



Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*

Benedita A. Pinheiro, Harry J. Gilbert, Kazutaka Sakka, Kazuo Sakka, Vânia O. Fernandes, José A.M. Prates, Victor D. Alves, David N. Bolam, Luís M.A. Ferreira, Carlos M.G.A. Fontes

► To cite this version:

Benedita A. Pinheiro, Harry J. Gilbert, Kazutaka Sakka, Kazuo Sakka, Vânia O. Fernandes, et al.. Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochemical Journal*, 2009, 424 (3), pp.375-384. <10.1042/BJ20091152>. <hal-00479238>

HAL Id: hal-00479238

<https://hal.science/hal-00479238v1>

Submitted on 30 Apr 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*

Benedita A. Pinheiro^{*}, Harry J. Gilbert^{†,1}, Kazutaka Sakka[‡], Kazuo Sakka[‡], Vânia O. Fernandes^{*}, José A.M. Prates^{*}, Victor D. Alves^{*}, David N. Bolam[†], Luís M.A. Ferreira^{*} and Carlos M.G.A. Fontes^{*,2}

Short title: Novel type I cohesin-dockerin specificities in *Clostridium thermocellum*

Key words: Cellulosome, protein:protein interactions, cell-surface attachment, cohesin, dockerin, molecular scaffoldin.

^{*} CIISA – Faculdade de Medicina Veterinária, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; [†] Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom; [‡] Applied Microbiology Laboratory, Graduate School of Bioresources, Mie University, Tsu, Japan

¹ Current address: Complex Carbohydrate Research Centre, University of Georgia, Athens, Georgia 30602-4712, USA.

² Corresponding author. Phone: (351) 213652876; Fax: (351) 213652889; E-mail: cafontes@fmv.utl.pt

SUMMARY

Cellulosomes, synthesized by anaerobic microorganisms such as *Clostridium thermocellum*, are remarkably complex nanomachines that efficiently degrade plant cell wall polysaccharides. Cellulosome assembly results from the interaction of type I dockerin domains, present on the catalytic subunits, and the cohesin domains of a large non-catalytic integrating protein that acts as a molecular scaffold. In general, type I dockerins contain two distinct cohesin binding interfaces that appear to display identical ligand specificities. Inspection of the *Clostridium thermocellum* genome reveals 72 dockerin-containing proteins. In four of these proteins, Cthe_0258, Cthe_0435, Cthe_0624 and Cthe_0918, there are significant differences in the residues that comprise the two cohesin-binding sites of the type I dockerin domains. In addition, a protein of unknown function (Cthe_0452), containing a C-terminal cohesin highly similar to the equivalent domains present in *C. thermocellum* integrating protein (CipA), was also identified. Here, the ligand specificities of the newly identified cohesin and dockerin domains are described. The results revealed that Cthe_0452 is located at the *C. thermocellum* cell surface and thus the protein was renamed as OlpC. The dockerins of Cthe_0258 and Cthe_0435 recognize, preferentially, the OlpC cohesin and thus these enzymes are believed to be predominantly located at the surface of the bacterium. By contrast, the dockerin domains of Cthe_0624 and Cthe_0918 are primarily cellulosomal since they bind preferentially to the cohesins of CipA. OlpC, which is a relatively abundant protein, may also adopt a “warehouse” function by transiently retaining cellulosomal enzymes at the cell surface before they are assembled onto the multi-enzyme complex.

INTRODUCTION

The plant cell wall represents the most abundant source of terrestrial organic carbon. The degradation of this composite structure by microbial enzymes is central to the carbon cycle. There is currently considerable interest in the application of these biocatalysts in the conversion of lignocellulosic biomass into biofuels such as ethanol and butanol. The complex chemical and physical structure of plant cell walls restricts enzyme access to the interlocking polysaccharides, primarily cellulose and hemicellulose, thus limiting carbon turnover. A common feature of anaerobic plant cell wall degrading microorganisms is that extracellular cellulases and hemicellulases are organized into high molecular weight multi-enzyme complexes, which are referred to as cellulosomes [1-3]. Cellulosomes enhance the hydrolytic activity of the assembled enzymes by i) potentiating the synergistic interactions between the biocatalysts and, ii) the presence of a central cellulose-specific carbohydrate binding module, which brings the catalytic subunits into intimate contact with their target substrates, thereby reducing the enzyme accessibility problem [4, 5].

Clostridium thermocellum displays one of the fastest known growth rates on cellulose and its cellulosome is the paradigm for plant cell wall degrading enzyme complexes. The protein that mediates the assembly of the *C. thermocellum* cellulosome is the scaffoldin subunit termed CipA – a 1853 amino acid non-catalytic polypeptide that contains nine highly conserved modules, known as type I cohesins, a family 3 carbohydrate-binding module (CBM) that attaches the cellulosome onto crystalline cellulose, and a type II dockerin [6]. Type I dockerins, located in cellulosomal enzymes, primarily glycoside hydrolases, but also carbohydrate esterases and polysaccharide lyases, bind extremely tightly to CipA cohesins thus anchoring the enzymes onto the macromolecular scaffold [1, 2]. The C-terminal type II dockerin of CipA, which recognizes, specifically, type II cohesins located in proteins bound to the bacterial peptidoglycan layer, maintains the cellulosome on the bacterial cell surface [7]. There is no cross specificity between type I and type II cohesin-dockerin partners. Significantly, a type I cohesin domain was also identified in a cell bound protein, OlpA, suggesting that cellulosomal enzymes can also adhere directly, and individually, onto the bacterial surface [8].

Crystal structures of type I cohesin-dockerin complexes have provided important insights into the mechanism of cellulosome assembly. Dockerins contain a tandem duplication of a 22-residue sequence that are organized into two F-hand calcium binding motifs and an α -helix that also display remarkable structural conservation – the N and C-terminal α -helices overlay with minor structural deviations [9-11]. Internal two-fold symmetry of the dockerin molecule has a profound influence on the mechanism of cohesin recognition. Structural and mutagenesis data revealed that type I dockerins contain two, essentially identical, cohesin-binding interfaces. Residues participating in cohesin recognition at the two binding interfaces, particularly a Ser-Thr motif at positions 11 and 12 and a Lys-Arg motif at positions 18 and 19, are highly conserved in the two segments of the majority of *C. thermocellum* dockerins (see below), suggesting that, in general, both binding interfaces display similar protein specificities [10, 12]. The dockerin dual binding mode may reduce the steric constraints that are likely to be imposed by assembling a large number of different catalytic modules into a single cellulosome. In addition, the switching of the binding mode between two conformations may also introduce quaternary flexibility into multi-enzyme complexes thus enhancing substrate targeting and the synergistic

interactions between some enzymes, particularly exo- and endo-acting cellulases [10, 11]. Similar structural observations were made for *C. cellulolyticum* dockerins, within cohesin-dockerin complexes, suggesting an evolutionary pressure for the retention of a dual binding mode in, at least, *Clostridial* dockerins [13]. Despite the structural similarity between the *C. thermocellum* and *C. cellulolyticum* cohesins and dockerins, there is no cross specificity between the protein partners of the two organisms [14]. Residues at dockerin positions 11 and 12 were shown to play a critical role in organism-specific recognition [15, 16], with *C. thermocellum* dockerins presenting, typically, a pair of hydroxy residues, while the corresponding amino acids in *C. cellulolyticum* dockerins are alanine and a second, larger, hydrophobic amino acid.

The genome of *C. thermocellum* ATCC 27405 encodes 72 polypeptides containing type I dockerin sequences. Inspection of dockerin sequences at the two ligand binding sites revealed a strong conservation of the amino acids that mediate cohesin recognition. There are, however, at least four dockerins that are components of the proteins Cthe_0435, Cthe_0918, Cthe_0258 and Cel9D-Cel44A (accession number Cthe_0624), which deviate from the canonical *C. thermocellum* motifs in one of the ligand binding interfaces. In these dockerins the usually conserved Ser-Thr pair, which dominates the hydrogen bond network with the cohesin, is replaced in one of the duplicated segments by non-hydroxy residues. The implications of these amino acid substitutions for cohesin recognition remain to be investigated. Here we have established the ligand specificity of *C. thermocellum* dockerins which display divergent cohesin binding interfaces.

EXPERIMENTAL

Cloning, expression and purification

Genes encoding dockerin domains were amplified from *C. thermocellum* genomic DNA using the thermostable DNA polymerase NZYPremium (NZYTech Ltd). Amplified DNA was directly cloned into pNZY28 (NZYTech Ltd) and sequenced to ensure that no mutations were accumulated during the amplification. Genes encoding the dockerin domains of Cthe_0435, termed Doc-435 (residues 32-112), Cthe_0918, termed Doc-918 (residues 1146-1209) and Cthe_0258, termed Doc-258 (residues 33-105), were subcloned into pET32a (Novagen) restricted with *Eco*R1 and *Xho*I (see Table 1). Recombinant Doc-435, Doc-918 and Doc-258 were expressed in fusion with thioredoxin, to improve dockerin solubility and stability. Similarly, to improve protein stability, the dockerin domain of Xyn10B was produced in fusion with the N-terminal CBM22 xylan-binding domain and the Cel9D-Cel44A dockerin was expressed with the GH44 catalytic domain at the N-terminus and the polycystic kidney-disease like module (PKD) followed by the family 44 carbohydrate-binding module (CBM44) at the C-terminus. The genes encoding the Xyn10B derivative, termed Doc-Xyn10B (residues 558-799, Cthe_0912) and Cel9D-Cel44A derivative, termed Doc-Cel44A (residues 774-1601, Cthe_0624), were subcloned into pET21a using primer pairs containing *Nhe*I and *Xho*I restriction sites, respectively. The recombinant plasmid encoding Doc-Cel44A was termed pD44-21a. Genes encoding the second cohesin of CipA, termed Coh-CipA2 (residues 182-328, Cthe_3077), the type I cohesin of OlpA, termed Coh-OlpA (residues 30-175, Cthe_3080), and type I cohesin of OlpC, termed Coh-OlpC (residues 100-258, Cthe_0452) were amplified as described above, cloned into pNZY28 (NZYTech Ltd.) and sequenced. Coh-OlpA and Coh-OlpC genes were subcloned into *Bgl*III/*Eco*RI and *Bam*HI/*Eco*RI digested pRSETa, respectively

(Invitrogen). The gene encoding Coh-CipA2 was sub-cloned into pET21a (Novagen). The DNA sequences encoding the N-terminal domain of Cthe_0452 (residues 25-112), termed CtUnk1, and the type II dockerin of CipA fused with the endogenous N-terminal X module (residues 1691-1853, Cthe_3077) were amplified using the procedures described above and the primers presented in Table 1. Both genes were subsequently subcloned into pET32a (Novagen). *Clostridium cellulolyticum* type I cohesin was produced using plasmid pHF1, as described by Pinheiro *et al.* [13]. All recombinant proteins contained an internal or a C-terminal His₆-tag.

