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 *C. 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. In the present study, 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 on to the multi-enzyme complex.

Key words: cellulosome, cell-surface attachment, cohesin, dockerin, molecular scaffoldin, protein–protein interaction.

**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-mass multienzyme 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 CBM (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 CBM that attaches the cellulosome on to 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 on to 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, on to the bacterial surface [8].

Crystal structures of type I cohesin–dockerin complexes have provided important insights into the mechanism of cellulosome assembly. Dockersins 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 displays remarkable structural conservation: the N- and C-terminal α-helices overlay with minor structural deviations [9–11]. Internal 2-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 serine-threonine motif at positions 11 and 12 and a lysine-arginine 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 multienzyme 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 similar structural complexity 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 hydroxyl residues, whereas 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 motif in one of the ligand-binding interfaces. In these dockerins the usually conserved serine-threonine pair, which dominates the hydrogen-bond network with the cohesin, is replaced in one of the duplicated segments by non-hydroxyl residues. The implications of these amino acid substitutions for cohesin recognition remain to be investigated. In the present study, 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). Amplified DNA was directly cloned into pNZY28 (NZYTech) 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 EcoRI and XhoI (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 (the family 22 CBM) xylan-binding domain and the Cel9D-Cel44A dockerin was expressed with the GH44 catalytic domain at the N-terminus and the PKD (polycystic kidney-disease-like) module followed by CBM44 (the family 44 CBM) 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 NheI and XhoI 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) and sequenced. Coh-OlpA and Coh-OlpC genes were subcloned into BglII/EcoRI- and BamHI/EcoRI-digested pRSETa respectively (Invitrogen). The gene encoding Coh-CipA2 was subcloned into pET21a (Novagen). The DNA sequences encoding the N-terminal domain of Cthe_0452 (residues 25–112), termed CUniK, 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 shown in Table 1. Both genes were subsequently subcloned into pET32a (Novagen). C. cellulolyticum type I cohesin was produced using plasmid pHFI, as described by Pinheiro et al. [13]. All recombinant proteins contained an internal or a C-terminal His-tag.

Escherichia coli Origami, BL21 and Tuner cells, transformed with pET32a, pET21a and pRSETa derivatives respectively, were grown at 37 °C to mid-exponential phase (\(D_{\text{max}} = 0.6\)). Recombinant protein expression was induced by adding 0.2 (Tuner) or 1 mM (all other E. coli strains) IPTG (isopropyl \(\beta\)-D-thiogalactoside) and incubation for a further 16 h at 19 °C. Soluble recombinant proteins were purified by immobilized metal-ion-affinity chromatography as described previously.

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

<table>
<thead>
<tr>
<th>Clone</th>
<th>Sequence (5′→3′)</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc-435</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-918</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-258</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Reverse</td>
</tr>
<tr>
<td>Doc-Xyn10B</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-Cel44A</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Reverse</td>
</tr>
<tr>
<td>Coh-CipA2</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Reverse</td>
</tr>
<tr>
<td>CUniK</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-CipAll</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-Cel44,1stQ0</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
<tr>
<td>Doc-Cel44,2ndQ0</td>
<td>CTGAAGTCGGATCCAAACATCAAGCCTTAATG</td>
<td>Forward</td>
</tr>
</tbody>
</table>

© The Authors. Journal compilation © 2009 Biochemical Society
Novel type I cohesin–dockerin specificities in Clostridium thermocellum

Mutagenesis

Site-directed mutagenesis was carried out using the PCR-based NZYMutagenesis site-directed mutagenesis kit (NZYTech) according to the manufacturer’s protocol, using plasmid pD44-21a as a 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

NPCS (native peptidoglycan-containing sacculi) and HF-EPCS (hydrofluoric acid-extracted cell-wall polymer) 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 (16 000g for 15 min at 4°C) and the supernatant and pellet fractions were analysed 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 h at room temperature (25°C) and the presence of complexes was evaluated by non-denaturing PAGE. 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 dockers and one cohesin (or two cohesins and one dockerin) and analysing the gel for complex formation. To determine the stability of cohesin–dockerin complexes, a cohesin–dockerin pair was assembled for 1 h. After this incubation period, a second cohesin or dockerin was added to the mixture and incubated for a further hour. Complex stability was evaluated in non-denaturing gels, as explained above.

