

Quantitative Mass Spectrometric Characterization of Substrate-Dependent Changes in the Cellulosome of *Clostridium thermocellum*

Gregory B. Hurst, Chongle Pan, Patricia K. Lankford, Babu Raman, Miguel Rodriguez Jr., Catherine K. McKeown, Steven D. Brown, Nagiza F. Samatova, Jonathan R. Mielenz

Oak Ridge National Laboratory, Oak Ridge TN



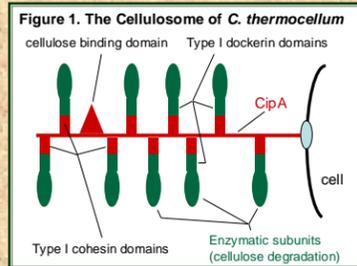
OVERVIEW

Cellulosome proteins expressed by the cellulose-degrading bacterium *Clostridium thermocellum* have been profiled on various carbohydrate substrates.

INTRODUCTION

Clostridium thermocellum is a bacterium with potential in the biofuel industry, as it can degrade cellulose and ferment the resulting sugars to ethanol. Cellulose degradation is aided by a large, extracellular complex of hydrolytic enzymes known as the cellulosome (Figure 1). This complex contains a backbone scaffoldin protein, CipA, which is attached to an anchor protein on the outer membrane of the cell. Various catalytic components for carbohydrate degradation contain dockerin domains that attach to cohesin domains on the CipA scaffoldin [1].

In this study, we investigated the changes in expression levels of the cellulosomal components during growth on a variety of carbohydrate substrates. Differential metabolic labeling of *C. thermocellum* cultures allowed the application of isotope ratio mass spectrometry for characterizing the substrate-dependent changes in the cellulosomal composition of *C. thermocellum*. This study expands the range of substrates on which the cellulosome composition has been characterized by mass spectrometry [2].



EXPERIMENTAL

Bacterial Growth: *C. thermocellum* fermentations were grown in duplicate in minimal medium with 5g/L total carbon, without any yeast extract, on six different carbon sources:

- Avicel (A) (crystalline cellulose)
- Cellobiose (Cb) (a disaccharide)
- Avicel-Pectin (3:2) (PA) (Pectin is a plant-produced polysaccharide)
- Avicel-Xylan (3:2) (AX) (Xylan is a complex plant-produced polysaccharide)
- Avicel-Pectin-Xylan (3:1:1) (PAX)
- Z-Trim (ZT) (A fat substitute prepared from plant fibers*)

* Additional fermentations were grown for use as quantitative protein standards on Avicel in a medium containing only ¹⁵N nitrogen compounds.

Cellulosome Purification: Cellulosomes were isolated from the supernatant at ~29hrs using the affinity digestion method [3]. Samples grown in ¹⁴N medium were mixed in 1:1 ratio with the ¹⁵N-labeled Avicel sample.

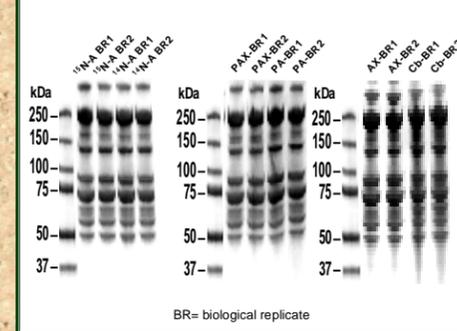
Sample Preparation: Samples were denatured, reduced, digested with trypsin, and desalted (C18 SepPak, Waters).

Mass Spectrometry: 2D HPLC separations were performed with six 2D cycles, consisting of a strong cation exchange salt step gradient followed by a reverse-phase linear gradient, for each sample [4]. Duplicate LC-MS-MS runs were performed for each fermentation. The eluent was introduced into a ThermoFinnigan LTQ mass spectrometer via nanospray. Tandem mass spectra were analyzed by SEQUEST [5] and identified peptides were filtered using DTASelect [6].

Data Analysis: ¹⁴N/¹⁵N ratios and confidence intervals for proteins were determined using ProRata software [7].

