) accumulated was subsequently converted to cellobiose-NIMS (purple down triangles), to glucose-NIMS (blue diamonds), and ultimately to aglycone-NIMS (black up triangles).

TABLE 2 | Apparent rate constants (min⁻¹) from numerical integration of time course reactions with cellotetraose-NIMS.

Rate	BglA	CelL	CelR	CelE	CelA	CelI	CelK
k1	0.05320	0.000	0.001	0.001	0.000001	0.0002	0.0000
k3	0.14805	0.0002	0.0000	0.0000	0.0000	0.0002	0.0000
k5	0.36803	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k7	0.36803	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k9	0.00047	0.0490	0.0170	0.0092	0.0004	0.0020	0.0439
k11	0.00006	0.0038	0.0453	0.0390	0.0116	0.0479	0.0000
k13	0.00006	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000
k15	0.00006	0.0000	0.0000	0.0000	0.0010	0.0000	0.0004
k17	0.00006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
k19	0.00006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

Rates for individual enzymes are color coded with smallest values in blue and larger values in red.

There are several features of the BglA reaction and simulation that warrant attention. The apparent rates k1, k3, k5, and k7, which correspond to successive removal of single glucose groups, dominate the numerical solution (**Table 2**; **Figure 4**). Under the reaction conditions used, BglA was able to completely convert cellotetraose-NIMS to aglycone-NIMS. It is also noteworthy that shortening the oligosaccharide chain led to an enhancement in the rate of hydrolysis, with reactions k5 (converting cellobiose-NIMS to glucose-NIMS) and k7 (converting glucose-NIMS to aglycone-NIMS) being fastest. Other apparent rates corresponding to side reactions for removal of cellobiose or larger oligosaccharides (e.g., k9 for removal of cellobiose from cellotetraose-NIMS) were less than 1/100th of the value observed for k1, the smallest of the central reactions. These simulation results are consistent with the assigned function of BglA as a β-glucosidase. Indeed, prior oxime-NIMS studies of the reaction of BglA with IL-SG revealed that glucose was the only product released from the biomass substrate (Deng et al., 2014). In the following paragraphs, these diagnostic behaviors of a beta-glucosidase are contrasted with two other classes of GHs, including five phylogenetically diverse endoglucanases and one cellobiohydrolase.

Endoglucanase and Cellobiohydrolase Reactions

Figure 5 shows time courses for reactions of endoglucanases CelA, CelI, CelE, CelR, CelL, and cellobiohydrolase CelK with cellotetraose-NIMS. The reactions of the individual enzymes were carried out and evaluated as described above for **Figure 4**. The appearance of the reaction time courses and the relative rates observed are markedly different than observed for BglA. Unlike the β-glucosidase reaction, no intermediates were observed to form and decay, and the central reactions corresponding to release of glucose units were negligible. This seemingly corresponds with the requirement of endoglucanases for a longer oligosaccharide chain to occupy the active site as a determinant of productive binding and catalysis. In effect, the endoglucanases and cellobiohydrolase primarily reacted only once with the cellotetraose-NIMS probe, leading to a markedly simpler cascade of products than observed for the beta-glucosidase. None of the enzymes characterized in **Figure 5** was able to carry out reactions that yielded the aglycone-NIMS product

(black up triangles), suggesting unproductive binding or blocking steric interactions of the NIMS product with adjacent features of the active site. In contrast, the β-glucosidase BglA (**Figure 4**) was able to successively remove all glucose groups from cellotetraose-NIMS to yield aglycone-NIMS.