E. coli Origami, BL21 and Tuner cells, transformed with pET32a, pET21a and pRSETa derivatives, respectively, were grown at 37°C to mid-exponential phase (OD₆₀₀=0.6). Recombinant protein expression was induced by adding 0.2 (Tuner) or 1 mM (all other *E. coli* strains) isopropyl-β-D-thiogalactopyranoside and incubation for a further 16 h at 19°C. Soluble recombinant proteins were purified by immobilized metal ion affinity chromatography as described previously [17, 18]. Fractions containing the purified proteins were buffer exchanged, in PD-10 Sephadex G-25M gel filtration columns (GE Healthcare), into 50 mM Hepes, pH 7.5, containing 100 mM NaCl and 5 mM CaCl₂. For isothermal calorimetry, a further purification step by gel exclusion chromatography was performed and the proteins were maintained in the same buffer excluding the NaCl. SDS/PAGE showed that all the recombinant proteins were more than 99 % pure.

Mutagenesis

Site-directed mutagenesis was carried out using the PCR-based NZYMutagenesis site-directed mutagenesis kit (NZYTech Ltd) according to the manufacturer's instructions, using plasmid pD44-21a as template. The sequences of the primers used to generate these mutants are displayed in Table 1. The mutated DNA sequences were sequenced to ensure that only the appropriate mutations had been incorporated into the nucleic acid.

Interaction of CtUnk1 with bacterial cell wall preparations

Native peptidoglycan-containing sacculi (NPCS) and hydrofluoric acid-extracted cell-wall polymer (HF-EPCS) were prepared from *E. coli* and *C. thermocellum* cells as described previously [19]. Recombinant CtUnk1 (10 µg) was separately incubated with NPCS and HF-EPCS (80 µg), in 50 µl of 20 mM sodium phosphate buffer (pH 7.4), at 4°C for 1 h with occasional shaking. Following incubation, the insoluble fractions were pelleted by centrifugation and the supernatant and pellet fractions were analyzed by SDS-PAGE.

Complex Formation in Solution

As an initial approach to test cohesin-dockerin affinity, the two proteins were combined in 50 mM Hepes buffer (pH 7.5), containing 100 mM NaCl and 5 mM CaCl₂ for 1 hour at room temperature and the presence of complexes was evaluated by non-denaturing polyacrylamide gel electrophoresis. For lower affinity interactions, complexes were detected by increasing concentrations of a dockerin against a fixed concentration of cohesin. New bands appearing in the native gel were used as an indication of complex formation. The differences in the affinities of the complexes were assessed by combining, in the same solution, equimolar concentrations of two dockerins and one cohesin (or two cohesins and one dockerin) and analyzing the gel for complex formation. To determine the stability of cohesin-dockerin complexes, a

cohesin-dockerin pair was assembled for an hour. After this incubation period a second cohesin or dockerin was added to the mixture and incubated for another hour. Complex stability was evaluated in non-denaturing gels, as explained above.

Isothermal Calorimetry of Coh-Doc Binding

Isothermal titration calorimetry was carried out essentially as described previously [10, 13], except that measurements were made at 55°C, and proteins were dialyzed into 50 mM NaHepes, pH 7.5, containing 2 mM CaCl₂. During titration, the dockerin (15 µM) was stirred at 300 rpm in the reaction cell, which was injected with 29 or 58 successive 5-µl aliquots of ligand comprising cohesin (180 µM) at 200-s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression by using a single site-binding model (Microcal ORIGIN, Ver. 5.0, Microcal Software, Northampton, MA). The fitted data yield the association constant (K_A) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated by using the standard thermodynamic equation: $\Delta RT \ln K_A = -\Delta G = \Delta H - T\Delta S$.

RESULTS AND DISCUSSION

Novel type I cohesin and dockerin domains in *C. thermocellum* proteins

The *C. thermocellum* proteome was searched using Xyn10B [20] and Cel9D-Cel44A [21-23] dockerin sequences through BLAST (<http://blast.ncbi.nlm.nih.gov>) to identify the complete repertoire of cellulosomal proteins. The data revealed that 72 *C. thermocellum* proteins contain type I dockerins. Alignment of all the identified dockerins revealed that at least four of these domains display a lack of conservation in the residues that interact with the type I cohesin partners present in CipA (data not shown). These domains belong to proteins Cthe_0258, Cthe_0435, Cthe_0624 and Cthe_0918. Cthe_0624 is the bifunctional cellulase Cel9D-Cel44A (initially referred to as CelJ [21]) where the dockerin is located internally between the two catalytic modules (GH9 and GH44) and a C-terminal CBM44 [23]. The other three proteins do not have an assigned function and the dockerin is positioned either at the N- (Cthe_0258 and Cthe_0435) or C-terminus (Cthe_0918) (Figure 1). Modules of unknown function in Cthe_0258, Cthe_0435 and Cthe_0918 have, respectively, 365, 218 and 1136 residues and may comprise catalytic domains, or possibly CBMs, which contribute to the deconstruction of the plant cell wall. Indeed, the 218 residue module in Cthe_0435 was recently shown to display cellulase activity against both crystalline and amorphous forms of the polysaccharide (H.J. Gilbert and C.M.G.A. Fontes, unpublished data). Alignment of the four divergent dockerins with the dockerin domain of Xyn10B (Doc-Xyn10B) revealed striking variations in positions 11 and 12 of one of the duplicated segments (Figure 2). The hydroxyl residues at positions 11 and 12 play an important role in the hydrogen bond network established with the cohesin [11]. Doc-Cel44A, Doc-258 and Doc-918, derived from Cel9D-Cel44A, Cthe_0258 and Cthe_0918, respectively, lack the serine-threonine pair in the first duplicated segment; these residues are substituted by an alanine and a hydrophobic residue (Doc-Cel44A and Doc-258) or by two acidic amino acids (aspartate and glutamate; Doc-918). At the second duplicated segment, only Doc-258 and Doc-435 do not present the conserved hydroxy-residues at position 11 and 12; Doc-258 contains a serine-isoleucine pair and Doc-435 an aspartate-isoleucine pair at these positions. However, in all dockerins, a high degree of conservation is evident in the

other residues that directly participate in cohesin recognition at both putative binding interfaces.

A similar bioinformatic strategy was employed to identify the complete array of type I cohesin domains encoded by the *C. thermocellum* genome. Three proteins were shown to contain type I cohesins: the scaffolding protein CipA (Cthe_3077), which contains nine cohesin modules, OlpA, with one cohesin module (Cthe_3080) and a previously unknown protein, Cthe_0452, which also contains a single type I cohesin domain (Figure 1). OlpA is a cell surface bi-modular protein presenting an N-terminal type I cohesin domain fused to three SLH (from S-layer homology) repeats [8]. The SLH repeats were previously shown to interact with the bacterial surface [8, 19, 24] and, therefore, cellulosomal enzymes may also be directed to the bacterial surface through the binding to the exposed type I cohesin domain of OlpA. In contrast, Cthe_0452 contains a C-terminal type I cohesin domain and an N-terminal module of unknown function. Alignment of the eleven *C. thermocellum* type I cohesin domains (Figure 1s, Supplementary Material) reveals that the dockerin-interacting residues of all nine CipA cohesins are highly conserved, suggesting that CipA cohesins cannot discriminate between dockerin modules, as indicated previously [10, 11]. In contrast, both the OlpA and Cthe_0452 cohesins reveal several substitutions in key residues involved in dockerin recognition. For example, Asn37, which is an important residue in the hydrogen bond network established with dockerins, is replaced by a serine in both proteins (this substitution also occurs in the ninth cohesin of CipA). In addition, the other two residues that make polar interactions with the dockerin, Asp39 and Glu131, are also replaced by an Asn and a Pro, respectively, in Cthe_0452 cohesin. The significance of these amino acid substitutions in cohesin-ligand specificity is explored below.

***C. thermocellum* Cthe_0452 is a cell-surface protein**

Cthe_0452 was shown to contain, in addition to a C-terminal type I cohesin, an 88 residue N-terminal domain of unknown function, designated CtUnk1. To determine the role of CtUnk1, in carbohydrate metabolism or cellulosomal function, the module was purified to electrophoretic homogeneity and its biochemical properties evaluated. Initially we explored the hypothesis that CtUnk1 constituted a CBM. However, affinity gel electrophoresis analysis revealed that CtUnk1 displays no significant affinity for xyloglucan, the β 1,4- β 1,3-mixed glucans barley β -glucan and lichenan, the β 1,4-glucan hydroxyethylcellulose (HEC), konjac glucomannan, oat spelt xylan, the β 1,3-glucans laminarin and curdlan, carob galactomannan, potato galactan, pullulan or pustulan (data not shown). In addition, the domain was unable to depolymerase these polysaccharides indicating that Cthe_0452 is not a plant cell wall degrading enzyme (data not shown).