* <http://www.ars.usda.gov/is/pr/1996/z-trim896.htm>

Figure 2. SDS-PAGE of cellulosomes isolated from duplicate *C. thermocellum* fermentations on various carbon sources.



RESULTS AND DISCUSSION

Isolated cellulosomes showed a relatively simple pattern of discrete bands in gel electrophoresis, with slight differences among the various carbohydrate substrates (Figure 2).

A number of proteins predicted to contain domains characteristic of the cellulosome (carbohydrate binding module—CBM, dockerins, and cohesins) were identified in the isolated cellulosome samples, and amounts of these proteins relative to growth on Avicel were measured by quantitative mass spectrometry (Table 1).

Increases > 2-fold (log2 > 1) are highlighted in red, while decreases > 2-fold (log2 < -1) are highlighted in green.

Key to other entries in Table 1:

- u, U: protein not consistently quantified in both biological replicates (BR), but result indicates up-regulation (lower Cb=0, log2ratio >=0.5)
- n, N: protein not consistently quantified in both BR, but result indicates up-regulation (-0.5 <= log2ratio <=0.5)
- i, I: protein identified only

Normalized Spectral Abundance Factors (NSAF) [8] were also calculated for ¹⁴N proteins (Table 2). A larger NSAF value indicate that a protein represents a larger fraction of the spectrum counts (corrected for protein length) summed across all proteins in a sample. Brighter green cells in Table 2 indicate proteins with higher NSAF:

Exoglucanase CelS represents the largest fraction of normalized spectral abundance (values > 5 for table)

Other proteins with notable NSAF include:

- Scaffoldin CipA and anchor protein SdbA
- Endoglucanases CelB, CelE, CelA, CelF, CelR, CelQ and CthE0821
- Xylanases XynC, XynA/U
- Serpin CthE0190

