JB Accepts, published online ahead of print on 30 November 2012
J. Bacteriol. doi:10.1128/JB.02014-12
Copyright © 2012, American Society for Microbiology. All Rights Reserved.

JB Accepts published online ahead of print

The role of the CipA scaffoldin protein in cellulose solubilization as determined by targeted gene deletion and complementation in *Clostridium thermocellum*

Running title: CipA deletion and complementation in *C. thermocellum*

Daniel G. Olson[1,4], Richard J. Giannone[2,4], Robert L. Hettich[2,4] and Lee R. Lynd[1,2,3,4]#

[1]Dartmouth College, Hanover, NH 03755

[2]Oak Ridge National Laboratory, Oak Ridge, TN 37830

[3]Mascoma Corporation, Lebanon, NH 03766

[4]BioEnergy Science Center, Oak Ridge, TN 37830

to whom correspondence should be addressed

lee.lynd@dartmouth.edu

JB Accepts published online ahead of print

Abstract

The CipA scaffoldin protein plays a key role in the *C. thermocellum* cellulosome. Previous studies have revealed that mutants deficient in binding or solubilizing cellulose also exhibit reduced expression of CipA. To confirm that CipA is, in fact, necessary for rapid solubilization of crystalline cellulose, the gene was deleted from the chromosome using targeted gene deletion technologies. The CipA deletion mutant exhibited a 100-fold reduction in cellulose solubilization rate, although it was eventually able to solubilize 80% of the 5 g/l cellulose initially present. The deletion mutant was complemented by a copy of *cipA* expressed from a replicating plasmid. In this strain, Avicelase activity was restored, although the rate was 2-fold slower than that of the wild type and the duration of the lag phase was increased. The *cipA* coding sequence is located at the beginning of a gene cluster containing several other genes thought to be responsible for the structural organization of the cellulosome, including *olpB*, *orf2p* and *olpA*. Tandem mass spectrometry revealed a 10-fold reduction in the expression of *olpB*, which may explain the slower growth rate. This deletion experiment adds further evidence that CipA plays a key role in cellulose solubilization by *C. thermocellum*, and raises interesting questions about the differential roles of the anchor scaffoldin proteins OlpB, Orf2p and SdbA.

Introduction

Clostridium thermocellum is an anaerobic thermophilic bacterium noted for its ability to rapidly solubilize crystalline cellulose, a process mediated by the cellulosome (1). The cellulosome is composed of tightly-bound enzymatic and structural components. At the heart of the

JB Accepts published online ahead of print

cellulosome is the scaffoldin protein, CipA (also known as S_L and S1) (2). This protein has been shown to be capable of crystalline cellulose solubilization in conjunction with cellulosomal cellulase Cel48S (3). Analysis of the DNA sequence of *cipA* has revealed a set of nine repeated elements known as type I cohesins (4). These cohesins, bind to the type I dockerins found on cellulosomal enzymes (5). Subsequent analysis of the CipA protein has revealed three additional modules, the type II dockerin, the cellulose binding module (CBM) and the x-domain.

The type II dockerin comprises a duplicated set of 22 amino acid residues located near the C-terminus of CipA (4). The type II dockerin binds to type II cohesins located on the anchor scaffoldin proteins, OlpB, Orf2p and SdbA. OlpB has seven type II cohesins, while Orf2p has two and SdbA has one. The anchor scaffoldins have a C-terminal sequence called the SLH domain that mediates binding to the cell surface(6). In CipA the CBM is located between the 2nd and 3rd type I cohesins and binds to crystalline cellulose with a K_D of 0.4 μM (1). Thus the current understanding of the adhesion of *C. thermocellum* to cellulose involves the following 3 interactions:

1. Binding of glycoside hydrolase enzymes in proximity to each other to promote enzyme-enzyme synergy.
2. Binding of enzymes to the cellulosic substrate via the CBM.
3. Anchoring the cellulosome to the cell surface. CipA binds to the anchor scaffoldin (OlpB, Orf2p or SdbA) via its type II dockerin. The anchor scaffoldins are attached to the cell by their SLH domains.

Finally, CipA has one additional module, located between the 9th type I cohesin and the type II dockerin called the x-module. Its function in *C. thermocellum* remains unknown, although it has been shown to improve the solubility of recombinantly expressed type II dockerins and seems to enhance the affinity of the type II cohesin-dockerin interaction (7).

JB Accepts published online ahead of print

Electron microscopy has revealed hemispherical protuberances on the outside of *C. thermocellum* cells which are known as polycellulosomes (8). In their resting state they are about 200 nm in diameter, but form a protracted conformation in the presence of cellulose (9). Immunolabeling has identified the presence of CipA (10) and OlpB (6) in the polycellulosomes, though the protuberances may contain other cellulosomal components as well.

There have been two previous reports of mutants of *C. thermocellum* deficient in cellulase activity. Both were isolated by screening for cells unable to adhere to cellulose. *C. thermocellum* AD2 was isolated by mixing cells with cellulose and allowing the cellulose to settle. Adherent cells were pulled out of solution upon binding to cellulose, thus enriching the supernatant for non-adherent cells. After five rounds of this sedimentation enrichment, strain AD2 was isolated by single colony purification (11). The AD2 strain was analyzed by SDS-PAGE and found to be missing a band associated with CipA when grown on cellobiose, although the band reappeared when the strain was grown on cellulose (8) . Further analysis of AD2 by scanning electron microscopy revealed the complete absence of polycellulosomes when grown on cellobiose (12).

Strains SM1, SM4, SM5 and SM6, also deficient in cellulase activity, were isolated using a procedure similar to that used for strain AD2, though augmented by an initial chemical mutagenesis step followed by a screen on cellobiose plates with an Avicel overlay (13). This final screen was designed to identify cells that were deficient in cellulose solubilization. These mutants were analyzed by SDS-PAGE and all were found to be missing a 210 kDa band associated with CipA. DNA sequence analysis revealed the presence of an IS1447 insertion

JB Accepts published online ahead of print

element disrupting the *cipA* coding sequence in each mutant. Strain SM1 had an insertion in the first type I cohesin and appeared to be completely lacking functional type I cohesins. This strain was unable to grow on MN300 cellulose and exhibited a 15-fold reduction in enzymatic activity compared to the wild type strain (13).

