Yoav *et al. Biotechnol Biofuels (2017) 10:222*
DOI 10.1186/s13068-017-0909-7

Biotechnology for Biofuels

RESEARCH

Open Access

How does cellulosome composition influence deconstruction of lignocellulosic substrates in *Clostridium* (*Ruminiclostridium*) *thermocellum* DSM 1313?

Shahar Yoav[1,2], Yoav Barak[3], Melina Shamshoum[4], Ilya Borovok[5], Raphael Lamed[5], Bareket Dassa[4], Yitzhak Hadar[1], Ely Morag[4] and Edward A. Bayer[4*]

Abstract

Background: Bioethanol production processes involve enzymatic hydrolysis of pretreated lignocellulosic biomass into fermentable sugars. Due to the relatively high cost of enzyme production, the development of potent and cost-effective cellulolytic cocktails is critical for increasing the cost-effectiveness of bioethanol production. In this context, the multi-protein cellulolytic complex of *Clostridium* (*Ruminiclostridium*) *thermocellum,* the cellulosome, was studied here. *C. thermocellum* is known to assemble cellulosomes of various subunit (enzyme) compositions, in response to the available carbon source. In the current study, different carbon sources were used, and their influence on both cellulosomal composition and the resultant activity was investigated.

Results: Glucose, cellobiose, microcrystalline cellulose, alkaline-pretreated switchgrass, alkaline-pretreated corn stover, and dilute acid-pretreated corn stover were used as sole carbon sources in the growth media of *C. thermocellum* strain DSM 1313. The purified cellulosomes were compared for their activity on selected cellulosic substrates. Interestingly, cellulosomes derived from cells grown on lignocellulosic biomass showed no advantage in hydrolyzing the original carbon source used for their production. Instead, microcrystalline cellulose- and glucose-derived cellulosomes were equal or superior in their capacity to deconstruct lignocellulosic biomass. Mass spectrometry analysis revealed differential composition of catalytic and structural subunits (scaffoldins) in the different cellulosome samples. The most abundant catalytic subunits in all cellulosome types include Cel48S, Cel9K, Cel9Q, Cel9R, and Cel5G. Microcrystalline cellulose- and glucose-derived cellulosome samples showed higher endoglucanase-to-exoglucanase ratios and higher catalytic subunit-per-scaffoldin ratios compared to lignocellulose-derived cellulosome types.

Conclusion: The results reported here highlight the finding that cellulosomes derived from cells grown on glucose and microcrystalline cellulose are more efficient in their action on cellulosic substrates than other cellulosome preparations. These results should be considered in the future development of *C. thermocellum*-based cellulolytic cocktails, designer cellulosomes, or engineering of improved strains for deconstruction of lignocellulosic biomass.

Keywords: Lignocellulosic biomass, Proteomics, Enzymatic hydrolysis, Scaffoldins, Biofuels

*Correspondence: ed.bayer@weizmann.ac.il
[4] Department of Biomolecular Sciences, The Weizmann Institute of Science, 76100 Rehovot, Israel
Full list of author information is available at the end of the article

Background

Developing cost-effective and renewable alternative energy resources capable of replacing currently used fossil fuel is an important challenge [1]. Cellulosic ethanol, one of the suggested solutions of this global issue, meets the necessary requirements of being renewable and environmentally friendly [2, 3]. Unlike the production process of the first-generation bioethanol alternative, which utilizes the edible parts of plants, the cellulosic ethanol alternative exploits the inedible polysaccharides of the plant, notably the cellulose, found in the cell walls of lignocellulosic biomasses [4, 5]. Agriculture or industrial lignocellulosic wastes can be used as sources of biomass, although removal of plant residues from the field could also have negative effects on soil fertility and quality [6].

The plant cell wall is a chemically complex structure composed of cellulose, hemicelluloses, and lignin as the main polymers. Those polymers, together with other components, provide the plant cell with the robustness required for its diverse functions [7, 8].

The production process for converting cellulosic biomass to ethanol involves three major steps [2]. The first includes chemical or physical pre-treatment, which is designed to loosen the rigid structure of the plant cell wall, to increase cellulose accessibility and to enrich the cellulose fraction. In the second step, the enriched cellulose fraction is hydrolyzed into soluble fermentable sugars. In the third step, the soluble sugar mixture is used as a carbon source for alcoholic fermentation. To date, the hydrolysis step is performed by enzymatic hydrolysis, rendering cellulosic ethanol economically infeasible, mainly due to the relatively high production costs of the carbohydrate-hydrolyzing enzymes [9, 10].

Cellulolytic microorganisms can utilize the cellulose as a carbon source. The natural arsenal of plant cell wall-degrading enzymes is diverse, and includes cellulases, hemicellulases, pectinases, ligninases, and additional accessory enzymes. Cellulose-hydrolyzing enzymes are classified into three major groups by their sequence homology and biochemical characteristics: (A) endoglucanases, which cleave bonds in the middle of the cellulose chains in random or semi-random fashion; (B) exoglucanases, which hydrolyze cellulose from either the reducing or non-reducing end in a processive manner, releasing soluble non-monomeric sugars; and (C) β-glucosidases, which hydrolyze the end-product (cellobiose) of cellulase hydrolysis to produce glucose [11, 12].

Clostridium thermocellum (recently reclassified as *Ruminiclostridium thermocellum*) is one of the best-explored and well-characterized cellulose-degrading bacteria in nature. Due to its characteristics, this anaerobic thermopilic bacterium was suggested to be the organism of choice for bioethanol production processes [13–17]. Its cellulolytic machinery, called cellulosome, is a multi-protein complex that contains a multiplicity of catalytic subunits, as well as structural proteins (scaffoldins) which are responsible for integrating the catalytic subunits into a well-ordered high-molecular-weight complex [14, 18]. Selected scaffoldins can bind cellulose by virtue of an integral cellulose-binding module (CBM), which are attached to the bacterium via an anchoring protein that contains an S-layer homology (SLH) module [19]. In this manner, the cellulosome creates proximity and substrate-targeting effects [20]. There are more than eighty genes in the genome of *C. thermocellum* that encode for cellulosomal subunits [21]. In addition to the cellulosome, *C. thermocellum* also utilizes soluble non-cellulosomal cellulolytic enzymes for deconstruction of cellulose [22–24].

The composition and structure of plant cell walls differ among different plant species as well as among different tissues in a given plant. In addition to their inherent variability, different pre-treatments can further alter the composition of the lignocellulosic biomass, leading to even higher diversity among the carbon sources used in the bioethanol production process [8]. Consequently, different enzyme compositions might be required for efficient hydrolysis of the different carbon sources. Indeed, proteomic and transcriptomic studies have shown that the expression pattern of the cell wall-degrading enzymes and the composition of the cellulosomes change in response to the carbon source of the medium. In other words, *C. thermocellum* senses the biomass in the medium and assembles a cellulosome preparation tailored to the requirements of the bacterium. Former studies have highlighted the need for understanding differential assembly of cellulosomal subunits in order to reveal key enzymes that are important for efficient hydrolysis [25–28].

In the current study, cellobiose (CB), microcrystalline cellulose (MCC), alkaline-pretreated switch grass (alSG), alkaline-pretreated corn stover (alCS), and dilute acid-pretreated corn stover (acCS), were used as sole carbon sources for growth of *C. thermocellum* DSM1313. In nature, *C. thermocellum* hydrolyzes cellulose into cellobiose units, which are in turn consumed by the bacterium. An adaptation process can enable some *C. thermocellum* strains to utilize monomeric glucose as a sole carbon source [13, 29–31]. In this study, such an adaptation process was conducted, and glucose was also used as a sole carbon source. The influence of various carbon sources on the structure and subunit composition of the resultant cellulosomes, and consequently on its hydrolysis activity, was investigated.

Methods

Alkaline pre-treatment of corn stover and switchgrass

Corn stover was collected after harvest from Moshav Kfar HaRif, Israel. Switchgrass was obtained from Notts Farms, Clinton, ON, Canada. Alkaline pre-treatment was carried out as previously described [32]. Briefly, 100 g of each feedstock were separately placed into 2-L glass beakers followed by the addition of 700 mL of 2% [wt/wt] NaOH solution. The beakers containing biomass and alkali solution were heated to boiling and allowed to proceed under this temperature for 1 h with continuous stirring. The pretreated biomasses were then washed by water through a glass Buchner funnel and adjusted to neutral pH. Finally, the pretreated biomasses were drained using vacuum through the funnel and brought to about 20–30% (wt/wt) solid content.

Dilute acid-pretreated corn stover

Dilute sulfuric acid-pretreated corn stover (160 °C for 1 min at an effective acid concentration of 1–2% [w/w]) was obtained from the National Renewable Energy Laboratory (NREL), Golden, CO (Batch Number P080828-CS-8. Manufactured: 26.11.13; Record No. 579).

Chemical composition of biomass

The chemical composition of the various biomasses was determined by conventional chemical analysis methods [33, 34]. Briefly, for delignification, the desired lignocellulosic biomass (1% slurry [w/v]) was supplemented with 1% acetic acid [v/v] and 1.5% [w/v] sodium chlorite and boiled for 1 h. The delignification process was then repeated. The obtained (white) holocellulose, i.e., complex of cellulose and hemicelluloses, was hydrolyzed by boiling in 1.5% hydrochloric acid for 2 h. The content of cellulose was calculated from the dry residue remaining after hydrolysis of the holocellulose, while the content of hemicelluloses was measured from weight loss of the hydrolyzed holocellulose sample. Lignin Klason was analyzed by means of standard TAPPI procedure T222 [34].

Anaerobic fermentation of *C. thermocellum*

Growth of the anaerobic thermophilic bacterium *C. thermocellum* (strain DSM 1313 obtained from the DSMZ collection) was performed as previously described [35] with minor changes. Briefly, GS-2 medium (0.5 g/L K_2HPO_4, 0.5 g/L $MgCl_2{\cdot}6H_2O$, 0.5 g/L KH_2PO_4, 1.3 g/L $(NH4)_2SO_4$, 0.002 g/L resazurin, 10.5 g/l 3-(*N*-morpholino) propanesulfonic acid (MOPS) buffer, 5 g/L yeast extract, 0.5 mM $CaCl_2$, and 1.25 mg/L iron(II) sulfate) was adjusted with 10 M NaOH to a final pH of 7.2. A portion (400 mL) of the medium was transferred into 0.5 L serum bottles containing 0.3% (wt/vol) of the different carbon sources (except 0.5% in the case of glucose), boiled, and extensively flushed with nitrogen. The bottles were sealed, autoclaved (121 °C, 20 min), and inoculated with a fresh CB-based *C. thermocellum* starter culture. For inoculation of glucose-containing media, a preliminary adaptation process was performed. Glucose-based media were thus inoculated with *C. thermocellum* followed by seven successive re-inoculation steps, which resulted in significant shortening of the lag phase rendering it comparable to the that of cellobiose-based growth media. Triplicate samples were prepared. Bottles were incubated for 48 h in a 60 °C shaking incubator.

Cellulosome purification

C. thermocellum growth media were centrifuged (10,900*g*, 7 min), and the supernatant fluids were carefully removed from the pellet and concentrated 40 times using a Pellicon XL biomax 300 cassette (Millipore, Cat. No. PXB300C50). Concentrated samples were fractionated by size exclusion chromatography using a Superdex S-200 prep grade 16/60 gel filtration column (GE Healthcare). Fractions (1 mL) were collected and analyzed by 6% SDS-PAGE. Fractions containing the cellulosomes (scaffoldin and identified enzymatic subunits) were pooled (Additional file 1: Figure S1). Cellulosome concentration was determined using a Pierce™ BCA Protein Assay Kit (Thermo scientific, Waltham, MA). Since samples from cellobiose- and glucose-based media showed relatively low concentrations, they were further concentrated by Vivaspin (Sartorius, Goettingen, Germany) with polyethylene sulfate (PES) membrane (30,000 MWCO). Samples were stored at −20 °C until use.

