

Novel *Clostridium thermocellum* Type I Cohesin-Dockerin Complexes Reveal a Single Binding Mode^{*§}

Received for publication, August 3, 2012, and in revised form, October 25, 2012. Published, JBC Papers in Press, November 1, 2012, DOI 10.1074/jbc.M112.407700

Joana L. A. Brás^{‡1,2}, Victor D. Alves^{‡1,3}, Ana Luísa Carvalho[§], Shabir Najmudin[‡], José A. M. Prates[‡], Luís M. A. Ferreira[‡], David N. Bolam[¶], Maria João Romão[§], Harry J. Gilbert[¶], and Carlos M. G. A. Fontes^{‡4}

From the [‡]Centro de Investigação Interdisciplinar em Sanidade Animal, Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal, the [§]REQUIMTE/CQFB, Departamento de Química, FCT-UNL, 2829–516 Caparica, Portugal, and the [¶]Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom

Background: In general, dockerins present two homologous cohesin-binding interfaces, which confer increased flexibility into cellulosomes.

Results: The structure of two novel Coh-Doc complexes reveals a dockerin single-binding mode.

Conclusion: Single-binding mode dockerins bind, preferentially, to cell surface cohesins.

Significance: The dual binding mode is a property of cellulosomal dockerins.

Protein-protein interactions play a pivotal role in a large number of biological processes exemplified by the assembly of the cellulosome. Integration of cellulosomal components occurs through the binding of type I cohesin modules located in a non-catalytic molecular scaffold to type I dockerin modules located at the C terminus of cellulosomal enzymes. The majority of type I dockerins display internal symmetry reflected by the presence of two essentially identical cohesin-binding surfaces. Here we report the crystal structures of two novel *Clostridium thermocellum* type I cohesin-dockerin complexes (CohOlpC-Doc124A and CohOlpA-Doc918). The data revealed that the two dockerins, Doc918 and Doc124A, are unusual because they lack the structural symmetry required to support a dual binding mode. Thus, in both cases, cohesin recognition is dominated by residues located at positions 11, 12, and 19 of one of the dockerin binding surfaces. The alternative binding mode is not possible (Doc918) or highly limited (Doc124A) because residues that assume the critical interacting positions, when dockerins are reoriented by 180°, make steric clashes with the cohesin. In common with a third dockerin (Doc258) that also presents a single binding mode, Doc124A directs the appended cellulase, Cell124A, to the surface of *C. thermocellum* and not to cellulosomes because it binds preferentially to type I cohesins located at the cell envelope. Although there are a few exceptions, such as Doc918 described here, these data suggest that there is considerable selective pressure for the evolution of a

dual binding mode in type I dockerins that direct enzymes into cellulosomes.

Biological nanomachines combining a range of complementary enzyme activities are critical to cellular function. Cellulosomes are one of nature's most elaborate and highly efficient multienzyme complexes that deconstruct cellulose and hemicellulose, two of the most abundant polymers on Earth (1–3). Thus, cellulosomes play a major role in carbon recycling and provide an opportunity to explore the largely untapped energy provided by plant biomass, by the bioenergy and bioprocessing sectors. It is now well established that the complex physical and chemical structure of plant cell walls restrict their access to hydrolytic enzymes. Aerobic microorganisms that utilize plant biomass as a significant nutrient express extensive repertoires of degradative enzymes, primarily glycoside hydrolases but also lyases and esterases, which attack the structural polysaccharides of the plant cell wall. In contrast, microbial anaerobes, due to environmental selective pressures, have a lower protein producing capacity and organize enzymes into cellulosomes, which enhance enzyme synergy and substrate targeting (see Refs. 1 and 2 for review).

The cellulosome of the thermophilic bacterium *Clostridium thermocellum* has been extensively explored (4, 5). It consists of a large non-catalytic multimodular protein, termed CipA, that contains nine tandemly repeated type I cohesins that recognize type I dockerins located in the cellulosomal enzymes (6, 7). Type I cohesins of CipA display a very high level of sequence identity. It was thus suggested that there is little discrimination by the dockerins and their protein receptors presented by the cellulosome scaffold (8). Primary scaffoldins, such as CipA, may also contain a C-terminal divergent type II dockerin that specifically recognizes type II cohesins located on the bacterium's envelope, thereby providing a mechanism for the cell surface attachment of cellulosomes (9). Thus, different cohesin-dock-

* This work was supported by Fundação para a Ciência e a Tecnologia Grants PEst-C/EQB/LA0006/2011, PTDC/BIA-PRO/69732/2006, PTDC/QUI-BIQ/100359/2008, and PTDC/BIA-PRO/103980/2008.

§ This article contains supplemental Table 1S and Figs. 1S–3S.

The atomic coordinates and structure factors (codes 4dh2 and 3ul4) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ Both authors contributed equally to this work.

² Supported by Fundação para a Ciência e a Tecnologia individual fellowship SFRH/BD/38667/2007.

³ To whom correspondence may be addressed. Tel.: 351-213652876; Fax: 351-213652889; E-mail: vdalves@fmv.utl.pt.

⁴ To whom correspondence may be addressed. Tel.: 351-213652876; Fax: 351-213652889; E-mail: cafontes@fmv.utl.pt.

erin (Coh-Doc)⁵ specificities (in *C. thermocellum* type I and type II) are responsible for the correct assembly of the multienzyme complex (type I) and its direct attachment to the organism (type II), respectively.

Structural studies on type I Coh-Doc complexes of *C. thermocellum* (10, 11) and *Clostridium cellulolyticum* (12), a mesophilic bacterium that produces a cellulosome analogous to the former microorganism, provided insights into the molecular determinants of protein-protein recognition that mediate the assembly of these protein complexes. Dockerins fold into two α -helices and EF-hand calcium-binding loop motifs, each corresponding to one of the two duplicated segments (10, 12). Thus, the structure of the N-terminal α -helix and EF-hand calcium-binding loop can be precisely superimposed over the equivalent structures at the C-terminal end, leading to an internal 2-fold symmetry in the dockerin molecule (11). The implications of this internal symmetry were realized when it was observed that type I dockerins present two cohesin binding surfaces because they can bind their cognate protein module either through the analogous N- or C-terminal α -helices (11). In *C. thermocellum* type I dockerins, residues that dominate the hydrogen bond network with cohesins are located at positions 11 and 12 of the calcium binding loop and are usually a Ser-Thr pair (10). When the dockerin is 180° reverse oriented, the equivalent residues (Ser⁴⁵ and Thr⁴⁶) in the C-terminal dockerin helix participate in cohesin recognition (11). The Ser-Thr dyad symmetry observed in *C. thermocellum* dockerins is replaced, in *C. cellulolyticum*, by hydrophobic residues, which accounts for the lack of affinity between protein partners from different species. 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.

Currently, it is unclear whether the dual-binding mode displayed by *C. thermocellum* and *C. cellulolyticum* dockerins is universal to all cellulosomal enzymes. The genome sequence of *C. thermocellum* ATCC 27405 encodes 72 polypeptides containing type I dockerin sequences. Alignment of the 72 dockerin sequences at the two ligand binding sites revealed a strong conservation of the amino acids that mediate cohesin recognition (particularly Ser¹¹, Thr¹², and a Lys-Arg motif at positions 18 and 19). Recently, we described the identification of four dockerins, of proteins Cthe_0435 (Cel124A), Cthe_0918, Cthe_0258, and Cthe_0624 (Cel9D-Cel44A), which deviate from the canonical *C. thermocellum* motifs at least in one of the cohesin binding interfaces (13). Here we describe the structure of two complexes in which two different type I cohesins are bound to these unusual dockerin modules. The data indicate that a cohort of *C. thermocellum* type I dockerins display a single binding mode. The possible biological significance for the single binding mode displayed by these dockerins is discussed.

TABLE 1

Primers used to obtain the genes encoding the cohesin and the dockerin derivatives used in the present study

Engineered restriction sites are shown in boldface type.

Clone	Sequence (5' → 3')	Direction
Doc124A	CTCCATATGTGGAATAAGGCAGTTATT	Forward
	CACGGATCCTTACGAATTGTAAGAGTTC	Reverse
Doc918	CTCCATATGGTTGTGCTTAATGGTGAC	Forward
	CACGGATCCCTATATAGTTATAAGTCC	Reverse
CohOlpC	CTCGCTAGCGTTGTGGCAATTCATG	Forward
	CACCTCGAGTTTTCATTTCAC	Reverse
CohOlpA	CTCGCTAGCCAAACAAACACCATGAA	Forward
	CACCTCGAGTGCCTCCGGAGCGGATGC	Reverse

EXPERIMENTAL PROCEDURES

Cloning and Expression

DNA encoding type I dockerins of Cthe_0435 (Cel124A, residues 31–112) and Cthe_0918 (residues 1146–1209) and type I cohesins of Cthe_0452 (OlpC, residues 108–258) and Cthe_3080 (OlpA, residues 30–177) were amplified by PCR from *C. thermocellum* genomic DNA using the thermostable DNA polymerase NZYDNAChange (NZYTech Ltd.) and primers described in Table 1. Genes encoding the type I dockerin modules of Cthe_0435 and Cthe_0918, here termed Doc124A and Doc918, respectively, were ligated into NdeI-BamHI-digested pET3a (Novagen). Genes encoding cohesin modules termed CohOlpC and CohOlpA, which derive from proteins Cthe_0452 and Cthe_3080, respectively, were ligated into NheI-XhoI-restricted pET21a (Novagen). Recombinant cohesins contained a C-terminal His₆ tag. To express the dockerin and the cohesin genes in the same plasmid, the recombinant pET3a derivative was digested with BglII and BamHI, to excise the dockerin gene under the control of the T7 promoter, which was subcloned into the BglII site of recombinant pET21a so that both genes were organized in tandem. Through this approach, it was possible to express both Doc124A and Coh452 and also Doc918 and Coh3080 in the same cells. Doc124A and Doc918 were also subcloned into pET32a vector (Merck) restricted with EcoRI and XhoI. Recombinant dockerins were expressed in fusion with thioredoxin to improve dockerin solubility and stability. OlpA and OlpC cohesins were also cloned into BglII- and EcoRI-digested pRSETa (Invitrogen). Mutant derivatives of both dockerins were synthesized (NZYTech Ltd.) with codon usage optimized for expression in *Escherichia coli* (Table 2). The synthesized genes contained engineered EcoRI and XhoI recognition sequences at the 5'- and 3'-ends, respectively, which were used for subsequent subcloning into pET-32a (Merck), as described above.

