



Published in final edited form as:

J Am Chem Soc. 2012 December 19; 134(50): 20521–20532. doi:10.1021/ja309790w.

A Suite of Activity-Based Probes for Cellulose Degrading Enzymes

Lacie M. Chauvigné-Hines^{#,‡}, Lindsey N. Anderson[‡], Holly M. Weaver, Joseph N. Brown, Phillip K. Koeh, Carrie D. Nicora, Beth A. Hofstad, Richard D. Smith, Michael J. Wilkins, Stephen J. Callister, and Aaron T. Wright^{*}

Biological Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, Washington, 99352

Abstract

Microbial glycoside hydrolases play a dominant role in the biochemical conversion of cellulosic biomass to high-value biofuels. Anaerobic cellulolytic bacteria are capable of producing multicomplex catalytic subunits containing cell-adherent cellulases, hemicellulases, xylanases, and other glycoside hydrolases to facilitate the degradation of highly recalcitrant cellulose and other related plant cell wall polysaccharides. *Clostridium thermocellum* is a cellulosome producing bacterium that couples rapid reproduction rates to highly efficient degradation of crystalline cellulose. Herein, we have developed and applied a suite of difluoromethylphenyl aglycone, *N*-halogenated glycosylamine, and 2-deoxy-2-fluoroglycoside activity-based protein profiling (ABPP) probes to the direct labeling of the *C. thermocellum* cellulosomal secretome. These activity-based probes (ABPs) were synthesized with alkynes to harness the utility and multimodal possibilities of click chemistry, and to increase enzyme active site inclusion for LC-MS analysis. We directly analyzed ABP-labeled and unlabeled global MS data, revealing ABP selectivity for glycoside hydrolase (GH) enzymes, in addition to a large collection of integral cellulosome-containing proteins. By identifying reactivity and selectivity profiles for each ABP, we demonstrate our ability to widely profile the functional cellulose degrading machinery of the bacterium. Derivatization of the ABPs, including reactive groups, acetylation of the glycoside binding groups, and mono- and disaccharide binding groups, resulted in considerable variability in protein labeling. Our probe suite is applicable to aerobic and anaerobic microbial cellulose degrading systems, and facilitates a greater understanding of the organismal role associated with biofuel development.

INTRODUCTION

The desire to identify biofuel alternatives to liquid fossil fuels has kindled worldwide interest in converting cellulosic biomass to high-value fuels and other small carbon compounds.¹ Cellulosic biomass is the most abundant renewable source of carbon and

^{*}Corresponding Author: aaron.wright@pnl.gov.

[#]Present Addresses

Lacie Chauvigné-Hines, Battelle Charlottesville, 1001 Research Park Blvd, Town Center Two, Suite 400, Charlottesville, VA 22911

[‡]Author Contributions

These authors contributed equally. The manuscript was written through contributions of all authors, and all authors have given approval to the final version of the manuscript.

The authors declare no competing financial interests.

ASSOCIATED CONTENT

Supporting Information. Probe syntheses and characterization. Gel with controls. NMR spectra for probes. Heatmap showing all carbohydrate active proteins labeled by **GH-ABPs**. Data for all ABP and global measurements. STRING interactions. Unique peptides identified for proteins. This material is available free of charge via the Internet at <http://pubs.acs.org>.

energy on the planet, with estimates of 180 million tons of feedstocks from agriculture and forest litter available per year.² Harnessing the catalytic prowess of glycoside hydrolases (GHs) within microbial cellulose degrading machinery for biofuels development is of high value and high interest.¹ Cellulolytic bacteria and fungi secrete complex suites of cellulases, hemicellulases, and accessory enzymes for hydrolysis of highly recalcitrant cellulose, and other related plant cell wall polysaccharides.

The rate-limiting step in converting cellulosic material to biofuels is the hydrolysis of polysaccharides by glycoside hydrolases. A gram-positive, thermophilic, anaerobic bacterium, *Clostridium thermocellum*, has the highest rate of cellulose utilization of any bacterium making it of particular interest for biofuel production.³ Among living organisms, *C. thermocellum* belongs to a unique group of bacterium that can utilize cellulosic materials to generate ethanol directly.⁴ *C. thermocellum* secretes a cell surface-bound macromolecular multienzyme complex, termed the cellulosome.⁵ This extensive consortia of extracellular enzymes synergistically degrades the recalcitrant polysaccharides of cellulosic materials. The cellulosome assembly is mediated by a noncatalytic scaffoldin (CipA) protein, which through high-affinity interactions can bear up to nine catalytic enzyme units, and contains a family 3 cellulose-binding module (CBM3) for attachment to cellulosic materials and a type II dockerin module for attachment of the cellulosome units to the cell surface.⁶ The cellulosomal enzyme units possess a type I dockerin module that bind strongly to cohesin modules on scaffoldin.⁷

The *C. thermocellum* genome encodes 72 proteins with type I dockerin modules.⁵ These include cellulases, β -glucanases, xylanases, pectinases, mannanases, xyloglucanases and proteinases, matching the complex diversity of the cellulosic material the bacterium acts upon.⁵ These catalytic activities can be generally categorized as: (i) endoglucanases, which hydrolyze internal sites on the cellulose chain, (ii) exoglucanases, which hydrolyze the reducing or nonreducing free chain ends of cellulose, and can act in a continuous, processive manner, and (iii) β -glucosidases, which hydrolyze soluble cellobioses and cellobiose to glucose.² These enzymes hydrolyze glycosidic bonds resulting in an inversion or retention of the stereochemical configuration of the anomeric center. Numerous glycoside hydrolase families within the catalytic machinery of cellulosomes fall within these categories.^{8,9}

Activity-based protein profiling (ABPP) has been successfully applied to the identification of enzymatic activities within complex proteomes.¹⁰ Glycoside hydrolases play a crucial role in biological functions from human health to biofuel generation, thus leading to the development of activity-based probes (ABPs) in order to profile these valuable enzymes.^{11–22} Herein, ABPs have been developed as affinity-based or mechanism-based irreversible inhibitors. Our affinity-based ABPs contain a region to impart specificity for glycoside hydrolases alongside a reactive group that forms a covalent bond to catalytic or noncatalytic regions of the enzyme. Mechanism-based ABPs form a covalent enzyme-ABP bond upon catalytic reaction.²³ Although multiple ABPs have been developed for GHs, no large-scale comparative analysis has been performed. Additionally, the array of diversity and variation of catalytic mechanisms found within GH families limits the profiling scope of a single ABP in a complex biological system.

To address the reactivity and selectivity of multiple GH-ABPs, we describe herein the synthesis and functional characterization of a suite of ABPs based on known scaffolds for affinity and mechanism-based inhibition of glycoside hydrolases (Figure 1A). All probes are click-chemistry compatible to reduce probe size and allow simultaneous fluorescent and LC-MS characterization of ABP-protein labeling events.²⁴ The ABP suite is applied directly to the complex cellulose degrading proteome of *C. thermocellum*. Not only do we show how these probes have complementary and distinct enzyme labeling profiles, we demonstrate the

utility of these probes for characterizing the cellulose degrading machinery of biofuel-relevant organisms.

EXPERIMENTAL SECTION

General Procedures and Materials

NMR spectra were recorded at 25 °C on a Varian Oxford 500 MHz spectrometer at the following frequencies: 499.8 MHz (^1H) and 125.7 MHz (^{13}C); spectra were calibrated to the chemical shift of tetramethylsilane ($\delta = 0$ ppm). Spectra were assigned with appropriate ^1H and ^{13}C NMR experiments. Chemical shifts are reported in ppm, coupling constants in Hertz (Hz), and multiplicities indicated with: singlet (s), doublet (d), triplet (t), doublet of doublets (dd), doublet of triplets (dt), doublet of doublet of doublets (ddd), and multiplet (m). ESI mass spectra were obtained with a LCQ Deca spectrometer. High-resolution mass spectra were obtained with an Exactive Orbitrap mass spectrometer (Thermo Scientific). Protein concentration measurements were made using a SpectraMax 96-well UV/Vis microplate reader. Fluorescent ABP-labeled proteins were separated by SDS-PAGE on Invitrogen 10% Tris-Glycine precast gels, and imaged with a Protein Simple FluorchemQ: 534 nm excitation, 606 nm emission filter. Unless otherwise noted, a Grace Davison Discovery Sciences Reveleris medium pressure liquid chromatographer fitted with commercial silica cartridges was used for purification of ABPs and intermediates. Starting mono- and disaccharide material for the synthesis of all **GH-ABPs** was purchased from Carbosynth Ltd. Other reagents were purchased from Sigma-Aldrich, Acros, or Alfa Aesar and used as received unless stated otherwise. Dry solvents, if not purchased, were obtained via a LC Technology Solutions, Inc., SP-1 solvent drying system; reactions were carried out in an inert nitrogen or argon environment.

Synthesis of GH-ABPs

See Supporting Information.

Bacterial Culture

Clostridium thermocellum (ATCC 27405) was grown in a modified Dehority medium,²⁵ and supplemented with 4 g L⁻¹ microcrystalline cellulose (without the addition of volatile fatty acids), and incubated at 60 °C under strict anaerobic conditions. Biomass was harvested between mid- and late-log phases by centrifugation, and flash frozen upon pelleting with liquid nitrogen. Spent growth media, containing secreted and cellulosome proteins (“cellulosomal secretome”), was concentrated using a Millipore Amicon centrifugal device. The buffer was exchanged to NaOAc (50 mM, pH 6), and the cellulosomal secretome collected. After initial protein content was determined by BCA, protein concentrations were normalized and stored at -80 °C until further analysis.

