JBC RESEARCH ARTICLE

# Tripartite binding mode of cohesin-dockerin complexes from *Ruminococcus flavefaciens* involving naturally truncated dockerins

Received for publication, January 23, 2025, and in revised form, May 16, 2025 Published, Papers in Press, June 2, 2025,
https://doi.org/10.1016/j.jbc.2025.110325

**Marlene Duarte[1,2], Ana Luísa Carvalho[3,4], Magda C. Ferreira[1,2], Beatriz Caires[1,2], Maria João Romão[3,4], José A. M. Prates[1,2], Shabir Najmudin[5], Edward A. Bayer[6,7], Carlos MGA. Fontes[1,8], and Pedro Bule[1,2,*]**

*From the [1]CIISA—Centre for Interdisciplinary Research in Animal Health, Faculty of Veterinary Medicine, University of Lisbon, Lisbon, Portugal; [2]Associate Laboratory for Animal and Veterinary Sciences (AL4AnimalS), Lisbon, Portugal; [3]UCIBIO, Chemistry Department, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal; [4]Associate Laboratory i4HB–Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, Portugal; [5]Randall Centre for Cell and Molecular Biophysics, Faculty of Life Sciences and Medicine, King's College London, London, UK; [6]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel; [7]Department of Life Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel; [8]NZYTech Genes & Enzymes, Lisbon, Portugal*

Reviewed by members of the JBC Editorial Board. Edited by Joseph Jez

**Polysaccharides in plant cell walls serve as a rich carbon and energy source, yet their structural complexity presents a barrier to efficient degradation. To address this, anaerobic microorganisms like *R. flavefaciens* have developed sophisticated multi-enzyme complexes known as cellulosomes, which enable the efficient breakdown of these recalcitrant polysaccharides. These complexes are assembled through high-affinity interactions between cohesin (Coh) modules in scaffoldin proteins and dockerin (Doc) modules in cellulosomal enzymes. *R. flavefaciens* FD-1 harbors one of the most intricate cellulosomes described to date, comprising over 200 Doc-containing proteins encoded in its genome. Despite substantial research on this cellulosome, the role of a group of truncated but functional dockerins, known as group-2 Docs, remains unclear. In this study, we present a detailed structural and binding analysis of a Coh-Doc complex involving the cohesin from the cell-anchoring scaffoldin ScaE and a group-2 Doc that bears only one of the two $Ca^{+2}$-coordinating loops that characterise the canonical Docs. Our findings reveal a novel tripartite binding mechanism, in which the cohesin can simultaneously bind two distinct dockerin units in three alternative conformations. This discovery provides new insights into the modular versatility of the *R. flavefaciens* cellulosome and sheds light on the mechanisms that enhance its efficiency in polysaccharide degradation.**

Polysaccharides are a highly diverse group of macromolecules with a multitude of biological roles across all domains of life. Among their many functions, complex carbohydrates are the main structural components of the plant cell wall, one of Earth's major carbon and energy reservoirs. However, because plant cell wall polysaccharides are extremely resistant to degradation, this energy source is largely inaccessible (1, 2). To promote carbohydrate depolymerization, several organisms have evolved a group of highly specialized enzymes capable of catalyzing the biosynthesis, modification and degradation of polysaccharides, collectively called carbohydrate active enzymes (CAZymes). Because of the biological significance of carbohydrates, CAZymes have an enormous biotechnological potential, notably in the conversion of lignocellulosic biomass into fermentable sugars to produce biofuels.

Bacteria are prominent producers of CAZymes, which they use to efficiently decompose and extract energy from numerous carbohydrate sources. Although most aerobic bacteria express and secrete CAZymes as free proteins, many anaerobic bacteria have evolved a highly efficient system to deconstruct polysaccharides, involving the assembly of an intricate multi-enzyme complex called cellulosome (Fig. 1). The cellulosome centralizes the cellulolytic efforts of these anaerobic bacteria by promoting enzyme synergy and protein stability, substantially improving degradation efficiency. Cellulosome assembly involves the high affinity and highly specific interactions between two conserved protein modules: cohesins (Coh) and dockerins (Doc) (3, 4). Cohesins are located within non-catalytic scaffold proteins, known as scaffoldins, which provide a structural backbone for cellulosome formation, while dockerins are mainly found at the C-terminus of CAZymes, anchoring these enzymes to the scaffoldins by interacting with the repeating Coh units (5). The specificity and affinity of these Coh–Doc interactions provide the basis for proper cellulosome assembly and function (3, 4).

Although basic cellulosome structure is mostly conserved across bacterial species, each presents its unique cellulosomal architecture with varying levels of complexity (6–10). One of the most elaborate and versatile cellulosomal systems described to date belongs to the rumen bacterium *Ruminococcus flavefaciens* strain FD-1, with a repertoire of 223 Doc-containing proteins. *R. flavefaciens* Docs have been classified into six groups based on primary sequence

* For correspondence: Pedro Bule, pedrobule@fmv.ulisboa.pt.

