# Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, *Ruminococcus champanellensis,* reveals a fine-tuned system for cohesin-dockerin recognition

Sarah Moraïs[1], Yonit Ben David[1], Lizi Bensoussan[1], Sylvia H. Duncan[2], Nicole M. Koropatkin[3], Eric C. Martens[3], Harry J. Flint[2] and Edward A. Bayer[1]*

[1]*Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel.*

[2]*Microbiology Group, Rowett Institute of Nutrition and Health, University of Aberdeen, Aberdeen, United Kingdom.*

[3]*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.*

*Corresponding author:
Edward A. Bayer, Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel. Tel: (+972)-8-934-2373. Fax: (+972)-8-934-4118. Email: ed.bayer@weizmann.ac.il

Running title: Fine-tuned cohesin-dockerin recognition



This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.13047

## Summary

***Ruminococcus champanellensis*** **is considered a keystone species in the human gut that degrades microcrystalline cellulose efficiently and contains the genetic elements necessary for cellulosome production. The basic elements of its cellulosome architecture, mainly cohesin and dockerin modules from scaffoldins and enzyme-borne dockerins, have been characterized recently. In this study, we cloned, expressed and characterized all of the glycoside hydrolases that contain a dockerin module. Among the 25 enzymes: 10 cellulases, 4 xylanases, 3 mannanases, 2 xyloglucanases, 2 arabinofuranosidases, 2 arabinanases and one β-glucanase were assessed for their comparative enzymatic activity on their respective substrates. The dockerin specificities of the enzymes were examined by ELISA, and 80 positives out of 525 possible interactions were detected. Our analysis reveals a fine-tuned system for cohesin-dockerin specificity and the importance of diversity among the cohesin-dockerin sequences. Our results imply that cohesin-dockerin pairs are not necessarily assembled at random among the same specificity types, as generally believed for other cellulosome-producing bacteria, but reveal a more organized cellulosome architecture. Moreover our results highlight the importance of the cellulosome paradigm for cellulose and hemicellulose degradation by** ***R. champanellensis*** **in the human gut.**

## Introduction

Interest in the human gut microbiota has increased considerably during recent years due to its influence upon human health. One of the main activities of the gut microbiota is to ferment fiber derived from the diet that remains undigested by host enzymes, yielding additional metabolites and energy sources that influence host metabolism, e.g., nutrient absorption and production (Goodman et al., 2009) and energy balance (Turnbaugh et al., 2006). In addition, the human gut microbiome plays a role in the regulation of the immune system (Lee and Mazmanian, 2010) and is an important parameter in many inflammatory and infectious diseases (Young et al., 2005; Kerckhoffs et al., 2011; Vaarala, 2012).

Although cellulose is the major constituent of plant fiber, there have been very few reports of bacteria from the human gut that are able to degrade cellulose. To date, the only human colonic bacterium reported to be capable of degrading crystalline cellulose is *Ruminococcus champanellensis*. This anaerobic, cellulolytic, gram-positive bacterium has been isolated from the human colon and characterized (Chassard et al., 2012). An additional strain closely related to *R. champanellensis*, *Ruminococcus* sp. CAG:379, was isolated independently from the human gut microbiota, suggesting that this bacterium could be widespread in humans. In view of *Ruminococcus champanellensis* remarkably efficient enzymatic activity on microcrystalline cellulose, its genome was sequenced (GenBank, FP929052.1), and this revealed numerous genes coding for elements of a cellulosomal enzyme complex (Ben David et al., 2015), including 12 scaffoldin proteins, collectively carrying 20 cohesins, and 65 dockerin-containing proteins.

Cellulosomes are high-molecular-weight multienzymes complexes which were first described in the anaerobic highly cellulolytic thermophilic bacterium, *Clostridium thermocellum* (Lamed et al., 1983). One of its basic cellulosomal components is a cell-associated scaffoldin subunit, which contains a single cellulose-binding module (CBM) for substrate binding and 9 cohesin modules that serve to integrate dockerin-containing enzymes. The high-affinity cohesin-dockerin interaction was demonstrated to be calcium dependent (Yaron et al., 1995) and species specific (Lytle et al., 1996). To date, three types of cohesins or dockerins have been defined, based on their amino-acid sequences (Bayer et al., 2004). The proximity of the

enzymes within the complex, the targeting of the complex to the substrate and its anchoring to the cell surface are believed to render the cellulosomal complex highly efficient in cellulose degradation. Cellulosomes with various architectures were also discovered in additional anaerobic bacteria within the Clostridiaceae and also within the Ruminococcaceae, specifically *Ruminococcus flavefaciens* isolated from the cow rumen (Ding et al., 2001; Rincon et al., 2010). The latter possesses a larger variety of cellulosomal components, including a large set of adaptor scaffoldins but its cellulosomal organization has yet to be fully determined.

In a recent study (Ben David et al., 2015), the *in-vitro* characterization of the various cellulosomal components of *R. champanellensis* was performed. The cohesin-dockerin interactions among the components revealed the possible assembly of a cell-associated cellulosomal complex that could assemble up to 11 enzymes. In addition, a scaffoldin cluster was described, displaying organizational similarities with the *R. flavefaciens* scaffoldin cluster. Moreover, most of the cohesins of the two species appeared to be phylogenetically related (in most cases type III cohesins). The reiterated sequences of the 65 dockerins were divided into 4 groups, using bioinformatic-based criteria. Twenty-four selected representatives of each group were examined for their specificities, among them 8 originating from scaffoldins and 11 derived from putative glycoside hydrolases. The enzymatic activity of each protein, however, remained undefined.

In the present report, we aimed to characterize the enzymatic activity of the 25 dockerin-containing glycoside hydrolases revealed by CAZy and bioinformatic analysis, along with their dockerin specificities, in order to expand our knowledge on the architecture and activity of the cellulosome from *R. champanellensis*, thus far the sole characterized cellulosome-producing bacterium in the human gut.

## RESULTS

**Dockerin-containing glycoside hydrolase production**. Bioinformatic analysis of the *R. champanellensis* genome revealed 25 putative dockerin-containing glycoside hydrolases (Ben David, 2015). Accordingly, these putative enzymes appeared to be related to GH5, GH8, GH9, GH10, GH11, GH16, GH26, GH30, GH43, GH44, GH48, GH74 and GH98 families (Cantarel et al., 2009). The modular organization

and molecular weights of these GHs are listed in Table 1. Altogether, 25 putative enzymes were cloned without their signal peptides, taking into account their inherent modular organization. The cloning of two multifunctional proteins, GH9B-Doc-GH16A and GH43C-Doc-CE, presented technical difficulties (no amplification could be obtained using either genomic DNA or whole cells as template, under various PCR conditions); therefore they were cloned in segments. In the first, GH9B-Doc-GH16A, the dockerin and the GH16 module only could be inserted in the plasmid, and in the second, GH43C-Doc-CE, the carbodydrate esterase module (CE) was omitted in the final construct. In both cases, the dockerin module, that may bear the most valuable information for our studies, could be preserved.

All of the proteins were produced in *Escherichia coli,* and SDS-PAGE analysis of the purified proteins revealed in most cases a major protein band in good agreement with the respective calculated molecular masses (Supplemental Figure S1).

**Enzymatic activity profile of the GH modules.** The enzymes GH5A, GH5B, GH8A, GH9A, GH9C, GH9D, GH9E, GH9F, GH9G were all active on CMC (carboxymethyl cellulose) and Avicel (microcrystalline cellulose) and thus classified as endoglucanases (Figure 1A and B). Only endoglucanases or processive endoglucanases are active on the soluble CMC substrate, whereas exoglucanase activity can be detected on microcrystalline cellulose. Three of the enzymes, GH8A, GH9A and GH9D, were the most active on the CMC substrate, and GH9A and GH9D also exhibited the highest levels of degradation on Avicel. As expected, the GH48A enzyme exhibited very low levels of enzymatic activity on Avicel by itself, but acted in synergy with GH5B and GH8A (2.7 and 4 fold, respectively), in accordance with other common cellulosomal GH48 cellobiohydrolases (Vazana et al., 2010; Zhang et al., 2010; Morais et al., 2012). Analysis of cell-associated proteins revealed that the two proteins most highly up-regulated during growth of *R. champanellensis* on filter paper cellulose compared with growth on cellobiose were Cel48A (GH48A, 364-fold increase) and Cel9F (GH9F 186-fold increase) enzymes (Table 2 and Supplemental Figure S2). In addition, a number of gene products showed decreased expression during growth on cellulose, which may be related to slower growth rate.

The GH10A, GH10B-GH43E, GH11A-CE and GH30A-CE were active on beechwood xylan and thus classified as xylanases (Figure 2A). The GH10A and GH11A-CE exhibited the highest level of degradation, while GH30A-CE was the least active of these four xylanases. Three mannanases, GH5C, GH26A and GH26B

were active on locust bean gum, and they exhibited similar levels of activity (Figure 2B). GH16A was active on β-D-glucan from barley (Figure 2C). Two arabinanases, GH43A and GH43D, were active on debranched arabinan, and GH43D was more active on this substrate (Figure 2D). Two xyloglucanases, GH44A and GH74A, were active on xyloglucan, GH44A being more active (Figure 2E). These two xyloglucanases were also active on CMC (but not PASC, phosphoric acid swollen cellulose) as sometimes observed for "non-xyloglucan-specific xyloglucanases" (Zverlov et al., 2005). Two arabinofuranosidases, GH43B and GH10B-GH43E, were active on pNP-α-L-arabinofuranoside; the highest activity was measured for the bifunctional xylanase-arabinofuranosidase GH10B-GH43E (Figure 2A and 2F). Since arabinofuranosidase activity can be attributed only to the GH43 module, we presume that the xylanase activity of this bifunctional enzyme is provided by the GH10 module.