Enzyme Assays

Cellulase (endo-1,4- β -glucanase) and xylanase (endo-1,4- β -xylanase) activity were determined according to manufacturer (Megazyme Intl. Ireland) specifications. Both assays release remazolbrilliant blue R coupled substrate, which is measured by absorbance at 590nm; cellulase activity: 140 milliU/mL; xylanase activity: 60 milliU/mL.

Proteome Labeling and Fluorescent Gel Analysis

Proteome samples (35 μL) were adjusted to a protein concentration of 1 mg/mL in NaOAc (50 mM, pH 6 containing SDS (0.42 μM)) buffer and labeled with an individual GH-ABP (**GH1** (300 μM), **GH2a-d** (75 μM), **GH3a-b** (75 μM), and **GH4a-b** (75 μM)), and incubated at 60 °C for 3 h with mild agitation.

Following ABP incubation, proteome samples treated with an azide-containing probe were treated with a rhodamine-alkyne fluorescent reporter group (30 μM , prepared in DMSO), dithiothreitol (1.4 mM, prepared fresh in water), tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (TBTA) (171 μM , prepared in DMF), and CuSO_4 (2.86 mM, prepared in water). Proteome samples treated with an alkyne-containing probe were treated with a rhodamineazide fluorescent reporter group (30 μM), tris(2-carboxyethyl)phosphine (TCEP) (429 μM , prepared fresh in water), TBTA (87 μM , prepared in 4:1 *tert*-butanol: DMSO), and CuSO_4 (857 μM , prepared in water). Samples were vortexed and incubated in the dark at rt for 1 h, at which time SDS-PAGE loading buffer (reducing) was added and samples were heated at 85 $^\circ\text{C}$ for 2 min prior to being loaded into precast 10% Tris-Glycine gels.

Global Sample Prep and LC-MS Analysis

Protein concentrations were determined by Coomassie assay (Thermo Scientific), then denatured and reduced by adding a final concentration of 8 M urea and fresh dithiothreitol (DTT) to a final concentration of 10 mM. Samples were incubated at 60 $^\circ\text{C}$ for 30 min, then diluted 8-fold with NH_4HCO_3 (100 mM, pH 8.4), to reduce salt concentration. CaCl_2 was added to the diluted samples to a final concentration of 1 mM, and digested for 3 h at 37 $^\circ\text{C}$ using sequencing grade trypsin (Promega) at a ratio of 1 unit of trypsin per 50 units of protein. Following incubation, digested samples were desalted using an appropriately sized C-18 SPE column (Supelco). Final peptide concentrations were determined by BCA assay.

Digested proteins from spent growth media and cell pellet fractions were analyzed using a custom-built 2D HPLC system coupled to an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher). The first dimension of separation occurred by way of a strong cation exchange (SCX) column (15-cm \times 360 μm o.d. \times 150 μm i.d.) containing 5- μm Polysulfoethyl A (PolyLC Inc) coupled to a trapping column (4-cm \times 360 μm o.d. \times 150 μm i.d.) containing 5- μm Jupiter C18 (Phenomenex). Fifteen fractions were trapped and individually separated by a reverse-phase column (35-cm \times 360 μm o.d. \times 75 μm i.d.) containing 3- μm Jupiter C18. Mobile phases consisted of 0.1 mM NaH_2PO_4 and 0.3 M NaH_2PO_4 for the first dimension, and 0.1% formic acid in water and 0.1% formic acid in acetonitrile for the second dimension. Columns were manufactured in-house by slurry packing media into fused silica (Polymicro Technologies Inc) using a 1 cm sol-gel frit for media retention.²⁶

Instrument operating conditions included heated capillary temperature of 200 $^\circ\text{C}$ and a spray voltage of 2.2 kV, respectively. Data was acquired for approximately 100 min per fraction, beginning 10 min into the reverse-phase gradient. Orbitrap spectra (AGC 1×10^6) were collected from 400–2000 m/z at a resolution of 100k followed by data-dependent ion trap MS/MS spectra (AGC 3×10^4) of the six most abundant ions using a collision energy of 35%.²⁷ A dynamic exclusion time of 60 sec was implemented to discriminate against previously analyzed ions.

Proteome Labeling and LC-MS Analysis

Proteome samples were adjusted to a protein concentration of 0.5 mg/mL in NaOAc buffer (50 mM, pH 6, containing SDS (0.42 μM)) and labeled using individual GH-ABPs: **GH1** (300 μM), **GH2a–d** (75 μM), **GH3a–b** (75 μM), and **GH4a–b** (75 μM). “No ABP” control samples were prepared by addition of DMSO (2 μL) rather than probe. All samples were incubated at 60 $^\circ\text{C}$ for 3 h with mild agitation. Following probe incubation, azide-containing **GH-ABP** labeled proteome samples were treated with biotin-alkyne (30 μM), dithiothreitol (DTT; 1.67 mM, prepared fresh in water), TBTA (100 μM , prepared in DMF), and CuSO_4 (1.65 mM final, prepared in water). Proteome samples treated with alkyne-containing **GH-ABPs** were treated with biotin-azide (30 μM), TCEP (250 μM , prepared fresh in water),

TBTA (51 μM , prepared in 4:1 *tert*-butanol:DMSO), and CuSO_4 (0.5 mM, prepared in water). Samples were vortexed and incubated at rt for 1.5 h.

Following click chemistry, samples were concentrated using Amicon Ultra-0.5 centrifugal filter devices at 14,000 $\times g$ for 20 min at 4 $^\circ\text{C}$ and washed with PBS (10 mM, pH 7.4, 2 \times 300 μL) before re-concentrating to 20 μL . PBS (0.5 mL, containing 1.2% SDS) was added to the filter device, then inverted into a clean recovery tube and collected at 1,000 $\times g$ for 4 min at 4 $^\circ\text{C}$. Samples were heated at 95 $^\circ\text{C}$ for 2 min and centrifuged at rt for 4 min at 6,000 $\times g$, removing particulates. Protein concentrations were determined for each sample using a BCA assay; all samples were normalized to 300 μg prior to enrichment. In order for the analysis to result in the strict evaluation of protein function on a per unit mass basis, protein content was normalized prior to loading proteome samples onto streptavidin resin.

For enrichment of ABP-labeled proteins, a 100 μL aliquot of streptavidin agarose resin (Thermo Fisher), per prepared sample, was placed in a BioRad Bio-Spin Chromatography Column using a vacuum manifold. The resin was washed with PBS (3 mL) and transferred to a 15 mL tube using two 0.5 mL aliquots of PBS. An additional 1.5 mL of PBS was added to each tube followed by the normalized proteome sample. Each falcon tube contained a final SDS concentration of \sim 0.2%. Tubes were then rotated for 4 h at rt.

Following streptavidin capture of probe-labeled proteins, each sample containing resin solution was transferred into a Bio-Spin column on the vacuum manifold. The resin was washed with 0.5% SDS in PBS (1 mL, repeat 3 \times), 6 M urea in PBS (1 mL, repeat 3 \times), MilliQ water (1 mL, repeat 3 \times), and PBS (1 mL, repeat 5 \times). The resin was then transferred to sterile 1.5 mL tubes using two 0.5 mL aliquots of ammonium bicarbonate (25 mM, pH 8), and the supernatant discarded.

To obtain peptides for MS analysis, 200 μL NH_4HCO_3 (25 mM, pH 8) was added to the resin for each sample and treated with a trypsin solution (2 μL , trypsin was reconstituted in 40 μL of NH_4HCO_3). Resin solutions were heated at 37 $^\circ\text{C}$ for 15 h with agitation. Following trypsin digestion, supernatant was collected by centrifugation at 6,000 $\times g$ and placed in a sterile tube. An additional amount of NH_4HCO_3 (150 μL) was added to sample resin and agitated on a thermal mixer at 37 $^\circ\text{C}$ for 10 min. Supernatants were again collected at 6,000 $\times g$ and added to corresponding tubes from prior collection. Volatiles were removed from tryptic peptide solutions by speedvac concentrator, and peptides were reconstituted in 40 μL NH_4HCO_3 , and heated for 10 min at 37 $^\circ\text{C}$ with mild agitation. Samples were centrifuged at 53,000 rpm in a Beckman TLA 120.1 rotor for 20 min at 4 $^\circ\text{C}$ then placed in vials for subsequent MS analysis using 25 μL aliquots.

LC-MS Proteomic Analysis of ABP-Labeled Proteins

Peptides were separated by high-resolution, reversed phase constant pressure capillary liquid chromatography as previously described.^{28,29} Briefly, MS analysis was performed using a Thermo Electron ion trap LTQ MS outfitted with a custom ion funnel and electrospray ionization (ESI) interface. Data was acquired for 100 min, beginning 65 min after sample injection (15 min into gradient). LTQ spectra were collected from 400–2000 m/z at a resolution of 100k, followed by data-dependent ion trap MS/MS spectra of the six most abundant ions using a 35% collision energy.²⁹ A dynamic exclusion time of 30 sec was used to discriminate against previously analyzed ions.

Data Analysis

LTQ-MS raw data was extracted using Decon_ MSn,³⁰ and analyzed with the SEQUEST algorithm (V27, revision 12)³¹ by aligning experimental MS spectra with theoretical spectra

generated from the *C. thermocellum* annotated genome (*C. thermocellum* ATCC 27405 Genbank Accession CP000568; downloaded_2008-04-07).