β-glucosidase

Thermoanaerobacter brockii thermostable β-glucosidase, CglT (GenBank: ADV80605.1), was a kind gift of CelDezyner LTD, Israel (alon@celdezyner.com). The concentration of CglT in the unpurified sample was determined based on comparative activity tests (1 mL of 50 mM sodium citrate buffer, pH 6.0, containing 5 mM *p*-nitrophenyl-β-D-1, 4-glucopyranoside [Sigma-Aldrich, Rehovot, Israel] was supplemented with 5 µL CglT sample dilutions, followed by incubation at 60 °C for 10 min. Optical densities were measured at a wavelength of 405 nm and compared to that of an assay mixture containing purified CglT.

Activity assay

All activities assays were conducted in a final volume of 1 mL solution, containing 20 mM citrate buffer (pH 6.0) supplemented with 10 mM $CaCl_2$, and substrate loadings of 7% for the MCC hydrolysis assay or 5% for the lignocellulosic biomasses. Cellulosome loadings of 20, 50, 3, or 50 µg/mL were used for MCC, alSG, alCS, and acCS

hydrolysis assays, respectively. The latter concentrations were found to be in the near-linear range of the reactions, as determined by preliminary calibration experiments (Additional file 2: Figure S2). CglT (equivalent to 0.33 mg/mL of purified enzyme) was added to the reaction mixture in order to prevent cellobiose feedback inhibition. To evaluate the activities, samples were incubated overnight at 70 °C with continuous shaking, centrifuged, and the supernatant fluids were separated from the undigested biomass. Released soluble sugar (reducing end) concentrations were analyzed by the dinitrosalicylic acid (DNS) method, as previously described [36]. Final soluble sugar concentrations were determined against a glucose calibration curve, and specific activity [μM reducing ends $(\mu g \text{ protein})^{-1} \text{ min}^{-1}$] was calculated.

Proteolysis

The purified cellulosome samples were dissolved in 8 M urea in 100 mM ammonium bicarbonate, reduced by dithiothreitol at a final concentration of 2.8 mM (60 °C for 30 min) and modified with 8.8 mM iodoacetamide in 100 mM ammonium bicarbonate (30 min, room temperature, in the dark). The reduced, modified samples were then digested by modified trypsin (Promega, Madison, WI) at a 1:50 enzyme-to-substrate ratio in 2 M urea, 25 mM ammonium bicarbonate, overnight, followed by a second digestion step (4 h).

Mass spectrometry analysis

Following the digestion step, the resultant peptide mixture was desalted, dried, and re-suspended in 0.1% formic acid. The peptides were resolved by reverse-phase chromatography on 0.075 × 180-mm fused silica capillaries (J&W) packed with Reprosil reversed phase material (Dr. Maisch GmbH, Germany). The peptides were eluted with a linear 60-min gradient of 5–28% acetonitrile with 0.1% formic acid, 5-min gradient of 28–95%, and 15 min at 95% acetonitrile with 0.1% formic acid in water, at a flow rate of 150 nL/min. MS analysis was performed by Q Exactive plus mass spectrometer (Thermo Scientific, Waltham, MA) in a positive mode using repetitively full MS scan followed by collision-induced dissociation (CID) of the ten most dominant ions selected from the first MS scan. The MS data were analyzed using MaxQuant v1.5.1.2 software (Cox and Mann [37]) versus the *C. thermocellum* DSM1313 section of the NCBI-nr database with 1% FDR, and further analyzed against the Carbohydrate-Active enzymes (CAZY) database. Due to the repetitive nature of some cellulosomal subunit sequences, we considered only proteins identified by at least one unique peptide. Data were statistically analyzed using Perseus v1.5.0.31 (part of the MaxQuant package). Intensities were normalized by the previously described intensity-based absolute quantification (iBAQ) method [38]. Average and standard deviations of duplicate samples of CB- and MCC-derived cellulosomes and triplicates of glucose-, alSG-, alCS-, and acCS-derived cellulosomes were analyzed.

Results and discussion

Purification of different cellulosomes

In order to investigate the influence of different carbon sources on the cellulosome composition and consequently on its activity, *C. thermocellum* strain DSM1313 was grown on cellobiose (CB), microcrystalline cellulose (MCC), alkaline-pretreated switchgrass (alSG), alkaline-pretreated corn stover (alCS), and dilute acid-pretreated corn stover (acCS). The latter lignocellulosic biomasses are representative of industrially relevant feedstocks. Both alkaline- and dilute acid-based pre-treatments are well established and common in the bioethanol field, designed to enrich the cellulosic fraction, and to increase accessibility of hydrolytic enzymes. The chemical composition of the different lignocellulosic biomasses is shown in Table 1. In nature, *C. thermocellum* hydrolyzes the cellulose into soluble cellobiose units, which in turn are actively taken up by the bacterium and further hydrolyzed into glucose units by a cell-associated β-glucosidase. Soluble glucose can be directly utilized and used as the sole carbon source by some strains of *C. thermocellum* only after a prolonged adaptation period [13, 29, 30]. In this study, such an adaptation period was used to generate cellulosomes from glucose-based growth media. Production of cellulosome samples was

Table 1 Chemical composition of lignocellulosic biomasses used in this work

	Abbreviation	Cellulose[a] (%)	Hemicellulose[a] (%)	Lignin[a] (%)	Non-ligno cellulose fraction[a] (%)
Untreated switchgrass	SG	37	28	18	17
Untreated corn stover	CS	36	27	20	17
Alkaline-pretreated switchgrass (alSG)	alSG	56	20	21	3
Alkaline-pretreated corn stover (alCS)	alCS	64	16	13	7
Dilute acid-pretreated corn stover (acCS)	acCS	60	5	30	5

[a] % dry matter

accomplished in triplicate for each of the six different carbon sources. The soluble cell-free cellulosomes were purified and analyzed for their cellulolytic activity and subunit composition.

Activity assays

In order to calculate specific activities [μM reducing ends (μg protein)$^{-1}$ min^{-1}], substrate degradation was quantified and total protein concentration in each sample was measured. MS analysis revealed that in addition to cellulosomal proteins, the "purified" high-molecular weight cellulosomal fraction also contained unrelated proteins, namely proteins without any known direct lignocellulolytic function (e.g., S-layer domain-containing proteins or flagellin domain-containing proteins). Consequently, the specific activity measured in a given sample would be biased by the presence of the unrelated proteins and would further lower the specific activity. In order to overcome this discrepancy, the relative content of cellulosomal proteins and non-cellulosomal enzymes in each sample was calculated using the MS data and used for calculations of the "true" specific activity.

Generally, the various isolated cellulosomes displayed varied specific activities on the different substrates, thus demonstrating the significant influence of the carbon source used in the growth media on the activity of the resulted cellulosome preparation (Fig. 1). One leading dogma assumes that cellulosomal subunit composition, generated from growth media supplemented with a specific carbon source, will lead to a superior activity

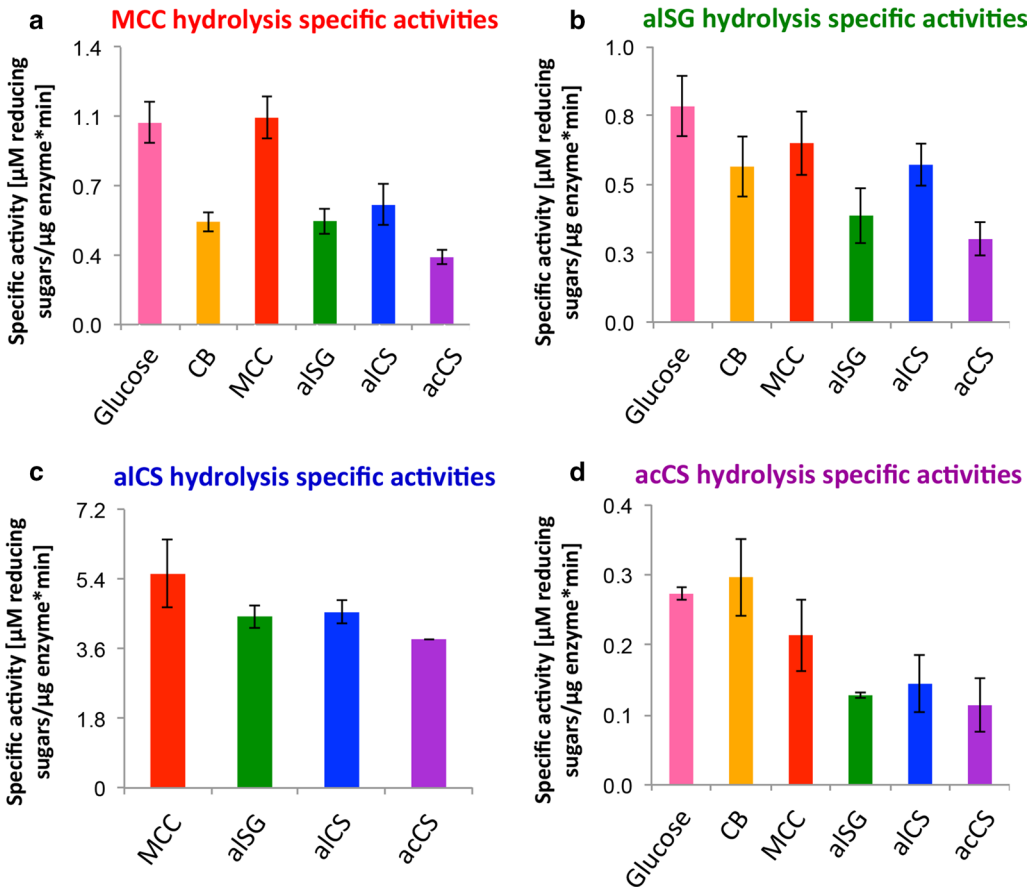

Fig. 1 Specific activities of the various cellulosome fractions: cellulosomes derived from glucose-, cellobiose (CB)-, microcrystalline cellulose (MCC)-, alkaline-pretreated switchgrass (alSG)-, alkaline-pretreated corn stover (alCS)-, and dilute acid-pretreated corn stover (acCS)-based growth media were applied on the different substrates and released soluble sugars were measured by the DNS method. The percent content of relevant enzymes (cellulosomal subunits and soluble carbohydrate-active enzymes) in each sample was calculated using MS, and specific activities were calculated [μM reducing ends μg protein^{-1} min^{-1}]. **a** Cellulosomes (20 μg) were applied to 70 mg MCC for 19 h; **b** cellulosomes (60 μg) were applied to 50 mg alSG for 16 h; **c** cellulosomes (3 μg) were applied to 50 mg alCS for 15 h. **d** Cellulosomes (50 μg) were applied to 50 mg acCS for 18 h. Cellulosomes and biomass dosages were determined in preliminary calibration assays in order to work in the near-linear range of the reaction (Additional file 2: Figure S2). All assays were performed at a final volume of 1 mL, at 70 °C, with the addition of 0.33 mg/mL equivalent of *Thermoanaerobacter brockii* β-glucosidase (CglT) in order to prevent feedback inhibition

towards that specific carbon source. For example, cellulosome preparations, generated from bacteria grown on corn stover, would be assumed to display relatively high hydrolytic activity towards the same substrate, compared to cellulosome preparations generated from wheat straw or switchgrass. The results obtained in this study do not entirely support this dogma. Although the specific activity of MCC-derived cellulosomes in hydrolyzing MCC was indeed higher than those of the other cellulosome preparations (Fig. 1), no advantage was found for the lignocellulosic biomass-derived cellulosome preparations in the hydrolysis of the same type of biomass used to generate them. Moreover, cellulosomes derived from glucose-based substrates (i.e., MCC, CB, and glucose itself) exhibited higher specific activities towards acCS, compared to cellulosomes from bacteria grown on lignocellulosic biomass, including acCS itself. In the same manner, glucose- and MCC-derived cellulosome preparations revealed higher specific activities towards hydrolysis of alSG compared to the alSG-derived cellulosome. Interestingly, glucose- and MCC-derived cellulosomes showed similar specific activities on all tested biomasses, including MCC.

