Catalytic subunit exchanges in the cellulosomes produced by Ruminiclostridium cellulolyticum suggest unexpected dynamics and adaptability of their enzymatic composition

Romain Borne, Minh-uyen Dao Ti, Henri-pierre Fierobe, Nicolas Vita, Chantal Tardif, Sandrine Pages

Catalytic subunits exchanges in the cellulosomes produced by *Ruminiclostridium cellulolyticum* suggest unexpected dynamics and adaptability of their enzymatic composition

Romain Borne, Minh-Uyen Dao Ti, Henri-Pierre Fierobe, Nicolas Vita, Chantal Tardif and Sandrine Pagès.

Aix-Marseille University, CNRS, LCB, Marseille, France.

Correspondance : LCB-CNRS, UMR7283, 31, Chemin Joseph Aiguier, 13402, Marseille Cedex 20, France. Tel: + 33 (0)4 91 16 45 47 E-mail: pages@imm.cnrs.fr

Running title : Unexpected dynamics in the cellulosomes of *R. cellulolyticum*.

Keywords

R. cellulolyticum, Type-I dockerins, cohesins, cellulosomes, dual binding mode.

Abbreviations

AA : Auxiliary Activities
CBM : Carbohydrate-Binding Modules
CE : Carbohydrate Esterases
GH : Glycoside Hydrolases
GT : GlycosylTransferases
MC1 : Miniscaffoldin MiniCipC-1
PL : Polysaccharide Lyases

Abstract

Cellulosomes are complex nano-machines produced by cellulolytic anaerobic bacteria such as *Ruminiclostridium cellulolyticum* (formerly known as *Clostridium cellulolyticum*). Cellulosomes are composed of a scaffoldin protein displaying several cohesin modules on which enzymatic components can bind to through their dockerin module. Even if the cellulosomes have been studied for decades, very little is known about the dynamics of the assembly. We have investigated the ability of some dockerin-bearing enzymes to chase the catalytic subunits already

bound onto a miniscaffoldin displaying a single cohesin. It appeared that the stability of the preassembled enzyme-scaffoldin complex depends of the dockerin nature, and we have identified a key position in the dockerin sequence involved in the stability of the complex with the cohesin. Depending upon the nature of the residue occupying this position, the dockerin can stringkingly establish with the cohesin partner whether a nearly irreversible or a reversible interaction, independently of the catalytic domain associated with the dockerin. Site directed mutagenesis of this residue converts a highly stable complex forming dockerin into a reversible complex forming one with the miniscaffoldin and vice versa. We also show that refunctionalization can occur with natural purified cellulosomes. Altogether, our results enlighten the dynamics of the cellulosomes, especially their capacity to be remodeled even after their assembly is "achieved", suggesting an unforeseen adaptability of their enzymatic composition over time.

Introduction

The microbial degradation of the plant cell wall involves a consortium of carbohydrate-active enzymes (CAZymes), acting in synergy, to degrade this highly recalcitrant substrate. These enzymes, referenced in the CAZY database [1,2], belong to families of Glycoside Hydrolases (GHs), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), GlycosylTransferases (GT) and Auxiliary Activities (AA) [3,4]. They are frequently appended with non-catalytic domains such as the Carbohydrate-Binding Modules (CBM) or dockerins. Most of the CAZymes secreted by anaerobic Gram positive bacteria such as *Ruminiclostridium cellulolyticum* (formerly known as *Clostridium cellulolyticum*), *Hungateiclostridium thermocellum and Hungateiclostridium clariflavum* (formerly known as *Clostridium thermocellum* and *Clostridium clariflavum)*, *Clostridium cellulovorans*, *Ruminococcus albus* or *Ruminococcus flavefaciens*, are gathered into large multi-enzyme complexes termed "cellulosomes" [5–10]. Most simple cellulosomes share similar architectures based on a central non catalytic scaffolding subunit, called the scaffoldin, harboring a Family 3a CBM, at least one "X" module of unknown function, and 3 to 15 repeated type-I cohesin modules [10–15]. Cohesins allow the anchoring of cellulosomal CAZymes through their specific dockerin module. The high affinity of the cohesin-dockerin interaction is the keystone of cellulosomes assembly. The reported K_D values of the interaction range from 10^{-10} to 10^{-8} M [16–18]. The mechanical stability of the interaction has also been recently addressed and for instance, forces ranging from 150 to 225pN are necessary to dissociate the Type-I cohesin-dockerin complex from *H. thermocellum* [19–21]. The measured rupture forces are comparable to the strongest known biomolecular interactions e.g., streptavidin–biotin (230pN) [22]. In some organisms, like *R. flavefasciens* or *H. thermocellum* Type-II and Type-III dockerins and cohesins

are involved in the formation of polycellulosomes [23–26]. These organized multi-enzyme complexes induce the concentration of various enzyme, promoting synergy and reducing the dispersion of soluble sugars.

Whole genome analysis of *R. cellulolyticum* indicates that 62 genes potentially encode Type-I dockerin-bearing proteins which can be incorporated into the cellulosomes [27,28]. The mature scaffoldin called CipC is composed of an N-terminal CBM3a, a X2 module of unknown function followed by 7 Type-I cohesins, a second X2 module and an eighth Type-I cohesin at the C-terminus [5]. Primary structures of the seven central cohesins of CipC are highly similar, sharing more than 80 % of identity, whereas the C-terminal cohesin is "only" 60% identical to the central cohesins (Supplementary Fig. 1) [11,29]. Type-I cohesins structure is composed of a nine-stranded flattened β-barrel with jelly-roll topology, defined by two β-sheets and stabilized by a tightly packed aromatic/hydrophobic core [30,31].

Type-I dockerins, which are well conserved among cellulosomes-synthesizing bacteria, are composed of a pair of a 22-residue segment spaced by a linker of 8–18 residues. Each segment contains an EF-hand calcium-binding loop [32], as illustrated in Figure 1a. The consensus logo sequence of the dockerins of *R. cellulolyticum* indicates a high level of conservation, in particular for the aspartate residues at positions 2, 6, 10, 13, 38, 42, 46, and 49, involved in Ca^{2+} coordination (Fig 1a and 1b). Two residues per segment, positions 11-12 in the first segment and 47-48 in the second one (Fig 1b), in most cases occupied by the dyad AI in *R. cellulolyticum* and the dyad ST in *H. thermocellum* drive the species specificity of the cohesin-dockerin interaction [33–35]. Each dockerin segment displays the same overall structure as reflected by the internal symmetry due to a 180° rotation between the N-terminal and C-terminal helices. These two helices can be structurally superimposed and play equivalent roles in cohesin recognition, but both segments of the dockerin are necessary to interact with the cohesin [16,36]. This structural symmetry confers a dual binding mode with two non-preferred possible orientations of the Type-I dockerin's α-helices [37–39], leading to a conformational heterogeneity of the cohesin-dockerin complexes population that long impeded the resolution of their crystal structure. Nevertheless, the use of specificity residue substitution in only one conserved segment of *R. cellulolyticum*'s and *H. thermocellum*'s dockerin (Cel5A and Xyn10B respectively) generates homogeneous crystallizable complexes by favorizing a single binding mode [38,39]. Despite the fact that the dockerin dual binding mode has been proposed to reduce the steric constraints imposed by assembling a large number of different enzymes into a single cellulosome, single binding mode have also been reported for some specific cases as the dockerin of Cel124A exclusively binding to the Type-I cohesin displayed by membrane-bound scaffoldin ScaG (previously named OlpC) in *H. thermocellum* [40,41].

In cellulosomes, Type-I dockerins appear able to interact with all scaffoldin's Type-I cohesins with none or few discrimination [29,42], suggesting that they are likely to assemble randomly. The enzyme content of cellulosomes is thus adaptable, depending on the growth substrate and the availability of dockerin-bearing enzymes. When *R. cellulolyticum* grows on wheat straw for instance, a specific *xyl-doc* gene cluster is expressed and the encoded hemicellulases are massively produced and incorporated in newly synthesized cellulosomes [28,43,44]. Besides the CAZymes availability, other mechanisms may govern the cellulosomes composition. Computational simulations suggested that depending upon their molecular mass, shape and modular architecture, some *H. thermocellum* cellulosomal enzymes would statistically bind more often than others [45]. Moreover, it has been reported that the preliminary binding of an enzyme onto a given cohesin exerts an influence on the occupancy of the adjacent cohesins [46]. Despite the studies mentioned above, many questions remain unanswered concerning the plasticity of the cellulosomes necessary to adapt to the plant cell wall structure and composition's variation during its deconstruction. It is not known if rearrangements can still occur even after the cellulosomes assembly is achieved: on the one hand, it has been proved that the lack of enzyme mobility, within artificial mini-cellulosomes negatively affects their activity [47]; on the other hand, enzymes displacements appear unlikely considering the strength of the cohesin-dockerin interaction, which seems nearly irreversible. Nevertheless, the production of the different CAZymes is regulated and depends on the available polymers during the plant cell wall decontruction. Are the newly synthesized dockerin-bearing CAZymes incorporated in the already formed cellulosomal particles or only recruited in *de novo* assembled complexes?