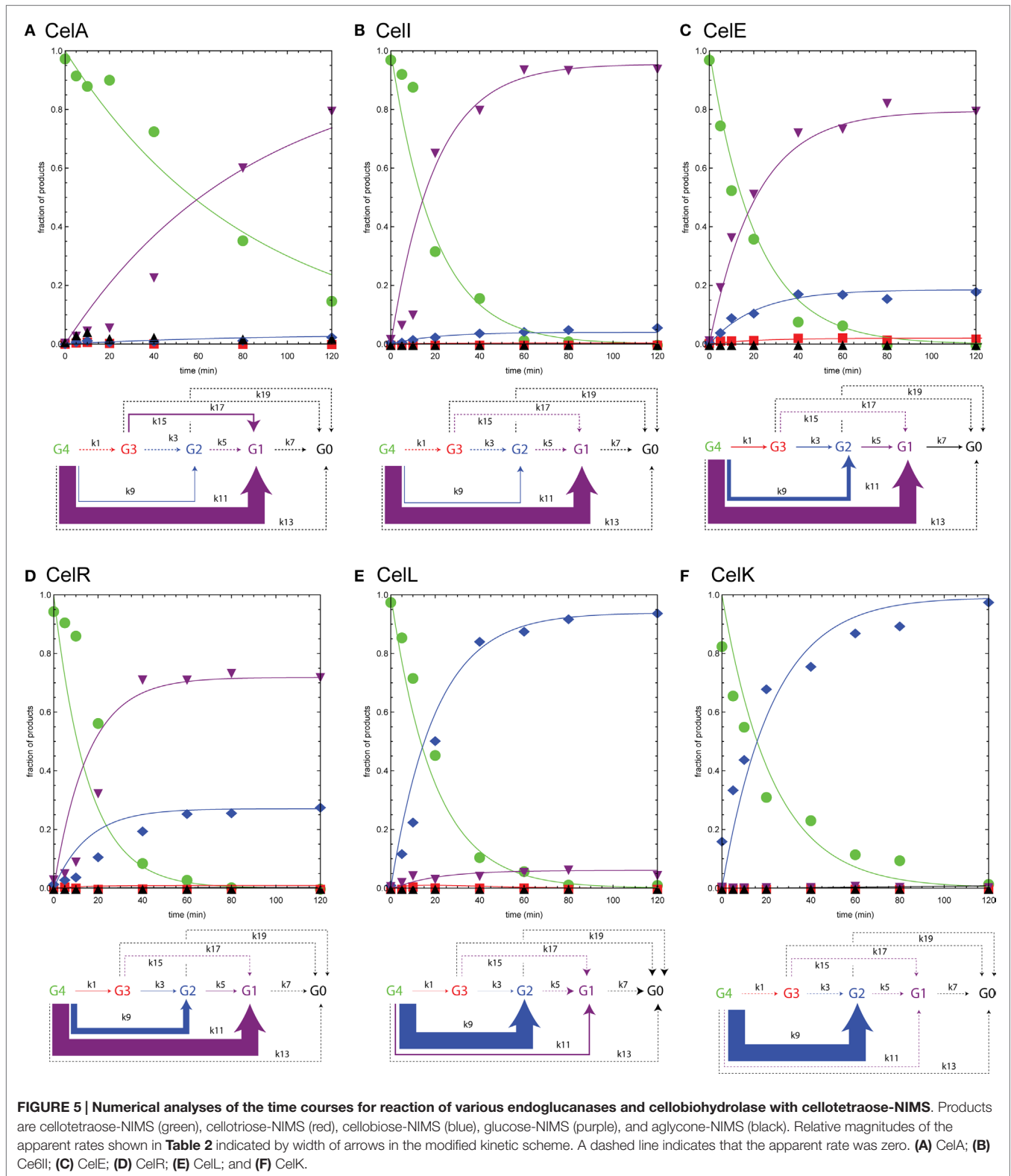## Endoglucanase CelA Reactions

CelA (Cthe_0269) is a GH8 endoglucanase. It is one of the most abundantly transcribed and secreted proteins in *C. thermocellum* during growth on cellulosic substrates (Brown et al., 2007; Gold and Martin, 2007; Raman et al., 2011; Riederer et al., 2011). Analysis of the crystal structure of the enzyme suggested that the substrate binding channel was optimally configured to bind a cellopentaose molecule (Alzari et al., 1996).

The functional characterizations of **Figure 5** demonstrate a progression in reaction selectivity among the enzymes studied. This is a unique power arising from the combination of time-dependent NIMS with numerical analysis. For CelA (**Figure 5A**), k11 governed removal of cellotriose from cellotetraose-NIMS, leading to the predominant accumulation of glucose-NIMS (88%, purple down triangles). The alternative removal of cellotriose via the two step pathway of k1 (removal of glucose) and k15 (removal of cellobiose) contributed ~9% to the overall product yields, while reaction via k9 (removal of cellobiose) added only ~3% of total products as cellobiose-NIMS (blue diamonds). It is worth noting that CelA gave the slowest hydrolysis of cellotetraose-NIMS of all enzymes tested, which is reflected in the values of apparent rates reported in **Table 2** and also in the shape of the plots in **Figure 5**. This may also reflect a partial rate diminution caused by a mismatch between cellotetraose-NIMS and a preferred cellopentaose occupying the active site channel.

