International Journal of Biological Macromolecules xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Structure-function studies can improve binding affinity of cohesin-dockerin interactions for multi-protein assemblies

Marlene Duarte [a,b], Victor D. Alves [a,b], Márcia Correia [c,d], Catarina Caseiro [a,b], Luís M.A. Ferreira [a,b], Maria João Romão [c,d], Ana Luísa Carvalho [c,d], Shabir Najmudin [e], Edward A. Bayer [f], Carlos M.G.A. Fontes [g], Pedro Bule [a,b,*]

[a] CIISA—Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, 1300-477 Lisbon, Portugal
[b] Associate Laboratory for Animal and Veterinary Sciences (AL4AnimalS), 1300-477 Lisbon, Portugal
[c] UCIBIO, Chemistry Department, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal
[d] Associate Laboratory i4HB – Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[e] Randall Centre for Cell and Molecular Biophysics, 3rd Floor New Hunt's House, Faculty of Life Sciences and Medicine, King's College London, London SE1 1UL, UK
[f] Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel
[g] NZYTech Genes & Enzymes, 1649-038 Lisbon, Portugal

ARTICLE INFO

Keywords:
Cohesin
Dockerin
Cellulosome
Protein complex
Biomass degradation
Carbohydrates

ABSTRACT

The cellulosome is an elaborate multi-enzyme structure secreted by many anaerobic microorganisms for the efficient degradation of lignocellulosic substrates. It is composed of multiple catalytic and non-catalytic components that are assembled through high-affinity protein-protein interactions between the enzyme-borne dockerin (Doc) modules and the repeated cohesin (Coh) modules present in primary scaffoldins. In some cellulosomes, primary scaffoldins can interact with adaptor and cell-anchoring scaffoldins to create structures of increasing complexity. The cellulosomal system of the ruminal bacterium, *Ruminococcus flavefaciens*, is one of the most intricate described to date. An unprecedent number of different Doc specificities results in an elaborate architecture, assembled exclusively through single-binding-mode type-III Coh-Doc interactions. However, a set of type-III Docs exhibits certain features associated with the classic dual-binding mode Coh-Doc interaction. Here, the structure of the adaptor scaffoldin-borne ScaH Doc in complex with the Coh from anchoring scaffoldin ScaE is described. This complex, unlike previously described type-III interactions in *R. flavefaciens*, was found to interact in a dual-binding mode. The key residues determining Coh recognition were also identified. This information was used to perform structure-informed protein engineering to change the electrostatic profile of the binding surface and to improve the affinity between the two modules. The results show that the nature of the residues in the ligand-binding surface plays a major role in Coh recognition and that Coh-Doc affinity can be manipulated through rational design, a key feature for the creation of designer cellulosomes or other affinity-based technologies using tailored Coh-Doc interactions.

1. Introduction

Plant cell wall hydrolysis is a key process in the global carbon cycle and a major step in the production of many value-added products, including biofuels, materials and chemicals. Due to its recalcitrant nature, the deconstruction of lignocellulosic biomass requires the combined effort of multiple enzymes, collectively termed carbohydrate-active enzymes, or CAZymes [1,2]. While many cellulolytic microorganisms have adopted different multi-enzyme strategies for biomass deconstruction, the cellulosome is likely the most sophisticated [3]. These are highly efficient multicomponent cellulolytic complexes produced by anaerobic bacteria, which centralize the action of multiple CAZymes onto a single mega-Dalton structure [4–7]. The proximity effect creates enzymatic synergy that increases degradation efficiency [8,9]. The assembly of these complex nanomachines is mediated by high-affinity calcium-dependent interactions between two conserved

* Corresponding author at: CIISA—Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, 1300-477 Lisbon, Portugal.
E-mail address: pedrobule@fmv.ulisboa.pt (P. Bule).

https://doi.org/10.1016/j.ijbiomac.2022.10.102
Received 5 August 2022; Received in revised form 28 September 2022; Accepted 11 October 2022
Available online 15 October 2022

protein modules, called cohesins (Coh) and dockerins (Doc) [10,11]. Cohs are repeated modular components in non-catalytic scaffold proteins, termed scaffoldins, whereas Docs are mostly present at the C-terminus of the modular cellulosomal CAZymes, with the tenacious Coh-Doc interaction serving to attach them securely onto the scaffoldin proteins. Scaffoldins can also contain a Doc module, which they use for cell-wall tethering or for articulation with other Coh-bearing scaffoldins in more complex assemblies. Some cellulosome components also incorporate carbohydrate-binding modules (CBMs) for substrate attachment, which further increase the proximity effect promoted by the cellulosome [12,13].

Each microorganism species presents its own cellulosomal molecular arrangement with different levels of complexity [12,14–17]. One of the most intricate and potentially versatile cellulosomal systems is that of the rumen bacterium *Ruminococcus flavefaciens* strain FD-1 [18–20]. This mesophilic cellulolytic bacterium possesses an unusually complex scaffoldin-encoding gene cluster, which encodes multiple scaffoldin proteins that can potentially interact with over 220 identified Doc-containing proteins [19,21,22]. This impressive array of Docs has been organized into 6 groups, based on primary sequence homology, a classification which was shown to translate into function [22,23]. As such, each group displays a distinct specificity profile, which is conserved along all members. This results in an especially intricate architecture, capable of centralizing the action of up to 14 enzymes in a single unit (Fig. 1). Most of these enzymes are incorporated into the cellulosome through the interaction of their group-1 Docs with the two Cohs of primary scaffoldin ScaA or Cohs 1 to 4 of the main adaptor scaffoldin ScaB [11]. In turn, ScaA interacts with Cohs 5 to 9 of ScaB, meaning that 5 ScaA units can bind to a single ScaB [22,24]. Another subset of enzymes, mostly hemicellulases, can also be indirectly recruited by means of their interaction with monovalent adaptor scaffoldin ScaC. ScaC possesses a group-1 Doc, that targets ScaA and B, and a single Coh capable of recognizing group-3 and -6 Docs, thereby expanding the array of enzymes that can integrate into the cellulosome [25]. This assembly is then tethered onto the cell wall of the bacterium by interaction of the C-terminal X-module–Doc dyad (XDoc) of ScaB, classified as a group-4 Doc, with the Coh of cell wall-bound ScaE [26]. Curiously, *R. flavefaciens* encodes many other proteins that can potentially interact with the bacterial surface through ScaE recognition, namely those bearing group-4 and group-2 Docs [22]. While group-2 Docs are functional truncated modules bearing a single Ca^{2+}-binding repeat, group-4 Docs display a distinctive symmetrical nature, unlike the other type-III Docs from *R. flavefaciens* [22]. Docs that exhibit symmetrical duplicated segments have been observed extensively in other bacterial species, such as *Clostridium thermocellum* (aka *Ruminiclostridium thermocellum*, *Hungateiclostridium thermocellum* or *Acetivibrio thermocellus*) [27,28] or *Bacteroides cellulosolvens* (aka *Pseudobacteroides cellulosolvens*) [29], and were shown to be able to bind their cognate

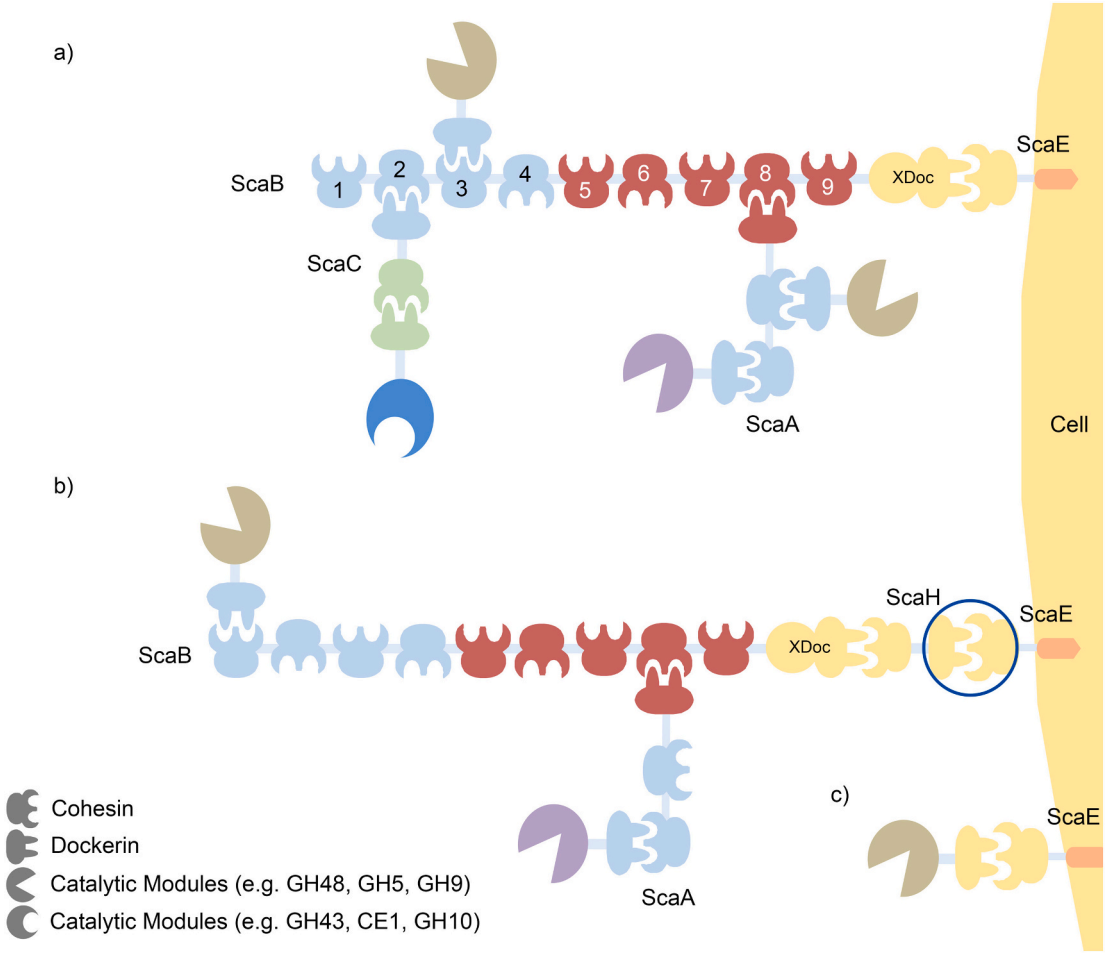

Fig. 1. Schematic representation of the *R. flavefaciens* strain FD-1 cellulosome. (a) Scaffoldin ScaB mediates the incorporation of several enzymes into the cellulosome, either directly through Cohs 1–4 or indirectly through interaction of Cohs 5–9 with the Docs of adaptor scaffoldin ScaC and ScaA. The assembly is tethered onto the cell wall by interaction of the C-terminal XDoc of ScaB with CohScaE. (b) Adaptor scaffoldin ScaH can insert itself between ScaB and ScaE, thereby introducing the dual-binding mode mechanism into the *R. flavefaciens* cellulosome. The interaction studied in the present work is highlighted with a blue circle. (c) Direct incorporation of non-cellulolytic proteins to the bacterial surface is mediated through ScaE recognition. Compatible Docs and Cohs are represented in the same color.

Cohs in either of two opposite, 180° related, orientations. This is referred to as the dual-binding mode and is thought to promote conformational flexibility, avoiding steric hindrance between the several cellulosomal components [7,11,22]. Interestingly, the best-characterized group-4 Docs are the XDocs of ScaB and CttA (a double-CBM-bearing protein) which, although also possessing duplicated segments, have their overall symmetry disrupted by three characteristic inserts that serve to physically support the upstream X module [26]. Nonetheless, most group-4 Docs do not possess an X module, and, as such, their marked conservation between each of the Ca^{2+}-binding repeats is not disrupted by the presence of the XDoc inserts. Thus, it is likely that these Docs can support a dual-binding mode, as previously suggested [22]. Within the cellulosomal system of *R. flavefaciens*, which, as previously demonstrated, is completely assembled through single-binding mode interactions, the dual-binding mode would then be a unique feature of group-4 Docs lacking an X-module.