No enzymatic activity could be detected for GH43C on xylans, arabinans, pNP-β-D-xylopyranoside and pNP-α-L-arabinofuranoside. The fact that we had to truncate the enzyme for cloning considerations could be a reason for the apparent absence of enzymatic activity. In addition, the enzymatic activity of GH98A remains undetermined (no enzymatic activity on pNP-β-D-galactopyranoside).

The schematic modular architecture of these 25 dockerin-containing glycoside hydrolases, along with their enzymatic activities and dockerin groupings and their proposed nomenclature, is presented in Figure 3. In total, the enzymatic activities of 10 cellulases, 4 xylanases, 3 mannanases, 2 xyloglucanases, 2 arabinofuranosidases, 2 arabinanases and one β-glucanase were established. The enzymatic activity of the putative carbohydrate-esterase modules remained undetermined.

**New insight into dockerin specificity: Regrouping of dockerin-containing proteins.** Dockerin structures are characterized by two segments, each of which contains a $Ca^{+2}$-binding loop and a cohesin-binding helix, which is coordinated by specific positions of amino acids (Pages et al., 1997; Mechaly et al., 2000; Mechaly et al., 2001; Carvalho et al., 2003; Bayer et al., 2004). Historically, in most of the type I dockerins, a clear two-fold symmetry has been observed between their two segments, wherein designated recognition residues are repeated (in identical or very similar fashion). This symmetry has proved to enable two separate cohesin-binding surfaces, with 180° rotation between them, and this phenomenon has been termed the dual

binding mode (Carvalho et al., 2007). However, it has also been observed that certain dockerins that lack structural symmetry display a single-binding mode (Bras et al., 2012). In those dockerin sequences each segment represents separate binding interfaces that can recognize a different cohesin (Pinheiro et al., 2009; Voronov-Goldman et al., 2015). The latter characteristics of known dockerin sequences reflect the complexity and diversity in cohesin-dockerin interactions that contribute to dockerin flexibility in case of steric interferences and improved response to the dynamic process of plant cell wall degradation (Carvalho et al., 2007).

In our previous study (Ben David, 2015), the 64 dockerins of *R. champanellensis* (not including the ScaL dockerin) were aligned, and then clustered into four groups. Dockerins of Group 1 were found to interact directly with the cell-anchoring scaffoldin, ScaE. Most of the proteins in Group 2 represent glycoside hydrolase enzymes, mainly cellulases or closely associated enzymes. In contrast, in the dockerin-containing enzymes of Groups 3 and 4, most appeared to be hemicellulases as well as dockerin-containing proteins that lack confirmed carbohydrate-degrading components.

Intriguingly, the dockerins from Groups 3 and 4 exhibited similar binding profiles. Clearly, the dockerins of these two groups are asymmetrical in their "repeated" segments, where the reputed recognition residues are clearly different in character (Ben David et al., 2015), thus indicating a single mode rather than a dual mode of binding. If so, then the two segments can theoretically be switched with retention of the same specificity characteristics. Therefore, a renewed alignment was performed, taking this possibility into account. Indeed, reexamination of the two dockerin segments in the sequences from groups 3 and 4 revealed remarkable, but reversed, similarities between the two groups. Thus, the first binding interface of Group 3 dockerins is highly similar to the second binding interface of Group 4 dockerins and vice versa (Figure 4). According to this new arrangement of the dockerins, a mechanism of an alternative-binding mode can be suggested; it seems that the first segment of Group 3 and the second segment of Group 4 could be responsible for the binding to cohesins C and D while the second segment of Group 3 and the first segment of Group 4 would allow interaction with cohesin I. This hypothesis can also be extended to the group 2 dockerins, which also bind to CohI. Similar motifs were thus found between the first segment of the group 2 dockerins and the segment that is considered to interact with CohI in groups 3 and 4, as

described below (see Figure 4). This analysis might therefore serve to explain why these groups are also associated with CohI.

The CohI-interacting dockerin segment is characterized by Val or Ile and Ala (or small uncharged residue, i.e., Asn or Ser) in positions 10 and 11, mainly Val or an aliphatic residue in positions 13 and 14 (positions 14 and 15 in the second helix), and hydrophilic resides followed by aromatic amino acids in positions 17 and 19 (Figure 4). In contrast, the sequences of the CohC- and CohD-interacting dockerin segment exhibit more variance in the amino acids found in the putative recognition positions. Yet, the basic amino acids, Arg and Lys, exclusively occupy positions 17 and 18. Notably, the dockerins of Xyn11A-CE and GH98 have hydrophilic and charged amino acids in positions 10, 11 and 14 (position 15 in the second helix), which can explain why they failed to interact with CohI.

According to these new findings, Groups 3 and 4 were redistributed into four groups of putative cohesin-dockerin interactions: (i) interaction with cohesins C, D and I, (ii) interaction with CohC and CohD only, (iii) interaction with CohI, and (iv) currently unknown interactions (Supplemental Figure S3).

**Affinity-based ELISA.** The dockerin-containing enzymes were examined for their interactions with the 21 cohesins of *R. champanellensis*. In a previous work (Ben David et al., 2015), 20 cohesins were described. The published sequence of the genome contains numerous gaps, and an additional scaffoldin, ScaL, was identified during the course of the present work. The ScaL gene, which was recovered by genome walking, includes a large N-terminal domain of unknown function, a nucleoporin-like module, a cohesin module and a dockerin module (Table 3).

The cohesin genes were all fused with a CBM cassette that has been employed earlier for antigen recognition (Ben David et al., 2015). All 25 dockerin-bearing proteins were tested for their binding affinity with the 21 cohesins known to date, including the additional CohL from ScaL, described in this work.

A total of 525 interactions were tested, among them 80 positives (Figures S4 and S5; negative interactions are not shown). Binding affinity partners of 24 dockerins were determined out of the 25 examined.

Glycoside-hydrolases from Group 2 (alignment in Supplemental Figure S6) presented various binding profiles (Table 4). Cel5A, Cel9A and Xeg74A dockerins could generally bind all of the cohesins partners (Cohesins A2, B1/B2/B3, B4, B5/B6, H, and I) except for the type I Cohs C and D. The Cel5B dockerin exhibited binding

affinity for cohesins B5/B6, H and I. The Cel9C, Cel48A, Man26B and Xyn10A dockerins were able to bind cohesins A2, B1/B2/B3, H and I. The Xeg44A dockerin has affinity only to cohesin H. The dockerin of Cel9D exhibited binding affinity for Cohesins B5/B6 only. Finally, the Cel9F and Man5A dockerins interacted with cohesins A2, B1/B2/B3 and H.

Glycoside-hydrolases from Groups 3 and 4 were all active except for Cel8A as previously reported (Ben David et al., 2015) (Table 4 and Supplemental Figures S4 and S5). The binding profile of the dockerins matched almost perfectly the above-predicted interactions. The dockerins of Glc16A (i.e., a cloned portion of the complete enzyme CBM4-Fn3-GH9B-Doc-GH16A), Cel9G, Xyn10B-Abf43E, Man26A, Abf43B, GH43C and Arb43D all interacted with CohC, CohD and CohI. The Arb43A dockerin, which would be predicted to share the same binding profile, interacted strongly only with CohD and weakly with CohC as previously reported (Ben David et al., 2015). However, based on the dockerin sequence, it would seem that it should bind to all three cohesins and not only to cohesin D. Since the interaction with cohesin D is relatively weak, it is therefore likely that the protein may not have been expressed and folded properly. One explanation for the lack of binding of the Cel8A dockerin and the weak binding of Arb43D dockerin could be the presence of cysteine residues in position 14 that could disturb proper folding in those particular cases.

Xyn11A-CE, Xyn30A-CE and GH98A dockerins interacted selectively with CohC and CohD as predicted from their amino-acid sequences.

The Cel9E dockerin, whose dockerin-binding profile remained uncharacterized in the previous work, interacted with CohC and CohI. The binding to CohC could be attributed to Arg and Lys residues in positions 18 and 19, respectively, and the ability to interact with CohI could be related to the valine residues in positions 10 and 14 and the aromatic Phe residue, in position 19.

## DISCUSSION

The microbial community that occupies the human gut habitat is known to produce an arsenal of enzymes that together degrade complex carbohydrates from the diet that cannot be hydrolyzed by human-based enzymes (Flint et al., 2008), thereby providing supplemental energy sources for the host. Bacteria within this community are believed to have evolved to specialize in certain types of carbohydrate degradation

and complement each other (Martens et al., 2011). Bacteroidetes display enzymatic activities for starch, hemicellulose, pectins and glucans (Xu et al., 2003a) but limited ability for cellulose degradation (Robert et al., 2007; McNulty et al., 2013). On the other hand, Firmicutes are able to utilize starch, cellulose and hemicelluloses. They are considered to be more substrate-specific (Salyers et al., 1977; Chassard et al., 2007; Walker et al., 2011; Chassard et al., 2012; Ze et al., 2012), and some species among the Firmicutes purportedly represent keystone species in polysaccharide degradation (Ze et al., 2013; Ben David et al., 2015).