Previous work has shown that the ability to both bind and solubilize cellulose are linked (11, 13). Thus, mutants deficient in cellulose-binding are also deficient in cellulose solubilization. The functional link between these abilities is consistent with our understanding of the component modules of CipA. In this study, we further evaluate the extent to which *cipA* is responsible for this dysfunctional phenotype and explore the cellulase activity of *C. thermocellum* in the absence of a complexed cellulase system.

Materials and methods

Strains and media

All *C. thermocellum* strains described here are derived from *C. thermocellum* strain DSM 1313 and were grown in modified DSM 122 broth as described previously (14). Cellobiose or Avicel-PH105 microcrystalline cellulose (Sigma-Aldrich) was used as the primary carbon source at a concentration of either 5 or 10 g/l. Cells were grown at 55°C. Strain M1354 was a generous gift from the Mascoma Corporation (Lebanon, NH)(15). This strain is derived from *C. thermocellum* strain DSM 1313 and has a deletion of the *hpt* gene (Clo1313_2927) to allow for use of the *hpt* gene as a counterselectable marker with the antimetabolite 8-azahypoxanthine.

JB Accepts published online ahead of print

Molecular biological methods

Plasmids were constructed using yeast mediated ligation (16), In-Fusion PCR cloning (Takara Bio Inc.), or standard cloning techniques(17). Plasmids were maintained in *E. coli* TOP10 cells (Invitrogen Corporation) and prepared using QIAGEN Plasmid Mini kit (QIAGEN Inc.). Sequences of chromosomal DNA were obtained by PCR using genomic DNA from *C. thermocellum* strain DSM 1313. Primers were designed using genome sequences provided by the Joint Genome Institute (http://www.jgi.doe.gov/). The *repB* and *cat* genes are derived from plasmid pMU102 (18). The yeast origin of replication is derived from plasmid pMQ87 (16). The pMB1 *E. coli* origin of replication is derived from plasmid pUC19 (Invitrogen Corp.). The p15A *E. coli* origin of replication and arabinose-inducible promoter are derived from plasmid pBAD30 (19). The *hpt* and *tdk* genes (Tsac_0936 and Tsac_0324, respectively) are derived from *Thermoanaerobacterium saccharolyticum* JW/SL-YS485. The $gapDH_p$ promoter consists of the 525 bp region upstream of the *C. thermocellum glyceraldehyde 3-phosphate dehydrogenase* gene (Clo1313_2095). The cbp_p promoter consists of the 621 bp region upstream of the cellobiose phosphorylase gene (Clo1313_1954).

Plasmid pDGO-37 (GenBank accession number JX966413) was created by combining the p15A *E. coli* origin of replication and P*bad* promoter with the thermophilic gram positive origin of replication from plasmid pMU102. Plasmid pDGO-40 (GenBank accession number JX966414) was created by inserting the *cipA* coding sequence, including 819 bp upstream of the start codon (putative promoter region) and 67 bp downstream of the stop codon (putative terminator region) into plasmid pDGO-37 (Fig. 1)(Table 1).

JB Accepts published online ahead of print

PCR was performed using either Taq or Phusion DNA polymerase (New England Biolabs Inc.) according to the directions provided by the manufacturer. When using whole cells as the PCR template, a 10-min heating step was included at the beginning of the thermocycling protocol to lyse the cells. When using Taq DNA polymerase, the lysing temperature was 95 °C. When using Phusion DNA polymerase, the lysing temperature was 98 °C. DNA sequencing was performed using standard techniques with an ABI Model 3100 genetic analyzer (Applied Biosystems).

Strain construction

Previously, two different deletions of *cipA* were made. In strain DS11, the *cipA* coding sequence was deleted from start to stop codon using plasmid pDGO-03 (GenBank accession number JX489218.1). In strain DS16, the *cipAp* promoter sequence was deleted in addition to the *cipA* coding region using plasmid pDGO-34 (Genbank accession number JX489219.1) (20). Plasmids were transformed into *C. thermocellum* using previously described techniques (21). To avoid the potential for homologous recombination between the plasmid and chromosomal copy of the *cipA* promoter region, plasmid pDGO-40 was only transformed into strain DS16.

Fermentation conditions

Strains were grown in modified DSM 122 broth (18) at 55°C with cellobiose or Avicel microcrystalline cellulose as the primary carbon source. When fermentations were performed in a 125 ml glass bottle sealed with a butyl rubber stopper (22), the fermentation volume was 50 ml, 5 g/l substrate (Avicel or cellobiose) was used, the headspace was purged with nitrogen and the bottles were shaken at 200 rpm. When fermentations were performed in a computer-controlled fermenter (Sartorius GmbH), the fermentation volume was 2 L, 10 g/l substrate

JB Accepts published online ahead of print

(Avicel or cellobiose) was used, the headspace was purged with a mixture of 20% CO2, 80% N2, the vessel was stirred at 200 rpm and pH was controlled to 7.0 with 4N potassium hydroxide. For some fermentations, an automated sampling device was used to take 6 ml samples at regular intervals (23).

Analytical techniques

Concentrations of cellobiose, glucose, lactate, acetate, ethanol and formate were measured by high performance liquid chromatography (HPLC) as previously described (24). Total carbon and total nitrogen concentrations were measured with a Shimadzu TOC-V CPH elemental analyzer with TNM-1 and ASI-V modules (Shimadzu Corp.) on 0.5-1.0 ml aliquots washed twice with water. Avicel concentration was determined from these measurements by assuming that Avicel contained no nitrogen and that cells contained carbon and nitrogen in a 4.67:1 molar ratio. A detailed description of the theory and calculations is being prepared for publication elsewhere (Holwerda and Lynd unpublished data).