In our earlier reactions of CelA with IL-SG (Deng et al., 2014), a mixture of glucose, cellobiose, triose, and tetraose was observed. Other than cellotetraose, whose release from cellotetraose-NIMS was probably prevented by improper binding of the NIMS moiety in the active site channel, the suite of products given by CelA reaction with cellotetraose-NIMS was comparable to that observed from reactions with the pretreated biomass (Deng et al., 2014).

Endoglucanase CelI, CelE, and CelR Reactions

For the reactions of CelI (**Figure 5B**), CelE (**Figure 5C**), and CelR (**Figure 5D**), the dominant pattern of preferred removal of cellotriose units to yield glucose-NIMS (purple down triangles) was retained. However, functional differences of these three enzymes were identified as the removal of cellobiose leading to cellobiose-NIMS (blue diamonds) assumed an increasing contribution to the total product distribution. For example, the observed change corresponds to an approximately eightfold increase in k9 between CelI and CelR. In the middle of these boundary enzymes, CelE was unique among the endoglucanases tested as it was also able to release a glucose unit from cellotetraose-NIMS in ~2% yield. In reactions with IL-SG and ammonia fiber expansion pretreated switchgrass (AFEX-SG) (Deng et al., 2014), these three enzymes released a mixture of glucose, cellobiose, and cellotriose, with the distribution of products in the biomass reaction shifted toward cellobiose and glucose. However, this shift is, in part, due

FIGURE 5 | Numerical analyses of the time courses for reaction of various endoglucanases and cellobiohydrolase with cellotetraose-NIMS. Products are cellotetraose-NIMS (green), cellotriose-NIMS (red), cellobiose-NIMS (blue), glucose-NIMS (purple), and aglycone-NIMS (black). Relative magnitudes of the apparent rates shown in **Table 2** indicated by width of arrows in the modified kinetic scheme. A dashed line indicates that the apparent rate was zero. **(A)** CelA; **(B)** Ce6lI; **(C)** CelE; **(D)** CelR; **(E)** CelL; and **(F)** CelK.

to the ability of these enzymes to cleave solubilized cellotriose into cellobiose and glucose. Subsequent hydrolysis of released oligosaccharides could not be detected when cellotetraose-NIMS was the substrate.

CelI (Cthe_0040) has a structure consisting of GH9 and two CBM3 domains (Hazlewood et al., 1993). It catalyzes the hydrolysis of 1,4-β-glucosidic linkages in cellulose and other glucans. The structure suggests the position of a tunnel that can permit the

release of either cellotriose or cellobiose from cellotetraose-NIMS (PDB 2XFG, no associated publication).

CelE (Cthe_0797) is a multidomain enzyme consisting of GH5, dockerin, and GSDL-lipase domains. Our work has shown that the GH5 domain has broad specificity for reaction with cellulose, xylan, mannan, xyloglucan, and other polysaccharides (Deng et al., 2014; Takasuka et al., 2014). The active site channel of this enzyme is open and tolerates the placement of each of these different linear and branched polysaccharides in a way that a glycosidic bond can be placed in the appropriate position for hydrolysis (Bianchetti et al., 2015). The release of cellotriose, cellobiose, and glucose from cellotetraose-NIMS is compatible with this broad specificity active site. Nevertheless, the active site is not sufficiently tolerant to remove cellotetraose, leading to the formation of aglycone-NIMS.

Previous studies have reported that CelR (Cthe_0578) is a β-glucanase with preference for release of cellotetraose in reactions with amorphous cellulose (Zverlov et al., 2005). Subsequently, CelR was able to convert the longer solubilized oligosaccharide to shorter oligosaccharides. The present studies provide support for this conclusion, as k11 for release of cellotriose was the predominant reaction with cellotetraose-NIMS. Our studies of CelR in reactions with IL-SG and AFEX-SG gave glucose and cellobiose as the dominant hydrolysis products (Deng et al., 2014), suggesting a kinetically rapid conversion of longer oligosaccharides to shorter during the duration of the reaction. Removal of cellotetraose was not observed from cellotetraose-NIMS, which as proposed above likely represents ineffective binding of the NIMS probe in the active site adjacent to the active site.