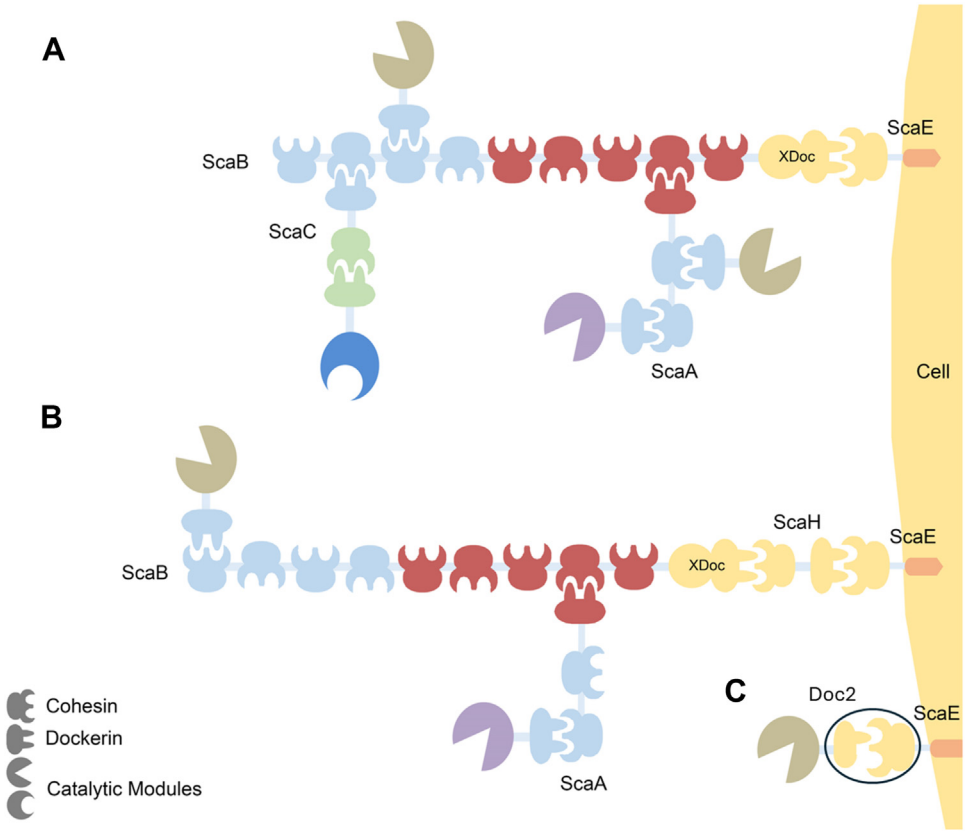


**Figure 1. Schematic representation of the *R. flavefaciens* strain FD-1 cellulosome**. *A*, Scaffoldin ScaB mediates the incorporation of several enzymes into the cellulosome, either directly through Cohs 1 to 4 or indirectly through the interaction of Cohs 5 to 9 with the Doc of ScaA. ScaC is a monovalent adaptor that mostly promotes the incorporation of hemicellulases bearing a Doc from either group 3 or 6. The resultant assembly is tethered onto the cell wall by the interaction of the C-terminal XDoc of ScaB with the cell wall attached CohScaE. *B*, adaptor scaffoldin ScaH can insert itself between ScaB and ScaE, thereby introducing the dual-binding mode mechanism into the *R. flavefaciens* cellulosome. *C*, direct incorporation of cellulosomal components to the bacterial surface can occur through the binding of group-2 dockerins to the cohesin domain of ScaE. Compatible Docs and Cohs are represented in the same colour. The interaction studied in the present work is highlighted with a *dark blue* circle.

homology—a classification system that was shown to translate into function (11–17). Briefly, Group 1 is the largest Doc group, with 96 members that bind to the Cohs of primary scaffoldins ScaA and ScaB and are responsible for the major enzyme recruitment in *R. flavefaciens* cellulosome. Docs from Groups 3 and 6 are mainly appended to hemicellulases that specifically bind to adaptor scaffoldin ScaC, and Group 5 contains the Doc from ScaA as its single member, which binds exclusively to ScaB Cohs. This later interaction thus has a central role in *R. flavefaciens* cellulosomal assembly, allowing the simultaneous integration of up to five ScaA primary scaffoldins to ScaB and incorporation of up to 10 additional enzymes into a single cellulosome. The multi-enzyme complex is then tethered to the bacteria's cell wall by the interaction of a group-4 Doc of ScaB to the single Coh of ScaE, a cell surface-attached scaffoldin. Non-cellulosomal proteins bearing Group 4 Docs can also bind directly to the bacterial cell surface through ScaE recognition, rather than through the primary cellulosomal scaffoldins (Fig. 1) (18, 19). Finally, Group 2 Docs are an intriguing subset of *R. flavefaciens* Docs that resemble truncated versions of Group 4 Docs with a single $Ca^{2+}$-binding repeat (17).

In most cellulosome-producing species, such as *Acetivibrio cellulolyticus* (6), *Clostridium thermocellum* (20) and *Bacteroides cellulosolvens* (21), Docs exhibit a remarkable internal symmetry in both their primary and tertiary structures. This allows them to bind their cognate Cohs in two opposite orientations in what has been termed the dual binding mode (DBM) (22, 23). This ability to adopt two alternative conformations is thought to promote structural flexibility to their assembly, thereby avoiding steric hindrance between the several cellulosomal components (11, 12, 16, 21, 24, 25). In contrast, *R. flavefaciens'* cellulosome was thought to fully rely on single-binding mode (SBM) interactions for its assembly (19), as the vast majority of its Docs do not display the required primary sequence symmetry to allow two alternative binding conformations. However, although *R. flavefaciens'* cellulosome can be fully assembled by SBM interactions, a unique mechanism allowing this cellulosome to selectively introduce the DBM by recruiting a set of highly symmetrical Group 4 Docs has been recently suggested (Fig. 1).

Although *R. flavefaciens'* cellulosomal assembly is now well documented, Group 2 dockerins remain poorly understood. Although highly homologous to Group 4 Docs, this distinctive group was initially thought to be non-functional due to the

naturally truncated structure of its members, which deviates from the conventional architecture of canonical dockerins. Unlike most dockerins, which are about 70 amino acids long and feature a tandem duplication of 22-residue segments and remarkable structural conservation, Group 2 dockerins are significantly shorter, consisting of approximately 40 amino acids. These shorter dockerins include only one of the conserved segments, retaining a single $Ca^{2+}$-binding repeat, while preserving the two critical Coh-recognizing residues at positions 10 and 11 within the single motif (4). Nonetheless, recent studies have shown that despite this truncation, Group 2 Docs retain their ability to bind cohesins with high affinity and a similar specificity to Group 4 Docs, challenging the assumption that both repeats are essential for function. However, the impact of their truncated nature on the binding mechanism remains largely unexplored (17).

This study presents the first structural and thermodynamic analysis of a natural cohesin-dockerin (Coh-Doc) complex involving truncated Docs, aiming to clarify the molecular basis of this unique interaction. To understand the interaction mechanism of Group-2 dockerins, which is crucial for uncovering their specific roles in the complex, the structure of a Coh-Doc complex involving the cohesin from cell-anchoring scaffolding ScaE in complex with a group-2 dockerin (*Rf*CohScaE-Doc2) was determined by X-ray crystallography. The structure allowed the identification of critical Coh recognizing residues while revealing an unprecedented trimer conformation, with two Docs binding simultaneously to a single cohesin unit. Additionally, we explored the possibility to transfer the shorter Doc form to other cellulosomal systems by using Doc2 as a model to produce truncated variants of previously characterized dockerins. Ultimately, these findings represent an expansion on our understanding about cellulosomal assembly strategies, while providing a new blueprint for ~~the~~ engineering tailored enzymatic complexes with enhanced adaptability, paving the way for next-generation cellulosome-based biocatalysts with greater efficiency in plant biomass degradation as well as to other affinity-based technologies for molecular biology and biomedical research.