To confirm and better understand the molecular mechanism governing the potential dual-binding mode in *R. flavefaciens*, an extensive structural and biochemical characterization was performed on a Coh-Doc complex of ScaE Coh (CohScaE) with the group-4 Doc from ScaH (DocScaH), a monovalent adaptor scaffoldin. ScaH was chosen because its Coh (CohScaH), like CohScaE, is also capable of recognizing group-4 Docs [22,27,29]. This means that, having both a group-4 Doc-binding Coh and a group-4 Doc not appended to an X module, ScaH can serve to introduce the dual-binding mode into the *R. flavefaciens*' cellulosome, by its positioning between the primary scaffoldin ScaB and the cell-surface anchoring scaffoldin ScaE (Fig. 1). This potential selective incorporation of the dual-binding mode into a cellulosomal system has not been previously observed in any other species. A biochemical analysis, guided by the crystal structure of the complex, demonstrated that DocScaH interacts with the cell-surface scaffoldin CohScaE through a dual-binding mode. Residues that determine Coh recognition were also identified. The data obtained were then used to explore whether the affinity between the two modules could be improved by manipulation of the Doc's electrostatic profile at the binding surface. The results revealed that the nature of the residues in the ligand-binding surface plays a major role in Coh recognition and that the Coh-Doc interaction presents a versatile platform for the rational design of improved high-affinity technologies, with potential applications in multiple industrial and research fields.

2. Results

2.1. Primary structure of R. flavefaciens group-4 Docs

Initially, only three XDoc-bearing proteins were thought to interact with the Coh of cell-wall attached scaffoldin ScaE [21,30]. Later, through a combination of orthogonal techniques, including microarrays, ELISA and *in-vivo* co-expression [22], it was revealed that the classification of *R. flavefaciens* Docs into homology groups is translated into function. As such, most group-4 Docs, including subgroups a and b, share similar binding preferences, as they specifically recognize the Cohs of anchoring scaffoldin ScaE and also of the adaptor scaffoldins ScaH and ScaG. This means that all group-4 Doc-bearing proteins can be tethered to the bacterial cell envelope, either directly, *via* recognition of ScaE, or indirectly through ScaH.

Presently, there are 46 Docs in group 4, 40 belonging to subgroup 4a and 6 to subgroup 4b, a subclassification based on primary sequence homology. Primary structure analysis revealed that there are key residues conserved along all group members, namely at the canonical Coh-recognition positions 11 and 12 (Fig. S1). Curiously, two sets of group-4 Docs, independent of the a and b subgrouping, were identified regarding the nature of the 11/12 dyad; one possesses a Gly-Arg pair (sometimes Ala-Arg on the second repeat) and contains most of the Docs, including DocScaH, while the other comprises an Ala-Val pair at the 11/12 positions. However, as previously mentioned, specificity is maintained across all members of the group, meaning that the 11/12 dyad, although important for Coh recognition, does not dictate specificity by itself. The regions flanking the canonical Coh-recognition pairs are also highly conserved along the group, as well as the majority of helix 1 and 3 residues. The most variable region is that connecting the two Coh-contacting helices, which does not participate in binding. Notably, not only are the two helices' sequences highly conserved in all group-4 Docs, but also conserved between themselves. That is to say that, unlike other type-III Docs from *R. flavefaciens*, group-4 Docs seem to display a distinctive symmetrical nature, much like that found in many type-I and -II Docs from other species [22,25,26,31].

Interestingly, most of the group-4-bearing proteins do not seem to be directly associated with carbohydrate degradation. A sequence homology search and alignment analysis of 43 *R. flavefaciens* group-4 Doc-containing ORFs, revealed a very diverse pool of putative functions, including transglutaminases, proteases, endoglucanases, lipases, as well as several proteins of unknown function and several containing leucine-rich repeats. Some glycoside hydrolase families were also identified, namely GH3 and GH24. Family GH3 comprises enzymes with a broad range of known activities, including beta-glucosidase, beta-xylosidase, *N*-acetyl beta-glucosaminidase, cellodextrinase and exo-1,3-1,4-glucanase. Based on homology search, the GH3 module found in group-4 Doc-bearing proteins (ZP_06143379.1) has 78 % identity with a known beta-glucosidase [32]. On the other hand, family GH24 is exclusively composed of lysozymes, which are more likely related to defensive strategies rather than complex polysaccharide hydrolysis for energy purposes.

2.2. Structure of the R. flavefaciens CohScaE-DocScaH complex

Presently, the only available structure of a Coh-Doc complex involving a group-4 Doc is that of the CohScaE in complex with the XDoc dyad of the CBM-bearing protein CttA [26]. Due to the three unique insertions within its structure, the Doc has a distorted internal symmetry, likely impairing its ability to bind in a dual mode [21,26,33]. In order to understand the molecular determinants governing the anchoring of proteins to the wall of *R. flavefaciens* FD-1 through putative dual-binding mode interactions, we have solved the structure of the Coh of cell- envelope-attached scaffoldin E in complex with the group-4 Doc from the monovalent adaptor scaffoldin H, using X-ray crystallography. Given that DocScaH is likely to bind CohScaE with a dual-binding mode, expression and purification of the wild-type complex would result in two different conformations in solution. The implied heterogeneity of such a solution was likely to impair crystal formation. Consequently, we have designed a mutant variant of DocScaH (DocScaH 79/80) by alanine substitution of the residues at the canonical Coh-recognition positions 11 and 12 of the C-terminal repeat (residues #79, and #80 of the constructs used in the present work) [7]. Thus, the DocScaH mutant was forced to bind CohScaE through a single interface, thereby promoting homogeneity of the purified protein. This strategy resulted in high-quality crystals, which diffracted to a maximum resolution of 1.7 Å and belonged to the monoclinic space-group $P2_1$, with unit-cell dimensions of a = 40.81 Å, b = 120.00 Å, c = 65.95 Å, and β = 97.773 Å. The structure of the CohScaE-DocScaH complex was solved following a molecular replacement strategy, using the available model of CohScaE in complex with the XDoc of CttA (PDB 4IU2). The final model obtained exhibited a resolution of 1.71 Å and possessed two copies of the complex in the asymmetric unit (Fig. 2). The crystallographic dimer resulted from interactions between two CohScaE copies (chains A and C), including 6H-bonds and 26 non-bonded contacts. The biological relevance of those interactions is unknown, but an analysis with the *Protein Interfaces, Surfaces and Assemblies* (PISA) [34] tool of the CCP4 suite [35] showed an interface area of 356 Å, a solvation energy of 2.6 kcal/mol and a binding energy of −0.3 kcal/mol, suggesting that the dimer is not a biological assembly. The two CohScaE-DocScaH complex molecules share 428 water molecules, and each Doc coordinates two calcium (Ca^{2+}) ions, while each Coh coordinates one. The final model was

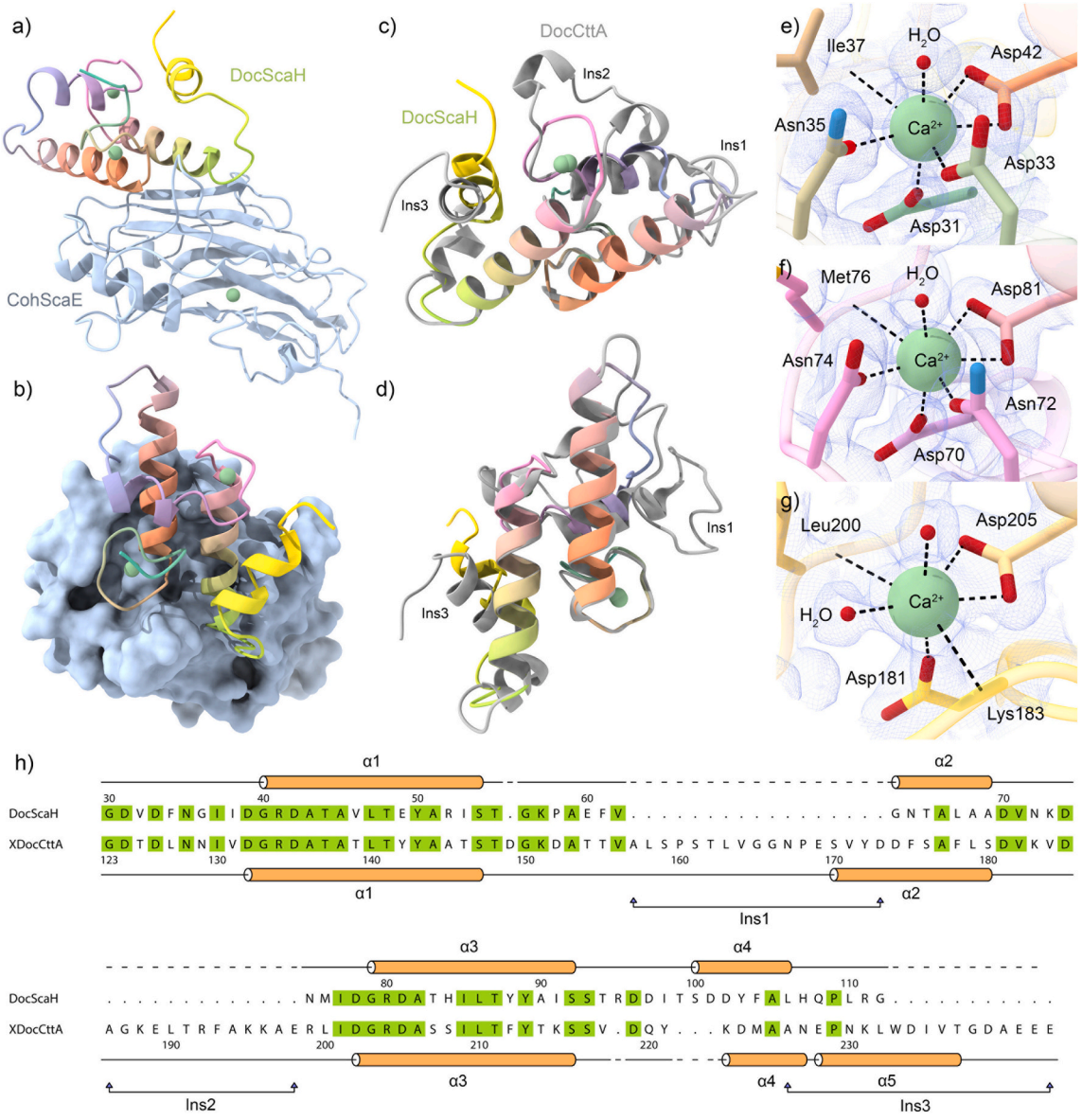

Fig. 2. 3D structure of the CohScaE-DocScaH complex. (a) Structure of the CohScaE-DocScaH complex, in ribbon representation, with the Coh in blue and the Doc color-ramped from red (N-terminus) to yellow (C-terminus). (b) CohScaE-DocScaH complex interface showing the van der Waals surfaces of the Coh module. (c) Overlay of the DocScaH and its group-4 structurally homologous XDoc from protein CttA (PBD code 4IU2), showing the absence of the three characteristic XDoc insertions in DocScaH, which creates a more symmetrical structure. (d) Cohesin-binding surface view of the DocScaH and XDocCttA overlay. (e, f) Detailed views of the respective N-terminal and C-terminal calcium-coordinating loops from DocScaH. (g) Detailed view of the C-terminal calcium-coordinating loop from XDocCttA, displaying an atypical coordination pattern with a second bridged water molecule. Calcium-coordinating residues are depicted in stick representation, hydrogen bonds are shown as dashed lines, water molecules as red spheres, and calcium ions as green spheres. The $2mF_{obs}$-DF_{calc} electron density map around the Ca^{2+} ions is contoured at 1 rmsd. (h) Primary structure alignment between DocScaH and XDocCttA, according to the 3D structure alignment performed with the matchmaker tool of ChimeraX [62]. Identical residues are highlighted in green. A cartoon representation of the DocScaH and XDocCttA secondary structures are displayed in orange, above and below the alignment, respectively. The three symmetry-disrupting inserts of the XDoc are labeled as Ins1–3.

deposited in the Protein Data Bank under the accession code 8AJY and contains residues 31–230 of DocScaH (Nonredundant RefSeq accession number ZP_06142361) and 270–363 of CohScaE (Nonredundant RefSeq accession number CAK18898). Data collection and refinement statistics are presented in Table 1.