In this present study we have investigated the dynamic of the cellulosome assembly using the miniscaffoldin MiniCipC-1 (MC1) composed of the first 3 modules of CipC encompassing the first cohesin. We have performed chase experiments on preassembled complexes to unravel the ability of some enzymes to take the place of already complexed ones. We have identified key residues in the dockerin sequence that can influence the stability of the interaction and shown that the composition of the cellulosomes can be modified even after their assembly.

Results

Displacement of preassembled enzymes in complex with MC1 by competitor enzymes

To address the question of the reorganization of cellulosomes, we have chosen to use a mini cellulosome composed of the miniscaffolding protein MC1 containing a single cohesin (Coh1),

and one enzyme bearing a dockerin (Cel5A, Cel48F, Cel9G and Man5K). The migration on native gel of each complex and each free enzyme is shown in Fig 2a. Results show clearly that free enzymes migrates on the bottom of the gel, and the various complexes on the top. We have investigated the stability of the interaction of the well-studied enzymes bearing a dockerin Cel5A, Cel48F, Cel9G, and Man5K complexed with the MC1, by challenging each individual complex with one of the three others free proteins by adding an equimolar amount of competitor to the pre-formed tested complex. The non-denaturing gel analyses of the protein mixtures shows that the pre-assembled MC1-Cel48F complex is not dissociated upon the addition of either Cel9G, Cel5A, or Man5K as the competitor and only trace amounts of complexed competitor can be observed meaning that MC1-Cel48F complex is stable (Fig 2b). For simplification only the top of the gels showing the complexes formed are presented. Similar results were obtained for MC1-Cel9G (Fig 2c) and MC1-Man5K (Fig 2d) contrary to the MC1-Cel5A complex for which the major complex detected in all cases is the MC1-competitor one (Fig 2e). Only few amounts of the original MC1-Cel5A complex can be visualized on the non-denaturing gel suggesting that Cel5A was almost completely chased by Cel48F, Cel9G or Man5K. Our results indicate that Cel48F, Cel9G and Man5K make a stable interaction with cohesin 1, whereas the MC1-Cel5A interaction seems rather unstable as it dissociates in presence of any of the tested competitors.

The dockerin of Cel5A displays a non-canonical species specificity pair of residues in the second segment where the dyad AF replaced the consensual AI (Supplementary Fig. 2). Comparison of the dockerins sequences of the 62 *R. cellulolyticum*'s dockerin bearing enzymes revealed that 24 of them display at least one non-canonical residue at the "specificity positions" 11-12 (first segment) or 47-48 (second segment), and 9 of them exhibit variations in the specificity residues in both segments (Supplementary Fig. 2). We focused on two enzymes encoded by the genes at loci Ccel_0286 and Ccel_1101 whose dockerins both display a Ser instead of Ala in the first segment. Since these proteins have not been yet characterized, they were named Rc0286 and Rc1101 according to their loci numbers in the genome of *R. cellulolyticum.* These two enzymes were overproduced in *E. coli*, purified and the stability of their interaction with MC1 was analyzed. The initial complex MC1-Rc1101 appeared to be destabilized by the tested competitors, since MC1-Cel48F, MC1-Cel9G, and MC1-Man5K became the major complexes, meanwhile Cel5A was unable to chase Rc1101 from MC1, as the most abundant complex remains MC1-Rc1101 in that particular case (Fig 3a). In contrast, the original complex MC1-Rc0286 appeared to be stable as none of the tested competitor enzymes was able to chase Rc0286, as shown in Fig 3b. Thus, variations in the specificity residues cannot explain the lower stability of certain MC1-enzyme complexes.

Involvement of the dockerin in the stability of the complex

To discriminate if the observed above instability of the cohesin-dockerin interaction could be due to the nature of the catalytic module associated to the dockerin and/or the dockerin module itself, the dockerin of Cel48F was grafted to the catalytic module of Cel5A and Rc1101. The resulting chimeric $Cel5A_{48}$ and $Rc1101_{48}$ were then pre-assembled with MC1, and subsequently challenged by either Cel48F, Cel9G, or Man5K as the competitor. MC1-$Cel5A_{48}$ and MC1-$Rc1101_{48}$ complexes display a more stable interaction compared to the corresponding proteins bearing their native dockerin (Fig 4a and 4b). Considering the fact that Cel5A failed to chase Rc1101 in the MC1-Rc1101complex, the same experiment was performed but with the chimeric $Cel5A_{48}$. In contrast to Cel5A bearing its native dockerin, the chimeric $Cel5A_{48}$ efficiently chased Rc1101 (Fig 4c) Altogether these data demonstrate that the dockerin module is the major determinant of the stability/instability trait of the interaction with MC1.

Few dockerin-bearing enzyme display an instable interaction with MC1 and are able to chase each other

21 dockerin-containing enzymes from *R. cellulolyticum* were used in chase assays to test their stability when complexed with MC1, using Cel48F or Cel9G as competitors (Table 1). Among this large panel, which represents a third of the 62 cellulosomal subunits of *R. cellulolyticum*, we have identified a "partially" unstable dockerin bearing protein, Cel9Q, for which half of the initially MC1-Cel9Q complexed is chased by addition of Cel48F (Fig 5). These results suggest that two populations of MC1-Cel9Q complexes with various stability were initially assembled: one interacting strongly and one much less stable where Cel9Q is completely replaced in the complex by Cel48F during the competition assay. Surprisingly, we only detected three dockerin-cohesin unstable complex forming proteins out of the 21 tested (Cel5A, Cel9Q and Rc1101). We then compared the stability level of MC1-Cel9Q, MC1-Cel5A and MC1-Rc1101 complexes using the same approach as described above with Cel5A, Cel9Q and Rc1101 as competitor. On figure 3a, the competition between Rc1101 bound onto MC1 and Cel5A has shown that Cel5A cannot expel Rc1101. On the opposite, when MC1-Cel5A is challenged by Rc1101, Cel5A is removed and MC1-Rc1101 becomes the major complex (Fig 6a). Thus, Cel5A seems to establish a less stable interaction than Rc1101 with MC1. Cel9Q is able to partially displace Cel5A and Rc1101 preassembled onto MC1 (Fig 6b), and these last two enzymes are also able to partially displace Cel9Q from its initial complex (Fig 6c). Then, this enzyme keeps its half-stable/half-unstable interaction when complexed to MC1. This demonstrates that a hierarchy still exist even among these dockerin-bearing enzyme displaying an instable interaction with MC1.

Stable dockerins share a consensus sequence

The sequences of the 18 identified dockerin generating highly stable complexes with MC1 were aligned using Clustal Omega software (Fig 7a), and a consensus logo sequence was generated with WebLogo software (Fig 7b). The comparison of the dockerin sequences from enzymes having a stable interaction with MC1 and the three proteins displaying a reversible binding to MC1 (Fig 7c) highlighted few differences, among which, the most evident is the conserved Leu22 in the first segment for the "stable dockerin" group, which is replaced by a Phe in Rc1101 or a Met in Cel5A and Cel9Q (framed in Fig 7). The equivalent position 58 in the second segment is exclusively occupied by a Leu residue, except for Cel9Q which displays a Gln residue. According to the available crystal structures [39] of the complexed cohesin 1 of CipC (the cohesin present in MC1) and Cel5A, the residues at position 22 and 58 are located at the extremity of the α helix in both segments (Fig 8). Due to the dual binding mode of the cohesin-dockerin interaction, one of this two residue is systematically located at the cohesin interface.

