Structure–function analyses generate novel specificities to assemble the components of multienzyme bacterial cellulosome complexes

Received for publication, December 4, 2017, and in revised form, January 11, 2018. Published, Papers in Press, January 24, 2018, DOI 10.1074/jbc.RA117.001241

Pedro Bule1, Kate Cameron1, José A. M. Prates1, Luís M. A. Ferreira1, Steven P. Smith8, Harry J. Gilbert9, Edward A. Bayer1, Shabir Najmudin, Carlos M. G. A. Fontes12, and Victor D. Alves13

From the CIISA-Faculdade de Medicina Veterinária, ULisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, the Department of Biomedical and Molecular Sciences, Queen’s University, Kingston, Ontario K7L 3N6, Canada, the Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom, and the Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100 Israel

Edited by Gerald W. Hart

The cellulosome is a remarkably intricate multienzyme nanomachine produced by anaerobic bacteria to degrade plant cell wall polysaccharides. Cellulosome assembly is mediated through binding of enzyme-borne dockerin modules to cohesin modules of the primary scaffoldin subunit. The anaerobic bacterium A. cellulolyticus produces a highly intricate cellulosome comprising an adaptor scaffoldin, ScaB, whose cohesins interact with the dockerin of the primary scaffoldin (ScaA) that integrates the cellulosomal enzymes. The ScaB dockerin selectively binds to cohesin modules in ScaC that anchors the cellulosome onto the cell surface. Correct cellulosome assembly requires distinct specificities displayed by structurally related type-I cohesin–dockerin pairs that mediate ScaC–ScaB and ScaA–enzyme assemblies. To explore the mechanism by which these two critical protein interactions display their required specificities, we determined the crystal structure of the dockerin of a cellulosomal enzyme in complex with a ScaA cohesin. The data revealed that the enzyme-borne dockerin binds to the ScaA cohesin in two orientations, indicating two identical cohesin–dockerin binding sites. Combined mutagenesis experiments served to identify amino acid residues that modulate type-I cohesin–dockerin specificity in A. cellulolyticus. Rational design was used to test the hypothesis that the ligand-binding surfaces of ScaA- and ScaB-associated dockerins mediate cohesin recognition, independent of the structural scaffold. Novel specificities could thus be engineered into one, but not both, of the ligand-binding sites of ScaB, whereas attempts at manipulating the

The atomic coordinates and structure factors (codes 5NRK and 5NRM) have been deposited in the Protein Data Bank (http://wwpdb.org/). This article contains Tables S1–S3 and Figs. S1–S3.

This work was supported by Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) through Grants PTDC/BIABIO/103980/2008, EXPL/BIA-MIC/1176/2012, and PTDC/BIA-MIC/5947/2014, from the European Union Seventh Framework Programme (FP7 2007–2013) under the WallTraC project (Grant Agreement 263916) and from BioStruct-X (Grant Agreement 283570). The authors declare that they have no conflicts of interest with the contents of this article. The atomic coordinates and structure factors (codes SNRK and SNRM) have been deposited in the Protein Data Bank (http://wwpdb.org/).

This is an Open Access article under the CC BY license.

Plant cell wall polysaccharides, primarily cellulose and hemicelluloses, are a major reservoir of carbon and energy (1), and the recycling of these complex carbohydrates by microorganisms is integral to the carbon cycle. Furthermore, as the demand for renewable sources of energy and novel molecules for the chemical industry increases, so does the environmental and industrial significance of these abundant structural macromolecules. The deconstruction of the plant cell wall requires, however, an extensive array of hydrolytic enzymes to attack this heterogeneous, predominantly insoluble and highly recalcitrant substrate (2). Specialized anaerobic bacteria have adopted an elaborate strategy to degrade structural plant carbohydrates, through the organization of enzymes into multiprotein complexes, termed cellulosomes (3). Typically, the molecular integration of microbial biocatalysts into these extremely elaborate nanomachines results from the binding of enzyme-borne type-I dockerin (Doc)4 modules to reiterated type-I cohesin (Coh) modules located in a large non-catalytic protein, termed scaffoldin, thus promoting enzyme synergism and protein stability. In addition, recruitment of cellulosomes to the bacterial cell surface via divergent type-II Coh–Doc interactions allows the immediate uptake of released sugars, which are used by microbes as an energy source (1, 4). The protein–protein interaction established between the Coh and Doc modules exhibits one of the strongest binding affinities found in nature, close to that of a covalent bond, and plays a crucial role in both cellulosome assembly and cell-surface attachment (3, 5, 6). In addition, the structural and architectural principles of cellulosomes are defined by the specificity of the different Coh and Doc modules (7).

This article contains Tables S1–S3 and Figs. S1–S3.

1 Supported by individual fellowships SFRH/BD/86821/2012 from Fundação para a Ciência e Tecnologia.
2 To whom correspondence may be addressed. E-mail: cafontes@fmv.ulisboa.pt.
3 To whom correspondence may be addressed. E-mail: vdalves@fmv.ulisboa.pt.
4 The abbreviations used are: Doc, dockerin; Coh, cohesin; ITC, isothermal titration calorimetry; NGE, non-denaturing gel electrophoresis; CBM, carbohydrate-binding module; IMAC, immobilized metal affinity chromatography; PDB, Protein Data Bank.
Dual-specificity Coh–Doc complex