2.3. Structure of DocScaH in the complex

The overall structure of DocScaH follows the classical Doc fold with two duplicated calcium-coordinating EF-hand-like motifs. Two α-helices (1 and 3), arranged in an antiparallel orientation, comprise the Coh-recognition surface and are connected by a third helix (α-helix 2). Unlike most Docs, including the majority of type-III Docs from *R. flavefaciens*, there is a C-terminal extension on DocScaH containing an extra α-helix (α-helix 4) (Fig. 2). Curiously, the extra helix disrupts the typical Doc clasp observed in many Docs, resulting from interactions between the N-terminal and C-terminal ends, which stabilizes the structure and contributes to their globular conformation [36]. Nonetheless, the DocScaH's structure is generally less elaborate than those of the known group-4 XDocs: its C-terminal extension has only one extra helix, as opposed to two in the XDocs, and it also lacks two other characteristic XDoc insertions that support the extended stalk-like conformation of the X-module, disrupting Doc symmetry. Likewise, calcium coordination also differs between DocScaH and the XDocs. The

Table 1
X-ray diffraction data collection and refinement statistics.

Crystal		RfCohScaE-DocScaH
Space group		$P12_11$
Unit cell parameters		
a, *b*, *c* (Å)		40.81, 120.00, 65.95
α, β, γ (°)		90.00, 97.77, 90.00
Data collection statistics		
X-ray source		ESRF, ID23-2
Wavelength, Å		0.873127
Total|unique no. of reflections		257,197|66,981
Resolution limits, Å		65.34–1.71 (1.74–1.71)
Completeness, %		99.6 (99.3)
Redundancy		3.8 (3.4)
Wilson B factor, Å^2		13.2
Matthews coefficient, Å^3/Da		2.37
Solvent content, %		48.1
Average I/σ(I)		8.0 (2.1)
R-merge(*I*)[a]		0.108 (0.500)
R-p.i.m.[b]		0.081 (0.389)
CC(1/2)		0.994 (0.789)
Structure refinement statistics		
Resolution used in refinement, Å		65.34–1.71
R-work[c]		0.194
R-free[c]		0.237
No. of protein residues in the asymmetric unit		569
No. of water molecules in the asymmetric unit		428
No. of atoms in the asymmetric unit		4810
R.M.S.Z., bond length, Å		0.88
R.M.S.Z., bond angles, °		0.98
Average temperature factor main chain|side chain, Å^2	Chain A	17.35|17.28
	Chain B	17.64|17.34
	Chain C	20.86|20.47
	Chain D	19.38|18.96
	Calcium atoms	15.17
	Solvent	24.38
Ramachandran plot	Residues in favored regions, %	99.6
	Residues in allowed regions, %	0.4
	Residues in forbidden regions, %	0.0
PDB ID		8AJY

Values in parentheses are for the highest resolution shell.

[a] $R_{merge} = \frac{\sum \mathrm{hkl} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum \mathrm{hkl} \sum_{i=1}^{n} I_i(hkl)}$, where I is the observed intensity, and $\bar{I}$ is the statistically-weighted average intensity of multiple observations.

[b] $R_{\mathrm{p.i.m.}} = \frac{\sum_{hkl} \sqrt{1/(n-1)} \sum_{i=1}^{n} |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_{i=1}^{n} I_i(hkl)}$, a redundancy-independent version of R_{merge}.

[c] $R_{work} = \frac{\sum \mathrm{hkl} ||F_{obs}(hkl)| - |F_{calc}(hkl)||}{\sum \mathrm{hkl} |F_{obs}(hkl)|}$, where F_{calc} $|F_{obs}|$ are the calculated and observed structure factor amplitudes, respectively. R_{free} is calculated for a randomly chosen 5% of the reflections.

present structure shows 2 Doc-bound calcium ions coordinated in a typical pentagonal bipyramid geometry, with the calcium-interacting residues following the N, N + 2, N + 4, N + 6, N + 11 pattern (Fig. 2). The first calcium is located near the N-terminus and is coordinated by residues Asp31 (Oδ1), Asp33 (Oδ1), Asn35 (Oδ1), Ile37 (backbone carbonyl), Asp42 (Oδ1 and Oδ2), and a water molecule, while the second calcium is coordinated by Asp70 (Oδ1), Asn72 (Oδ1), Asp74 (Oδ1), Met76 (backbone carbonyl), Asp81 (Oδ1 and Oδ2) and a water molecule. On the other hand, the C-terminal calcium-binding loop observed in the XDocs of *R. flavefaciens*' CttA and ScaB is distorted by a 13-residue long insertion, that displaces the traditional positions of the coordinating residues, forcing a second bridged water molecule to complete the pentagonal bipyramid geometry (Fig. 2). The third major difference between DocScaH and the group-4 XDocs is the absence of an insert between α-helices 2 and 3 that extends the linker connecting the two motifs and the C-terminal portion of α-helix 2. The absence of this insert creates a much more symmetrical structure on DocScaH than on the XDocs, evidenced by the low root mean square deviation (rmsd) when overlaying a 180° rotated copy of α-helix 1 with α-helix 3 (0.283 Å between 16 atom pairs). Overall, coupled with its primary sequence symmetry, the DocScaH 3D structure suggests a potential dual-binding mode for Coh interaction. A structure similarity search using the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) revealed that, as expected, the closest functionally relevant structural homologues of DocScaH are the group-4 XDoc of CttA (PDB code 4IU2, Z-score of 10.5, rmsd of 2.1 Å and 40 % identity over 83 aligned residues) followed by the type-I Docs of *Acetivibrio cellulolyticus* (4UYP, Z-score of 10.5, rmsd of 2.1 Å and 40 % identity over 83 aligned residues) and *C. thermocellum* (2CCL, Z-score of 9.5, rmsd of 1.9 Å and 30 % identity over 61 aligned residues), both of which interact with their cognate Coh partners through a dual-binding mode interaction. Curiously, the closest structurally related *R. flavefaciens* Doc, apart from CttA's XDoc, is a group−3 Doc, bound to adaptor scaffoldin C (5LXV, Z-score of 8.2, rmsd of 2.1 Å and 26 % identity over 62 aligned residues), which is divergent from the majority of *R. flavefaciens* type-III Docs and more closely related to the type-I Docs from other species, albeit lacking the ability to function with a dual-binding mode [25,37].

2.4. Structure of CohScaE in the complex

The structure of CohScaE in complex with DocScaH is nearly identical to the one reported in the complex with the XDoc of CttA (PDB code 4IU2), as suggested by the low rmsd of 0.198 among 199 atoms [38]. Like all reported Coh structures, CohScaE displays a classical β-sandwich jellyroll topology, with 9 antiparallel β-strands forming the two faces of the sandwich (4 strands in the Doc-interacting face and 5 strands in the "back" face) (Fig. 2). Despite the classical structure, CohScaE possesses several unique motifs, including a particularly large α-helix between strands 8 and 9, a long N-terminal extension before strand 1 which interacts extensively with the prominent α-helix [26], and a calcium-coordination loop between the α-helix and strand 9 which coordinates a calcium ion in a classical pentagonal bipyramid geometry. A structure similarity search using the DALI server revealed that the closest functionally relevant homologue of CohScaE is the isolated Coh of ScaG (PDB code 4N2O, Z-score of 18.3, rmsd of 2.4 and 21 % identity over 165 aligned residues), which has been shown to interact with group-4 Docs [39]. Curiously, like DocScaH, CohScaE also bears structural homology with type-I complexes of both *C. thermocellum* (PDB code 4DH2, Z-score of 13.5, rmsd of 2.3 Å and 16 % identity over 140 aligned residues) and *A. cellulolyticus* (PDB code 4UYP, Z-score of 13.5, rmsd of 2.5 Å and 20 % identity over 134 aligned residues). The closest *R. flavefaciens*' structural homologue to CohScaE is the group-3 and -6 Doc-binding Coh of adaptor scaffoldin C (PDB code 5LXV, Z-score of 12.7, rmsd of 2.9 Å and an identity of 14 % across 139 aligned residues), much like DocScaH is related to group-3 dockerins. This suggests that the insert-lacking type-III complex of CohScaE–DocScaH, like the complexes of group-3 Docs with the Coh of ScaC, are reminiscent of type-I complexes from the canonical cellulosomes of *C. thermocellum* and *A. cellulolyticus*.

2.5. CohScaE–DocScaH complex interface

The CohScaE–DocScaH interface comprises the Coh face defined by β-strands 5-6-3-8 and α-helices 1 and 3 of the Doc. Although β-strands 5, 6 and 3 of the Coh create a predominantly flat Doc-binding surface, a particularly long flap protrudes from β-strand 8 at almost a 90° angle, creating an L-shaped groove and positioning itself between the N-

terminal end of the Doc's helix 1 and the C-terminal portion of helix 3 (Fig. 2). This flap makes extensive contacts with the Doc, thus contributing to complex formation, as was also observed with the Coh of *R. flavefaciens* ScaC. A network of both polar and hydrophobic interactions was identified at the complex interface (Fig. 3, Table 2).

Curiously, unlike most dual-binding Docs, the interaction with CohScaE does not seem to be dominated by either of the contacting helices, but with both helix 1 and 3 making a similar number of contacts with the Coh partner. Thus, both helices lie flat on top of the Coh and parallel to the binding interface. This contrasts with type-I complexes from the canonical *C. thermocellum* cellulosome, where the non-dominant interacting helix has its low contact end slightly detached from the binding surface (Fig. S2). Although the ITC results with the wild-type complex suggest an enthalpy-driven and therefore H-bond-dominated interaction, the number of polar contacts formed between the two modules is less abundant than those reported for most *R. flavefaciens* complexes. On the other hand, an extensive number of hydrophobic interactions is observed, with an important hydrophobic binding surface created by a pair of leucine (Leu47 and Leu86) and tyrosine (Tyr50 and Tyr89) residues. Modelling a 180°-rotated DocScaH, bound to CohScaE by matching the duplicated residues at the same relative positions of each interacting helix, gives a nearly identical network of contacts in both complexes. Once again, this suggests that DocScaH can bind its protein partner in either of the two distinct orientations (Fig. 4).

2.6. CohScaE–DocScaH binding mechanism

To better understand the mechanism underlying the Coh-Doc interaction, the contribution of each key Doc residue, identified in the 3D complex structure, was evaluated using site-directed mutagenesis coupled with native-gel electrophoresis (Fig. S3) and ITC. Residues establishing direct hydrogen bonds as well as those with the most extensive hydrophobic contacts were mutated to alanine and the resulting variants were tested against CohScaE (Table S1). Alanine was chosen (unless otherwise stated below) because it typically represents a deletion of the side chain at the β-carbon, maintaining the main-chain conformation. Because of the possible dual-binding mode, all mutations were performed on both helices, at equivalent positions. ITC data indicated a 1:1 binding and an association constant (*K*a) of 3.51×10^6 M^{-1} between the wild-type Doc and Coh. Of all variants, the 41/80 variant had the highest impact on binding (to Ka = 6.75×10^5 M^{-1}) (Table 3). This is consistent with observations of the structure, where Arg41 makes numerous polar and hydrophobic contacts. However, ITC experiments showed that the single-repeat mutant variants 41 and 80, separately, exhibited a similar affinity to that of the wild-type Doc, (Ka = 1.04×10^6 M^{-1} and 1.58×10^6 M^{-1}, respectively), suggesting that the two positions have different contributions to binding, with one dominating relative to the other.