Data filtering criteria based on the MS-GF score and precursor ion mass accuracy were used to limit false positive identifications to <1% at the peptide level, estimated employing a reverse database search.³² Using peptides that match the given criteria, we then stipulated that in order to classify a protein as specifically labeled by a **GH-ABP**, we required the following criteria: (i) 2 unique peptides per protein; (ii) 2 peptides measured per protein in at least two replicates; (iii) the protein exhibits 3-fold more abundance in the **GH-ABP**-labeled samples relative to the “No probe” control. All filter passing peptide identifications were tallied to provide quantitative spectral count data for each protein. Relative protein abundance was estimated by averaging peptide spectral counts for a protein across the three **GH-ABP**-labeling replicates, with standard deviation under 10%.^{30,31,33}

RESULTS AND DISCUSSION

Probe Design and Synthesis

Activity-based probes **GH1-GH4b** (Figure 1A) are designed to facilitate two-step bioorthogonal chemistry, whereby probes first label enzymes, followed by Cu(I)-catalyzed cycloaddition (click chemistry(CC)) to a reporter group. Our CC-enabled **GH-ABPs** are composed of three general elements to facilitate capture and identification of cellulosomal GHs: (1) a reactive group as either an affinity-based or mechanism-based inhibitor for targeting active enzymes, (2) a mono- or disaccharide as the binding group that facilitates probe-enzyme adduction, and (3) a latent alkyne or azide handle for CC attachment of a reporter group to visualize binding events by fluorescence, or identify ABP targets by biotin enrichment and MS measurement.²⁴ Incorporation of a CC-compatible alkyne or azide to the **GH-ABP** replaces bulky biotin or fluorophores that would not normally be part of the natural substrate for GH enzymes.

At the onset of our studies we evaluated a number of known classes of affinity and mechanism-based inhibitors for comparison, paying particular attention to the likelihood of the probe class to capture elements of the cellulosome. 2-Deoxy-2-fluoroglycosides were first reported as glycosidase inhibitors in 1987 by Withers and coworkers,³⁴ and first as ABPs by Vocadlo and Bertozzi.¹⁵ These inhibit retaining GHs with discriminating specificity for the active site. The 2-deoxy-2-fluoro substitution slows the formation and hydrolysis of the covalent glycosyl-enzyme adduct by stabilizing the oxocarbenium like transition state.^{23,35} However, the glycosyl-enzyme adduct can slowly hydrolyze with a lifetime ranging from seconds to months; thus it may be described as a very slow enzymatic substrate rather than a mechanism-based inhibitor.²³ Despite contrasting interpretations, the long life of the adduct was deemed suitable for our proteomic experiments; in turn, we synthesized the glucose-based **GH1-ABP** (Scheme 1A) according to literature procedures^{15,18,19} and utilized selective protection and reactivity manipulation of the carbohydrate functional groups.

Probes **GH2a-d-ABPs** are affinity-based inhibitors, in which the glycosyl-binding group imparts selectivity toward enzymes acting on sugar moieties; however, these probes can label the acid/base catalytic residues of retaining β -glycosidases, as well as amino acid residues within the active site of inverting and retaining enzymes through nucleophilic displacement of the acetylhalide.^{23,36} *N*-bromo- and *N*-iodoacetyl glycosylamines have been used to label and inactivate several glycosidases and cellulases, and show sufficient stability toward decomposition.^{36,37} For broad characterization of the components of the cellulosome, it can be said that these probes offer an advantage by covering inverting and retaining enzymes, and without specific activity toward a single GH family. In the design of

probes **GH2b-** and **GH2d-ABP** the acetyl group was retained, as the lipophilic property of the acetylated sugar probes may be efficacious for binding to certain GH families.³⁷ In a single step synthesis of probes **GH2a-ABP** and **GH2c-ABP**, 6-azido-1,6-dideoxy- β -D-glucopyranosylamine was reacted directly with either bromo- or iodoacetic anhydride, respectively (Scheme 1B).³⁸ To obtain the acetylated **GH2b-ABP** and **GH2d-ABP** probes, 2,3,4-tri-*O*-acetyl-1-amino-6-azido-1,6-dideoxy- β -D-glucose was reacted directly with bromo- or iodoacetic anhydride, respectively (Scheme 1B).

Probes of structure **GH3-ABP** and **GH4-ABP** are activated difluoromethylphenyl aglycones, which upon glycoside hydrolase cleavage of the glycosidic bond release an activated aglycone that subsequently irreversibly reacts with the GH. Unlike the **GH1-** and **GH2-ABPs**, these probes are solely activated after GH enzymatic cleavage.^{12,13} Hydrolysis of the glycosidic bond by a GH liberates a difluoromethyl phenolate, followed by rapid elimination of a fluorine atom to form a reactive quinone methide. Any nucleophile within the active site, including catalytic residues, can perform a Michael reaction with the quinone methide yielding a covalent adduct. A drawback from prior studies, but an advantage to broadly identifying cellulosome activity, is that the intermediate quinone methide can transiently diffuse from the GH active site, and label a reactive GH outside the active site, or label neighboring cellulosome components. However, biotin and fluorophore containing ABPs have been successfully applied to the labeling of GHs, and in biologically relevant systems.^{12,13,16,39} We designed our CC enabled phenyl-dimethyl aglycone probes as mono- and disaccharides, with and without acetylation. Beginning with 1-bromo-1-deoxy-2,3,4,6-tetra-*O*-acetyl- α -D-galactopyranose (**GH3-ABPs**) or 1-bromo-1-deoxy-2,3,4,2',3',4',6'-hepta-*O*-acetyl- α -D-lactopyranose (**GH4-ABPs**), the compounds were reacted with nitrophenol, **10**, followed by fluorination of the aldehyde, **11/15**,^{39a} and reduction of the nitro moiety to an amine, **12/16**.^{39a} An alkyne for click chemistry was appended to form **GH3b-ABP** and **GH4b-ABP**. Removal of the acetyl groups provided **GH3a-ABP** and **GH4a-ABP**. Prior studies have demonstrated that background labeling of proteins during click chemistry is heightened when the probe contains the azide, and the reporter group the alkyne.^{21,24} Due to the ease of installation, we appended the alkyne groups.

Probe Reactivity and Gel Analysis in *C. thermocellum*

Effective degradation of lignocellulosic materials for production of ethanol and biofuel precursors is a hallmark of the anaerobic bacterium *C. thermocellum*. This organism produces an extensive consortium of proteins and enzymes that work synergistically to degrade cellulosic matter. Some proteins are secreted directly to the medium, while others contain a dockerin module, which strongly bind cohesin modules on scaffoldin (CipA) forming a macromolecular “cellulosome” structure adherent to the cell surface (Figure 2). *C. thermocellum* secrete a large number of GHs from multiple inverting and retaining families, including: exoglucanases, endoglucanases, hemicellulases, xylanases, pectinases, and other minor enzyme activities exhibited. To determine probe reactivity and selectivity toward GHs, including those associated with the cellulosome, we screened our suite of **GH-ABPs** against the *C. thermocellum* spent growth culture medium, which contains secreted proteins and cellulosome proteins (aka, “cellulosomal secretome”). Probes were incubated with the cellulosomal secretome sample for 3 h at 60 °C at a final concentration of 75 μ M with the exception of **GH1-ABP**. Due to the lower reactivity of **GH1-ABP** and its potential for hydrolysis, the post-enzyme reaction was used at final concentration of 300 μ M instead.¹⁵ Following **GH-ABP** incubation, proteome samples were treated with either rhodamine-azide or rhodamine-alkyne under click chemistry conditions, separated by SDS-PAGE, and visualized by fluorescence (Figure 1B).

The probe labeling profiles observed by fluorescent gel analysis (Figure 1B) show diverse protein reactivity mediated by the reactive group selection, acetylation status of the hydroxyl groups on the sugar binding groups, and **GHABP** size. Unfortunately, **GH1-ABP**, displayed no observable binding by fluorescent gel. Increases to the concentration and alterations to labeling conditions including time and temperature consistently failed to produce any labeling with **GH1-ABP**. Within **GH2-ABPs**, reactivity was increased by acetylation (**GH2b-** and **GH2d-ABPs**), but the effect of a Br or I in the reactive group had minimal impact. Within **GH3-** and **GH4-ABPs**, acetylation decreased reactivity, and the larger disaccharide **GH4-ABPs** were considerably less reactive. It is interesting that the effect of acetylation upon protein labeling is reversed between **GH2-ABPs** and **GH3-** and **GH4-ABPs**. **GH3-** and **GH4-ABPs** require enzymatic cleavage of the glycosidic bond to release the reactive quinone methide for enzyme labeling. Acetylation likely disrupts specific binding events required for glycosidic cleavage by the enzyme's catalytic amino acid residues. However, the **GH2-ABPs** only require the presence of a nucleophilic amino acid residue to displace the halogen from the reactive group for covalent protein binding. Alternatively, the difference may be due to the bulky reactive group of **GH3-** and **GH4-ABPs**; combination with the added steric bulk from the acetyl moieties may preclude active-site inclusion; whereas, the smaller reactive groups of **GH2-ABPs** may mitigate this effect. Intriguingly, the labeling profiles for the most reactive probes, **GH2b/d-ABPs** and **GH4a-ABP**, are nearly identical. This suggests that the protein targets are identical, and that acetylation facilitates **GH2-ABP** labeling, but contributes to an overall steric bulk detrimental to **GH3-** and **GH4-ABP** enzyme labeling.

Probe Reactivity and LC-MS/MS Analysis in *C. thermocellum*

To quantitatively assess the differences in **GH-ABP** labeling profiles and selectivity with higher sensitivity and resolution, (i) the cellulosomal secretomes were treated with biotin-azide or biotin-alkyne under click chemistry conditions, (ii) ABP-labeled proteins were enriched on streptavidin agarose resin, (iii) digested with trypsin, and (iv) analyzed by LC-MS/MS. Measured spectral counts were averaged across replicates for all peptides unique to a single protein to determine protein abundance. At least two unique peptides were measured per protein, and the average of spectral counts for that protein was required to be 2-fold the average of the spectral counts for that same protein in the no probe control samples. Data passing these filter criteria were used to compare probe reactivity and selectivity (Table 1; see also Supporting Information Figure S2).