Endoglucanase CelL and Cellobiohydrolase CelK Reactions

We tested the cellotetraose-NIMS reactions with an additional endoglucanase, CelL (Cthe_0405, **Figure 5E**), and a reducing end cellobiohydrolase, CelK (Cthe_0212, **Figure 5F**). These enzymes show a shift in reaction specificity so that removal of cellobiose to produce cellobiose-NIMS (blue diamonds) became the dominant pattern of reaction. Notably, CelL had an approximately threefold enhanced ability to remove cellobiose relative to CelR because of a higher k9 value and also an ~10-fold decrease in the ability to remove cellotriose associated with a lower k11 value (**Table 2**). CelL reacted with IL-SG also showed preference for release of cellobiose (Deng et al., 2014). Furthermore, although CelK also had an approximately threefold enhanced ability to remove

cellobiose relative to CelR because of a higher k9 value, it showed no ability to produce either cellotriose or glucose (e.g., k1 and k11 = 0; **Table 2**).

The high specificity for release of cellobiose by a cellobiohydrolase is a characteristic reactivity (Amano et al., 1996; Barr et al., 1996; Divne et al., 1998), including CelK (Kataeva et al., 1999) and also CelK reacted with IL-SG (Deng et al., 2014). Thus, cellotetraose-NIMS clearly reports on this catalytic function of CelK. There are no previously published reactivity studies or crystal structures of CelL, beyond our studies of reaction with IL-SG, where CelL showed strong preference for release of cellobiose and xylobiose from the pretreated biomass (Deng et al., 2014).

CONCLUSION

This work establishes the utility of a chemically synthesized mass spectral probe for characterization of GHs. We have shown remarkable correspondence between the products obtained from enzyme reactions with the synthetic cellotetraose-NIMS probe and IL- and AFEX-pretreated switchgrass (Deng et al., 2014). Because of the emerging success of robotic cell-free translation to provide active enzyme samples from synthesized genes (Takasuka et al., 2014; Bianchetti et al., 2015), the substantial advantages of automation and miniaturization afforded by the Nimzyme platform (Deng et al., 2012, 2014; de Rond et al., 2013; Heins et al., 2014), and the predictive power inherent in numerical analysis of enzyme reaction time courses (Cleland, 1975; Orsi and Tipton, 1979; Duggleby, 1995; Marangoni, 2003), our combination offers a powerful new approach for functional annotation of bioenergy phylogenetic space.

AUTHOR CONTRIBUTIONS

KD, TT, CB, LB, PA, TN, and BF designed experiments, carried out experimental work, analyzed results, and prepared the manuscript. All authors read and approved the final manuscript.

FUNDING

REFERENCES

Aceti, D. J., Bingman, C. A., Wrobel, R. L., Frederick, R. O., Makino, S., Nichols, K. W., et al. (2015). Expression platforms for producing eukaryotic proteins: a comparison of *E. coli* cell-based and wheat germ cell-free synthesis, affinity and solubility tags, and cloning strategies. *J. Struct. Funct. Genomics* 16, 67–80. doi:10.1007/s10969-015-9198-1

Alzari, P. M., Souchon, H., and Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* 4, 265–275. doi:10.1016/S0969-2126(96)00031-7

Amano, Y., Shiroishi, M., Nisizawa, K., Hoshino, E., and Kanda, T. (1996). Fine substrate specificities of four exo-type cellulases produced by *Aspergillus niger*, *Trichoderma reesei*, and *Irpex lacteus* on (1 – >3), (1 – >4)-beta-D-glucans and

xyloglucan. *J. Biochem.* 120, 1123–1129. doi:10.1093/oxfordjournals.jbchem. a021531

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. doi:10.1016/j.biotechadv.2009.06.005

Barr, B. K., Hsieh, Y. L., Ganem, B., and Wilson, D. B. (1996). Identification of two functionally different classes of exocellulases. *Biochemistry* 35, 586–592. doi:10.1021/bi9520388

Beebe, E. T., Makino, S., Markley, J. L., and Fox, B. G. (2014). Automated cell-free protein production methods for structural studies. *Methods Mol. Biol.* 1140, 117–135. doi:10.1007/978-1-4939-0354-2_9