Similar to OlpA, it is possible that Cthe_0452 constitutes a cell surface protein based on the assumption that CtUnk1 binds to the bacterial cell envelope. To explore this hypothesis the capacity of CtUnk1 to interact with *C. thermocellum* cell-wall fractions was investigated. Proteins located at the outermost cell envelop of Gram-positive bacteria, known as the bacterial S-layer, bind secondary cell wall polymers (SCWP) rather than the peptidoglycan cell wall matrix. The data from pull-down experiments, displayed in Figure 3, revealed that CtUnk1 bound to the insoluble *C. thermocellum* peptidoglycan-containing sacculi (NPCS); only a small amount of the protein was detected in the unbound soluble fraction. In contrast, CtUnk1, was unable to interact

with *C. thermocellum* hydrofluoric acid-extracted cell wall polymer (HF-EPCS). Hydrofluoric extraction removes SCWP from NPCS [19, 24] and, therefore, HF-EPCS consists mainly of peptidoglycan. Therefore, the data suggest that CtUnk1 displays affinity for *C. thermocellum* SCWP. To further confirm that CtUnk1 was unable to interact with the peptidoglycan fraction of the cell wall, an HF-EPCS fraction was prepared from *Escherichia coli*, which is free of Gram-positive SCWP. The data (Figure 3) revealed that CtUnk1 was unable to interact with the *E. coli* HF-EPCS preparation, strongly suggesting that CtUnk1 specifically interacts with SCWP present in *C. thermocellum*, or at least in Gram-positive bacterial cell walls. Since it appears that CtUnk1 directs Cthe_0452 onto the bacterial envelope we propose defining Cthe_0452 as OlpC (Outer Layer Protein analogous to OlpA). Interestingly, CtUnk1 displays a specificity that is similar to the OlpA SLH domain, and other SLH domains, in recognizing SCWP [19, 24]. In contrast, the SLH domains of SdbA, OlpB and Orf2, unusually for such modules, bind to the peptidoglycan layer [19, 24]. Therefore, the data suggest that proteins that are responsible for binding cellulosomes, such as SdbA, OlpB and Orf2, are bound to the peptidoglycan layer while OlpA and OlpC, which bind individual cellulosomal components, interact with SCWP. The biological significance of these identified differences in cell wall specificity remain, however, unclear. In addition, these data demonstrate that cellulosomal enzymes can either be bound to cellulosomes (through the binding of CipA cohesins) or interact indirectly with the cell surface of the bacterium, by binding to OlpA or OlpC. However, it is presently unknown if differences in cohesin-dockerin affinity can modulate the distribution of *C. thermocellum* enzymes over the cell surface or onto the cellulosomes.

Ligand specificity of novel *C. thermocellum* cohesin and dockerins

As described above, a set of cohesins and dockerins, whose sequences diverge from the consensus cellulosomal enzymes, were identified in *C. thermocellum*. The impact of these primary sequence differences on cohesin-dockerin specificity was analyzed by ITC. Titration experiments, exemplified in Figure 4, were performed at 55 °C. The data, presented in Table 2, revealed a range of affinities with K_d s varying from 10^5 to $>10^9$ M⁻¹ (Figure 4). In general, binding was enthalpically driven with the change in entropy slightly decreasing affinity, which is consistent with the thermodynamic interactions reported for most type I cohesin-dockerin complexes [10, 11, 13]. Thus, the data suggest that all the atypical *C. thermocellum* dockerins, Doc-435, Doc-918 and Doc-258, bound to the *C. thermocellum* cohesins. This was further confirmed, for most of the cohesin-dockerin pairs, by native gel electrophoresis as exemplified in Figure 5a; cohesin and dockerins were mixed together and complex formation was revealed by the appearance of an extra band in the gel.

The interaction between Doc-Cel44A and the *C. thermocellum* cohesins in CipA (Coh-CipA2) and OlpA is very tight, with an affinity constant probably higher than 10^9 M⁻¹, which is the limit for quantifying molecular interactions by ITC. In contrast DocCel44A displays lower affinity for Coh-OlpC. A similar pattern is observed for Doc-918 which binds to the CipA cohesin ~60-fold more tightly than Coh-OlpC, although the K_a for Coh-OlpA is also significantly lower (~20-fold) than for Coh-CipA2. By contrast the cellulosomal dockerin Doc-Xyn10B does not appear to distinguish between the three *C. thermocellum* cohesins. Dockerin primary sequences suggest that Doc-Cel44A and Doc-918 will recognize cohesins preferentially through the C-terminal binding site since there is significant deviation from the canonical Ser-

Thr or Ser-Ser motifs at positions 11 and 12 at the N-terminal ligand interface. Indeed the preference of these two dockerins for the CipA cohesin is consistent with the presence of two hydroxy residues at the critical 11 and 12 positions of the putative C-terminal ligand binding site

When compared with the cellulosomal dockerin Doc-Xyn10B, Doc-258 displays lower affinities for the three type I *C. thermocellum* cohesins. While Doc-435 also displayed lower affinity than the xylanase dockerin for Coh-CipA2 and Coh-OlpA, its affinity for Coh-OlpC was comparable to Doc-Xyn10B. The primary sequence of Doc-258 suggests that the module will recognize cohesins preferentially through the C-terminal binding. In contrast, Doc-435 is likely to interact with type I cohesins through the N-terminal protein binding site. Interestingly, contrasting with Doc-Cel44A and Doc-918 (which present two hydroxy residues at the critical 11 and 12 positions of the putative C-terminal binding site; see above) and Doc-Xyn10B, dockerins Doc-258 and Doc-435 contain an isoleucine and a hydroxyl residue in these positions of their predicted binding sites. Thus, the presence of a hydrophobic residue at positions 11 or 12 may account for the lower affinities that Doc-258 and Doc-435 dockerins exhibit for Coh-CipA2. Dockerin-cohesin selectivity will be further explored below.

The binding interfaces of Doc-258 and Doc-Cel44A that lack a hydroxyl residue and are thus predicted not to bind type I CipA cohesins, resembles the ligand binding interfaces of *C. cellulolyticum* type I dockerin (Figure 2). Work by Bayer, Fierobe, Lamed and colleagues revealed that a single Thr to Leu substitution at position 12 of a *C. thermocellum* dockerin allows the protein to bind *C. cellulolyticum* cohesins [15]. Therefore, the *C. thermocellum* dockerins were also probed against a type I cohesin from *C. cellulolyticum*, termed Coh-Ccel. The data (Table 2), revealed that Doc-Cel44A presented a high affinity for the *C. cellulolyticum* cohesin Coh-Ccel (K_a 10^8 M⁻¹), while all other *C. thermocellum* dockerins bound much more weakly (100-1000 lower affinities) to the non-cognate cohesin. Overall the data suggest that apart from Doc-Cel44A, *C. thermocellum* dockerins displayed species restricted ligand specificity. However, there are no obvious amino acid substitutions in the Doc-258 primary sequence, when compared with Doc-Cel44A (Figure 2), which can explain its inability to recognize the *C. cellulolyticum* cohesin.

Complex stability and cohesin-dockerin selectivity

In order to gain further insights into the mechanism of cohesin-dockerin recognition, the capacity of individual cohesins or dockerins to destabilize previously assembled complexes was evaluated. Thus, individual cohesin or dockerin domains were added to an assembled complex and the stability of the original assembly, or the formation of a new complex, was evaluated by native gel electrophoresis (examples in Figure 5b and 5c). If a previously formed complex was reconfigured, a new band would appear on the gel. Since detection of several complexes was not straightforward through gel electrophoresis, due to low cohesin-dockerin affinity, some protein domains were excluded from the analysis. The data, displayed in Table 3, revealed that cohesin exchange occurred readily, although a dockerin bound to its cohesin partner was more rarely displaced by the addition of an unliganded dockerin. For example, addition of Coh-OlpA to a Coh-CipA2/Doc-Xyn10B complex led to the formation of a Coh-OlpA/Doc-Xyn10B complex, while no dockerin was able to destabilize the Coh-CipA2/Doc-Xyn10B or the Coh-CipA2/Doc-Cel44A complexes. The existence of two

cohesin binding interfaces in dockerins may explain this observation as they may allow these domains to switch more easily from one cohesin to another and yet maintain a high affinity protein:protein interaction. Significantly, dockerin switching seems to occur only towards higher affinity cohesins. Since cohesins contain only one binding interface, their switching from one dockerin to another is more difficult considering the high affinity interaction established in cohesin-dockerin complexes. Therefore, taken together the data suggest that free dockerins cannot easily displace already bound dockerins. However, bound dockerins can easily bind to an unliganded cohesin if this is energetically favored. Thus, dockerin cohesin switching may reflect the existence of two cohesin-binding interfaces in dockerins. Significantly, complexes containing Doc-258 were dissociated in the presence of Doc-Xyn10B or Doc-918, which likely reflects their substantially higher affinity for cohesins, compared to Doc-258.

In addition, dockerin and cohesin preference for protein partners was analyzed by mixing one dockerin with two cohesins (or one cohesin with two dockerins when cohesin selectivity was evaluated) and assessing complex formation by native gel electrophoresis (see Figure 5d for an example). Each cohesin and dockerin was mixed with the combination of all possible protein partners. The resulting data were integrated and allowed, for each cohesin or dockerin domain, ordering their protein partners by binding preference. The data, presented in Table 4, revealed that subtle differences in the affinity of the various cohesin-dockerin pairs, as evaluated by ITC, can have a significant effect in cohesin-dockerin preference. For example, dockerins Doc_Xyn10B and Doc_Cel44A, which were shown to bind cohesins Coh-CipA2 and Coh-OlpA with similar thermodynamic parameters and affinities, bind preferentially Coh-OlpA rather than Coh-CipA2. In addition, the two dockerins displayed a lower preference for Coh-OlpC as a protein partner. When compared with Doc_Xyn10B and Doc_Cel44A, Dockerin Doc-918 displayed a similar cohesin selectivity, although it did not display a preference for either Coh-CipA2 or Coh-OlpA. In contrast to the cohesin preferences expressed by dockerins Doc-918, Doc_Xyn10B and Doc_Cel44A, dockerins Doc-435 and Doc-258 bound, preferentially, Coh-OlpC. In addition, these dockerins revealed the lowest affinities for cohesin Coh-OlpA. Thus, the data suggest that substitution of one of the hydroxyl residues of the conserved Ser-Thr pair of the putative cohesin binding interface with an isoleucine (Figure 2) leads to a switch in cohesin preference in dockerins Doc-435 and Doc-258 from cohesins Coh-CipA2 or Coh-OlpA to cohesin Coh-OlpC [14]. In the cohesin side, Coh-CipA2 and Coh-OlpA displayed identical dockerin preferences and bound, with similar affinities, dockerins Doc_Cel44A, Doc_Xyn10B and Doc-918. These two cohesins display a lower selectivity for Doc-258 and particularly for Doc-435. Thus, amino acid substitutions in Coh-OlpA, and most probably in the ninth cohesin of CipA (see above), do not seem to affect the capacity of OlpA cohesin to interact with cellulosomal dockerin partners such as Doc_Xyn10B. In contrast, cohesin Coh-OlpC interacts, preferentially, with dockerin Doc-435, suggesting that Coh-OlpC amino acid substitutions at the dockerin binding platform (such as D39N and E131P) are particularly suited for the recognition of dockerin Doc-435. Surprisingly, Coh-OlpC still recognizes dockerins Doc_Cel44A and Doc-Xyn10B with considerable affinity.