Confirmation of the dual-binding mode was performed by replacing the characteristic group-4 Gly-Arg pair at key Coh-recognition positions 11 and 12 by a Glu-Ala pair. The swap was made in either one or both putative duplicated recognition sites (residues 40/41 and 79/80 of the construct used in this study). Glu was used to replace Gly as its bulky

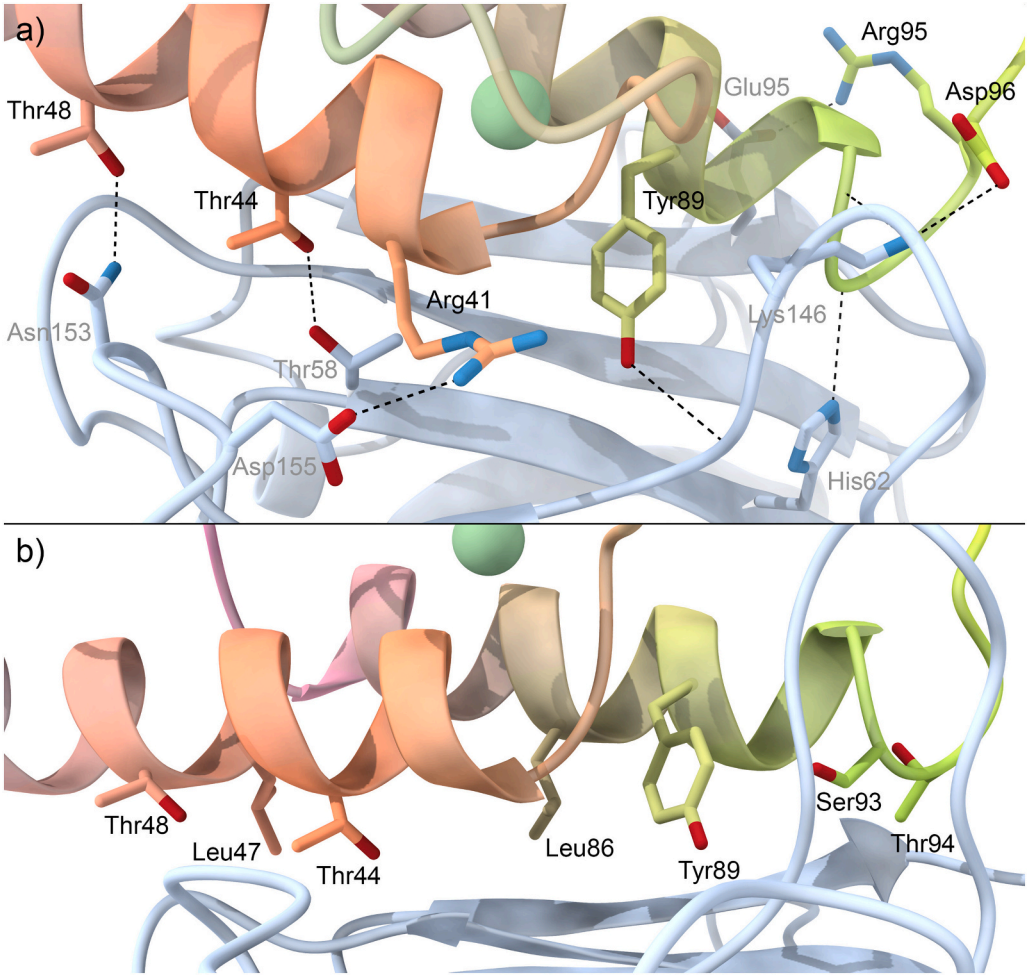

Fig. 3. Detailed view of the network of CohScaE-DocScaH interface interactions. Panels (a) and (b) highlight of the main polar interactions and the main hydrophobic contacts, respectively, involved in Coh-Doc recognition and complex formation. In both panels the main contacting amino-acid residues are depicted in the stick configuration. Dashed black lines represent hydrogen-bond interactions. The CohScaE-DocScaH complex is shown in ribbon representation, with the Doc color-ramped (helix 1 in orange and helix 2 in pink to yellow) and the Coh in blue. Ca^{2+} ions are depicted as green spheres.

Table 2
Main interactions between CohScaE and DocScaH.

Direct hydrogen bonds					
	DocScaH			CohScaE	
#	Residue	Atom	Distance (Å)	Residue	Atom
	ARG41	NH1	2.96	ASP155	OD2
	THR44	OG1	2.81	THR58	OG1
	THR48	OG1	2.81	ASN153	ND2
	TYR89	OH	2.76	GLY147	O
	SER92	O	2.74	LYS146	NZ
	SER93	O	2.78	HIS62	NE2
	ARG95	NH1	2.84	GLU95	OE1
	ASP96	OD2	3.27	LYS146	NZ

Salt bridges					
	DocScaH			CohScaE	
#	Residue	Atom	Distance (Å)	Residue	Atom
	ARG41	NH1	2.96	ASP155	OD2
	ARG95	NH1	2.84	GLU95	OE1
	ASP96	OD2	3.27	LYS146	NZ

Non polar interactions/non bonded contacts		
	DocScaH	CohScaE
#	Residue	Residues
	ILE37	SER145 (2)
	ASP39	LEU149
	GLY40	THR58 (2), LEU149
	ARG41	THR58, LEU149, ASP155 (3), LYS160
	ALA43	LEU91
	THR44	THR58 (4), GLY106, ALA107 (2), ASN153
	LEU47	SER90 (2), LEU91 (2), ALA107 (2), ASP108
	THR48	ALA107, ASP108 (4), ASN153 (4)
	ALA51	ASP108 (2)
	THR83	LEU91
	LEU86	LEU91, ALA92, LYS93, ALA104
	THR87	LYS93
	TYR89	GLY60, PHE102 (3), GLY147 (4), ASP148 (2), LEU149 (4)
	ALA90	GLU95 (5), PHE102 (3)
	SER92	LYS146 (8)
	SER93	HIS62 (4), PHE102 (9), LYS146 (2)
	THR94	GLU95 (2), ASN96, ASN97 (7), PHE102
	ARG95	GLU95 (3)
	ASP96	LYS146 (2)

side-chain better contrasts with the single hydrogen in Gly than the relatively small methyl group of Ala. In both cases, the single-site mutations maintained the affinity towards the Coh, while the double-site mutant lost the ability to recognize CohScaE. This confirms the hypothesis that DocScaH, a group-4 Doc, lacking the atypical inserts of the XDocs of ScaB and CttA, can act in a dual-binding mode.

Most of the other mutated residues did not have a significant impact on Coh-Doc affinity, likely because the extensive network of intermolecular contacts compensated for their absence. Nonetheless, the mutation of the Leu47/Leu86 pair, part of the above-mentioned hydrophobic patch on the binding surface of DocScaH, resulted in a slight decrease in affinity, thus suggesting a central role of that hydrophobic area for binding.

2.7. Changing the hydrophobic properties on the Doc's surface to explore its effects in binding affinity

Interestingly, replacing Thr44 and Thr83, which are adjacent to the hydrophobic patch, resulted in an increased binding affinity (Ka = 3.17 $\times$ 10^7 M^{-1}) between the modules. This effect was also noticed with the single Thr83Ala substitution (1.14 $\times$ 10^7 M^{-1}) (Table 3), although not with Thr44Ala alone (Ka = 6.82 $\times$ 10^6 M^{-1}). Due to the hydrophobic properties of Ala, it was hypothesized that the resulting extended hydrophobic patch with the Thr-to-Ala substitutions were responsible for the increased binding affinity, while the lack of effect of the single Thr44 substitution can probably be explained by the loss of an H-bond interaction that compensates for the slight increase in hydrophobicity. To test this hypothesis, three new variants were constructed: Thr44Leu (T44L), Thr83Leu (T83L) and Thr44Leu + Thr83Leu (T44L/T83L). The first two were used to evaluate the individual effect of each of the duplicated residues, while the double substitution led, together with Leu47 and Leu86, to a quartet of closely positioned leucine residues that significantly extended the hydrophobic patch at the Doc's surface (Fig. 5). As proposed, since leucine has more pronounced hydrophobic properties than alanine, the single mutations resulted in higher affinities than those observed with equivalent alanine substitutions: T44L (Ka = 7.50 $\times$ 10^7 M^{-1}) and T83L (Ka = 6.55 $\times$ 10^7 M^{-1}). In addition, the double mutation resulted in a binding constant of 1.53 $\times$ 10^8 M^{-1}, *i.e.*, two orders of magnitude higher than that of the wild-type Doc.

3. Discussion

Initially, only three *R. flavefaciens* XDoc-bearing proteins had been reported to interact with the ScaE Coh. These included CttA, which is a double-CBM-bearing protein, the nine-Coh primary scaffoldin, ScaB, and a cysteine peptidase of the NlpC/P60 superfamily (accession number: ZP_06142651) [21,30]. With the sequencing of the entire *R. flavefaciens* genome and the identification of over 200 Doc-bearing proteins [19], further work was developed to understand the full extent of this species' cellulosome assembly. This led to the classification of the Docs into 6 groups according to sequence homology [23]. Later [22], this classification was found to be functionally relevant, as members of the same group share the same specificity. ScaB, CttA and the cysteine peptidase were included in group-4, which now has a total of 46 identified members. Thus, many other group-4 Doc-bearing proteins can also presumably be directed towards the bacterial cell surface through their interaction with ScaE.

3.1. ScaH introduces the dual-binding mode into the cellulosome of R. flavefaciens

The targeting of cellulolytic enzymes directly to the bacterial cell wall has been previously demonstrated in *C. thermocellum*, through type-I Docs that lacked the structural symmetry needed for a dual-binding mode [27]. It has been suggested that those enzymes might work in synergy with the cellulosome towards the efficient hydrolysis of the polysaccharide substrates. Dual-binding enzyme-borne Docs were also observed interacting with anchoring scaffoldins in *C. thermocellum*, while still preferring the main cellulosomal Cohs. One suggested hypothesis is that the secreted cellulosomal enzymes might transiently interact with the cell surface, to ensure that they are retained, even if the Cohs on the cellulosomes are saturated. These enzymes would then integrate into a cellulosome once a scaffoldin becomes available. Interestingly, group-4 Docs from *R. flavefaciens* display the internal symmetry required for a dual-binding mode, unlike other type-III Docs, which generally exhibit atypical second repeats. Until now, the best-characterized group-4 interactions were those of the ScaB and CttA XDocs, responsible for tethering the cellulosome to the bacterial cell wall and for mediating recognition and cell adhesion to polysaccharide substrates [20,21,40]. However, these XDocs have unusual structural motifs that disrupt their symmetry, which are not present in most group-4 Docs. Therefore, it was considered pertinent to understand the structural determinants governing the interaction of group-4 Docs that lack the X module but exhibit the potential ability to interact with the respective Coh according to the dual-binding mode. The chosen Doc for the present work is a particularly interesting subject, as it belongs to the adaptor scaffoldin ScaH. ScaH is a monovalent scaffoldin, meaning it