Probe reactivity was varied considerably, particularly between *N*-bromo- or -iodoacetylglycosylamines (**GH2-ABPs**) and the activated difluoromethylphenyl aglycone (**GH3-** and **GH4-ABPs**). As with the gel results, no statistically significant probe labeling was identified with **GH1-ABP**, and therefore no data is shown in Table 1. Correlating with gel-based fluorescent analyses, **GH2-ABPs** have increased reactivity when acetylated (**2b**, **2d**). This difference is most pronounced between **GH2c-** and **GH2d-ABPs**, with nearly 70% reduction in peptide spectral counts from the acetylglycoside to the free glycoside ABPs. The reduction between **GH2a-** and **GH2b-ABPs** is subtler, only showing a 25% difference in measured peptides, and 20% difference in proteins (Table 1). Differences in the size of the halogen displaced by the enzymatic reaction resulted in an ~50% difference in reactivity between brominated glycoside **GH2a-ABP** and iodinated glycoside **GH2c-ABP**. Acetylated glycoside ABPs **GH2b** and **GH2d** do not follow this trend, rather they are similarly reactive. The halogen's strong impact on free glycoside probe labeling, versus no impact on acetylglycoside probe labeling, may be attributed to two potential factors: (i) acetylation drives active-site alignment of the probe, regardless of the halogen, and/or (ii) the higher reactivity of iodinated **GH-ABPs**, particularly toward hydrolysis, is limited by the presence of acetyl groups that confer greater hydrophobicity to the probe.⁴⁰

The activated difluoromethylphenyl aglycones (**GH3-** and **GH4-ABPs**) require mechanism-based glycosidic cleavage to release a reactive quinone methide, subsequently forming a covalent bond to a reactive amino acid residue within the active site. Alternatively, the reactive quinone methide may transiently diffuse and label a neighboring protein. The diffusion of the quinone methide is a concern, as non-ABP targets are subsequently labeled. The distribution of total spectral counts (Table 1) reveals that **GH3a-** and **GH3b-ABPs** are most reactive. High correlation was found between protein observations and the fluorescent gel data in Figure 1B. As observed, acetylation (**GH3b** and **GH4b**) decreases probe reactivity, opposite of the halogenated acetylglycosylamine probes. However, the size of the probe severely impacts reactivity, such that the disaccharide **GH4-ABPs** are >10-fold less reactive than the monosaccharide **GH3-ABPs**. The larger size restricts access to the active site and limits access to only the enzymes with activity toward cellobiose or cellulose (polysaccharides), reducing the concentration of quinone methide generated.

To validate that ABP labeling is selective, rather than driven solely by labeling of highly abundant proteins, ABP labeling was compared directly to the global MS measurement of the *C. thermocellum* cellulosomal secretome (Table 1). The global MS resulted in the identification of 499 proteins (42,254 spectral counts); notably greater than labeling by any **GH-ABP**.

Probe Selectivity for Cellulosome and Secreted Carbohydrate Active Enzymes

To determine selectivity of **GH-ABPs** for carbohydrate active enzymes, we evaluated the spectral count distribution of total spectral counts from global proteomics measurements and ABP-labeled spectral counts (Figure 3A). Selectivity for carbohydrate active enzymes, including GHs, glycosyl transferases, and carbohydrate kinases (see Supporting Information for a complete list of proteins identified in **GH-ABP** and global measurements), was determined by calculating the percentage of carbohydrate active peptide spectral counts versus the total measured peptide spectral counts. The probes with the most selectivity for carbohydrate-active enzymes were predominantly found in the **GH2c-ABP** (84%) and **GH4b-ABP** (71%). The less reactive **GH-ABPs** were far more selective than the highly reactive **GH3-ABPs** (**3a** – 33%; **3b** – 38%). All other **GH-ABPs** are ~60% selective for carbohydrate-active enzymes; even the least selective probes are >2-fold more selective than the distribution revealed by the global MS (only 17% carbohydrate active) measurements.

We also evaluated the selectivity of **GH-ABPs** for GH enzymes, including: cellulases, xylanases, and hemicellulases. The GH enzymes represent our primary interest for our probe suite, and given the reactive nature of these probes we anticipated cross-reactivity, primarily with cellulosome components. Coverage of GH enzymes as a percentage of total peptides reveals the same trends as carbohydrate-active enzymes, with less reactive probes more selective for GH enzymes (e.g., **GH2c-ABP** – 56%). **GH3-ABPs** were least selective, ~20%. However, only 5% of the total peptides measured in the global analyses are for GHs, revealing a 4- to 10-fold increase in selectivity for GHs using ABPs. Next we evaluated the percent peptide spectral counts for proteins containing a type I dockerin (Figure 3A), which is indicative of a role in the cellulosome. The percentiles mirror the GH distribution, however the more reactive probes (**GH2d-**, **3a-**, **3b-ABPs**) also labeled type I dockerin containing carbohydrate esterases and a serpin protease. When compared to the global data, **GH-ABPs** are 4- to 14-fold more selective for type I dockerin containing proteins.

Lastly, we determined the distribution of type I dockerin containing GHs as a percentage of all measured GHs. The less reactive probes (**2a-**, **2b-**, **2c-**, **4a-**, and **4b-ABPs**) labeled almost exclusively (~100%) cellulosome type I dockerin containing GHs. The more reactive probes ranged from ~85–90%; the remaining 10–15% of labeling is of secreted GHs, known to

closely interact with cellulosome components. Approximately three-quarters of all spectral counts for GHs in the global data contain the type I dockerin module. All-in-all, the data in Figure 3B reveals selectivity for cellulosome components, including GH enzymes, with a significant distribution of peptide spectral counts assigned to these categories when compared to the global distribution.

Characterization of Glycoside Hydrolase Enzymes Identified in GH-ABP and Global MS Measurements

Activated difluoromethylphenyl aglycones and *N*-bromo- and *N*-iodoacetyl glycosylamines are both able to react with glycoside hydrolases showing inverting- or retaining-type mechanisms.^{13,23b} This dual specificity is an advantage for high-throughput characterization of the GH complement of a proteome. Within the *C. thermocellum* cellulosomal secretome, GH enzymes are prevalent in both the cellulosome complexes, and as freely secreted enzymes. All **GH-ABPs** showed an approximate 60:40 inverting:retaining labeling distribution (Figure 4). This is the result of a large array of GH9 (inverting) cellulase enzymes, and a number of GH5 (retaining) endoglucanases, and GH10 and GH11 (retaining) xylanases (Figure 3B and Supporting Information). The global MS measurement revealed a near 50:50 inverting: retaining distribution, due to the identification of a large number of secreted retaining GHs. The lack of labeling of these secreted and retaining GHs by any **GH-ABP** likely indicates they were in an inhibited or zymogen (proenzyme) non-functional state.

A closer inspection of the GH enzymes labeled reveals broad coverage of cellulose degrading GH families (Figure 3B). We determined the total average peptide spectral counts for only GH family enzymes in the **GH-ABP** labeled and global MS data. We then normalized the contribution of each individual GH family to the total (Figure 3B). GH9, endoglucanases (EC 3.2.1.4) that catalyze the hydrolysis of (1→4)-β-D-glucosidic linkages in cellulose, is the dominant GH family represented. All 12 GH9 enzymes expressed in *C. thermocellum* were labeled by **GH-ABPs** (e.g., **GH3a-ABP**; see Supporting Information for all GHs labeled by individual probes). GH9 enzymes are unique in that they display a processive mechanism, in which they are able to cleave internal regions of cellulose, and continue to degrade processively down the cleaved chain. GH family 5, which catalyzes the endohydrolysis of (1→4)-β-D-glucosidic linkages in cellulose (EC 3.2.1.4), and hydrolysis of (1→4)-β-D-glucosidic linkages in cellulose and cellotetraose to release cellobiose from the non-reducing ends of the polymeric chain (EC 3.2.1.91), were the second most prevalent GH family labeled for all ABPs, with labeling of 5 of the 9 proteins encoded in the genome. It has been demonstrated previously that GH5 and GH9 enzymes make up half the catalytic components of the cellulosome, and abundance of GH5 enzymes increase when the carbon source for growth is crystalline cellulose.⁴¹

The order of labeling followed with GH94, cellobiose phosphorylases (EC 2.4.1.20), and GH families 10 and 11, xylanases that catalyze the endohydrolysis of (1→4)-β-D-xylosidic linkages in xylans (EC 3.2.1.8). It is compelling to note that *C. thermocellum* cannot grow on xylan as a carbon source, but contains a suite of xylanases (hemicellulases). Other families labeled were GH48 processive cellulases (EC 3.2.1.4), GH74 xyloglucanases (EC 3.2.1.151), and GH26 mannanases. An additional six GH families were identified, but with less than 5% peptide spectral count distribution when compared to the total GH peptide spectra measured (Figure 3B).