Protein Purification

Cohesin-Dockerin Complexes—The Coh-Doc complexes CohOlpC-Doc124A and CohOlpA-Doc918 were expressed in *E. coli* Tuner cells, grown at 37 °C to A₆₀₀ of 0.5. Recombinant protein expression was induced by adding isopropyl- β -D-thiogalactopyranoside to a final concentration of 0.2 mM and incubation for 16 h at 19 °C. The recombinant proteins were purified by immobilized metal ion affinity chromatography using Sepharose columns charged with nickel (HisTrapTM). Fractions containing the purified Coh-Doc complexes were buffer-exchanged, using PD-10 Sephadex G-25 M gel filtration col-

⁵ The abbreviations used are: Coh, cohesin; Doc, dockerin; ITC, isothermal titration calorimetry; r.m.s., root mean square; H-bond, hydrogen bond.

TABLE 2

Protein sequences of dockerins Doc124A and Doc918 and respective mutant derivatives

Amino acid changes are shown in boldface type.

	Protein sequence (N → C terminus)
Doc124A	WNKAVIGDVNADGVVNISDYVLMKRYILRIADFPADDDMWVGDVNGDNVINDIDCNLYLKRY LLHMIREFPKNSYNSA
Doc124A_m1	WNKAVIGDVNADGVVNISDYVLMMAAYILRIADFPADDDMWVGDVNGDNVINDIDCNLY LKRYLLHMIREFPKNSYNSA
Doc124A_m2	WNKAVIGDVNADGVVNISDYVLMMAAYILRIADFPADDDMWVGDVNGDNVINDIDC NYLAAYLLHMIREFPKNSYNSA
Doc124A_m3	WNKAVIGDVNADGVVNISDYVLMKRYILRIADFPADDDMWVGDVNGDNV INDIDCNLYLKRYLLHMIREFPKNSYNSA
Doc124A_m4	WNKAVIGDVNADGVVNISDYVLMKRYILRIADFPADDDMWVGDVNGDNV IIDCVLLKRYLLHMIREFPKNSYNSA
Doc124A_m5	WNKAVIGDVNADGVVNISDYVLMKRYILRIADFPADDDMWVGDVNGDNVINDIDCNLYLK YQQHMIREFPKNSYNSA
Doc124A_m6	WNKAVIGDVNADGVVNISDYVLMKRYQQRIADFPADDDMWVGDVNGDNVINDIDCNLYLK RYQQHMIREFPKNSYNSA
Doc918	VVLNGDLNRNGIVNDEDEYILLKKNYLLRGKLVLDLNVADVNDKGKVNSTDCFLKKNYILGLITI
Doc918_m1	VVLNGDLNRNGIVNDEDEYILLKKNYLLRGKLVLDLNVADVNDKGKVNDEDCFLKKNYILGLITI
Doc918_m2	VVLNGDLNRNGIVNSTDYILLKKNYLLRGKLVLDLNVADVNDKGKVNDEDCFLKKNYILGLITI
Doc918_m3	VVLNGDLNRNGIVNDEDEYILLKKNYLLRGKLVLDLNVADVNDKGKVNQQDCFLKKNYILGLITI

umns (Amersham Biosciences), into 20 mM Tris-HCl buffer, pH 8.0, containing 2 mM CaCl₂. A further purification step by anionic exchange chromatography was performed by using a column loaded with Source 30Q matrix and a gradient elution of 0–1 M NaCl (GE Healthcare). Fractions containing the purified complex were then concentrated with Amicon 10-kDa molecular mass centrifugal membranes and washed three times with 2 mM CaCl₂. The final protein concentration was adjusted to 21 g/liter in 2 mM CaCl₂ for CohOlpC-Doc124A complex and was 16 g/liter for CohOlpA-Doc918 complex.

Unbound Cohesins and Dockerins—Dockerins Doc124A, Doc918, and the respective mutant derivatives cloned in pET32a were expressed in *E. coli* Origami cells. CohOlpC and CohOlpA cloned in pRSETa vector were expressed in *E. coli* Tuner cells. Growth was performed at 37 °C to midexponential phase ($A_{600} = 0.5$) in Luria broth. Recombinant protein expression was induced with 1 mM (Origami) or 0.2 mM (Tuner) isopropyl- β -D-thiogalactopyranoside and incubation for 16 h at 19 °C. The recombinant proteins were purified by immobilized metal ion affinity chromatography as described above and buffer-exchanged into 50 mM Na-Hepes buffer, pH 7.5, containing 2 mM CaCl₂ and then subjected to gel filtration using a HiLoad 16/60 Superdex 75 column (GE Healthcare) at a flow rate of 1 ml/min.

Isothermal Titration Calorimetry (ITC)

ITC experiments were carried out essentially as described previously (11, 12), except that the titrations were at 55 °C, and proteins were in 50 mM Na-HEPES buffer, pH 7.5, containing 2 mM CaCl₂. During titration, the dockerin (40 μ M) was stirred at 300 rpm in the reaction cell, which was injected with 28 successive 10- μ l aliquots of ligand comprising cohesin (180 μ M) at 200-s intervals. Integrated heat effects, after correction for heats of dilution, were analyzed by non-linear regression using a single site-binding model (Microcal ORIGIN, version 5.0, Microcal Software). The fitted data yielded the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated by using the standard thermodynamic equation, $\Delta RT \ln K_a = \Delta G = \Delta H - T\Delta S$.

Crystallization and Data Collection

Protein crystals were obtained using the hanging drop, vapor diffusion method. CohOlpC-Doc124A complex crystals grew in 2 M ammonium sulfate, pH 4.6 (condition 32 of Crystal Screen HR2-110 from Hampton Research) in drops with 7 g/liter protein and were harvested after 5–7 weeks at 19 °C. CohOlpA-Doc918 complex crystals grew in 0.2 M lithium sulfate, 10% (w/v) PEG 8000+, 10% (w/v) PEG 1000, pH 7.5 (condition 14 of Clear Strategy Screen I MD1-14 from Molecular Dimensions) in drops with 16 g/liter protein and were harvested after 3–5 weeks at 19 °C. Crystals were cryocooled with paratone in liquid nitrogen prior to data collection at beamline ID14-2 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) at 100 K using an ADSC QUANTUM 4R CCD detector and at a wavelength of 0.9330 Å.

The two data sets were integrated using MOSFLM (14) and scaled with SCALA (15) from the CCP4 suite (16). CohOlpC-Doc124A crystals belong to the $P3_221$ space group, with cell constants $a = b = 90.76$ Å, $c = 135.07$ Å and diffracted beyond 1.75 Å resolution. Matthews (17) coefficient calculations suggested the presence of two molecules in the asymmetric unit (2.84 Å³/Da and 56.8% of solvent content). CohOlpA-Doc918 crystals belong to $I4_122$ space group, with cell constants $a = b = 130.05$ Å, $c = 70.19$ Å and diffracted beyond 1.95 Å resolution. Matthews coefficient calculations suggested the presence of one molecule in the asymmetric unit (3.0 Å³/Da and 59% of solvent content).