ITC (isothermal titration calorimetry) of cohesin–dockerin binding

ITC was carried out essentially as described previously [10,13], except that measurements were made at 55°C, and proteins were dialysed into 50 mM sodium Hepes (pH 7.5) containing 2 mM CaCl₂. During titration, the dockerin (15 μM) was stirred at 300 rev/min 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 analysed by non-linear regression using a single site-binding model (Microlab ORIGIN, Version 5.0; Microlab Software). The fitted data yield the association constant (K_a) and the enthalpy of binding (∆H). Other thermodynamic parameters were calculated by using the standard thermodynamic equation:

\[ \Delta R T \ln K_a = -\Delta G = \Delta H - T \Delta S \]

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 dockers. Alignment of all of the identified dockers 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 (results not shown). These domains belong to proteins Cthc_0258, Cthc_0435, Cthc_0624 and Cthc_0918. Cthc_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- (Cthc_0258 and Cthc_0435) or C- (Cthc_0918) terminus (Figure 1). Modules of unknown function in Cthc_0258, Cthc_0435 and Cthc_0918 have 365, 218 and 1136 residues respectively, and may comprise catalytic domains, or possibly CBMs, which contribute to the deconstruction of the plant cell wall. Indeed, the 218 residue module in Cthc_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 work). Alignment of the four divergent dockers 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 cohesins [11]. Doc-Cel44A, Doc-258 and Doc-918, derived from Cel9D-Cel44A, Cel9D-Cel258 and Cel9D-Cel918 respectively, lack the serine-threonine pair in the first duplicated segment; these residues are substituted by an alanine residue and a hydrophobic residue (Doc-Cel258 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 hydroxyl 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 dockers, 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 (Cthc_3077), which contains nine cohesin modules, OlpA, with one cohesin module (Cthc_3080), and a previously unknown protein, Cthc_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 (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, Cthc_0452 contains a C-terminal type I cohesin domain and an N-terminal
Figure 1  Simplified representation of C. thermocellum cellulosome

The cell-bound anchoring scaffoldins OlpB, Orf2 and Cthe_0736 are excluded from the Figure for simplicity. The bacterium expresses two cell-anchor ed proteins, OlpA and OlpC (described in the present study), 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 the present paper are 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 highlighted in grey participate in cohesin hydrophobic interactions.

module of unknown function. Alignment of the eleven C. thermocellum type I cohesin domains (Supplementary Figure S1 at http://www.BiochemJ.org/bj/424/bj4240375add.htm) 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 residue 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, Asp59 and Glu131, are also replaced by an asparagine and a proline residue 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
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 × g for 30 min and then the supernatant (lanes 1 and 3) and precipitated (lanes 2 and 4) fractions were subjected to SDS/PAGE.

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 envelope of Gram-positive bacteria, known as the bacterial S-layer, bind SCWPs (secondary cell wall polymers) rather than the peptidoglycan cell-wall matrix. The results from pulldown 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 HF-EPICS. Hydrofluoric extraction removes SCWPs from NPCS [19,24] and, therefore, HF-EPICS consists mainly of peptidoglycan. Therefore the results suggest that CtUnk1 displays affinity for C. thermocellum SCWPs. To further confirm that CtUnk1 was unable to interact with the peptidoglycan fraction of the cell wall, an HF-EPICS fraction was prepared from E. coli, which is free of Gram-positive SCWPs. The data (Figure 3) revealed that CtUnk1 was unable to interact with the E. coli HF-EPICS preparation, strongly suggesting that CtUnk1 specifically interacts with SCWPs present in C. thermocellum, or at least in Gram-positive bacterial cell walls. Since it appears that CtUnk1 directs Cthe_0452 on to 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 SCWPs [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 results suggest that proteins that are responsible for binding cellulosomes, such as SdbA, OlpB and Orf2, are bound to the peptidoglycan layer, whereas OlpA and OlpC, which bind individual cellulosomal components, interact with SCWPs. The biological significance of these identified differences in cell-wall specificity remain, however, unclear. In addition, these results 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 whether differences in cohesin–dockerin affinity can modulate the distribution of C. thermocellum enzymes over the cell surface or on to 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 analysed by ITC. Titration experiments, shown in Figure 4, were performed at 55 °C. The data, presented in Table 2, revealed a range of affinities with K_d 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 of 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 shown in Figure 5(A); 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 \( C. \) \textit{thermocellum} cohesins in CipA (Coh-CipA2) and OlpA is very tight, with an affinity constant probably higher than \( 10^8 \) M\(^{-1}\), which is the limit for quantifying molecular interactions by ITC. In contrast, DocCel44A displays a lower affinity for Coh-OlpC. A similar pattern is observed for Doc-918 which binds to the CohA cohesin \( \sim 60\)-fold more tightly than Coh-OlpC, although the \( K_a \) for Coh-OlpC is also significantly lower \( (\sim 20\)-fold) than for Coh-CipA2. By contrast the cellulosomal dockerin Doc-Xyn10B does not appear to distinguish between the three \( C. \) \textit{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 serine-threonine or serine-serine motifs at positions 11 and 12 at the N-terminal ligand interface. Indeed the preference of these two dockerins for the CohA cohesin is consistent with the presence of two hydroxyl 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 \( C. \) \textit{thermocellum} cohesins. Although Doc-435 also displayed lower affinity than the xylanase dockerin for Coh-CipA2 and Coh-OlpA, its affinity for Coh-OlpC was comparable with 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 hydroxyl 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 residue 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 dockers exhib 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 CohA cohesins, resemble the ligand-binding interfaces of \( C. \) \textit{cellulolyticum} type I dockerin (Figure 2). Work by Bayer and colleagues [15] revealed that a single threonine to leucine substitution at position 12 of a \( C. \) \textit{thermocellum} dockerin allows the protein to bind \( C. \) \textit{cellulolyticum} cohesins. Therefore the \( C. \) \textit{thermocellum} dockerins were also probed against a type I cohesin from \( C. \) \textit{cellulolyticum}, termed Coh-Ccel. The data (Table 2) revealed that Doc-Cel44A presented a high affinity for the \( C. \) \textit{cellulolyticum} cohesin Coh-Ccel \( (K_a = 10^8 \text{ M}^{-1}) \), whereas all other \( C. \) \textit{thermocellum} dockerins bound much more weakly \( (100–1000\)-fold lower affinities) to the non-cognate cohesin. Overall the results suggest that, apart from Doc-Cel44A, \( C. \) \textit{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 \( C. \) \textit{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 dockers to destabilize previously assembled complexes was
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 h 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 of Coh-OlpC and Coh-CipA2 respectively. 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 of Doc-Xyn10B and Doc-Cel44A respectively. 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 was 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.