Taken together the data suggest that most cellulosomal enzymes, which contain dockerins similar to Doc_Xyn10B, will bind preferentially to the type I cohesin of the cell surface protein OlpA rather than to CipA cohesins. However, Coh-OlpA is

relatively scarce on the *C. thermocellum* cell surface [25] and, therefore, cellulosomal enzymes are preferentially located in cellulosomes. In addition, dockerin Doc_Xyn10B will bind less tightly to Coh-OlpC. In contrast to OlpA, OlpC is relatively abundant on the bacterium's surface [25]. Thus, this protein may transiently maintain cellulosomal enzymes, which contain Doc_Xyn10B like dockerins, at the bacterial surface before these components are recruited for cellulosome assembly. A clearly different binding preference is expressed by dockerins Doc-258 and Doc-435, which seem to favor the binding to the type I cohesin of OlpC. Thus, it is suggested that these dockerins target their appended modules to the bacterium cell surface rather than to cellulosomes.

Doc-Cel44A contains two cohesin-binding interfaces presenting different ligand specificities

The dockerin Doc-Cel44A was shown to bind both *C. thermocellum* and *C. cellulolyticum* type I cohesins (Table 2). The Doc-Cel44A primary sequence (Figure 2) suggests that the dockerin N-terminal binding interface will recognize *C. cellulolyticum* type I cohesins. In contrast, the Doc-Cel44A C-terminal ligand binding site is typical of *C. thermocellum* type I dockerins, and thus is unlikely to interact with *C. cellulolyticum* cohesins. To investigate the ligand specificity of the two Doc-Cel44A binding interfaces, residues at positions 11 and 12 of the two binding sites (Ala-Val and Ser-Ser, respectively) were mutated to two glutamines, separately or together, in order to inactivate the binding site [13]. The affinities of the variants generated for the *C. cellulolyticum* and *C. thermocellum* cohesins were determined by ITC. The data, presented in Table 5, suggest that mutation of the Ala-Val motif at the N-terminal binding site of Doc-Cel44A, decreases affinity for the *C. cellulolyticum* cohesin. A slightly larger reduction in affinity for the *C. cellulolyticum* cohesin was observed when residues at both binding interfaces were altered. In contrast, modification of only the C-terminal ligand binding site of the dockerin did not affect recognition of the *C. cellulolyticum* cohesin. By contrast, mutation of the N-terminal Ala-Val motif had little influence on affinity for the *C. thermocellum* cohesin, while replacing the C-terminal Ser-Ser sequence with Gln-Gln resulted in a 30-fold reduction in the K_a . Therefore, taken together, the data suggest that Doc-Cel44A displays affinity for *C. cellulolyticum* cohesins predominantly through its N-terminal binding interface, while the C-terminal ligand binding site primarily recognizes *C. thermocellum* cohesins.

Data described above suggest that the dual specificity revealed by Doc-Cel44A may allow the formation of a tri-modular complex through the simultaneous binding of the dockerin to both a *C. cellulolyticum* cohesin, at the N-terminal, and a *C. thermocellum* cohesin, at the C-terminal binding faces. To probe this possibility, the three domains were mixed in solution and the formation of a tri-modular complex was investigated through non-denaturing electrophoresis. The data, presented in Figure 6, suggests that Doc-Cel44A can only bind to a single cohesin, and displays a preference for the *C. thermocellum* protein partner, consistent with the ITC data. This observation is also in agreement with studies by Carvalho and colleagues demonstrating that steric clashes prevents the binding of one dockerin to two cohesins [10]. The biological rationale for the recognition of *C. cellulolyticum* cohesins by a *C. thermocellum* dockerin remains unknown. This observation may reflect a fortuitous interaction of an yet unknown function.

Conclusions

The genome sequence of *C. thermocellum* allowed the identification of the complete set of cohesins and dockerins produced by this bacterium. Data presented here revealed that type I cohesins are located in the scaffoldin protein CipA and in the cell-surface proteins OlpA and OlpC. Therefore, enzymes containing type I dockerins may be targeted to cellulosomes or directly to the *C. thermocellum* cell envelop. However, Xyn10B like dockerins, which are the most common in *C. thermocellum*, seem to display a much higher affinity for CipA cohesins than to OlpC, the dominant type I cohesin-containing cell surface protein [25]. These data suggest that cellulosomal enzymes may transiently interact with the bacterium's surface, through the binding to OlpC, before they are assembled into the multi-enzyme complexes. Similar to the Xyn10B dockerin, the majority of *C. thermocellum* dockerins present two highly conserved cohesin binding sites, which are likely to display identical ligand affinities [10]. The dockerin dual binding mode may facilitate dockerin switching, through the recognition of unbound cohesins, thus leading to a continuous reorganization of the cellulosome. The potential switching of dockerins between various protein partners within the cellulosome may introduce the required conformational flexibility in the quaternary structure of the multi-enzyme complex. Nevertheless, *C. thermocellum* expresses a set of dockerins that have not retained this molecular conservation at the two binding sites. Significantly, two of these dockerins, Doc-258 and Doc-435, appear to bind preferentially to the type I cohesin of OlpC rather than to CipA cohesins. Thus, the data suggest that a particular set of enzymes might preferentially bind directly to the bacterium cell surface rather than the cellulosome. One of these enzymes, Cthe_0435, was recently shown to display cellulase activity. The biological significance of *C. thermocellum* targeting a set of enzymes to the cell envelop instead of the cellulosome remains unknown. However, it should be noted that the cellulosome, when appended to the surface of *C. thermocellum*, is a more efficient cellulose degrading nanomachine than when the complex is released into the culture media [26]. It is possible that the increased activity reflects synergistic interactions between catalytic components of the cellulosome and enzymes directly appended to the surface of the bacterium.

ACKNOWLEDGEMENTS

This work was supported by grants PPCDT/BIA-PRO/59118/2004 and PTDC/BIA-PRO/69732/2006 from the Fundação para a Ciência e a Tecnologia, Portugal, and by the individual grant SFRH/BD/25439/05 (to B.A.P.).

FIGURE LEGENDS

Figure 1 Simplified representation of *Clostridium thermocellum* cellulosome

The cell bound anchoring scaffoldins OlpB, Orf2 and Cthe_0736 were excluded from the figure for simplicity. The bacterium expresses two cell anchored proteins, OlpA and OlpC (described in this work), which contain type I cohesins. In addition, CipA functions as the cellulosome primary scaffoldin and contains nine type I cohesin domains. The molecular architecture of enzymes whose dockerins were studied in this work is displayed.

Figure 2 Structural alignment of *C. thermocellum* and *C. cellulolyticum* representative type I dockerins

Alignment of *C. thermocellum* dockerin domains of Xyn10B (Doc-Xyn10B), Cel9-GH44 (Doc-Cel44A), Cthe_0258 (Doc-258), Cthe_0435 (Doc-435) and Cthe_0918 (Doc-918) with the *C. cellulolyticum* dockerin of Cel5A (Doc-cellulolyticum). The secondary structure elements of Doc-Xyn10B are shown above the alignment. The alignment on the top represents the cohesin interacting residues at the N-terminal binding interface. The alignment below represents the cohesin interacting residues at the C-terminal binding interface. Boxes indicate the residues participating in direct hydrogen bonds with cohesin residues. The triangles at the top indicate the amino acids residues involved in water mediated hydrogen bonds with cohesins. Residues in grey participate in cohesin hydrophobic interactions.

Figure 3 OlpC is a cell surface protein

Adsorption of CtUnk1, the N-terminal domain of OlpC, to *C. thermocellum* and *E. coli* cellular preparations. CtUnk1 was incubated with (lanes 1 and 2) or without (lanes 3 and 4) the cell wall extracts in 20 mM sodium phosphate buffer (pH 7.5) on ice for 30 min with occasional shaking. Protein mixtures were centrifuged at 40,000 x g for 30 min and then the supernatant (lanes 1 and 3) and precipitated (lanes 2 and 4) fractions were subjected to SDS-PAGE.

Figure 4 Examples of the isothermal titration calorimetry (ITC) experiments with selected cohesin and dockerin domains expressing different affinities

The upper parts of each panel show the raw heats of binding, whereas the lower parts are the integrated heats after correction for heat dilution. The curve represents the best fit to a single-site binding model. (A) Coh-OlpA plus Doc-258. (B) Coh-Cel plus Doc-Cel44A. (C) Coh-CipA2 plus Doc-Xyn10B. (D) Coh-OlpA plus Doc-Cel44A.