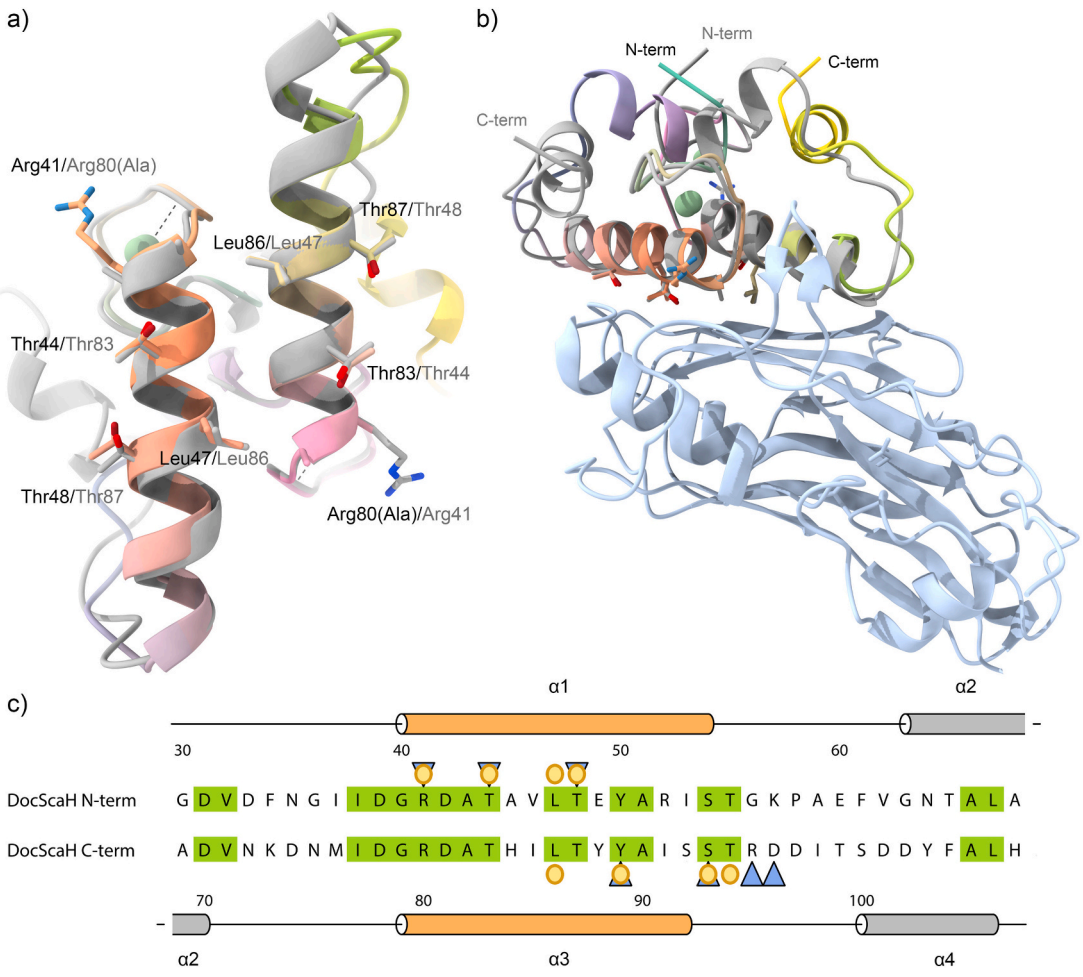

Fig. 4. The highly symmetric nature of DocScaH supports a dual-binding mode of interaction. (a) DocScaH structure overlaid with a 180°-rotated version of itself (gray), showing the matched duplicated interacting residues at the same relative positions of each helix. Key Coh-interacting residue pairs are depicted in stick configuration. (b) Overlay of the CohScaE-DocScaH complex with a rotated version of DocScaH (gray) demonstrating that the overall structure of the complex as well as key contacts are maintained. (c), Primary structure alignment of the DocScaH N-terminal portion with its C-terminal portion showing the remarkable symmetry provided by a high degree of conservation. A cartoon representation of DocScaH secondary structure is displayed in orange, above the alignment. Conserved residues are highlighted in green. Residues involved in hydrogen bonding with CohScaE are marked with a blue triangle and those establishing hydrophobic contacts with yellow circles.

only has a single Coh module in addition to the Doc, which is itself able to recognize group-4 Docs. As such, this scaffoldin can also interact with the ScaE Coh. The obtained structure of DocScaH–CohScaE revealed that the absence of an X-module was accompanied by the absence of the three symmetry-disrupting inserts of the XDoc and that the major Coh-contacting residues are duplicated and occupy the same relative positions in each of the Doc repeats. This resulted in a 3D structure that is almost perfectly superposable with a 180°-rotated copy of itself and therefore capable of interacting in a dual-binding mode. This was confirmed by ITC experiments. Thus, ScaH can serve as an adaptor to introduce the dual-binding mode into the *R. flavefaciens* cellulosome by positioning itself between the main scaffoldin ScaB and the anchoring scaffoldin ScaE. Furthermore, since ScaH's Doc can also bind the Coh of another ScaH molecule (CohScaH) [22], it can potentially create multi-modular chains by assembling tandem copies that work as highly flexible spacers of variable lengths. Consequently, different cellulosome units would be positioned at different distances from the bacterial cell wall, thereby creating a more dynamic and spatially efficient mega structure (Fig. 6). In states of high cellulolytic activity with increased expression of cellulosomal components, this incorporation of the dual-binding mode through adaptors of varying sizes could serve to displace the multiplicity of cell wall-attached cellulosomes from each other, thus reducing the likelihood of non-productive steric clashes.

3.2. *The R. flavefaciens dual-binding-mode Coh-Doc interactions recruit a variety of enzyme activities to the cell surface*

It is well established that Coh-Doc interactions mediate cellulosome assembly and attachment to the bacterial surface, thereby maintaining a central role in complex polysaccharide degradation in many different species of anaerobic bacteria. Interestingly, among the 37 *R. flavefaciens* group-4 Doc-bearing proteins investigated in the present study, only a few seem to be directly involved in polysaccharide degradation or cellulosome assembly. The incorporation of non-cellulolytic proteins into the cellulosome, including serpins, expansins, peptidases and transglutaminases, has been previously observed in some bacterial species [16]. Although it is still unclear which role these proteins exert as cellulosomal components, it has been suggested that they may have protective or regulatory activities, or even act in synergy with CAZymes to degrade highly complex substrates [41–43]. The most curious aspect of the non-cellulolytic group-4 Doc-bearing proteins is that they appear to be directly targeted to the bacterial cell wall, as opposed to a cellulosomal scaffoldin. This suggests that these activities might not be associated with polysaccharide degradation at all, but rather that the bacteria might use the Coh-Doc system in other biological processes, as a strategy to attach key proteins to its cellular envelope. These can include the extracellular degradation of proteins and lipids, as suggested by the

Table 3
Thermodynamics of interaction between CohScaE and DocScaH variants. Indicated residues were replaced with alanine unless otherwise specified, *i.e.*, by Glu or Leu (E or L).

Doc	Ka M^{-1}	ΔG° kcal mol^{-1}	ΔH kcal mol^{-1}	−TΔS° kcal mol^{-1}	N
DocScaH WT	3.51E6	−3.88E4	−3.89E4	96.3	1.12
DocScaH 41	1.04E6	−2.39E4	−2.39E4	50.0	0.88
DocScaH 80	1.58E6	−3.87E4	−3.88E4	97.4	0.80
DocScaH 41/80	1.57E5	−6480	−6477	−2.75	0.93
DocScaH 40E/41	4.41E6	−4.14E4	−4.15E4	104	1.06
DocScaH 79E/80	1.42E7	−5.11E4	−5.12E4	133	1.05
DocScaH 40E/41/79E/80	Nb	Nb	Nb	Nb	Nb
DocScaH 44	6.82E6	−3.51E4	−3.52E4	83.0	1.07
DocScaH 83	1.14E7	−3.77E4	−3.78E4	90.6	0.96
DocScaH 44/83	3.17E7	−3.25E4	−3.25E4	71.1	1.13
DocScaH 44L	7.50E7	−5.61E4	−5.62E4	146	1.05
DocScaH 83L	6.55E7	−6.92E4	−6.94E4	190	1.03
DocScaH 44/83L	1.53E8	−1.01E4	−1.01E4	−4.50	0.95
DocScaH 47	7.87E5	−1.03E4	−1.03E4	6.28	0.88
DocScaH 86	3.14E5	−1.52E4	−1.52E4	24.1	0.88
DocScaH 47/86	5.36E5	−7370	−7368	−2.30	0.91
DocScaH 48	2.78E6	4.64E4	−4.65E4	121	1.05
DocScaH 87	5.35E6	−5.05E4	−5.06E4	133	1.09
DocScaH 48/87	1.03E7	−4.03E4	−4.04E4	98.9	1.06
DocScaH 55	1.05E7	−4.66E4	−4.67E4	119	0.97
DocScaH 94	3.11E6	−3.63E4	−3.64E4	88.3	0.92
DocScaH 55/94	6.04E6	−4.51E4	−4.52E4	116	0.99

Nb, No binding.
WT, wild-type.
All thermodynamic parameters were determined at 308 K.

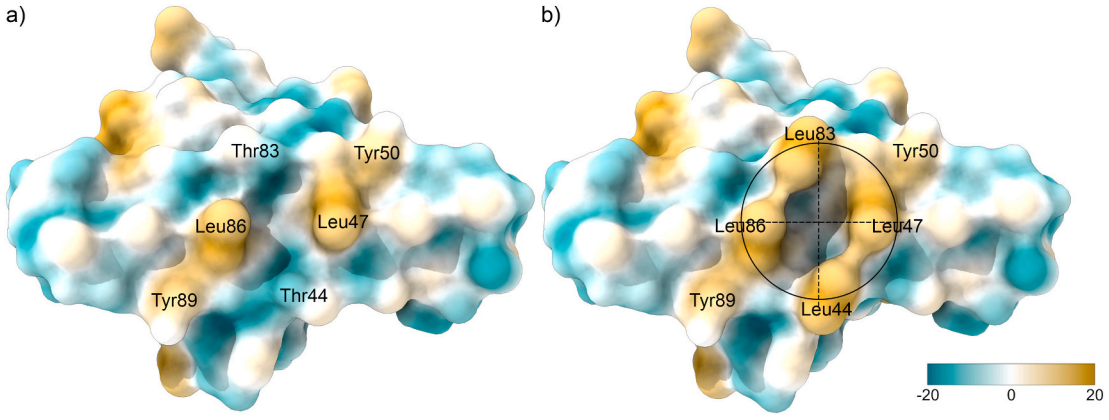

Fig. 5. Increasing the hydrophobic profile of DocScaH's binding surface. (a) Topology of the wild-type hydrophobic patch at the binding surface of DocScaH. (b) Topology of the mutated hydrophobic patch. The black circle depicts the quartet of leucine residues, highlighting the hydrophobic hot spot at the DocScaH surface created by combining mutations T44L and T83L.
The van der Waals surfaces are colored according to their molecular lipophilicity potential using the Fauchère method in ChimeraX, from blue (−20 units, hydrophilic) to yellow (+20 units, hydrophobic).

identified peptidase and lipase modules, or even in defense-and-attack mechanisms in microbial interactions, as suggested by the presence of lysozymes and a D-alanyl-D-alanine carboxypeptidase within the group-4 Doc-bearing proteins. Coh- and/or Doc-encoding genes have also been reported in many organisms from several domains of life that do not produce cellulosomes, but their function(s) remain speculative [44]. Therefore, it would not be surprising to eventually find these modules associated with a wide variety of cellular processes that are unrelated to classic cellulosome function.