The mesophilic anaerobic bacterium *Acetivibrio cellulolyticus* produces a highly efficient celluloseme capable of hydrolyzing a range of cellulosic materials (8, 9). Within the genome of *A. cellulolyticus* there is a cluster of four tandem scaffoldin genes (*scaA*, *scaB*, *scaC*, and *scaD*) (10, 11). The primary scaffoldin ScA (where the enzymes of the cellulosome are recruited) shares the main traits found in the primary ScA scaffoldin of the canonical cellulosome of *Clostridium thermocellum*. Thus, *A. cellulolyticus* ScA contains an internal family-type 3 carbohydrate-binding module (CBM3), flanked by seven type-I Coh modules and a divergent C-terminal type-II Doc (12). Downstream of ScA are genes encoding for an adaptor and an anchoring scaffoldin, ScB and ScA C, respectively. ScB contains four type-II Coh modules, which interact with the type-II Doc of ScA, and a divergent C-terminal type-I Doc that binds to the type-I Coh modules of the ScA scaffoldin. ScB essentially plays the role of an adaptor protein that mediates the interaction between ScA (and its incorporated enzymes) and ScA C, the bacterial cell-surface anchoring protein. ScB is the first example of a cellulosomal adaptor protein (10) (Fig. 1) (10). The genome sequence of *A. cellulolyticus* CD2 revealed numerous additional cellulosomal components, gene-regulatory elements, and cell-anchoring modules (identified by the presence of signature Doc or Coh sequences), suggestive of a much more elaborate and sophisticated cellulosome system than originally observed (13). In total, 143 Doc-containing *A. cellulolyticus* proteins were identified.

There is no evidence of cross-specificity between type-I and type-II Coh–Doc partners (14–16), explaining why ScB, through its Coh and Doc modules, interacts with two distinct proteins (ScA and ScA C). Structural studies on type-I complexes, primarily from *C. thermocellum* (5, 17, 18) and *Clostridium cellulolyticum* (19), revealed that the primary sequence duplication displayed by type-I Docs supports a dual-binding mode, based on the interaction of two 180° symmetry-related binding interfaces. It was recently shown that the sequence and structural symmetry within the ScB *A. cellulolyticus* type-I Doc allows it to bind ScA C Cohs in two different orientations (20). This sequence symmetry is also evident in the enzyme-borne Docs of *A. cellulolyticus* that interacts with ScA, therefore suggesting a putative dual-binding mode capability for these interactions.

Although very closely related, the enzyme-borne and ScA type-I Docs do not display cross-specificity. Thus, Coh-contacting residues at positions 11 and 12 (numbering established considering the first Gly of each calcium-binding loop as residue 1), which are thought to be the major specificity determinants of all type-I Docs (3), are different in the ScA B and enzyme type-I Docs. Differences at these key residues may explain why there is a lack of cross-specificity between the type-I Doc interactions that modulate the binding of ScA onto ScA C or the cellulosomal enzymes onto ScA (10, 21). It is possible, however, that other elements of the two type-I Doc species confer their observed distinct specificities. Furthermore, although sequence symmetry of the duplicated Doc segments of *A. cellulolyticus* cellulosomal enzymes supports a potential dual-binding mode, this hypothesis remains to be tested.

Here, we have explored the mechanism by which the cellulosome assembles and is anchored onto the bacterial surface. We report the structure of a type-I complex in which a cellulosomal enzyme Doc (AcDocCel5) is bound to the sixth Coh of ScA (AcCohScaA6). Biochemical analysis guided by the crystal structure demonstrated that the enzyme-borne Doc modules of *A. cellulolyticus* interact with the Cohs of ScA through a dual-binding mode. Residues that determine the different specificities between the type-I Coh–Doc complexes of *A. cellulolyticus* were identified. The data informed the use of rational design to explore whether the binding surface alone confers ligand specificity. The data show that whereas the nature of the residues in the ligand-binding surface plays a major role in Coh recognition, the topology of the Doc modules also influences specificity.

Results and discussion

Previous studies have shown that the type-I Doc of *A. cellulolyticus* ScB binds specifically to the type-I Cohs of ScA, but not to those of ScA (10–12). Similarly, enzyme-borne type-I Docs specifically bind to the seven type-I Cohs of ScA (and to one in ScA D), but not to those of ScA C (21). In adherence to the canonical cellulosomal organizational framework, there are two distinct specificities within type-I Coh–Doc complexes of the *A. cellulolyticus* cellulosome system, one responsible for recruiting enzymes to ScA and a second one responsible for the anchoring of cellulosomes to the cell wall surface (Fig. 1). A recent study explored the structural and biochemical nature of one of these specificities by studying the interaction between the Doc of adaptor scaffoldin ScB (AcDocScB) and the third Coh of anchoring scaffoldin (AcCohScA C) (20). To dissect the mechanisms of enzyme–Doc ScA Coh recognition and whether the Doc scaffold contributes to ligand specificity, the structure of the Coh–Doc complex that recruits cellulosomal enzymes to ScA was investigated by solving the X-ray crystal structure of the sixth ScA C Coh (AcCohScA6) in complex with the Doc of a family 5 glycoside hydrolase (AcDocCel5). Established co-expression strategies for the production and purification of Coh–Doc complexes (5) had previously generated sufficient amounts of highly pure protein complexes that gave good quality crystals. This strategy, detailed below, was used to seek crystals of the ScA–enzyme Coh–Doc complexes, to investigate the interface residues that govern the specificity and mode of interaction.