Beebe, E. T., Makino, S., Nozawa, A., Matsubara, Y., Frederick, R. O., Primm, J. G., et al. (2011). Robotic large-scale application of wheat cell-free translation to

structural studies including membrane proteins. *N Biotechnol.* 28, 239–249. doi:10.1016/j.nbt.2010.07.003

Bianchetti, C. M., Takasuka, T. E., Deutsch, S., Udell, H. S., Yik, E. J., Bergeman, L. F., et al. (2015). Active site and laminarin binding in glycoside hydrolase family 55. *J. Biol. Chem.* 290, 11819–11832. doi:10.1074/jbc.M114.623579

Blommel, P. G., Martin, P. A., Seder, K. D., Wrobel, R. L., and Fox, B. G. (2009). Flexi vector cloning. *Methods Mol. Biol.* 498, 55–73. doi:10.1007/978-1-59745-196-3_4

Brown, S. D., Raman, B., Mckeown, C. K., Kale, S. P., He, Z., and Mielenz, J. R. (2007). Construction and evaluation of a *Clostridium thermocellum* ATCC 27405 whole-genome oligonucleotide microarray. *Appl. Biochem. Biotechnol.* 13, 663–674. doi:10.1007/s12010-007-9087-6

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi:10.1093/nar/gkn663

Chundawat, S. P., Balan, V., and Dale, B. E. (2008). High-throughput microplate technique for enzymatic hydrolysis of lignocellulosic biomass. *Biotechnol. Bioeng.* 99, 1281–1294. doi:10.1002/bit.21805

Cleland, W. W. (1975). Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry* 14, 3220–3224. doi:10.1021/bi00685a029

Dashtban, M., Maki, M., Leung, K. T., Mao, C., and Qin, W. (2010). Cellulase activities in biomass conversion: measurement methods and comparison. *Crit. Rev. Biotechnol.* 30, 302–309. doi:10.3109/07388551.2010.490938

de Rond, T., Peralta-Yahya, P., Cheng, X., Northen, T. R., and Keasling, J. D. (2013). Versatile synthesis of probes for high-throughput enzyme activity screening. *Anal. Bioanal. Chem.* 405, 4969–4973. doi:10.1007/s00216-013-6888-z

Decker, S. R., Adney, W. S., Jennings, E., Vinzant, T. B., and Himmel, M. E. (2003). Automated filter paper assay for determination of cellulase activity. *Appl. Biochem. Biotechnol.* 105–108, 689–703. doi:10.1385/ABAB:107:1-3:689

Deng, K., George, K. W., Reindl, W., Keasling, J. D., Adams, P. D., Lee, T. S., et al. (2012). Encoding substrates with mass tags to resolve stereospecific reactions using Nimzyme. *Rapid Commun. Mass Spectrom.* 26, 611–615. doi:10.1002/rcm.6134

Deng, K., Takasuka, T. E., Heins, R., Cheng, X., Bergeman, L. F., Shi, J., et al. (2014). Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem. Biol.* 9, 1470–1479. doi:10.1021/cb5000289

Ding, S. Y., Xu, Q., Crowley, M., Zeng, Y., Nimlos, M., Lamed, R., et al. (2008). A biophysical perspective on the cellulosome: new opportunities for biomass conversion. *Curr. Opin. Biotechnol.* 19, 218–227. doi:10.1016/j.copbio.2008.04.008

Divne, C., Stahlberg, J., Teeri, T. T., and Jones, T. A. (1998). High-resolution crystal structures reveal how a cellulose chain is bound in the 50 A long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J. Mol. Biol.* 275, 309–325. doi:10.1006/jmbi.1997.1437

Dowe, N. (2009). Assessing cellulase performance on pretreated lignocellulosic biomass using saccharification and fermentation-based protocols. *Methods Mol. Biol.* 581, 233–245. doi:10.1007/978-1-60761-214-8_15

Duggleby, R. G. (1995). Analysis of enzyme progress curves by nonlinear regression. *Meth. Enzymol.* 249, 61–90. doi:10.1016/0076-6879(95)49031-0

Eklof, J. M., Ruda, M. C., and Brumer, H. (2012). Distinguishing xyloglucanase activity in endo-beta(1 – >4)glucanases. *Meth. Enzymol.* 510, 97–120. doi:10.1016/B978-0-12-415931-0.00006-9