3.3. Coh-Doc affinity can be manipulated through rational structure-guided protein engineering

The structure-function studies on the present complex have revealed a novel understanding of the molecular mechanisms driving Coh-Doc interactions. Although the main contacts identified between the two modules, as well as the thermodynamic profile of the interaction, clearly show that polar contacts are the main drivers of binding, the data also show that there are numerous hydrophobic forces which are crucial for complex formation. Particularly, an important hydrophobic area was identified at the binding surface of DocScaH, which led to a decrease in complex affinity once disrupted. The topology of the surrounding area suggested that this hydrophobic patch could be extended, while maintaining the overall conformation of the Doc's surface. Specifically, a pair of two threonine residues that are perpendicularly positioned relatively to the two leucine residues forming the hydrophobic patch, provided an attractive target to engineer an improved hydrophobic interface (Fig. 5). Interestingly, in early work [5,45], the positions of these two residues were also suggested to play a Coh-binding role. By changing the threonine residues to leucine residues, it was possible to form a quartet of leucine residues that increased the hydrophobic profile of the Doc's Coh-interacting surface. This led to a 100-fold increase in the affinity between the two modules. These results are a testament to the importance of having detailed information regarding these unusually strong protein:

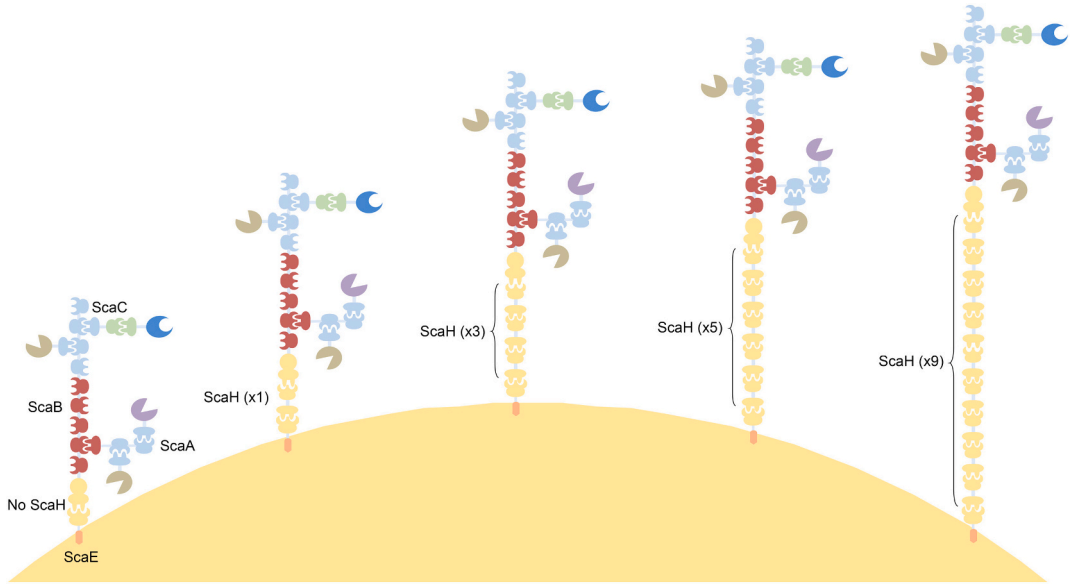

Fig. 6. ScaH as a spacer to avoid steric hindrance. Because the Doc of ScaH can recognize the Coh of another ScaH unit, it is possible for *R. flavefaciens* to assemble chains of tadem copies of ScaH with variable lengths. These ScaH chains could then position themselves between ScaE and ScaB to introduce the dual-binding mode and to arrange different cellulosome units at different distances from the bacterial cell wall, thereby creating a more dynamic and spatially efficient mega structure.

protein interactions. Guided by these data, it is possible to manipulate the affinity between the two modules, either making it weaker or stronger, or even changing the specificity of the Docs towards different Cohs [46]. This could potentially inform attempts at creating tailored Coh-Doc interactions with various applications, such as affinity-based molecular tools, or designer cellulosomes for biomass conversion [6].

3.4. Concluding remarks

The present study revealed the mechanism by which the particularly intricate *R. flavefaciens* cellulosome can incorporate the dual-binding mode of the Coh-Doc interaction into its cell-surface attachment. Unlike previously characterized type-III *R. flavefaciens* Coh-Doc complexes, such as those governing the interaction between scaffoldins, those responsible for recruiting enzymes into the cellulosome and even those tethering the cellulosome to the cell-wall, the structure of DocScaH in complex with CohScaE revealed a remarkable structural symmetry capable of supporting two 180°-related conformations. Upon characterization of the archetypal cellulosome of *C. thermocellum*, it was believed that the dual-binding mode was tied to Coh-Doc types and their role in the cellulosome. Until recently, the prevailing notion was that type-I complexes responsible for enzyme recruitment operated in a dual-binding mode, while type-II interactions, anchoring the cellulosome to the cell-wall, had a single-binding mode. This was recently proven inaccurate, since, depending on the species, there are both type-I and -II complexes responsible for enzyme recruitment, as well as for cell wall tethering. Likewise, there are both type-I and -II complexes capable of interacting in a dual-binding mode, as well as in a single-binding mode. The dual-binding mode now seems much more widely distributed than initially proposed [10]. The present findings further expand this notion, by confirming the presence of the dual-binding mode in type-III complexes and by proposing a mechanism by which it can be selectively introduced into the *R. flavefaciens* cellulosome, a structure which was previously considered to be completely assembled through single-binding mode interactions.

Unfortunately, the exact significance of the dual-binding mode is still somewhat elusive. The best explanation might relate to the addition of an extra layer of flexibility, by allowing the cellulosomes to organize their components in different orientations and thus avoid steric hindrance [47,48]. This hypothesis is consistent with the prevalence of the single-binding mode in the *R. flavefaciens* cellulosome which, albeit complex, is a relatively small cellulosome that can only assemble up to a maximum of 14 enzymes in a single unit. This is considerably less than the *B. cellulosolvens* or *C. clariflavum* systems, which support the assembly of up to 110 and 160 enzymes, respectively, in a single unit, all through dual-binding mode interactions. Nonetheless, periods of intense cellulolytic activity may lead to several cellulosome units to be expressed on the surface of *R. flavefaciens*. At this stage it would make sense to incorporate the dual-binding mode in newly generated complexes by means of ScaH, which would improve the overall flexibility of the enzyme complexes. This hypothesis remains to be proved, but it is unlikely that such a widespread feature would be conserved throughout evolution without presenting a selective advantage. In addition, recent findings suggest the existence of active regulatory mechanisms for the dual-binding mode, which rely on pH or the isomerization state of key proline residues to switch between binding orientations, suggesting a functional role for the dual-binding mode [49,50].

The present study also revealed how a deep understanding of the fine mechanisms governing Coh-Doc interactions can be used to manipulate the affinity between the two modules. A combination of structural and thermodynamic data was applied to a protein-engineering strategy, creating a Doc variant with improved affinity for its cognate Coh. Previous studies [45,46,51–53] have been successful in switching Doc specificity or in decreasing affinity, but to our knowledge, this is the first successful attempt at improving Coh-Doc affinity. Many studies have previously explored the development of tailored cellulosomes for the conversion of lignocellulosic biomass or other technologies relying on the Coh-Doc interaction, such as biosensors or protein purification strategies [52,54] and the manipulation of the Coh-Doc interaction is a key aspect in devising such strategies. The present results offer a rational basis for engineering more stable and effective complexes that will inform future efforts aiming to harvest the potential of the Coh-Doc interaction.

4. Methods

4.1. Gene synthesis and DNA cloning

Docs are small unstable protein modules when expressed individually in heterologous hosts. To promote stability, *R. flavefaciens* FD-1

DocScaH (ZP_06142361) was co-expressed *in vivo* with CohScaE (CAK18898). 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, Germany) under the control of separate T7 promoters. The DocScaH-encoding gene was positioned at the 5′ end and the CohScaE-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 also contained a His6 tag introduced at the N-terminus of the Doc required for protein purification by immobilized metal affinity chromatography (IMAC). We use the polyhistidine tag at the Doc N-terminus, because the expression levels of both Coh and Doc are higher, contrary to Coh tagging that results in the accumulation of large levels of unbound Coh in the purification product.

To produce recombinant DocScaH and CohScaE individually, the recombinant complex was digested with *Bgl*II to excise the Doc-encoding amplicon. This strategy gave a pET28a derivative encoding the recombinant CohScaE fused to a C-terminal hexa-histidine tag, maintaining plasmid integrity by re-ligating. The DocScaH-encoding gene was subcloned into the pHTP8 plasmid by homologous recombination (NZYTech Ltd) following the manufacturer's protocol. The resulting expressed product consisted of a His-tagged DocScaH fused to Thioredoxin (Trx) for increased solubility and stability.

To produce the mutants used in the binding experiments, several DocScaH protein derivatives were constructed (Table S1) using site-directed mutagenesis and the primers shown in Table S2. Each of the newly generated gene sequences was fully sequenced to verify that only the desired mutation accumulated in the nucleic acid chain.

4.2. Protein expression and purification

E. coli BL21 (DE3) cells were transformed with vectors containing the constructs of interest and grown at 37 °C to an OD_{600} of 0.4–0.6. Recombinant protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by incubation at 19 °C for 16 h. After harvesting the cells by centrifuging 15 min at 5000 ×*g*, the cells were resuspended in 10 mL of immobilized-metal affinity chromatography (IMAC) binding buffer (50 mM HEPES, pH 7.5, 10 mM imidazole, 1 M NaCl, 5 mM $CaCl_2$). Sonication was used to disrupt the cells, and the cell-free supernatant fluids were then recovered by centrifuging for 30 min at 15,000 ×*g*. After loading the soluble fraction into a HisTrap nickel-charged Sepharose column (GE Healthcare, UK), initial purification was carried out by IMAC in an FPLC system (GE Healthcare, UK) using conventional protocols with a 35 mM imidazole wash and a 35–300 mM imidazole elution gradient. After selecting the fractions containing the Coh–Doc complex, the buffer of the purified samples was changed to 50 mM HEPES, pH 7.5, containing 200 mM NaCl, 5 mM $CaCl_2$, using a PD-10 Sephadex G-25 M gel-filtration column (Amersham Pharmacia Biosciences, UK). Gel-filtration chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare, UK) was used to ensure a high-level of purification, required for crystallography. The purified complex samples were concentrated in an Amicon Ultra-15 centrifugal device with a 10-kDa cutoff membrane (Millipore, USA) and washed three times with molecular biology grade water (Sigma) containing 0.5 mM $CaCl_2$. The final protein concentration was adjusted to 30 mg·mL^{-1}. Protein concentration was estimated in a NanoDrop 2000c spectrophotometer (Thermo Scientific, USA). SDS–PAGE gels (14 % w/v) were used to confirm the purity and molecular mass of the recombinant complexes. CohScaE, DocScaH and its mutant derivatives, used in ITC and native PAGE experiments, were expressed as described above and purified with His GraviTrap gravity-flow nickel-charged Sepharose columns (GE Healthcare, UK). After IMAC, the recombinant proteins were buffer exchanged to 50 mM HEPES pH 7.5, 0.5 mM $CaCl_2$ and 0.5 mM TCEP, using PD-10 Sephadex G-25 M gel filtration columns (GE Healthcare, UK).

4.3. Nondenaturing gel electrophoresis (NGE)

For the NGE experiments, each of the DocScaH and mutant variants, at a concentration of 50 μM, were incubated in the presence and absence of 50 μM CohScE for 1 h 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 distinct electrophoretic mobility from the one presented by the individual modules.

4.4. Crystallization of CohScaE-DocScaH

Several crystallization conditions were tested by using the sitting-drop vapor-diffusion method, with the aid of an Oryx8 robotic nano-drop dispensing system (Douglas Instruments, UK). The commercial kits JCSG+ HT96, MIDAS (Molecular Dimensions, UK), Crystal Screen and PEG/Ion (Hampton Research, California, USA), and in-house prepared 80 factorial solutions were used for the screening. Drops (1 μL) of 10 and 20 mg·mL^{-1} CohScaE-DocScaH were mixed with 1 μL reservoir solution at room temperature. The resulting plates were then stored at 293 K. Crystal formation was observed under nine different conditions after a period of approximately 30 days from setting up the plates. Crystals were cryoprotected in well solution with 25 % glycerol and flash cooled in liquid N_2 for data collection. Preliminary in-house X-ray diffraction experiments (microfocus IμS Bruker D8 Venture CuKα diffractometer operated at 50 kV and 1 mA and coupled to a Photon 100 CMOS detector) revealed that the best diffracting crystals were formed in a solution of 0.2 M Lithium sulfate monohydrate, 0.1 M TRIS hydrochloride pH 8.5 and 30 % w/v Polyethylene glycol 4000. Optimization plates based on this condition were set up, which resulted in the formation of several good quality crystals.