Expression and crystallization of *A. cellulolyticus* Coh–Doc complexes

Analysis of the AcDocCel5 sequence revealed a high degree of internal symmetry, which suggested that this Doc contained two identical Coh-binding interfaces. Because a dual-binding mode implies that two different complex conformations will be present in solution, this would probably compromise protein crystallization. It is well established that in type-I Docs, residues at positions 11 and 12 of each one of the two duplicated segments play a key role in Coh recognition and act as specificity determinants (3). Thus, to force a single binding mode and therefore promote homogeneity in the final product, two Doc mutants were created. AcDocCel5 mutations used for the crys-
tallization experiments were designed to replace the putative recognition residues in relative positions 11 and 12 (Ser-15/Ile-16 and Ser-51/Leu-52) with those of the ScaB Doc (Ile-Asn), rather than the commonly applied alanine substitution. These amino acid changes were performed based on previous data that revealed a lack of cross-reaction between these two type-I Coh–Doc complexes (21). The sequence of the resulting Docs is displayed in Table 1. Recombinant plasmids encoding Doc-tagged AcCohScaA6–DocCel5 complexes were selected to produce highly pure Coh–Doc complexes for crystallization. Both AcCohScaA6–DocCel5 variants, containing the S15I/I16N (M1) or S51I/L52N (M2) amino acid substitutions, resulted in high-quality crystals.

Structure of AcCohScaA6 in complex with AcDocCel5

AcCohScaA6 type-I Coh in complex with its cognate Doc presents an elliptical structure comprising two β-sheets aligned in an elongated β-sandwich in a classic jellyroll fold. The two sheets are composed of β-strands 9, 1, 2, 7, and 4 on one face and β-strands 8, 3, 6, and 5 on the other face. Strands 1 and 9 align parallel to each other, thus completing the jellyroll topology, whereas the other β-strands are antiparallel (Fig. 2). β-Strand 8 is interrupted by a small β-hairpin, which spans residues Gly-118 to Pro-120, and there is a small α-helix N-terminal to β-strand 5. The two closest functionally relevant structural homologues to AcCohScaA6 were type-I Coh modules from Clostridia (Table S1).

Structure of AcDocCel5 in complex with AcCohScaA6

In both complexes, the AcDocCel5 Doc displays an identical structure that comprises two α-helices arranged in an antiparallel orientation ranging from residue Ile-15 to Leu-25 (helix 1) and from Ser-51 to Leu-61 (helix 3), respectively. These two helices comprise portions of the two classic Doc repeating segments, each containing a bound calcium ion in loops located at opposite ends of the module. The loop connecting these secondary structures contains a six-residue α-helix extending from Asp-37 to Gly-41 (helix 2). The overall structure of A. cellulolyticus AcDocCel5 is very similar to other type-I Docs (Table S1) The Ca²⁺ ion located at the Doc N terminus is coor-
dinated by the side chains of residues Asp-6, Asp-8, Asn-10, and Asp-17 (both the O\textsubscript{1} and O\textsubscript{2}), the latter belonging to the N-terminal \(\alpha\)-helix (helix 1) of this module. The octahedral geometry of the coordination of this Ca\textsuperscript{2+} ion is fulfilled by the main-chain carbonyl of Ser-12 and by a water molecule. The second Ca\textsuperscript{2+} site stabilizes the loop connecting the internal and C-terminal \(\alpha\)-helix (helices 2 and 3) of the Doc module. This Ca\textsuperscript{2+} ion is coordinated by the side chains of residues Asp-42, Asp-44, Asn-46, and Asp-53 (both the O\textsubscript{1} and O\textsubscript{2}) as well as by the carbonyl of Ser-48, with the octahedral geometry also completed by a water molecule. Thus, both Ca\textsuperscript{2+} comprise \(n\), \(n + 2\), \(n + 4\), \(n + 6\) (main-chain oxygen atom), \(n + 11\), and a water molecule completing the coordination pattern. There is a third calcium atom bound to AcDocCel5 which is coordinated by a loop between helix 1 and helix 2. This calcium is distant from the ligand-binding surface and thus probably plays a stabilizing role in protein structure. This third Ca\textsuperscript{2+}, not previously observed in Coh–Doc complexes, presents the typical octahedral geometry coordination through the side chains of Asp-31 and Asp-37, the main chain carbonyl oxygen atoms of Phe-32 and Ala-34, and by two water molecules (Fig. S1).