Previous proteomic and transcriptomic studies have suggested that understanding the relationship between a specific carbon source and the resultant cellulosomal subunit composition will enable selection of key potent enzymes for efficient hydrolysis of specific substrates important for industry [25, 28, 39]. Conversely, the results reported here suggest that glucose- and MCC-derived cellulosomes are comparable or superior in their polysaccharase activity on all cellulosic substrates tested, to those obtained from cells grown on cellobiose or lignocellulosic substrates. Consequently, glucose- and MCC-derived cellulosomes are better sources for determining optimal compositions of key enzymes than those derived from the other substrates tested. The finding that differential assembly processes and consequent cellulosome compositions do not necessarily display more efficient cellulosomes is supported by similar findings of hydrolyzing pretreated switchgrass by *Clostridium clariflavum* cellulosomes and partially by the hydrolysis of untreated wheat straw by *Clostridium cellulolyticum* cellulosomes [40, 41].

Several explanations can be suggested in order to explain these findings. For example, an optimized tailored cellulosome, assembled by the bacterium to fit a carbon source, is not necessarily the most active one. Alternatively, controlled hydrolysis rate, and the resultant controlled soluble sugar release, may serve to avoid competition and/or interact with surrounding satellite microorganisms [42]. In the same manner, uncontrolled biomass degradation can increase the release of inhibitors. Furthermore, the hydrolysis rate has to be synchronized with the cellobiose uptake rate, since high cellobiose concentrations can inhibit cellulosome activity [43]. Thus, it is likely that differential assembly of cellulosomal subunits in response to carbon source is designed to regulate the hydrolysis rate rather to achieve the highest rate of hydrolysis. Another possible explanation lies in the nature of the biomass used in the original bacterial cell culture. Pure cellulose chains, as well as soluble cellobiose or glucose, are not common in nature (except for wastes derived from human society). Even the lignocellulosic biomasses used in the present study have undergone pre-treatment steps, as required for the bioethanol production process. Therefore, the carbon sources used in this study are 'unnatural' substrates (i.e., not available to the bacteria in nature). The bacterial regulation apparatus, designed to respond to the natural carbon source in the environment, might be ineffective or even a source of interference for degradation of the 'unnatural' cellulosic substrates used here. The resultant cellulosome preparation would thus show no advantage in their hydrolysis [41].

Mass spectrometry analysis

The compositions of purified cellulosomes were analyzed by label-free LC–MS/MS. Peptide sequences from MS data were compared to the annotated protein database of *C. thermocellum* DSM1313 (NCBI, RefSeq NC017304.1) and further compared to the CAZY database for characterized carbohydrate-active enzymes (http://www.cazy.org/) [44]. Measured intensities were normalized using the intensity-based absolute quantification (iBAQ) method. Since mass spectrometry measurements depend not only on the concentration of each protein but also on its amino acid sequence, different proteins of the same concentration might reveal different total intensity measurements. The iBAQ method normalizes the results in a way that two different proteins with the same molar ratio will show similar iBAQ intensity, thus enabling internal comparison among samples (i.e., between different proteins in a given sample) [38].

Scaffoldins

Cellulosomes are heterogeneous sets of high-molecular weight multi-enzyme complexes composed of multidomain structural scaffoldins and several dozen catalytic subunits. The binding of the catalytic subunits is mediated by the high-affinity noncovalent interactions between the multiple scaffoldin-borne cohesin modules and a dockerin module located in each catalytic subunit. The primary scaffoldin in *C. thermocellum,* CipA (Clo1313_0627), herein termed as ScaA (Fig. 2) according to Bras et al. [45], serves as a binding platform for the

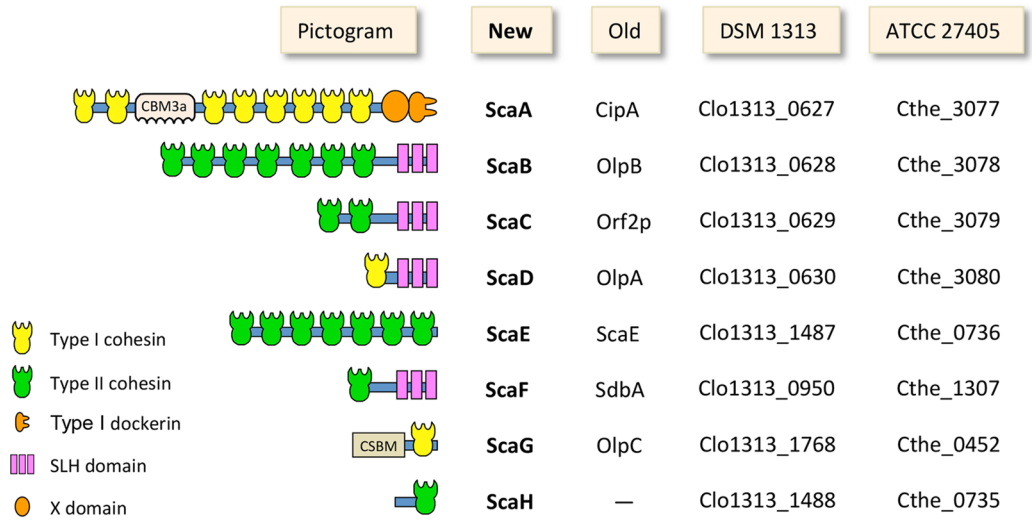

Fig. 2 Schematic diagram of the eight *C. thermocellum* DSM1313 scaffoldins. The genome of *C. thermocellum* DSM1313 (RefSeq NC017304.1) was screened for cohesin-containing proteins, and annotated by NCBI annotation software. New (as reported in Bras et al. [45]) and old terminologies are shown for the scaffoldins, together with their gene loci in strain DSM 1313 and their equivalents in the ATCC 27405 strain. CBM, carbohydrate-binding module; SLH, surface-layer homology; CSBM, cell-surface binding module. The numbers of cohesins in ScaA and ScaB (8 and 7, respectively) shown in the figure are according to the recent report of Hong et al. [49]

catalytic subunits via type I cohesin–dockerin interactions. In addition, *C. thermocellum* cellulosomes can be assembled into more complex cellulosomal suprastructures (polycellulosome complexes) via secondary scaffoldins. This type of assembly is mediated by the interaction of a type II dockerin module (located at the C terminus of the scaffoldin) with type II cohesins located in the secondary scaffoldin. Those secondary scaffoldins might be either cell-associated anchoring scaffoldins (mediating attachment of cellulosomes to the bacterial cell wall via their S-layer homology [SLH] domain [18, 19]) or soluble cell-free scaffoldins (lacking the SLH domain). Secondary scaffoldins, which bear several type II cohesins, can bind an equivalent number of primary scaffoldins and thus can assemble poly-cellulosomal complexes of different size, architecture, and content. The organization of the catalytic subunits on structural scaffoldins create a proximity effect that enhances synergy among neighboring enzymes [20, 46].

In order to analyze the structure of the different cellulosome samples, the genome of *C. thermocellum* DSM1313 (NC017304.1) [47] was screened for cohesin-bearing proteins. A schematic diagram of all *C. thermocellum* DSM1313 scaffoldins is shown in Fig. 2. Previously identified scaffoldins are now coined herein according to new terminology as reported in Bras et al. [45]. The scaffoldins of strain DSM1313 are similar in architecture and content to those of strain ATCC 27405 [48], with two noticeable differences: (A) according to the available genome sequence, DSM1313 ScaA contains six type I cohesin domains while ATCC 27405 ScaA (CipA) contains nine type I cohesin modules, and (B) the DSM1313 secondary scaffoldin ScaB contains four type II cohesin modules, whereas ATCC 27405 ScaB (OlpB) contains seven type II cohesin modules. While those discrepancies arise from sequence annotation, it is relevant to note that Hong et al. [49] have claimed that the genomic sequencing of the *cipA* and *olpB* genes in the DSM1313 strain was incorrect. Based on PCR reaction studies, these authors claimed that *C. thermocellum* DSM1313 ScaA contains eight cohesins (one less than that reported for *C. thermocellum* ATCC 27405 [50]) and ScaB contains seven cohesins (similar to that of *C. thermocellum* ATCC 27405 OlpB). In the present work, we have accepted the latter claims and used a value of eight type I cohesins for ScaA and seven type II cohesins for ScaB.

MS analysis revealed the presence of all eight known scaffoldins in all cellulosome samples, albeit some (i.e., ScaH and ScaE in some samples) were present only in relatively minor amounts.

Secondary scaffoldins

In order to shed light on the polycellulosomes suprastructures, the relative proportion of each secondary scaffoldin, derived from cells grown on each substrate, was assessed (Fig. 3). The different cellulosome samples exhibited differences in distribution of the secondary scaffoldins.

The most abundant secondary scaffoldin in most of the tested cellulosomal samples (i.e., CB-, MCC-, alSG-, and

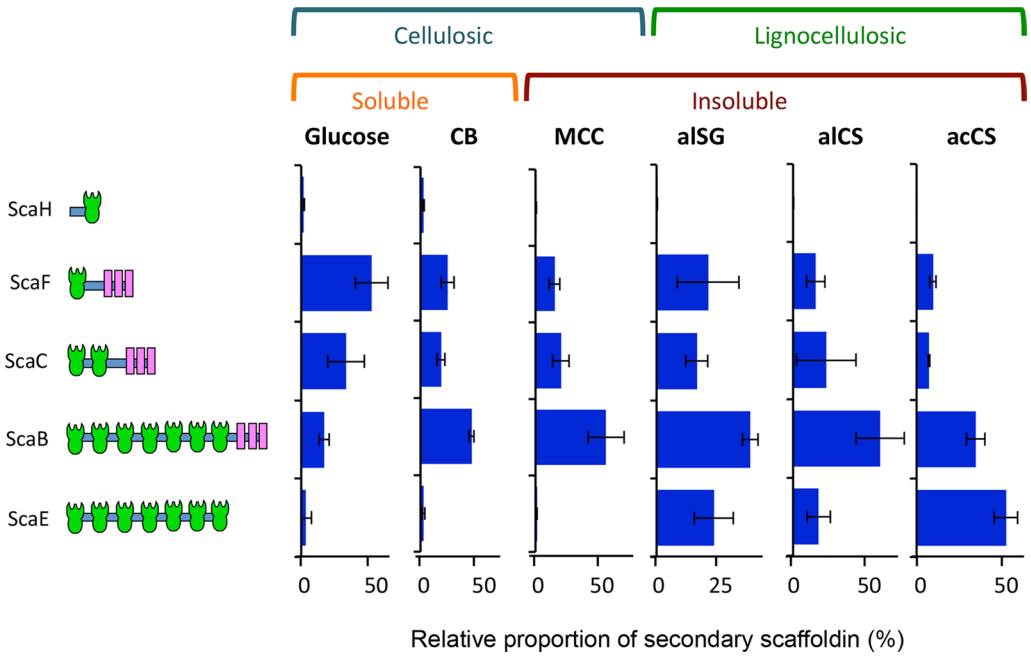

Fig. 3 Differential distribution of secondary type II cohesin-containing scaffoldins. iBAQ intensities of all secondary scaffoldins in each sample were summed up, and the relative proportions of each secondary scaffoldin in each sample was calculated. The carbon source (see legend to Fig. 1 for definitions) used for the respective cellulosome production is shown in the graph