Structure Determination and Refinement

Structure determination of CohOlpC-Doc124A was based on two data sets that were processed in MOSFLM (14) merged and combined with SORTMTZ from the CCP4 suite (16) and scaled in SCALA (15). Phasing was performed by molecular replacement with the program BALBES (18) using a search model based on the Protein Data Bank structures 2ccl, 1aoh, 2vn6, 1nv8, and 1ixh, mainly related to cohesin and dockerin modules from *C. thermocellum* and *C. cellulolyticum* (11, 12). Density modification, together with non-crystallographic symmetry averaging, was done with the DM program from the

TABLE 3

Data collection and refinement statistics

	CohOlpC-Doc124A	CohOlpA-Doc918
Space group	P3 ₁ 21	I4 ₁ 22
Unit cell parameters $a = b, c$ (Å); α, β, γ (degrees)	90.76, 135.07; 90, 90, 120	130.05, 70.19; 90, 90, 90
Matthews parameter (Å ³ /Da)	2.81	3.0
Data collection statistics		
X-ray source	ESRF, ID14-2	ESRF, ID14-2
Wavelength (Å)	0.933	0.933
No. of unique reflections	66027	22341
Resolution limits (Å)	78.60–1.75 (1.80–1.75) ^a	91.96–1.95 (2.05–1.95)
Completeness (%)	99.9 (99.6)	100 (100)
Redundancy	14.9 (8.8)	14.4 (13.8)
Average $I/\sigma(I)$	18.7 (2.4)	23.3 (1.8)
R_{sym} (%)	0.090 (0.887)	0.080 (0.420)
Refinement statistics		
Resolution limits (Å)	39.30–1.80	32.79–1.95
R -work	0.181	0.175
R -free	0.215	0.206
No. of protein residues in the asymmetric unit	456	205
No. of water molecules in the asymmetric unit	461	178
No. of atoms in the asymmetric unit	7844	1824
r.m.s. deviation bond length (Å)	0.014	0.014
r.m.s. deviation bond angles (degrees)	1.296	1.314
Average temperature factor (Å ²)		
Protein main chain	26.4	28.2
Protein side chain	34.5	33.8
Water molecules	41.9	37.5
Sulfates	65.5	50.6
Calcium	24.5	20.0
Ramachandran plot		
Residues in most favored regions (%)	91.1	89.4
Residues in additionally allowed regions (%)	8.4	10.6
Residues in generously allowed regions (%)	0.5	0
Protein Data Bank code	4dh2	3ul4

^a Values in parentheses are for the highest resolution shell.

CCP4 suite (16). ARP/wARP (19) was used to automatically build the protein model. Model completion, editing, and initial validation were carried out in COOT (21). Initial restrained refinement of the molecular model was done using REFMAC version 5.5 (22), and water molecules were added/validated according to the following criteria: a compatible water-shaped peak ($F_o - F_c > \sim 3 \sigma$, $2F_o - F_c > \sim 1.5 \sigma$), whose center was within acceptable hydrogen bond distance to the closest protein atoms or other waters (~ 2.4 – 3.2 Å), and B -factors similar to neighboring atoms but less than 80 Å^2 . The final cycles of refinement were done with the program PHENIX.REFINE from the PHENIX suite (23). The two molecules in the asymmetric unit are arranged as a dimer of heterodimers, the later composed of chains A/B or C/D. All atoms in the protein could be properly assigned and refined, apart from a few initial (first 6 and 7 residues from chains A and C, respectively, and the first 6 residues from chains B and D) and final residues (last 6 and 7 residues from chains A and C, respectively, and the last 6 and 8 residues from chains B and D, respectively) in the polypeptide chains. The final model also includes 461 water molecules and four calcium ions. R -work and R -free converged to 18.1 and 21.5%, respectively. Model assessment and validation were carried out by PHENIX.POLYGON (24) and MOLPROBITY (25) from the PHENIX suite and PROCHECK (26). According to these programs, the final model contains 99.5% of the residues in most favored and allowed regions of the Ramachandran plot and 0.5% of the residues in generously allowed regions of the plot.

Structure determination of CohOlpA-Doc918 was similarly done by molecular replacement with BALBES using structures

with Protein Data Bank entries 2ccl and 2vn6 as models (11, 12). Density modification with non-crystallographic symmetry was done with the PARROT (27) program from the CCP4 suite. ARP/wARP was used to automatically build the protein model. Model completion, editing, and initial validation were also carried out in COOT. The dockerin start model had to be manually rebuilt due to a mistracing error by ARP/wARP, originating from the presence of dockerin's two duplicated segments that share a striking sequence similarity and strong structural conservation. Refinement procedures were done as described for CohOlpC-Doc124A. R -work and R -free converged to 17.5 and 20.6%, respectively. Two chains were found in the asymmetric unit, arranged as a dimer (A/B). Protein residues could be properly assigned and refined, apart from the first 5 and last 10 residues from chain A and the first 2 residues from chain B. The final model includes 224 water molecules and two calcium ions. Model assessment and validation using the above-mentioned tools produced a final model with 100% of the residues in the most favored and allowed regions of the Ramachandran plot. Data collection and refinement details data for the two complete structures are summarized in Table 3.

RESULTS AND DISCUSSION

Expression and Crystallization of Novel Coh-Doc Complexes

In a previous study (13), *C. thermocellum* type I dockerins were shown to bind to the nine type I cohesins of CipA and the single cohesin modules of the cell surface protein OlpA or OlpC (Fig. 1). The majority of *C. thermocellum* dockerins (68 of 72), exemplified by the well characterized dockerin of Xyn10B (10,

11), display a distinctive internal symmetry that is compatible with a dual binding mode. These dockerins display preferential recognition for OlpA and CipA cohesins. In contrast, two *C. thermocellum* dockerins, from the protein of unknown function, Cthe_0258, and the recently described cellulase, Cel124A (28), display a 2- and 10-fold preferential binding, respectively, to the cell envelope cohesin of OlpC (13). A third dockerin, of the bifunctional cellulase Cel9D-Cel44A, displays two cohesin-binding interfaces with different specificities; the dockerin can interact with *C. cellulolyticum* cohesins through the N-terminal interface and with *C. thermocellum* counterparts through the C-terminal binding site (13). A fourth dockerin, from the protein of unknown function Cthe_0918, binds equally well to CipA and OlpA and displays a lower affinity for OlpC. The primary sequences of these four dockerins lack the distinctive symmetry at the binding interfaces, which may explain, at least for dockerins of Cthe_0258, Cel124A and Cel9D-Cel44A, the observed differences in ligand specificity (13). The dockerin dual binding mode does not favor crystallization of protein complexes, and the usual approach used to study the Coh-Doc interaction involves the inactivation of one of the dockerin

cohesin-binding interfaces through site-directed mutagenesis (11, 12). Because the four unusual dockerins described above seem to present a single binding interface, wild-type proteins were used for these structural studies. Here, we have used established strategies for the production and purification of Coh-Doc complexes, which involve the co-expression of both proteins in *E. coli* cells (29).

Structure of Type I Coh-Doc Complexes

The structures of OlpA type I cohesin bound to the dockerin of the protein Cthe_0918 (CohOlpA-Doc918) and of the OlpC type I cohesin in complex with the dockerin of Cel124A (CohOlpC-Doc124A) were solved to 1.95 and 1.75 Å resolution, respectively (Fig. 2). In *C. thermocellum*, OlpA and OlpC cohesins are the only two type I cohesins that do not belong to CipA and show significant deviations in the putative residues that participate in dockerin recognition, when compared with the nine highly homologous cohesins of CipA (13) (for details, see supplemental Fig. 1S).

Structure of OlpA and OlpC Type I Cohesins—The 146-residue OlpA cohesin in complex with its cognate dockerin displays an elongated nine-stranded flattened β -sandwich structure, defined by two β -sheets (A and B) in a classical jelly roll topology (supplemental Fig. 2S) (10). β -Sheet A is composed of the β -strands (and respective residues) 4 (residues 51–58), 7 (residues 99–108), 2 (residues 22–30), 1 (residues 9–17), and 9 (residues 138–146), whereas β -sheet B includes β -strands 5 (residues 69–74), 6 (residues 79–85), 3 (residues 37–45), and 8 (residues 115–129). β -Strand 8 is the longest and partly integrates both β -sheets. β -Sheet B forms a distinctive planar plateau amid a molecule with a flattened and overall curved cylindrical shape (supplemental Fig. 2SC). The OlpC cohesin module with 159 amino acid residues exhibits the same type I cohesin architecture, where β -sheet A is composed of the same β -strands (and respective residues), 4 (residues 60–67), 7 (residues 108–117), 2 (residues 31–39), 1 (residues 15–27), and 9 (residues 149–161), whereas β -sheet B includes β -strands 5 (residues 78–83), 6 (residues 88–94), 3 (residues 48–54), and 8 (residues 129–142), the latter also contributing to both

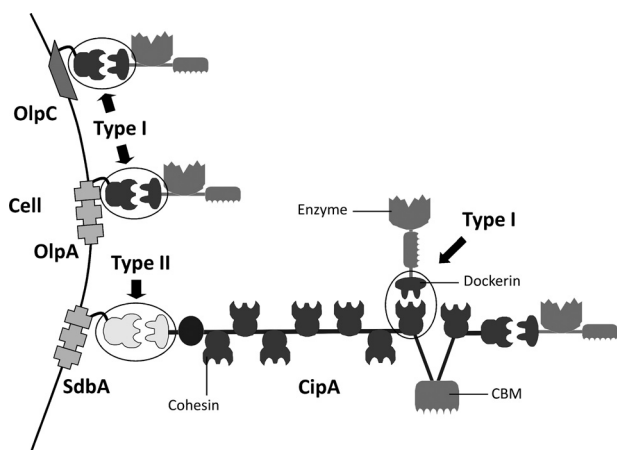


FIGURE 1. ***C. thermocellum* cellulosome.** *C. thermocellum* scaffoldin (CipA) contains nine type I cohesins and thus organizes a multienzyme complex that incorporates nine enzymes. The C-terminal type II dockerin of CipA binds specifically to type II cohesin modules found in cell surface proteins. Individual enzymes may also adhere directly to the bacterium cell envelope by binding the single type I cohesins found in OlpA and OlpC.