**A. cellulolyticus type-I CohScaA6-DocCel5 M1 and CohScaA6-DocCel5 M2 interfaces**

In the two AcCohScaA6-DocCel5 complexes, AcDocCel5 interacts with the 8-3-5-6 \(\beta\)-sheet of the AcCohScaA6 \(\beta\)-sandwich via helices 1 and 3. The Doc-contacting surface of AcCohScaA6 presents a predominantly flat rectangular shape, whose angles are slightly elevated toward the Doc and corre-
and 9 and the /H9252 spond to the loops between /H9252 loops between Values in parentheses are for the highest-resolution shell.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>AcCohScaA6-DocCel5 M1</th>
<th>AcCohScaA6-DocCel5 M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beamline</td>
<td>ESRF ID29</td>
<td>ESRF ID29</td>
</tr>
<tr>
<td>Space Group</td>
<td>P2,2,2;</td>
<td>P2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9763</td>
<td>0.9763</td>
</tr>
<tr>
<td>Unit-cell parameters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>46.54, 79.81, 112.16</td>
<td>30.49, 59.95, 51.26</td>
</tr>
<tr>
<td>a, β, γ (degrees)</td>
<td>90, 90, 90</td>
<td>90, 106.88, 90</td>
</tr>
<tr>
<td>V_m (Å³ Da⁻¹)</td>
<td>2.11</td>
<td>2.67</td>
</tr>
<tr>
<td>Solvent content (%)</td>
<td>42</td>
<td>53.98</td>
</tr>
<tr>
<td>Resolution limits (Å)</td>
<td>65.03–1.45 (1.502–1.45)</td>
<td>49.05–1.4 (1.45–1.4)</td>
</tr>
<tr>
<td>No. of observations</td>
<td>1,417,284 (78,998)</td>
<td>485,764 (47,950)</td>
</tr>
<tr>
<td>No. of unique observations</td>
<td>74,084 (7157)</td>
<td>33,519 (3358)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>19.0 (11.0)</td>
<td>14.2 (14.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.00 (95.16)</td>
<td>96.33 (94.58)</td>
</tr>
<tr>
<td>(I/σ(I))</td>
<td>18.9 (3.5)</td>
<td>7.0 (2.2)</td>
</tr>
<tr>
<td>CCb</td>
<td>0.998 (0.912)</td>
<td>0.995 (0.503)</td>
</tr>
<tr>
<td>Wilson B-factor</td>
<td>10.75</td>
<td>12.78</td>
</tr>
<tr>
<td>Rmerge</td>
<td>0.057 (0.203)</td>
<td>0.083 (0.599)</td>
</tr>
</tbody>
</table>

Structure refinement

| PDB accession code | 5NRK | 5NRM |

* Matthews coefficient.
* The correlation between intensities from a random half data set.
* $R_{merge} = \frac{\sum_{i} | I_i(hkl) - \langle I(hkl) \rangle |}{\sum_{i} I_i(hkl)}$, where $I_i(hkl)$ is the $i$th intensity measurement of reflection $hkl$, including symmetry-related reflections and $(I(hkl))$ is its average.
* $R_{work} = \frac{\sum_{i} | F_{o,i} - F_{c,i} |}{\sum_{i} F_{c,i}}$.
* $R_{free}$ as $R_{work}$ but summed over a 5% test set of reflections.

Dual-specificity Coh–Doc complex

A large network of hydrophobic interactions plays a key role in AcCohScaA6–DocCel5 M1 and M2 complex assembly (Table S2 and Fig. 3 (C and D)). The intermolecular interfaces also include several hydrogen bonds (Table 3 and Fig. 3 (A and B)). The DocCel5 residues at the complex interface located in helices 1 and 3 remain largely unchanged upon the 180° rotation of the Doc module over the CohScaA6 surface, reflecting the internal symmetry of the ScaB Doc (Figs. 3 and 4). Therefore, the interactions between the dominant Doc helix and the Coh are mainly established by residues Asn-50/14, Ser-51/15, Leu-52/Ile-16, Phe-54/18, Ala-55/19, and Arg-58/22 (on the AcCohScaA6–DocCel5 M1/M2 structures, respectively), whereas the main contacting residues in the non-dominant helix are Phe-18/54, Arg-22/58, and Leu-26/62 (again from M1/M2 structures, respectively). The side chains of Phe-18/54, Leu-26/62, Leu-52/16, Phe-54/18, and Ala-55/19 dominate the hydrophobic recognition of the Coh. Phe-18/54, Phe-54/18, and Ala-55/19 are at the core of the interaction, where, together with the aliphatic regions of Arg-58/22 and Ser-51/15, they establish numerous van der Waal contacts with β-sheets 3 and 6 of AcCohScaA6, including interactions with the side chains of residues Thr-35, Leu-80, and Tyr-82. Leu-26/62 and Leu-52/16 are located in opposing extremities of the interaction interface.
Dual-specificity Coh–Doc complex

Leu-26/62, together with the aliphatic region of Arg-22/58, contacts mainly with H9252-sheet 5 of AcCohScaA6, whereas Leu-52/16, together with the aliphatic region of Asn-50/14, establishes several van der Waals contacts with H9252-sheet 8 of AcCohScaA6. The hydrogen-bonding network established by the dominant helix of AcDocCel5 is supported by the interactions between the O γ atom of Ser-51/15 and AcCohScaA6 residues Thr-35 and Asn-37, between the nitrogen atom of Ser-51/15 and AcCohScaA6 Asn-37, and between Arg-58/22 (both Ne and NH2 atoms) and Tyr-82 of AcCohScaA6 (Table 3). In the non-dominant interacting helix, the hydrogen bonds are established between AcDocCel5 Arg-22/58 and Leu-26/62 and AcCohScaA6 Tyr-82 and Asn-68, respectively. In the AcCohScaA6–DocCel5 M1 structure, one extra hydrogen bond was observed between Doc Asn-28 and Coh Asn-68 (Table 3).

**Thermodynamics of the dual-binding mode**

Previous studies revealed that type-I Coh–Doc complexes of other cellulosome systems that display a dual-binding mode, such as those of *C. thermocellum* or *C. cellulolyticum*, have no preference for a particular binding orientation (17, 19). Thus, affinity between Cohs and Docs is similar whether the Doc
module binds its protein partner via the N-terminal or the C-terminal helix. To establish whether a similar mechanism operates during AcCohScaA6–DocCel5 recognition, the binding thermodynamics between AcCohScaA6 and the wildtype, M1, M2, and M1 + M2 variants of AcDocCel5 were determined using isothermal titration calorimetry (ITC). The data, presented in Table 4 and exemplified in Fig. S2, revealed a macro-dynamic architecture.