## Results

### *Primary structure of* R. flavefaciens *Group 2 Docs*

A primary sequence-based analysis organised *R. flavefaciens'* dockerins into six homology groups (17). These were later found to be functionally relevant, with members of each group exhibiting similar Coh-binding preferences [13,18]. Five Doc sequences have been assigned to group 2 (17), although only three sequences were accessible at the time of writing. These include a Doc associated with a module of unknown function (*Rf*Doc2), a Doc from a leucine-rich repeat (LRR) containing protein, and a third Doc associated with a cohesin-like module, reminiscent of an adaptor scaffoldin (17). A primary sequence alignment revealed that these short-length truncated Docs are highly homologous to the C-terminus of group-4 Docs, particularly to subgroup 4a, a group of dockerins involved in cellulosome cell-wall anchoring, some of which were recently shown to be able to bind in a DBM, due to their remarkable internal symmetry (16, 26). Notably, Group 2 Docs retain the same Gly-Arg pair at the canonical Coh-recognition positions (positions 10/11) as Group 4 Docs, suggesting a similar Coh preference (Fig. S1). Despite the similarities, the truncated form of Group 2 Docs suggests a distinct binding mechanism with potential functional implications not yet explored. When a group 2 Doc binds to its corresponding Coh, an additional binding platform may remain accessible. This creates a potential opportunity for a second Group 2 Doc to simultaneously bind to the available site, introducing an additional layer of complexity to the assembly of the *R. flavefaciens* cellulosome.

### *Structure of the* R. flavefaciens *CohScaE–Doc2 complex*

Despite the significant progress towards understanding the structural mechanisms underlining the assembly of *R. flavefaciens* cellulosome, very little is known about the interaction of group-2 Docs with their cognate Cohs. To date, no experimental structures of complexes involving these naturally truncated Docs have been reported. To understand the molecular determinants governing the binding mechanism and the role of these Group 2 Docs within cellulosomal assembly, we have solved the structure of the Coh from the cell-envelope-attached ScaE (*Rf*CohScaE) in complex with a Group 2 Doc (*Rf*Doc2), using X-ray crystallography.

Obtaining the first crystals of the *Rf*CohScaE-Doc2 complex posed a challenging task. Despite a high yield and good stability, initial trials were not successful, prompting further analysis. Careful inspection of the chromatograms from the size-exclusion chromatography purification step was key to adjusting the strategy. The SEC chromatograms revealed two distinct peaks (Fig. S2), which are indicative of assemblies with different molecular sizes within the purified sample. This variability likely impaired the formation of well-ordered crystals, requiring careful fractionation to isolate a homogeneous population suitable for crystallization.

The short length of Group 2 Docs, which possess only one of the dockerin binding helices, suggests that they should only partially occupy the Doc-binding surface of the Coh, leaving the remaining site accessible for potential interaction. It was then hypothesized that the Coh could interact with either a single Doc (1:1 complex) or with two Docs simultaneously (2:1 complex), in a conformation resembling a classic Coh-Doc interaction involving a full-sized Doc. This would mean the coexistence of multiple conformations in the sample, creating conformational heterogeneity that would likely impair the formation of good-quality diffracting crystals. To avoid this, each peak was isolated and crystallized separately, which led to the formation of crystals with the fractions corresponding to the earlier eluting peak (larger molecular weight), likely containing the 2:1 complex. The absence of crystals for the latter eluting peak can be attributed to the heterogeneity resulting from the two possible conformations for the 1:1 complex, both of which are of equal size (Fig. S2). The crystals obtained with the high Mw fractions diffracted to a maximum resolution of

2.6 Å, belonging to space group $P6_4$ with unit cell dimensions of $a$ = 142.390 Å, $b$ = 142.390 Å, $c$ = 57.5 Å, and $\alpha = \beta = 90°$, $\gamma = 120°$. The structure of the *Rf*CohScaE-Doc2 complex was solved following a molecular replacement strategy, using the available model of CohScaE in complex with the Doc of adaptor scaffoldin ScaH (*Rf*CohScaE-DocScaH, PDB: 8AJY). The final model obtained was refined to a resolution of 2.97 Å with a single unit of the Coh-Doc complex in the asymmetric unit. As suspected, two Doc units (chain B – Doc2B, chain C – Doc2C) can be observed interacting with the same Coh, each coordinating one calcium ($Ca^{2+}$) ion, while the Coh coordinates a third one (Fig. 2, *A* and *B*). The final model was