alCS-derived cellulosomes) was ScaB. In contrast, the most abundant secondary scaffoldin for acCS-derived cellulosomes was ScaE. Both ScaB and ScaE (containing seven cohesin models, respectively) enable the formation of large polycellulosome superstructures. Conversely, the most abundant secondary scaffoldin in the glucose-derived cellulosomes was ScaF (formerly SdbA), containing only one cohesin module. The combined amount of both ScaB and ScaE in the glucose-derived samples was only 19% (compared to 51–83% in the other samples). These results suggest the formation of significantly smaller cellulosomal complexes in the glucose-derived cellulosomes. It was therefore of interest to consider whether the multiple polycellulosome suprastructures assembled by the various secondary scaffoldins would directly influence their activity on (ligno) cellulosic substrates. For example, cellulosome architecture could possibly increase the synergistic effect by assembling several individual cellulosomes into a higher order structure, each of which would bear a different composition of catalytic subunits. In addition, the size of the polycellulosome complex may influence its accessibility to the crystalline cellulosic regions of the substrate [51]. However, no connection was observed in the present study between the distribution of secondary scaffoldins and the activity of the various cellulosome preparations. Intriguingly, glucose- and MCC-derived samples revealed similar levels of activity (Fig. 1), despite of the differences in the distribution of their secondary scaffoldins. In addition, CB- and MCC-derived cellulosomes exhibited similar patterns of secondary scaffoldins but significantly different specific activities when assayed for hydrolysis of MCC. These results indicate that the distribution of secondary scaffoldin per se does not significantly influence specific activity. Our results correspond to previously described reports, demonstrating that secondary scaffoldins exhibit only a minor effect on the hydrolysis of cellulosic substrates compared to the critical effect of ScaA [49, 52–55].

ScaE contains seven type II cohesins but lacks an SLH domain. The relative proportion of ScaE was higher (one order of magnitude) in lignocellulosic substrate-derived cellulosomes (i.e., alSG, alCS, and acCS) compared to that of cellulosomes derived from cells grown on pure homogeneous cellulose and its degradation products (i.e., MCC, CB, and glucose). ScaE was found to mediate the assembly of large polycellulosome complexes [52]. Those complexes were termed long-range, cell-free cellulosomes due to the lack of SLH or any other known module or sequence that would facilitate their binding to the *C. thermocellum* cell wall [52]. It was suggested that such "cell-free cellulosomes" can diffuse away from the cell and degrade polysaccharide substrates remotely from the bacterial cell [52]. It was further suggested that such a system can accommodate and target catalytic subunits

involved in the hydrolysis of "non-cellulosic" [25] (i.e., hemicellulosic) polysaccharides via the CBM domain of the scaffoldin. In this context, the hemicellulase subunits would thus expose the cellulose microfibrils of complex lignocellulosic substrates, thereby enabling its sequential hydrolysis by the cell-surface cellulosomes. Interestingly, our results demonstrate relatively high portions of ScaE when *C. thermocellum* DSM1313 was grown on lignocellulosic substrates (where exposure of cellulose fibrils is necessary) and relatively low portions of ScaE when *C. thermocellum* DSM1313 was grown on pure cellulose or its degradation products (where such exposure of the microfibrils is not required).

Another "cell free" secondary scaffoldin produced by this bacterium is ScaH (that contains only a single type II cohesin without an SLH domain). This secondary scaffoldin was the least abundant in all samples. Nevertheless, the relative portion of ScaH in cellulosomes derived from soluble substrates (i.e., CB and glucose) was one order of magnitude higher compared to those of cellulosomes derived from the lignocellulosic substrates (i.e., alSG, alCS, and acCS) and ~2.5-fold higher than that of MCC-derived cellulosome.

Primary scaffoldins

Among all structural proteins, the relative proportion of primary scaffoldins (type I cohesin-containing proteins) was larger than that of the secondary scaffoldins in all samples, accounting for 58–63% in the glucose-, CB-, and MCC-derived cellulosomes, and 76–80% for the lignocellulose-derived cellulosomes (Table 2).

Figure 4 summarizes the distribution of the primary scaffoldins. ScaA, by far, comprised the great majority of the primary scaffoldins, accounting for 67% (in CB-derived cellulosomes) up to 98% (in alSG-derived cellulosomes), followed by ScaD, and ScaG. The relative proportion of ScaD and ScaG (single type I cohesin) was significantly lower in the insoluble substrate-derived cellulosomes (7.7, 2.3, 5.1, and 6.5% for MCC-, alSG-, alCS-, and acCS-derived cellulosomes, respectively) compared to that of the glucose- and cellobiose-derived cellulosomes (18.3 and 33% for glucose- and CB-derived cellulosomes, respectively). These results correlate with the measured intensities of Cthe_0452 (ScaG homolog) in the ATCC 27405 CB-derived cellulosomes versus those derived from insoluble substrates [25].

Table 2 Relative proportion of primary and secondary scaffoldins in the different samples

Carbon source	Primary scaffodins (%)	Secondary scaffoldins (%)
Glucose	60 ± 3	40 ± 3
CB	63 ± 1	37 ± 1
MCC	58 ± 6	42 ± 6
alSG	76 ± 2	24 ± 2
alCS	78 ± 6	22 ± 6
acCS	80 ± 2	20 ± 2

Occupancy of type II cohesins

ScaA is the only primary scaffoldin that contains a type II dockerin that can bind to the type II cohesins of secondary scaffoldins. Comparison of the amount of ScaA to that of the available type II cohesins can also provide insight into the suprastructure of the different cellulosome samples. In order to calculate the type II dockerin-to-cohesin ratio, the relative portions of the secondary scaffoldins were multiplied by the number of their cohesins, and the ScaA/type II cohesin ratio was calculated. The ratio for CB- and MCC-derived cellulosomes was 0.26 ± 0.0 and 0.27 ± 0.07, respectively. Significantly higher ratios were found for the lignocellulosic biomass-derived cellulosomes (0.61 ± 0.03, 0.65 ± 0.2, and 0.63 ± 0.09 for alSG-, alCS-, and acCS-derived cellulosomes, respectively), thus indicating a higher degree of occupancy and complexity in the lignocellulosic biomass-derived cellulosomes. Interestingly, cellulosomes that exhibited higher occupancies for type II cohesins (i.e., lignocellulosic substrate-derived cellulosomes) generally showed relatively lower levels of specific activities (Fig. 1). Surprisingly, the ratio of glucose-derived cellulosomes was 0.51 ± 0.09, similar to those of the lignocellulosic substrate-derived cellulosomes. Yet, the glucose-derived cellulosomes exhibited relatively high specific activity values. This result may be related to the distribution of secondary scaffoldins in the glucose-derived cellulosomes, which shows a majority of the monovalent scaffoldin ScaF, as opposed to the other types of cellulosome. Thus, in this case, the degree of occupancy may only have a negligible influence on activity. Incomplete occupancy of type II cohesins was previously reported for *C. thermocellum* ATCC 27405 cellulosomes, with even lower occupancy levels at the proteome [25] and transcriptome levels [56].

Cellulosomal subunits

In order to analyze subunit composition of the different cellulosome samples, the *C. thermocellum* DSM1313 genome was screened for type I dockerin-containing proteins. Seventy-five hypothetical cellulosomal subunits were thus revealed. Altogether, the proteomic study revealed 67 different dockerin-containing proteins in the cellulosome samples. For comparison among the different cellulosome samples, iBAQ intensities of cellulosomal subunits of a given sample were normalized with that of the primary scaffoldin ScaA (CipA) from the same sample, thus creating a relative abundance index. The

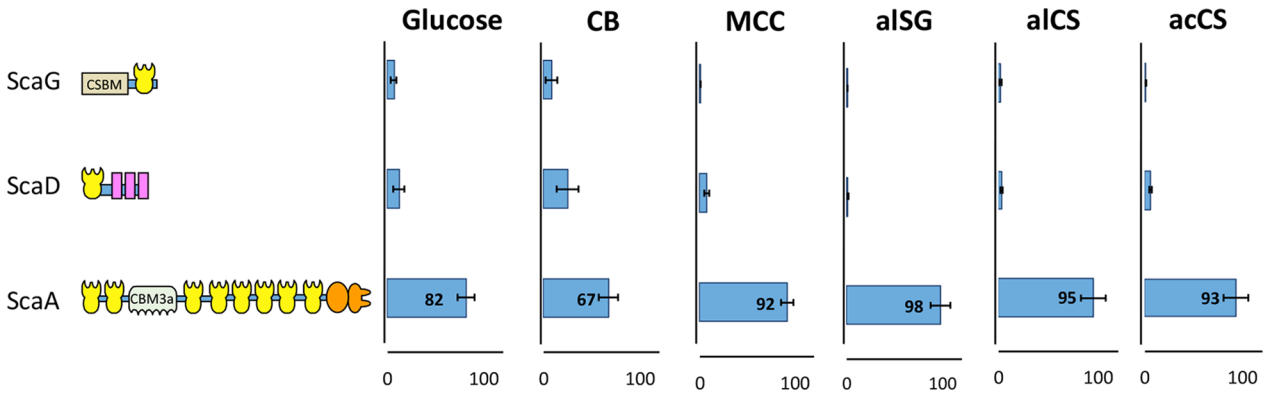

Fig. 4 Differential distribution of primary (type I cohesin-containing) scaffoldins. iBAQ intensities of the primary scaffoldins in each sample were summed up, and the relative proportions of each primary scaffoldin in each sample was calculated. The carbon source (see legend to Fig. 1 for definitions) used for the respective cellulosome production is shown in the graph

use of ScaA as an internal standard provides information regarding the composition of the cellulosome subpopulations within each sample (as described in previous proteomic studies [25, 26]). The values obtained reveal differences between samples, by comparing the relative abundance of the individual subunits. It is important to emphasize that ScaA is a glycosylated protein [57–60]. Glycosylation changes the molecular weight of a peptide, which interferes with its identification by the LC–MS/MS. Since iBAQ normalization takes into account the number of theoretical peptides, the iBAQ intensity of ScaA is probably underestimated. Nevertheless, the relative abundance index enables comparison *within* samples, and, assuming identical glycosylation of ScaA in all samples, *among* samples as well [25].

The table in Fig. 5 summarizes the relative abundance values of the cellulosomal subunits detected in the different samples. The twenty most abundant enzymes are marked and rated using the color scale shown in the table. The five most abundant proteins accounted for at least 50% of total cellulosomal subunits in each sample, with a higher percentage (60–64%) in the lignocellulosic biomass-derived cellulosomes, compared to those (50–56%) in the glucose-, cellobiose-, and MCC-derived cellulosomes. The most abundant subunits in all samples included exoglucanases Cel48S (Clo1313_2747) and Cel9K (Clo1313_1809), and endoglucanases Cel9Q (Clo1313_1603), Cel9R (Clo1313_16590), and Cel5G (Clo1313_0413), indicating their importance in biomass degradation. Out of the 20 most abundant catalytic subunits, 12 were common among all samples. The detected cellulosomal subunits were analyzed for their functional class distribution (Fig. 6; Additional file 3: Table S1). The cellulase (endo- and exoglucanase) portion was about 76–78% of the total cellulosomal subunits in all samples, except for the CB-derived samples, in which the cellulase portion was somewhat lower (71%). The endoglucanase-to-exoglucanase ratio was found to be higher in glucose-, cellobiose-, and MCC-derived cellulosomes (1.4, 2, and 1.65, respectively), compared to that of the lignocellulosic biomass-derived cellulosomes (0.71, 0.9 and 0.64 for alSG-, alCS-, and acCS-derived cellulosomes, respectively). Since glucose and MCC-derived cellulosomes showed higher specific activity, we may hypnotize that an endo-to-exo ratio of ~1.5 is recommended for assembly of cellulolytic cocktails based on *C. thermocellum* enzymes.