The 22th residue of the dockerin is involved in the stability of the complex with the cohesin

To investigate the role of the residue at position 22, we have generated the following variants: Cel5A$_{dockM22L}$, Rc1101$_{dockF22L}$, and Cel48F$_{dockL22M}$. Chase assays with Cel48F as a competitor and the preformed complexes MC1-Cel5A$_{dockM22L}$ (Fig 9a), MC1-Rc1101$_{dockF22L}$ (Fig 9b), and MC1 complexes containing the corresponding wild type dockerin-bearing proteins (Figs 9d and e) showed that the single mutation in the dockerins of Cel5A and Rc1101 drastically increased the strength of their interaction with MC1, as these modified proteins are virtually no longer chased by the competitor Cel48F, in contrast to their wild type counterparts. Consistently, the Cel48F$_{dockL22M}$ variant exhibited a lower stability than wild type Cel48F, as shown in Fig 9c and 9f. Approximately 50 % of Cel48F$_{dockL22M}$ was displaced by the Cel9G competitor. Altogether these data confirm the essential role of this residue in the stability of the interaction.

***In vitro* reorganization of preassembled cellulosomes**

The ability of some proteins to be incorporated in already formed cellulosomes was tested. Cellulosomes produced by *R. cellulolyticum* during the growth on cellulose-based medium were extracted from residual cellulose by filtration and further purified by gel filtration. Purified cellulosomes were incubated with Rc0286, Man5K or Rc1101 and then purified again by gel filtration. SDS-PAGE analysis of the cellulosomes composition showed that each protein was incorporated in cellulosomes (Fig10a). The Rc1101Δdoc variant, that lacks the dockerin, was used as a negative control and as expected could not be incorporated in the cellulolytic complexes. To rule out the possibility that free cohesins are still available in the cellulosomes extracted from cellulose-grown culture, and may interact with the added competitor enzymes, cellulosomes were

first incubated with Man5K, purified by gel filtration and then incubated again with Rc0286 or Rc1101. As shown in Fig10b, Rc0286 and Rc1101 are still incorporated in cellulosomes enriched by Man5K thereby showing that the cellulosomes' composition can evolve even after their assembly. Moreover, Cel5A, which has been identified in this study as exhibiting the lowest stable interaction with MC1, is unable to integrate the cellulosomes (Fig10c), in contrast to the chimeric $Cel5A_{48}$ bearing the dockerin of Cel48F.

Results

In the present study, we have explored an unexpected property of cellulosomes with respect to the cohesin/dockerin interaction which is considered as one of the strongest protein-protein interactions. The two well-known Cel5A and Cel48F proteins were characterized for their interaction with the first cohesin of the scaffoldin CipC from *R. cellullolyticum* by Surface Plasmon Resonance, and displayed similar sensorgrams and K_D values (1–2.5 10^{-10} M) [16]. Nevertheless, in this study it was observed that the k_{off} rate during the dissociation was slightly higher for Cel5A compared to Cel48F [16]. These data already suggested that complexed Cel5A displayed a higher propensity to dissociate from the immobilized scaffoldin. To our knowledge, cohesin-dockerin interaction studies reported to date were aimed at mixing the interacting partners simultaneously, as it is hypothesized to occur during the assembly of cellulosomes *in vivo*.

Here, the mini-scaffoldin MinicipC1 (MC1) harboring the first cohesin of CipC from *R. cellulolyticum*, were pre-mixed with a dockerin-bearing enzyme and the resulting complex was subsequently incubated with another dockerin-containing enzyme. This approach showed that the prominent cellulase Cel48F was able to chase Cel5A from the complex, whereas Cel5A was unable to displace Cel48F in the preassembled MC1-Cel48F complex. Our results demonstrated that the exchange depends only on some specific residues of the dockerin module since the chimeric $Cel5A_{48}$ composed of the catalytic domain of Cel5A fused to the dockerin of Cel48F is no longer expelled by the competitor. We have experimentally identified only two other dockerins among the 21 tested that display a behavior similar to dockerin of Cel5A, namely the dockerins of Cel9Q and not yet characterized Rc1101. Sequence analysis of the 21 dockerins suggested that the residue at position 22 in the first conserved segment may play an important role with respect to the stability of the interaction with the cohesin: a Leu residue is exclusively found at this position in the "stable dockerins", whereas "unstable dockerins" display either a Met or a Phe residue at this position. It is noteworthy that at the corresponding position in the second segment (position 58) a Leu is observed, even in two of the "unstable dockerins" harbored by Cel5A and Rc1101 (Fig 7).

According to the structures available of the dockerin of Cel5A in complex with cohesin 1 of CipC, these residues are located at the extremity of the α-helix (Fig 8) and one of them is at proximity of the cohesin, depending on the orientation of the dockerin (dual binding mode). The three dockerins from Cel5A, Rc1101, and Cel9Q display variations for at least one of these positions, Met22, Phe22, and Met22/Gln58, respectively. The Leu residue at position 22 was thus suspected to play a role in the stability of the complex, which was confirmed by the much higher stability of the MC1-complexes containing the variants $Cel5A_{dockM22L}$ or $Rc1101_{dockF22L}$.

While the dual binding mode is certainly the rule for most species, there is strong evidence that ruminococcal cellulosomes are assembled almost exclusively through single-binding mode of cohesin-dockerin interactions. The dual binding mode of interaction has been demonstrated for both Xyn10B from *H. thermocellum* [38] and Cel5A from *R. cellulolyticum* [39]. In the case of Cel5A, the structure of the complex made of the wild type dockerin of Cel5A (DocA) and the cohesin 1 from CipC (Coh1) could not be solved, due to conformational heterogeneity. This issue was overcome by using complexes containing mutated dockerins for their specificity residues Coh1-DocA A11S/L12T and Coh1-DocA A47F/S48T, whose crystal structures could be solved, and were shown to be superimposable. With this approach, only one structural conformation was generated during the complexation, as the mutations in the dockerins generate a single binding mode. In figure 8a is reported the structure of the Coh1-DocA A47F/S48T complex that shows that the first segment is primarily involved in the cohesin recognition. In this configuration, Met22 is located at the extremity of the helix bearing the specificity residues, outlying the cohesin-dockerin contact surface, although L58 on the second segment is close to the direct binding surface of the cohesin. In contrast, in figure 8b showing the structure of the Coh1-DocA A11S/L12T complex, Met22 is brought at the cohesin surface whereas the corresponding Leu58 in the second segment is outlying the interface. Indeed this observation revealed that the dockerin of Cel5A can bind to the cohesin 1 of CipC in two different conformations. Nevertheless, no information is available concerning the proportion of each conformation in a wild type context. According the available structures and our results we can hypothesize that on the second conformation (Fig 8b), the residue Met22 should establish a weaker interaction with the cohesin than the one established by the residue dockL58 on the first conformation (Fig 8a). The dual binding mode interaction between Cohesin 1 and wild type dockerin of Cel5A dockerin might occur in an “imbalanced” way, leading to a major proportion of a given predominant conformation. In the case of Cel5A, the second conformation showed in Figure 8b would be predominant as Cel5A is easily chased by all tested competitors, whereas the conformation illustrated in Figure 8a, which is presumably more stable, would be minor.

Cel9Q and Cel48F$_{dockL22M}$ both display a Met at the 22th position of the dockerin. Our data indicate that these two dockerins complexed with MC1 are "half-stable/half-unstable", when competition with a stable dockerin is made. In the population of complexes, one conformation engendered by the dual binding mode would lead to a stable interaction with MC1 whereas the other conformation would be sensitive to the chase. The dockerins of Cel48F$_{dockL22M}$ and Cel9Q could display a "balanced" dual binding mode to the cohesin, both conformations having roughly the same proportion.

The scaffoldin CipC synthesized by *R. cellulolyticum* contains 8 cohesins. Our study focused on the interaction between various dockerin-bearing proteins and the first cohesin from CipC. The other central cohesins share 82% to 98% of identity for Cohesin 1. Nevertheless, the last cohesin (Cohesin 8) located at the C-terminus is more divergent and displays only 57% to 61% identity with the other cohesins (Supplementary Fig. 2). It would be interesting to study if the residues located at the positions 22 and 58 of the dockerins play a similar role in the stability of the complexes with the other cohesins harbored by CipC, especially Cohesin 8.