Samples for protein identification were prepared as previously described (14). Briefly, cell pellets were separated from supernatant by centrifugation (2000 g) and washed twice with Tris-Buffered Saline (100 mM Tris HCl, 150 mM NaCl, pH 8.0) to remove residual supernatant proteins. Cells were lysed by the addition of SDS lysis buffer (SDS LB; 4% SDS [w/v] in 100 mM Tris HCl, pH 8.0), boiled (5 min), sonically disrupted (Branson), and boiled again. Supernatants were concentrated 50-fold (50 ml to 1 ml) via spin filtration using a 3 KDa MW cutoff membrane (Vivaspin 20, 3 kDa, PES [GE Healthcare]), adjusted to 2% SDS with 1 ml of SDS LB, and boiled (5 min). Both fractions, whole-cell (WC) and supernatant (SN), were precleared by

JB Accepts published online ahead of print

centrifugation (21000 g) and protein concentrations determined by BCA assay (Pierce). Sample fractions were then combined in a 2:1 WC to SN ratio (w/w), reduced with 25 mM dithiothreitol (DTT), and TCA precipitated (3 mg of combined crude lysate adjusted to 20% TCA on ice for 1 hr). Precipitated proteins were then washed, resolubilized in denaturation buffer (8 M urea, 100 mM Tris HCl, 5 mM DTT, pH 8.0), digested with trypsin and prepared for MudPIT LC-MS/MS as previously described (14). In total, 50 ug of peptides were analyzed per sample via a 24-hr MudPIT analysis using an LTQ XL mass spectrometer (Thermo Scientific). Resulting peptide fragmentation data was then searched with the MyriMatch database search algorithm (25) against the *C. thermocellum* DSM 1313 proteome (with decoy sequences) as previously described. Identified peptides were then score-filtered (FDR < 2% peptide spectrum match) and assembled into proteins identifications (minimum of 2 distinct peptides per protein call) by IDPicker 3 (26). Proteins were then spectrally balanced to deal with non-unique peptides, normalized by NSAF (normalized spectral abundance factors), and abundance values adjusted to nSpC (normalized spectral counts) as previously described (27). Protein-to-protein abundance was then assessed across all samples to identify those that were differentially expressed (Student's T test).

Categorization of cellulase and cellulosomal proteins

A list of all proteins that could participate in cellulose solubilization was generated, based on membership in the Carbohydrate Active Enzyme (CAZy) database (28) or presence of a cohesin or dockerin domain as determined by the Pfam database (29) (Supplemental dataset S1). Proteins with Pfam domain PF00963 were labeled as "cohesin-containing." Proteins with Pfam domain PF00404 (and no cohesin domain) were labeled as "dockerin-containing." Three

JB Accepts published online ahead of print

additional proteins, Clo1313_1300, Clo1313_2479 and Clo1313_2861, were added to this list, based on analysis done by Ed Bayer and colleagues (Ed Bayer, personal communication). Proteins in the CAZy database that did not have a cohesin or dockerin domain were labeled as "CAZy, no cohesin, no dockerin."

Mathematical analysis of fermentation data

To determine the rate of substrate consumption, the substrate consumption data points were fitted with the 5-parameter sigmoidal Richards equation(30) as described by Holwerda and Lynd (unpublished data).

$$s(t) = A_0 + \frac{A_t - A_0}{\left(1 + e^{\frac{t_0 - t}{slope}}\right)^{asymm}} \qquad [1]$$

Where

A_0= lower horizontal asymptote
A_t= higher horizontal asymptote
t = time
x_0=inflection point
slope=slope at the inflection point
asymm= asymmetry parameter

The time (t) when the slope of the fitted curve was greatest was determined by taking the 2nd derivative with respect to time, setting it equal to zero and solving for t, which yielded the following equation

$$t_{\max slope} = x_0 + \ln(asymm) * slope \qquad [2]$$

The first derivative of equation [1] with respect to time was then evaluated at the time determined by equation [2] to determine the maximum rate of substrate consumption. To

JB Accepts published online ahead of print

allow ready comparison between Avicel and cellobiose, the substrate consumption rate was determined in mM glucose equivalent/hour.

Results

Comparing growth rates of various mutants

As expected, the wild type (WT) and *cipA* deletion strains (DS11 and DS16) have similar substrate consumption rates when grown on cellobiose (Fig. 2). The metabolic burdens of plasmid maintenance and thiamphenicol inactivation do not have an effect on Avicel consumption, as can be seen by comparing the wild type (strain WT) to the empty-vector control (strain DS19) (Fig. 2). The effect of *cipA* overexpression can be seen by comparing the empty-vector control (strain DS19) with the *cipA* overexpression strain (DS22). The rate of Avicel consumption was unchanged (Fig. 2). Deletion of *cipA* (strains DS11 and DS16) resulted in a 100-fold decrease in Avicel consumption rate when compared to the wild type strain (WT) (Fig. 2). Both cipA deletion strains were able to consume 80% of the Avicel initially present after ~2000 hours (Supplemental table S1 and supplemental figure S1). Fermentation products were similar to the parent strain (Supplemental table S2) and there was no significant accumulation of glucose or cellobiose. To see if the rate of Avicel consumption could be improved by adaptation, strain DS11 was subsequently passaged two additional times on Avicel, but no change in rate was detected.

Transforming the *cipA* deletion strain (DS16) with the *cipA* expression plasmid (pDGO-40, resulting in strain DS20) dramatically increased the rate of Avicel solubilization, although it was

JB Accepts published online ahead of print

still only 1/3 as fast as the empty vector control (DS19)(Fig. 2)(Supplemental table S1). Interestingly, there was a 100-hour lag phase before the start of rapid cellulose solubilization. The technique of measuring Avicel concentration by TOCN resulted in a larger degree of measurement variation during the early parts of fermentation. It is therefore difficult to determine whether slight negative trend observed in the first 100 hours with strain DS20 represents a physical phenomenon or is simply an artifact of the measurement technique (Fig. 3).

Comparing protein abundance in various mutants

Although the abundance was measured for all proteins (Supplemental dataset S1), only those proteins thought to be able to participate in cellulose solubilization (due to presence of a cohesin, dockerin or CAZy domain) were analyzed. The effect of the metabolic burdens of plasmid maintenance and antibiotic inactivation can be determined by comparing the empty vector control (strain DS19) to the wild type (WT) (Fig. 4, column 1). Among the cohesin-containing proteins, none are significantly differentially expressed. Among the dockerin-containing proteins, only Cel9F is significantly differentially expressed. Among the other CAZy proteins, only LicA is significantly differentially expressed.