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 Figures 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, owing to low cohesin–dockerin affinity, some protein domains were excluded from the analysis. The results, 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, whereas 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 results suggest that free dockerins cannot easily displace already bound dockerins. However, bound dockerins can easily bind to an unliganded cohesin if this is energetically favoured. 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 probably reflects their substantially higher affinity for cohesins, compared with Doc-258.

In addition, dockerin and cohesin preference for protein partners was analysed 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 results, 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 cohesive selectivity, although it did not display a preference for either Coh-CipA2 or Coh-OlpA. In contrast, with 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 serine-threonine pair of the putative cohesin-binding interface with an isoleucine residue (such as D39N and E131P) are particularly suited for the recognition of dockerin Doc-435.

Taken together these results suggest that most 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, which contain dockerins similar to Doc_Xyn10B, with considerable affinity. Coh-OlpC prefers Dockerin Doc-258, 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.

Figur e 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.

Coh-OlpA and Coh-OlpC are relatively scarce on the bacterial surface before these enzymes are present. 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 with OlpA, OlpC is relatively abundant on the surface of the bacterium [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-Cel44A.

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

<table>
<thead>
<tr>
<th>Complex</th>
<th>Coh-CipA2</th>
<th>Coh-OlpA</th>
<th>Coh-OlpC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc-Xyn10B</td>
<td>–</td>
<td>R</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>NR</td>
<td>–</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>R</td>
<td>R</td>
<td>–</td>
</tr>
<tr>
<td>Doc-Cel44A</td>
<td>nd</td>
<td>nd</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>nd</td>
<td>nd</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>R</td>
<td>R</td>
<td>–</td>
</tr>
<tr>
<td>Doc-435</td>
<td>–</td>
<td>nd</td>
<td>R</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>nd</td>
<td>–</td>
<td>nd</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>NR</td>
<td>nd</td>
<td>–</td>
</tr>
<tr>
<td>Doc-918</td>
<td>–</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>NR</td>
<td>–</td>
<td>NR</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>R</td>
<td>nd</td>
<td>–</td>
</tr>
<tr>
<td>Doc-258</td>
<td>–</td>
<td>nd</td>
<td>R</td>
</tr>
<tr>
<td>Coh-OlpA</td>
<td>nd</td>
<td>–</td>
<td>R</td>
</tr>
<tr>
<td>Coh-OlpC</td>
<td>NR</td>
<td>NR</td>
<td>–</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Dockerin/cohens</th>
<th>Binding preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc-Xyn10B</td>
<td>Coh-OlpA–Coh-CipA2–Coh-OlpC</td>
</tr>
<tr>
<td>Doc-Cel44A</td>
<td>Coh-OlpA–Coh-CipA2–Coh-OlpC–Coh-Ccel</td>
</tr>
<tr>
<td>Doc-435</td>
<td>Coh-OlpA–Coh-CipA2–Coh-OlpC–Coh-Ccel</td>
</tr>
<tr>
<td>Doc-918</td>
<td>Coh-OlpA–Coh-OlpC</td>
</tr>
<tr>
<td>Doc-258</td>
<td>Coh-OlpA–Coh-CipA2–Coh-OlpC</td>
</tr>
</tbody>
</table>

**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.
and Doc-435, which seem to favour 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 cellulomes.