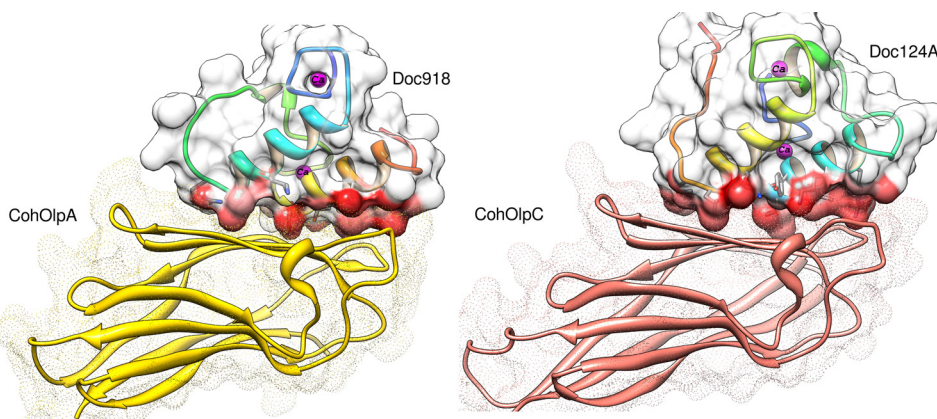


FIGURE 2. **Structure of the novel type I Coh-Doc complexes, CohOlpA-Doc918 and CohOlpC-Doc124A.** The dockerin structure is rainbow-colored (blue, N terminus; red, C terminus). Calcium ions are modeled as magenta spheres. The cohesin surface is depicted as dots, whereas the dockerin surface is solid white with the main contact surface area highlighted in red. CohOlpA-Doc918 shows a C terminus (helix-3)-dominated Coh-Doc interface, whereas CohOlpC-Doc124A reveals an N terminus (helix-1)-dominated Coh-Doc interface.

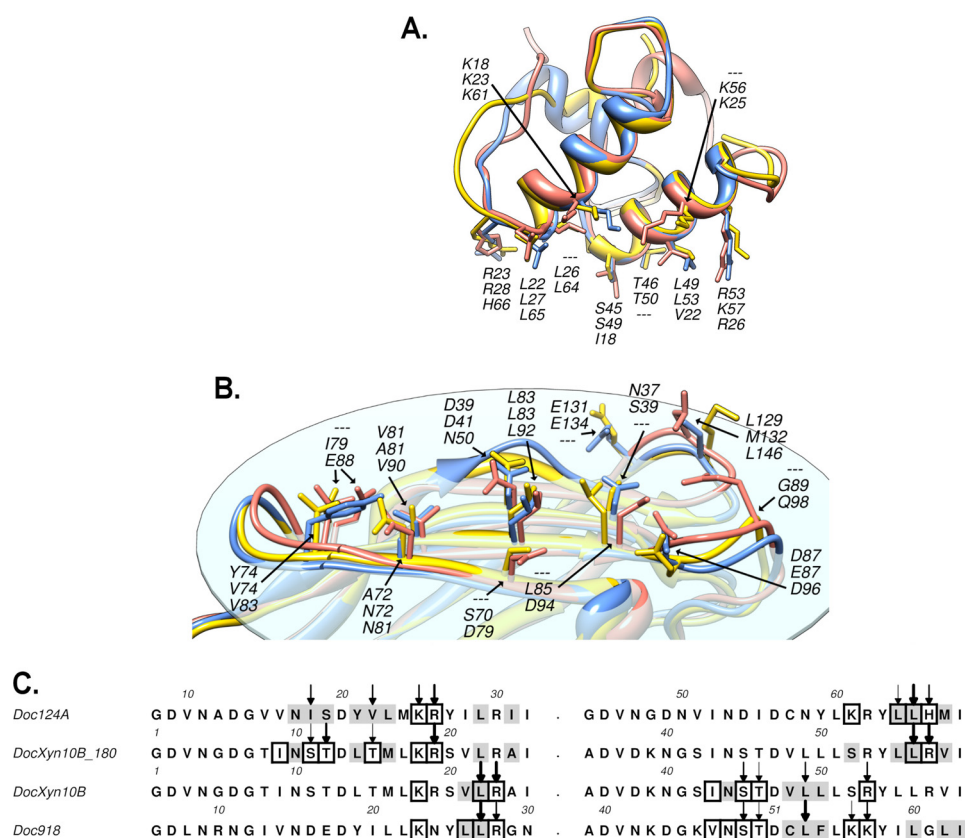


FIGURE 3. Structure superposition and important contact residues. *A*, dockerin superposition between DocXyn10B (blue), Doc918 (yellow), and a 180°-rotated Doc124A (salmon). Residues with a significant contribution to the Coh-Doc contact surface area are shown in stick representations and numbered according to the Protein Data Bank entries 1ohz, 3ul4, and 4dh2. *B*, cohesin superposition between CohCipA2 (blue), CohOlpa (yellow), and CohOlpc (salmon). Residues important to the Coh-Doc contact interface are shown above the β -sheet B plane, in stick representations and numbered according to the respective Protein Data Bank entries. *C*, dockerin sequence alignment and interacting residues. Residues with a significant contribution to the Coh-Doc contact surface area are marked with a top variable width small arrow. Residues involved in hydrophobic interactions are shown with a gray background, whereas a box highlights residues with polar interactions. DocXyn10B_180 denotes a 180° binding interface rotation.

β -sheets. Both cohesins contain a highly hydrophobic core (supplemental Fig. 2SB).

Both type I cohesin structures in complex with their respective protein partners reveal a striking similarity (supplemental Fig. 2SA). In comparison with the second cohesin of CipA (CohCipA2), a structure superposition with CohOlpa has an r.m.s. deviation of 1.04 Å (between 136 C_{α} pairs with a sequence identity of 35.3%) and 1.18 Å with CohOlpc (133 C_{α} pairs, 38.3% sequence identity). The two novel cohesins superpose with each other with an r.m.s. deviation of 1.14 Å (139 C_{α} pairs, 34.5% sequence identity). Noteworthy structural divergences occur between β -strands 4 and 5 (which include a small α -helix), where both CohOlpa and CohOlpc have a shorter loop than CohCipA2, and on the loop between β -strands 7 and 8 that, compared with CohCipA2, is slightly longer in CohOlpa and considerably larger in CohOlpc, increasing the main longitudinal axis length of these two proteins by around 2 and 10 Å, respectively (supplemental Fig. 2SA). The β -sheet B interface area, evaluated on the basis of its solvent-accessible area when in complex with its cognate dockerin (PDBePISA) (20), was 686 Å² for CohCipA2, 803 Å² for CohOlpa, and 729 Å² for CohOlpc.

Structure of Type I Dockerins—The structures of dockerins of Cthe_0918 and Cel124A, here termed Doc918 and Doc124A, respectively, are organized in two α -helices, arranged in an

antiparallel orientation (N-terminal or helix-1 and C-terminal or helix-3) connected through an extended loop displaying a small helix (helix-2) (Fig. 3). In Doc918, helix-1 is composed of residues 15–27, helix-2 extends between residues 35 and 39, and helix-3 extends from residue 48 to 60. In Doc124A, the respective residues are as follows: helix-1 (residues 17–29), helix-2 (residues 39–45), and helix-3 (residues 53–65). Helix-2 connects the other two helices, both of which provide the two putative cohesin binding interfaces. In fact, this region, limited by the distal end of helix-1 and the C terminus of helix-2, contains a large amount of the structural variability found among the core C_{α} trace of these dockerins. In Doc918, the region connecting the two helices is less structured than in DocXyn10B, presenting a single turn on its α -helix, similar to a type I *C. cellulolyticum* dockerin (12). The internal sequence duplication and nearly perfect 2-fold symmetry was quantified by an internal superposition between helix-1 and -3 within each structure. Doc918 shows an r.m.s. deviation of 0.57 Å for 23 C_{α} pairs, and in Doc124A, both segments overlap almost as well, with an r.m.s. deviation of 0.66 Å for 26 C_{α} pairs. Lack of conservation in the key contacting residues when the two putative binding surfaces are compared should prevent a dual binding mode, which is explored below. Both dockerins contain two Ca²⁺ ions coordinated by several residues in the canonical EF-hand calcium-binding loop. The coordination of the two cal-

Novel Type I Coh-Doc Complexes from *C. thermocellum*

cium ions is similar to the metal ions observed in the type I dockerins of *C. thermocellum* and *C. cellulolyticum* in complex with their cognate protein partners (supplemental Fig. 3SA).

Novel Type I Coh-Doc Complex Interfaces

In contrast with what was previously observed for other type I complexes, the dockerins described here in complex with their protein partners seem to present a single binding mode. Thus, in the CohOlpA-Doc918 complex, binding is dominated by the Doc918 C-terminal helix. In contrast, in the CohOlpC-Doc124A complex, binding is orchestrated by the dockerin N-terminal helix (Fig. 2). In these two novel Coh-Doc structures, the complex interface has a significant hydrophobic nature. Using the solvation free energy gain at complexation, calculated by PDBEPIA (Δ^iG in kcal/mol (20)), the CohOlpA-Doc918 interaction is more hydrophobic (−10.6 kcal/mol) than that of CohOlpC-Doc124A (−7.7 kcal/mol), which in turn exceeds the CohCipA-DocXyn10B value of −6.4 kcal/mol. However, the negative values upon binding are less significant than those of the highly hydrophobic *C. cellulolyticum* type I complex (Protein Data Bank code 2vn6) with −14.9 kcal/mol. These differences reflect the numerous hydrophobic residues, involved in the Coh-Doc complex interface, enumerated in detail in supplemental Table 1S and highlighted in supplemental Figs. 1S and 3S. Thus, the numbers of cohesin and dockerin hydrophobic residues implicated in the interface of the CohOlpA-Doc918 are greater than in the CohCipA2-DocXyn10B complex. Although the hydrophobic contact network of CohOlpC-Doc124A is also extensive, the hydrophobic residues that contribute to the heterodimer interface are contributed primarily by Doc124A.