Discussion

Our results have unraveled an unsuspected putative dynamics in the composition of cellulosomes of *R. cellulolyticum*. The position 22 in the first segment of the dockerin, as well as the symmetrical position 58 in the second segment, seems highly involved in the stability of the cohesin-dockerin interaction in a context of competition for the occupancy of the cohesin. Our data suggest that the conserved Leu22 should provide a more stable interaction than the Met22 residue. In contrast to the species-specificity residues at positions 11-12 and 47-48, which vary among the cellulosome-producing bacteria to fit with the corresponding cohesin interface specificity [48], Leu at positions 22 and 58 are highly conserved among the dockerin-bearing proteins of other bacterial species since the consensus sequences of the dockerins from *R. papyrosolvens, R. josui, C. sufflavum, R. sp.* BNL1100*, C. cellobioparum, R. termitidis, C. cellulovorans,* and *H. thermocellum* harbor the same conserved Leu residue in the two segments [11]. It is interesting to notice that Met residue is mainly the second residue at the critical position of the dockerin consensus in all these species, suggesting our data on *R. cellulolyticum* may apply to other bacterial species, and possibly be the basis of a general mechanism governing cellulosomes dynamics.

At the present time, the number of dockerin-bearing proteins having an unstable mode of binding in *R. cellulolyticum* is unclear. The sequence comparison of all dockerins found in *R. cellulolyticum* reveals that in total 14 dockerins display another residue than Leu at position 22 or

58 (Table 2). All these enzymes, except Rc2530, were identified by mass spectrometry analysis in the cellulosomes produced by *R. cellulolyticum* by Blouzard *et al.* [28]. The presence of these enzymes in the cellulosomes, suggest multiple possibilities of cellulosomes remodeling.

In the present study we propose a functional model of the interaction between the dockerin and the cohesin in which each conserved segment of the dockerin plays a specific and distinct role depending on the conformation: one segment displays its specificity residues (Ala-Ile dyad for *R. cellulolyticum*) at the interface, leading to a dedicated role of the whole segment for the species specificity recognition, whereas the other segment modulates the stability of the interaction and harbors the residues involved in the stability (Leu, Met…) at the interface (Fig 8). Our data showed that fully assembled cellulosomes complexes could be refunctionalized by the replacement of some subunits, depending on the type of dockerin involved in this competition. In addition, it has been shown that the catalytic domain of cellulosomal enzymes has influence on the type of enzyme occupying adjacent cohesins on the scaffoldin during *de novo* assembly of the cellulosomes, without any involvement of the dockerin sequence in this process (Fig 11). In this perspective, these two complementary phenomenons could be considered as a possible adaptation to the heterogeneous plant cell wall substrate evolution during its degradation. For instance, is has been shwon that the *xyl-doc* gene cluster encoding 14 cellulosomal CAZymes is controlled by a two component system sensing hemicelluloses components [44]. The newly synthetized cellulosomal CAZymes would be gathered onto freshly produced scaffoldin CipC to create *de novo* cellulosomes, but some of them could integrate and refunctionalize the previously assembled cellulosomes anchored to the substrate, thus allowing a faster adaptation to the novel polysaccharides emerging during the plant cell wall deconstruction.

Materials and methods

Strains and plasmids

Vector encoding Rc1101, and Rc0286 from R. cellulolyticum. Using genomic DNA from *R. cellulolyticum* as the template, the DNA regions encoding the mature products of the genes at loci Ccel_0286 and Ccel_1101 were amplified using the forward and reverse primers listed in Supplementary Table 1, introducing an NcoI and XhoI, or NdeI and a XhoI (underlined) sites at the extremities of the amplicon, respectively. The amplicons were cloned in NcoI-XhoI linearized pET28b vector (Novagen, Madison, WI), to generate pET-Rc0286, and in NdeI-XhoI linearized pET22b vector (Novagen), to generate pET-Rc1101, respectively. Each vector will allow the production of a C-terminal 6-Histidine tagged recombinant protein. Since the dockerin of both

enzymes is located at the N-terminus, the 6 His codons were added in the forward cloning primer (Supplementary Table 1). Ligation mixture was used to transform competent Neb5α (New England Biolabs, Ipswich, MA). Clones containing the insert were selected by PCR on colony, and checked by sequencing. All the vectors used in this study are listed in Supplementary Table 2.

Vector encoding Rc1101 with dockerin from Cel48F. The vector pET-Rc1101$_{48}$ encoding Rc1101 appended with the native *R. cellulolyticum* dockerin of Cel48F at the N-terminus was obtained by overlap extension PCR. The DNA encoding the catalytic module of Rc1101 was amplified using the template pET-Rc1101 and the DNA encoding the dockerin of Cel48F was amplified from pET-Cel48Fc using the primers listed in Supplementary Table 1. The two overlapping fragments (overlapping region in italics) were mixed, and a combined fragment was synthesized using the external primers. Since the dockerin is located at the N-terminus, a 6 His codons were also added in the forward external primer (Supplementary Table 1). The fragment was cloned into NdeI-XhoI-linearized pET22b(+) (Novagen,), thereby generating pET-Rc1101$_{48}$.

Modifications of the dockerins. To construct Cel5A$_{dockM22L}$, the codon corresponding to residue Met22 of the dockerin from Cel5A was replaced by Leu codon by a one-step PCR method [49] using plasmid pET-5Ac as a template and overlapping divergent primers (Supplementary Table 1). After DpnI treatment, the PCR product were used to transform competent Neb5α. The sequence accuracy of the construction was verified by sequencing.

The vectors pET-Rc1101$_{dockF22L}$, and pET-Cel48F$_{dockL22M}$, in which the codon corresponding to residue Phe22 of the dockerin from Rc1101 was replaced by Leu codon, and the codon corresponding to Leu22 of the dockerin from Cel48F was replaced by Met codon were constructed similarly.

Production and purification of dockerin-containing proteins

The production and purification of Cel5A, Cel5A$_{48}$, Cel48F, Cel9G, Man5K, and miniCipC-1 (MC1) were formerly described [46,50–54]. All the Cel5A variants were produced and purified using the same procedure than Cel5A.

The BL21(DE3) strains overproducing Rc0286 and Rc1101 were grown in 2.5-liter toxin flasks at 37°C in Lysogenic-Broth medium supplemented with glycerol (12 g/L) and kanamycin (50 mg/L) or Ampicillin (100 mg/L) until A_{600nm} = 1.5. The cultures were cooled down and induction of the expression was performed overnight at 20°C with 200 µM isopropyl-thio-*β*-D-galactoside for the strain BL21(DE3) carrying the recombinant vector pET-Rc0286 and 18°C with 100 µM isopropyl-thio-*β*-D-galactoside for the strain BL21(DE3) carrying pET-Rc1101. After 16h, the cells were harvested by centrifugation (5000 *g*, 15 min), resuspended in 30 mL of 30 mM Tris-HCl, pH 8.0, 1 mM $CaCl_2$, supplemented with a few milligrams of DNase I (Roche, Mannheim,

Germany), and broken using a French press. The crude extract was centrifuged 15 min at 15,000 *g* and loaded onto 1 mL of nickel-nitrilotriacetic acid resin (Qiagen, Vanloo, The Netherlands) equilibrated in the same buffer. The proteins of interest were then eluted with 100 mM imidazole in 30 mM Tris-HCl, pH 8.0, 1 mM $CaCl_2$. The purification of the recombinant proteins was achieved on monoQ-Sepharose column (GE Healthcare, Chicago, US) equilibrated in 30 mM Tris-HCl, pH 8.0, 1 mM $CaCl_2$. The proteins of interest were eluted by a linear gradient of 0–500 mM NaCl in 30 mM Tris-HCl, pH 8.0, 1 mM $CaCl_2$.

The purified proteins were dialyzed by ultrafiltration against 10 mM Tris-HCl pH 8.0, 1 mM $CaCl_2$, and stored at -80°C. The concentration of the proteins was estimated by absorbance at 280 nm in 25 mM sodium phosphate, pH 6.5, using the program ProtParam tool (www.expasy.org/tools/protparam.html, Lausanne, Switzerland).

Code de champ modifié

Complex formation and chase monitored by electrophoresis.

Assembly of the minicellulosomes. To ensure the saturation of MC1 with its dockerin-bearing partner, the same amount of MC1 (10µM final concentration) was mixed with the partner protein at concentration ranging from 8 to 12 µM for 5 minutes. Saturation of MC1 was determined on non denaturing PAGE, by loading 4 µL on a 4-15% gradient gel using a Phastsystem apparatus (GE Healthcare). Migration was made using a global applied power of 115AVh, which allows full separation of MC1-Cel48F, MC1-Cel9G, MC1-Cel5A, and MC1-Man5K complexes. The amount of dockerin-containing partner requested to saturate 10µM of MC1 was thus defined in each case and subsequently used to perform the chase assays as it is illustrated for MC1-Cel48F in Supplementary Fig. 3.