The distribution of GH family labeling was near identical across the suite of **GH-ABPs**. This distribution was different from the global data, which had less measurement of the primary lignocellulose degrading enzyme families (GH5, 9, and 10). Differences in probe reactivity are particularly evident within the disaccharide **GH4-ABPs**; labeling of GH

families 9 and 10 are dominant with these probes. Acetylation (**GH4b-ABP**) severely limited the labeling of cellobiose GH family 5 when compared to the free glycoside (**GH4a-ABP**), but this trend is only within **GH4-ABPs**. Labeling of GH94 cellobiose phosphorylases is clearly reactivity driven, with **GH2d-**, **GH3a-**, and **GH3b-ABPs** having the largest peptide distribution within this family. **Profiling the *C. thermocellum* Cellulosome.** *C. thermocellum* is a remarkable anaerobic cellulosome producing organism, with a specific activity against crystalline cellulose that is 50-fold higher than that of *Trichoderma reesei*, a commercially employed fungal strain for biofuel production.² The cellulosome macromolecular structure is organized by the association of cell wall proteins that bind extracellular components, including GH enzymes, carbohydrate esterases, and carbohydrate binding modules (CBMs) (Figure 2). Binding between cell wall proteins and extracellular components is mediated by type I and II dockerin and cohesin modules; with high affinity protein-protein interactions ($>10^9 \text{ M}^{-1}$).⁴² An extracellular scaffoldin (CipA) protein contains nine type I cohesins to which nine enzymes bearing type I dockerin groups can bind, and a carbohydrate binding module for direct attachment of the entire cellulosome complex to the cellulosic substrate. Scaffoldin also contains a type II cohesin for direct attachment to cell surface anchoring proteins (SdbA, OlpB, and Orf2). OlpB and Orf2 contain multiple type II cohesin domains, permitting the assembly of polycellulosome components with as many as 63 enzymatic subunits. Direct adherence of enzymes to the cell surface is mediated by anchoring proteins OlpA and OlpC, which contain a single type I cohesin (Figure 2).

C. thermocellum produces 72 cellulosome enzymes (polypeptides with type I dockerins), although only approximately half have been identified when the organism is cultured on microcrystalline cellulose (avicel).⁴³ Prior global MS-based proteomic analysis identified 35 cellulosome type I dockerin-containing proteins when grown on avicel;⁴³ in a series of seven growth conditions 59 type I dockerin-containing proteins were found.⁴¹ Both prior experiments employed targeted MS analyses in which they only measured an enriched cellulosome. Our ABPP approach focused on characterizing the cellulosomal secretome and other secreted non-cellulosomal GH enzymes. We identified 35 cellulosome type I dockerin-containing enzymes, scaffoldin (CipA), and four cell anchoring proteins with our **GH-ABP** suite (Figure 5).

The type I dockerin-containing enzymes measured by **GH-ABPs** are primarily endoglucanases, exoglucanases, and xylanases. Minor selectivity was shown for accessory enzymes such as hemicellulases, mannanases, and carbohydrate esterases, though these are minor components as seen in the global data. Given this array of cellulosome functions, differences in coverage are seen between ABPs (Table 2). Several **GH-ABPs** had a peptide spectral count ratio of ~2:2:1 endoglucanase:exoglucanase:xylanase labeling; which is significantly different than the ~1:1:1 global MS distribution (Table 2). **GH4a-ABP**, a disaccharide, showed highest reactivity toward exoglucanases, but little xylanase coverage. The acetylated **GH4b-ABP** showed high selectivity for exoglucanases, with nearly no labeling of endoglucanases, and marginal coverage of xylanases.

Spectral counts measured for unique peptides assigned to CipA (scaffoldin) are concomitant with coverage for either endo- or exoglucanases for both global and **GH-ABP** data (Table 2). Scaffoldin does not have specific GH activity, but rather contains nine type I cohesin domains to bind nine type I dockerin-containing enzymes, and a specific CBM for cellulose. However, CipA has been shown to enhance the specific activity of cellulosome enzymes upon binding, which includes CelS, a major cellulosome exoglucanase (GH48).⁴⁴ Our probes labeled CipA at a higher level than the 9:1 ratio of enzymes:CipA. The strong enzyme- CipA complex likely mediated this. In the case of **GH2-ABPs** the tight enzyme-CipA interactions likely facilitated a close proximity between the probes and nucleophilic

amino acid residues on CipA. Alternatively, for **GH3-** and **GH4- ABPs**, a reactive quinone methide transiently diffused from the enzyme to label CipA. This is made evident as the less reactive **GH4-ABPs** labeled CipA far less than **GH3- ABPs**, but at a level concomitant with endo- and exoglucanase activities.

Significant differences were seen between global and **GH-ABP**-labeled cellulosome components, providing evidence of selectivity. Evaluation of a particularly interesting set of cellulosome proteins, the cellobiohydrolases (CelK, CbhA, CelO, and CelS), reveals global versus ABP commonalities and significant differences. CelK and CbhA are highly homologous cellobiohydrolases, encoding CBM4 domains, an Ig-like domain, a GH9, and a type I dockerin. A 50:50 abundance identified in the global MS data mirrors the activity-dependent ABP data. However, CelO, encoding a CBM3, a GH5, and a type I dockerin, was not in the global data, and only identified by **GH2b-** and **GH3a-ABPs**. CelS, with a CBM3, a GH48, and a type I dockerin, was the most abundant cellobiohydrolase in the global data, but was only half as reactive as CelK and CbhA when ABP-labeled (Figure 5 and Supporting Information). Two other highly abundant proteins in the global data were Cthe_0821, encoding a CBM32, GH35, and type I dockerin, and XynC, encoding a CBM22, GH10, and type I dockerin. All probes label both of these proteins less extensively. Additionally, three proteins were found only in the ABPP data: Cthe_0279, CelW, and CelO. Our data shows that selectivity is found for **GH-ABPs** when compared against the global analyses.

Tuning the **GH-ABPs** with different reactive groups, acetylation of the glycosides, and mono- versus disaccharides, resulted in disparities in reactivity toward cellulosomal components. The reactivity differences are complementary to those discussed earlier, with the difluoromethylphenyl aglycones being highly reactive. However, **GH-ABP** derivations do result in selectivity (Figure 5). Compared to monosaccharide **GH3-ABPs**, disaccharide **GH4-ABPs** are particularly selective toward exoglucanase cellobiohydrolases (CelK, CelS, and CbhA) and xylanases. Comparing the deacetylated **GH4a-ABP** to acetylated **GH4b-ABP** shows **GH4a** as selective for two endoglucanases (CelW, Cthe_0821), and **GH4b** as selective for CelS, which acts in a processive fashion starting from the reducing end of the cellulose chain.

A final examination of the cellulosome data reveals labeling of a number of proteins for which specific function is hypothetical or has not been identified (“undefined” in Figure 5), and enzymes that do not contain type I dockerin units (see Supporting Information for the specific targets of each **GH-ABP**). However, employing STRING (Search Tool for the Retrieval of Interacting Genes) to identify protein- protein interactions revealed that enzymes without type I dockerins are closely integrated with cellulosome components (see Supporting Information for STRING interactions).⁴⁵ Four glycoside hydrolases, BglA, LicA, Cthe_1787, and Cthe_2119, have high scores for interactions with cellulosomal cellulases and xylanases. Therefore, the ABPP approach applied to cellulose degrading systems can identify key proteins with undefined functions, and cellulosome and cellulosome-associated enzymes.

Together, our research reveals the dynamic assembly and reactivity of the *C. thermocellum* cellulosome. As we have identified by **GH-ABP** labeling, a diverse array of type I dockerin containing GH enzyme families and additional accessory proteins function in concert. The high rate of cellulose degradation by this anaerobic bacterium is clearly a function of the sum of its parts.

CONCLUSIONS

Here, we have developed and applied a suite of difluoromethylphenyl aglycone, 2-deoxy-2-fluoroglycoside, and *N*-halogenated glycosylamine **GH-ABPs** to the direct labeling of the *C. thermocellum* cellulosomal secretome. These **GH-ABPs** were synthesized with alkynes to harness the utility and multimodal possibilities of click chemistry, and to increase enzyme active site inclusion. Comparing **GHABP**-labeled to global MS data reveals probe selectivity for GH enzymes, including a large array included in the cellulosome. Derivatization of the **GH-ABPs**, including reactive groups, acetylation of the glycoside binding groups, and mono- and disaccharide binding groups, resulted in considerable variability in protein labeling. Our probe suite is applicable to aerobic and anaerobic cellulose degrading systems, and facilitates greater understanding of roles for organisms in biofuel development. Future work will include the optimization of the **GH-ABPs** to increase selectivity, and application of the probes to aerobic bacteria and fungal cellulose degrading organisms. Clearly, the combination of **GH-ABPs** and MS-based proteomics provides a high-throughput avenue to characterizing GH enzymes in complex lignocellulose degrading systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Biological Separations & Mass Spectrometry group for helpful discussions and critical reading of the document. This work was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. DOE under contract DE-AC05-76RL01830. This work used instrumentation and capabilities developed under support from the National Institutes of Health (NIH) National Center for Research Resources (5P41RR018522) and the National Institute of General Medical Sciences (8P41GM103493-10), and the U.S. DOE Office of Biological and Environmental Research (DOE-BER). Mass spectrometry-based proteomic measurements were performed in the Environmental Molecular Sciences Laboratory, a DOE-BER national scientific user facility at PNNL.