Table 1: log2 ratio (with 90% confidence interval[7]) of amounts of protein measured for various carbohydrate substrates, relative to growth on Avicel (crystalline cellulose)									
Locus	Gene	Structural, Catalytic and/or Binding Modules	Avicel	Avicel+Xylan	Cellobiose	Pectin+Avicel	Pectin+Avicel +Xylan	Z-Trim	Protein expression reported on Avicel vs. Cellobiose [2]
CthE3077	CipA	9 type I cohesins - Scaffoldin	-0.1 (-0.2, 0.0)	+0.1 (0.0, +0.2)	+0.7 (+0.6, +0.8)	+0.1 (0.0, +0.2)	0.0 (-0.1, +0.1)	+0.4 (+0.3, +0.5)	Increase
CthE452		1 type I cohesin	u	-0.5 (-0.7, -0.2)	+1.8 (+1.8, +2.0)	0.0 (-0.2, +0.2)	-0.1 (-0.3, +0.2)	+0.2 (0.0, +0.5)	
CthE1307	SdbA	1 type II cohesin, anchor protein	+0.8 (+0.6, +0.9)	-0.1 (-0.2, +0.1)	+3.2 (+3.1, +3.4)	+0.1 (0.0, +0.3)	+0.4 (+0.3, +0.6)	+3.2 (+3.0, +3.3)	
CthE3078	OpB	7 type II cohesins, anchor protein	-0.1 (-0.2, +0.1)	+0.1 (0.0, +0.2)	0.0 (-0.1, +0.1)	-0.1 (-0.2, 0.0)	-0.2 (-0.3, 0.0)	+0.4 (+0.3, +0.5)	Increase
CthE3079	Orf2p	2 type II cohesins, anchor protein	+0.3 (+0.1, +0.5)	-0.4 (-0.5, -0.2)	+0.2 (0.0, +0.4)	+0.8 (+0.9, -0.6)	-0.1 (-0.3, 0.0)	-0.4 (-0.6, -0.1)	
CthE0738		7 type II cohesins	-0.7 (-0.8, -0.5)	+1.1 (+1.0, +1.2)	+1.3 (+1.1, +1.5)	+1.6 (+1.4, +1.7)	+0.6 (+0.5, +0.8)	+0.8 (+0.6, +0.9)	
CthE2089	CelS	GH48, exoglucanase	+0.1 (0.0, +0.2)	0.0 (-0.1, +0.1)	-0.6 (-0.7, -0.5)	+0.1 (-0.1, +0.2)	-0.2 (-0.3, -0.1)	-0.5 (-0.6, -0.4)	Increase
CthE2147	CelO	GH5, CBM3, cellobiohydrolase	0.0 (-0.4, +0.4)	+0.1 (-0.2, +0.5)	+0.4 (0.0, +0.7)	-0.2 (-0.5, +0.1)	-0.3 (-0.6, 0.0)	-0.3 (-0.6, +0.1)	
CthE0412	CelK	GH9, CBM4, cellobiohydrolase	-0.1 (-0.2, +0.1)	+0.2 (+0.1, +0.3)	-0.4 (-0.5, -0.2)	+0.1 (0.0, +0.2)	-0.1 (-0.2, 0.0)	-0.7 (-0.9, -0.6)	Increase
CthE0413	CthA	GH9, CBM4, CBM3, cellobiohydrolase	0.0 (-0.2, +0.1)	-0.2 (-0.3, -0.1)	+0.6 (+0.5, +0.7)	-0.4 (-0.5, -0.3)	-0.3 (-0.5, -0.2)	-1.3 (-1.4, -1.2)	None
CthE0536	CelB	GH5, endoglucanase	0.0 (-0.2, +0.2)	-0.3 (-0.4, -0.2)	+0.4 (+0.2, +0.6)	-0.2 (-0.4, -0.1)	-0.3 (-0.4, -0.2)	+0.5 (+0.4, +0.6)	Decrease
CthE2872	CelG	GH5, endoglucanase	-0.1 (-0.1, +0.3)	-0.2 (-0.3, -0.1)	+0.2 (0.0, +0.4)	-0.2 (-0.3, 0.0)	-0.3 (-0.5, -0.2)	+0.4 (+0.3, +0.6)	Decrease
CthE405	CelL	GH5	-0.4 (-0.7, -0.1)	-0.2 (-0.4, +0.1)	+0.2 (-0.2, +0.6)	+0.1 (-0.2, +0.3)	-0.1 (-0.4, +0.1)	-0.5 (-0.8, -0.2)	
CthE0821		GH5, CBM32	+0.4 (+0.2, +0.5)	0.0 (-0.2, +0.1)	0.0 (-0.2, +0.1)	+0.2 (+0.1, +0.3)	+0.1 (0.0, +0.2)	+0.2 (+0.1, +0.3)	None
CthE193		GH5, CBM6, CBM13	-0.0 (-0.3, +0.4)	-0.3 (-0.6, -0.1)	+0.1 (-0.3, +0.4)	0.0 (-0.3, +0.3)	-0.5 (-0.8, -0.3)	+0.1 (-0.2, +0.4)	Decrease
CthE3012		GH5, CBM6	-0.5 (-1.2, +0.1)	-0.4 (-0.9, 0.0)	+1.2 (+0.6, +1.4)	+0.2 (-0.4, +0.7)	+0.6 (+0.1, +1.2)	n	