Chase assay. Complexes (at a concentration of 20 µM) were assembled as described above at room temperature, in 20 mM Tris-maleate, pH 6.0, 1 mM $CaCl_2$. The dockerin bearing competitor (at 20 µM) was prepared the same buffer. Then, 5µL of 2X MC1-partner were mixed to 5µL of 2X competitor or 1X buffer solution. The mixtures containing MC1-Cel5A or MC1-Cel9G preassembled complexes with Cel48F as a competitor were incubated for five minutes to 2 hours at room temperature , before 4 µL were subjected to non denaturing PAGE 4-15% gradient with a global power of 115AVh (Supplementary Fig. 3). These experiments indicated that a competition duration of 5 minutes with an equimolar amount of competitor was found sufficient to achieve the chase experiments. Moreover, it was observed that adding 3- or 9-times excess of competitor does not increase the chase (Supplementary Fig. 3).

Dockerin sequences logos.

The sequences of dockerins of the 62 cellulosomal subunits produced by *R. cellulolyticum* were aligned using Clustal Omega website (https://www.ebi.ac.uk/Tools/msa/clustalo/, EMBL-EBI,

Code de champ modifié

Cambridgeshire, UK). Aligned sequences (Supplementary Fig. 2) were then combined on WebLogo website to obtain the graphical representation of the *R. cellulolyticum* dockerins amino-acid sequences (https://weblogo.berkeley.edu/logo.cgi, University of California, Berkeley, US). The Sequences of the dockerins from the proteins having a stable interaction with MC1 were extracted from the original alignment to get the stable dockerin logo.

Code de champ modifié

Structural analysis of Cel5A dockerin – MC1 cohesin complex using PyMOL.

Crystal structures 2VN5 and 2VN6 [Pinheiro 2008] corresponding to cohesin 1 module from CipC complexed with the dockerin modules of Cel5A$_{dockA11S-L12T}$, and Cel5A$_{dockA47S-L48T}$, were analysed with PyMOL software 1.1 (Delano Scientific LLC, Portland, US), respectively.

Cellulosomes preparation

R. cellulolyticum ATCC 35319 was grown anaerobically at 32˚C on 500 mL basal medium supplemented with 5 g/L cellulose (Sigmacell 20, Sigma-Aldrich St Louis, MO). After 6 days of growth, the cell culture was filtered through a 3 µm pore size GF⁄D glass filter (Whatman-GE Healthcare). The residual cellulose was washed with 50 and 12.5 mM phosphate buffer (pH 7.0). Cellulosomes fraction was then eluted with 100mL water, dialyzed and concentrated in 20 mM Tris-HCl buffer (pH 8.0), 150 mM NaCl and 1 mM $CaCl_2$. The protein concentration of the cellulosome preparation was determined by Lowry assay, and aliquots of 800 µg were used for each gel filtration experiment.

Gel filtration

800 µg of cellulosomes were incubated with 200 µg of dockerin-bearing protei for 1 hour at 4˚C in the total volume of 2 ml in 20 mM Tris-HCl buffer (pH 8.0), 150 mM NaCl and 1 mM $CaCl_2$. New cellulosomes were separated by gel filtration. Gel-filtration chromatography was performed using a HiLoad 26⁄60 Superdex 200 column (Amersham Biosciences) equilibrated with 20mM Tris-HCl buffer (pH 8.0), 150 mM NaCl and 1 mM $CaCl_2$. Eluted cellulosomes were concentrated and dialyzed against 20 mM Tris-HCl buffer (pH 8.0), 1 mM $CaCl_2$ and analyzed by SDS-PAGE.

References

1 Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233–D238.

2 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM & Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-495.

3 Nguyen STC, Freund HL, Kasanjian J & Berlemont R (2018). Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. *Appl. Microbiol. Biotechnol.* **102**, 1629–1637.
4 André I, Potocki-Véronèse G, Barbe S, Moulis C & Remaud-Siméon M (2014) CAZyme discovery and design for sweet dreams. *Current Opinion in Chemical Biology* **19**, 17–24.
5 Gal L, Pagès S, Gaudin C, Bélaïch A, Reverbel-Leroy C, Tardif C & Bélaïch JP (1997) Characterization of the cellulolytic complex (cellulosome) produced by *Clostridium cellulolyticum*. *Appl. Environ. Microbiol.* **63**, 903–909.
6 Bayer EA, Setter E & Lamed R (1985) Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J. Bacteriol.* **163**, 552–559.
7 Artzi L, Dassa B, Borovok I, Shamshoum M, Lamed R & Bayer EA (2014) Cellulosomics of the cellulolytic thermophile *Clostridium clariflavum*. *Biotechnol Biofuels* **7**, 100.
8 Doi RH, Goldstein M, Hashida S, Park JS & Takagi M (1994) The *Clostridium cellulovorans* cellulosome. *Crit. Rev. Microbiol.* **20**, 87–93.
9 Ohara H, Karita S, Kimura T, Sakka K & Ohmiya K (2000) Characterization of the cellulolytic complex (cellulosome) from *Ruminococcus albus*. *Biosci. Biotechnol. Biochem.* **64**, 254–260.
10 Ding SY, Rincon MT, Lamed R, Martin JC, McCrae SI, Aurilia V, Shoham Y, Bayer EA & Flint HJ (2001) Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens*. *J. Bacteriol.* **183**, 1945–1953.
11 Dassa B, Borovok I, Lombard V, Henrissat B, Lamed R, Bayer EA & Moraïs S (2017) Pan-cellulosomics of mesophilic *Clostridia*: variations on a theme. *Microorganisms* **5**, 74.
12 Artzi L, Bayer EA & Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95.
13 Ding SY, Bayer EA, Steiner D, Shoham Y & Lamed R (2000) A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. *J. Bacteriol.* **182**, 4915–4925.
14 Bayer EA, Belaich J-P, Shoham Y & Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–554.
15 Ding S-Y, Lamed R, Bayer EA & Himmel ME (2003) The bacterial scaffoldin: structure, function and potential applications in the nanosciences. *Genet. Eng. (N.Y.)* **25**, 209–225.
16 Fierobe HP, Pagès S, Bélaïch A, Champ S, Lexa D & Bélaïch JP (1999) Cellulosome from *Clostridium cellulolyticum*: molecular study of the Dockerin/Cohesin interaction. *Biochemistry* **38**, 12822–12832.
17 Mechaly A, Fierobe HP, Bélaïch A, Bélaïch JP, Lamed R, Shoham Y & Bayer EA (2001) Cohesin-dockerin interaction in cellulosome assembly: a single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition. *J. Biol. Chem.* **276**, 9883–9888.
18 Fierobe HP, Mechaly A, Tardif C, Bélaïch A, Lamed R, Shoham Y, Bélaïch JP & Bayer EA (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. *J. Biol. Chem.* **276**, 21257–21261.
19 Stahl SW, Nash MA, Fried DB, Slutzki M, Barak Y, Bayer EA & Gaub HE (2012) Single-molecule dissection of the high-affinity cohesin–dockerin complex. *PNAS* **109**, 20431–20436.
20 Jobst MA, Milles LF, Schoeler C, Ott W, Fried DB, Bayer EA, Gaub HE & Nash MA (2015) Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLife* **4**, e10319.
21 Milles LF & Gaub HE (2019) Is mechanical receptor ligand dissociation driven by unfolding or unbinding? *bioRxiv*, 593335.
22 Sedlak SM, Bauer MS, Kluger C, Schendel LC, Milles LF, Pippig DA & Gaub HE (2017) Monodisperse measurement of the biotin-streptavidin interaction strength in a well-defined pulling geometry. *PLOS ONE* **12**, e0188722.