ABBREVIATIONS

ABP	Activity-based probe
ABPP	Activity-based protein profiling
AGC	automatic gain control
CBM	carbohydrate binding module
CC	click chemistry
EC	enzyme classification
i.d	internal diameter
GH	glycoside hydrolase
LC	liquid chromatography
MS	mass spectrometry
o.d	outer diameter
SC	spectral count
STRING	Search Tool for the Retrieval of Interacting Genes

References

1. (a) Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA. *Biofuels*. 2010; 1:323–341.(b) Ohmiya K, Sakka K, Kimura T, Morimoto K. *J Biosci Bioeng*. 2003; 95:549–561. [PubMed: 16233456] (c) Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. *Microbiol Mol Biol Rev*. 2002; 66:506–577. [PubMed: 12209002]
2. Demain AL, Newcomb M, Wu JHD. *Microbiol Mol Biol Rev*. 2005; 69:124–154. [PubMed: 15755956]
3. McBee RH. *J Bacteriol*. 1954; 67:505–506. [PubMed: 13152068]
4. Lamed R, Zeikus JG. *J Bacteriol*. 1980; 144:569–578. [PubMed: 7430065]
5. (a) Fontes CM, Gilbert HJ. *Annu Rev Biochem*. 2010; 79:655–681. [PubMed: 20373916] (b) Doi RH, Kosugi A. *Nature Rev Microbiol*. 2004; 2:541–551. [PubMed: 15197390] (c) Bayer EA, Belaich JP, Shoham Y, Lamed R. 2004; 58:521–554.(d) Lamed R, Setter E, Kenig R, Bayer EA. *Biotechnol Bioeng Symp*. 1983; 13:163–181.
6. (a) Poole DM, Morag E, Lamed R, Bayer EA, Hazlewood GP, Gilbert HJ. *FEMS Microbiol Lett*. 1992; 78:181–186. [PubMed: 1490597] (b) Wu JHD, Orme-Johnson WH, Demain AL. *Biochemistry*. 1988; 27:1703–1709.
7. Kruus K, Lua AL, Demain AL, Wu JHD. *Proc Natl Acad Sci USA*. 1995; 92:9254–9258. [PubMed: 7568112]
8. Rye CS, Withers SG. *Curr Opin Chem Biol*. 2000; 4:573–580. [PubMed: 11006547]
9. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. *Nucl Acids Res*. 2009; 37:D233–D238. [PubMed: 18838391]
10. For concise reviews: Li N, Overkleeft HS, Florea BL. *Curr Opin Chem Biol*. 2012; 16:227–233. [PubMed: 22325363] Cravatt BF, Wright AT, Kozaarich JW. *Annu Rev Biochem*. 2008; 77:383–414. [PubMed: 18366325]
11. For concise reviews: Witte MD, van der Marel GA, Aerts JMFG, Overkleeft HS. *Org Biomol Chem*. 2011; 9:5908–5926. [PubMed: 21769371] Stubbs KA, Vocadlo DJ. *Aust J Chem*. 2009; 62:521–527.
12. Tsai CS, Li YK, Lo LC. *Org Lett*. 2002; 4:3607–3610. [PubMed: 12375899]
13. Kuroguchi M, Nishimura S, Lee YC. *J Biol Chem*. 2004; 43:5338–5342.
14. Romaniouk AV, Silva A, Feng J, Vijay IK. *Glycobiology*. 2004; 14:301–310. [PubMed: 14718372]
15. Vocadlo DJ, Bertozzi CR. *Angew Chem Int Ed*. 2004; 43:5338–5342.
16. Hinou H, Kuroguchi M, Shimizu H, Nishimura S. *Biochemistry*. 2005; 44:11669–11675. [PubMed: 16128567]
17. Williams SJ, Hekmat O, Withers SG. *ChemBioChem*. 2006; 7:116–124. [PubMed: 16397879]
18. Hekmat O, Florizone C, Kim Y-W, Eltis LD, Warren RAJ, Withers SG. *ChemBioChem*. 2007; 8:2125–2132. [PubMed: 17952884]
19. Stubbs KA, Scaffidi A, Debowski AW, Mark BL, Stick RV, Vocadlo DJ. *J Am Chem Soc*. 2008; 130:327–335. [PubMed: 18067297]
20. Witte MD, Kallemeijn WW, Aten J, Li KY, Strijland A, Donker-Koopman WE, van den Nieuwendijk AM, Bleijlevens B, Kramer G, Florea BL, Hooibrink B, Hollak CE, Ottenhoff R, Boot RG, van der Marel GA, Overkleeft HS, Aerts JM. *Nat Chem Biol*. 2010; 6:907–913. [PubMed: 21079602]
21. Witte MD, Walvoort MTC, Li K-Y, Kallemeijn WW, Donker-Koopman WE, Boot RG, Aerts JMFG, Codee JDC, van der Marel GA, Overkleeft HS. *ChemBioChem*. 2011; 12:1263–1269. [PubMed: 21538758]
22. Gandy MN, Debowski AW, Stubbs KA. *Chem Commun*. 2011; 47:5037–5039.
23. (a) Walvoort MTC, Kallemeijn WW, Willems LI, Witte MD, Aerts JMFG, van der Marel GA, Codee JDC, Overkleeft HS. *Chem Comm*. 2012; 48:10386–10388. [PubMed: 22932951] (b) Hinou H, Nishimura S-I. *Curr Top Med Chem*. 2009; 9:106–116. [PubMed: 19199999] (c) Rempel BP, Withers SG. *Glycobiology*. 2008; 18:570–586. [PubMed: 18499865]
24. Speers AE, Cravatt BF. *Chem Biol*. 2004; 11:535–546. [PubMed: 15123248]

25. Weimer P, Odt CL. *Enzymatic Degradation of Insoluble Carbohydrates*. 1996; 618:291–304.
26. Maiolica A, Borsotti D, Rappsilber J. *Proteomics*. 2005; 5:3847–3850. [PubMed: 16130174]
27. Lipton MS, Pasa-Tolic L, Gordon AA, Anderson DJ, Auberry DL, Battista JR, Daly MJ, Fredrickson J, Hixson KK, Kostandarithes H, Masselon C, Markillie LM, Moore RJ, Romine MF, Shen Y, Stritmatter E, Tolic N, Udseth HR, Venkateswaran A, Wong KK, Zhao R, Smith RD. *Proc Natl Acad Sci USA*. 2002; 99:11049–11054. [PubMed: 12177431]
28. Livesay EA, Tang K, Taylor BK, Buschbach MA, Hopkins DF, LaMarche BL, Zhao R, Shen Y, Orton DJ, Moore RJ, Kelly RT, Udseth HR, Smith RD. *Anal Chem*. 2008; 80:294–302. [PubMed: 18044960]
29. Chundawat SPS, Lipton MS, Purvine SO, Uppugundla N, Gao D, Balan V, Dale BE. *J Proteome Res*. 2011; 10:4365–4372. [PubMed: 21678892]
30. Mayampurath AM, Jaitly N, Purvine SO, Monroe ME, Auberry KJ, Adkins JN, Smith RD. *Bioinformatics*. 2008; 24:1021–1023. [PubMed: 18304935]
31. Eng J, McCormack A, Yates J. *J Am Soc Mass Spectrom*. 1994; 5:976–989.
32. Kim S, Gupta N, Pevzner PA. *J Proteome Res*. 2008; 7:3354–3363. [PubMed: 18597511]
33. Elias JE, Gygi SP. *Nat Methods*. 2007; 4:207–214. [PubMed: 17327847]
34. Withers SG, Street IP, Bird P, Dolphin DH. *J Am Chem Soc*. 1987; 109:7530–7531.
35. Street IP, Kempton JB. *Biochemistry*. 1992; 31:9970–9978. [PubMed: 1390781]
36. (a) Black TS, Kiss L, Tull D, Withers SG. *Carbohydr Res*. 1993; 250:195–202. [PubMed: 8143291] (b) Tull D, Burgoyne DL, Chow DT, Withers SG, Aebersold R. *Anal Chem*. 1996; 234:119–125. (c) Chir J, Withers SG, Wan CF, Li YK. *Biochem J*. 2002; 365:857–863. [PubMed: 11978178] (d) Kiss T, Erdei A, Kiss L. *Arch Biochem Biophys*. 2002; 399:188–194. [PubMed: 11888205] (e) Vocadlo DJ, Wicki J, Rupitz K, Withers SG. *Biochemistry*. 2002; 41:9736–9746. [PubMed: 12146939] (f) Jager S, Kiss L. *World J Microbial Biotechnol*. 2005; 21:337–343.
37. Yan Q, Cao R, Yi W, Liang Y, Chen Z, Ma L, Song H. *Bioorg Med Chem Lett*. 2009; 19:4055–4058. [PubMed: 19564107]
38. Likhoshervostov LM, Novikova OS, Zheltova AO, Shibaev VN. *Russ Chem Bull, Int Ed*. 2004; 53:709–713.
39. (a) Janda KD, Lo L-C, Lo C-HL, Sim M-M, Wang R, Wong CH, Lerner RA. *Science*. 1997; 275:945–948. [PubMed: 9020070] (b) Ichikawa M, Ichikawa Y. *Bioorg Med Chem Lett*. 2001; 11:1769–1773. [PubMed: 11425557] (c) Lu CP, Ren C-T, Lai Y-N, Wu S-H, Wang W-M, Chen J-Y, Lo L-C. *Angew Chem Int Ed*. 2005; 44:6888–6892. (d) Kwan DH, Chen H-M, Ratananikom K, Hancock SM, Watanabe Y, Kongsaree PT, Samuels AL, Withers SG. *Angew Chem Int Ed*. 2011; 50:300–303.
40. Sarkar AK, Fritz TA, Taylor WH, Esko JD. *Proc Natl Acad Sci USA*. 1995; 92:3323–3327. [PubMed: 7724561]
41. Raman B, Pan C, Hurst GB, Rodriguez M, McKeown CK, Lankford PK, Samatova NF, Mielenz JR. *PLoS One*. 2009; 4:e5271. [PubMed: 19384422]
42. Schaeffer F, Matuschek M, Guglielmi G, Miras I, Alzari PM, Beguin P. *Biochemistry*. 2002; 41:2106–2114. [PubMed: 11841200]
43. Gold ND, Martin VJJ. *J Bacteriol*. 2007; 189:6787–6795. [PubMed: 17644599]
44. Kruus K, Lua AC, Demain AL, Wu JHD. *Proc Natl Acad Sci USA*. 1995; 92:9254–9258. [PubMed: 7568112]
45. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M, Muller J, Bork P, Jensen LJ, von Mering C. *Nucleic Acids Res*. 2011; 39:D561–D568. [PubMed: 21045058]