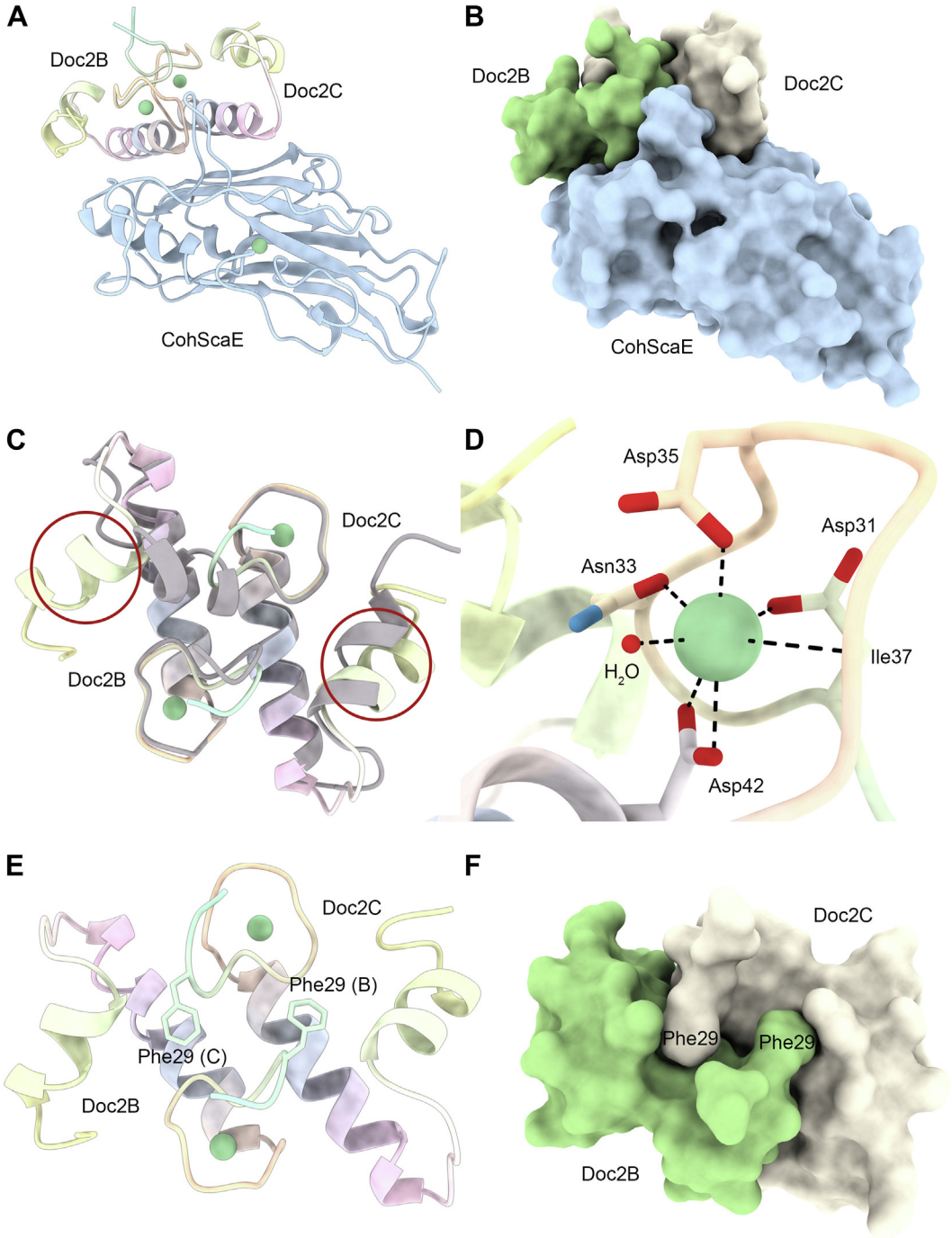


**Figure 2. 3D structure of the *Rf*CohScaE-Doc2 complex**. *A*, structure of the *Rf*CohScaE-Doc2 complex, in ribbon representation, with the Coh in *blue* and the two Doc copies colour-ramped from *green* (N-terminus) to *yellow* (C-terminus). *B*, *Rf*CohScaE-Doc2 complex showing the van der Waals surfaces of the Coh module (*blue*) and the two Doc copies (*green* and *light yellow*, respectively). *C*, overlay of the two *Rf*Doc2 units and group-4 DocScaH XDoc (PBD code 8AJY) with *red* circles highlighting that the second helix of *Rf*Doc2 matches better with the fourth helix of DocScaH, rather than with its second helix. *D*, detailed view of *Rf*Doc2's single calcium-coordinating loop. Calcium-coordinating residues are depicted in stick representation, hydrogen bonds are shown as dashed lines, the water molecules as a *red* sphere, and the calcium ion as a *green* sphere. *E*, the two copies of *Rf*Doc2 in ribbon representation with the Phe29 residue highlighted in stick representation and (*F*) the two Docs with the van der Waals surfaces showing the interlocking mechanisms uniting the four ends.

deposited in the Protein Data Bank under the accession code 9GVB and contains residues 27 to 71 of *Rf*Doc2 (779–823 of Nonredundant RefSeq accession number WP_009984933) and 2 to 200 of CohScaE (30–228 of Nonredundant RefSeq accession number CAK18898). Data collection and refinement statistics are presented in Table 1.

### *Structure of* Rf*Doc2 in the complex*

Docs typically consist of two duplicated calcium-coordinating EF-hand-like motifs, contributing with two α-helices arranged in an antiparallel orientation to form the Coh-recognition surface, with the two motifs connected by a third α-helix (3, 27). However, group-2 Docs are unconventionally short, containing a single calcium-binding repeat. As such, *Rf*Doc2 does not exhibit the characteristic of a highly symmetric 3D Doc fold. Instead, it adopts a simpler helix-loop-helix structure comprising a calcium-coordinating loop, followed by a single Coh-interacting α-helix, a loop, and a second C-terminal α-helix (Fig. 2, *A* and *B*, Fig. S3). Interestingly, although *Rf*Doc2's primary sequence shows slightly higher homology with the N-terminal repeat of the group 4 Doc *Rf*DocScaH (46% vs. 40%), a 3D structure alignment with this Doc (PDB code 8AJY) reveals a closer fit with its C-terminal portion (rmsd 5.8 Å vs. 2.7 Å) (Fig. 2*C*). This is primarily due to the orientation of *Rf*Doc2's second helix, which matches that of *Rf*DocScaH's α-helix 4. α-helix four is a distinctive feature of group-4 Docs, meaning that *Rf*Doc2 only possesses one of the canonical α-helices from the classic Doc fold, followed by an additional helix that is unique to *R. flavefaciens'* group-2 and group-4 Docs. A calcium ion is also seen in the Doc structure, coordinated in a typical pentagonal bipyramid geometry, with the calcium-interacting residues following the typical N, N + 2, N + 4, N + 6, N + 11 pattern (Fig. 2*D*).

As previously mentioned, two copies of *Rf*Doc2 are seen interacting with CohScaE in an antiparallel arrangement that mimics the structure of a classical full-sized Doc. This is possible because *Rf*Doc2 is homologous to the dual-binding-mode Docs of Group 4, giving it the ability to bind to either of the Doc-recognition interfaces on CohScaE. The structures of the two *Rf*Doc2 copies are highly similar, with an rmsd of 0.569 Å across 45 aligned Ca atoms, indicating that *Rf*Doc2 retains its conformation regardless of which portion of the Coh binding site it engages with. A network of 39 non-bonded interactions is formed between the two Doc copies, involving 12 residues from each, which stabilizes the structure and compensates the entropic penalty of the 2:1 vs 1:1 conformation. While most contacts occur between the Coh-binding helices, an N-terminal phenylalanine (Phe29) is seen establishing critical interactions with the C-terminal region of the opposing Doc, potentially stabilizing the complex structure (Fig. 2, *E* and *F*). Interestingly, this phenylalanine is not conserved in either repeat of group-4 Docs, suggesting it may have evolved specifically to support this double-Doc arrangement. Interactions between the N- and C-terminal ends of Docs have been observed in various species, where they form a clasp that stabilizes the structure and contributes to a globular conformation (28). However, in Group 4 Docs, the presence of the additional helix four disrupts this mechanism. Notably, in *Rf*Doc2, the same extra helix, in conjunction with Phe29, appears to restore the stabilizing mechanism by bringing the four ends of the two Docs into proximity (Fig. 2, *E* and *F*).