The *R. champanellensis* genome contains a repertoire of 12 scaffoldins (Table 3), each of which contains various numbers of cohesins from one to seven. In most cases the scaffoldins also possess a dockerin that will allow interactions with other scaffoldins. The cohesin-dockerin interactions among the various components revealed the possible assembly of a cell-associated cellulosomal complex that could assemble up to 11 enzymes.

In this study, we conducted an extensive, near-complete analysis of the cellulosomal enzymatic system of *Ruminococcus champanellensis*. In addition, the dockerin specificities of 25 enzymes were revealed and were found to be consistent with our overall predictions, based on the sequence similarity between dockerins and recognition residues. The *R. champanellensis* genome contains 65 dockerin-bearing proteins, among which 25 enzymes were characterized in the present study, and 8 scaffoldin-borne dockerins were characterized in our previous study (Ben David et al., 2015) in addition to 31 non-glycoside-hydrolase dockerin-containing proteins. The dockerin specificities of the latter remain to be elucidated. As in our previous study (Ben David et al., 2015), none of the dockerins examined in this study interacted with 5 cohesins (namely, CohB6, CohB7, CohF, CohG and CohK). Consequently, their respective binding partner(s) remain as yet unknown.

The set of cellulosomal enzymes in *R. champanellensis* comprises both cellulose- (endoglucanases and exoglucanases) and hemicellulose-degrading activities, the latter of which include xylanases, mannanases, arabinanases, xyloglucanases and arabinofuranosidases. One interesting fact is that all members of the cellulase families, i.e. GH8, GH9 and GH48, contain a dockerin module. Moreover, all members of the hemicellulase families, including GH10, GH11, GH30, GH43, GH44 and GH74, are also cellulosomal enzymes. The *R. champanellensis* genome also contains eight GH5 enzymes, but only three of them are cellulosomal.

Sequence alignment and phylogenetic tree analysis of the additional five GH5s with other characterized GH5 enzymes demonstrated that 4 of them are predicted cellulases and one is consistent with mannanases.

Cellulosome-producing bacteria frequently possess two sets of enzymes, cellulosomal and non-cellulosomal. *C. thermocellum,* for example, produces two highly active cellulases, Cel48Y and Cel9I, which contain cellulose-specific CBMs instead of dockerins, and are therefore not part of the cellulosome system (Berger et al., 2007). The non-cellulosomal system also includes two GH5s, three GH10s (at least one of which exhibited xylanase activity (Zverlov et al., 2005)), one GH43 and several others (Dassa et al., 2012). Intriguingly, *R. champanellensis* produces cellulosome complexes for its main strategy for both cellulose and hemicellulose degradation, with only a few free enzymes confined to GH family 5.

It is interesting to note that representatives of the GH48 and GH9 families were highly upregulated in the proteome of both *R. champanellensis* and *R. flavefaciens* cells (Vodovnik et al., 2013) when grown on cellulose rather than cellobiose. Both of these highly expressed proteins carry dockerins and are thus assumed to be cellulosomal in these two species. Both types of enzymes are typically abundant in cellulosomes, particularly when the parent bacterium is grown on cellulosic substrates (Dror et al., 2003; Berg Miller et al., 2009).

It is also interesting to note, that the *R. champanellensis* genome codes for a GH98 enzyme, which is rare, and this enzyme is also part of the cellulosomal machinery. Thus far, a dockerin-containing GH98 was reported previously only in *Clostridium cellulovorans* (Cantarel et al., 2009). *Ruminococcus albus* also produces a GH98, but without a dockerin (Dassa et al., 2014). GH98 enzymes have previously been shown to exhibit blood group endo-β-galactosidase activity in pathogenic bacteria, although in our particular case, the enzyme appeared to be inactive on a colorimetric galactopyranoside-containing substrate.

The enzymes examined in this study exhibit two types of cohesin-dockerin specificities. The specificity type seems to be unrelated to the molecular weight of the proteins but could perhaps be linked to the enzymatic activity, i.e., Group 2 enzymes representing mostly cellulases and Group 3 and 4 mostly hemicellulases and non-glycoside hydrolase proteins. These results raise the question why certain enzymes need an adaptor scaffoldin to be integrated into the cellulosomal complex? An

interesting observation is that most of the enzymes that bind directly to the scaffoldin (from Group 2) have a particularly long Thr-rich linker that links the dockerin to the catalytic module, which may infer that the adaptor scaffoldins (ScaC and ScaD) also serve as a linker for proteins that lack these types of linker (from Groups 3 and 4).

Similar to the *R. champanellensis* Sca's C and D, the *R. flavefaciens* ScaC also serves as an adaptor scaffoldin, which allows many proteins that are not recognized directly by the ScaA cohesins to be bridged into the cellulosome assembly (Rincon et al., 2004). These types of adaptors are different from adaptor scaffoldins that serve to amplify the number of enzymes in the cellulosomal complex (Xu et al., 2003b; Dassa et al., 2012). In contrast, monovalent adaptor scaffoldin may be part of a regulatory mechanism for cellulosomal composition.

An interesting fact is that in each of the two specificities, the dockerins did not interact similarly with the various cohesins but presented diverse patterns of affinity. This phenomenon is especially intriguing considering that Group 2 dockerin sequences are very similar. This could reflect an organized manner of integrating enzymes or cellulosomal components in the complex and not a random assembly of the enzymes on the scaffoldin as suggested for cellulosome assembly in other bacteria. Multiple cohesin-dockerin binding specificities have also been demonstrated for different dockerin-carrying enzymes in the phylogenetically related *R. flavefaciens* (Rincon et al., 2003; Jindou et al., 2006). These results for both Ruminococcus spp. essentially contradict those of a recent study by Hirano and colleagues (Hirano et al., 2015), in which it was suggested that preferential binding of cellulosomal enzymes to the cohesin modules did not result from slight differences in binding affinity but from differences in the length of the inter-cohesin linker: a shorter inter-cohesin linker promoting preferential binding**.**

Our analyses contribute to a better understanding of the enzymatic degradation of complex carbohydrates by *R. champanellensis* in the human gut. Our findings highlight the importance of the cellulosome paradigm for cellulose and hemicellulose degradation and the controlled assembly of the complex via fine-tuned cohesin-dockerin recognition.

## Experimental Procedures

**Cloning.** Dockerin-containing glycoside hydrolases were cloned from *R. champanellensis* genomic DNA using appropriate primers (Supplemental Table 1) and Phusion High Fidelity DNA polymerase F530-S (New England Biolabs, Inc). The genes were restricted using Fastdigest enzymes (Thermo scientific, USA) and ligated into either in pET21a or pET28a using T4 DNA ligase (Fermentas UAB, Vilnius, Lithuania). The constructs were designed to contain a His-tag for subsequent purification.

The CBM-Coh gene cassette (Barak et al., 2005) consists of a family 3a CBM from the *C. thermocellum* CipA scaffoldin cloned into plasmid pET28a (Novagen Inc., Madison, WI, USA), into which any cohesin gene can be introduced between BamHI and XhoI restriction sites of the plasmid. The Coh-CBM gene cassette is the same as the CBM-Coh cassette, only in reverse order of the modules. Any cohesin gene can be introduced between NcoI and BamHI restriction sites of the plasmid. The full list of fused cohesins used in this article is given in Table 3.

PCR products were purified using a HiYield$^{TM}$ Gel/PCR Fragments Extraction Kit (Real Biotech Corporation, RBC, Taiwan) and plasmids were extracted using Qiagen miniprep kit (Netherlands). The cloning of each gene was confirmed by DNA sequencing. Competent *E. coli* XL1 competent cells were used for plasmid transformation.

**Recombinant protein expression and purification**. *E. coli* BL21 (DE3) cells were transformed with the desired plasmid and plated onto LB-kanamycin plates. The cells producing GH5B-, GH8A-, GH9A-, GH9C-, GH9D-, GH9E-, GH9F-, GH9G-, GH10A-, GH10B/GH43E-, GH11A/CE-, GH16-, GH43C- and GH74A-containing enzymes and ScaL were grown in 50 ml LB (Luria Broth) and 2 mM $CaCl_2$ (to facilitate proper folding of the dockerin) at 37°C until $A_{600}$≈0.8-1 and induced by adding 0.1 mM (final concentration) isopropyl-1-thio-β-D-galactoside (IPTG) (Fermentas UAB Vilnius, Lithuania). Cell growth was continued at 16°C overnight. Cells producing GH5A, GH5C, GH26A, GH26B, GH30A-CE, GH43A, GH43B, GH43D, GH44A, GH48A or GH98A were grown in 50 ml TYG (Tryptone Yeast Glucose) medium supplemented with 2 mM $CaCl_2$ at 37°C until $A_{600}$≈0.8-1 and induced by adding 0.1 mM IPTG. Growth was continued 3 h at 37°C. Cells were harvested by centrifugation at 5000 rpm for 5 min. Pelleted cells were resuspended in 1 ml TBS (Tris-buffered saline, 137 mM NaCl, 2.7 mM KCL, 25 mM Tris-HCl, pH=7.4).