The major hydrophobic contact residues located at the surface of cohesins CipA2, OlpA, and OlpC include a completely conserved leucine (Leu⁸³, Leu⁸³, and Leu⁹², respectively), which is assisted by upstream hydrophobic residues Val⁸¹, Ala⁸¹, and Val⁹⁰ and downstream by Ala⁸⁵, Leu⁸⁵, and a divergent Asp⁹⁴ in OlpC, respectively. Other important contributors correspond to Leu¹²⁹, Met¹³², and Leu¹⁴⁶, respectively. With respect to the dockerins Xyn10B- α 3/Xyn10B- α 1, 918, and 124A, the major hydrophobic contact residues are Leu²²/Leu⁵⁶, Leu²⁷, and Leu⁶⁵, respectively, at position 22 of the less interacting binding interface. In addition, in position 15 of the dominating interface, residues Leu⁴⁹/Thr¹⁵, Leu⁵³, and Val²² make a significant contribution to cohesin recognition. The above mentioned conserved leucine located at the surface of the three cohesins is part of an important hydrophobic pocket formed in CohCipA2 by Ala⁷², Tyr⁷⁴, Val⁸¹, and Leu⁸³, which is occupied by Leu²² or Leu⁵⁶ from DocXyn10B in the two possible binding modes, respectively. Using the same relative structural positioning order, for CohOlpA, we find Asn⁷², Ala⁸¹, and Leu⁸³, which accommodate Leu²⁷ from Doc918. As for CohOlpC, residues Asn⁸¹, Val⁹⁰, and Leu⁹² form a hydrophobic pocket that is occupied by the equivalent Doc124A residue, Leu⁶⁵, found in the opposite C-terminal interface.

The heterodimer interfaces are assisted by a network of direct and bridged hydrogen bonds and salt bridge interactions (described in detail in Table 4). Compared with DocXyn10B(α 3) in a similar C-terminal binding conformation

TABLE 4

Network of polar interactions in novel type I Coh-Doc complex interfaces

	Dockerin Doc124A	HB: H Bond SB: Salt bridge Wat: H ₂ O bridge	Cohesin OlpC
Helix-1	Lys B25 NZ	SB	Asp A94 OD2
	Lys B25 NZ	SB	Asp A96 OD1
	Arg B26 NE	HB	Gln A98 OE1
	Arg B26 NH2	HB	Gln A98 OE1
	Arg B26 NH1	HB	Phe A144 O
	Arg B26 NH2	HB	Phe A144 O
Helix-3	Lys B61 NZ	SB	Asp A79 OD2
	Leu B64 O	HB	Asn A50 ND2
	Leu B65 O	HB	Asn A81 ND2
	His B66 ND1	SB	Glu A88 OE1
	Dockerin Doc918	HB: H Bond SB: Salt bridge Wat: H ₂ O bridge	Cohesin OlpA
Helix-1	Lys B23 NZ	Wat A193	Ser A70 OG
	Leu B26 O	Wat A179	Val A82 O
	Leu B27 O	HB	Asn A72 ND2
	Arg B28 NH1	Wat A188	Asp A41 OD2
	Val B47 O	Wat A199	Asp A41 OD2
Helix-3	Asn B48 ND2	Wat A189	Glu A134 OE2
	Ser B49 N	HB	Asp A41 OD1
	Ser B49 OG	HB	Asp A41 OD1
	Ser B49 OG	HB	Ala A126 O
	Ser B49 OG	Wat A167	Ser A39 OG
	Thr B50 OG1	HB	Glu A134 OE2
	Thr B50 OG1	Wat A167	Ser A39 OG
	Lys B56 NZ	HB	Glu A87 O
	Lys B56 NZ	SB	Glu A87 OE2
	Lys B57 NZ	HB	Gly A89 O
	Lys B57 NZ	Wat A196	Asn A130 OD1
	Lys B57 NZ	Wat A207	Asn A130 OD1

(10), Doc918 reveals a more imbalanced distribution of polar bonds, favoring helix-3 residues. Although the Ser/Thr dyads of both complexes share an equivalent contribution, the main difference occurs at the Lys⁵⁶/Lys⁵⁷ pair of Doc918 that contribute with one salt bridge and two direct H-bonds, whereas in DocXyn10B(α 3), the equivalent Ser⁵² makes no polar bonds, and Arg⁵³ establishes a single salt bridge. Again, in comparison with the N-terminal bound DocXyn10B(α 1), Doc124A reveals some striking differences with respect to the relevant Ser-Thr pair, which is replaced by a divergent Ile¹⁸-Ser¹⁹ motif. In Doc124A the N-terminal binding face interacts with CohOlpC, through significant hydrophobic contacts. The only direct polar interactions mediated by helix-1 occur via positions 18 and 19 (Lys²⁵-Arg²⁶), through six direct bonds (two salt bridges from Lys²⁵ and four H-bonds from Arg²⁶) and a couple of water bridged H-bonds involving Ile¹⁸. In contrast, the N-terminal bound configuration of DocXyn10B reveals a hydrogen bond network around, and dominated by, the conserved Ser-Thr pair and also some involvement of residues 18 and 19. Also in contrast to DocXyn10B(α 1), Doc124A presents in the opposite interface (α 3) a stronger polar contribution participated in by four residues, Lys⁶¹ (one salt bridge), Leu⁶⁴ (one H-bond), Leu⁶⁵ (one H-bond), and His⁶⁶ (one salt bridge), whereas in DocXyn10B, Leu⁵⁶ and mainly Arg⁵⁷ make polar contacts with the CipA cohesin (Fig. 3, A and C; detailed contacts in supplemental Figs. 1S and 3S). Extending the comparison of Doc124A to *C. cellulolyticum* type I complex (2vn5/2vn6) in an analogous

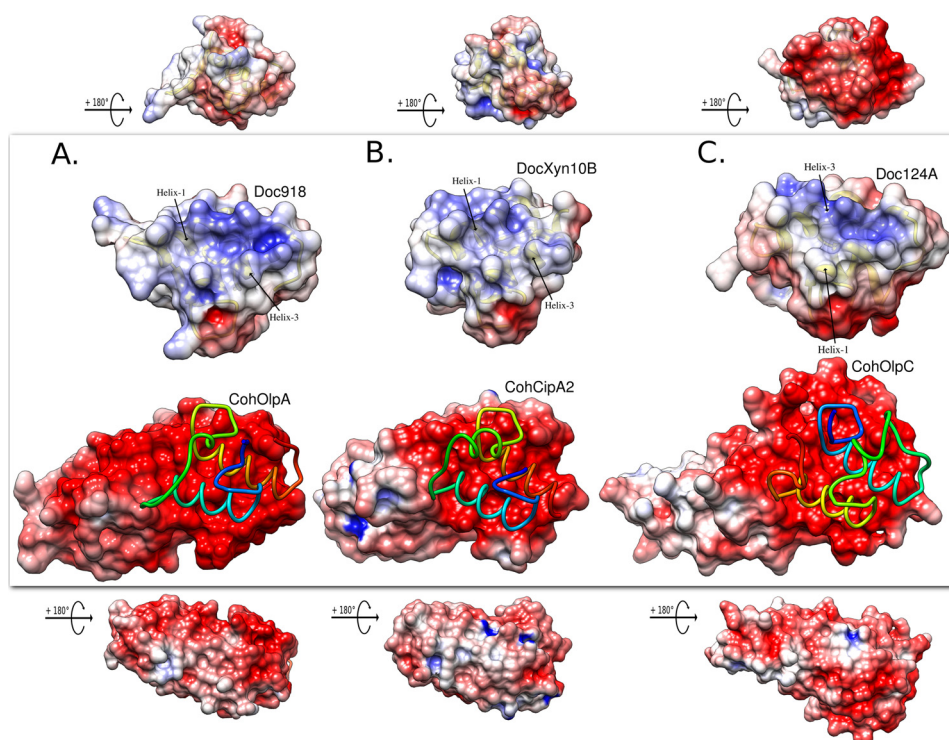


FIGURE 4. **Electrostatic surface potential for the Coh-Doc molecules.** In each panel, the *central image* shows the top cohesin binding plateau with a licorice model of the bound dockerin in N to C terminus *rainbow color ramped style*. Above, there is a view of the molecular dockerin binding surface. The *top and bottom images* are smaller scaled and *orthogonal representations*, respectively, of the dockerin and cohesin. A, CohOlpA and Doc918. B, CohCipA2 and DocXyn10B. C, CohOlpC and Doc124A. The electrostatic potential is contoured in UCSF Chimera from -6 (red) to $+6$ (blue) (arbitrary Chimera units).

binding conformation, the major difference consists of a much subdued polar interaction network found in the latter, especially at the $\alpha 3$ interface, where only positions 22 and 23 reveal direct contacts (12).