The effect of *cipA* overexpression is demonstrated by comparing the empty vector control (strain DS19) with the *cipA* overexpression strain (strain DS22) (Fig. 4, column 2). Among the cohesin-containing proteins, none of them were significantly different at the 0.01 level. Although the increase in *cipA* expression was not significant at the 0.01 level ($p=0.015$), when the values for the wild type (WT) are included as well, the significance increases to 0.0002,

JB Accepts published online ahead of print

suggesting that the effect would likely be confirmed if we were to perform more replicates. *CipA* expression increased by 3-fold in strain DS22, although this was not significantly different from the mean *cipA* expression in the empty-vector control (strain DS19) at the 0.01 level (p=0.015). Among the dockerin-containing enzymes, none were significantly differently expressed at the 0.01 level. Among the other CAZy proteins, LicA and Clo1313_0647 (CBM16-Domain of unknown function) were significantly lower in abundance.

The effect of *cipA* complementation can be determined by comparing the empty vector control (strain DS19) with the complemented *cipA* deletion strain (DS20) (Fig. 4, column 3). Among the cohesin-containing proteins, OlpB showed significantly reduced expression and was 11-fold less abundant in the complemented deletion strain. Among dockerin-containing proteins, Cel9P and Clo1313_2861 (GH2-CBM6) showed increased expression. The significance of the Clo1313_2861 result is difficult to interpret because of the low number of spectra identified for this protein (≤ 7 for all samples). Among the other CAZy proteins, Clo1313_2460 (GH15) showed significantly increased abundance.

Since OlpB contains 7 type II cohesins, a dramatic decrease in OlpB levels could result in a decrease in type II cohesin availability. Type II cohesin availability was calculated by multiplying the abundance of each anchor scaffoldin (SdbA, Orf2p and OlpB) by the number of type II cohesins it contains (Table 2).

JB Accepts published online ahead of print

Discussion

In agreement with Zverlov et al., we have shown that *cipA* is essential for rapid solubilization of crystalline cellulose(13). However, contrary to what was reported, we observed that *cipA* deletion strains are able to solubilize Avicel microcrystalline cellulose (which is similar to the MN300 microcrystalline cellulose used by Zverlov et al.). Furthermore, since the solubilization of Avicel resulted in the production of lactate, acetate and ethanol and this ability was maintained despite serial transfer, it appears that the strain was able to grow on Avicel. What is the explanation for residual ability of the *cipA* deletion strains (DS11 and DS16) to solubilize crystalline cellulose? One possibility is that components of the non-complexed cellulase system (i.e. CelI and CelY), which have been shown to synergistically solubilize crystalline cellulose (31), can compensate for the expected loss of activity. Since each enzyme has its own CBM, there is no need for CipA to mediate binding with the cellulosic substrate. These enzymes were found at very low levels in all strains (<0.01% of cell protein) as determined by nSpC values (Supplemental dataset S1), which reduces support for this explanation. Another possibility is that the cellulosomal components are bound directly to the cell surface via OlpA, which contains both a type I cohesin (for binding a cellulase enzyme containing a type I dockerin) and an s-layer homology (SLH) binding domain (for binding to the cell surface). Levels of OlpA were about 40% higher in the complemented *cipA* deletion strain (DS20) compared with the empty vector control (DS19), which supports this hypothesis.

Why does the *cipA* deletion and complementation strain (DS20) grow more slowly than the wild-type strain (WT) and have a longer lag phase? *CipA* expression doesn't seem to be a likely explanation, since minor variations in CipA abundance do not appear to be correlated with

JB Accepts published online ahead of print

289 growth rate (Fig. 5). On the other hand, OlpB levels were unexpectedly low in this strain.
290 Compared to the empty-vector control strain (DS19), the *cipA* deletion and complementation
291 strain (DS20) had 30% fewer type II cohesins, since it seems to have partly compensated for the
292 reduction in *olpB* expression with higher levels of *sdbA* and *orf2p* expression. Furthermore, the
293 wild type strain (WT) and the *cipA* overexpression strain (DS22) both had 15% fewer type II
294 cohesins and this change in type II cohesin number did not have a substantial effect on
295 fermentation performance, thus it seems unlikely that the change in type II cohesin number is
296 the full explanation. Another possibility is that the anchor scaffoldins (OlpB, Orf2p and SdbA)
297 are not, in fact, interchangeable. For example, if Orf2p and SdbA are primarily used during
298 cellulosome assembly (as has been suggested for ORFXp in *Clostridium cellulolyticum* (32)) and
299 OlpB is the final destination for the assembled cellulosome, then a change in the abundance of
300 OlpB might have a greater impact on cellulosome function than would be indicated simply by
301 the overall change in type II cohesin availability. Further investigations will be required to
302 determine the exact molecular role of the various type II cohesin-containing proteins in
303 cellulosome assembly.

304 Why is *olpB* expression changed in the *cipA* deletion strain? Although *cipA* and *olpB* have been
305 reported to be transcribed independently (33), they may, in fact, be co-transcribed. Even if
306 *cipA* and *olpB* are expressed from individual promoters, the 1 kb region upstream of *cipA* may
307 contain other regulatory elements that affect *olpB* expression. Replacing the native *cipA*
308 promoter on plasmid pDGO-40 with a different promoter would allow *cipA* to be expressed
309 from a replicating plasmid in a *cipA* deletion strain where the native *cipA* promoter has been
310 left on the chromosome (i.e. strain DS11).

JB Accepts published online ahead of print

The *cipA* deletion and complementation system described here will be useful for systematic understanding of the cellulolytic capabilities of *C. thermocellum*. The ability to express *cipA* from a replicating plasmid will enable the rapid exploration of the roles of its subcomponents including: elucidating the function of individual modules of *cipA*, exploration of alternative cellulosomal architectures and characterization of its non-complexed cellulase system.

Acknowledgements

We wish to thank Mr. Kyle Saltsman for his assistance with experimental work related to measuring growth rates.

The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research, Genome Sciences Program in the DOE Office of Science.

References

1. **Demain AL, Newcomb M, and Wu JHD.** 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. **69:**124-154.
2. **Lamed R, Setter E, and Bayer EA.** 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. **156:**828-836.
3. **Wu JHD, Orme-Johnson WH, and Demain AL.** 1988. Two components of an extracellular protein aggregate of *Clostridium thermocellum* together degrade crystalline cellulose. Biochemistry **27:**1703-1709.
4. **Gerngross UT, Romaniec MPM, Kobayashi T, Huskisson NS, and Demain AL.** 1993. Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal S(L) protein reveals an unusual degree of internal homology. Mol. Microbiol. **8:**325-334.