Feinberg, L., Foden, J., Barrett, T., Davenport, K. W., Bruce, D., Detter, C., et al. (2011). Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. *J. Bacteriol.* 193, 2906–2907. doi:10.1128/JB.00322-11

Fontes, C. M., and Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79, 655–681. doi:10.1146/annurev-biochem-091208-085603

Fox, B. G., Goulding, C., Malkowski, M. G., Stewart, L., and Deacon, A. (2008). Structural genomics: from genes to structures with valuable materials and many questions in between. *Nat. Methods* 5, 129–132. doi:10.1038/nmeth0208-129

Gerlt, J. A., Allen, K. N., Almo, S. C., Armstrong, R. N., Babbitt, P. C., Cronan, J. E., et al. (2011). The enzyme function initiative. *Biochemistry* 50, 9950–9962. doi:10.1021/bi201312u

Gold, N. D., and Martin, V. J. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* 189, 6787–6795. doi:10.1128/JB.00882-07

Grabnitz, F., Seiss, M., Rucknagel, K. P., and Staudenbauer, W. L. (1991). Structure of the beta-glucosidase gene bglA of *Clostridium thermocellum*. Sequence analysis reveals a superfamily of cellulases and beta-glycosidases including human lactase/phlorizin hydrolase. *Eur. J. Biochem.* 200, 301–309. doi:10.1111/j.1432-1033.1991.tb16186.x

Hazlewood, G. P., Davidson, K., Laurie, J. I., Huskisson, N. S., and Gilbert, H. J. (1993). Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. *J. Gen. Microbiol.* 139, 307–316. doi:10.1099/00221287-139-2-307

Heins, R. A., Cheng, X., Nath, S., Deng, K., Bowen, B. P., Chivian, D. C., et al. (2014). Phylogenomically guided identification of industrially relevant GH1 beta-glucosidases through DNA synthesis and nanostructure-initiator mass spectrometry. *ACS Chem. Biol.* 9, 2082–2091. doi:10.1021/cb500244v

Henrissat, B., and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7, 637–644. doi:10.1016/S0959-440X(97)80072-3

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi:10.1126/science.1137016

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. doi:10.1128/AEM.00772-15

Hirano, N., Hasegawa, H., Nihei, S., and Haruki, M. (2013). Cell-free protein synthesis and substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest multi-enzyme subunit of the *Clostridium thermocellum* cellulosome. *FEMS Microbiol. Lett.* 344, 25–30. doi:10.1111/1574-6968.12149

Horn, S. J., Sorlie, M., Varum, K. M., Valjamae, P., and Eijsink, V. G. (2012). Measuring processivity. *Meth. Enzymol.* 510, 69–95. doi:10.1016/B978-0-12-415931-0.00005-7

Kataeva, I., Li, X. L., Chen, H., Choi, S. K., and Ljungdahl, L. G. (1999). Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum*: evidence for gene duplication and recombination. *J. Bacteriol.* 181, 5288–5295.

Kosik, O., Bromley, J. R., Busse-Wicher, M., Zhang, Z., and Dupree, P. (2012). Studies of enzymatic cleavage of cellulose using polysaccharide analysis by carbohydrate gel electrophoresis (PACE). *Meth. Enzymol.* 510, 51–67. doi:10.1016/B978-0-12-415931-0.00004-5

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

Madono, M., Sawasaki, T., Morishita, R., and Endo, Y. (2011). Wheat germ cell-free protein production system for post-genomic research. *N Biotechnol.* 28, 211–217. doi:10.1016/j.nbt.2010.08.009

Makino, S., Beebe, E. T., Markley, J. L., and Fox, B. G. (2014). Cell-free protein synthesis for functional and structural studies. *Methods Mol. Biol.* 1091, 161–178. doi:10.1007/978-1-62703-691-7_11

Marangoni, A. G. (2003). *Enzyme Kinetics*. Hoboken, NJ: John Wiley & Sons, Inc.