Figure 5 The cohesin-dockerin interaction as evaluated by non-denaturing gel electrophoresis

(A) Detection of cohesin-dockerin complexes. Equimolar amounts of Coh-CipA2 and Doc-Xyn10B were mixed for 1 hour and the resulting protein complex (Doc-Xyn10B/Coh-CipA2) was separated by electrophoresis. (B) The stability of Doc-Cel44A/Coh-CipA2 and Doc-Cel44A/Coh-OlpC complexes in the presence, respectively, of Coh-OlpC and Coh-CipA2. Note that the addition of Coh-CipA2 to the previously assembled Doc-Cel44A/Coh-OlpC complex led to dockerin switching and formation of a novel Doc-Cel44A/Coh-CipA2 complex. Under the same circumstances, the Doc-Cel44A/Coh-CipA2 complex remains stable in the presence of Coh-OlpC. (C) The stability of Doc-Cel44A/Coh-CipA2 and Doc-Xyn10B/Coh-CipA2 complexes in the presence, respectively, of Doc-Xyn10B and Doc-Cel44A. Note that the addition of the dockerins to the previously assembled complexes had no effect on complex stability.

(D) The preference for cohesin partners by dockerin Doc-Cel44A. The dockerin was incubated with Coh-CipA2 or Coh-OlpC, individually or in combination with the two cohesins, and complex formation monitored as described above. When the three proteins are mixed together only the complex Doc-Cel44A/Coh-CipA2 is formed demonstrating that Doc-Cel44A has a preference for Coh-CipA2.

Figure 6 Doc-Cel44A cannot bind *C. thermocellum* and *C. cellulolyticum* cohesins simultaneously

Doc-Cel44A was mixed with Coh-CipA2 or Coh-Ccel and the electrophoretic mobility of the assembled complexes was monitored. When the dockerin was mixed, simultaneously, with the two cohesins there is only evidence for the formation of the Doc-Cel44A/Coh-CipA2 complex, suggesting that the dockerin is unable to bind the two cohesins at the same time and has a preference for Coh-CipA2.

Supplementary figure

Figure 1s Structural alignment of *C. thermocellum* and *C. cellulolyticum* representative type I cohesins

Alignment of *C. thermocellum* type I cohesin domains of CipA, OlpA and OlpC with a representative cohesin domain from *C. cellulolyticum* (Coh-cellulolyticum). The secondary structure elements of the second cohesin of CipA (Coh-CipA2) are shown above the alignment. Boxes indicate the residues participating in direct hydrogen bonds. The triangles at the top indicate the amino acids residues involved in water mediated hydrogen bonds. Residues in grey participate in cohesin hydrophobic interactions.

REFERENCES

- 1 Bayer, E. A., Morag, E. and Lamed, R. (1994) The cellulosome--a treasure-trove for biotechnology. *Trends in biotechnology*. **12**, 379-386
- 2 Beguin, P. and Alzari, P. M. (1998) The cellulosome of *Clostridium thermocellum*. *Biochemical Society transactions*. **26**, 178-185
- 3 Doi, R. H. and Tamaru, Y. (2001) The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. *Chemical record* (New York, N.Y.). **1**, 24-32
- 4 Fierobe, H. P., Bayer, E. A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y. and Belaich, J. P. (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. *The Journal of biological chemistry*. **277**, 49621-49630
- 5 Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P. and Bayer, E. A. (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. *The Journal of biological chemistry*. **276**, 21257-21261
- 6 Gerngross, U. T., Romaniec, M. P., Kobayashi, T., Huskisson, N. S. and Demain, A. L. (1993) Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal SL-protein reveals an unusual degree of internal homology. *Molecular microbiology*. **8**, 325-334
- 7 Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. and Beguin, P. (1995) OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. *Journal of bacteriology*. **177**, 2451-2459
- 8 Salamitou, S., Lemaire, M., Fujino, T., Ohayon, H., Gounon, P., Beguin, P. and Aubert, J. P. (1994) Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. *Journal of bacteriology*. **176**, 2828-2834
- 9 Adams, J. J., Pal, G., Jia, Z. and Smith, S. P. (2006) Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America*. **103**, 305-310
- 10 Carvalho, A. L., Dias, F. M., Nagy, T., Prates, J. A., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M., Romao, M. J., Fontes, C. M. and Gilbert, H. J. (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America*. **104**, 3089-3094
- 11 Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J. and Fontes, C. M. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 13809-13814
- 12 Karpol, A., Barak, Y., Lamed, R., Shoham, Y. and Bayer, E. A. (2008) Functional asymmetry in cohesin binding belies inherent symmetry of the dockerin module: insight into cellulosome assembly revealed by systematic mutagenesis. *The Biochemical journal*. **410**, 331-338
- 13 Pinheiro, B. A., Proctor, M. R., Martinez-Fleites, C., Prates, J. A., Money, V. A., Davies, G. J., Bayer, E. A., Fontesm, C. M., Fierobe, H. P. and Gilbert, H. J. (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *The Journal of biological chemistry*. **283**, 18422-18430

- 14 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
- 15 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 nonrecognition and high affinity recognition. *The Journal of biological chemistry*. **276**, 9883-9888
- 16 Mechaly, A., Yaron, S., Lamed, R., Fierobe, H. P., Belaich, A., Belaich, J. P., Shoham, Y. and Bayer, E. A. (2000) Cohesin-dockerin recognition in cellulosome assembly: experiment versus hypothesis. *Proteins*. **39**, 170-177
- 17 Carvalho, A. L., Goyal, A., Prates, J. A., Bolam, D. N., Gilbert, H. J., Pires, V. M., Ferreira, L. M., Planas, A., Romao, M. J. and Fontes, C. M. (2004) The family 11 carbohydrate-binding module of *Clostridium thermocellum* Lic26A-Cel5E accommodates beta-1,4- and beta-1,3-1,4-mixed linked glucans at a single binding site. *The Journal of biological chemistry*. **279**, 34785-34793
- 18 Carvalho, A. L., Pires, V. M., Gloster, T. M., Turkenburg, J. P., Prates, J. A., Ferreira, L. M., Romao, M. J., Davies, G. J., Fontes, C. M. and Gilbert, H. J. (2005) Insights into the structural determinants of cohesin-dockerin specificity revealed by the crystal structure of the type II cohesin from *Clostridium thermocellum* SdbA. *Journal of molecular biology*. **349**, 909-915
- 19 Zhao, G., Ali, E., Sakka, M., Kimura, T. and Sakka, K. (2006) Binding of S-layer homology modules from *Clostridium thermocellum* SdbA to peptidoglycans. *Applied microbiology and biotechnology*. **70**, 464-469
- 20 Fontes, C. M., Hazlewood, G. P., Morag, E., Hall, J., Hirst, B. H. and Gilbert, H. J. (1995) Evidence for a general role for non-catalytic thermostabilizing domains in xylanases from thermophilic bacteria. *The Biochemical journal*. **307** (Pt 1), 151-158
- 21 Ahsan, M., Matsumoto, M., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (1997) Purification and characterization of the family J catalytic domain derived from the *Clostridium thermocellum* endoglucanase CelJ. *Bioscience, biotechnology, and biochemistry*. **61**, 427-431
- 22 Ahsan, M. M., Kimura, T., Karita, S., Sakka, K. and Ohmiya, K. (1996) Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *Journal of bacteriology*. **178**, 5732-5740
- 23 Najmudin, S., Guerreiro, C. I., Carvalho, A. L., Prates, J. A., Correia, M. A., Alves, V. D., Ferreira, L. M., Romao, M. J., Gilbert, H. J., Bolam, D. N. and Fontes, C. M. (2006) Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. *The Journal of biological chemistry*. **281**, 8815-8828
- 24 Zhao, G., Li, H., Wamalwa, B., Sakka, M., Kimura, T. and Sakka, K. (2006) Different binding specificities of S-layer homology modules from *Clostridium thermocellum* AncA, Slp1, and Slp2. *Bioscience, biotechnology, and biochemistry*. **70**, 1636-1641
- 25 Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., Jr., McKeown, C. K., Lankford, P. K., Samatova, N. F. and Mielenz, J. R. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PloS one*. **4**, e5271

- 26 Lu, Y., Zhang, Y. H. and Lynd, L. R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. Proceedings of the National Academy of Sciences of the United States of America. **103**, 16165-16169

Table 1 Primers used to obtain the genes encoding the cohesin and dockerin derivatives used in this work and for the mutagenesis of Doc-Cel44A

Engineered restriction sites and mutation points are depicted in bold.

| <i>Clone</i> | <i>Sequence (5'→3')</i> | <i>Direction</i> |
|-----------------|---|------------------|
| Doc-435 | CTC GAA TTC TGG AAT AAG GCA GTT ATT | FOR |
| | CAC CTC GAG TTA TGC CGA ATT GTA AGA G | REV |
| Doc-918 | CTC GAA TTC GTT GTG CTT AAT GGT GAC | FOR |
| | CAC CTC GAG CTA TAT AGT TAT AAG TCC | REV |
| Doc-258 | CTC GAA TTC GCG GAC GGT AAA AAT GTG | FOR |
| | CAC CTC GAG TTA TCC GGG AAA TTT CTC | REV |
| Doc-Xyn10B | CTC GCT AGC AAA CCG GAA GAG CCG GAT G | FOR |
| | CAC CTC GAG GAT TAC TCT TAA AAG GTA G | REV |
| Doc-Cel44A | CTC GCT AGC GCA AAA GTG GTT GAC ATC | FOR |
| | CAC CTC GAG CCA GTC AAT AGC ATC TAC | REV |
| Coh-CipA2 | CTC GCT AGC GAC GGT GTG GTA GAT GGC | FOR |
| | CAC CTC GAG CTT GGT CGG TGT TGC ATT G | REV |
| Coh-OlpA | CTC GGA TCC CAA ACA AAC ACC ATT GAA | FOR |
| | CAC GAA TTC TGC CTC CGG AGC GGA TGC | REV |
| Coh-OlpC | CTC AGA TCT GTT GTG GCA ATT CAT GAA | FOR |
| | CAC GAA TTC TTA TTT TTC AAT TTC CAC | REV |
| CtUnk1 | CTC GAA TTC GCT GAG GCA AAC CAG TGG ACG | FOR |
| | CAC CTC GAG TTA TAT ATA GTC CGC AGT TTC | REV |
| DocCipAII | CTC GAA TTC AAT AAA CCT GTA ATA GAA | FOR |
| | CAC CTC GAG TTA CTG TGC GTC GTA ATC | REV |
| Doc-Cel44,1stQQ | C TCT AAA GTA AAC CAA CAA GAC ATT ATG ATG | FOR |
| | CAT CAT AAT GTC TTG TTG GTT TAC TTT AGA G | REV |
| Doc-Cel44,2sdQQ | GGT GTT GTA AAT CAA CAA GAC TAT AAT ATA ATG | FOR |
| | CAT TAT ATT ATA GTC TTG TTG ATT TAC AAC ACC | REV |