JB Accepts published online ahead of print

5. **Tokatlidis K, Salamitou S, Beguin P, Dhurjati P, and Aubert JP.** 1991. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett. **291:**185-188.
6. **Lemaire M, Ohayon H, Gounon P, Fujino T, and Beguin P.** 1995. OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of Its S-layer-like domains to components of the cell-envelope. J. Bacteriol. **177:**2451-2459.
7. **Adams JJ, Pal G, Jia Z, and Smith SP.** 2006. Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. Proc. Natl. Acad. Sci. U. S. A. **103:**305-310.
8. **Bayer EA, Setter E, and Lamed R.** 1985. Organization and distribution of the cellulosome in *Clostridium thermocellum*. J. Bacteriol. **163:**552-559.
9. **Bayer EA, Shimon LJW, Shoham Y, and Lamed R.** 1998. Cellulosomes--structure and ultrastructure. J. Struct. Biol. **124:**221-234.
10. **Bayer EA, and Lamed R.** 1986. Ultrastructure of the cell-surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J. Bacteriol. **167:**828-836.
11. **Bayer EA, Kenig R, and Lamed R.** 1983. Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. **156:**818-827.
12. **Lamed R, Naimark J, Morgenstern E, and Bayer EA.** 1987. Specialized cell-surface structures in cellulolytic bacteria. J. Bacteriol. **169:**3792-3800.
13. **Zverlov VV, Klupp M, Krauss J, and Schwarz WH.** 2008. Mutations in the scaffoldin gene, *cipA*, of *Clostridium thermocellum* with Impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS*1447*, and implications for cellulase synergism on crystalline cellulose. J. Bacteriol. **190:**4321-4327.
14. **Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, and Lynd LR.** 2010. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. Proc. Natl. Acad. Sci. U. S. A. **107:**17727-17732.
15. **Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, and Caiazza NC.** 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. Appl. Environ. Microbiol. **77:**8288-8294.
16. **Shanks RMQ, Caiazza NC, Hinsa SM, Toutain CM, and O'Toole GA.** 2006. *Saccharomyces cerevisiae*-based molecular tool kit for manipulation of genes from gram-negative bacteria. Appl. Environ. Microbiol. **72:**5027-5036.
17. **Maniatis T, Fritsch EF, and Sambrook J.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
18. **Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR, Hogsett DA, and Caiazza NC.** 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. Appl. Environ. Microbiol. **76:**6591-6599.
19. **Guzman LM, Belin D, Carson MJ, and Beckwith J.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J. Bacteriol. **177:**4121-4130.
20. **Waller B, Olson DG, Currie DH, Guss A, and Lynd LR.** 2012. Exchange of type II dockerin-containing subunits of the *Clostridium thermocellum* cellulosome. FEMS Microbiol. Lett. **in press***.
21. **Olson DG, and Lynd LR.** 2012. Chapter seventeen - transformation of *Clostridium thermocellum* by electroporation, p. 317-330. *In* H. J. Gilbert (ed.), Methods Enzymol., vol. Volume 510. Academic Press.

JB Accepts published online ahead of print

22. **Hogsett DA.** 1995. Cellulose hydrolysis and fermentation by *Clostridium thermocellum* for the production of ethanol. Ph.D. thesis. Dartmouth College, Hanover, NH.
23. **Ellis LD.** 2011. Mass balances applied to cellulosic fermentations by *Clostridium thermocellum* (ATCC 27405) and the development of an automatic sampling system. M.S. thesis. Dartmouth College, Hanover.
24. **Zhang Y-HP, and Lynd LR.** 2005. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. **187:**99-106.
25. **Tabb DL, Fernando CG, and Chambers MC.** 2007. MyriMatch: Highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J. Proteome Res. **6:**654-661.
26. **Ma ZQ, Dasari S, Chambers MC, Litton MD, Sobecki SM, Zimmerman LJ, Halvey PJ, Schilling B, Drake PM, Gibson BW, and Tabb DL.** 2009. IDPicker 2.0: improved protein assembly with high discrimination peptide identification filtering. J. Proteome Res. **8:**3872-3881.
27. **Lochner A, Giannone RJ, Keller M, Antranikian G, Graham DE, and Hettich RL.** 2011. Label-free quantitative proteomics for the extremely thermophilic bacterium *Caldicellulosiruptor obsidiansis* reveal distinct abundance patterns upon growth on cellobiose, crystalline cellulose, and switchgrass. J. Proteome Res. **10:**5302-5314.
28. **Cantarel BL, Coutinho PM, 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:**233-238.
29. **Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, and Bateman A.** 2010. The Pfam protein families database. Nucleic Acids Res. **38:**D211-D222.
30. **Richards FJ.** 1959. A flexible growth function for empirical use. J. Exp. Bot. **10:**290-301.
31. **Berger E, Zhang D, Zverlov VV, and Schwarz WH.** 2007. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol. Lett. **268:**194-201.
32. **Desvaux M.** 2005. The cellulosome of *Clostridium cellulolyticum*. Enzyme. Microb. Technol. **37:**373-385.
33. **Dror TW, Rolider A, Bayer EA, Lamed R, and Shoham Y.** 2003. Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. J. Bacteriol. **185:**5109-5116.

Figure legends

FIGURE 1 Diagram of genetic elements used in this work.

JB Accepts published online ahead of print

418 **FIGURE 2** Substrate consumption rate for strains of *C. thermocellum* grown on either cellobiose
419 (cb) or Avicel (Av) at initial concentrations of 5 or 10 g/l. Antibiotic selection was used to
420 maintain plasmid in plasmid-containing strains. The presence of the *cipA* coding sequence is
421 indicated as either chromosomal (C), plasmid-based (P) or both. Error bars represent one
422 standard deviation and were determined based on biological replicates, where $n \geq 2$.

423 *Due to the difficulties of growing strains DS11 and DS16 on Avicel in fermenters, they were
424 grown in sealed glass bottles instead.