**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 (alanine-valine and serine-serine respectively) were mutated to two glutamine residues, 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 results, shown in Table 5, suggest that mutation of the alanine-valine 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 alanine-valine motif had little influence on affinity for the *C. thermocellum* cohesin, whereas replacing the C-terminal serine-serine sequence with glutamine-glutamine resulted in a 30-fold reduction in the Ks. Therefore, taken together, the results suggest that Doc-Cel44A displays affinity for *C. cellulolyticum* cohesins predominantly through its N-terminal-binding interface, whereas the C-terminal ligand-binding site primarily recognizes *C. thermocellum* cohesins.

The results 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 results, shown in Figure 6, suggest 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 et al. [10] demonstrating that steric clashes prevent the binding of one dockerin to two cohesins. The biological rational for the recognition of *C. cellulolyticum* cohesins by a *C. thermocellum* dockerin remains unknown. This observation may reflect a fortuitous interaction of an as 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. The results of the present study 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 cellulomes or directly to the *C. thermocellum* cell envelope. 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 cellulomalous enzymes may transiently interact with the surface of the bacterium, through the binding to OlpC, before they are assembled into the multienzyme 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 cellulome. The potential switching of dockerins between various protein partners within the cellulome may introduce the required conformational flexibility in the quaternary structure of the multienzyme 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 cellulome. One of these enzymes, Cte_0435, was recently shown to display cellulase activity (H. J. Gilbert and C. M. G. A. Fontes, unpublished work). The biological significance of *C. thermocellum* targeting a set of enzymes to the cell envelope instead of the cellulome remains unknown. However, it should be noted that the cellulome, when appended to the surface of *C. thermocellum*, is a more efficient cellulose-degrading nanomachine than when the complex is released into the culture medium [26]. It is possible that the increased activity reflects synergistic interactions between catalytic components of the cellulome and enzymes directly appended to the surface of the bacterium.

**AUTHOR CONTRIBUTION**

Benedita Pinheiro, Harry Gilbert, Kazuo Sakka and Carlos Fontes conceived and designed the experiments. Benedita Pinheiro, Kazutaka Sakka, Vânia Fernandes, José Prates, Victor Alves, David Bolam, Luis Ferreira and Carlos Fontes performed the experiments. Benedita Pinheiro, Harry Gilbert, Kazuo Sakka and Carlos Fontes managed the project, and Benedita Pinheiro and Carlos Fontes wrote the manuscript.

---

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

<table>
<thead>
<tr>
<th></th>
<th>Coh-CipA2</th>
<th>Coh-Ccel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc-Cel44A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wild-type</td>
<td>(1.6 ± 0.9) × 10^3</td>
<td>−13.82 ± 0.08</td>
</tr>
<tr>
<td>1st QQ</td>
<td>(5.6 ± 1.2) × 10^3</td>
<td>−13.08 ± 2.80</td>
</tr>
<tr>
<td>2nd QQ</td>
<td>(4.8 ± 0.8) × 10^2</td>
<td>−11.49 ± 0.97</td>
</tr>
<tr>
<td>Glutamine-glutamine</td>
<td>(3.72 ± 1.1) × 10^2</td>
<td>−11.32 ± 0.31</td>
</tr>
</tbody>
</table>

Thermodynamic parameters were determined at 55 °C; nd means that the values were too low to be determined. QQ, pairs of glutamine residues.
Pinheiro, Harry Gilbert, Kazuo Sakka, David Bolam and Carlos Fontes analysed the data. Benedita Pinheiro, Harry Gilbert and Carlos Fontes wrote the paper.

**FUNDING**

This work was supported by the Fundação para a Ciência e a Tecnologia (Portugal) [grant numbers PPRC/DIB/PRO/59118/2004, PTDC/DBI-PRO/69732/2006, SFRH/BD/25439/05 (to B. A. P.)].

**REFERENCES**

SUPPLEMENTARY ONLINE DATA

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* and Carlos M. G. A FONTES*‡

*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, The Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, U.K., and ‡Applied Microbiology Laboratory, Graduate School of Bioresources, Mie University, Tsu 514-8507, Japan

**Figure S1** 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 highlighted in grey participate in cohesin hydrophobic interactions.

Received 28 July 2009/7 September 2009; accepted 17 September 2009
Published as BJ Immediate Publication 17 September 2009, doi:10.1042/BJ20091152

1 Present address: Complex Carbohydrate Research Centre, University of Georgia, Athens, GA 30602-4712, U.S.A.
2 To whom correspondence should be addressed (email cafontes@fmv.utl.pt).