Exoglucanases

Excluding CB-derived cellulosomes, all samples revealed a similar pattern of exoglucanases, in which the relative abundance value of Cel48S was the highest, cell9K showed the second highest, followed by CbhA, and finally Cel5O with the lowest level of abundance. In the CB-derived cellulosomes, Cel9K showed the highest relative abundance among the four exoglucanases. Cel48S and CbhA showed lower levels, but the differences were not statistically significant. Similar trends of lower Cel48S abundance in the CB-derived cellulosomes compared to that of MCC-derived cellulosomes was previously reported for *C. thermocellum* ATCC 27405, both at the transcriptome and protein levels [21, 25, 61, 62], thus confirming our present data. The lignocellulosic biomass-derived cellulosomes revealed modest decreases in the relative abundance of Cel48S, compared to that of the MCC-derived cellulosomes. Only glucose-derived cellulosomes revealed a similar level of relative abundance of Cel48S to those of MCC-derived cellulosomes.

Cel9K showed similarly, higher relative abundance in the MCC- and glucose-derived cellulosomes, and the

			Glucose	CB	MCC	SG-AL	CS -AL	CS AC
Clo1313_2747	Cel48S	GH48	4.04E+00 ± 0.38	1.62E+00 ± 0.25	4.00E+00 ± 0.15	2.88E+00 ± 0.50	3.52E+00 ± 0.30	3.04E+00 ± 0.54
Clo1313_1809	Cel9K	CBM4,GH9	2.45E+00 ± 0.39	1.70E+00 ± 0.69	2.83E+00 ± 0.39	2.33E+00 ± 0.45	2.14E+00 ± 0.46	1.91E+00 ± 0.44
Clo1313_1603	Cel9Q	GH9,CBM3	2.09E+00 ± 0.38	1.25E+00 ± 0.34	2.66E+00 ± 0.06	6.65E-01 ± 0.05	1.23E+00 ± 0.30	5.29E-01 ± 0.06
Clo1313_1659	Cel9R	GH9,CBM3	7.55E-01 ± 0.13	1.32E+00 ± 0.17	1.69E+00 ± 0.32	4.59E-01 ± 0.03	6.93E-01 ± 0.15	4.57E-01 ± 0.09
Clo1313_0413	Cel5G	GH5	1.67E+00 ± 0.05	2.24E+00 ± 0.32	1.56E+00 ± 0.33	7.70E-01 ± 0.08	1.02E+00 ± 0.10	5.77E-01 ± 0.08
Clo1313_1477	Cel9W	GH9,CBM3	3.79E-01 ± 0.13	7.28E-01 ± 0.13	1.48E+00 ± 0.08	1.26E-01 ± 0.02	2.32E-01 ± 0.06	1.46E-01 ± 0.03
Clo1313_1398		GH5,CBM32	7.76E-02 ± 0.05	2.59E-02 ± 0.02	7.91E-01 ± 0.07	7.33E-02 ± 0.02	2.24E-01 ± 0.04	2.03E-01 ± 0.04
Clo1313_3023	Cel9U	GH9,CBM3,CBM3	1.04E+00 ± 0.25	2.77E-01 ± 0.12	5.93E-01 ± 0.01	2.85E-01 ± 0.02	4.12E-01 ± 0.10	2.64E-01 ± 0.05
Clo1313_0349	Cel9V	GH9,CBM3,CBM3	2.47E-01 ± 0.09	4.34E-01 ± 0.15	5.91E-01 ± 0.10	2.73E-01 ± 0.01	2.54E-01 ± 0.05	1.23E-01 ± 0.00
Clo1313_0400	Cel9T	GH9	4.84E-01 ± 0.04	2.35E-01 ± 0.03	5.56E-01 ± 0.00	1.66E-01 ± 0.02	3.58E-01 ± 0.06	2.14E-01 ± 0.03
Clo1313_1701	Cel5B	GH5	1.19E+00 ± 0.32	7.83E-01 ± 0.01	5.31E-01 ± 0.04	2.17E-01 ± 0.04	3.94E-01 ± 0.07	1.49E-01 ± 0.03
Clo1313_0177		CBM22,GH10	2.17E-01 ± 0.05	1.37E-01 ± 0.03	5.09E-01 ± 0.27	1.67E-01 ± 0.11	2.27E-01 ± 0.22	8.41E-02 ± 0.07
Clo1313_1604	Cel 9/44 J	CBM30,GH9,GH44,CBM44	6.08E-01 ± 0.13	2.56E-01 ± 0.06	4.86E-01 ± 0.07	2.81E-01 ± 0.04	2.95E-01 ± 0.09	1.79E-01 ± 0.03
Clo1313_1808	CbhA	CBM4,GH9,CBM3	8.66E-01 ± 0.11	9.32E-01 ± 0.35	4.02E-01 ± 0.06	2.90E-01 ± 0.02	6.82E-01 ± 0.18	3.30E-01 ± 0.08
Clo1313_2189	Cel9N	GH9,CBM3	3.31E-02 ± 0.01	1.39E-01 ± 0.03	3.87E-01 ± 0.04	1.04E-01 ± 0.04	1.76E-01 ± 0.03	1.17E-01 ± 0.01
Clo1313_2635		CE1,CBM6,GH10	3.12E-01 ± 0.06	1.42E-01 ± 0.03	3.65E-01 ± 0.02	6.29E-02 ± 0.01	1.13E-01 ± 0.01	6.34E-02 ± 0.00
Clo1313_1955	Cel9P	GH9	4.52E-01 ± 0.05	3.04E-01 ± 0.08	3.31E-01 ± 0.02	6.90E-02 ± 0.01	1.50E-01 ± 0.01	1.15E-01 ± 0.02
Clo1313_0399		CBM35,GH26	3.88E-02 ± 0.01	3.52E-02 ± 0.00	3.28E-01 ± 0.02	1.07E-01 ± 0.03	1.34E-01 ± 0.02	8.28E-02 ± 0.01
Clo1313_1960	Cel8A	GH8	5.83E-01 ± 0.10	3.57E-01 ± 0.02	3.07E-01 ± 0.02	1.01E-01 ± 0.00	1.58E-01 ± 0.02	8.11E-02 ± 0.02
Clo1313_0350		GH9,CBM3	1.22E-01 ± 0.03	3.77E-01 ± 0.19	2.97E-01 ± 0.00	1.56E-01 ± 0.02	1.61E-01 ± 0.03	7.33E-02 ± 0.01
Clo1313_1788		GH9,CBM3	6.06E-01 ± 0.03	3.64E-01 ± 0.18	2.87E-01 ± 0.03	4.07E-01 ± 0.01	3.89E-01 ± 0.09	1.73E-01 ± 0.04
Clo1313_2530	XynC	CBM22,GH10	5.05E-01 ± 0.08	8.75E-01 ± 0.05	2.70E-01 ± 0.04	3.13E-01 ± 0.04	2.93E-01 ± 0.01	2.11E-01 ± 0.03
Clo1313_1694	Cel9F	GH9,CBM3	2.97E-01 ± 0.12	1.96E-01 ± 0.05	1.93E-01 ± 0.02	2.06E-01 ± 0.07	1.35E-01 ± 0.06	2.77E-01 ± 0.02
Clo1313_2693			3.57E-01 ± 0.09	3.24E-01 ± 0.09	1.84E-01 ± 0.06	1.03E-01 ± 0.01	1.12E-01 ± 0.02	8.35E-02 ± 0.01
Clo1313_0521	Xyn11A	GH11,CBM6,CE4	6.91E-02 ± 0.04	1.42E-01 ± 0.01	1.76E-01 ± 0.04	1.32E-01 ± 0.05	1.38E-01 ± 0.03	8.56E-02 ± 0.04
Clo1313_1396	Cel9D	GH9	1.77E-01 ± 0.05	1.63E-01 ± 0.09	1.54E-01 ± 0.00	3.97E-02 ± 0.01	6.69E-02 ± 0.00	6.45E-02 ± 0.01
Clo1313_1816	Cel5L	GH5	1.82E-01 ± 0.05	4.56E-02 ± 0.00	1.52E-01 ± 0.01	2.37E-02 ± 0.01	5.22E-02 ± 0.01	3.18E-02 ± 0.01
Clo1313_1305		CBM22,GH10,CBM22,CE1	2.36E-01 ± 0.03	1.82E-01 ± 0.00	1.09E-01 ± 0.01	1.00E-02 ± 0.00	2.33E-02 ± 0.01	1.75E-02 ± 0.01
Clo1313_0851	Xgh74A	GH74	1.40E-01 ± 0.06	8.91E-02 ± 0.03	8.65E-02 ± 0.02	1.05E-01 ± 0.03	2.05E-01 ± 0.01	6.65E-02 ± 0.03
Clo1313_1587			6.35E-02 ± 0.01	2.13E-01 ± 0.04	4.60E-02 ± 0.01	3.52E-02 ± 0.01	6.56E-02 ± 0.01	1.99E-02 ± 0.00
Clo1313_2805	Cel5O	CBM3,GH5	4.25E-02 ± 0.01	1.10E-01 ± 0.04	4.39E-02 ± 0.00	2.55E-02 ± 0.00	3.86E-02 ± 0.01	2.08E-02 ± 0.00
Clo1313_2043			1.52E-03 ± 0.00	8.89E-03 ± 0.00	3.15E-02 ± 0.02	1.68E-02 ± 0.00	1.49E-02 ± 0.00	1.22E-02 ± 0.00
Clo1313_2564			1.87E-02 ± 0.01	6.94E-03 ± 0.00	3.15E-02 ± 0.01	4.30E-03 ± 0.00	2.46E-03 ± 0.00	2.05E-04 ± 0.00
Clo1313_1425	Cel5E	GH5,CE2	1.42E-02 ± 0.01	5.37E-03 ± 0.00	2.38E-02 ± 0.00	1.73E-02 ± 0.01	1.51E-02 ± 0.01	2.26E-02 ± 0.01
Clo1313_0849		GH53	1.03E-02 ± 0.00	1.24E-02 ± 0.00	2.36E-02 ± 0.00	8.46E-03 ± 0.00	1.01E-02 ± 0.00	4.41E-03 ± 0.00
Clo1313_1563		GH43,CBM13	1.58E-02 ± 0.01	3.25E-02 ± 0.00	2.29E-02 ± 0.02	7.14E-04 ± 0.00	7.00E-04 ± 0.00	2.37E-03 ± 0.00
Clo1313_2202		CBM35,GH26	1.51E-01 ± 0.05	2.41E-02 ± 0.00	2.22E-02 ± 0.01	3.79E-02 ± 0.01	9.20E-02 ± 0.02	2.56E-02 ± 0.01
Clo1313_2856		GH5,CBM6,CBM13,CBM62	6.01E-02 ± 0.01	1.30E-01 ± 0.03	1.95E-02 ± 0.00	1.46E-02 ± 0.00	1.92E-02 ± 0.00	3.08E-02 ± 0.00
Clo1313_0522		GH11,CBM6	2.62E-03 ± 0.00	5.65E-03 ± 0.00	1.87E-02 ± 0.00	5.79E-03 ± 0.00	6.88E-03 ± 0.00	3.21E-03 ± 0.00
Clo1313_0420		CEnc	3.91E-03 ± 0.00	4.32E-03 ± 0.00	1.69E-02 ± 0.00	4.16E-04 ± 0.00	9.93E-04 ± 0.00	1.03E-03 ± 0.00
Clo1313_0689			4.02E-03 ± 0.00	4.27E-02 ± 0.00	1.38E-02 ± 0.00	3.28E-02 ± 0.01	1.88E-02 ± 0.01	2.67E-02 ± 0.01
Clo1313_0563		GH30,CBM6	1.60E-02 ± 0.00	8.53E-02 ± 0.01	1.23E-02 ± 0.00	3.42E-03 ± 0.00	8.75E-03 ± 0.00	8.82E-03 ± 0.00
Clo1313_1959		GH18	1.31E-02 ± 0.00	7.77E-03 ± 0.00	1.22E-02 ± 0.01	2.89E-03 ± 0.00	7.62E-03 ± 0.00	6.05E-03 ± 0.00
Clo1313_1983		CBM35,PL11	7.82E-03 ± 0.00	6.20E-03 ± 0.00	1.17E-02 ± 0.00	1.74E-03 ± 0.00	1.67E-03 ± 0.00	2.42E-03 ± 0.00
Clo1313_1990			2.33E-02 ± 0.01	1.40E-02 ± 0.00	1.07E-02 ± 0.01	1.49E-02 ± 0.00	1.30E-02 ± 0.00	5.04E-02 ± 0.02
Clo1313_2795		GH30,CBM42,GH43	3.96E-03 ± 0.00	2.66E-03 ± 0.00	1.01E-02 ± 0.00	5.84E-03 ± 0.00	4.98E-03 ± 0.00	4.06E-03 ± 0.00
Clo1313_2793		GH39,CBM35,CBM35	6.83E-02 ± 0.01	2.20E-02 ± 0.00	9.62E-03 ± 0.00	9.60E-04 ± 0.00	1.98E-03 ± 0.00	4.99E-03 ± 0.00
Clo1313_0987		GH43,CBM6,CBM6	6.17E-02 ± 0.01	2.28E-02 ± 0.00	9.29E-03 ± 0.00	1.29E-03 ± 0.00	9.72E-03 ± 0.01	4.28E-03 ± 0.00
Clo1313_1971			1.73E-03 ± 0.00	1.47E-02 ± 0.00	7.93E-03 ± 0.00	8.70E-03 ± 0.01	5.05E-03 ± 0.00	1.20E-02 ± 0.00
Clo1313_0685			2.86E-03 ± 0.00	2.23E-02 ± 0.01	7.88E-03 ± 0.00	1.75E-02 ± 0.00	1.49E-02 ± 0.01	1.96E-02 ± 0.01
Clo1313_0501		PL1,CBM35	7.30E-03 ± 0.00	4.44E-03 ± 0.00	6.46E-03 ± 0.00	3.06E-04 ± 0.00	2.67E-04 ± 0.00	5.63E-04 ± 0.00
Clo1313_2857		GH43,CBM6	2.43E-02 ± 0.00	1.47E-02 ± 0.00	5.57E-03 ± 0.00	1.19E-04 ± 0.00	1.38E-03 ± 0.00	3.23E-04 ± 0.00
Clo1313_1564		GH81	9.86E-03 ± 0.00	9.22E-03 ± 0.00	5.02E-03 ± 0.01	4.17E-03 ± 0.00	4.72E-03 ± 0.00	2.08E-03 ± 0.00
Clo1313_2022		GH16	2.58E-02 ± 0.01	1.39E-02 ± 0.00	5.01E-03 ± 0.00	8.56E-03 ± 0.00	6.71E-03 ± 0.00	2.88E-03 ± 0.00
Clo1313_1424		CE3,CE3	1.47E-03 ± 0.00	3.01E-03 ± 0.00	3.35E-03 ± 0.00	2.48E-03 ± 0.00	2.29E-03 ± 0.00	7.42E-02 ± 0.02
Clo1313_2479			3.89E-05 ± 0.00	2.09E-04 ± 0.00	1.91E-03 ± 0.00	1.73E-03 ± 0.00	1.25E-03 ± 0.00	2.74E-03 ± 0.00
Clo1313_1786		GH124	7.30E-03 ± 0.01	3.37E-02 ± 0.00	1.76E-03 ± 0.00	6.32E-04 ± 0.00	7.67E-04 ± 0.00	1.33E-03 ± 0.00
Clo1313_2188			8.88E-05 ± 0.00	7.34E-05 ± 0.00	1.62E-03 ± 0.00	1.40E-04 ± 0.00	1.26E-04 ± 0.00	1.28E-04 ± 0.00
Clo1313_1494			2.86E-03 ± 0.00	1.89E-04 ± 0.00	9.57E-04 ± 0.00	4.79E-03 ± 0.00	2.42E-03 ± 0.00	9.79E-03 ± 0.00
Clo1313_0693		CE12,CBM35,CE12	1.92E-04 ± 0.00	1.09E-04 ± 0.00	3.08E-04 ± 0.00	1.84E-04 ± 0.00	7.09E-05 ± 0.00	1.20E-03 ± 0.00
Clo1313_2858		CE1,CBM6	4.24E-03 ± 0.00	1.85E-03 ± 0.00	2.88E-04 ± 0.00	1.33E-04 ± 0.00	3.30E-04 ± 0.00	5.54E-04 ± 0.00
Clo1313_2859		CBM6	3.54E-03 ± 0.00	3.82E-04 ± 0.00	2.79E-04 ± 0.00	1.84E-05 ± 0.00	1.94E-04 ± 0.00	1.00E-04 ± 0.00
Clo1313_2234	Cel 26/5 H	GH26,GH5,CBM11	2.14E-03 ± 0.00	2.68E-03 ± 0.00	2.40E-04 ± 0.00	2.13E-03 ± 0.00	7.63E-03 ± 0.00	4.25E-03 ± 0.00
Clo1313_2794		CBM42,GH43	3.70E-04 ± 0.00	1.28E-03 ± 0.00	1.58E-04 ± 0.00	2.95E-05 ± 0.00	4.90E-04 ± 0.00	5.39E-04 ± 0.00
Clo1313_2861		GH2,CBM6	7.83E-04 ± 0.00	2.07E-04 ± 0.00	6.99E-05 ± 0.00	6.04E-06 ± 0.00	1.00E-04 ± 0.00	8.90E-06 ± 0.00
Clo1313_2860		GH43,CBM6	4.10E-04 ± 0.00	2.34E-04 ± 0.00	0.00E+00 ± 0.00	0.00E+00 ± 0.00	2.03E-04 ± 0.00	0.00E+00 ± 0.00
Clo1313_2216		CBM42,GH43	4.07E-04 ± 0.00	0.00E+00 ± 0.00	0.00E+00 ± 0.00	7.56E-05 ± 0.00	1.83E-04 ± 0.00	5.68E-05 ± 0.00