As expected, a structural similarity search with DALI (http://ekhidna2.biocenter.helsinki.fi/dali/) identified the Group 4 Doc from scaffoldin ScaH (PDB code 8AJY) as the closest functionally relevant structural homologue of *Rf*Doc2 (Table S1), with a Z-score of 6.1, an rmsd of 1.9 Å, and 36% identity over 44 aligned backbone residues (26). The only other functionally relevant structural homolog of *Rf*Doc2 identified with the DALI search was the Doc from the primary *R. flavefaciens* scaffoldin ScaA (PDB code 5N5P) (Table S1), with a Z-score of 2.4, an rmsd of 2.3 Å, and 23% identity over 30 aligned residues, which mediates interactions between the major scaffoldins ScaA and ScaB (23). Most other DALI hits corresponded to nucleic acid-binding proteins, likely due to the helix-turn-helix conformation of *Rf*Doc2 (29, 30).

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

| Crystal | *Rf*CohScaE-Doc2 |
|---|---|
| Space Group | $P6_4$ |
| Unit cell parameters *a*, *b*, *c* (Å) | 142.39, 142.39, 57.596 |
| α, β, γ (°) | 90, 90, 120 |
| Data collection statistics | |
| X-ray source | ESRF ID23–1 |
| Wavelength, Å | 0.97626 |
| Total \| unique no. of reflections | 13,690 (2765) |
| Resolution limits, Å | 71.19–2.971 (3.2–2.97) |
| Completeness, % | 97.87 (99.00) |
| Wilson B factor, $Å^2$ | 31.01 |
| Matthews coefficient, $Å^3$/Da | 2.25 |
| Solvent content, % | 45.51 |
| Average I/σ(I) | 15.2 |
| R-merge(I)[a] | 0.039 |
| R-p.i.m.[b] | 0.028 |
| CC(1/2) | 0.998 |
| Structure refinement statistics | |
| R-work[c] | 0.1986 (0.2814) |
| R-free[c] | 0.2133 (0.3041) |
| No. of protein residues in the asymmetric unit | 286 |
| No. of water molecules in the asymmetric unit | 16 |
| No. of non-H atoms in the asymmetric unit | 2182 |
| R.M.S.Z., bond length, Å | 0.011 |
| R.M.S.Z., bond angles, ° | 2.15 |
| Average temperature factor, $Å^2$ | |
| Macromolecules | 37.42 |
| Calcium atoms | 28.34 |
| Solvent | 12.52 |
| *Ramachandran plot* | |
| Residues in favoured regions, % | 92.50 |
| Residues in allowed regions, % | 6.79 |
| Residues in forbidden regions, % | 0.71 |
| PDB ID | 9GVB |

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 $R_{\mathrm{merge}}$

[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.

## *R. flavefaciens'* Coh-Doc complex involving two Doc copies

### *Structure of CohScaE in the complex*

The structure of CohScaE has been previously reported as part of two Coh-Doc complexes involving group-4 Docs (CohScaE-DocScaH, PDB code 8AJY and CohScaE-XdocCttA, PDB code 4IU2), one of which is with an associated X-module (Xdoc). The CohScaE structure of this *Rf*CohScaE-Doc2 complex is nearly identical to the ones reported for the other two complexes, as indicated by a low root mean square deviation (rmsd) of 0.5 Å over 199 backbone atoms. The fact that CohScaE maintains its structure, regardless of whether it binds to a classical Doc, an XDoc, or two truncated Docs simultaneously, highlights its structural stability and seems to be ligand independent. Apart from some previously reported unique motifs, like a calcium-coordinating loop and a large α-helix between strands eight and 9, CohScaE possesses a 9-stranded β-sandwich jellyroll topology, with four strands in the Doc-interacting face and five strands in the "back" face, similar to all reported Cohs (Fig. 2*A*) (26, 31). The closest functionally relevant homologue of CohScaE is the isolated autonomous Coh of ScaG, which is also capable of interacting with group-4 Docs (32).

### *CohScaE*–Rf*Doc2 complex interface*

The double Doc interaction observed in the CohScaE–RfDoc2 complex represents a unique binding profile among known Coh-Doc complexes. In this arrangement, two copies of the truncated Doc (chain B, Doc2B, and chain C, Doc2C) simultaneously bind their cognate Coh, resulting in all Doc residues at the Coh interface being repeated. This contrasts with DBM Docs, which, while displaying remarkable internal symmetry with key residues repeated at equivalent positions, show minor differences in composition between their Coh-contacting helices.

The residues used by *Rf*Doc2 to engage one Coh binding site differ from those used for the second site (Fig. S4). This is analogous to full DBM Docs, where the residues involved in the binding alternate between helices 1 and 3, depending on the Doc orientation. The binding interface is defined by the N-terminal helices of each Doc and by β-strands 5-6-3-8 on the Coh surface. The Coh surface at the binding interface is mostly flat, with a prominent loop extending between the two Doc helices and contributing to binding *via* extensive contacts with both helices—similar to what is observed in Group 4 Doc complexes.

Each Doc establishes an extensive network of polar and hydrophobic interactions at the Coh interface, forming two distinct interaction networks for the two binding sites (Fig. 3, Table 2). Despite maintaining a comparable number of stabilizing interactions at both Coh sites, *Rf*Doc2 utilizes different residues for binding at each site. Specifically, at the first binding site, direct hydrogen bonds involving Arg41 and Thr44 of Doc2B mediate the interaction, whereas at the second binding site, Tyr50 and Ser54 of Doc2C play the primary role.

A comparative analysis with the group-4 Doc, DocScaH (26), highlights significant similarities in the residues employed for Coh binding, suggesting a conserved interaction strategy despite their structural differences (Fig. 4). Like in the CohScaE–DocScaH, the binding is dominated by hydrophobic interactions, particularly involving Leu47, Tyr50, and Thr44 within a conserved hydrophobic patch. Furthermore, as in the DocScaH complex, neither Doc copy dominates the interaction; instead, both lie flat and parallel atop the Coh binding surface. This configuration contrasts with typical type-I Coh-Doc complexes, where the non-dominant helix often appears slightly detached from the binding interface (33).