Developing a specificity hybrid A. cellulolyticus type-I Doc

Overall, the structure and mode of interaction of the AcCohScaA6–DocCel5 complex are very similar to those of the previously characterized AcCohScaC3–DocScaB heterodimer that displays a different specificity (20). Both AcDocCel5 and AcDocScaB possess the ability to bind their Coh partners in two different orientations, resulting in Coh–Doc complex configurations that are superimposable (root mean square deviation of 1.1 and 1.0 Å for the helix 1–dominated interaction and the helix 3–dominated interactions, respectively) (Fig. 4). Crucial interacting residues are generally located at the same relative positions, and, especially at the N-terminal Doc repeats, there is a high degree of sequence conservation between AcDocCel5 and AcDocScaB, exemplified by the key interacting residues Arg-22, Leu-26, and Arg-58. Despite these similarities, AcDocScaB displays a distinct Coh specificity when compared with AcDocCel5, whose binding properties should represent those of the remaining A. cellulolyticus type-I Doc modules that recruit enzymes to the cellulosome. Overall, these observations suggest that the two different type I Coh–Doc specificities identified in A. cellulolyticus are modulated by the nature of the contacting residues located at the protein interfaces, although differences in the topography of the two protein partners may also contribute to specificity.

To probe the hypothesis that differences in Coh–Doc specificity in A. cellulolyticus are modulated by the nature of the residues identified at the surface of the two protein modules, we attempted to alter the specificity of type-I Docs based on the structure of the two available protein–protein complexes. The aim was to create an AcDocScaB mutant capable of recognizing ScaA Cohs, an AcDocCel5 variant capable of binding to ScaC Cohs, and a hybrid type-I Doc that exhibited both specificities via two distinct Coh-binding interfaces. Structural alignment between AcDocCel5 and AcDocScaB (Fig. 5) revealed the main divergent residues between both Docs in the first repeat. This information was used to design an AcDocCel5 mutant where Asn-14, Ser-15, Ile-16, Phe-18, Ala-19, Met-21, and Asn-23 were replaced by the corresponding residues in AcDocScaB (Arg-14, Ile-15, Asn-16, Ala-18, Val-19, Ile-21, and Asp-23).
DocCel5, a consensus sequence based on 137 Dockerin sequences (data not shown). To confirm these results and further explore the thermodynamics of these interactions, ITC was carried out (Fig. S3) (13). The residues involved in the molecular interactions with the respective Coh partner are highlighted as follows: blue arrow for polar contacts and yellow circles for hydrophobic contacts. Residues involved in the coordination with the first and second calcium ions are marked with a blue star. Below the alignment, the Clustalo consensus symbols represent the position conservation status. The primary sequence background is colored according to the ALSCRIPT Concast convention, implemented in ALINE (22): red, identical residues; orange to blue, lowering color-ramped scale of conservation.

Figure 5. Multiple-sequence alignment of DocCel5 and DocScaB in a C terminus (helix 3)–dominated Coh–Doc interface. Based on the AcCohScaB6-DocCel5 M1 complex (PDB code 5NRK; top) and AcCohScaC3-DocScaB (PDB code 4UYP; bottom), a schematic representation of the secondary structure is shown. The residues involved in the molecular interactions with the respective Coh partner are highlighted as follows: blue arrow for polar contacts and yellow circles for hydrophobic contacts. Residues involved in the coordination with the first and second calcium ions are marked with a blue star. Below the alignment, the Clustalo consensus symbols represent the position conservation status. The primary sequence background is colored according to the ALSCRIPT Concast convention, implemented in ALINE (22): red, identical residues; orange to blue, lowering color-ramped scale of conservation.

Table 4
Thermodynamics of the interaction established between wildtype AcCohScaA6 and AcCohScaC3 and various mutant variants of AcDocGh5 and AcDocScaB

<table>
<thead>
<tr>
<th>Dockerin</th>
<th>K Values</th>
<th>∆G°</th>
<th>∆H</th>
<th>−T∆S</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcCohScaA6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AcDocGh5 WT</td>
<td>1.12E8 ± 3.17E7</td>
<td>−11.35</td>
<td>−21.91 ± 0.20</td>
<td>10.56</td>
<td>1.00</td>
</tr>
<tr>
<td>AcDocGh5M1 (S15I L16N)</td>
<td>6.72E7 ± 2.75E7</td>
<td>−11.04</td>
<td>−21.02 ± 0.52</td>
<td>9.97</td>
<td>1.00</td>
</tr>
<tr>
<td>AcDocGh5M2 (S51I L52N)</td>
<td>6.89E5 ± 4.89E4</td>
<td>−8.31</td>
<td>−20.63 ± 0.52</td>
<td>12.32</td>
<td>1.00</td>
</tr>
<tr>
<td>AcDocGh5M1+M2</td>
<td>NB†</td>
<td>−9.12</td>
<td>−21.75 ± 2.59</td>
<td>12.62</td>
<td>0.97</td>
</tr>
<tr>
<td>AcDocGh5 RINAVID</td>
<td>2.04E6 ± 4.77E5</td>
<td>−7.92</td>
<td>−18.17 ± 0.25</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>AcDocScaB WT</td>
<td>NB</td>
<td>−9.12</td>
<td>−21.75 ± 2.59</td>
<td>12.62</td>
<td>0.97</td>
</tr>
<tr>
<td>AcDocScaB M7</td>
<td>4.18E5 ± 4.73E4</td>
<td>−7.92</td>
<td>−18.17 ± 0.25</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>AcDocScaB M8</td>
<td>NB</td>
<td>−9.12</td>
<td>−21.75 ± 2.59</td>
<td>12.62</td>
<td>0.97</td>
</tr>
</tbody>
</table>