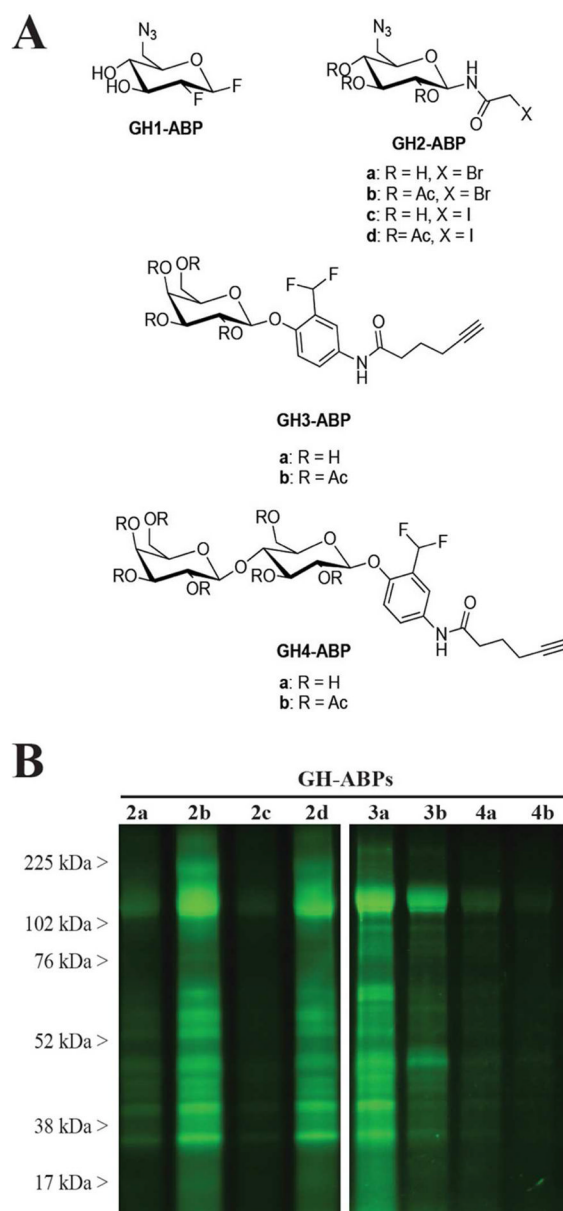
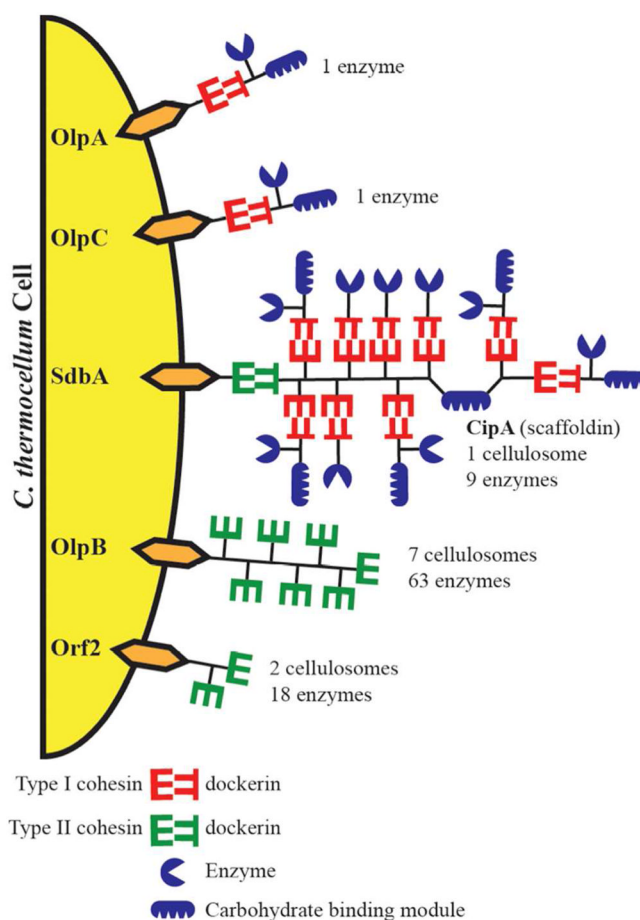


Figure 1. (A) Structures of glycoside hydrolase-directed activity-based probes (**GH-ABPs**). (B) *C. thermocellum* secretome samples, containing cellulosome proteins, were incubated at 1 mg/mL with individual **GH-ABPs** (75 μ M). Click chemistry was used to append the fluorophores rhodamine-azide and -alkyne, and proteins were separated by SDS-PAGE and imaged to reveal the fluorescent ABP-labeled proteins. Control gels and protein abundance stains are shown in the Supporting Information.

**Figure 2.**

Organization of the *Clostridium thermocellum* cellulosome structure. The scaffoldin protein (CipA) contains nine type I cohesins to which nine enzymes bearing type I dockerin groups can bind, and a carbohydrate binding module for direct attachment of the entire cellulosome complex to the cellulosic substrate. The scaffoldin also contains a type II cohesin for direct attachment to cell surface anchoring proteins (SdbA, OlpB, and Orf2). OlpB and Orf2 contain multiple type II cohesin domains, permitting the assembly of polycellulosome components with as many as 63 enzymatic subunits. Direct adherence of enzymes to the cell surface is mediated by anchoring proteins OlpA and OlpC, which contain a single type I cohesin.

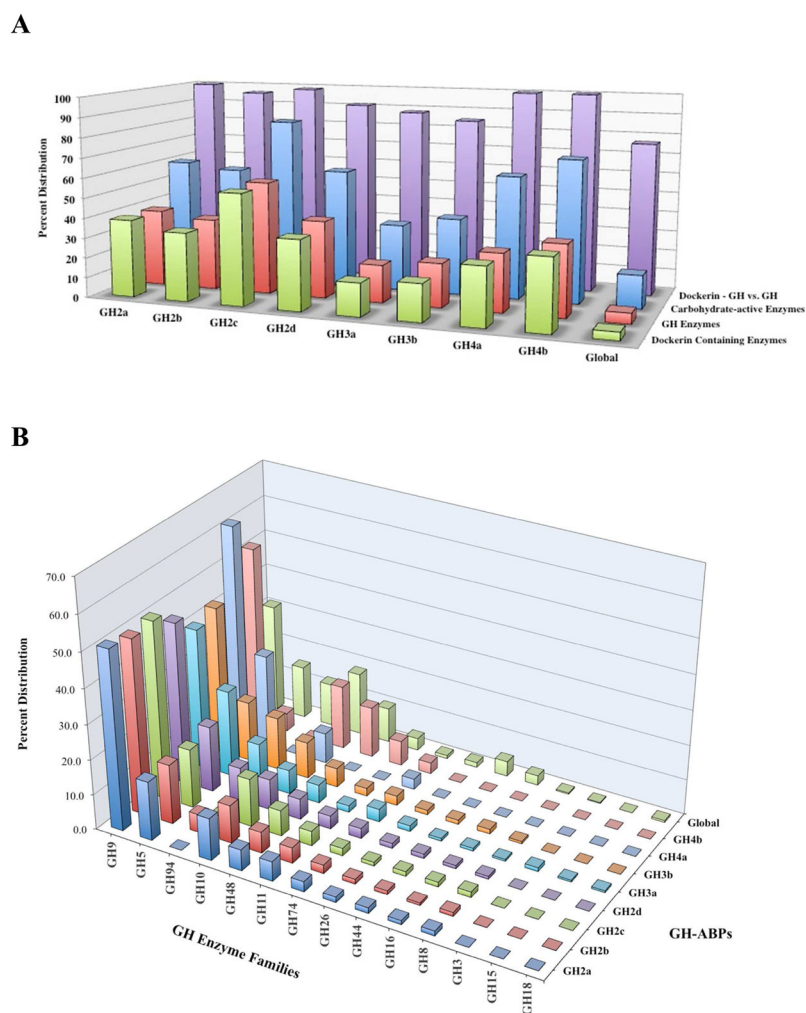


Figure 3. (A) Percentage of total spectral counts for **GH-ABP**-labeled and global samples attributed to: (i) Dockerin Containing Enzymes – any enzyme containing a type I dockerin module involved in cellulosome interaction, (ii) GH Enzymes – glycoside hydrolase enzymes with and without a type I dockerin module, (iii) Carbohydrate-active Enzymes – all enzymes with and without type I dockerin modules that are active toward glycosides, including GHs, glycosyl transferases, and carbohydrate kinases. The final group, (iv) Dockerin-GH vs. GH – represents the distribution of spectral counts attributed to type I dockerin containing GH enzymes as a percentage of all measured GH enzyme spectral counts. (B) Percent distribution of total spectral counts for **GH-ABP** labeled and global samples attributed to glycoside hydrolase (GH) enzymes. Peptide spectral counts were averaged for each GH family; GH families were summed and distributed according to each **GH-ABP**. Percent distribution was calculated by dividing the average peptide spectral counts for a GH family by the total spectral counts measured for all GH families, for a particular **GH-ABP**. See Supporting Information for raw data. All measurements represent three sample replicates per **GH-ABP**.