### *CohScaE*–Rf*Doc2 binding mechanism*

The contribution of each key Doc residue identified in the 3D complex structure was assessed using site-directed mutagenesis and ITC. Residues forming direct hydrogen bonds or participating in extensive hydrophobic contacts were mutated to alanine, and the resulting variants were tested against CohScaE (Table 2). Alanine was selected because it removes the side chain at the β-carbon while preserving the main-chain conformation. However, Ser54 was mutated to tryptophan to induce a more pronounced effect; substituting Ser54 with alanine would likely result in minimal structural changes, providing limited insight into its functional role. By contrast, introducing tryptophan creates substantial steric hindrance and electronic changes due to its larger aromatic side chain.

ITC data (Table 3, Fig. S2) revealed a strong interaction between the wild-type Doc and Coh with an association constant (Ka) of $6.86 \times 10^{7}$ $M^{-1}$ ± $6.77 \times 10^{6}$ $M^{-1}$. Among all variants, mutations in residues Arg41 and Tyr50 had the most significant impact on binding, reducing Ka values by nearly 100-fold, followed by Leu47 and Ser54. These findings align with the 3D structure, where these residues form numerous polar and hydrophobic contacts with the Coh. Close inspection of the interactions established by Arg41 and Ser54 highlights their distinct roles depending on their binding orientations: Arg41 seems to primarily mediate interactions at one CohScaE binding site, while Ser54 assumes a similar role at the other. This intriguing dichotomy highlights the flexibility within the molecular interaction landscape, where the same Doc employs different residues to bind distinct Coh sites. To confirm this, a double mutant targeting both Arg41 and Ser54 was created, which resulted in a complete loss of binding by compromising both binding orientations, emphasizing the critical importance of these residues (Table 3, Fig. S3). Accurately interpreting the binding stoichiometry based on the N-values obtained by ITC requires an accurate determination of the concentration of active molecules. In our experience, this can be very challenging in the case of Coh-Doc interactions, since the structural stability of the unbound dockerins is usually low, making the measured concentration to differ significantly from the "active" concentration. Classical dockerins interact with a 1:1 stoichiometry, so active concentration can be inferred by adjusting the Doc concentration, when processing the ITC results, until N = 1 is obtained. Because the Doc is in the

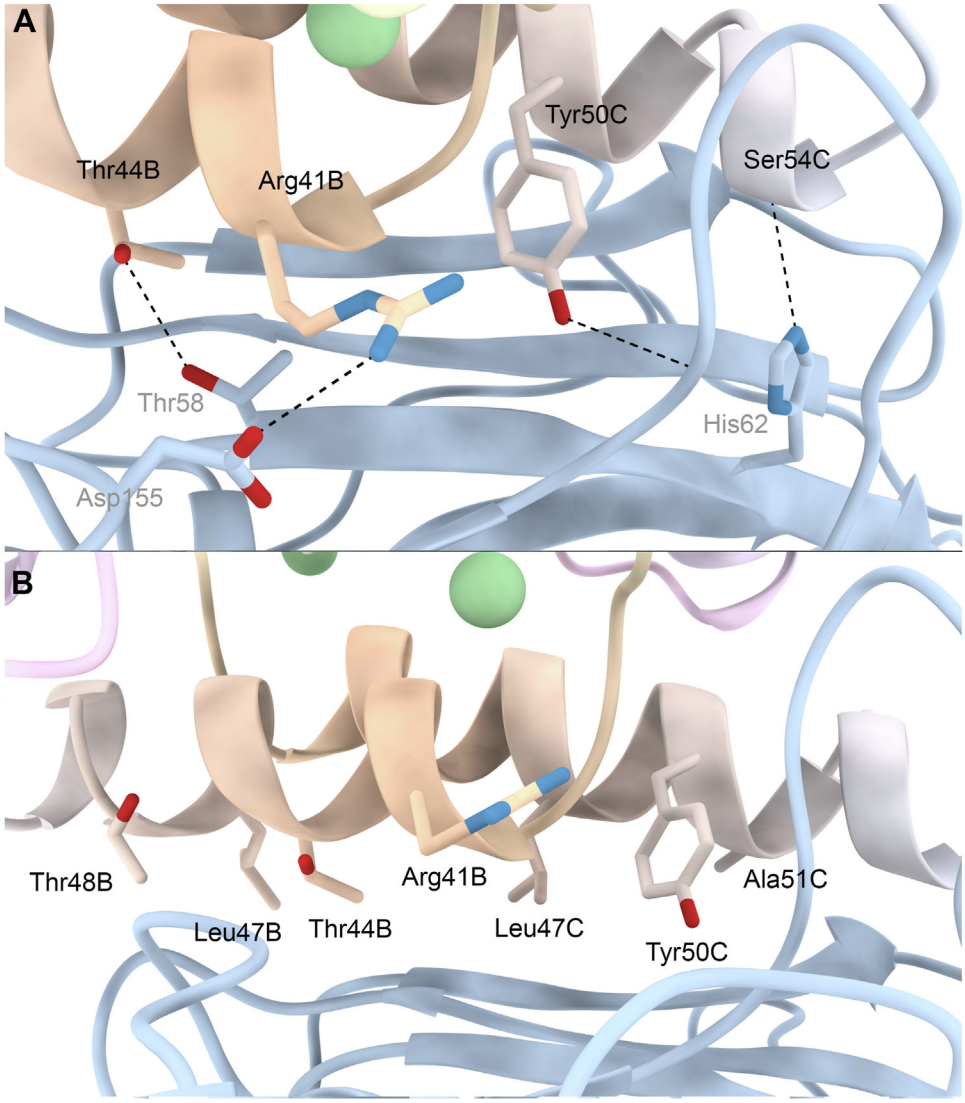


**Figure 3. Detailed view of the *Rf*CohScaE-Doc2 binding interface.** Panels (*A*) and (*B*) highlight the main polar contacts and the main hydrophobic contacts, respectively, involved in Coh-Doc recognition. In panel (*A*), the amino acid residues involved in the main polar interactions are depicted in stick configuration. Dashed *black* lines represent hydrogen-bond interactions. In panel (*B*), the main Doc residues making hydrophobic contacts with the Coh are depicted in stick configuration. The *Rf*CohScaE-Doc2 complex is shown in ribbon representation, with the two Doc copies colour-ramped from *blue* to *orange* and the Coh in *blue*. $Ca^{2+}$ ions are depicted as *green* spheres. The Doc residues are labeled according to which copy of *Rf*Doc2 they belong to (either chain *B* or *C*).

experimental cell while the Coh is in the syringe, this adjustment has little to no impact on the binding constant and thermodynamic parameters. With Doc2, due to the possibility of a 2:1 binding mode, this is no longer a valid approach. Nonetheless, assuming that the Arg41 mutation forces a specific Doc orientation and therefore a 1:1 interaction, we performed this adjustment while processing the data for the Arg41 mutant interaction and used the obtained active Doc concentration to calculate the N values for all variants. Interestingly, they were all close to one suggesting a similar stoichiometry for all interactions.