CthE2147	CelO	GH5, endoglucanase	-0.1 (-0.2, 0.0)	-0.3 (-0.5, -0.2)	+0.1 (0.0, +0.3)	-0.5 (-0.6, -0.3)	-0.6 (-0.7, -0.5)	-0.7 (-0.9, -0.6)	Decrease
CthE2825	CelD	GH9, endoglucanase	-0.1 (-0.4, +0.2)	+0.4 (+0.2, +0.6)	+0.3 (+0.1, +0.5)	+0.3 (+0.1, +0.6)	+0.2 (-0.1, +0.4)	-0.2 (-0.5, 0.0)	
CthE2812	CelT	GH9, endoglucanase	0.0 (-0.2, +0.2)	-0.1 (-0.2, +0.1)	-0.2 (-0.4, 0.0)	+0.2 (0.0, +0.3)	-0.1 (-0.3, 0.0)	-0.1 (-0.2, +0.1)	None
CthE0403		GH9, CBM3, endoglucanase	+0.1 (-0.2, +0.3)	+0.6 (+0.4, +0.8)	-0.1 (-0.4, +0.2)	+0.3 (+0.1, +0.5)	+0.2 (0.0, +0.4)	-0.4 (-0.7, -0.2)	
CthE0543	CelF	GH9, CBM3, endoglucanase	-0.2 (-0.4, -0.1)	+0.4 (+0.3, +0.5)	-0.1 (-0.2, +0.1)	+0.6 (+0.4, +0.7)	+0.2 (+0.1, +0.4)	+0.7 (+0.6, +0.8)	None
CthE0578	CelR	GH9, CBM3, endoglucanase	-0.0 (-0.1, +0.1)	-0.2 (-0.4, -0.1)	-0.3 (-0.5, -0.2)	-0.1 (-0.3, 0.0)	-0.5 (-0.6, -0.3)	-0.3 (-0.4, -0.1)	None
CthE0625	CelQ	GH9, CBM3, endoglucanase	0.0 (-0.2, +0.2)	0.0 (-0.1, +0.1)	-0.1 (-0.2, +0.1)	+0.1 (0.0, +0.3)	-0.3 (-0.4, -0.2)	-0.1 (-0.2, 0.0)	
CthE274	CelP	GH9	+0.1 (-0.1, +0.4)	+0.7 (+0.4, +1.0)	+0.7 (+0.4, +1.0)	+0.1 (-0.1, +0.2)	-0.1 (-0.3, +0.1)	-0.8 (-1.1, -0.6)	
CthE2760	CelV	GH9, CBM3	+0.1 (-0.1, +0.4)	+0.1 (0.0, +0.3)	0.0 (-0.2, +0.3)	+0.3 (+0.1, +0.5)	+0.3 (+0.1, +0.5)	+1.4 (+1.3, +1.6)	
CthE0745	CelW	GH9, CBM3	+0.2 (0.0, +0.4)	+0.2 (0.0, +0.3)	+0.1 (-0.1, +0.3)	+0.2 (0.0, +0.3)	0.0 (-0.1, +0.1)	-0.7 (-0.8, -0.5)	Decrease
CthE0433		GH9, CBM3	+0.3 (+0.1, +0.6)	+0.2 (0.0, +0.4)	+3.3 (+3.0, +3.6)	+0.4 (+0.2, +0.6)	+0.4 (+0.2, +0.5)	+0.7 (+0.5, +0.9)	None
CthE2761		GH9, CBM3	-0.1 (-0.4, +0.2)	-0.3 (-0.6, +0.0)	-0.4 (-0.8, 0.0)	0.0 (-0.2, +0.3)	-0.2 (-0.4, 0.0)	+1.2 (+0.9, +1.5)	None
CthE1398	XghA	GH74, xyloglucanase	+0.2 (+0.1, +0.4)	-0.3 (-0.5, -0.2)	+1.8 (+1.7, +1.9)	-0.1 (-0.2, +0.1)	-0.5 (-0.6, -0.3)	+2.0 (+1.9, +2.2)	Decrease
CthE1838	XynC	GH10, CBM22, xylanase	0.0 (-0.1, +0.2)	-0.1 (-0.3, 0.0)	+1.3 (+1.2, +1.5)	+0.1 (0.0, +0.3)	-0.2 (-0.4, -0.1)	+0.5 (+0.3, +0.5)	Decrease
CthE2590	XynD	GH10, CBM22, xylanase	+0.2 (+0.1, +0.5)	+0.7 (+0.3, +1.0)	n	n	+0.1 (-0.3, +0.5)		
CthE0912	XynY	GH10, CE1, CBM22, multifunctional component	0.0 (-0.4, +0.3)	-0.2 (-0.4, 0.0)	+0.8 (+0.5, +1.0)	+0.2 (-0.1, +0.5)	+0.1 (-0.1, +0.3)	-0.3 (-0.5, 0.0)	
CthE1963	XynZ	GH10, CE1, CBM2, multifunctional component	+0.2 (0.0, +0.5)	-0.1 (-0.3, +0.1)	+2.1 (+1.9, +2.3)	-0.1 (-0.3, +0.1)	+0.1 (-0.1, +0.3)	+0.2 (0.0, +0.4)	Decrease
CthE2972	XynAU	GH11, CE4, CBM6, xylanase, multifunctional component	-0.1 (-0.2, +0.1)	+0.5 (+0.4, +0.6)	+1.5 (+1.4, +1.7)	+0.2 (+0.1, +0.4)	+0.5 (+0.4, +0.6)	+0.1 (0.0, +0.3)	Decrease
CthE0624	CelJ	GH9, GH4, CBM30, CBM44, multifunctional component	0.0 (-0.2, +0.1)	-0.1 (-0.2, 0.0)	+0.2 (0.0, +0.3)	+0.4 (+0.2, +0.5)	-0.3 (-0.4, -0.2)	+1.3 (+1.2, +1.4)	Increase