425 **FIGURE 3** Avicel consumption of 4 strains of *C. thermocellum* growing on 10 g/l Avicel. In order
426 to allow subsequent comparison with growth rates on cellobiose, the rate was reported in mM
427 glucose equivalents per hour. Based on an assumed monomer mass of 162 g/mole and a 5%
428 moisture content of Avicel, 58.6 mM glucose equivalents were present initially. Avicel
429 consumption was measured by elemental analysis of the pellet fraction of fermentation broth
430 corrected for cell carbon. Error bars represent one standard deviation, n=3, for Avicel
431 measurement of a representative fermentation. Solid lines represent the best fit of a 5-
432 parameter logistic equation. Equation parameters are given in Supplemental table S1.

433 **FIGURE 4** Comparison of protein abundance as determined by normalized spectral abundance
434 factor (nSpC) from tandem mass spectrometry measurements of fermentation broth (combined
435 cells and supernatant) at the end of Avicel fermentations. nSpC measurements were taken
436 from biological duplicate experiments. Pairwise comparisons were made and proteins with
437 significant changes ($p < 0.01$) are indicated by filled symbols. Other proteins are indicated by

JB Accepts published online ahead of print

438 unfilled symbols. The presence of cohesins, dockerins and carbohydrate-binding modules
439 (CBMs) were determined by searching the Pfam database (29).

440 **FIGURE 5** Comparison of substrate consumption rate with the abundance of the CipA scaffoldin
441 protein for duplicate fermentations with strains DS1, DS19, DS20 and DS22 grown on 10 g/l
442 Avicel.

443 ## Tables

444 **TABLE 1** Description of strains

Strain	Chromosomal genetic elements			Plasmid genetic elements		Genotype	Source
	cipAp	*cipA*	*hpt*	*cat*	*cipAp-cipAt*		
WT	+	+	+			Wild type *C. thermocellum* DSM1313	DSMZ[a]
M1354	+	+				Δ*hpt*	(15)
DS11	+					M1354 Δ*cipA*	(20)
DS16						M1354 Δ(*cipAp-cipA*)	(20)
DS18				+		DS16 /pDGO-37	This study
DS19	+	+		+		M1354/pDGO-37	This study
DS20				+	+	DS16/pDGO-40	This study
DS22	+	+		+	+	M1354 /pDGO-40	This study

445 [a]Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Germany.

JB Accepts published online ahead of print

446

447 **TABLE 2** Abundance of type II cohesins

Name	Type II coh per molecule	Protein abundance (nSpC) DS19	WT	DS20	DS22
SdbA	1	35.9	30.1	67.7	46.2
OlpB	7	44.1	42.0	3.8	38.7
Orf2p	2	106.3	70.8	148.1	77.9
Total type II cohesins		557.4	466.0	390.8	473.5

448

449

450

JB Accepts published online ahead of print

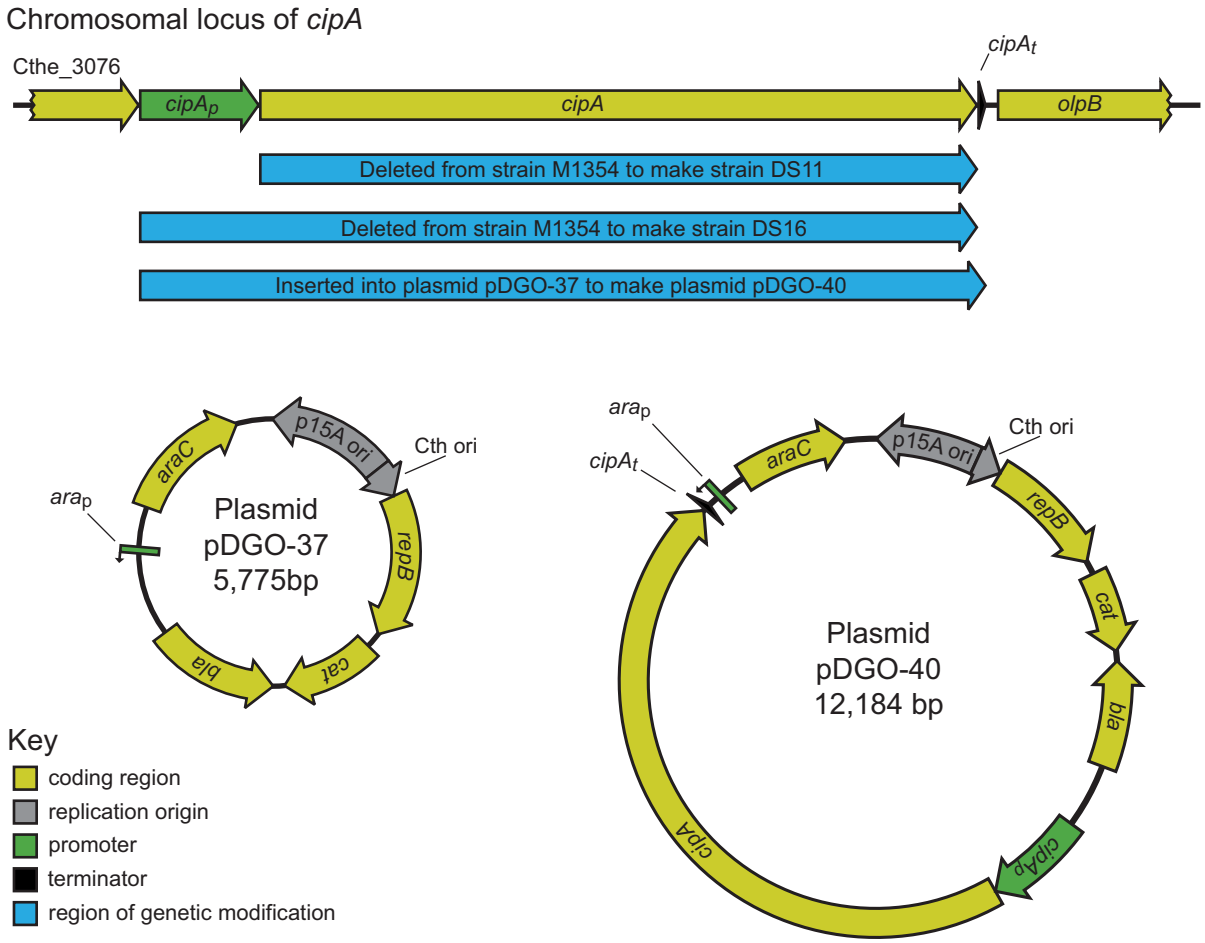
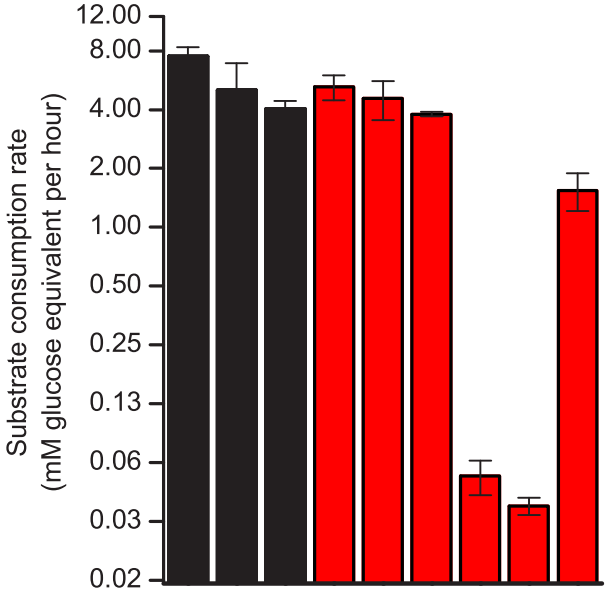