4.5. 3D structure determination and refinement

The crystal structure of CohScaE-DocScaH was determined by molecular replacement. X-ray diffraction data were collected on beamline ID23-2 at the ESRF, Grenoble, France, using a PILATUS3 X 2 M (Dectris Ltd.). A systematic grid search was carried out on all crystals to select the best-diffracting part of each crystal. iMosflm [55] was used for strategy calculation during data collection. All data sets were processed using the Fast_dp and xia2 packages, which use the programs XDS [56], POINTLESS and SCALA from the CCP4 suite. Data-collection statistics are given in Table 1. The best-diffracting crystal diffracted to a resolution of 1.5 Å and belonged to the monoclinic spacegroup $P2_1$. Phaser MR was used to carry out molecular replacement, using the structure of CohScaE in complex with the XDoc from CttA (4 IU2) [26]. Two copies of the CohScaE-DocScaH heterodimer were present in the asymmetric unit. The partially obtained model was completed with Buccaneer [57] and with manual modelling in COOT [58]. It was then refined using REFMAC5 [59] and PDB REDO [60] interspersed with model adjustment in COOT. The final round of refinement was performed using the TLS/restrained refinement procedure, using each module as a single group, giving the final model (Protein Data Bank code 8AJY). The rmsd of bond lengths, bond angles, torsion angles and other indicators were continuously monitored using validation tools in COOT and MOLPROBITY [61]. A summary of the refinement statistics is provided in Table 1. wwPDB Validation Service was used to validate the structures before deposition in the PDB. 3D structure figures were generated using UCSF ChimeraX [62].

4.6. Isothermal titration calorimetry

All ITC experiments were carried out at 308 K in a Microcal VP-ITC

system (Malvern Panalytical). The purified wild-type DocScaH and mutant variants were diluted to 50 μM, and CohScE was diluted to 150 μM. All diluted proteins were filtered using a 0.45 μm syringe filter (PALL). During titrations the Doc constructs were stirred at 307 revolutions/min in the reaction cell and titrated with 28 successive 10-μL injections of CohScaE 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, USA). The fitted data yielded the association constant (Ka) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic equation: $\Delta RTlnKa = \Delta G = \Delta H\text{-}T\Delta S$. Binding isotherms are shown in Fig. S4.

CRediT authorship contribution statement

PB and CMGAF have planned the experiments. PB, VA, ALC, MD and SN have interpreted the experiments. PB, MD and MC have purified and crystalized the complex. ALC and PB have collected the X-ray diffraction data and solved the Coh-Doc complex structure. MD has produced the mutants. MD and CC performed the binding experiments. PB and MD have written and illustrated the manuscript. CMGAF, ALC, MJR, LMAF, EAB and SN have revised the manuscript.

Funding

The authors would like to acknowledge the financial support of FCT - Fundação para a Ciência e a Tecnologia, I.P., in the scope of the Centro de Investigação Interdisciplinar em Sanidade Animal (CIISA) grant UIDB/00276/2020, the Associate Laboratory for Animal and Veterinary Sciences (AL4AnimalS) grant LA/P/0059/2020, the project grants PTDC/BIA-MIC/5947/2014 and RECI/BBB-BEP/0124/2012, the Research Unit on Applied Molecular Biosciences - UCIBIO grants UIDP/04378/2020 and UIDB/04378/2020 and the LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy - i4HB.

MS is supported by a PhD studentship (SFRH/BD/146965/2019) from FCT-MCTES.

Conflict of interest

The Co-author Carlos M.G.A. Fontes is Founder and CSO of Nzytech Lda, which has manufactured some of the reagents used in this work.

Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 8AJY [https://www.rcsb.org/structure/8AJY]. All further data supporting the findings of this study are available from the corresponding author, upon reasonable request.

Acknowledgements

We acknowledge the European Synchrotron Radiation Facility (ESRF) for access to the synchrotron facility through BAG-Portugal. Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases [62].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2022.10.102.

References

[1] G. Paës, D. Navarro, Y. Benoit, S. Blanquet, B. Chabbert, B. Chaussepied, P. M. Coutinho, S. Durand, I.V. Grigoriev, M. Haon, L. Heux, C. Launay, A. Margeot, Y. Nishiyama, S. Raouche, M.-N. Rosso, E. Bonnin, J.-G. Berrin, Tracking of enzymatic biomass deconstruction by fungal secretomes highlights markers of lignocellulose recalcitrance, Biotechnol. Biofuels 12 (2019) 76, https://doi.org/10.1186/s13068-019-1417-8.

[2] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) D490–D495, https://doi.org/10.1093/nar/gkt1178.

[3] L. Artzi, E.A. Bayer, S. Moraïs, Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides, Nat. Rev. Microbiol. 15 (2017) 83–95, https://doi.org/10.1038/nrmicro.2016.164.

[4] C.M.G.A. Fontes, H.J. Gilbert, Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates, Annu. Rev. Biochem. 79 (2010) 655–681, https://doi.org/10.1146/annurev-biochem-091208-085603.

[5] E.A. Bayer, J.-P. Belaich, Y. Shoham, R. Lamed, The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides, Annu. Rev. Microbiol. 58 (2004) 521–554, https://doi.org/10.1146/annurev.micro.57.030502.091022.

[6] E.A. Bayer, E. Morag, R. Lamed, The cellulosome–a treasure-trove for biotechnology, Trends Biotechnol. 12 (1994) 379–386, https://doi.org/10.1016/0167-7799(94)90039-6.

[7] E.A. Bayer, R. Lamed, B.A. White, H.J. Flint, From cellulosomes to cellulosomics, Chem. Rec. 8 (2008) 364–377, https://doi.org/10.1002/tcr.20160.

[8] J. Jalak, M. Kurašin, H. Teugjas, P. Väljamäe, Endo-exo synergism in cellulose hydrolysis revisited, J. Biol. Chem. 287 (2012) 28802–28815, https://doi.org/10.1074/jbc.M112.381624.

[9] S. Malgas, M. Thoresen, J.S. van Dyk, B.I. Pletschke, Time dependence of enzyme synergism during the degradation of model and natural lignocellulosic substrates, Enzym. Microb. Technol. 103 (2017) 1–11, https://doi.org/10.1016/j.enzmictec.2017.04.007.

[10] P. Bule, V.M. Pires, C.M. Fontes, V.D. Alves, Cellulosome assembly: paradigms are meant to be broken!, Curr. Opin. Struct. Biol. 49 (2018) 154–161, https://doi.org/10.1016/j.sbi.2018.03.012.

[11] P. Bule, V.D. Alves, V. Israeli-Ruimy, A.L. Carvalho, L.M.A. Ferreira, S.P. Smith, H. J. Gilbert, S. Najmudin, E.A. Bayer, C.M.G.A. Fontes, Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes, Sci. Rep. 7 (2017) 759, https://doi.org/10.1038/s41598-017-00919-w.

[12] Y. Hamberg, V. Ruimy-Israeli, B. Dassa, Y. Barak, R. Lamed, K. Cameron, C.M.G. A. Fontes, E.A. Bayer, D.B. Fried, Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions, PeerJ. 2 (2014), e636, https://doi.org/10.7717/peerj.636.

[13] Q. Xu, M.G. Resch, K. Podkaminer, S. Yang, J.O. Baker, B.S. Donohoe, C. Wilson, D. M. Klingeman, D.G. Olson, S.R. Decker, R.J. Giannone, R.L. Hettich, S.D. Brown, L. R. Lynd, E.A. Bayer, M.E. Himmel, Y.J. Bomble, Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities, Sci. Adv. 2 (2016), e1501254, https://doi.org/10.1126/sciadv.1501254.

[14] B. Dassa, S. Utturkar, R.A. Hurt, D.M. Klingeman, M. Keller, J. Xu, Y.H.K. Reddy, I. Borovok, I. Rozman Grinberg, R. Lamed, O. Zhivin, E.A. Bayer, S.D. Brown, Near-complete genome sequence of the cellulolytic bacterium Bacteroides (Pseudobacteroides) cellulosolvens ATCC 35603, Genome Announc. 3 (2015), https://doi.org/10.1128/genomeA.01022-15.

[15] I. Cann, R.C. Bernardi, R.I. Mackie, Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm, Environ. Microbiol. 18 (2016) 307–310, https://doi.org/10.1111/1462-2920.13152.

[16] L. Artzi, B. Dassa, I. Borovok, M. Shamshoum, R. Lamed, E.A. Bayer, Cellulosomics of the cellulolytic thermophile Clostridium clariflavum, Biotechnol. Biofuels 7 (2014) 100, https://doi.org/10.1186/1754-6834-7-100.

[17] J.J. Adams, M.A. Currie, S. Ali, E.A. Bayer, Z. Jia, S.P. Smith, Insights into higher-order organization of the cellulosome revealed by a dissect-and-build approach: crystal structure of interacting Clostridium thermocellum multimodular components, J. Mol. Biol. 396 (2010) 833–839, https://doi.org/10.1016/j.jmb.2010.01.015.

[18] S. Jindou, J.M. Brulc, M. Levy-Assaraf, M.T. Rincon, H.J. Flint, M.E. Berg, M. K. Wilson, B.A. White, E.A. Bayer, R. Lamed, I. Borovok, Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens, FEMS Microbiol. Lett. 285 (2008) 188–194, https://doi.org/10.1111/j.1574-6968.2008.01234.x.

[19] M.E. Berg Miller, D.A. Antonopoulos, M.T. Rincon, M. Band, A. Bari, T. Akraiko, A. Hernandez, J. Thimmapuram, B. Henrissat, P.M. Coutinho, I. Borovok, S. Jindou, R. Lamed, H.J. Flint, E.A. Bayer, B.A. White, Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of Ruminococcus flavefaciens FD-1, PLoS One 4 (2009), e6650, https://doi.org/10.1371/journal.pone.0006650.

[20] S. Jindou, I. Borovok, M.T. Rincon, H.J. Flint, D.A. Antonopoulos, M.E. Berg, B. A. White, E.A. Bayer, R. Lamed, Conservation and divergence in cellulosome architecture between two strains of Ruminococcus flavefaciens, J. Bacteriol. 188 (2006) 7971–7976, https://doi.org/10.1128/JB.00973-06.

[21] M.T. Rincon, T. Cepeljnik, J.C. Martin, Y. Barak, R. Lamed, E.A. Bayer, H.J. Flint, A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of Ruminococcus flavefaciens, J. Bacteriol. 189 (2007) 4774–4783, https://doi.org/10.1128/JB.00143-07.

[22] V. Israeli-Ruimy, P. Bule, S. Jindou, B. Dassa, S. Moraïs, I. Borovok, Y. Barak, M. Slutzki, Y. Hamberg, V. Cardoso, V.D. Alves, S. Najmudin, B.A. White, H.J. Flint,

H.J. Gilbert, R. Lamed, C.M.G.A. Fontes, E.A. Bayer, Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions, Sci. Rep. 7 (2017) 42355, https://doi.org/10.1038/srep42355.

[23] M.T. Rincon, B. Dassa, H.J. Flint, A.J. Travis, S. Jindou, I. Borovok, R. Lamed, E. A. Bayer, B. Henrissat, P.M. Coutinho, D.A. Antonopoulos, M.E.Berg Miller, B. A. White, Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, Ruminococcus flavefaciens FD-1, PLoS ONE 5 (2010), e12476, https://doi.org/10.1371/journal.pone.0012476.

[24] P. Bule, V.M.R. Pires, V.D. Alves, A.L. Carvalho, J.A.M. Prates, L.M.A. Ferreira, S. P. Smith, H.J. Gilbert, I. Noach, E.A. Bayer, S. Najmudin, C.M.G.A. Fontes, Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction, Sci. Rep. 8 (2018), https://doi.org/10.1038/s41598-018-25171-8.

[25] P. Bule, V.D. Alves, A. Leitão, L.M.A. Ferreira, E.A. Bayer, S.P. Smith, H.J. Gilbert, S. Najmudin, C.M.G.A. Fontes, Single binding mode integration of hemicellulose-degrading enzymes via adaptor scaffoldins in Ruminococcus flavefaciens cellulosome, J. Biol. Chem. 291 (2016) 26658–26669, https://doi.org/10.1074/jbc.M116.761643.