23 Stern J, Moraïs S, Lamed R & Bayer EA (2016) Adaptor Scaffoldins: An original strategy for extended designer cellulosomes, inspired from nature. *MBio* **7**, e00083.
24 Currie MA, Adams JJ, Ali S, Smith SP & Jia Z (2010) Purification and crystallization of a multimodular heterotrimeric complex containing both type I and type II cohesin-dockerin interactions from the cellulosome of *Clostridium thermocellum*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 327–329.
25 Alber O, Noach I, Rincon MT, Flint HJ, Shimon LJW, Lamed R, Frolow F & Bayer EA (2009) Cohesin diversity revealed by the crystal structure of the anchoring cohesin from *Ruminococcus flavefaciens*. *Proteins* **77**, 699–709.
26 Adams JJ, Pal G, Jia Z & Smith SP (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 305–310.
27 Blouzard J-C, Bourgeois C, de Philip P, Valette O, Bélaïch A, Tardif C, Bélaïch J-P & Pagès S (2007) Enzyme diversity of the cellulolytic system produced by *Clostridium cellulolyticum* explored by two-dimensional analysis: identification of seven genes encoding new dockerin-containing proteins. *J. Bacteriol.* **189**, 2300–2309.
28 Blouzard J-C, Coutinho PM, Fierobe H-P, Henrissat B, Lignon S, Tardif C, Pagès S & de Philip P (2010) Modulation of cellulosome composition in *Clostridium cellulolyticum*: Adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *PROTEOMICS* **10**, 541–554.
29 Pagès S, Bélaïch A, Fierobe H-P, Tardif C, Gaudin C & Bélaïch J-P (1999) Sequence analysis of scaffolding protein CipC and ORFXp, a new cohesin-containing protein in *Clostridium cellulolyticum*: comparison of various cohesin domains and subcellular localization of ORFXp. *Journal of Bacteriology* **181**, 1801–1810.
30 Tavares GA, Béguin P & Alzari PM (1997) The crystal structure of a type I cohesin domain at 1.7 A resolution. *J. Mol. Biol.* **273**, 701–713.
31 Spinelli S, Fiérobe HP, Bélaïch A, Bélaïch JP, Henrissat B & Cambillau C (2000) Crystal structure of a cohesin module from *Clostridium cellulolyticum*: implications for dockerin recognition. *J. Mol. Biol.* **304**, 189–200.
32 Lytle BL, Volkman BF, Westler WM, Heckman MP & Wu JHD (2001) Solution structure of a Type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. *Journal of Molecular Biology* **307**, 745–753.
33 Mechaly A, Yaron S, Lamed R, Fierobe HP, Bélaïch A, Bélaïch JP, Shoham Y & Bayer EA (2000) Cohesin-dockerin recognition in cellulosome assembly: experiment versus hypothesis. *Proteins* **39**, 170–177.
34 Pagès S, Bélaïch A, Bélaïch JP, Morag E, Lamed R, Shoham Y & Bayer EA (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.
35 Jindou S, Soda A, Karita S, Kajino T, Béguin P, Wu JHD, Inagaki M, Kimura T, Sakka K & Ohmiya K (2004) Cohesin-dockerin interactions within and between *Clostridium josui* and *Clostridium thermocellum*: binding selectivity between cognate dockerin and cohesin domains and species specificity. *J. Biol. Chem.* **279**, 9867–9874.
36 Lytle B & Wu JH (1998) Involvement of both dockerin subdomains in assembly of the *Clostridium thermocellum* cellulosome. *J. Bacteriol.* **180**, 6581–6585.
37 Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romão MJ & Fontes CMGA (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13809–13814.
38 Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith N, Bayer EA, Davies GJ, Ferreira LMA, Romão MJ, Fontes CMGA & Gilbert HJ (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3089–3094.

39 Pinheiro BA, Proctor MR, Martinez-Fleites C, Prates JAM, Money VA, Davies GJ, Bayer EA, Fontesm CMGA, Fierobe H-P & Gilbert HJ (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J. Biol. Chem.* **283**, 18422–18430.
40 Brás JLA, Pinheiro BA, Cameron K, Cuskin F, Viegas A, Najmudin S, Bule P, Pires VMR, Romão MJ, Bayer EA, Spencer HL, Smith S, Gilbert HJ, Alves VD, Carvalho AL & Fontes CMGA (2016) Diverse specificity of cellulosome attachment to the bacterial cell surface. *Sci Rep* **6**.
41 Brás JLA, Alves VD, Carvalho AL, Najmudin S, Prates JAM, Ferreira LMA, Bolam DN, Romão MJ, Gilbert HJ & Fontes CMGA (2012) Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode. *J. Biol. Chem.* **287**, 44394–44405.
42 Park J-S, Matano Y & Doi RH (2001) Cohesin-Dockerin Interactions of cellulosomal subunits of *Clostridium cellulovorans*. *J Bacteriol* **183**, 5431–5435.
43 Perret S, Maamar H, Bélaïch J-P & Tardif C (2004). Use of antisense RNA to modify the composition of cellulosomes produced by *Clostridium cellulolyticum*. *Mol. Microbiol.* **51**, 599–607.
44 Celik H, Blouzard J-C, Voigt B, Becher D, Trotter V, Fierobe H-P, Tardif C, Pagès S & de Philip P (2013) A Two-Component System (XydS/R) Controls the expression of genes encoding CBM6-containing proteins in response to straw in *Clostridium cellulolyticum*. *PLOS ONE* **8**, e56063.
45 Bomble YJ, Beckham GT, Matthews JF, Nimlos MR, Himmel ME & Crowley MF (2011) Modeling the self-assembly of the cellulosome enzyme complex. *J. Biol. Chem.* **286**, 5614–5623.
46 Borne R, Bayer EA, Pagès S, Perret S & Fierobe H-P (2013) Unraveling enzyme discrimination during cellulosome assembly independent of cohesin-dockerin affinity. *FEBS J.* **280**, 5764–5779.
47 Mingardon F, Chanal A, Tardif C, Bayer EA & Fierobe H-P (2007) Exploration of new geometries in cellulosome-like chimeras. *Appl. Environ. Microbiol.* **73**, 7138–7149.
48 Slutzki M, Reshef D, Barak Y, Haimovitz R, Rotem-Bamberger S, Lamed R, Bayer EA & Schueler-Furman O (2015) Crucial roles of single residues in binding affinity, specificity, and promiscuity in the cellulosomal cohesin-dockerin interface. *J Biol Chem* **290**, 13654–13666.
49 Ansaldi M, Lepelletier M & Méjean V (1996) Site-specific mutagenesis by using an accurate recombinant polymerase chain reaction method. *Anal. Biochem.* **234**, 110–111.
50 Fierobe HP, Gaudin C, Bélaïch A, Loutfi M, Faure E, Bagnara C, Baty D & Bélaïch JP (1991) Characterization of endoglucanase A from *Clostridium cellulolyticum*. *J. Bacteriol.* **173**, 7956–7962.
51 Reverbel-Leroy C, Bélaïch A, Bernadac A, Gaudin C, Bélaïch JP & Tardif C (1996) Molecular study and overexpression of the *Clostridium cellulolyticum* CelF cellulase gene in *Escherichia coli*. *Microbiology (Reading, Engl.)* **142 (Pt 4)**, 1013–1023.
52 Gal L, Gaudin C, Bélaïch A, Pagès S, Tardif C & Bélaïch JP (1997) CelG from *Clostridium cellulolyticum*: a multidomain endoglucanase acting efficiently on crystalline cellulose. *J. Bacteriol.* **179**, 6595–6601.
53 Perret S, Bélaïch A, Fierobe H-P, Bélaïch J-P & Tardif C (2004) Towards designer Cellulosomes in *Clostridia*: mannanase enrichment of the cellulosomes produced by *Clostridium cellulolyticum*. *J Bacteriol* **186**, 6544–6552.
54 Pagès S, Bélaïch A, Tardif C, Reverbel-Leroy C, Gaudin C & Bélaïch JP (1996) Interaction between the endoglucanase CelA and the scaffolding protein CipC of the *Clostridium cellulolyticum* cellulosome. *J Bacteriol* **178**, 2279–2286.

Tables

Table 1. Stability of MC1 complexes with 21 dockerin-bearing subunits from *R. cellulolyticum,* explored by the chase assay.