| AcCohScaC3   |          |     |     |      |    |
| AcDocGh5WT   | NB       | NB  | NB  | NB   | NB |
| AcDocGh5M1 (S15I L16N) | NB       | NB  | NB  | NB   | NB |
| AcDocGh5M2 (S51I L52N) | NB       | NB  | NB  | NB   | NB |
| AcDocGh5M1+M2 | NB       | NB  | NB  | NB   | NB |
| AcDocGh5RINAVID | NB       | NB  | NB  | NB   | NB |
| AcDocScaB WT | ND#      | ND  | ND  | ND   | ND |
| AcDocScaB M7 | 3.63E6 ± 1.68E6 | −9.24 | −4.937 ± 0.14 | −4.31 | 0.89 |
| AcDocScaB M8 | NB       | −9.24 | −4.937 ± 0.14 | −4.31 | 0.89 |

| AcDocCel5 (5NRK) |          |     |     |      |    |

Table 4: Thermodynamics of the interaction established between wildtype AcCohScaA6 and AcCohScaC3 and various mutant variants of AcDocGh5 and AcDocScaB. All thermodynamic parameters were determined at 308 K.
enough space to accommodate the mutations on both Doc repeats without steric clashes, especially between Phe-18 and Phe-54 (Fig. 6). Therefore, AcDocScaB M8 specificity is probably compromised through steric hindrance, which explains why this Doc derivative was unable to bind to AcCohScaA6. The inability of AcDocCel5 RINAVID to bind to the ScaB Coh may also reflect steric constraints imposed by the topography of the enzyme Doc. Thus, whereas the data show that the nature of the residues at the interface of Coh–Doc complexes is an important modulator of specificity, the topology of the Docs also contributes to ligand specificity through steric effects.

**Conclusions**

It is now well-established that type-I Coh–Doc interactions are essential to recruit cellulosomal enzymes onto primary scaffolds, which in turn are attached to the cell surface via a type-II Coh–Doc pair. In *A. cellulolyticus*, a second type-I Coh–Doc specificity is responsible for the attachment of an unusual adaptor scaffoldin ScaB to the bacterial cell surface. Previous work revealed that the type-I Coh–Doc complexes that recruit ScaB to the cell envelope present a dual-binding mode resulting from the presence of two identical Coh-binding faces, characteristic of the majority of cellulosomal type-I Coh–Doc complexes. Here, we reveal that the type-I Coh–Doc complexes that recruit enzymes to the cellulosome of *A. cellulolyticus* also present a dual-binding mode, suggesting that flexibility in the orientation of Coh recognition seems to be a general feature of type-I Doc modules, including those that recruit cellulosomes onto the cell surface. The structure of AcDocCel5 revealed an internal symmetry that supports the presence of two virtually identical Coh-binding faces. Due to the high degree of homology shared by the two different type-I Coh–Doc specificities discovered in *A. cellulolyticus*, an engineered Doc was designed with each ligand-binding site recognizing a different Coh. The data showed that although the residues in the two ligand-binding sites made a major contribution to Coh recognition, it should also be recognized that the topology of the Doc modules, through steric effects, also influenced Coh recognition. Thus, the evolution of Coh recognition by Doc modules requires modulation of both the ligand-binding surface and topology of the complete protein module.

**Experimental procedures**

**Gene synthesis and DNA cloning**

Docs are inherently unstable when produced in *E. coli*. To promote Doc stability, *A. cellulolyticus* DocCel5 of protein WP_010249057 (residues 502–573) was co-expressed *in vivo* with the sixth Coh of ScaA, AcCohScaA6 (AAF06064; residues 1472–1611). The immediate binding of AcDocCel5 to AcCohScaA6 is believed to confer the necessary Doc stabilization. The genes encoding the two proteins were designed with a codon usage optimized to maximize expression in *E. coli*, synthesized *in vitro* (NZYTEch Ltd., Lisbon, Portugal) and cloned into pET28a (Merck Millipore, Darmstadt, Germany) under the control of separate T7 promoters. The AcDocCel5-encoding gene was positioned at the 5′-end and the AcCohScaA6-encoding gene at the 3′-end of the artificial DNA. A T7 terminator sequence (to terminate transcription of the Doc gene) and a T7 promoter sequence (to control transcription of the Coh gene) were incorporated between the sequences of the two genes. This construct contained specifically tailored Nhel and NcoI recognition sites at the 5′-end and Xhol and Sall at the 3′-end to allow subcloning of the nucleic acid into pET-28a (Merck Millipore), such that the sequence encoding a six-residue His tag could be introduced either at the N terminus of the Doc (through digestion with Nhel and Sall, incorporating the additional sequence MGSSHHHHHHRSSGLVPRGSHMAS at the N terminus of the AcDocCel5) or at the C terminus of the AcCohScaA6 (by cutting with NcoI and Xhol, which incorporates the additional sequence LEHHHHHHH at the C terminus of the Coh). To block the dual-binding mode and promote the structural homogeneity required for protein crystallization, two different genes were synthesized, each with a distinct Doc mutant: mutant M1 with the S15I and I16N amino acid changes to the wild-type Doc and a T7 promoter sequence (to control transcription of the Coh gene) were incorporated between the sequences of the two genes. This construct contained specifically tailored Nhel and NcoI recognition sites at the 5′-end and Xhol and Sall at the 3′-end to allow subcloning of the nucleic acid into pET-28a (Merck Millipore), such that the sequence encoding a six-residue His tag could be introduced either at the N terminus of the Doc (through digestion with Nhel and Sall, incorporating the additional sequence MGSSHHHHHHRSSGLVPRGSHMAS at the N terminus of the AcDocCel5) or at the C terminus of the AcCohScaA6 (by cutting with NcoI and Xhol, which incorporates the additional sequence LEHHHHHHH at the C terminus of the Coh). To block the dual-binding mode and promote the structural homogeneity required for protein crystallization, two different genes were synthesized, each with a distinct Doc mutant: mutant M1 with the S15I and I16N amino acid changes and mutant M2 with the S51I and L52N replacements. These substitutions represent residue changes to amino acids present in type-I Docs of *A. cellulolyticus* that do not bind to ScaA but rather to the cell-surface anchoring scaffoldin ScaC. In addition, these residues are located, respectively, at the N-terminal...
**Dual-specificity Coh–Doc complex**