FIGURE 1 Diagram of genetic elements used in this work

JB Accepts published online ahead of print



strain	WT	DS11	DS16	WT	DS19	DS22	DS11*	DS16*	DS20
substrate	cb	cb	cb	Av	Av	Av	Av	Av	Av
amount (g/l)	10	10	10	10	10	10	5	5	10
antibiotic					yes	yes			yes
cipA	C			C	C	C,P			P

FIGURE 2 Substrate consumption rate for strains of *C. thermocellum* grown on either cellobiose (cb, black bars) or Avicel (Av, red bars) at initial concentrations of 5 or 10 g/l. Antibiotic selection was used to maintain plasmid in plasmid-containing strains. The presence of the *cipA* coding sequence is indicated as either chromosomal (C), plasmid-based (P) or both. Error bars represent one standard deviation and were determined based on biological replicates, where $n \geq 2$.

*Due to the difficulties of growing strains DS11 and DS16 on Avicel in fermenters, they were grown in sealed glass bottles instead.

JB Accepts published online ahead of print

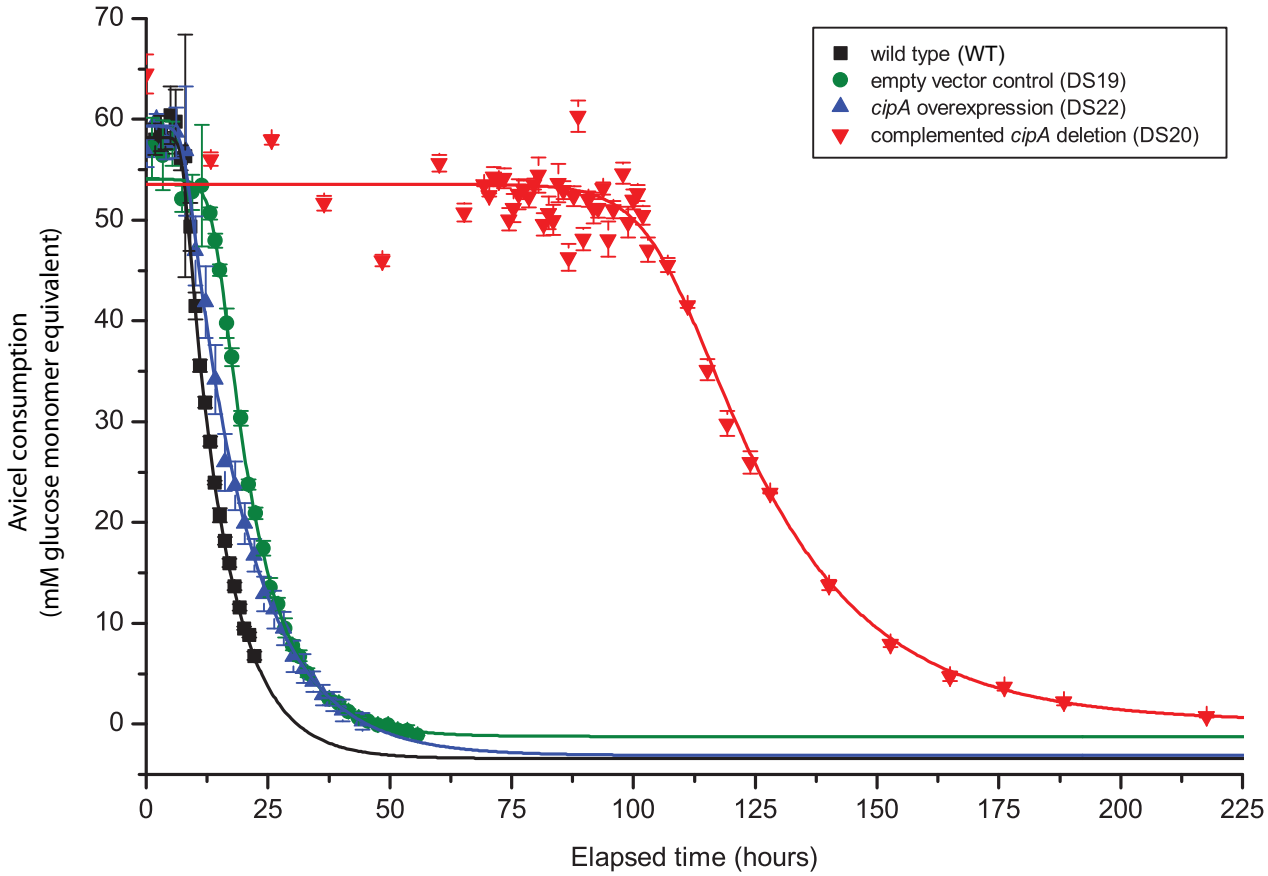


FIGURE 3 Avicel consumption of 4 strains of *C. thermocellum* growing on 10 g/l Avicel. In order to allow subsequent comparison with growth rates on cellobiose, the rate was reported in mM glucose equivalents per hour. Based on an assumed monomer mass of 162 g/mole and a 5% moisture content of Avicel, 58.6 mM glucose equivalents were present initially. Avicel consumption was measured by elemental analysis of the pellet fraction of fermentation broth corrected for cell carbon. Error bars represent one standard deviation, n=3, for Avicel measurement of a representative fermentation. Solid lines represent the best fit of a 5-parameter logistic equation. Equation parameters are given in Supplementary material Table S1.