Fig. 5 Cellulosomal subunit composition. Subunit compositions of the different cellulosomes were analyzed by label-free LC–MS/MS mass spectrometry. The intensities were normalized by intensity-based absolute quantification (iBAQ) method. The resultant iBAQ intensities of type I dockerin-containing subunits were divided by the iBAQ intensity of ScaA in each sample, thereby generating a relative abundance index. Average and standard deviations of duplicate samples of CB- and MCC-derived cellulosomes and triplicates of glucose-, aISG-, aICS-, and acCS-derived cellulosomes were analyzed. The 20 most abundant subunits in each sample were rated by a colored scale. Gene ID and CAZY annotation of the subunits are mentioned in the table too. GH, glycoside hydrolase; CBM, carbohydrate-binding module; CE, carbohydrate esterase

lowest level in CB-derived cellulosomes. Nevertheless, these differences were not statistically significant (A similar pattern was reported for *C. thermocellum* ATCC 27450 [26]). A different pattern was found for CbhA; however, which revealed higher relative abundance in the CB- and glucose-derived cellulosomes compared to

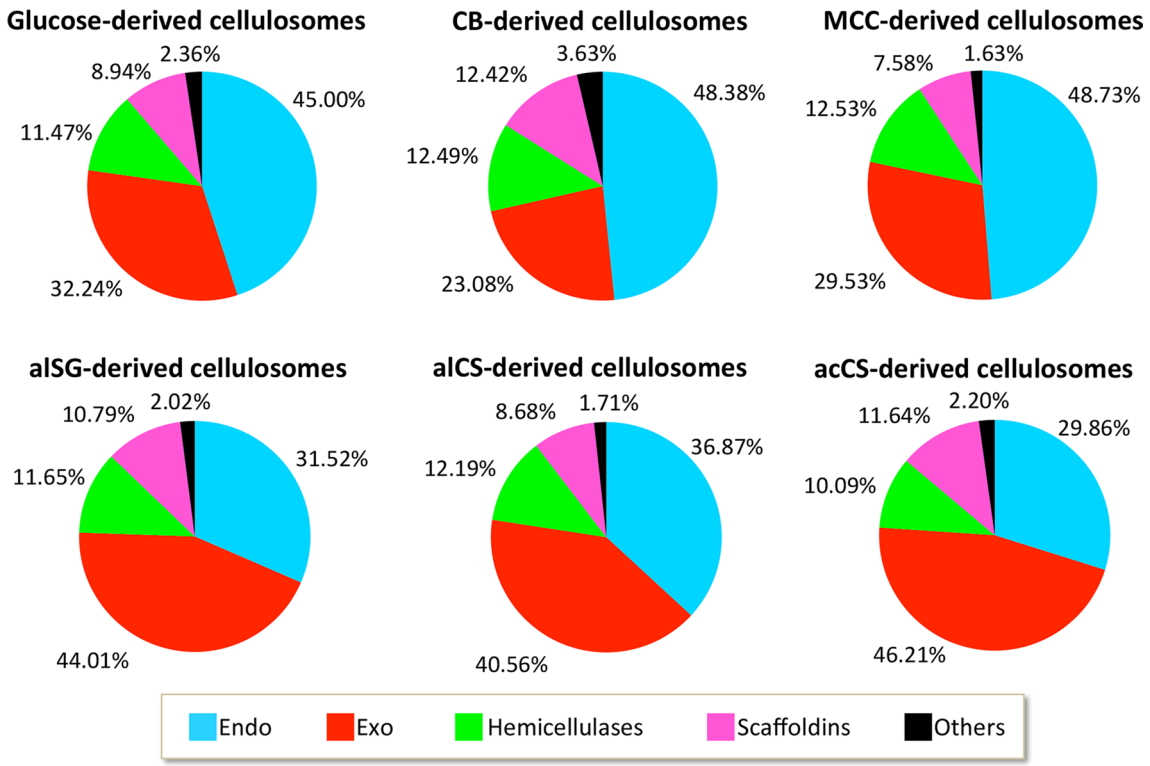

Fig. 6 Cellulosomal subunit distribution: Relative abundance of all type I dockerin-containing subunits were summed up and defined as 100%. The subunits were divided into five groups according to their activity as detailed in Additional file 3: Table S1. The relative proportion of each group was calculated. Blue, endoglucanase; red, exoglucanase; green, hemicellulase; pink, scaffoldins; and black, others

the MCC- and the lignocellulosic biomass-derived cellulosomes. Cel9 K and CbhA are tandem genes, containing a similar type of modular structure (except the lack of X1-like domains and the CBM3 domain in Cel9K), with 94% identity among the modules, and therefore suggested to arose by gene duplication [63]. Due to the similarity in enzyme structure, it would be tempting to assume a similar function. Thus, the non-similar abundance pattern of the two proteins was somewhat surprising. The different role of CbhA and Cel9 K remains to be explored.