Proteins in bold and italics displayed an unstable status when complexed to MC1 and competed by Cel9G or Cel48F. The suspension points indicate that the dockerin is followed by another module in C-terminal.

locus number_Protein	chased by Cel48F or Cel9G	Complex stability
0231_Cel9Q	+	***unstable***
0286_CE4…	-	stable
0429_Cel44O…	-	stable
0729_Cel48F	-	stable
0731_Cel9G	-	stable
0732_Cel9E	-	stable
0734_Cel9H	-	stable
0735_Cel9J	-	stable
0736_Man5K…	-	stable
0737_Cel9M	-	stable
0753_Cel9P	-	stable
0755_Cel9U	-	stable
1099_Cel5A	+	***unstable***
1101_UNK…	+	***unstable***
1230_GH10	-	stable
1234_GH62	-	stable
1249_Cel9T	-	stable
1648_Cel9R	-	stable
1656_UNK	-	stable
2392_Cel9V	-	stable
2621_Cel9X	-	stable

Table 2. Dockerins exhibiting divergent residues from the consensus at positions 22 and 58.

Amino acids are listed for the positions 22 and 58 of the dockerin sequence. "-" indicates the absence of residue 58 in the alignment (Supplementary Fig. 2). Next residue at position 59 indicated into brackets when residue 58 is missing from sequence alignment. Suspension points indicate that the dockerin is followed by another module in C-terminal.

locus number_Protein	Position 22	Position 58
0231_Cel9Q	M	Q
0649_GH5	I	L
0750_GH11	I	- (G59)
0752_Man26A	L	T
0920_CE1	L	- (G59)
0931_Xyn10A	L	I
1060_UNK	M	L
1099_Cel5A	M	L
1101_UNK...	F	L
1232_CE1	L	I
1809_chagasine...	L	M
2017_UNK	M	L
2162_CE2...	M	L
2530_UNK...	M	L

Mis en forme : Centré, Aucun(e), Espace Avant : 0 pt, Après : 0 pt, Taquets de tabulation : Pas à 8 cm + 16 cm

Tableau mis en forme

Legends

Figure 1. The dockerin organization. a) Schematic representation of the dockerin secondary structure. Species-specificity residues positions (AI) are indicated for each segment. **b)** Logo of the 62 dockerins sequences from *Ruminiclostridium cellulolyticum.* Specificity residues are framed in a) and b). Alignments were performed using ClustalOmega.

Figure 2. Chase assay of the preassembled complexes MC1-Cel48F, MC1-Cel9G, MC1-Man5K, and MC1-Cel5A with free Cel5A, Man5K, Cel9G and Cel48F. a) Migration in non-denaturing gel of the complexes and free enzymes. **b)** MC1-Cel48F (in red) challenged with competitors Cel9G (in green), Cel5A (in blue), and Man5K (in purple). **c)** MC1-Cel9G (in green) challenged with competitors Cel48F (in red), Cel5A (in blue), and Man5K (in purple). **d)** MC1-Man5K (in purple) challenged with competitors Cel48F (in red), Cel9G (in green), and Cel5A (in blue). **e)** MC1-Cel5A (in blue) challenged with competitors Cel48F (in red), Cel9G (in green), and Man5K (in purple). The type of complex corresponding to each band, according to the standard gel migration is summarized on the right side of the gels, and the name of the initial complex is framed. Experiments were performed in triplicate.

Figure 3. Chase assay of the preassembled complexes MC1-Rc0286 and MC1-Rc1101 with free Cel5A, Man5K, Cel9G and Cel48F. a) MC1-Rc0286 (in black) challenged with competitors Cel48F (in red), Cel9G (in green), Cel5A (in blue), and Man5K (in purple). **b)** MC1-Rc1101 (in black) challenged with competitors Cel48F (in red), Cel9G (in green), Cel5A (in blue), and Man5K (in purple). Complexes MC1-Cel48F, MC1-Cel9G, MC1-Man5K, and MC1-Cel5A are used as controls. The type of complex corresponding to each band is summarized on the right side of the gels, and the name of the initial complex is framed.Experiments were performed in triplicate.

Figure 4. Chase assay of the preassembled complexes MC1-Cel5A$_{48}$ and MC1-Rc1101$_{48}$ with free Cel5A, Cel5A$_{48}$, Man5K, Cel48F and Cel9G. The chimeric Cel5A and Rc1101 catalytic domains fused to Cel48F dockerin are named Cel5A$_{48}$ and Cel5A$_{48}$, respectively. **a)** MC1-Cel5A$_{48}$ (in blue and red) challenged with competitors Cel48F (in red), Cel9G (in green), and Man5K (in purple). **b)** MC1-Rc1101$_{48}$ (in black and red) challenged with competitors Cel48F (in red), Cel9G (in green), and Man5K (in purple). **c)** MC1-Rc1101 (in black) challenged with competitors Cel5A (in blue) and Cel5A$_{48}$ (in blue and red). The type of complex corresponding to each band is

summarized on the right side of the gels, and the name of the initial complex is framed. Experiments were performed in triplicate.

Figure 5. Chase assay the preassembled complex MC1-Cel9Q with free Cel48F enzyme. MC1-Cel9Q (in orange) challenged with competitor Cel48F (in red). The type of complex corresponding to each band is summarized on the right side of the gel, and the name of the initial complex is framed. Experiments were performed in triplicate.

Figure 6. Chase assay of the preassembled complexes MC1-Cel5A, MC1-Rc1101, and MC1-Cel9Q with free Cel5A, Rc1101, or Cel9Q proteins. a) MC1-Cel5A (in blue) challenged with competitor Rc1101 (in black). **b)** MC1-Cel9Q (in orange) challenged with competitors Cel5A (in blue) and Rc1101 (in black). **c)** MC1-Cel5A (in blue) and MC1-Rc1101 (in black) challenged by competitors Cel9Q (in orange). The type of complex corresponding to each band is summarized on the right side of the gels, and the name of the initial complex is framed. Experiments were performed in triplicate.

Figure 7. Alignment of the sequences of the dockerins. a) Sequences of the dockerins from the proteins displaying a stable interaction with MC1. Characterized proteins are named, and CAZy family was given for not yet characterized protein (CE: Carbohydrate Esterase, GH: Glycoside Hydrolase…). UNK corresponds to a domain of unknown function. Ellipsis indicates that the dockerin is followed by another module at its C-terminus. **b)** Logo of the sequences from enzymes exhibiting a stable interaction with MC1. **c)** Sequences of the dockerins of the three enzymes showing an unstable binding with MC1. Specific residues of the unstable dockerins are labeled in red. Dockerins' residues at positions 22 and 58 are framed in black. Species-specificity residues at positions 11-12 and 47-48 are in black. Alignments were performed using ClustalOmega. NCBI-proteinID: 0729_Cel48F: ACL75108; 0731_Cel9G: ACL75110; 0739_Man5K: ACL75115; 0286_CE4: ACL74673; 0732_Cel9E: ACL75111; 0735_Cel9J: ACL75114; 0734_Cel9H: ACL75113; 0737_Cel9M: ACL75116; 0753_Cel9P: ACL75131; 1648_Cel9R: ACL75999; 1249_Cel9T: ACL75605; 0755_Cel9U: ACL75133; 2392_Cel9V: ACL76725; 2621_Cel9X: ACL76949; 0429_Cel44O: ACL74812; 1230_GH10: ACL75586; 1234_GH62: ACL75590; 1656_UNK: ACL76007; 1101_UNK: ACL75460; 0231_Cel9Q: ACL74618; 1099_Cel5A: ACL75458.

Figure 8. Location of the residues at positions 22 and 58 in the dockerin of Cel5A in complex with the cohesin 1 of CipC. The crystal structures 2VN5 and 2VN6 corresponding to cohesin 1 module from CipC (white) in complex with the dockerin module of $Cel5A_{dockA11S-L12T}$ or

Cel5A$_{dockA47S-L48T}$ (inverted rainbow), respectively, were colored using Pymol software [39]. **a)** Key residues are labelled as follow: species-specificity residues A11 and L12 involved in the cohesin-dockerin recognition on the first dockerin segment (in purple), M22 residue ending the first segment α-helix (in red), species-specificity residues mutated A47S and F48T in the second segment of the dockerin (in blue), and L58 residue ending the second segment α-helix at the interface with the cohesin (dark red). **b)** Key residues are labelled as follow: specificity residues mutated A11S and L12T in the first segment of the dockerin (in orange), M22 residue ending the first segment α-helix at the interface with the cohesin (in red), specificity residues A47 and F48 involved in the cohesin-dockerin recognition in the second segment of the dockerin (in purple), and L58 residue ending the second segment α-helix (dark red). NCBI-proteinID: 1099_Cel5A: ACL75458; 0728_CipC ACL75107.