Table 2 Recognition of *C. thermocellum* and *C. cellulolyticum* cohesins by *C. thermocellum* dockerins as evaluated by ITC.
Thermodynamic parameters were determined at 55°C

| Dockerin | Cohesin | | | | | | | |
|------------|-----------------------------|------------------------|------------------------|------------------------|-----------------------------|------------------------|------------------------|------------------------|
| | Coh-CipA2 | | | | Coh-OlpA | | | |
| | K_a | ΔG° | ΔH° | $T\Delta S^\circ$ | K_a | ΔG° | ΔH° | $T\Delta S^\circ$ |
| | M ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | M ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ |
| Doc-Xyn10B | $(2.1 \pm 0.6) \times 10^8$ | -12.51 ± 0.22 | -26.06 ± 0.22 | -13.55 | $(2.2 \pm 0.3) \times 10^8$ | -12.53 ± 0.19 | -26.47 ± 0.19 | -13.94 |
| Doc-Cel44A | $(1.6 \pm 3.0) \times 10^9$ | -13.82 ± 0.08 | -12.66 ± 0.08 | 1.16 | $(5.9 \pm 0.4) \times 10^9$ | -14.68 ± 0.08 | -21.60 ± 0.08 | -6.92 |
| Doc-435 | $(9.7 \pm 1.1) \times 10^6$ | -10.49 ± 0.19 | -19.90 ± 0.19 | -9.41 | $(9.5 \pm 2.2) \times 10^6$ | -10.49 ± 0.35 | -25.32 ± 0.35 | -14.83 |
| Doc-918 | $(1.9 \pm 0.4) \times 10^8$ | -12.42 ± 0.20 | -32.59 ± 0.20 | -20.17 | $(9.7 \pm 2.7) \times 10^6$ | -10.57 ± 62.3 | -45.01 ± 0.12 | -34.44 |
| Doc-258 | $(4.2 \pm 1.4) \times 10^6$ | -9.95 ± 0.11 | -21.92 ± 0.11 | -11.97 | $(4.8 \pm 1.7) \times 10^6$ | -10.04 ± 0.20 | -17.45 ± 0.50 | -7.41 |
| Dockerin | Cohesin | | | | | | | |
| | Coh-OlpC | | | | Coh-Ccel | | | |
| | K_a | ΔG° | ΔH° | $T\Delta S^\circ$ | K_a | ΔG° | ΔH° | $T\Delta S^\circ$ |
| | M ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | M ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ | kcal mol ⁻¹ |
| Doc-Xyn10B | $(1.4 \pm 0.1) \times 10^8$ | -12.23 ± 0.15 | -34.01 ± 0.15 | -21.78 | $(9.8 \pm 1.6) \times 10^4$ | -7.48 ± 1.31 | -12.50 ± 1.31 | -5.02 |
| Doc-Cel44A | $(2.2 \pm 0.4) \times 10^8$ | -12.63 ± 0.21 | -22.90 ± 0.21 | -10.27 | $(9.2 \pm 0.9) \times 10^7$ | -11.96 ± 0.06 | -12.69 ± 0.06 | -0.73 |
| Doc-435 | $(1.1 \pm 0.1) \times 10^8$ | -12.15 ± 0.27 | -48.84 ± 0.27 | -36.69 | $(3.3 \pm 0.2) \times 10^5$ | -8.31 ± 0.63 | -19.79 ± 0.63 | -11.41 |
| Doc-918 | $(3.5 \pm 0.3) \times 10^6$ | -9.84 ± 0.56 | -33.36 ± 0.56 | -23.52 | $(3.1 \pm 0.5) \times 10^5$ | -8.24 ± 0.80 | -9.66 ± 0.80 | -1.42 |
| Doc-258 | $(9.2 \pm 1.4) \times 10^6$ | -10.46 ± 0.37 | -31.48 ± 0.37 | -21.02 | $(4.6 \pm 0.3) \times 10^5$ | -8.50 ± 0.22 | -13.39 ± 0.22 | -4.89 |

Table 3 Stability of previously assembled cohesin-dockerin complexes to the presence of newly introduced cohesins or dockerins

Cohesin-dockerin complexes were assembled and mixed with unbound cohesin or dockerins. R indicates that the complex was reverted and that a second complex was formed. NR indicates that the complex was not reverted with the addition of the cohesin/dockerin. Nd means that the stability of the complex was not determined.

| Complex | | Dissociating Cohesin | | | | |
|------------|------------|-----------------------|------------|----------|---------|---------|
| Dockerin | Cohesin | Coh-CipA2 | Coh-OlpA | Coh-OlpC | | |
| Doc-Xyn10B | Coh-CipA2 | - | R | NR | | |
| | Coh-OlpA | NR | - | NR | | |
| | Coh-OlpC | R | R | - | | |
| Doc-Cel44A | Coh-CipA2 | - | nd | NR | | |
| | Coh-OlpA | nd | - | NR | | |
| | Coh-OlpC | R | R | - | | |
| Doc-435 | Coh-CipA2 | - | nd | R | | |
| | Coh-OlpA | nd | - | nd | | |
| | Coh-OlpC | NR | nd | - | | |
| Doc-918 | Coh-CipA2 | - | NR | NR | | |
| | Coh-OlpA | NR | - | NR | | |
| | Coh-OlpC | R | nd | - | | |
| Doc-258 | Coh-CipA2 | - | nd | R | | |
| | Coh-OlpA | nd | - | R | | |
| | Coh-OlpC | NR | NR | - | | |
| Cohesin | + Dockerin | Dissociating Dockerin | | | | |
| | | Doc-Xyn10B | Doc-Cel44A | Doc-435 | Doc-918 | Doc-258 |
| Coh-CipA2 | Doc-Xyn10B | - | NR | nd | NR | nd |
| | Doc-Cel44A | NR | - | nd | NR | nd |
| | Doc-435 | nd | nd | - | nd | nd |
| | Doc-918 | NR | NR | nd | - | nd |
| | Doc-258 | nd | nd | nd | nd | - |
| Coh-OlpA | Doc-Xyn10B | - | NR | nd | nd | nd |
| | Doc-Cel44A | NR | - | nd | NR | NR |
| | Doc-435 | nd | nd | - | nd | nd |
| | Doc-918 | nd | NR | nd | - | NR |
| | Doc-258 | nd | PR | nd | R | - |
| Coh-OlpC | Doc-Xyn10B | - | NR | nd | nd | NR |
| | Doc-Cel44A | NR | - | nd | nd | nd |
| | Doc-435 | nd | nd | - | nd | NR |
| | Doc-918 | nd | nd | nd | - | nd |
| | Doc-258 | R | nd | NR | nd | - |

Table 4 Identification of preferred cohesin and dockerin partners as evaluated by native gel electrophoresis

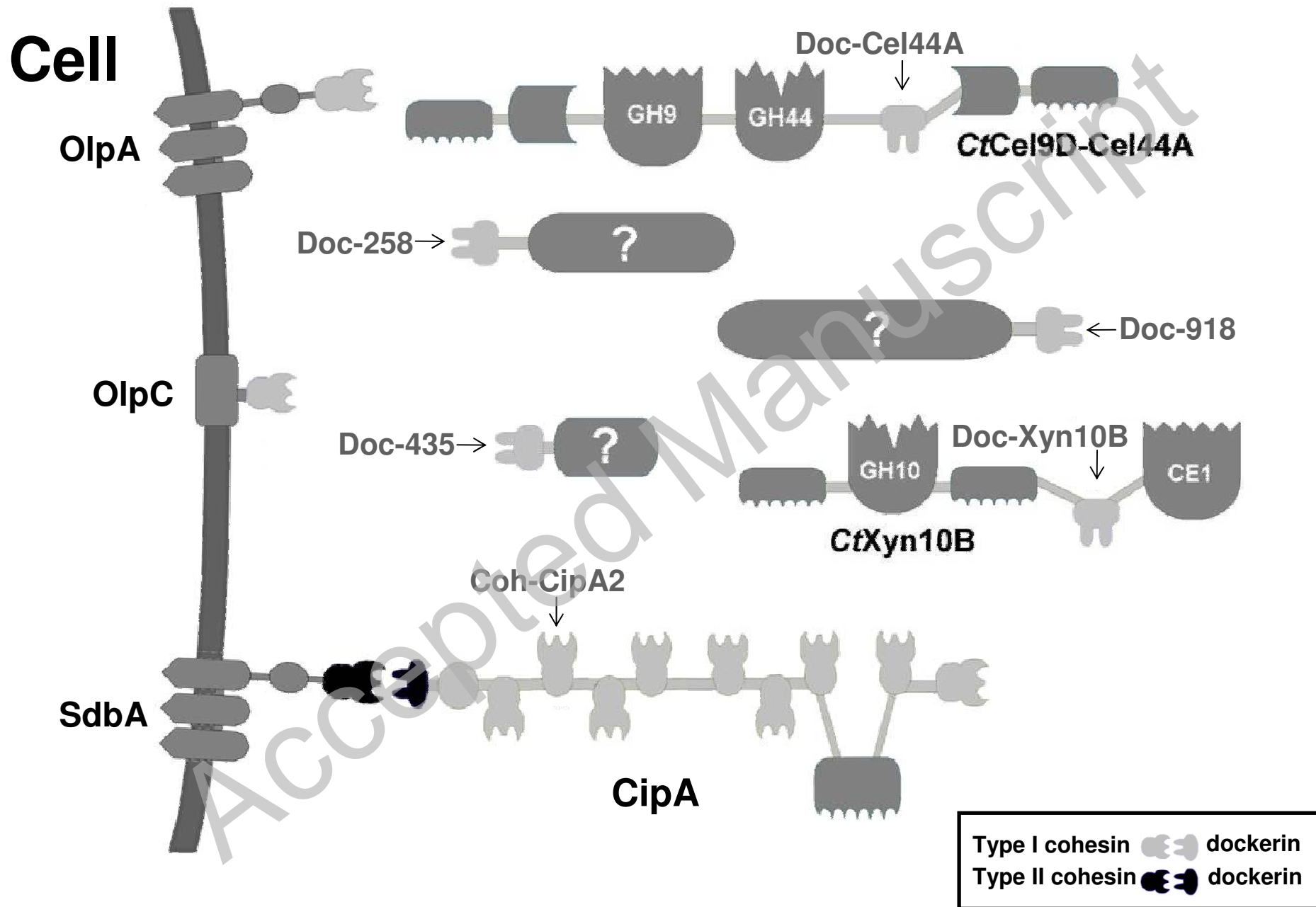
One dockerin and two cohesins or *vice versa* (one cohesin and two dockerins) were mixed and the resulting cohesin-dockerin complexes formed were analysed by native gel electrophoresis. Using this strategy it was possible to test complex formation derived from the mixture of one dockerin with all possible combination of cohesins and *vice versa* for cohesins. Thus, using this strategy it was possible to analyze the preference for protein partners of both cohesins and dockerins.