Endoglucanases

Among the different cellulosomal samples examined in this work, the twenty most abundant catalytic endoglucanase subunits always included the following enzymes: Cel9Q, Cel9R, Cel5G, Cel9W, Cel9U, Cel9V, Cel9T, Cel5B, and Cel9/44J. In recent preliminary data (Additional file 4: Table S2), Cel9/44J and Cel9D exhibited the highest endoglucanase activity of 25 *C. thermocellum* cellulosomal subunits examined. Surprisingly, Cel9D showed low relative abundance and therefore did not appear in the latter list in any of the analyzed samples. The data support the above claim that the cellulosomal subunit composition, derived from cells grown on the different substrates, is not necessarily optimized for efficient substrate hydrolysis.

Type I dockerin-to-ScaA ratio

Type I dockerin-to-ScaA (type I dockerin/scaffoldin) ratio was calculated by summing up the relative abundance values for all of the type I dockerin-containing subunits in each sample. The ratio was almost double in the MCC- and glucose-derived cellulosomes (22.7 and 20.9 molecules per ScaA, respectively) compared to those of the alSG- and acCS-derived cellulosomes (11.2 and 10.1 molecules per ScaA, respectively) and higher than the alCS- and CB-derived cellulosomes (14.3 and 16.5 molecules per ScaA, respectively). The results revealed higher numbers of cellulosomal subunits per scaffoldin in the MCC- and glucose-derived cellulosomes, which might thus explain the observed higher hydrolytic activity of the latter cellulosomes.

Soluble carbohydrate-active enzymes

Besides cellulosomal (dockerin-containing) enzymes, *C. thermocellum* produces a repertoire of soluble carbohydrate-active enzymes without dockerins as well. In order to characterize these non-cellulosomal enzymes

and their relative content in the different samples, MS results were analyzed and compared to the *C. thermocellum* DSM1313 CAZY database (for GH/PL/GT/CE/CBM containing enzymes). In total, 28 additional non-cellulosomal carbohydrate-active enzymes were identified in the various cellulosome samples (Additional file 5: Table S3). The presence of un-complexed subunits was also supported by the detection of the structural scaffoldin ScaG. The present results are in line with previously described reports for *C. clariflavum* cellulosomes [40]. The detection of soluble carbohydrate-active enzymes in the high-molecular-weight fractions may reflect the result of non-specific interactions or unknown specific interactions that are not mediated by cohesin-dockerin interactions.

Unique expression patterns

Three cellulosomal subunits were significantly upregulated in cellulosomes derived from the defined cellulosic substrates, that is, glucose-, cellobiose-, and MCC-derived cellulosomes: Clo1313_1305 (CBM22-GH10-CBM22-dockerin-CE1), Clo1313_1563 (GH43-CBM13-dockerin), and Clo1313_0501 (PL1-dockerin-CBM35). Comparison of cellulosomes derived from the soluble substrates (i.e., cellobiose and glucose) versus those from the insoluble substrates revealed significant upregulation of the non-cellulosomal protein Clo1313_0397 (SLH-CBM54-GH16-CBM4-CBM4-CBM4-CBM4-CBM4) and the cellulosomal subunit Clo1313_2857 (GH43-CBM6-dockerin). The cellulosomal subunit Clo1313_2479 (containing no identified CAZy module) was significantly downregulated. The role of these enzymes is as yet unknown.

Comparison of *C. thermocellum* strains

The 'classic' reference, *C. thermocellum* strain ATCC 27405, was intensively studied in the past at the proteomic and transcriptomic level, while growing the cells on several different carbon sources [21, 25, 28, 39]. A recent release of the complete *C. thermocellum* strain DSM 1313 genome sequence [47] enabled differential omic studies, including both transcriptome and proteome. Moreover, the only successful gene-directed mutagenesis has been reported for strain DSM 1313, while strain ATCC 27405 is not easily amenable to genetic manipulations [64]. The composition of MCC-derived cellulosomes from strain DSM1313 reported here showed differences in subunit composition compared to that previously described for equivalent preparations from strain ATCC 27405. At least nine undetected cellulosomal enzymes in strain ATCC 27405 cellulosome [25] were detected in this study using the DSM 1313 strain: Clo1313_3023 (Cel9U, which was also detected earlier by Zverlov et al. [65]), Clo1313_2793 (GH39), Clo1313_1564 (GH81), Clo1313_0501 (PL10, Clo1313_2794, Clo1313_2858 (CE1), Clo1313_2859 (GH141), Clo1313_2860 (GH43), and Clo1313_2861 (GH2) (Fig. 5). The last five proteins showed relatively low but measurable relative abundance. In this context, Clo1313_2858, Clo1313_2859, Clo1313_2860, and Clo1313_2861 are located on the same operon. The orthologous operon in strain ATCC 27405 was found to be disrupted by a putative 2419 bp insertion sequence element located within the 5′ end of the Clo1313_2861 orthologue (annotated as two different genes: Cthe_2197 and Cthe2200). Such an insertion sequence does not exist in the DSM 1313 strain. The most abundant endoglucanases in the DSM1313 cellulosome, i.e., Cel9Q, Cel9R, Cel5G, and Cel9W, showed 3.7-fold to fivefold higher relative abundance, compared to the strain ATCC 27405-derived cellulosome [25]. In contrast, Cel8A, and Cel5E showed relatively lower abundance (fivefold and tenfold, respectively) in the DSM1313 cellulosomes, compared to those of strain ATCC 27405. Xyn11A (Clo1313_0521), which was the third most abundant protein in the ATCC 27405 cellulosome system, showed fivefold lower relative abundance in the DSM1313 system. We assume that higher expression of the Xyn11A homologue in ATCC 27405 (Cthe_2972) might be explained by a different gene organization in the two *C. thermocellum* genomes. While the *xyn11A* gene of DSM 1313 is a second ORF of a two-cistronic operon (*xyn11B*/Clo1313_0522-*xyn11A*/Clo1313_0521), the ATCC 27405 genome has only one gene, *xyn11A*, while a *xyn11B* homologue appears to be omitted from this genome.

Conclusions

Our results demonstrate that cells grown on a given lignocellulosic substrate do not necessarily produce cellulosomes that exhibit enhanced activity levels on that particular substrate. Surprisingly, MCC- and glucose-derived cellulosomes showed superior performance even towards degradation of complex lignocellulosic substrates. The latter cellulosome preparations exhibited distinctive characteristics, e.g., elevated endoglucanase-to-exoglucanase ratios and enzyme-versus-ScaA ratios. The most abundant subunits in all tested cellulosomes included Cel48S, Cel9K, Cel9Q, Cel9R, and Cel5G, indicating their preferential contribution and importance to deconstruction of complex cellulosic substrates. Our results should be implemented in the future for fabrication of efficient designer cellulosomes, for formulation of recombinant cellulolytic cocktails based on *C. thermocellum* enzymes or for engineering *C. thermocellum* strains with improved lignocellulosic biomass-converting abilities.

Additional files

Additional file 1: Figure S1. Purification profile of the different cellulosomes by Gel filtration chromatography. *C. thermocellum* growth media were centrifuged (10,900 g, 7 min), and the supernatant fluids were carefully removed from the pellet and concentrated 40 times using a Pellicon XL biomax 300 cassette (Millipore, Cat. No. PXB300C50). Concentrated samples were fractionated by size exclusion chromatography using a SuperdexS-200 prep grade 16/60 gel filtration column (GE Healthcare). (A) Chromatogram of glucose-derived cellulosomes (B) Chromatogram of CB-derived cellulosomes (C) Chromatogram of MCC-derived cellulosomes (D) Chromatogram of alSG-derived cellulosomes (E) Chromatogram of alCS-derived cellulosomes (F) Chromatogram of acCS-derived cellulosomes. Fractions containing the cellulosomes (according to SDS-PAGE analysis) are marked by black arrows. All cellulosomes-containing fractions were eluted immediately after void value due to the separation range of the column.

Additional file 2: Figure S2. Calibration of the near-linear range of the various substrate hydrolyses. Increased cellulosome dosages were applied on (**A**) 7% microcrystalline cellulose [MCC], (**B**) 5% alkaline-pretreated switch grass [alSG], (**C**) 5% alkaline-pretreated corn stover [alCS] and (**D**) 5% dilute acid-pretreated corn stover [acCS], and the samples were incubated overnight at 70 °C. Released sugar concentrations were measured by *dinitro*salicylic acid (DNS) method, as previously described [36]. All assays were performed with the addition of 0.33 mg/ml equivalent of *Thermoanaerobacter brockii* β-glucosidase (CglT) in order to prevent feedback inhibition. Enzyme loadings of 20, 50, 3 and 50 μg/ml for MCC, alSG, alCS and acCS hydrolysis assays, respectively, were chosen for activity measurements (Black arrows).

Additional file 3: Table S1. Functional classes distribution of *C. thermocellum* DSM 1313 cellulosomal subunits. Cellulosomal subunits detected in the *C. themocellum D*SM 1313 proteome were sorted into five functional classes: endoglucanases, exoglucanases, hemicellulases, scaffoldins and others.

Additional file 4: Table S2. Activities of *C. thermocellum* cellulosomal enzymes. Recombinant *C. themocellum* cellulosomal (type I dockerin-containing) enzymes were a kind gift of CelDezyner Ltd. (Rehovot, Israel). Activity assays were conducted in a final volume of 1 ml, containing 50 mM acetate buffer, 1% carboxymethyl cellulose (sodium salt, low viscosity CMC, BDH chemicals) and 7 nM enzyme. Samples were incubated with shaking for 3 h at 60 °C. Released soluble sugar (reducing ends) concentrations were analyzed by the dinitrosalicylic acid (DNS) method, as previously described [36]. Final soluble sugar concentrations were determined against a glucose calibration curve, and CMCase activities [μM reducing ends·μmol enzyme^{-1}·min^{-1}] were calculated. *A thermostable clone of Cel8A [66] was used.

Additional file 5: Table S3. Soluble (non-cellulosomal) carbohydrate-active enzyme composition. Soluble (dockerin-lacking) carbohydrate-active enzyme compositions of the different cellulosomes were analyzed by label-free LC–MS/MS mass spectrometry. The intensities were normalized by the intensity-based absolute quantification (iBAQ) method. The resultant iBAQ intensities were divided by the iBAQ intensity of ScaA in each sample, thereby generating a relative abundance index. Standard deviations of duplicate samples of CB- and MCC-derived cellulosomes and triplicates of glucose-, alSG-, alCS-, and acCS-derived cellulosomes were analyzed. Gene ID and CAZy annotation of the subunits are designated. Acronyms: GH, glycoside hydrolase; CBM, carbohydrate-binding module; CE, carbohydrate esterase; GT, glycosyl transferase.

Abbreviations

CB: cellobiose; MCC: microcrystalline cellulose; alSG: alkaline-pretreated switch grass; alCS: alkaline-pretreated corn stover; acCS: dilute acid-pretreated corn stover; MS: mass spectrometry; iBAQ: intensity-based absolute quantification; CBM: carbohydrate-binding module; GH: glycoside hydrolase; CE: carbohydrate esterase; GT: glycosyl transferase; PL: polysaccharide lyase; Sca: scaffoldin; SLH: S-layer homology; Xyn: xylanase; SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis.

Authors' contributions

SY, YH, EM, and EAB designed the research. SY performed the experiments and analyzed the results. SY, YB, and MS purified the cellulosomes. IB and RL analyzed the genome data. SY and BD analyzed mass spectrometry data. SY, EM, and EAB wrote the manuscript. All authors read and approved the final manuscript.