The cohesin-interacting residues can be grouped into three regions corresponding to β -strands $\beta 3$, β -strands $\beta 5/\beta 6$, and the loop between β -strands $\beta 8$ and $\beta 9$ (Fig. 3B; detailed contacts in supplemental Fig. 1S). Around the $\beta 3$ region, the important interactions are quite similar among CohCipA2/CohOlpA, because equivalent residues Asn³⁷/Ser³⁹ and Asp³⁹/Asp⁴¹, respectively, establish relevant polar contacts with the dockerin Ser/Thr pair. Conversely, the equivalent CohOlpC residue Ser⁴⁸ does not display any polar contacts, and Asn⁵⁰, equivalent to CohCipA2 Asp³⁹, establishes a single H-bond with the dockerin. In the $\beta 5/\beta 6$ cohesin region, notable differences between CohCipA2, CohOlpA, and CohOlpC occur, respectively, at Arg⁷⁷, Asp⁷⁷, and Asp⁸⁶ residues; Arg⁷⁷ makes an H-bond with its target dockerin, whereas the equivalent acidic residues of the other two cohesins are not implicated on the interface. In the $\beta 8$ -loop- $\beta 9$ region, the corresponding residues Asn¹²⁷, Asn¹³⁰, and Phe¹⁴⁴ in CohCipA2, OlpA, and OlpC, respectively, reveal some differences in their capacity to recognize the dockerin protein partner. In a helix-3-dominated binding, CohCipA2-Asn¹²⁷ does not exhibit any contacts with its dockerin, whereas CohOlpA-Asn¹³⁰ makes two bridged H-bonds. However, in a helix-1-dominated binding, CohCipA2 uses its Asn¹²⁷ to make two H-bonds with Doc-Arg¹⁹, whereas in CohOlpC, the backbone of Phe¹⁴⁴ establishes two H-bonds with Doc-Arg²⁶. In addition, in the Glu¹³¹/Glu¹³⁴/Pro¹⁴⁸ position of the cohesins, both acidic residues from CohCipA2 and

CohOlpA form an H-bond with the critical threonine found at position 12 of the dockerin, whereas CohOlpC-Pro¹⁴⁸ does not contribute to dockerin recognition.

Further analysis of the differences between the canonical type I cohesin and this work on cell-bound cellulosomal cohesins was based on the predicted negative hydrogen bond-accepting regions in an electrostatic surface potential evaluation using the Poisson-Boltzmann electrostatics calculation on the PDB2PQR server (30) and visualization of the results in UCSF Chimera (31) (Fig. 4). As reported previously (32), cohesins are strikingly negatively charged in the binding interface plateau, whereas dockerins present a suitable complementary positive-to-neutral surface. Compared with CohOlpC and CohCipA2, CohOlpA shows an elongated polar region that extends beyond the binding interface. As described for the type II cohesin of SdbA (32), the opposite cohesin surfaces in CohOlpC and CohOlpA are more positively/neutrally charged, which was suggested to be important to promote a tighter interaction of cell surface cohesins to the negatively charged peptidoglycan layer.

Analysis of the type I Coh-Doc interfaces provides significant insights into the previously described tight binding of Doc124A to CohOlpC, in comparison with the lower affinity displayed by this dockerin toward the cohesins of CipA and OlpA (13). The hydrophobic nature of Ile¹⁸ at the critical position 11 of the Doc124A interface, establishes a strong network of apolar contacts with CohOlpC, namely with Asn⁵⁰, Asn¹⁴⁰, and Cys¹⁴². These CohOlpC pivotal residues are replaced by an aspartate (position 50) and by small residues, namely glycine or alanine, at the other two positions in CohCipA and CohOlpA cohesins.

A)

Doc918 GDLNRNGIVNDEDYILLKNYLLRGN...ADV NKDGKVNSTDCLFLKKYILGLITI

Doc918_m1 GDLNRNGIVNDEDYILLKNYLLRGN...ADV NKDGKVNDEDCFLFLKNYILGLITI

Doc918_m2 GDLNRNGIVNSTDYILLKKYLLRGN...ADV NKDGKVNDEDCFLFLKNYILGLITI

Doc918_m3 GDLNRNGIVNDEDYILLKNYLLRGN...ADV NKDGKVNQQDCFLFLKKYILGLITI

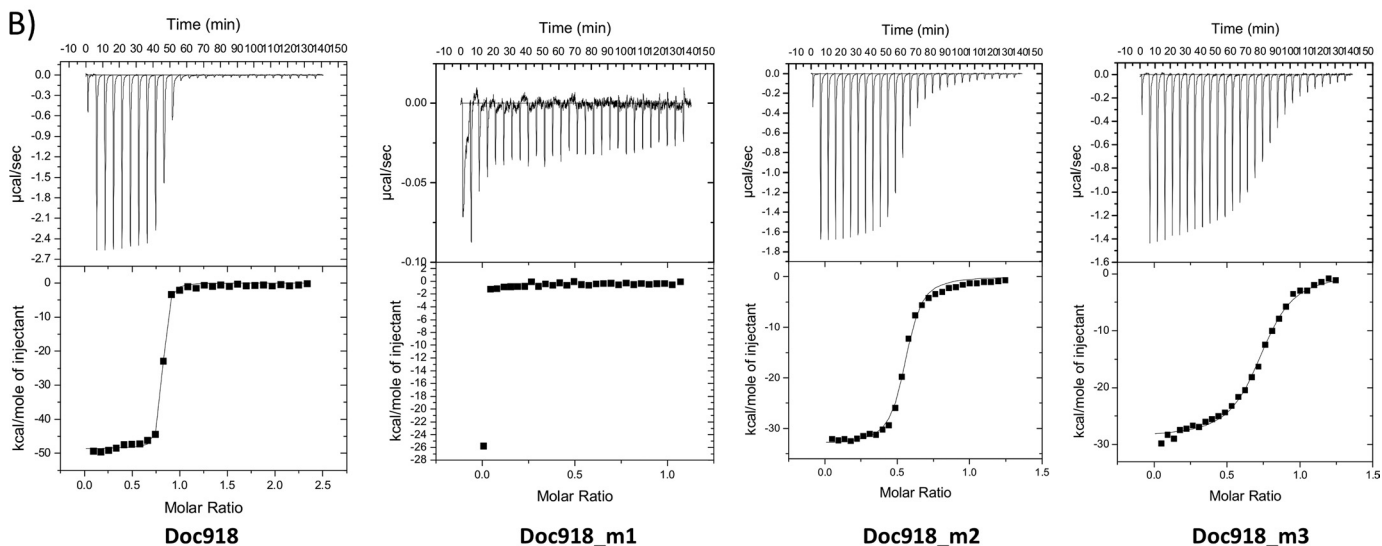


FIGURE 5. Alignment of Doc918 sequence and its mutant derivatives (A) and representative ITC experiments using the proteins generated in this study (B). In A, mutated residues are highlighted in gray. In B, the upper parts of each panel show the raw heats of binding, whereas the lower parts show the integrated heats after correction for heat dilution. The curve represents the best fit to a single-site binding model.

TABLE 5

Thermodynamics of the binding between *C. thermocellum* dockerins and their mutant derivatives and type I cohesins of OlpC and OlpA

Thermodynamic parameters were determined at 328.15 K. ND, values too low to be determined.

Cohesin	Dockerin	K_a M^{-1}	ΔG $kcal\ mol^{-1}$	ΔH $kcal\ mol^{-1}$	$T\Delta S$ $kcal\ mol^{-1}$
OlpC	Doc124A	$3.07E7 \pm 3.57E6$	-11.24 ± 0.18	-40.70 ± 0.18	-29.46
	Doc124A_m1	$6.91E4 \pm 1.10E4$	-7.16 ± 0.47	-9.08 ± 0.47	-1.92
	Doc124A_m2	ND	ND	ND	ND
	Doc124A_m3	$7.40E6 \pm 3.43E5$	-10.44 ± 0.09	-28.29 ± 0.09	-17.85
	Doc124A_m4	$6.12E7 \pm 7.20E6$	-11.84 ± 0.04	-28.90 ± 0.04	-17.06
	Doc124A_m5	$1.68E6 \pm 1.50E5$	-9.34 ± 0.34	-32.38 ± 0.34	-23.04
OlpA	Doc124A_m6	$1.50E5 \pm 1.39E4$	-7.77 ± 2.80	-39.86 ± 2.80	-32.09
	Doc918	$8.77E7 \pm 1.31E7$	-11.80 ± 0.22	-47.89 ± 0.22	-36.09
	Doc918_m1	ND	ND	ND	ND
	Doc918_m2	$6.07E6 \pm 5.42E5$	-10.18 ± 0.22	-30.49 ± 0.22	-20.31
	Doc918_m3	$2.47E6 \pm 3.52E5$	-9.62 ± 0.39	-27.24 ± 0.39	-17.62

The aspartate residue, equivalent to CohOlpC-Asn⁵⁰, found in both OlpA and CipA cohesins is also highly relevant for the recognition of typical type I dockerins because, together with Asn³⁷, it establishes conserved hydrogen bonds with the canonical serine residues usually found at position 11. In CohOlpC, the latter residues are replaced with Ser⁴⁸ and Asn⁵⁰, respectively, whose side chains were found more than 4 Å apart and thus presumably unavailable for H-bond formation. In addition, dockerin position 12 of the binding interface makes some relevant polar contacts with the mentioned residue of CohCipA2-Asn³⁷ and also Glu¹³¹. Again, in CohOlpC, the equivalent Ser⁴⁸ side chain orientation is unsuitable for those contacts, and the Pro¹⁴⁸, which substitutes Glu¹³¹, is manifestly non-reactive.