### *Doc 2 as a blueprint for the design of short-length Docs*

Considering the functional nature of *R. flavefaciens'* truncated Group 2 Docs, it remains unclear whether the duplicated structure observed in most other Docs is essential for Coh recognition. Using a smaller module for cellulosome assembly while maintaining functionality would be a more cost-effective strategy both not only in biological systems but also for the production of designer cellulosomes for biomass conversion. To explore whether the truncated format could be transposed to other Coh-Doc interactions, the structure of *Rf*Doc2 was used as a template to design truncated versions of Docs from various cellulosome-producing species. The goal was to evaluate whether these engineered single-repeat Docs could retain functionality. Four Docs with confirmed dual-binding mode and inherent internal symmetry were selected from four different species: *R. flavefaciens* DocScaH (PDB code 8AJY) (26), *B. cellulosolvens* DocCel48 (PDB code 2Y3N) (21), *A. cellulolyticus* DocCel5 (PDB code 5NRM) (34), and *C. thermocellum* DocXyn10 B (PDB code 2CCL) (22). Thus, well documented

**Table 2**
**Main interactions between CohScaE binding sites and the two *Rf*Doc2 units (chain B - Doc2B, chain C - Doc2C)**

| Direct hydrogen bonds | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ChainB-Doc2B | | | CohScaE | | | | | ChainC-Doc2C | |
| Residue | Atom | Distance (Å) | Residue | Atom | Residue | Atom | Distance (Å) | Residue | Atom |
| ARG41 | NH1 | 3.20 | ASP155 | OD2 | GLY147 | O | 2.66 | TYR50 | OH |
| THR44 | OG1 | 3.16 | THR58 | OG1 | HIS62 | NE2 | 2.62 | SER54 | O |

| Salt Bridges | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| ChainB-Doc2B | | | CohScaE | | | | | ChainC-Doc2C | |
| Residue | Atom | Distance (Å) | Residue | Atom | Residue | Atom | Distance (Å) | Residue | Atom |
| ARG41 | NH1 | 3.20 | ASP155 | OD2 | - | - | - | - | - |

| Nonpolar interactions/Non bonded contacts | | | |
|---|---|---|---|
| ChainB-Doc2B | CohScaE | | ChainC-Doc2C |
| Residue | Residues | | Residue |
| ILE37 | PRO144, SER145 (5) | LEU91 | ALA43 |
| ILE38 | SER145, LYS146 | LEU91 | THR44 |
| GLY40 | THR58, LEU149 (2) | LEU91 (3), ALA92, ALA104 | LEU47 |
| ARG41 | THR58, LEU149 (3), ASP155 (4) | LYS93 (2) | THR48 |
| THR44 | THR58 (4), GLY106, ALA107 (2) | PHE102 (3), GLY147 (4), ASP148 (2), LEU149 (2) | TYR50 |
| LEU47 | SER90 (3), LEU91 (2), ALA107 (3), ASP108 | GLU95 (5), PHE102 (2) | ALA51 |
| THR48 | ALA107, ASP108 (4), ASN153 (3) | GLU95 (2) | LYS52 |
| ALA51 | ASP108 | | |
| GLN68 | LYS154 | | |
| ASN69 | LYS154 (2) | | |