CthE0797	CelE	GH5, CE2, multifunctional component	-1.0 (-1.2, -0.8)	+2.2 (+2.1, +2.4)	-0.2 (-0.4, 0.0)	+2.2 (+2.0, +2.4)	+1.3 (+1.1, +1.5)	u	Decrease
CthE1472	CelH	GH5, GH26, CBM11, multifunctional component	-0.2 (-0.7, +0.3)	0.0 (-0.3, +0.3)	+1.5 (+1.2, +1.6)	-0.5 (-0.9, -0.1)	0.0 (-0.4, +0.3)	+1.7 (+1.4, +2.0)	
CthE0270	ChIA	GH18, other hemicellulases (chitinase)	+0.2 (-0.3, +0.7)	n	u	-0.4 (-0.8, -0.1)	-0.5 (-0.9, -0.1)	0.0 (-0.5, +0.6)	None
CthE0211	LcbB	GH16, other hemicellulases (lichenase)	0.0 (-0.5, +0.5)	+0.2 (-0.3, +0.7)	I	n	-0.4 (-0.9, +0.1)	+4.2 (+3.9, +4.3)	
CthE2811	ManA	GH26, CBM35, other hemicellulases (mannanase)	0.0 (-0.2, +0.2)	+0.1 (-0.1, +0.2)	-1.3 (-1.4, -1.1)	0.0 (-0.1, +0.2)	-0.1 (-0.2, +0.1)	-0.4 (-0.6, -0.3)	
CthE0032		GH26, CBM35, other hemicellulases	-0.3 (-0.7, +0.1)	0.0 (-0.4, +0.3)	+1.3 (+1.0, +1.7)	-0.2 (-0.6, +0.1)	-0.3 (-0.6, 0.0)	+1.5 (+1.2, +1.7)	Decrease
CthE1400		GH53, other hemicellulases	0.0 (-0.5, +0.6)	-0.2 (-0.6, +0.3)	+1.4 (+1.0, +1.8)	-0.1 (-0.4, +0.2)	-0.3 (-0.6, +0.1)	+0.2 (-0.1, +0.6)	None
CthE3141		CE12, CBM35, putative carbohydrate esterase	+0.2 (-0.5, +0.8)	+0.7 (+0.2, +1.1)	+1.8 (+1.5, +2.2)	+0.6 (+0.2, +1.0)	+0.8 (+0.4, +1.1)	+0.8 (+0.2, +1.4)	
CthE0798		CE3, putative carbohydrate esterase	+1.6 (+1.0, +2.1)	u	u				
CthE0661		GH43, CBM13, putative glycosidase	-0.1 (-0.5, +0.4)	-0.6 (-0.9, -0.2)	+0.2 (-0.1, +0.5)	I	-0.9 (-1.3, -0.4)	-0.2 (-0.5, +0.2)	
CthE1271		GH43, CBM6, putative glycosidase	-0.1 (-0.6, +0.5)	n	+1.2 (+0.8, +1.6)	+0.1 (-0.3, +0.6)	+0.3 (-0.2, +0.7)	+1.0 (+0.6, +1.4)	
CthE2446		PL11, CBM35, putative pectinase	+0.2 (0.0, +0.5)	-0.4 (-0.6, -0.2)	+1.6 (+1.3, +1.9)	-0.3 (-0.6, -0.1)	-0.2 (-0.5, 0.0)	0.0 (-0.3, +0.2)	
CthE0190		protease inhibitor, serpin	-0.6 (-0.8, -0.5)	+1.2 (+1.1, +1.3)	+1.1 (+0.9, +1.3)	+0.6 (+0.5, +0.8)	+0.6 (+0.5, +0.8)	-0.7 (-0.9, -0.5)	
CthE191		protease inhibitor, serpin	I	n	n	n	n	n	
CthE0239		unknown function	0.0 (-0.4, +0.3)	+0.6 (+0.4, +0.8)	+1.6 (+1.3, +1.8)	+0.8 (+0.6, +1.1)	+0.7 (+0.4, +0.9)	+0.4 (+0.2, +0.6)	
CthE0258		unknown function	-0.1 (-0.6, +0.3)	u	+0.9 (+0.6, +1.2)	+0.5 (+0.1, +0.8)	+0.5 (+0.1, +0.8)	+1.3 (+0.9, +1.6)	
CthE0435		unknown function	+0.6 (+0.3, +0.8)	-0.1 (-0.5, +0.2)	+1.5 (+1.2, +1.7)	+0.3 (+0.1, +0.5)	+0.2 (0.0, +0.4)	+0.2 (-0.1, +0.4)	
CthE0640		unknown function	-0.2 (-0.3, +0.6)	-0.2 (-0.5, +0.2)	+1.1 (+0.8, +1.5)	-0.2 (-0.6, +0.3)	-0.1 (-0.4, +0.3)	+0.5 (0.0, +1.0)	
CthE2879		unknown function	I	n	+1.4 (+0.9, +1.9)	I	n	+0.2 (-0.2, +0.7)	
CthE3132		unknown function	n	u	+3.4 (+3.0, +3.9)	u	+2.8 (+1.7, +3.9)	n	
CthE1890		unknown function	n	u	+3.5 (+2.9, +3.9)	n			