Author details

[1] Department of Plant Pathology and Microbiology, Robert H. Smith Faculty of Agriculture, Food and Environment, The Advanced School for Environmental Studies, The Hebrew University of Jerusalem, 76100 Rehovot, Israel. [2] Designer Energy Ltd, 2 Bergman Street, Rehovot, Israel. [3] Bio-Nano Unit, Chemical Research Support, The Weizmann Institute of Science, 761000 Rehovot, Israel. [4] Department of Biomolecular Sciences, The Weizmann Institute of Science, 76100 Rehovot, Israel. [5] Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Israel.

Acknowledgements

The authors appreciate the contribution of the Smoler Proteomics Center, Faculty of Biology, Technion, Israel, and especially the contribution of Keren Bendalak for the proteomic analysis. The authors also acknowledge CelDezyner LTD, Rehovot, Israel, for supplying thermostable β-glucosidase samples and *C. thermocellum* recombinant enzymes.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Not applicable.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

This research was supported by a research grant from the Ministry of Science, Technology & Space, Israel. In addition, Grant No. 1349 from the Israel Science Foundation (ISF), Jerusalem, Israel, is gratefully acknowledged. The authors also appreciate the support of the European Union, Area NMP.2013.1.1-2: self-assembly of naturally occurring nanosystems: CellulosomePlus Project Number: 604530 and European Union Horizon 2020 contract: Sustainable production of next generation biofuels from waste streams: Waste2Fuels. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

Publisher's Note

Received: 8 June 2017 Accepted: 7 September 2017
Published online: 18 September 2017

References

1. Goldemberg J. Ethanol for a sustainable energy future. Science. 2007;315:808–10.
2. Hadar Y. Sources for lignocellulosic raw materials for the production of ethanol. In: Faraco V, editor. Lignocellulose conversion. Berlin: Springer; 2013. p. 21–39.
3. Sun Y, Cheng J. Hydrolysis of lignocellulosic materials for ethanol production: a review. Bioresour Technol. 2002;83:1–11.

4. Bayer EA, Lamed R, Himmel ME. The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol. 2007;18:237–45.
5. Mittal A, Decker SR. Special issue: application of biotechnology for biofuels: transforming biomass to biofuels. 3Biotech. 2013;3:341–3.
6. Reijnders L. Ethanol production from crop residues and soil organic carbon. Resour Conserv Recycl. 2008;52:653–8.
7. Pettolino FA, Walsh C, Fincher GB, Bacic A. Determining the polysaccharide composition of plant cell walls. Nat Protoc. 2012;7:1590–607.
8. Pauly M, Keegstra K. Cell-wall carbohydrates and their modification as a resource for biofuels. Plant J. 2008;54:559–68.
9. Viikari L, Vehmaanperä J, Koivula A. Lignocellulosic ethanol: from science to industry. Biomass Bioenerg. 2012;46:13–24.
10. Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnol Bioeng. 2012;109:1083–7.
11. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. Novel enzymes for the degradation of cellulose. Biotechnol Biofuels. 2012;5:45.
12. Himmel ME, Xu Q, Luo Y, Ding S-Y, Lamed R, Bayer EA. Microbial enzyme systems for biomass conversion: emerging paradigms. Biofuels. 2010;1:323–41.
13. Koeck DE, Koellmeier T, Zverlov VV, Liebl W, Schwarz WH. Differences in biomass degradation between newly isolated environmental strains of *Clostridium thermocellum* and heterogeneity in the size of the cellulosomal scaffoldin. Syst Appl Microbiol. 2015;38:424–32.
14. Fontes CMGA, Gilbert HJ. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem. 2010;79:655–81.
15. Demain AL, Newcomb M, Wu JHD. Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev. 2005;69:124–54.
16. Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev. 2002;66:506–77.
17. Lynd LR, Van Zyl WH, Mcbride JE, Laser M. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol. 2005;16:577–83.
18. Bayer EA, Lamed R, White BA, Flint HJ. From cellulosomes to cellulosomics. Chem Rec. 2008;8:364–77.
19. Kosugi A, Murashima K, Tamaru Y, Doi RH. Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J Bacteriol. 2002;184:884–8.
20. Fierobe H, Bayer EA, Tardif C, Czjzek M, Mechaly A, Lamed R, et al. Degradation of cellulose substrates by cellulosome chimeras. J Biol Chem. 2002;277:49621–30.
21. Raman B, Mckeown CK Jr, Rodriguez M, Brown SD, Mielenz JR. Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. BMC Microbiol. 2011;11:134.
22. Hazlewood GP, Davidson K, Laurie JI, Huskisson NS, Gilbert HJ. Gene sequence and properties of CelI, a family E endoglucanase from *Clostridium thermocellum*. J Gen Microbiol. 1993;139:307–16.
23. Fuchs K, Zverlov VV, Velikodvorskaya GA, Lottspeich F, Schwarz WH. Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-β-1, 3-glucanase bound to the outer cell surface. Microbiology. 2003;149:1021–31.
24. Berger E, Zhang D, Zverlov VV, Schwarz WH. Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. FEMS Microbiol Lett. 2007;268:194–201.
25. Raman B, Pan C, Hurst GB, Rodriguez M, Mckeown CK, Lankford PK, et al. Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PLoS ONE. 2009;4:e5271.
26. Gold ND, Martin VJJ. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. J Bacteriol. 2007;189:6787–95.
27. Wilson CM Jr, Rodriguez M, Johnson CM, Martin SL, Chu TM, Wolfinger RD, et al. Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass. Biotechnol Biofuels. 2013;6:179.
28. Wei H, Fu Y, Magnusson L, Baker JO, Maness P, Xu Q, et al. Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. Front Microbiol. 2014;5:142.
29. Johnson EA, Bouchot F, Demain AL. Regulation of cellulase formation in *Clostridium thermocellum*. J Gen Microbiol. 1985;131:2303–8.
30. Hernandez PE. Transport of D-glucose in *Clostridium thermocellum* ATCC-27405. J Gen Appl Microbiol. 1982;28:469–77.
31. Strobel HJ, Caldwell FC, Dawson KA. Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQRI. Appl Environ Microbiol. 1995;61:4012–5.
32. Ioelovich M, Morag E. Study of enzymatic hydrolysis of mild pretreated lignocellulosic biomass. BioResources. 2012;7:1040–52.
33. Ioelovich M. Methods for determination of chemical composition of plant biomass. J SITA. 2015;17:208–14.
34. TAPPI. Test method T 222 om-02. Acid-insoluble lignin in wood and pulp. Atlanta: TAPPI; 2002.
35. Bayer EA, Kenig R, Lamed R. Adherence of *Clostridium thermocellum* to cellulose. J Bactertiol. 1983;156:818–27.
36. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Biochem. 1959;31:426–8.
37. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26:1367–72.
38. Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, et al. Global quantification of mammalian gene expression control. Nature. 2011;473:337–42.
39. Rydzak T, Mcqueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, et al. Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiol. 2012;12:214.
40. Artzi L, Morag E, Barak Y, Lamed R, Bayer EA. *Clostridium clariflavum*: key cellulosome players are revealed by proteomic analysis. mBio. 2015;6:e00411–5.
41. De Philip P, Lignon S, Tardif C, Page S. Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. Proteomics. 2010;10:541–54.
42. Bayer EA, Morag E, Lamed R. The cellulosome—a treasure-trove for biotechnology. Trends Biotechnol. 1994;12:379–86.
43. Lamed R, Kenig R, Setter E, Bayer EA. Major characteristics of the cellulolytic system of *Clostridium thermocellum* coincide with those of the purified cellulosome. Enzyme Microb Technol. 1985;7:37–41.
44. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–5.
45. Brás JLA, Pinheiro BA, Cameron K, Cuskin F, Viegas A, Najmudin S, et al. Diverse specificity of cellulosome attachment to the bacterial cell surface. Sci Rep. 2016;6:38292.
46. Fierobe H-P, Mingardon F, Mechaly A, Belaich A, Rincon MT, Pages S, et al. Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J Biol Chem. 2005;280:16325–34.
47. Feinberg L, Foden J, Barrett T, Davenport KW, Bruce D, Detter C, et al. Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. J Bacteriol. 2011;193:2906–7.
48. Dassa B, Borovok I, Lamed R, Henrissat B, Coutinho P, Hemme CL, et al. Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. BMC Genom. 2012;13:210.
49. Hong W, Zhang J, Feng Y, Mohr G, Lambowitz AM, Cui G, et al. The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. Biotechnol Biofuels. 2014;7:1–16.
50. Gerngross UT, Romaniec MPM, Kobayashi T, Huskisson NS, Demain AL. Sequencing of a Clostridium thermocellum gene (cipA) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. Mol Microbiol. 1993;8:325–34.
51. Ding SY, Liu YS, Zeng Y, Himmel ME, Baker JO, Bayer EA. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? Science. 2012;338:1055–60.
52. Xu Q, Resch MG, Podkaminer K, Yang S, Baker JO, Donohoe BS, et al. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. Sci Adv. 2016;:e1501254.
53. Olson DG, Giannone RJ, Hettich RL, Lynd LR. Role of the CipA scaffoldin protein in cellulose solubilization, as determined by targeted gene

deletion and complementation in *Clostridium thermocellum*. J Bacteriol. 2013;195:733–9.
54. Zverlov VV, Klupp M, Krauss J, Schwarz WH. Mutations in the scaffoldin gene, CipA, of *Clostridium thermocellum* with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose. J Bacteriol. 2008;190:4321–7.
55. Krauss J, Zverlov VV, Schwarz WH. In vitro reconstitution of the complete *Clostridium thermocellum* cellulosome and synergistic activity on crystalline cellulose. Appl Environ Microbiol. 2012;78:4301–7.
56. Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. J Bacteriol. 2003;185:5109–16.
57. Gerwig GJ, De Waard P, Kamerling JP, Vliegenthart JFG, Morgenstern E, Lamed R, et al. Novel O-linked carbohydrate chains in the cellulase complex (cellulose) of *Clostridium thermocellum*. 3-O-methyl-N-acetylglucosamine as a constituent of a glycoprotein. J Biol Chem. 1989;264:1027–35.
58. Gerwing GJ, Kamerling JP, Vliegenthart JFG. Morag (Morgenstern) E, Lamed R, Bayer EA. Primary structure of O-linked carbohydrate chains in the cellulosome of different *Clostridium thermocellum* strains. Eur J Biochem. 1991;196:115–22.
59. Gerwig GJ, Kamerling JP, Vliegenthart JFG, Morag E, Lamed R, Bayer EA. Novel oligosaccharide constituents of the cellulase complex of *Bacteroides* cellulosolvens. Eur J Biochem. 1992;205:799–808.
60. Gerwig GJ, Kamerling JP, Vliegenthart JFG, Morag E, Lamed R, Bayer EA. The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of *Clostridium thermocellum* and Bacteroides cellulosolvens. J Biol Chem. 1993;268:26956–60.
61. Dror TW, Morag E, Rolider A, Bayer EA, Lamed R, Shoham Y. Regulation of the cellulosomal celS (cel48A) gene of *Clostridium thermocellum* is growth rate dependent. J Bacteriol. 2003;185:3042–8.
62. Stevenson DM, Weimer PJ. Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. Appl Environ Microbiol. 2005;71:4672–8.
63. McGrath CE, Wilson DB. Endocellulolytic activity of the *Clostridium thermocellum* Cel9C (formerly CbhA) catalytic domain. Ind Biotechnol. 2008;4:99–104.
64. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, et al. Development of pyrF-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. Appl Environ Microbiol. 2010;76:6591–9.
65. Zverlov VV, Kellermann J, Schwarz WH. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. Proteomics. 2005;5:3646–53.
66. Anbar M, Gul O, Lamed R, Sezerman UO, Bayer EA. Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. Appl Environ Microbiol. 2012;78:3458–64.