The His-tagged proteins were either purified on a nickel-nitrilotriacetic acid (Ni-NTA) column (Qiagen), as reported earlier (Caspi et al., 2006) or small-scale purified using Ni-NTA spin columns (Qiagen). The cohesin-containing protein supernatant fluids were added to 2 g of macroporous bead cellulose preswollen gel (IONTOSORB, Usti nad Labem, Czech Republic), and incubated for 1 h, with rotation at 4°C. The mixture was then loaded onto a gravity column, and washed with 100 ml of TBS containing 1 M NaCl, and then washed with 100 ml TBS. Three 5 ml elutions of 1% triethanolamine (TEA) were then collected. The fractions were subjected to SDS-PAGE, in order to assess protein purity, and then dialyzed with TBS.

Purity of the recombinant proteins was tested by SDS-PAGE on 10% acrylamide gels. Protein concentration was estimated by absorbance (280 nm) based on the known amino acid composition of the protein using the Protparam tool (http://www.expasy.org/tools/protparam.html). Proteins were stored in 50% (v/v) glycerol at -20°C.

**Enzymatic activity assay.** All assays were performed at least twice in triplicate. The different proteins were tested against several potential substrates according to the GH family (Cantarel et al., 2009) and at a pH corresponding to the optimal pH generally observed for these enzymatic activities in previous studies. All enzymes were tested at a concentration of 0.5 μM at 37°C. Cellulases were tested at pH 5 (buffer acetate 50 mM final concentration), for either 1 h with 2% carboxymethyl cellulose (CMC) (VWR International, Ltd, England) or in 10% Avicel for 24 h (FMC, Delaware USA). Xylanases were tested at pH 6 (buffer citrate 50 mM final concentration) for 1 h with 2% beechwood xylan (Sigma). β-glucanase were tested on β-D glucan from barley (Sigma) for 1h at pH 5 (buffer acetate 50 mM final concentration) Arabinanases were tested at pH 6 (buffer citrate 50 mM final concentration), for 1 h with 2% debranched arabinan (Megazyme, Ireland). Mannanase were tested at pH 5 (buffer acetate 50 mM final concentration) for 1 h with 1% locus bean gum. The xyloglucanase was examined with 2% xyloglucan (Megazyme, Ireland) for 1 h at pH 6 (buffer citrate 50 mM final concentration). Enzymatic reactions were terminated by transferring the tubes to an ice-water bath, and the tubes were centrifuged for 2 min at 14000 rpm at room temperature. Enzymatic activity was then determined quantitatively by measuring the soluble

reducing sugars released from the polysaccharide substrates by the dinitrosalicyclic acid (DNS) method (Miller, 1959; Ghose, 1987). DNS solution (150 μl) was added to 100 μl of sample, and after boiling the reaction mixture for 10 min, absorbance at 540 nm was measured. Sugar concentrations were determined using a glucose standard curve. The colorimetric substrate, pNP-α-L-arabinofuranoside (pNPA) (Sigma), was used at 12.5 mM and pH 6 (50 mM citrate buffer) in a reaction mixture containing 0.5 micromolar enzyme, and the tubes were incubated for 20 min at 37°C.

Chitin, laminarin, pNP-β-D-glucopyranoside and pNP-β-D-cellobioside (Sigma) were also used for substrate specificity determination.

**Affinity-based ELISA.** The matching fusion-protein procedure of Barak et al (Barak et al., 2005; Caspi et al., 2006) was followed to determine cohesin-dockerin specificity of interaction. Dockerin-containing enzymes were immobilized on the plate at a concentration of 1 μg/ml (100 μl/well) in 0.1 M sodium carbonate (pH 9) and incubated at 4°C overnight. The following steps were performed at room temperature for 1 h with all reagents at a volume of 100 μl/well, with a three-times repeated washing step (300 μl/well blocking buffer without BSA) included after each step. The coating solution was discarded, and blocking buffer (TBS, 10 mM $CaCl_2$, 0.05% Tween 20, 2% BSA) was added. The blocking buffer was discarded, and the desired CBM-Coh(s), diluted to concentrations of 100 ng/ml in blocking buffer, were added. Rabbit anti-CBM antibody (diluted 1:3000) was used as the primary antibody preparation and the secondary antibody preparation was HRP-labeled anti-rabbit antibody diluted 1:10000 in blocking buffer. Substrate-Chromogen TMB (Dako, Agilent Technologies, USA) was added at 100 μl/well and the reaction was carried out for 2 min before color formation was terminated upon addition of 1 M $H_2SO_4$ (50 μl/well), and the absorbance was measured at 450 nm using a tunable microplate reader.

**Proteomic analysis of *R. champanellensis* 18P13.** *R. champanellensis* 18P13 cultures were grown anaerobically (37°C) in 800 ml of basal YCFA medium (Lopez-Siles et al., 2012) containing 1% clarified rumen fluid with either 0.5 % cellobiose or 0.5% of filter paper cellulose cut into 1 cm squares (Whatman No.1) for 48 h and 96 h, respectively. Samples were analyzed from duplicate biological repeats, with three technical replicates for each gel separation, such that comparison was made between six gel separations from each growth condition. The cellulose-grown cells were

harvested following vigorous shaking and allowing the substrate to sediment for a period of 10 min. The cells from both the cellobiose- and cellulose-grown cultures were harvested as described by Vodovnik et al. (Vodovnik et al., 2013). Equivalent levels of proteins in Rabilloud buffer were separated by two dimensional gel electrophoresis, and gels were imaged as described previously (Vodovnik et al., 2013). The gels were analyzed with PD Quest software (Bio-Rad). Spots of interest were excised from the gels manually then processed and identified by Nano LC MS/MS as described previously (Vodovnik et al., 2013).

**Acknowledgements**

This research was supported by the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel and by a grant (No. 1349) to EAB from the Israel Science Foundation (ISF). Additional support was obtained from the establishment of an Israeli Center of Research Excellence (I-CORE Center No. 152/11) managed by the Israel Science Foundation. 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 an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. In addition, EAB is grateful for a grant from the F. Warren Hellman Grant for Alternative Energy Research in Israel in support of alternative energy research in Israel administered by the Israel Strategic Alternative Energy Foundation (I-SAEF). HJF acknowledges support from BBSRC grant no BB/L009951/1, from the Scottish Government Food, Land and People program and from the Society for Applied Microbiology. Thanks are due to Fergus Nicol and Louise Cantlay for proteomic analysis. EAB is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.



## Figure legends

**Figure 1:** Enzymatic activity of *R. champanellensis* cellulosomal cellulases. A. Comparative enzymatic activity of the cellulases at a concentration of 0.5 μM. Cellulases were tested at pH 5 and 37°C for 1 h with 2% carboxymethyl cellulose. B. Comparative enzymatic activity of cellulases at a concentration of 0.5 μM and synergism with GH48A. Cellulases were tested at pH 5 and 37°C for 24 h with 10% Avicel. Reactions were performed at least twice in triplicate, standard deviations are indicated.

**Figure 2:** Enzymatic profile of *R. champanellensis* cellulosomal glycoside hydrolases. A. Comparative enzymatic activity of xylanases at a concentration of 0.5 μM at pH 6 and 37°C for 1 h with 2% beechwood xylan. B. Comparative enzymatic activity of mannanases at a concentration of 0.5 μM at pH 5 and 37°C for 1 h with 1% locus bean gum. C. β-glucanase activity at a concentration of 0.5 μM at pH 5 and 37°C for 1 h with β-D glucan from barley D. Comparative enzymatic activity of arabinanases at a concentration of 0.5 μM at pH 6 and 37°C 1 h on 2% debranched arabinan. E. Comparative enzymatic activity of xyloglucanases at a concentration of 0.5 μM at pH6 and 37°C 1 h on 2% xyloglucan. F. Comparative enzymatic activity of *R. champanellensis* cellulosomal arabinofuranosidases. The enzymes were tested at pH 6 and 37°C for 20 min with 12.5 mM pNP-α-L-arabinofuranoside. Reactions were performed at least twice in triplicate, standard deviations are indicated.

**Figure 3:** Schematic representation and proposed nomenclature of the dockerin-containing glycoside hydrolases from *R. champanellensis.* Enzyme activity and dockerin-specificity are color-coded. GH and CBM families are indicated numerically. Modules shown in white were not expressed in this study.

**Figure 4:** Proposed twofold alternative specificity mechanism of *R. champanellensis* cohesins C, D and I. Red boxes indicate the residues suspected as responsible for specific cohesin recognition. Residues highlighted in cyan and yellow are involved in the two forms of binding to a cohesin. Note that the two segments of group 3 dockerins (blue and green boxes, arrows) appear in reversed order, such that their predicted recognition residues align with those of the group 4 dockerins (yellow). Positions of calcium binding residues are shown in gray. Numbering indicates the residue positions in the two duplicated segments.

## Supplementary information legends

**Supplemental Table 1:** Primers used in the study (restrictions sites represented in upper cases)

**Supplemental Figure S1:** Purity of the recombinant enzymes after Ni-NTA purification as assessed by SDS-PAGE gels (10% acrylamide).

**Supplemental Figure S2:** Comparative proteome of A. cellobiose and B. filter paper cellulose grown *Ruminococcus champanellensis* 18P13. Spot F1 = Cel9F and Spot F2 = Cel48A.

**Supplemental Figure S3. New division of *R. champanellensis* Group 3 and 4 dockerins.** The dockerins of Groups 3 and 4 were redivided based on the finding of the alternative-binding mode (Figure 4). Positions of the putative cohesin recognition residues are highlighted in cyan for the first helix and in yellow for the second helix. Proteins highlighted in green were examined in our previous study (Ben David et al., 2015), and proteins highlighted in blue were topics of the present study.