Table 2: Normalized spectral abundance factors (NSAF) [8]

Locus	Gene	Domains	NSAF [8]						
			A	AX	Cb	PA	PAX	ZT	
CthE3077	CipA	9 type I cohesins - Scaffoldin							
CthE452		1 type I cohesin							
CthE1307	SdbA	1 type II cohesin, anchor protein							
CthE3078	OpB	7 type II cohesins, anchor protein							
CthE3079	Orf2p	2 type II cohesins, anchor protein							
CthE0738		7 type II cohesins							
CthE2089	CelS	GH48, exoglucanase							
CthE2147	CelO	GH5, CBM3, cellobiohydrolase							
CthE0412	CelK	GH9, CBM4, cellobiohydrolase							
CthE0413	CthA	GH9, CBM4, CBM3, cellobiohydrolase							
CthE0536	CelB	GH5, endoglucanase							
CthE2872	CelG	GH5, endoglucanase							
CthE405		GH5							
CthE0821		GH5, CBM32							
CthE193		GH5, CBM6, CBM13							
CthE3012		GH5, CBM6							
CthE2147	CelO	GH5, endoglucanase							
CthE2825	CelD	GH9, endoglucanase							
CthE2812	CelT	GH9, endoglucanase							
CthE0403		GH9, CBM3, endoglucanase							
CthE0543	CelF	GH9, CBM3, endoglucanase							
CthE0578	CelR	GH9, CBM3, endoglucanase							
CthE0625	CelQ	GH9, CBM3, endoglucanase							
CthE274		GH9							
CthE2760		GH9, CBM3							
CthE0745		GH9, CBM3							
CthE0433		GH9, CBM3							
CthE2761		GH9, CBM3							
CthE1398	XghA	GH74, xyloglucanase							
CthE1838	XynC	GH10, CBM22, xylanase							