Markley, J. L., Aceti, D. J., Bingman, C. A., Fox, B. G., Frederick, R. O., Makino, S., et al. (2009). The center for eukaryotic structural genomics. *J. Struct. Funct. Genomics* 10, 165–179. doi:10.1007/s10969-008-9057-4

McCleary, B. V., Mckie, V., and Draga, A. (2012). Measurement of endo-1,4-beta-glucanase. *Meth. Enzymol.* 510, 1–17. doi:10.1016/B978-0-12-415931-0.00001-X

Nair, R., Liu, J., Soong, T. T., Acton, T. B., Everett, J. K., Kouranov, A., et al. (2009). Structural genomics is the largest contributor of novel structural leverage. *J. Struct. Funct. Genomics* 10, 181–191. doi:10.1007/s10969-008-9055-6

Nelder, J. A., and Mead, R. (1965). A simplex-method for function minimization. *Comput. J.* 7, 308–313. doi:10.1093/comjnl/7.4.308

Northen, T. R., Lee, J. C., Hoang, L., Raymond, J., Hwang, D. R., Yannone, S. M., et al. (2008). A nanostructure-initiator mass spectrometry-based enzyme activity assay. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3678–3683. doi:10.1073/pnas.0712332105

Orsi, B. A., and Tipton, K. F. (1979). Kinetic analysis of progress curves. *Meth. Enzymol.* 63, 159–183. doi:10.1016/0076-6879(79)63010-0

Pena, M. J., Tuomivaara, S. T., Urbanowicz, B. R., O'neill, M. A., and York, W. S. (2012). Methods for structural characterization of the products of cellulose- and xyloglucan-hydrolyzing enzymes. *Meth. Enzymol.* 510, 121–139. doi:10.1016/B978-0-12-415931-0.00007-0

Pieper, U., Schlessinger, A., Kloppmann, E., Chang, G. A., Chou, J. J., Dumont, M. E., et al. (2013). Coordinating the impact of structural genomics on the human alpha-helical transmembrane proteome. *Nat. Struct. Mol. Biol.* 20, 135–138. doi:10.1038/nsmb.2508

Raman, B., Mckeown, C. K., Rodriguez, M. Jr., Brown, S. D., and Mielenz, J. R. (2011). Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol.* 11:134. doi:10.1186/1471-2180-11-134

Riederer, A., Takasuka, T. E., Makino, S., Stevenson, D. M., Bukhman, Y. V., Elsen, N. L., et al. (2011). Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl. Environ. Microbiol.* 77, 1243–1253. doi:10.1128/AEM.02008-10

Roberts, S. B., Gowen, C. M., Brooks, J. P., and Fong, S. S. (2010). Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst. Biol.* 4:31. doi:10.1186/1752-0509-4-31

Selig, M. J., Tucker, M. P., Law, C., Doeppke, C., Himmel, M. E., and Decker, S. R. (2011). High throughput determination of glucan and xylan fractions in lignocelluloses. *Biotechnol. Lett.* 33, 961–967. doi:10.1007/s10529-011-0526-7

Sharrock, K. R. (1988). Cellulase assay methods: a review. *J. Biochem. Biophys. Methods* 17, 81–105. doi:10.1016/0165-022X(88)90040-1

Smith, S. P., and Bayer, E. A. (2013). Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* 23, 686–694. doi:10.1016/j.sbi.2013.09.002

Takasuka, T. E., Walker, J. A., Bergeman, L. F., Vander Meulen, K. A., Makino, S., Elsen, N. L., et al. (2014). Cell-free translation of biofuel enzymes. *Methods Mol. Biol.* 1118, 71–95. doi:10.1007/978-1-62703-782-2_5

Watson, J. D., Sanderson, S., Ezersky, A., Savchenko, A., Edwards, A., Orengo, C., et al. (2007). Towards fully automated structure-based function prediction in structural genomics: a case study. *J. Mol. Biol.* 367, 1511–1522. doi:10.1016/j.jmb.2007.01.063

Whitehead, C., Gomez, L. D., and Mcqueen-Mason, S. J. (2012). The analysis of saccharification in biomass using an automated high-throughput method. *Meth. Enzymol.* 510, 37–50. doi:10.1016/B978-0-12-415931-0.00003-3

Wischmann, B., Toft, M., Malten, M., and Mcfarland, K. C. (2012). Biomass conversion determined via fluorescent cellulose decay assay. *Meth. Enzymol.* 510, 19–36. doi:10.1016/B978-0-12-415931-0.00002-1

Zverlov, V. V., Schantz, N., and Schwarz, W. H. (2005). A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-be-ta-1,4-glucanase producing cellotetraose. *FEMS Microbiol. Lett.* 249, 353–358. doi:10.1016/j.femsle.2005.06.037