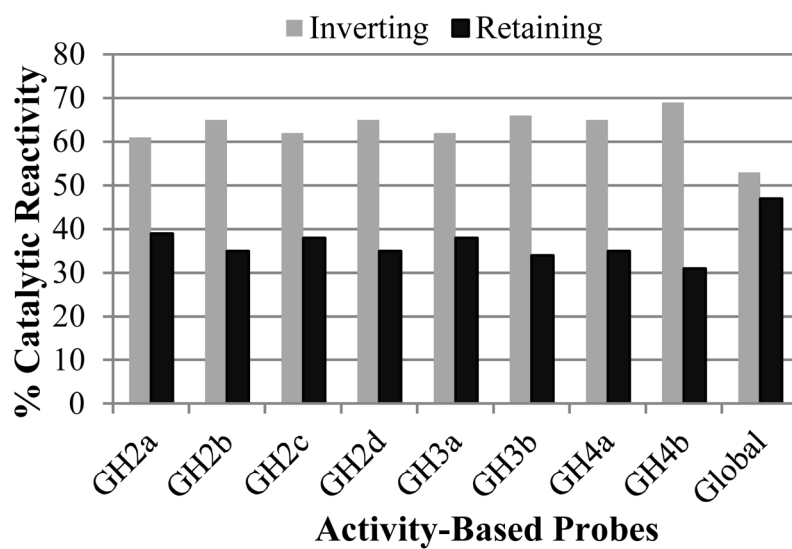


Figure 4. Percent distribution of catalytic reactivity of GH enzymes resulting in an overall inversion or retention of the glycoside anomeric configuration according to each **GH-ABP**. Three replicates per probe.

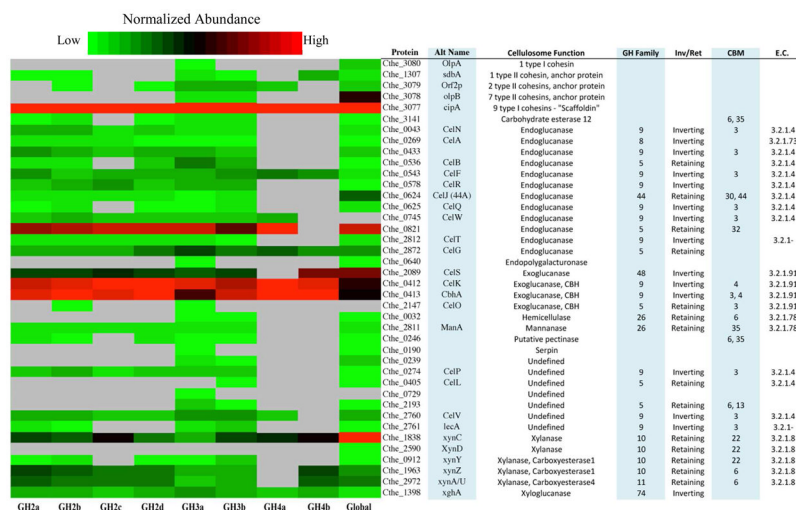
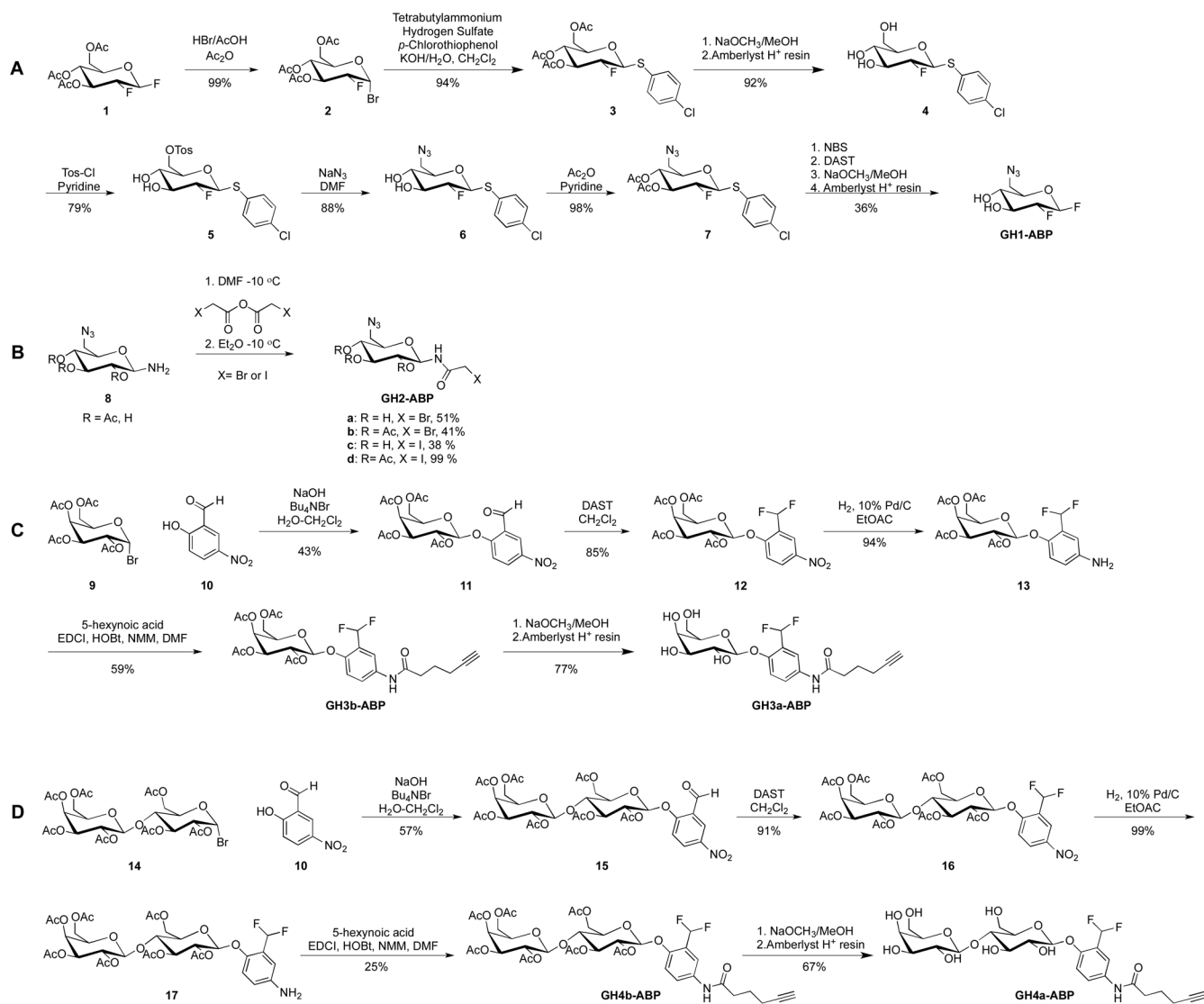


Figure 5. Activity-based probe and global profiling of the *C. thermocellum* cellulosome. Components include scaffoldin (CipA), cell wall anchoring proteins (OlpA, OlpB, SdbA, Orf2p), and type I dockerin-containing enzymes. Heatmap data is normalized within each column (i.e., each **GH-ABP** or global data), and represents the average spectral counts for each protein normalized on a 0–100% scale. The associated table shows protein IDs and names (if previously assigned), functions within the cellulosome, GH families, inverting or retaining catalytic mechanisms, carbohydrate binding modules (CBM), and EC number. See Supporting Information for associated raw data.



Scheme 1.
GH-ABP Syntheses.

Table 1

Spectral count (SC) and protein count data for labeling by each individual **GH-ABP** and for the global analysis of the *C. thermocellum* cellulosomal secretome.

Data Category	Activity-Based Probe										
	GH2a	GH2b	GH2c	GH2d	GH3a	GH3b	GH4a	GH4b	Global		
Total Protein Count	74	91	38	84	219	177	41	23	499		
Total Peptide SC	783	1044	338	1061	4035	2872	345	200	42254		
Carbohydrate Active Protein Count	36	41	24	39	60	52	17	12	84		
Carbohydrate Active SC ^A	471	599	284	630	1344	1101	211	142	6997		
Dockerin Containing Protein Count	27	27	20	25	33	30	9	9	41		
Dockerin Containing SC ^B	303	358	188	374	667	534	100	71	1705		
GH Family Protein Count	25	27	21	27	33	31	9	9	52		
GH Family SC ^C	301	373	191	410	757	630	100	71	2233		

^A Includes all enzyme families that act upon glycoside substrates, including glycoside hydrolases, glycosyl transferases, and carbohydrate kinases.

^B Includes all enzymes that contain a type I dockerin module for binding within the cellulosome macromolecular structure.

^C Includes all glycoside hydrolase enzymes, including those that are secreted but do not contain a type I dockerin module.

Table 2

Spectral counts for groups of cellulosome functions for each **GH-ABP** and global data.[#]

Cellulosome Function	Activity-Based Probe											Global	
	GH2a	GH2b	GH2c	GH2d	GH3a	GH3b	GH3c	GH3d	GH4a	GH4b	GH4c		
Carbohydrate Esterase	3	5	2	2	7	6							11
Endoglucanase	108	122	62	141	274	188			37	7			557
Exoglucanase	105	135	80	140	211	187			50	44			524
Hemicellulase					2	2							5
Mannanase	5	5	2	6	10	6							21
Pectinase	3	3		4	4	2							2
Serpin					6								8
Xylanase/Xyloglucanase	62	66	38	63	86	94			12	20			452
Scaffoldin - Cell Anchoring	79	101	61	106	221	174			57	44			543
Undefined	18	24	7	17	66	50			2				125

[#] Spectral counts represent the sum of the average peptide SCs for all proteins within a category for a given ABP or the global data. See the Supporting Information for raw data for individual proteins.