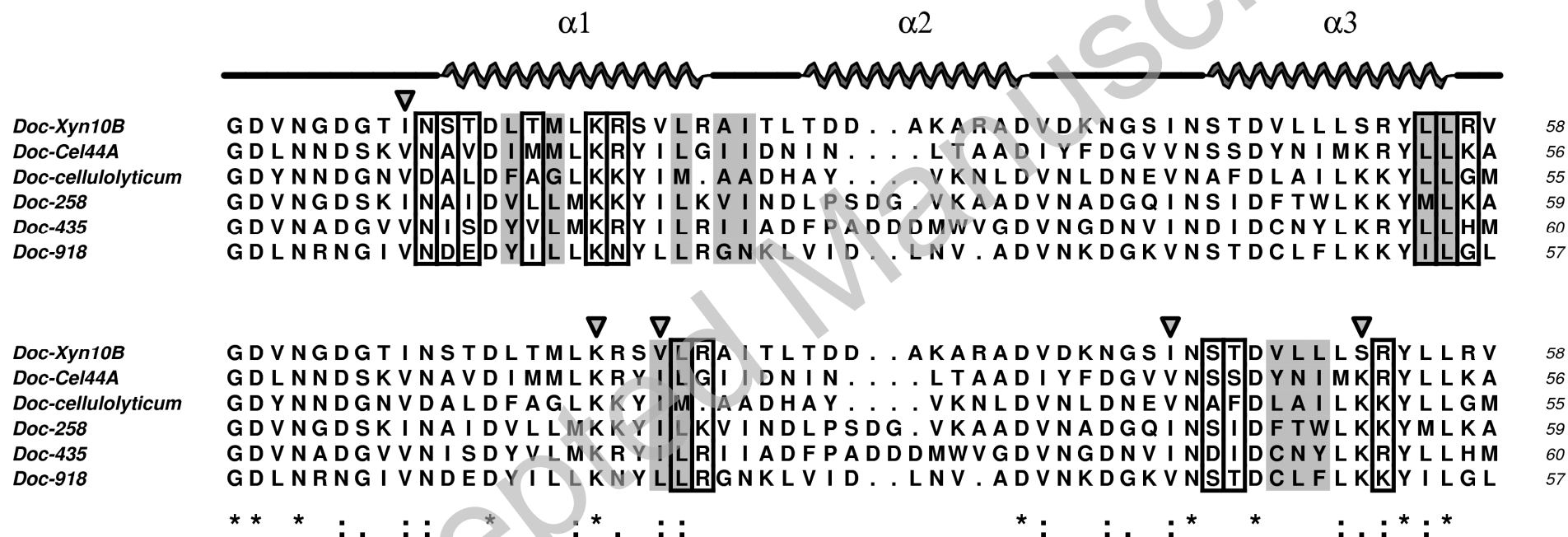
| Dockerin | Dockerin binding preference |
|------------|---------------------------------------|
| Doc-Xyn10B | Coh-OlpA>Coh-CipA2>Coh-OlpC |
| Doc-Cel44A | Coh-OlpA>Coh-CipA2>Coh-OlpC>Coh-Ccel |
| Doc-435 | Coh-OlpC>Coh-CipA2>Coh-OlpA |
| Doc-918 | Coh-CipA2=Coh-OlpA>Coh-OlpC |
| Doc-258 | Coh-OlpC>Coh-CipA2>Coh-OlpA |
| Cohesin | Cohesin binding preference |
| Coh-CipA2 | Xyn10B=Cel44A=Doc-918>Doc-258>Doc-435 |
| Coh-OlpA | Xyn10B=Cel44A=Doc-918>Doc-258>Doc-435 |
| Coh-OlpC | Doc-435>Xyn10B=Cel44A>Doc-258>Doc-918 |

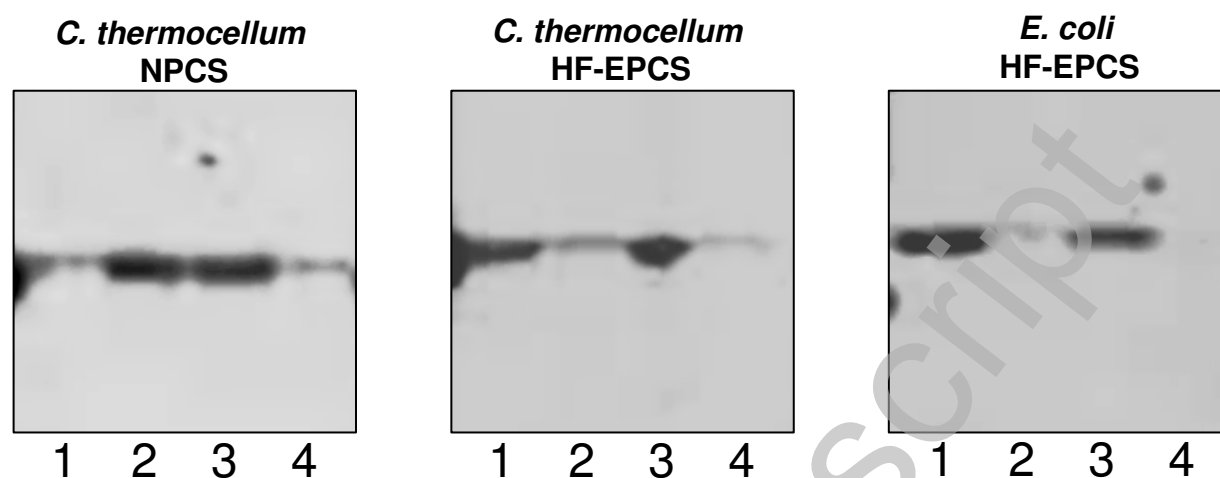
Table 5 Thermodynamics of the binding between wild type dockerin Doc-Cel44A and its mutant derivatives and *C. cellulolyticum* (Coh-Ccel) and *C. thermocellum* (Coh-CipA2) cohesins

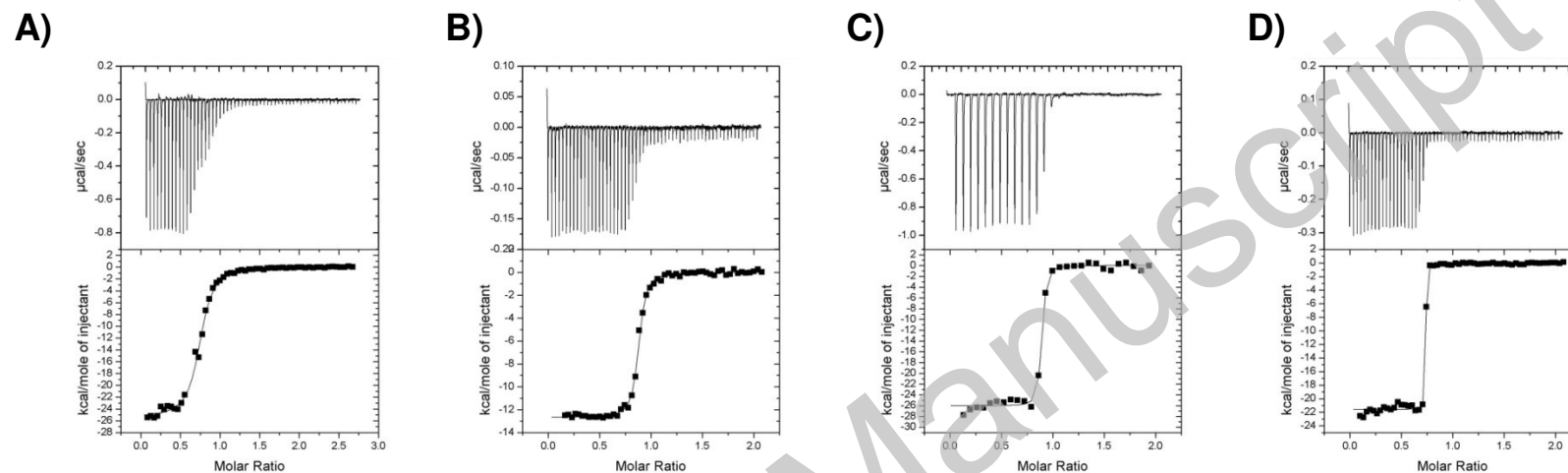
Thermodynamic parameters were determined at 55°C. Nd means that the values were too low to be determined.