**Supplemental Figure S4:** Affinity-based ELISA with Group 2 enzymes. The dockerin-containing enzymes were coated at 1 μg/ml and the CBM fused to CohH, CohI, CohA2, CohB1/B2/B3, CohB4, CohB5/B6 or CohCc (from *Clostridium cellulolyticum* as negative control) were used at 100 ng/ml. Reactions were performed at least three times in triplicate, standard deviations are indicated.

**Supplemental Figure S5** Affinity-based ELISA with Group 3 and 4 enzymes. The dockerin-containing enzymes were coated at 1 μg/ml and the CBM fused to CohC, CohD, CohH or CohCc (from *C. cellulolyticum* as negative control) were used at 100 ng/ml. Reactions were performed at least three times in triplicate, standard deviations are indicated.

**Supplemental Figure S6: *R. champanellensis* dockerin Group 2 alignment.** The 17 dockerin sequences of *R. champanellensis* were aligned, using bioinformatics-based criteria. Dockerins selected for this study are highlighted in blue and those highlighted in green were also assayed in our previous study (Ben David et al., 2015) (see Table 1 for GI number of the parent proteins). Positions of calcium binding residues are shown in cyan, and putative recognition residues are shown in yellow.
Protein names highlighted green were examined in our previous study (Ben David et al., 2015), and protein names highlighted in blue were topics of the present study.

## References


Barak, Y., Handelsman, T., Nakar, D., Mechaly, A., Lamed, R., Shoham, Y., and Bayer, E.A. (2005) Matching fusion protein systems for affinity analysis of two interacting families of proteins: the cohesin-dockerin interaction. *J Mol Recognit* **18**: 491-501.

Bayer, E.A., Belaich, J.-P., Shoham, Y., and Lamed, R. (2004) The cellulosomes: Multi-enzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* **58**: 521-554.

Ben David, Y., Dassa, B., Borovok, I., Lamed, R., Koropatkin, N.M., Martens, E.C. et al. (2015) Ruminococcal cellulosome systems from rumen to human. *Environ Microbiol*.

Berg Miller, M.E., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T. et al. (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PloS one* **4**: e6650.

Berger, E., Zhang, D., Zverlov, V.V., and Schwarz, W.H. (2007) Two noncellulosomal cellulases of *Clostridium thermocellum*, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically. *FEMS Microbiol Lett* **268**: 194-201.

Bras, J.L., Alves, V.D., Carvalho, A.L., Najmudin, S., Prates, J.A., Ferreira, L.M. et al. (2012) Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode. *J Biol Chem* **287**: 44394-44405.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**: D233-238.

Carvalho, A.L., Dias, F.M., Prates, J.A., Nagy, T., Gilbert, H.J., Davies, G.J. et al. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc Natl Acad Sci USA* **100**: 13809-13814.

Carvalho, A.L., Dias, F.M., Nagy, T., Prates, J.A., Proctor, M.R., Smith, N. et al. (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proc Natl Acad Sci USA* **104**: 3089-3094.

Caspi, J., Irwin, D., Lamed, R., Shoham, Y., Fierobe, H.-P., Wilson, D.B., and Bayer, E.A. (2006) *Thermobifida fusca* family-6 cellulases as potential designer cellulosome components. *Biocat Biotransform* **24**: 3-12.

Chassard, C., Goumy, V., Leclerc, M., Del'homme, C., and Bernalier-Donadille, A. (2007) Characterization of the xylan-degrading microbial community from human faeces. *FEMS Microbiol Ecol* **61**: 121-131.

Chassard, C., Delmas, E., Robert, C., Lawson, P.A., and Bernalier-Donadille, A. (2012) *Ruminococcus champanellensis sp. nov*., a cellulose-degrading bacterium from human gut microbiota. *Int J Syst Evol Microbiol* **62**: 138-143.

Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C.L. et al. (2012) Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. *BMC genomics* **13**: 210.

Dassa, B., Borovok, I., Ruimy-Israeli, V., Lamed, R., Flint, H.J., Duncan, S.H. et al. (2014) Rumen cellulosomics: divergent fiber-degrading strategies revealed by comparative genome-wide analysis of six ruminococcal strains. *PloS one* **9**: e99221.

Ding, S.Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V. et al. (2001) Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J Bacteriol* **183**: 1945-1953.

Dror, T.W., Morag, E., Rolider, A., Bayer, E.A., Lamed, R., and Shoham, Y. (2003) Regulation of the cellulosomal CelS (cel48A) gene of *Clostridium thermocellum* is growth rate dependent. *J Bacteriol* **185**: 3042-3048.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. (2008) Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **6**: 121-131.

Ghose, T.K. (1987) Measurments of cellulase activity. *Pure Appl Chem* **59**: 257-268.

Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A. et al. (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**: 279-289.

Hirano, K., Nihei, S., Hasegawa, H., Haruki, M., and Hirano, N. (2015) Stoichiometric Assembly of Cellulosome Generates Maximum Synergy for the Degradation of Crystalline Cellulose, as Revealed by In Vitro Reconstitution of the *Clostridium thermocellum* Cellulosome. *Appl Environ Microbiol*.

Jindou, S., Borovok, I., Rincon, M.T., Flint, H.J., Antonopoulos, D.A., Berg, M.E. et al. (2006) Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*. *J Bacteriol* **188**: 7971-7976.

Kerckhoffs, A.P., Ben-Amor, K., Samsom, M., van der Rest, M.E., de Vogel, J., Knol, J., and Akkermans, L.M. (2011) Molecular analysis of faecal and duodenal samples reveals significantly higher prevalence and numbers of *Pseudomonas aeruginosa* in irritable bowel syndrome. *J Med Microbiol* **60**: 236-245.

Lamed, R., Setter, E., and Bayer, E.A. (1983) Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *J Bacteriol* **156**: 828-836.

Lee, Y.K., and Mazmanian, S.K. (2010) Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* **330**: 1768-1773.

Lopez-Siles, M., Khan, T.M., Duncan, S.H., Harmsen, H.J., Garcia-Gil, L.J., and Flint, H.J. (2012) Cultured representatives of two major phylogroups of human colonic *Faecalibacterium prausnitzii* can utilize pectin, uronic acids, and host-derived substrates for growth. *Appl Environ Microbiol* **78**: 420-428.

Lytle, B., Myers, C., Kruus, K., and Wu, J.H.D. (1996) Interactions of the CelS binding ligand with various receptor domains of the *Clostridium thermocellum* cellulosomal scaffolding protein, CipA. *J Bacteriol* **178**: 1200-1203.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P. et al. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**: e1001221.

McNulty, N.P., Wu, M., Erickson, A.R., Pan, C., Erickson, B.K., Martens, E.C. et al. (2013) Effects of diet on resource utilization by a model human gut microbiota containing *Bacteroides cellulosilyticus* WH2, a symbiont with an extensive glycobiome. *PLoS Biol* **11**: e1001637.

Mechaly, A., Fierobe, H.-P., Belaich, A., Belaich, J.-P., Lamed, R., Shoham, Y., and Bayer, E.A. (2001) Cohesin-dockerin interaction in cellulosome assembly: A

single hydroxyl group of a dockerin domain distinguishes between non-recognition and high-affinity recognition (Erratum). *J Biol Chem* **276**: 19678.

Mechaly, A., Yaron, S., Lamed, R., Fierobe, H.P., Belaich, A., Belaich, J.P. et al. (2000) Cohesin-dockerin recognition in cellulosome assembly: experiment versus hypothesis. *Proteins* **39**: 170-177.

Miller, G.L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Biochem* **31**: 426-428.

Morais, S., Barak, Y., Lamed, R., Wilson, D.B., Xu, Q., Himmel, M.E., and Bayer, E.A. (2012) Paradigmatic status of an endo- and exoglucanase and its effect on crystalline cellulose degradation. *Biotechnol Biofuels* **5**: 78.

Pages, S., Belaich, A., Belaich, J.-P., Morag, E., Lamed, R., Shoham, Y., and Bayer, E.A. (1997) Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum:* Prediction of specificity determinants of the dockerin domain. *Proteins* **29**: 517-527.

Pinheiro, B.A., Gilbert, H.J., Sakka, K., Sakka, K., Fernandes, V.O., Prates, J.A. et al. (2009) Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem J* **424**: 375-384.

Rincon, M.T., Ding, S.Y., McCrae, S.I., Martin, J.C., Aurilia, V., Lamed, R. et al. (2003) Novel organization and divergent dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *J Bacteriol* **185**: 703-713.

Rincon, M.T., Martin, J.C., Aurilia, V., McCrae, S.I., Rucklidge, G.J., Reid, M.D. et al. (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the *Ruminococcus flavefaciens* 17 cellulosome. *J Bacteriol* **186**: 2576-2585.

Rincon, M.T., Dassa, B., Flint, H.J., Travis, A.J., Jindou, S., Borovok, I. et al. (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PloS one* **5**: e12476.

Robert, C., Chassard, C., Lawson, P.A., and Bernalier-Donadille, A. (2007) *Bacteroides cellulosilyticus sp. nov.*, a cellulolytic bacterium from the human gut microbial community. *Int J Syst Evol Microbiol* **57**: 1516-1520.

Salyers, A.A., West, S.E., Vercellotti, J.R., and Wilkins, T.D. (1977) Fermentation of mucins and plant polysaccharides by anaerobic bacteria from the human colon. *Appl Environ Microbiol* **34**: 529-533.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027-1031.