[26] O. Salama-Alber, M.K. Jobby, S. Chitayat, S.P. Smith, B.A. White, L.J.W. Shimon, R. Lamed, F. Frolow, E.A. Bayer, Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses, J. Biol. Chem. 288 (2013) 16827–16838, https://doi.org/10.1074/jbc.M113.466672.

[27] J.L.A. Brás, B.A. Pinheiro, K. Cameron, F. Cuskin, A. Viegas, S. Najmudin, P. Bule, V.M.R. Pires, M.J. Romão, E.A. Bayer, H.L. Spencer, S. Smith, H.J. Gilbert, V. D. Alves, A.L. Carvalho, C.M.G.A. Fontes, Diverse specificity of cellulosome attachment to the bacterial cell surface, Sci. Rep. 6 (2016) 38292, https://doi.org/10.1038/srep38292.

[28] A.L. Carvalho, F.M.V. Dias, T. Nagy, J.A.M. Prates, M.R. Proctor, N. Smith, E. A. Bayer, G.J. Davies, L.M.A. Ferreira, M.J. Romão, C.M.G.A. Fontes, H.J. Gilbert, Evidence for a dual binding mode of dockerin modules to cohesins, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 3089–3094, https://doi.org/10.1073/pnas.0611173104.

[29] M. Duarte, A. Viegas, V.D. Alves, J.A.M. Prates, L.M.A. Ferreira, S. Najmudin, E. J. Cabrita, A.L. Carvalho, C.M.G.A. Fontes, P. Bule, A dual cohesin–dockerin complex binding mode in bacteroides cellulosolvens contributes to the size and complexity of its cellulosome, J. Biol. Chem. 296 (2021), 100552, https://doi.org/10.1016/j.jbc.2021.100552.

[30] M. Levy-Assaraf, M. Voronov-Goldman, I.Rozman Grinberg, G. Weiserman, L.J. W. Shimon, S. Jindou, I. Borovok, B.A. White, E.A. Bayer, R. Lamed, F. Frolow, Crystal structure of an uncommon cellulosome-related protein module from Ruminococcus flavefaciens that resembles papain-like cysteine peptidases, PLoS ONE 8 (2013), e56138, https://doi.org/10.1371/journal.pone.0056138.

[31] J.Y. Weinstein, M. Slutzki, A. Karpol, Y. Barak, O. Gul, R. Lamed, E.A. Bayer, D. B. Fried, Insights into a type III cohesin-dockerin recognition interface from the cellulose-degrading bacterium Ruminococcus flavefaciens, J. Mol. Recognit. 28 (2015) 148–154, https://doi.org/10.1002/jmr.2380.

[32] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402, https://doi.org/10.1093/nar/25.17.3389.

[33] I. Venditto, P. Bule, A. Thompson, J. Sanchez-Weatherby, J. Sandy, L.M.A. Ferreira, C.M.G.A. Fontes, S. Najmudin, Expression, purification, crystallization and preliminary X-ray analysis of CttA, a putative cellulose-binding protein from Ruminococcus flavefaciens, Acta Crystallogr. F Struct. Biol. Commun. 71 (2015) 784–789, https://doi.org/10.1107/S2053230X15008249.

[34] E. Krissinel, Stock-based detection of protein oligomeric states in jsPISA, Nucleic Acids Res. 43 (2015) W314–W319, https://doi.org/10.1093/nar/gkv314.

[35] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R. M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G. N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin, K. S. Wilson, Overview of the CCP4 suite and current developments, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 235–242, https://doi.org/10.1107/S0907444910045749.

[36] M. Slutzki, M.K. Jobby, S. Chitayat, A. Karpol, B. Dassa, Y. Barak, R. Lamed, S. P. Smith, E.A. Bayer, Intramolecular clasp of the cellulosomal Ruminococcus flavefaciens ScaA dockerin module confers structural stability, FEBS Open Bio. 3 (2013) 398–405, https://doi.org/10.1016/j.fob.2013.09.006.

[37] O. Alber, I. Noach, M.T. Rincon, H.J. Flint, L.J.W. Shimon, R. Lamed, F. Frolow, E. A. Bayer, Cohesin diversity revealed by the crystal structure of the anchoring cohesin from Ruminococcus flavefaciens, Proteins 77 (2009) 699–709, https://doi.org/10.1002/prot.22483.

[38] O. Salama-Alber, Y. Gat, R. Lamed, L.J.W. Shimon, E.A. Bayer, F. Frolow, Crystallization and preliminary X-ray characterization of a type III cohesin-dockerin complex from the cellulosome system of Ruminococcus flavefaciens, Acta Crystallogr.Sect. F Struct. Biol. Cryst. Commun. 68 (2012) 1116–1119, https://doi.org/10.1107/S1744309112033088.

[39] M. Voronov-Goldman, M. Levy-Assaraf, O. Yaniv, G. Wisserman, S. Jindou, I. Borovok, E.A. Bayer, R. Lamed, L.J.W. Shimon, F. Frolow, Structural characterization of a novel autonomous cohesin from Ruminococcus flavefaciens, Acta Crystallogr. F Struct. Biol. Commun. 70 (2014) 450–456, https://doi.org/10.1107/S2053230X14004051.

[40] M.T. Rincon, T. Cepeljnik, J.C. Martin, R. Lamed, Y. Barak, E.A. Bayer, H.J. Flint, Unconventional mode of attachment of the Ruminococcus flavefaciens cellulosome to the cell surface, J. Bacteriol. 187 (2005) 7569–7578, https://doi.org/10.1128/JB.187.22.7569-7578.2005.

[41] C. Chen, Z. Cui, X. Song, Y.-J. Liu, Q. Cui, Y. Feng, Integration of bacterial expansin-like proteins into cellulosome promotes the cellulose degradation, Appl. Microbiol. Biotechnol. 100 (2016) 2203–2212, https://doi.org/10.1007/s00253-015-7071-6.

[42] S. Moraïs, D.W. Cockburn, Y. Ben-David, N.M. Koropatkin, E.C. Martens, S. H. Duncan, H.J. Flint, I. Mizrahi, E.A. Bayer, Lysozyme activity of the Ruminococcus champanellensis cellulosome, Environ. Microbiol. 18 (2016) 5112–5122, https://doi.org/10.1111/1462-2920.13501.

[43] L. Bensoussan, S. Moraïs, B. Dassa, N. Friedman, B. Henrissat, V. Lombard, E. A. Bayer, I. Mizrahi, Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome, Environ. Microbiol. 19 (2017) 185–197, https://doi.org/10.1111/1462-2920.13561.

[44] A. Peer, S.P. Smith, E.A. Bayer, R. Lamed, I. Borovok, Noncellulosomal cohesin- and dockerin-like modules in the three domains of life, FEMS Microbiol. Lett. 291 (2009) 1–16, https://doi.org/10.1111/j.1574-6968.2008.01420.x.

[45] A. Mechaly, H.P. Fierobe, A. Belaich, J.P. Belaich, R. Lamed, Y. Shoham, E. A. Bayer, Cohesin-dockerin interaction in cellulosome assembly: a single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition, J. Biol. Chem. 276 (2001) 9883–9888, https://doi.org/10.1074/jbc.M009237200.

[46] D. Nakar, T. Handelsman, Y. Shoham, H.-P. Fierobe, J.-P. Belaich, E. Morag, R. Lamed, E.A. Bayer, Pinpoint mapping of recognition residues on the cohesin surface by progressive homologue swapping, J. Biol. Chem. 279 (2004) 42881–42888, https://doi.org/10.1074/jbc.M407350200.

[47] M.A. Nash, S.P. Smith, C.M. Fontes, E.A. Bayer, Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes, Curr. Opin. Struct. Biol. 40 (2016) 89–96, https://doi.org/10.1016/j.sbi.2016.08.002.

[48] J. Dorival, S. Moraïs, A. Labourel, B. Rozycki, P.-A. Cazade, J. Dabin, E. Setter-Lamed, I. Mizrahi, D. Thompson, A. Thureau, E.A. Bayer, M. Czjzek, Mapping the deformability of natural and designed cellulosomes in solution, Biotechnol. Biofuels 15 (2022) 68, https://doi.org/10.1186/s13068-022-02165-3.

[49] X. Yao, C. Chen, Y. Wang, S. Dong, Y.-J. Liu, Y. Li, Z. Cui, W. Gong, S. Perrett, L. Yao, R. Lamed, E.A. Bayer, Q. Cui, Y. Feng, Discovery and mechanism of a pH-dependent dual-binding-site switch in the interaction of a pair of protein modules, Sci. Adv. 6 (2020), eabd7182, https://doi.org/10.1126/sciadv.abd7182.

[50] A.M. Vera, A. Galera-Prat, M. Wojciechowski, B. Różycki, D.V. Laurents, M. Carrión-Vázquez, M. Cieplak, P. Tinnefeld, Cohesin-dockerin code in cellulosomal dual binding modes and its allosteric regulation by proline isomerization, Structure (2021), https://doi.org/10.1016/j.str.2021.01.006.

[51] P. Bule, K. Cameron, J.A.M. Prates, L.M.A. Ferreira, S.P. Smith, H.J. Gilbert, E. A. Bayer, S. Najmudin, C.M.G.A. Fontes, V.D. Alves, Structure–function analyses generate novel specificities to assemble the components of multi-enzyme bacterial cellulosome complexes, J. Biol. Chem. (2018), https://doi.org/10.1074/jbc.RA117.001241.

[52] A. Karpol, L. Kantorovich, A. Demishtein, Y. Barak, E. Morag, R. Lamed, E.A. Bayer, Engineering a reversible, high-affinity system for efficient protein purification based on the cohesin-dockerin interaction, J. Mol. Recognit. 22 (2009) 91–98, https://doi.org/10.1002/jmr.926.

[53] A. Mechaly, S. Yaron, R. Lamed, H.P. Fierobe, A. Belaich, J.P. Belaich, Y. Shoham, E.A. Bayer, Cohesin-dockerin recognition in cellulosome assembly: experiment versus hypothesis, Proteins 39 (2000) 170–177, https://doi.org/10.1002/(SICI)1097-0134(20000501)39:2<170::AID-PROT7>3.0.CO;2-H.

[54] J.E. Hyeon, D.H. Kang, S.O. Han, Signal amplification by a self-assembled biosensor system designed on the principle of dockerin-cohesin interactions in a cellulosome complex, Analyst 139 (2014) 4790–4793, https://doi.org/10.1039/c4an00856a.

[55] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie, iMOSFLM : a new graphical interface for diffraction-image processing with MOSFLM, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 271–281, https://doi.org/10.1107/S0907444910048675.

[56] W. Kabsch, XDS, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 125–132, https://doi.org/10.1107/S0907444909047337.

[57] K. Cowtan, The buccaneer software for automated model building. 1. Tracing protein chains, Acta Crystallogr. D Biol. Crystallogr. 62 (2006) 1002–1011, https://doi.org/10.1107/S0907444906022116.

[58] P. Emsley, B. Lohkamp, W.G. Scott, K. Cowtan, Features and development of coot, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 486–501, https://doi.org/10.1107/S0907444910007493.

[59] G.N. Murshudov, P. Skubák, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, M.D. Winn, F. Long, A.A. Vagin, REFMAC 5 for the refinement of macromolecular crystal structures, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 355–367, https://doi.org/10.1107/S0907444911001314.

[60] R.P. Joosten, F. Long, G.N. Murshudov, A. Perrakis, The PDB_REDO server for macromolecular structure model optimization, IUCrJ. 1 (2014) 213–220, https://doi.org/10.1107/S2052252514009324.

[61] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L. W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21, https://doi.org/10.1107/S0907444909042073.

[62] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J. H. Morris, T.E. Ferrin, UCSF ChimeraX : structure visualization for researchers, educators, and developers, Protein Sci. 30 (2021) 70–82, https://doi.org/10.1002/pro.3943.