JB Accepts published online ahead of print

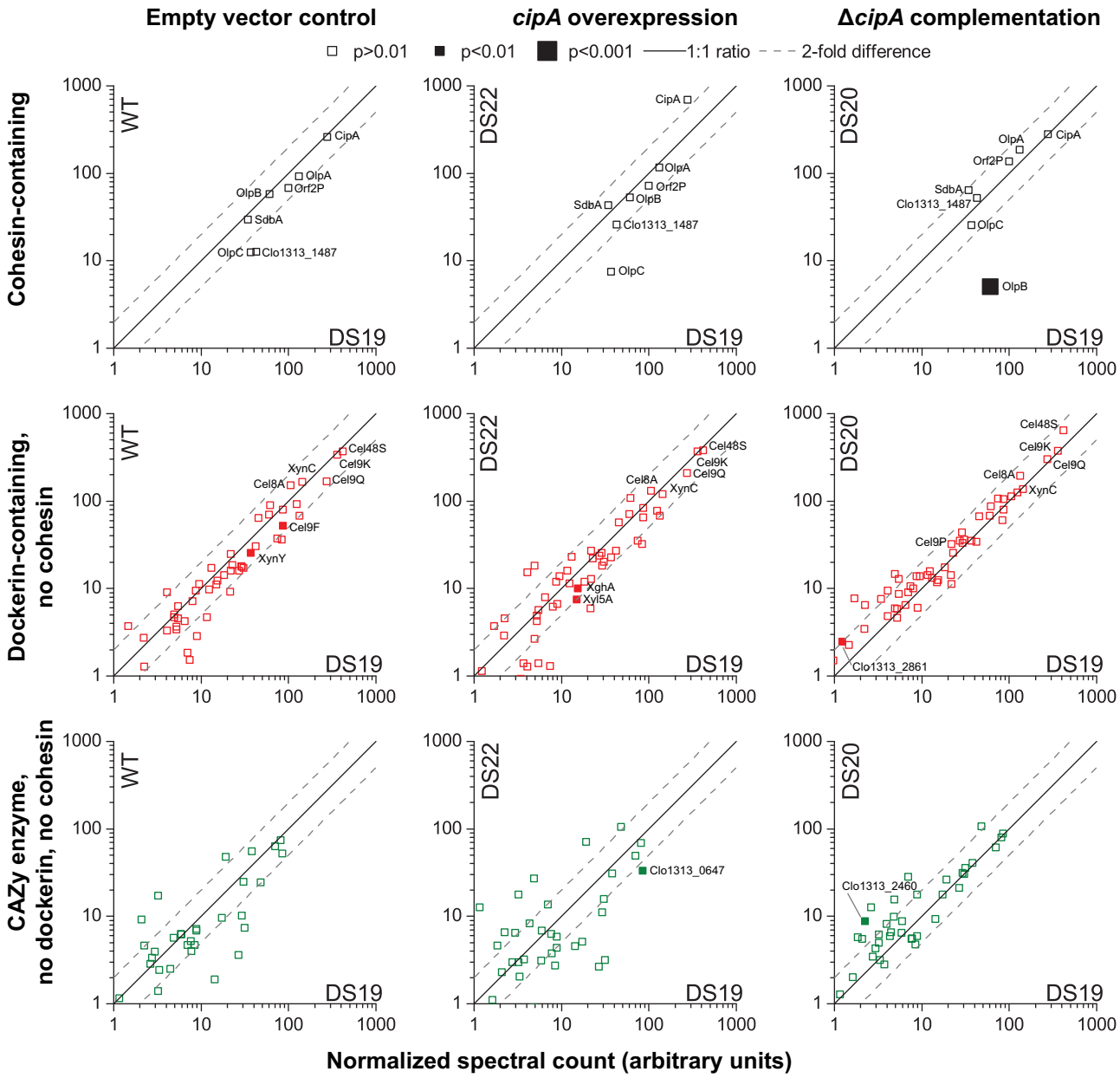
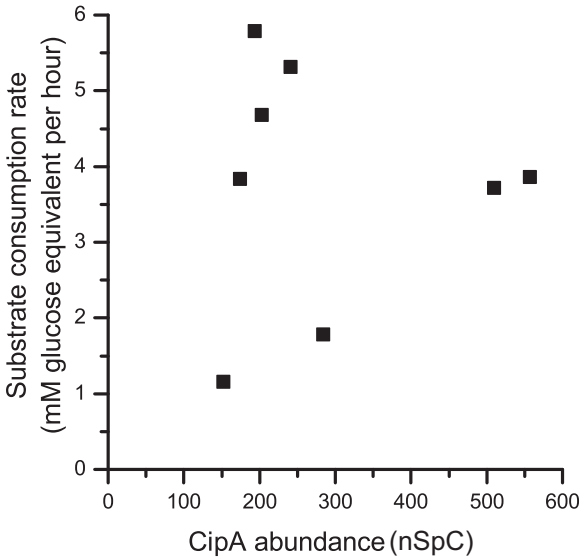


FIGURE 4 Comparison of protein abundance as determined by normalized spectral abundance factor (nSpC) from tandem mass spectrometry measurements of fermentation broth (combined cells and supernatant) at the end of Avicel fermentations. nSpC measurements were taken from biological duplicate experiments. Pairwise comparisons were made and proteins with significant changes ($p < 0.01$) are indicated by filled symbols. Other proteins are indicated by unfilled symbols. The presence of cohesins, dockerins and carbohydrate-binding modules (CBMs) were determined by searching the Pfam database.

JB Accepts published online ahead of print



Comparison of substrate consumption rate with the abundance of the CipA scaffoldin protein for duplicate fermentations with strains WT, DS19, DS20 and DS22 grown on 10 g/l Avicel.