Vaarala, O. (2012) Gut microbiota and type 1 diabetes. *Rev Diabet Stud* **9**: 251-259.

Vazana, Y., Morais, S., Barak, Y., Lamed, R., and Bayer, E.A. (2010) Interplay between *Clostridium thermocellum* family 48 and family 9 cellulases in cellulosomal versus noncellulosomal states. *Appl Environ Microbiol* **76**: 3236-3243.

Vodovnik, M., Duncan, S.H., Reid, M.D., Cantlay, L., Turner, K., Parkhill, J. et al. (2013) Expression of cellulosome components and type IV pili within the extracellular proteome of *Ruminococcus flavefaciens 007*. *PloS one* **8**: e65333.

Voronov-Goldman, M., Yaniv, O., Gul, O., Yoffe, H., Salama-Alber, O., Slutzki, M. et al. (2015) Standalone cohesin as a molecular shuttle in cellulosome assembly. *FEBS letters*.

Walker, A.W., Ince, J., Duncan, S.H., Webster, L.M., Holtrop, G., Ze, X. et al. (2011) Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* **5**: 220-230.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C. et al. (2003a) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* **299**: 2074-2076.

Xu, Q., Gao, W., Ding, S.Y., Kenig, R., Shoham, Y., Bayer, E.A., and Lamed, R. (2003b) The cellulosome system of *Acetivibrio cellulolyticus* includes a novel type of adaptor protein and a cell surface anchoring protein. *J Bacteriol* **185**: 4548-4557.

Yaron, S., Morag, E., Bayer, E.A., Lamed, R., and Shoham, Y. (1995) Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. *FEBS Lett* **360**: 121-124.

Young, G.P., Hu, Y., Le Leu, R.K., and Nyskohus, L. (2005) Dietary fibre and colorectal cancer: a model for environment-gene interactions. *Mol Nutr Food Res* **49**: 571-584.

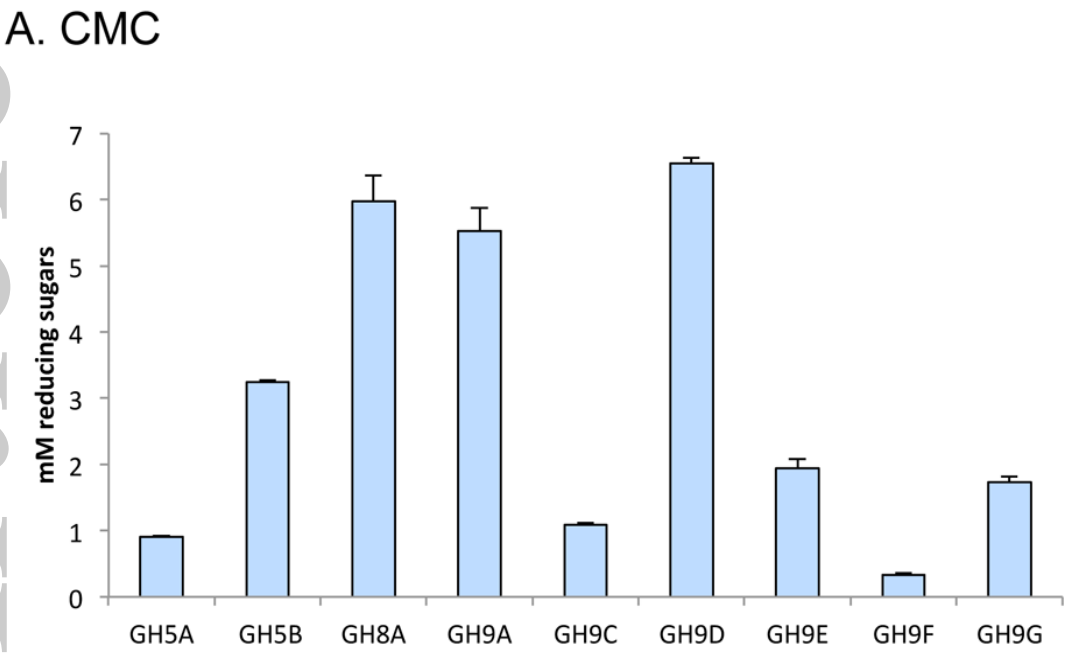
Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**: 1535-1543.

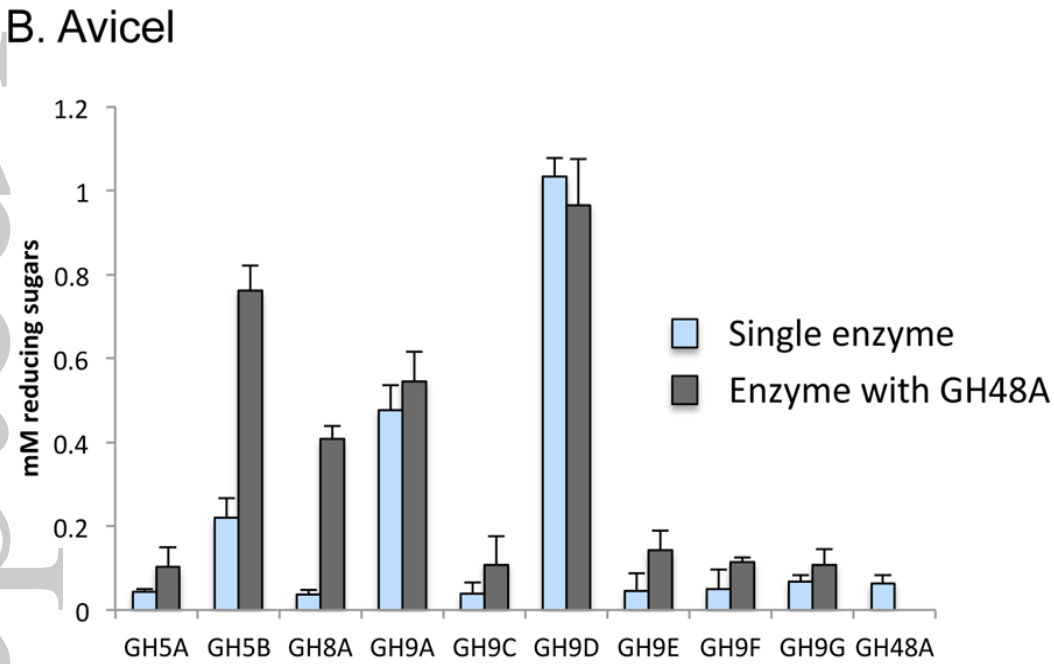
Ze, X., Le Mougen, F., Duncan, S.H., Louis, P., and Flint, H.J. (2013) Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant substrates. *Gut microbes* **4**: 236-240.

Zhang, X.Z., Sathitsuksanoh, N., and Zhang, Y.H. (2010) Glycoside hydrolase family 9 processive endoglucanase from *Clostridium phytofermentans:* heterologous expression, characterization, and synergy with family 48 cellobiohydrolase. *Bioresour Technol* **101**: 5534-5538.

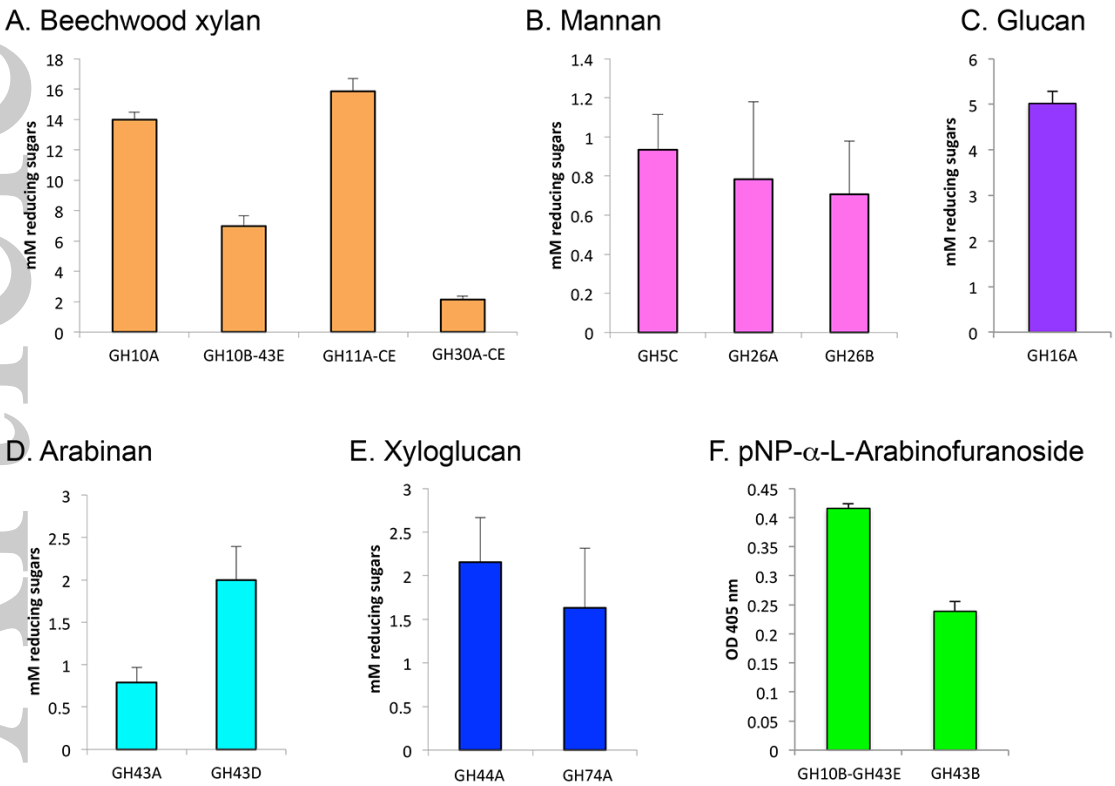
Zverlov, V.V., Schantz, N., Schmitt-Kopplin, P., and Schwarz, W.H. (2005) Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. *Microbiology* **151**: 3395-3401.