Probing the Importance of Contact Residues in Dockerins

To identify the dockerin residues that are involved in cohesin recognition, a mutagenesis study informed by the previously

described type I complex structures was implemented. Previous data suggest that the implications of single changes in dockerin activity may be relatively modest, so the strategy used here involved the change of particular groups of residues that are believed to play a cooperative role in cohesin recognition (10–13).

CohOlpA-Doc918 Complex—Site-directed mutagenesis and ITC data of the CohOlpA-Doc918 complex (Fig. 5 and Table 5), show that the replication of the relevant residue environment from dockerin helix-1 into the C-terminal helix-3, which dominates ligand recognition (mutant Doc918_m1: S49D, T50E, and K57N), precluded any binding, which reinforces a vital role for the Ser-Thr motif, similar to the canonical type I Coh-Doc interaction (10). The drastic decrease in affinity obtained with mutant Doc918_m3 (S49Q and T50Q) also supports a major role for the C-terminal dockerin Ser-Thr motif. The K57N

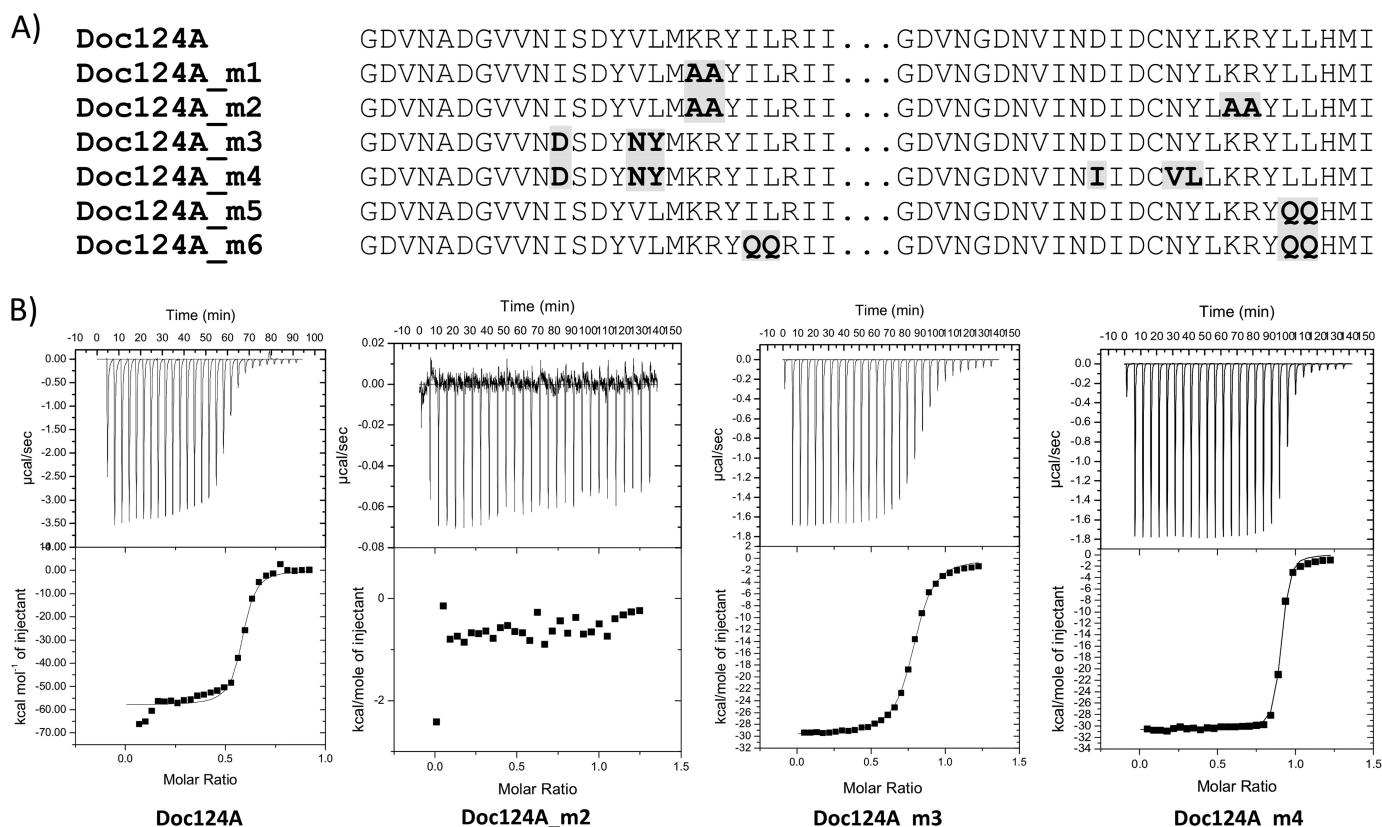


FIGURE 6. Alignment of Doc124A primary sequence and its mutant derivatives (A) and examples of ITC experiments using the proteins generated in this study (B). In A, mutated residues are highlighted in gray. In B, the upper parts of each panel show the raw heats of binding, whereas the lower parts are the integrated heats after correction for heat dilution. The curve represents the best fit to a single-site binding model.

mutation also emphasizes the relative importance of a basic residue, such as lysine or arginine (equivalent to Arg⁵³ in DocXyn10B), in this position for efficient binding.

The Doc918_m2 mutant design provided additional insights into the pivotal residues mediating cohesin recognition. Essentially, using the inactive dockerin, Doc918_m1, an attempt was made to force the alternate helix-1 binding mode. Thus, the non-functional N-terminal helix of Doc918_m1 was engineered to restore an N-terminal cohesin-binding interface by introducing the three pivotal residues (mutant Doc918_m2: D16S, E17T, and N24K) identified in helix-3 in the corresponding positions in helix-1. ITC data showed that, although with a 10-fold reduction in affinity, this strategy indeed allowed binding through helix-1. In CohOlpC-Doc124A and CohCipA2-DocXyn10B (*i.e.* the S45A/T46A mutant of DocXyn10B) (11), where helix-1 dominates the binding interface, there is a bulky positively charged residue in helix-3 (His⁶⁶ and Arg⁵⁷, respectively), which provides polar and hydrophobic interactions to the interface but which was replaced by a Gly⁶¹ in Doc918 when binding was engineered at the N-terminal face (Fig. 3). This divergent substitution could thus contribute to the reduced affinity displayed by the Doc918_m2 mutant. Overall, the data presented here confirm that Doc918 presents a single protein-binding interface that is dominated by the C-terminal helix and where Ser⁴⁹, Thr⁵⁰, and Lys⁵⁷ dominate cohesin recognition.

CohOlpC-Doc124A Complex—As described above, the Doc124A Lys²⁵-Arg²⁶ pair dominates the polar binding network with OlpC cohesin, whereas Lys⁶¹ makes an important

salt bridge with Asp⁷⁹ present at the surface of the cohesin. Thus, Doc124A mutants m1 and m2 were used to explore the importance of helix-1 Lys²⁵-Arg²⁶ and helix-3 Lys⁶¹-Arg⁶² pairs, by mutating them separately (m1) or simultaneously (m2) to Ala (Fig. 6 and Table 5). As expected, based on these multiple polar contacts, the lesion in helix-1 (m1) caused an ~400-fold decrease in affinity. In addition, the additive effect of mutating the two basic pairs at helix-1 and helix-3 simultaneously (m2) led to complete loss in cohesin recognition, confirming the importance of Lys⁶¹ in heterodimer formation. Thus, the basic pair at helix-1 plays a key role in cohesin recognition, and the massive reduction in affinity suggests a single binding mode for Doc124A. However, because the helix-3 Lys⁶¹-Arg⁶² pair is in a position symmetry-related to that of Lys²⁵-Arg²⁶ in helix-1, it is also possible that, following a 180° rotation of the dockerin, these latter residues could participate in a lower affinity cohesin recognition mediated by helix-3. Under these circumstances, the lower affinity of m1 would result from substitution of the critical Ile by an Asp at position 11 and by the loss of a putative Lys²⁵-mediated salt bridge at the other helix.

Data presented above suggest that Doc124A could eventually present two cohesin binding interfaces expressing different affinities. To explore this possibility, Doc124A Ile¹⁸, Val²², and Leu²³, which are part of the hydrophobic platform of the helix-1 binding interface, were mutated to replicate their symmetry-related counterparts in helix-3 (m3) (Fig. 6 and Table 5). The data revealed that these mutations lead to a reduction in the capacity of Doc124A to bind its cohesin partner. The

Doc124A_m4 mutant introduces into the m3 background, in which helix-1 binding is reduced, the mutations D54I, N58V, and Y59L, with the intention of promoting a reversal in binding through the C-terminal helix (Fig. 6 and Table 5). ITC results show an 8-fold increase in affinity over the m3 mutant, similar to the wild type dockerin, suggesting that although a dual binding mode is not feasible in the native form of Doc124A, in the m4 mutant, binding is probably dominated by the C-terminal interface. Thus, overall, the data suggest that Doc124A presents a single binding mode driven by helix-1.