| Cohesins | Constants | Doc-Cel44A | | | |
|-----------|---|-----------------------------|-----------------------------|-----------------------------|------------------------------|
| | | Wild type | 1 st QQ | 2 nd QQ | QQ QQ |
| Coh-CipA2 | $K_a \text{ M}^{-1}$ | $(1.6 \pm 3.0) \times 10^9$ | $(5.6 \pm 1.2) \times 10^8$ | $(4.8 \pm 0.8) \times 10^7$ | $(3.72 \pm 1.1) \times 10^7$ |
| | $\Delta G^\circ \text{ kcal mol}^{-1}$ | -13.82 ± 0.08 | -13.08 ± 2.80 | -11.49 ± 0.97 | -11.32 ± 0.31 |
| | $\Delta H^\circ \text{ kcal mol}^{-1}$ | -12.66 ± 0.08 | -26.10 ± 2.80 | -16.25 ± 0.97 | -14.30 ± 0.31 |
| | $T\Delta S^\circ \text{ kcal mol}^{-1}$ | 1.16 | -13.02 | -4.76 | -2.98 |
| Coh-Ccel | $K_a \text{ M}^{-1}$ | $(5.3 \pm 0.7) \times 10^7$ | $(1.4 \pm 0.3) \times 10^5$ | $(6.1 \pm 1.0) \times 10^7$ | $< 10^4$ |
| | $\Delta G^\circ \text{ kcal mol}^{-1}$ | -11.61 ± 0.13 | -7.76 ± 8.34 | -11.70 ± 0.24 | Nd |
| | $\Delta H^\circ \text{ kcal mol}^{-1}$ | -16.04 ± 0.13 | -19.96 ± 8.34 | -21.80 ± 0.24 | Nd |
| | $T\Delta S^\circ \text{ kcal mol}^{-1}$ | -4.43 | -12.20 | -10.10 | Nd |

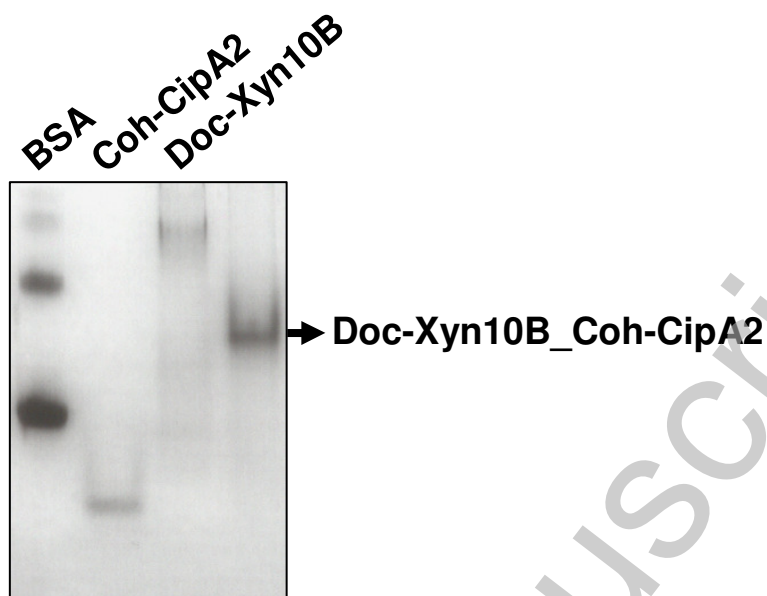




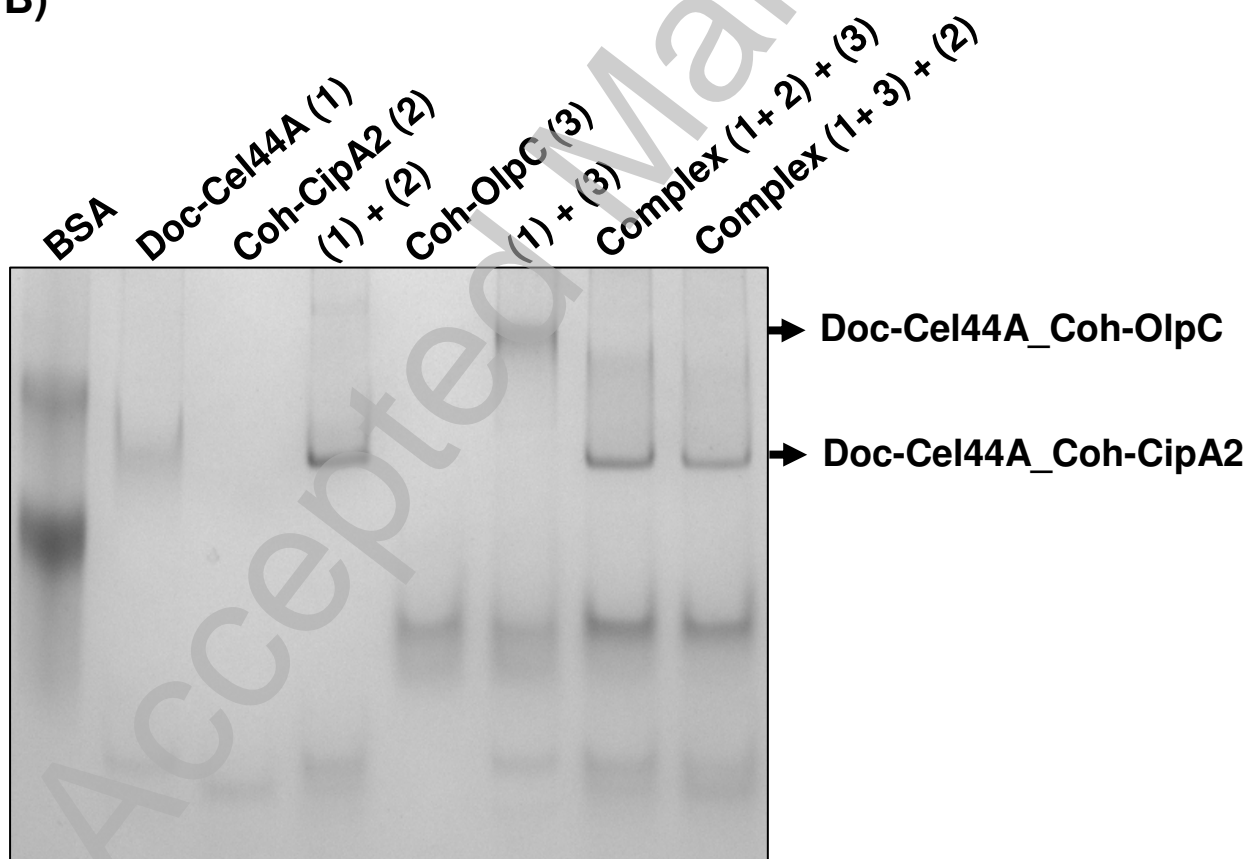




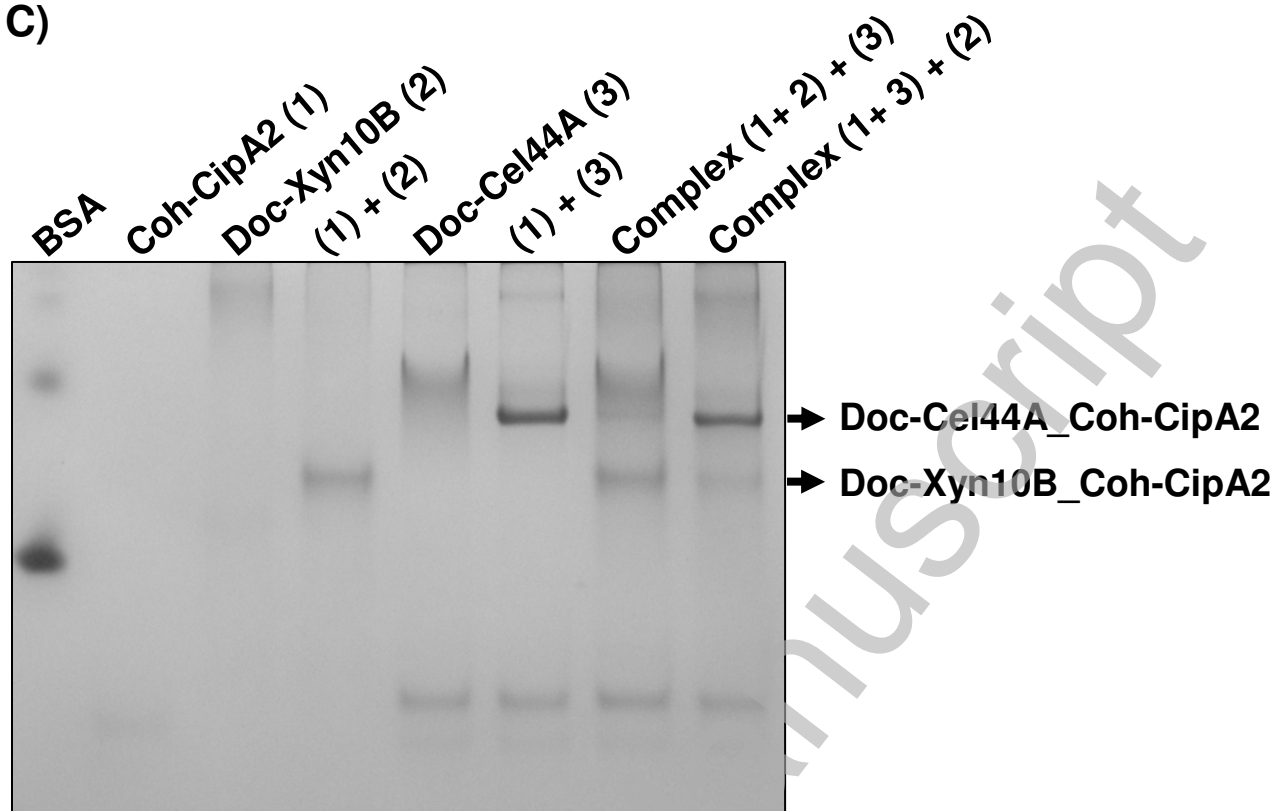
A)



B)



C)



D)