Figure 9. Chase assay of the preassembled complexes MC1-Cel5A$_{dockM22L}$, MC1-Rc1101$_{dockF22L}$, and MC1-Cel48F$_{dockL22M}$ with free enzymes. a) MC1-Cel5A$_{dockM22L}$ (in blue) challenged with competitor Cel48F (in red). **b)** MC1-Rc1101$_{dockF22L}$ (in black) challenged with competitor Cel48F (in red). **c)** MC1-Cel48F$_{dockL22M}$ (in red) challenged with competitor Cel9G (in green). **d)** MC1-Cel5A (in blue) challenged with competitor Cel48F (in red). **e)** MC1-Rc1101 (in black) challenged with competitor Cel48F (in red). **f)** MC1-Cel48F (in red) challenged with competitor Cel9G (in green). The type of complex corresponding to each band is summarized on the right side of the gels, and the name of the initial complex is framed.

.

Fig 10. Incorporation of free proteins in *R. cellulolyticum* cellulosomes. SDS-PAGE analysis of the cellulosomes **a)** incubated with Rc0286, Man5K, Rc1101 and Rc1101Δdoc and purified by size-exclusion chromatography **b)** incubated with Man5K, purified by size-exclusion chromatography, subsequently incubated with Rc0286 or Rc1101 and purified again by size-exclusion chromatography **c)** incubated with Cel5A and Cel5A$_{48}$ and purified by size-exclusion chromatography. Experiments were performed in duplicate.

Fig 11. Model of the cellulosome assembly and refunctionalization. Enzymes incorporated in the cellulosomes are at least composed of a catalytic domain providing the enzymatic activity and a dockerin module interacting with the scaffoldin cohesin. The segments' α-helices of the dockerin fold into an anti-parallel conformation in which the two α-helices interact differently with the cohesin. One helix presents the specificity AI residues involved in the species recognition, whereas the other helix exposes the stability residue (Leu, Met…). Depending upon

the nature of the residue involved in the stability, enzyme can be replaced by another enzyme displaying a more stable dockerin. Moreover, the catalytic domain has been shown to impact the neighborhood composition during *de novo* assembly of the cellulosome. Thus, both the catalytic domain and the dockerin module of cellulosomal enzymes would display complementary roles in the dynamic assembly or refunctionalization of the cellulosomes.

Legends - Supporting Information

Supplementary Figure 1. Comparison of the 8 cohesins of the scaffoldin CipC from *R. cellulolyticum*. Amino-acid sequences of the 8 cohesins from CipC were blasted individually against *R. cellulolyticum* organism. Sequence identities are summarized in the table and color heatmap was applied by the Excel "Color scale" tool. Color scale used is indicated on the right side of the table. NCBI-proteinID: 0728_CipC ACL75107. Alignments were performed using ClustalOmega.

Supplementary Figure 2. Alignment of the 62 *Ruminiclostridium cellulolyticum* dockerin sequences Amino-acid sequence alignment was done with Clustal Omega software. Characterized proteins are named, or putative functions are indicated (CE, GH, PL…). UNK corresponds to a domain of unknown function. Dockerins are usually present in the C-terminal region of the protein. The ellipsis indicates that the dockerin is followed by another module at the C-terminus. Alignments were performed using ClustalOmega. NCBI-proteinID: 0231_Cel9Q: ACL74618; 0286_CE4: ACL74673; 0379_GH5: ACL74764; 0417_GH11: ACL74801; 0429_Cel44O: ACL74812; 0649_GH5: ACL75030; 0729_Cel48F: ACL ACL75108; 0730_Cel8C: ACL75109; 0731_Cel9G: ACL75110; 0732_Cel9E: ACL75111 0734_Cel9H: ACL75113; 0735_Cel9J: ACL75114; 0736_Man5K: ACL75115; 0737_Cel9M: ACL75116; 0739_Rlg11Y: ACL75119; 0740_Cel5N: ACL75118; 0750_GH11: ACL75128; 0752_Man26A: ACL75130; 0753_Cel9P: ACL75131; 0755_Cel9U: ACL75133; 0840_Cel5D: ACL75216; 0920_CE1: ACL75288; 0931_Xyn10A: ACL75297; 1060_UNK: ACL75420; 1099_Cel5A: ACL75458; 1101_UNK: ACL75460; 1207_GH74: ACL75564; 1229_GH43: ACL75585; 1230_GH10: ACL75586; 1231_GH43: ACL75587; 1232_CE1: ACL75588; 1233_GH43 ACL75589; 1234_GH62: ACL75590; 1235_GH43: ACL75591; 1236_UNK: ACL75592; 1237_GH27: ACL75593; 1238_GH59: ACL75594; 1239_GH2: ACL75595; 1240_GH62: ACL75596; 1241_GH95: ACL75597; 1242_GH5: ACL75598; 1245_PL10: ACL75601; 1246_CE8: ACL75602; 1249_Cel9T: ACL75605; 1298_GH8: ACL75653; 1550_GH18: ACL75902; 1551_GH26: ACL75903; 1597_GH27: ACL75949; 1648_Cel9R: ACL75999; 1655_UNK: ACL76006; 1656_UNK: ACL76007; 1809_chagasine: ACL76159; 2017_UNK: ACL76363; 2123_GH53: ACL76467; 2162_CE2: ACL76504; 2243_PL1: ACL76585; 2337_GH5: ACL76673; 2392_Cel9V: ACL76725; 2395_CE3: ACL76728; 2442_protease: ACL76771; 2530_UNK: ACL76858; 2621_Cel9X: ACL76949.

Supplementary Figure 3.

a) Saturation of the miniscaffoldin MC1 by Cel48F Miniscaffoldin MC1 at 10µM (in black) was mixed with various concentrations of Cel48F ranging from 8µM to 12 µM (in red). Protein mixtures were loaded on a non denaturing PAGE. Cel48F and MC1 alone were used as a control. In the first lane free MC1 is still visible, although an excess of free Cel48F is observed for other complexes and free MC1 is no longer detected. So the suitable quantity of Cel48F was set at 9.3 µM for a concentration of MC1 at 10µM. Similarly, suitable quantities of Cel5A, Cel9G, and Man5K were determined and combined with MC1 scaffoldin to assemble MC1-Cel9G, MC1-Cel5A, and MC1-Man5K, respectively. **b) Kinetic of the chase assay for the complex MC1-Cel5A** MC1-Cel5A (in blue) challenged by competitor Cel48F (in red) for 5 minutes, 1 or 2 hours before loading on a non denaturing gel. **c) Kinetic of the chase assay for the complex MC1-Cel9G** MC1-Cel9G (in green) challenged by competitor Cel48F (in red for 5 minutes, 1 or 2 hours before loading on a non denaturing gel. The type of complex corresponding to each band is summarized on the right side of the gels, according to the standard gel migration (Suppl. 1), and the name of the initial complex is framed. **d) Effect on an excess of competitor** 10µM of MC1-Cel48F (in red) are challenged for 5 minutes with the competitor Man5K (in purple) at a concentration ranging from 10µM (1X) to 90µM (9X) before loading on a non denaturing gel. The type of complex corresponding to each band is summarized on the right side of the gels, according to the standard gel migration (Suppl. 1), as well as the free (released) Cel48F or free competitor Man5K, and the name of the initial complex is framed. Experiments were performed in triplicate.

Supplementary Table 1. Primers. Mutated codons are indicated in bold and italic. Cloning sites are underlined. His-Tag sequence is indicated in italics. Homology sequences for the fusion PCR are underlined and italics.

Supplementary Table 2. Plasmids.

Acknowledgments

The authors are indebted to Pascale de Philip, Stéphanie Perret, Nathalie Franche for helpful discussions.

This study was supported by grants from the Innovative Training Networks Marie Curie CLOSTNET grant N°237942, the Centre National de la Recherche Scientifique and Aix-Marseille Université

Author contributions

RB, MHDT, SP and CT made the genetic constructions, production of recombinant proteins in *E. coli*, purification of the recombinant proteins and native gel experiments.
MHDT made the cellulosomes extraction and gel filtration. RB wrote the manuscript. The manuscript was read and corrected by CT, NV, HPF and SP.

Funding sources

This research was supported by the Innovative Training Networks Marie Curie CLOSTNET grant N°237942, the Centre National de la Recherche Scientifique and Aix-Marseille Université.

Conflicts and Interest

The authors declare no conflict of interest.