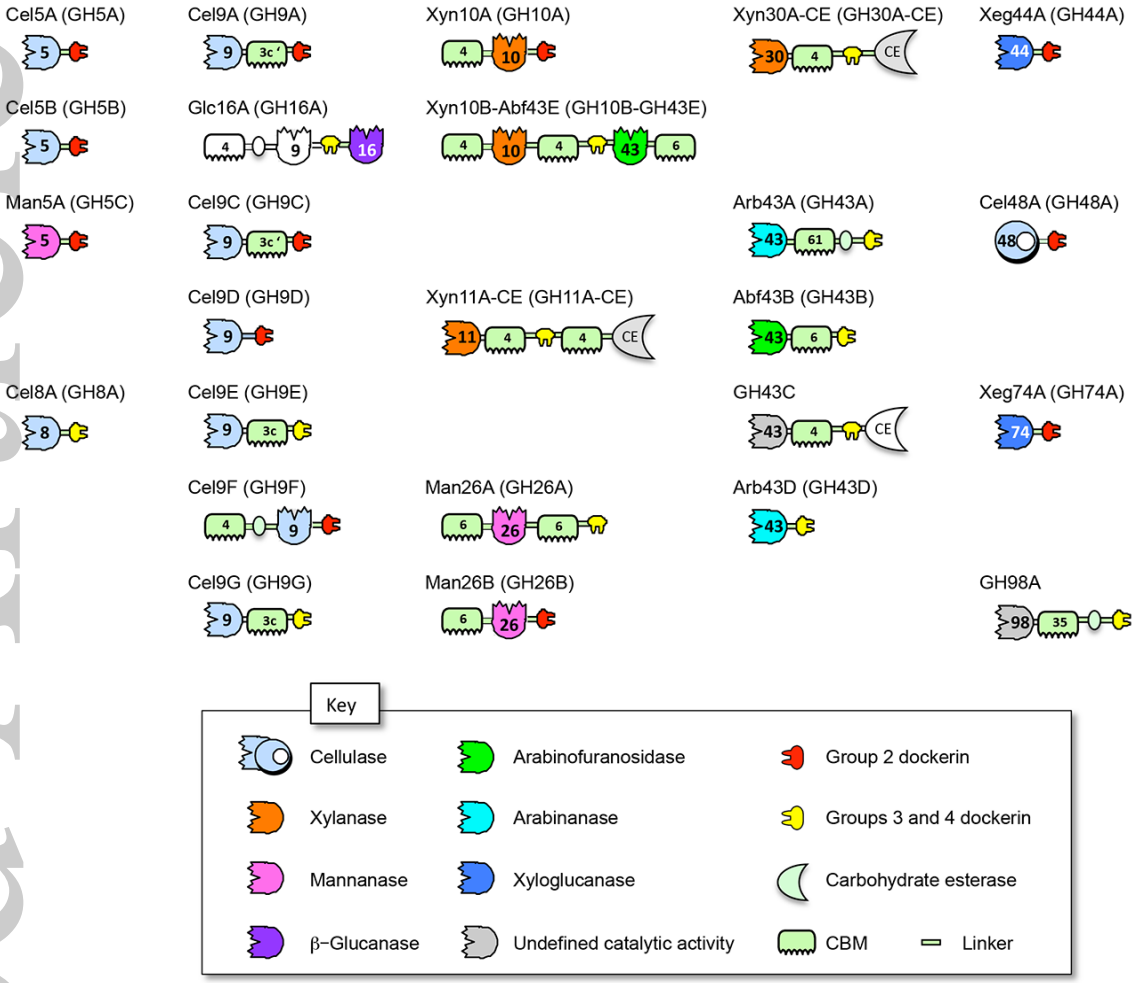


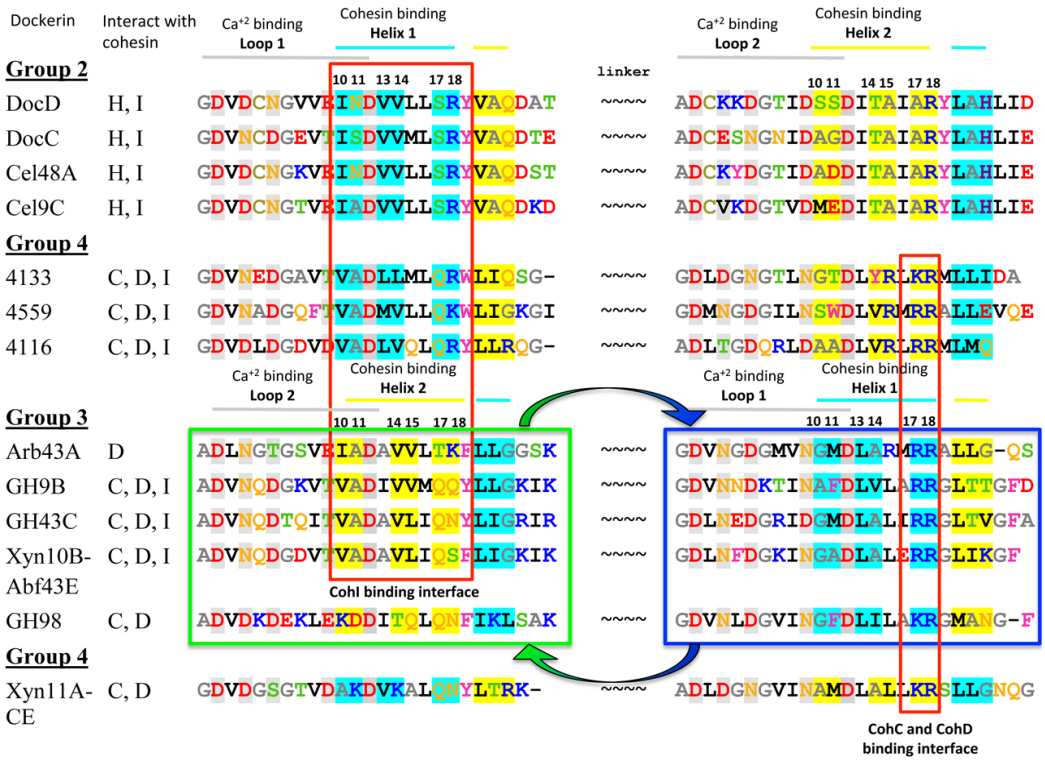

EMI_13047_F1




EMI_13047_F2




EMI_13047_F3




EMI_13047_F4

**Table 1:** Putative dockerin-containing glycoside hydrolases of *R. champanellensis*

| GH Family | Current Nomenclature | Modular Organization | Molecular weight[#] | GI number |
|---|---|---|---|---|
| **GH5** | GH5A | GH5A-Doc | 68695 Da | 291543414 |
| | GH5B | GH5B-Doc | 63394 Da | 291543738 |
| | GH5C | GH5C-Doc | 68852 Da | 291545071 |
| **GH8** | GH8A | GH8A-Doc | 51881 Da | 291543899 |
| **GH9** | GH9A | GH9A-CBM3c'-Doc | 93340 Da | 291543282 |
| | GH9B | CBM4-Fn3-GH9B-Doc-GH16A (Doc-GH16A)* | 114666 Da 36367 Da | 291543673 |
| | GH9C | GH9C-CBM3c'-Doc | 98050 Da | 291543938 |
| | GH9D | GH9D-Doc | 64377 Da | 291544445 |
| | GH9E | GH9E-CBM3c-Doc | 82111 Da | 291544574 |
| | GH9F | CBM4-Fn3-GH9F-Doc | 104833 Da | 291544575 |
| | GH9G | GH9G-CBM3c-Doc | 79955 Da | 291545280 |
| **GH10** | GH10A | CBM4-GH10A-Doc | 69424 Da | 291543470 |
| | GH10B-GH43E | CBM4-GH10B-CBM4-Doc-GH43E-CBM6 | 137621 Da | 291544573 |
| **GH11** | GH11A-CE | GH11A-CBM4-Doc-CBM4-CE | 94529 Da | 291545196 |
| **GH26** | GH26A | CBM6-GH26A-CBM6-Doc | 79544 Da | 291544512 |
| | GH26B | CBM6-GH26B-Doc | 68166 Da | 291545037 |
| **GH30** | GH30A | GH30A-CBM4-Doc-CE | 104949 Da | 291544794 |
| **GH43** | GH43A | GH43A-CBM61-X157-Doc | 79531 Da | 291543994 |
| | GH43B | GH43B-CBM6-Doc | 80395 Da | 291543991 |
| | GH43C | GH43C-CBM4-Doc-CE† GH43C-CBM4-Doc† | 118020 Da 83891 Da | 291544122 |
| | GH43D | GH43D-Doc | 83137 Da | 291544405 |
| **GH44** | GH44A | GH44A-Doc | 81929 Da | 291543699 |
| **GH48** | GH48A | GH48A-Doc | 88132 Da | 291544207 |
| **GH74** | GH74A | GH74A-Doc | 92496 Da | 291543413 |
| **GH98** | GH98A | GH98A-CBM35-X157-Doc | 114519 Da | 291544973 |

Abbreviations used in the table: GH, glycoside hydrolase; Doc, dockerin; CBM, carbohydrate binding module; Fn3, fibronectin type 3 motif; CE, carbohydrate esterase; X157, domain of unknown function.