The importance of the hydrophobic network established between Doc124A and OlpC was further explored in the mutant m5, which investigated the role of a second residue pair, Leu⁶⁴-Leu⁶⁵, in the interactions established with the cohesin (Fig. 6 and Table 5). As described above, the Doc124A dockerin presents a symmetry-related pair at helix-1, Ile²⁸-Leu²⁹, which could be involved in a similar interaction if binding was mediated by the helix-3 lower affinity interface. The importance of this pair was explored in m6. The knock-out of the Leu⁶⁴-Leu⁶⁵ helix-3 pair (m5) induced a 10-fold decrease in affinity, confirming the relevance of these residues in binding the cohesin when helix-1 is the dominant binding face. Indeed, it is reasonable to assume that the loss of Leu⁶⁴ and Leu⁶⁵ in m5 reorients the major binding face to helix-3. Consistent with this view is the further reduction in affinity by the concurrent mutation of the proximal helix-1 residue pair (m6).

CONCLUSIONS

The structure of two type I Coh-Doc complexes presented here revealed that unlike the large majority of *C. thermocellum* dockerins, the dockerins of cellulase Cel124A and of Cthe_0918 protein, presently of unknown function, display a single cohesin-binding surface. The structures of the two dockerins were solved in complex with the two unique cell surface type I cohesins of *C. thermocellum*, OlpA and OlpC, which direct plant cell wall hydrolytic enzymes directly to the cell surface. A recent study (33) revealed that cellulosomes act in synergy with enzymes located at the bacterium cell envelope, which include the abundant Cel124A endocellulase that targets cellulose crystalline-amorphous junctions. The fact that high quality crystals for both complexes were obtained using wild type dockerins was an initial good indication that these dockerins present essentially a single interacting surface. The structures of the two complexes revealed that the critical positions 11 and 12 of the dockerin non-interacting interface are occupied predominantly with acidic residues (Glu and Asp). Acid residues are not suitable for interacting with the highly negatively charged cohesin platform. Site-directed mutagenesis data demonstrate the importance of the Ser/Ile-Thr motif at positions 11 and 12 and the Lys/Lys-Arg pair at positions 18 and 19 in cohesin recognition. Inspection of the primary sequences of dockerins of Cthe_0258, which recognizes OlpC with higher affinity, and cellulase Cel9D-Cel44A, which binds both *C. thermocellum* and *C. cellulolyticum* cohesins, also revealed unsuitable substitutions at one of the dockerin binding faces, which should result in only one binding face capable of recognizing *C. thermocellum* cohesins. It is presently unclear why a subset of four dockerins, the two described here and those from Cthe_0258 and

Cel9D-Cel44A, have not evolved the dual binding mode characteristic of the other 68 *C. thermocellum* cellulosomal enzymes and extensively described for Xyn10B dockerin. Whereas the Cel124A dockerin directs the appended enzyme to the cell surface, because it binds predominantly to the OlpC cohesin, the dockerin of Cel9D-Cel44A is believed to present two cohesin binding interfaces with different cohesin specificities (the N-terminal face binds *C. cellulolyticum*-like cohesins, and the C-terminal interface binds their *C. thermocellum* counterparts). Thus, together, these data suggest that a dual binding mode is of primary importance for enzymes binding CipA, the multimodular cohesin scaffolding responsible for cellulosome assembly in *C. thermocellum*. An exception to this general rule is the dockerin of Cthe_0918, which recognizes CipA cohesins with higher affinity. The elucidation of the functional role of the protein domain appended to Cthe_0918 dockerin would help to clarify this issue. Nevertheless, the presence of two cohesin binding interfaces in dockerins integrated in multienzyme complexes may contribute to the capacity of the cellulosome to adjust its catalytic machinery to a highly insoluble and recalcitrant substrate.

Acknowledgment—We acknowledge the European Synchrotron Radiation Facility (Grenoble, France) (beamline ID14-2) for access and technical support during data collection.

REFERENCES

1. Fontes, C. M., and Gilbert, H. J. (2010) Cellulosomes. Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681
2. Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004) The cellulosomes. Multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–554
3. Bayer, E. A., Lamed, R., White, B. A., and Flint, H. J. (2008) From cellulosomes to cellulosomics. *Chem. Rev.* **8**, 364–377
4. Bayer, E. A., and Lamed, R. (1986) Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J. Bacteriol.* **167**, 828–836
5. Béguin, P., and Alzari, P. M. (1998) The cellulosome of *Clostridium thermocellum*. *Biochem. Soc. Trans.* **26**, 178–185
6. Salamitou, S., Raynaud, O., Lemaire, M., Coughlan, M., Béguin, P., and Aubert, J. P. (1994) Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. *J. Bacteriol.* **176**, 2822–2827
7. Tokatlidis, K., Salamitou, S., Béguin, P., Dhurjati, P., and Aubert, J. P. (1991) Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Lett.* **291**, 185–188
8. Salamitou, S., Tokatlidis, K., Béguin, P., and Aubert, J. P. (1992) Involvement of separate domains of the cellulosomal protein S1 of *Clostridium thermocellum* in binding to cellulose and in anchoring of catalytic subunits to the cellulosome. *FEBS Lett.* **304**, 89–92
9. Leibovitz, E., and Béguin, P. (1996) A new type of cohesin domain that specifically binds the dockerin domain of the *Clostridium thermocellum* cellulosome-integrating protein CipA. *J. Bacteriol.* **178**, 3077–3084
10. Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romão, M. J., and Fontes, C. M. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13809–13814
11. 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., Romão, M. J., Fontes, C. M., and Gilbert, H. J. (2007) Evidence for a dual binding mode of dockerin modules

- to cohesins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3089–3094
12. Pinheiro, B. A., Proctor, M. R., Martinez-Fleites, C., Prates, J. A., Money, V. A., Davies, G. J., Bayer, E. A., Fontes, C. M., Fierobe, H. P., and Gilbert, H. J. (2008) The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. *J. Biol. Chem.* **283**, 18422–18430
13. Pinheiro, B. A., Gilbert, H. J., Sakka, K., Sakka, K., Fernandes, V. O., Prates, J. A., Alves, V. D., Bolam, D. N., Ferreira, L. M., and Fontes, C. M. (2009) Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem. J.* **424**, 375–384
14. Leslie, A. G. W., and Powell, H. R. (2007) Processing Diffraction Data with MOSFLM. in *Evolving Methods for Macromolecular Crystallography* (Read, R., and Sussman, J. L., eds) pp. 41–51, NATO Series II, Vol. 245, Springer, Dordrecht, The Netherlands
15. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82
16. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Pottterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
17. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497
18. Long, F., Vagin, A. A., Young, P., and Murshudov, G. N. (2008) BALBES. A molecular-replacement pipeline. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 125–132
19. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for x-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179
20. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
21. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
22. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
23. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
24. Urzhumtseva, L., Afonine, P. V., Adams, P. D., and Urzhumtsev, A. (2009) Crystallographic model quality at a glance. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 297–300
25. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity. All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
26. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK. A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
27. Zhang, K. Y., Cowtan, K., and Main, P. (1997) Combining constraints for electron-density modification. *Methods Enzymol.* **277**, 53–64
28. Brás, J. L., Cartmell, A., Carvalho, A. L., Verze, G., Bayer, E. A., Vazana, Y., Correia, M. A., Prates, J. A., Ratnaparkhe, S., Boraston, A. B., Romão, M. J., Fontes, C. M., and Gilbert, H. J. (2011) Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5237–5242
29. Brás, J. L., Carvalho, A. L., Viegas, A., Najmudin, S., Alves, V. D., Prates, J. A., Ferreira, L. M., Romão, M. J., Gilbert, H. J., and Fontes, C. M. (2012) *Escherichia coli* expression, purification, crystallization, and structure determination of bacterial cohesin-dockerin complexes. *Methods Enzymol.* **510**, 395–415
30. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR. An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–667
31. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera. A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
32. Carvalho, A. L., Pires, V. M., Gloster, T. M., Turkenburg, J. P., Prates, J. A., Ferreira, L. M., Romão, 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. *J. Mol. Biol.* **349**, 909–915
33. Lu, Y., Zhang, Y. H., and Lynd, L. R. (2006) Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16165–16169

Novel *Clostridium thermocellum* Type I Cohesin-Dockerin Complexes Reveal a Single Binding Mode

Joana L. A. Brás, Victor D. Alves, Ana Luísa Carvalho, Shabir Najmudin, José A. M. Prates, Luís M. A. Ferreira, David N. Bolam, Maria João Romão, Harry J. Gilbert and Carlos M. G. A. Fontes

J. Biol. Chem. 2012, 287:44394-44405.

doi: 10.1074/jbc.M112.407700 originally published online November 1, 2012

Access the most updated version of this article at doi: [10.1074/jbc.M112.407700](https://doi.org/10.1074/jbc.M112.407700)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2012/11/12/M112.407700.DC1>

This article cites 33 references, 9 of which can be accessed free at

<http://www.jbc.org/content/287/53/44394.full.html#ref-list-1>