and C-terminal Coh recognition sites. Thus, as a result of this strategy, four pET28a plasmid derivatives were produced: pET28Dtc M1 and M2 with the engineered tag in the Doc and pET28Dct M1 and M2 where the engineered tag is attached to the Coh. The four plasmids were used to express AcCohScaA6–DocCel5 M1 and M2 complexes in *E. coli*. Recombinant AcDocCel5 and AcCohScaA6 primary sequences are presented in Table 1.

To produce recombinant *AcCohScaA6* and *AcDocCel5* individually, an ELISA-based system designed to probe Coh–Doc affinities that requires fusion with xylanase or carbohydrate-binding modules (CBMs) was selected, as it allows production of highly stable and functional Coh and Doc derivatives (23). Thus, sequences encoding each of the two modules were amplified from *A. cellulolyticus* genomic DNA by PCR, using NZY-Proof polymerase (NZYTech Ltd.) and the primers shown in Table S3. The M1 and M2 Doc mutants were amplified from the previously described synthesized DNA constructs. Following gel purification, the *AcDocCel5* encoding ampiclon was inserted into a xylanase-Doc cassette in pET9d plasmid after digestion with KpnI and BamHI and ligation with T4-ligase. The resulting expressed products consist of His-tagged *Xyn* fused to the xylanase T-6 from *Geobacillus stearothermophilus* at the N terminus of the polystyridine tag (Xyn AcDocCel5). The *AcCohScaA6*-encoding gene was cloned into CBM-Coh cassettes in pET28a after digestion with BamHI and Xhol restriction enzymes. This resulted in His-tagged *AcDocCel5* recombinant derivative fused to a CBM3a from the *C. thermocellum* scaffoldin CipA (CBM AcCohScaA6) (24). *Xyn AcDocScaB* and *CBM AcCohScaC3* were produced for a previous study, following the same approach (20).

For the specificity switch experiments, several *XynAcDocCel5* protein derivatives were produced using site-directed mutagenesis (Table S3). Each of the newly generated gene sequences was fully sequenced to verify that only the desired mutation accumulated in the nucleic acid chain. The *AcDocScaB* mutants (M7 and M8) were produced for a previous study using previously published primers (20).

**Expression and purification of recombinant proteins**

Preliminary expression screens revealed that when the polystyridine tag was located at the Doc N-terminal end of the *AcCohScaA6–DocCel5* complex, the expression levels of both Coh and Doc were higher. Tagging the Coh resulted in the accumulation of high levels of unbound Coh in the purification product, suggesting that the Coh was expressed at higher levels than the Docs. Consequently, the plasmid pET28Dtc was used to transform *E. coli* BL21 (DE3) cells to produce the *AcCohScaA6–DocCel5* complex in large quantities. Transformed *E. coli* were grown at 37 °C to an *A*₅₆₀ of 0.5. Recombinant protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside followed by incubation at 19 °C for 16 h. Cells were harvested by 15 min centrifugation at 5000 × *g* and resuspended in 20 ml of IMAC binding buffer (50 mM HEPES, pH 7.5, 10 mM imidazole, 1 mM NaCl, 5 mM CaCl₂). Cells were then disrupted by sonication, and the cell-free supernatant was recovered by a 30-min centrifugation at 15,000 × *g*. After loading the soluble fraction into a HisTrap™ nickel-charged Sepharose column (GE Healthcare), initial purification was carried out by IMAC in an FPLC system (GE Healthcare) using conventional protocols with a 35 mM imidazole wash and a 35–300 mM imidazole gradient. The buffer of all recovered fractions containing the purified Coh–Doc complex was exchanged into 50 mM HEPES, pH 7.5, containing 200 mM NaCl, 5 mM CaCl₂ using a PD-10 Sephadex G-25 M gel-filtration column (Amersham Pharmacia Biosciences). A further purification step by gel-filtration chromatography was performed by loading the samples onto a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml min⁻¹.

Fractions containing the purified complex were then concentrated with Amicon Ultra-15 centrifugal devices with a 10-kDa cutoff membrane (Millipore) and washed three times with molecular biology grade water (Sigma) containing 0.5 mM CaCl₂. The protein concentration was estimated in a NanoDrop 2000c spectrophotometer (Thermo Scientific) using a molar extinction coefficient (ε) of 8,940 M⁻¹ cm⁻¹. The final protein concentration was adjusted to 12 mg/ml for *XynAcDocCel5* M2 and 15 mg/ml for *XynAcDocCel5* M1, in molecular biology grade water containing 0.5 mM CaCl₂. The purity and molecular mass of the recombinant complex was confirmed by 14% (w/v) SDS-PAGE.