[#] Based on the known amino acid composition of the desired protein using the Protparam tool (http://www.expasy.org/tools/protparam.html)

* Entire enzyme could not be cloned, only the GH16 and dockerin modules.

†The complete protein could not be cloned, and the CE module was thus omitted.

**Table 2.** Cell-associated proteins from *R. champanellensis* 18P13 cultures showing differential expression during growth on cellobiose or filter paper cellulose as sole energy sources

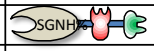
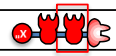
| Spot ID | Fold-change | p-value | Protein hit | Score | % Coverage | Closest match to *R. champanellensis* 18P13 |
|---|---|---|---|---|---|---|
| ***Filter paper cellulose > cellobiose*** | | | | | | |
| 3620 | 364.2 | <0.001 | gi:291544207 | 960 | 23 | Cel48A |
| 3606 | 185.6 | <0.001 | gi:291544575 | 1262 | 30 | Cel9F |
| 4607 | 97.35 | <0.001 | gi:291543571 | 84 | 2 | pyruvate, phosphate dikinase |
| 2406 | 70.19 | <0.001 | gi:291544494 | 573 | 34 | cell division protein FtsZ |
| 7310 | 61.00 | 0.002 | gi:291544534 | 327 | 21 | deoxy-D-arabinoheptulosonate phosphate synthase |
| 6301 | 36.56 | 0.006 | gi:291544494 | 93 | 8 | cell division protein FtsZ |
| 7312 | 28.53 | 0.003 | gi:291543397 | 603 | 31 | carbohydrate ABC transporter ATP BP CUT1 family |
| 4108 | 12.32 | <0.001 | gi:291543600 | 308 | 37 | hypothetical protein RUM04790 |
| 4609 | 8.73 | 0.011 | gi:291544244 | 2207 | 9 | elongation factor |
| 7311 | 5.82 | <0.001 | gi:291545194 | 670 | 33 | glutamate dehydrogenase |
| 6307 | 2.10 | 0.002 | gi:291543339 | 175 | 12 | branched chain amino acid transferase apoenzyme |
| ***Cellobiose > Filter paper cellulose*** | | | | | | |
| 4212 | 117.50 | 0.002 | gi:291543615 | 187 | 21 | CheY receiver domain |
| 4403 | 81.50 | 0.039 | gi:524639232 | 716 | 39 | elongation factor Tu |
| 4206 | 77.17 | 0.026 | gi:291543975 | 343 | 42 | tryptophan synthase |
| 4404 | 67.63 | 0.031 | gi:291545113 | 712 | 18 | hydroxy-methyl but-2-enyl phosphate reductase |
| 6105 | 42.65 | 0.002 | gi:291544482 | 170 | 17 | SSU ribosomal protein S13P |
| 4103 | 40.00 | 0.001 | gi:291544325 | 238 | 30 | translation elongation factor EFP |
| 5005 | 35.63 | 0.003 | gi:291544576 | 173 | 32 | hypothetical protein RUM15970 |
| 6204 | 11.86 | 0.009 | gi:291544387 | 251 | 28 | transcription elongation factor GreA |
| 6408 | 7.18 | 0.010 | gi:291543396 | 883 | 51 | pyridoxal phosphate-dependent Trp B-like protein |
| 5101 | 5.36 | 0.014 | gi:291544471 | 201 | 27 | SSU ribosomal protein S8P |

| | | | | | |
|---|---|---|---|---|---|
| 3203 | 4.21 | 0.009 | gi:291544054 | 1041 | 63 |
| fructose 1,6 bisphosphate aldolase | | | | | |
| 6107 | 3.14 | 0.005 | gi:291544476 | 232 | 22 |
| LSU ribosomal protein L15P | | | | | |
| 5004 | 3.03 | 0.010 | gi:291544048 | 146 | 22 |
| uncharacterized protein | | | | | |

Table 3: List of the *R. champanellensis* CBM-fused cohesin proteins used in this article. Name and modular architecture of the original scaffoldin are given. Abbreviations: CBM, CBM3a from the *C. thermocellum* CipA scaffoldin; SIGN, signal peptide; Doc, dockerin; Coh, cohesin; GH, glycoside hydrolase; SGNH, lipases or esterases; SORT, sortase motif; NUC, nucleoporin like module; UNK, X, unknown.

| Fused cohesin | Emerging scaffoldin | Modular architecture |
|---|---|---|
| CBM-CohA2 | ScaA | SIGN X Coh Coh Doc |
| CBM-CohB1/B2/B3 | ScaB | SIGN Coh Coh Coh Coh Coh Coh Coh X Doc |
| CBM-CohB4 | ScaB | SIGN Coh Coh Coh Coh Coh Coh Coh X Doc |
| CBM-CohB5/B6 | ScaB | SIGN Coh Coh Coh Coh Coh Coh Coh X Doc |
| CBM-CohB6 | ScaB | SIGN Coh Coh Coh Coh Coh Coh Coh X Doc |
| CBM-CohB7 | ScaB | SIGN Coh Coh Coh Coh Coh Coh Coh X Doc |
| CBM-CohC | ScaC | SIGN Coh UNK Doc |
| CBM-CohD | ScaD | SIGN Coh Doc |
| CBM-CohE | ScaE | SIGN Coh SORT |
| CohF-CBM | ScaF | SIGN Coh Doc |
| CBM-CohG | ScaG | SIGN Coh Doc |
| CBM-CohH | ScaH | SIGN SGNH Coh Doc |
| CBM-CohI | ScaI | SIGN Coh |
| CohJ1-CBM | ScaJ | SIGN Coh Coh Coh Doc |
| CBM-CohJ2 | ScaJ | SIGN Coh Coh Coh Doc |
| CBM-CohJ3 | ScaJ | SIGN Coh Coh Coh Doc |
| CBM-CohK | ScaK | SIGN Coh GH25 |
| CBM-CohL | ScaL | SIGN UNK NUC Coh Doc |
| CBM-CohCc (-) | *Clostridium cellulolyticum* CipC | (Negative control) |

**Table 4:** Cohesin-dockerin interactions in *R. champanellensis*: Summary of ELISA experiments. Twenty-five dockerin-containing enzymes from Groups 2, 3 and 4 (rows) were checked against 22 cohesins (only reactive cohesins and negative control are presented in the columns). The schematic modular architecture of the original scaffoldin is represented, cohesins interacting with Group 2 dockerins are represented in red and cohesins interacting with Groups 3 and 4 dockerins are represented in yellow, color-coded according to the scheme in Figure 3. Interaction intensity (number of pluses) was defined as the intensity of the absorbance at 450 nm. The dockerin-containing enzymes were coated at 1 □g/ml, and the CBM-fused CohH, CohI, CohA2, CohB1/B2/B3, CohB4, CohB5/B6, CohC, CohD or CohCc (from *Clostridium cellulolyticum* as negative control) were used at 100 ng/ml. Reactions were performed at least three times in triplicate.

| | | CBM-CohH | CBM-CohI | CBM-CohA2 | CBM-CohB1/B2/B3 | CBM-CohB4 | CBM-CohB5/B6 | CBM-CohC | CBM-CohD | CBM-CohCc |
|---|---|---|---|---|---|---|---|---|---|---|
| | |  |  |  |  |  |  |  |  | (-) |
| Group 2 | Cel5A | ++++ | +++ | +++ | +++ | +++ | +++ | | | |
| | Cel5B | +++ | +++ | | | | + | | | |
| | Man5A | +++ | | +++ | +++ | | | | | |
| | Cel9A | ++++ | +++ | ++++ | ++++ | +++ | +++ | | | |
| | Cel9C | +++ | + | + | + | | | | | |
| | Cel9D | | | | | | + | | | |
| | Cel9F | + | | + | + | | | | | |
| | Xyn10A | ++ | + | + | + | | | | | |
| | Man26B | ++++ | + | ++++ | ++++ | | + | | | |
| | Xeg44A | + | | | | | | | | |
| | Cel48A | ++++ | + | ++++ | ++++ | | | | | |
| | Xeg74A | ++++ | +++ | ++++ | ++++ | ++ | ++ | | | |
| Groups 3 and 4 | Cel8A | | | | | | | | | |
| | Glc16A | | ++++ | | | | | ++++ | ++++ | |
| | Cel9G | | +++ | | | | | +++ | +++ | |
| | Xyn10B-Abf43E | | + | | | | | +++ | +++ | |
| | Man26A | | ++ | | | | | +++ | +++ | |
| | Arb43A | | + | | | | | + | ++ | |
| | Abf43B | | + | | | | | ++++ | ++++ | |
| | GH43C | | ++++ | | | | | ++++ | ++++ | |
| | Arb43D | | ++ | | | | | ++ | ++ | |
| | Xyn11A-CE | | | | | | | ++++ | ++++ | |
| | Xyn30A-CE | | | | | | | ++ | +++ | |
| | GH98A | | | | | | | ++ | ++ | |
| | Cel9E | | ++ | | | | | ++ | | |