CBMCohs, XynDocs, and respective protein derivatives used in ITC and native PAGE experiments were expressed as described above and purified with IMAC by nickel-charged Sepharose His GraviTrap gravity-flow columns (GE Healthcare). After IMAC, the recombinant Coh and Docs were buffer-exchanged to 50 mM HEPES, pH 7.5, 0.5 mM CaCl₂, and 0.5 mM tris(2-carboxyethyl)phosphine using PD-10 Sephadex G-25 M gel filtration columns (GE Healthcare).

**NGE**

For the NGE experiments, each of the *XynAcDocCel5* and *XynAcDocScaB* variants, at a concentration of 30 μM, was incubated in the presence and absence of 30 μM CBM4CohScaA6 or CBM4CohScaC3 for 30 min at room temperature and separated on a 10% native polyacrylamide gel. Electrophoresis was carried out at room temperature. The gels were stained with Coomassie Blue. Complex formation was detected by the presence of an additional band displaying a lower electrophoretic mobility than the individual modules.

**Isothermal titration calorimetry**

All ITC experiments were carried out at 308 K. The purified *XynAcDocCel5*, *XynAcDocScaB*, CBM4CohScaA6, or CBM4CohScaC3 variants were diluted to the required concentrations and filtered using a 0.45-μm syringe filter (PALL). During titrations, the Doc constructs were stirred at 307 rpm in the reaction cell and titrated with 28 successive 10-μl injections of Coh at 220-s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single-site model (Microcal ORIGIN version 7.0, Microcal Software). The fitted data yielded the association constant (Kₐ) and the enthalpy of binding (∆H). Other thermodynamic parameters were calculated using the standard thermodynamic equation, ∆RTlnKₐ = ∆G = ∆H − T∆S.
**X-ray crystallography, structure determination, and refinement**

The crystallization conditions were set up using the sitting-drop vapor-diffusion method with an Oryx8 robotic nanodrop-dispensing system (Douglas Instruments (25)). The commercial kits Crystal Screen, Crystal Screen 2, PEG/Ion, and PEG/Ion 2 (Hampton Research), JCSG+ HT96 (Molecular Dimensions), and an in-house screen (80 factorial) were used for the screening. Precisely 0.7-μl drops of 15 and 12 mg/ml mg ml⁻¹ of AcCohScaA6–DocCel5 M1 and M2, respectively, were mixed with 0.7 μl of reservoir solution at room temperature per well containing 50 μl of the crystallization solution. The resulting plates were then stored at 292 K. Crystal formation was observed in four different conditions for AcCohScaA6–DocCel5 M1 and in one condition for AcCohScaA6–DocCel5 M2, within ~15 days (maximum dimension ~120 × 100 × 30 μm). All of the crystals were obtained from the initial screens. These crystals were cryo-protected with mother solution containing 20–30% glycerol or with 100% Paratone-N (Hampton Research) and flash-cooled in liquid nitrogen. Data were collected on beamline ID29 at the European Synchrotron Radiation Facility (Grenoble, France), using a PILATUS 6M detector (Dectris Ltd.) from crystals cooled to 100 K using a Cryostream (Oxford Cryosystems Ltd). iMOSFLM (26) was used for strategy calculation during data collection. All data sets were processed using iMOSFLM (26) and AIMLESS (27) from the CCP4 suite (Collaborative Computational Project, Number 4 (28)).

Data collection statistics are given in Table 2.

The best-diffracting AcCohScaA6–DocCel5 M1 crystals were the ones formed in the condition composed of 0.2 M sodium thiocyanate, 20% (w/v) PEG 3350, pH 6.9, and diffracted to a resolution of 1.57 Å. The crystals from the other three conditions did not diffract at all. The crystal belongs to the orthorhombic space group P2_12_12_1. The best-diffracting AcCohScaA6–DocCel5 M2 crystals were those formed under the condition composed by 0.2 M CaCl₂, 0.1 M HEPES, pH 7.5, and 28% PEG 400. The crystal belongs to the monoclinic space group P2₁, BALBES was used to carry out molecular replacement (29). The best solution for AcCohScaA6–DocCel5 M1 was found using the type-I Coh–Doc complex from C. cellulosypticum (PDB entries 2VN5 and 2VN6 with sequence identity of 36.9% with the Coh and 32.8% with the Doc (19)), producing at the end of the BALBES run an R_free of 35.7 and 40.6%, respectively, and a Q-factor of 0.719 after REFMAC5 refinement (30). An ARP/wARP (31) run after BALBES gave a model of 400 residues in six chains, with an estimated correctness of 99.9%. Two copies of the heterodimer AcCohScaA6–DocCel5 M1 complex are present in the asymmetric unit. This model was adjusted and refined using REFMAC5 and PDB REDO (32) interspersed with model adjustment in COOT to give the final structure (Protein Data Bank code 5NRM; Table 2). The final round of refinement was performed using the TLS/restrained refinement procedure, using each module as a single group. The root mean square deviation of bond lengths, bond angles, torsion angles, and other indicators were continuously monitored using validation tools in COOT and MOLPROBITY (33). A summary of the refinement statistics is shown in Table 2. The best solution for AcCohScaA6–DocCel5 M2 was found using the AcCohScaA6–DocCel5 M1 refined model. The refinement process was as described above for AcCohScaA6–DocCel5 M1 (Protein Data Bank code 5NRM; Table 2).

**References**

Acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. BMC Genomics 13, 210 CrossRef Medline