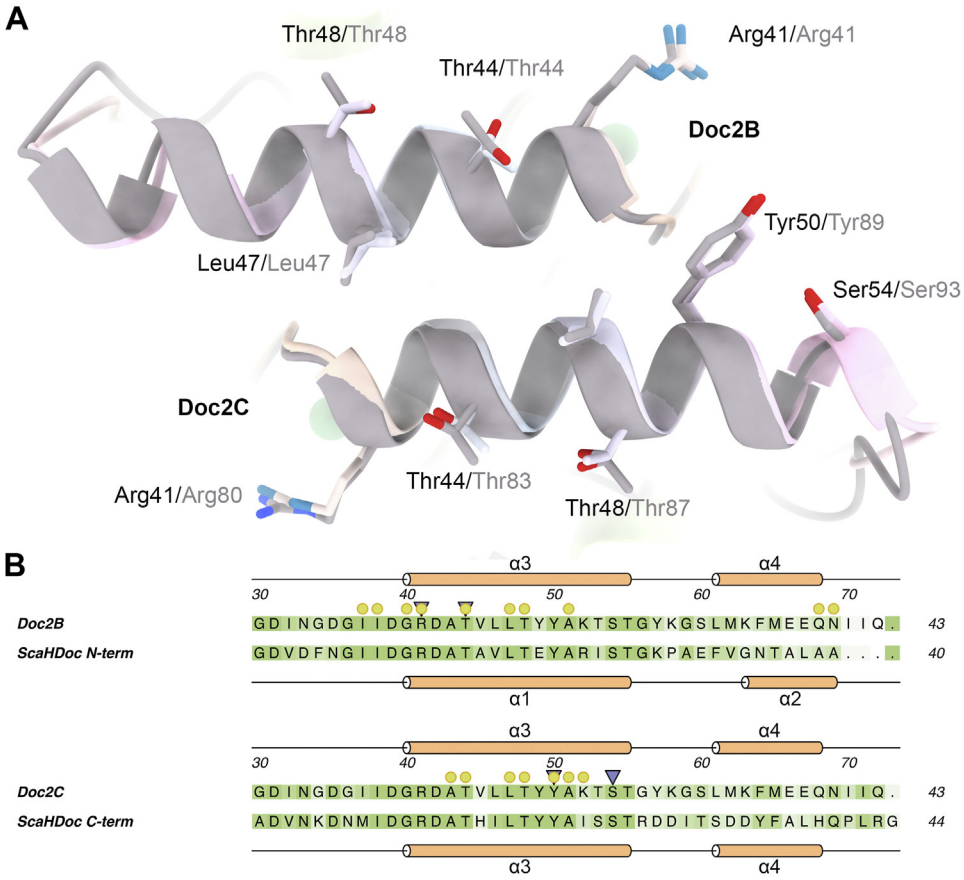


**Figure 4. Conservation of Coh binding residues between group 4 and group 2 Docs**. *A*, overlay of DocScaH with the two *Rf*Doc2 copies in the *Rf*CohScaE-Doc2 complex. DocScaH is seen in *gray* while both copies of *Rf*Doc2 are colour ramped from *yellow* to *pink*. Key Coh-contacting residues are shown in stick representation and are labelled in *black* for *Rf*Doc2 and in *gray* for DocScaH. Chain B *Rf*Doc2 is seen matching with DocScaH's α-helix one while Chain C *Rf*Doc2 matches DocScaH's α-helix 3. *B*, primary structure alignment between Doc2B and the N-terminal half of DocScaH and between Doc2C and the C-terminal half of DocScaH. Hydrogen bonding residues are indicated by a *blue* arrow while residues making hydrophobic contacts with the Coh are indicated by a *yellow* circle. The sequences are coloured according to conservation from *white* (*low*) to *green* (*high*). The secondary structure of each sequence is displayed in *orange* either above or below the corresponding sequence. This comparison highlights how the key Coh-contacting residues of both DocScaH α-helices one and three are present at the same relative positions in the single α-helix of *Rf*Doc2.

**Table 3**
**Thermodynamics of interaction between CohScaE and *Rf*Doc2 variants**

| RfDoc2 variants | Ka $M^{-1}$ | ΔG° kcal $mol^{-1}$ | ΔH kcal $mol^{-1}$ | −TΔS° kcal $mol^{-1}$ | N |
|---|---|---|---|---|---|
| WT | 6.86E7 ± 6.77E6 | −30.572 | −20.8 ± 0.132 | 9.772 | 1.04 ± 0.002 |
| R41 A | 1.55E6 ± 2.98E5 | −8.731 | −7.605 ± 0.393 | 1.126 | 0.94 ± 0.033 |
| T44 A | 1.28E8 ± 2.77E7 | −25.598 | −18.5 ± 0.155 | 7.098 | 0.99 ± 0.003 |
| L47 A | 1.06E7 ± 8.46E5 | −26.887 | −18.4 ± 0.171 | 8.487 | 1.05 ± 0.005 |
| T48 A | 1.23E8 ± 4.93E7 | −12.595 | −12.0 ± 0.186 | 0.595 | 1.12 ± 0.007 |
| Y50 A | 2.15E6 ± 2.59E5 | −28.968 | −18.95 ± 0.475 | 10.018 | 1.13 ± 0.017 |
| S54 W | 1.24E7 ± 1.12E6 | −25.093 | −17.54 ± 0.233 | 7.553 | 0.777 ± 0.007 |
| R41 A + S54 W | Nb | Nb | Nb | Nb | Nb |
| T55 A | 1.64E8 ± 5.64E7 | −31.91 | −21.75 ± 0.226 | 10.16 | 1.02 ± 0.004 |

Docs that function in a dual-binding mode were selected. This is because these dockerins usually have a dominating Coh-interacting helix (either α-helix one or 3, depending on the orientation of the binding), while single-binding mode Docs interact with the cohesin using both α-helices one and 3. Therefore, loss of function after truncation would be more likely on a single binding mode dockerin. Using a 3D structural alignment of the selected Docs with *Rf*Doc2, a 'cut-site' was identified to generate N-terminal and C-terminal truncated variants for each Doc (Fig. S4). The binding capabilities of these 'half Docs' were evaluated against their cognate Coh modules using ITC (Table 4, Fig. 5).

Both the N-terminal truncated variant of *C. thermocellum*' DocXyn10 B and the C-terminal variant of *B. cellulosolvens*' DocCel48 retained their Coh-binding capability with a similar affinity as their wild-type versions. Additionally, *A. cellulolyticus*' DocCel5 N-terminal variant was also able to bind its Coh counterpart, albeit with a 100-fold reduction in the affinity constant. The other engineered Docs failed to recognise their binding partners. This result was particularly unexpected for the C-terminal 'half-Doc' of *Rf*DocScaH, given its high homology to *Rf*Doc2, emphasizing the distinct functional role of these truncated Docs within *R. flavefaciens* cellulosome complex. The absence of stabilizing features, such as the helix–4–Phe29 clasp observed in *Rf*Doc2, could explain the results with *Rf*DocScaH.

## Discussion

With one of the most intricate cellulosomes described to date, *R. flavefaciens* has unveiled numerous surprising strategies for cellulosomal assembly, significantly expanding our understanding of Coh-Doc interactions in recent years. This bacterium can assemble an entire cellulosome exclusively through single-binding mode interactions (19), features specificity-switching adaptor scaffoldins that expand the enzyme repertoire bound to the cellulosome (35) and possesses adaptor scaffoldins that selectively introduce the dual-binding mode while functioning as variable-length spacers (26). These characteristics have reshaped the original paradigm established by the cellulosome of *C. thermocellum*.

Group-2 Docs, a unique class of short-length Docs from *R. flavefaciens*, were initially suspected to be non-functional partial Docs (17). Through a series of orthogonal techniques, they were later shown to bind to the cell-bound scaffoldin ScaE, directing their associated proteins to the bacterial surface. These represent yet another unusual assembling strategy in *R. flavefaciens* cellulosome. However, their binding mechanism remains elusive (16).

### R. flavefaciens *Group 2 Docs can interact through a tripartite binding mode*

Despite their shorter length compared to canonical Docs, Group 2 Docs demonstrate robust binding affinity and specificity for their cognate Coh partners (16). Nevertheless, their smaller size suggests an alternative mode of interaction.

In this study of the *Rf*CohScaE–Doc2 complex, a single Coh domain was observed binding two Doc modules simultaneously—an interaction mode not previously reported. While most Doc residues involved in binding interact with the Coh regardless of the binding platform they occupy, their

**Table 4**
**Thermodynamics of interaction between the truncated dockerin variants of *R. flavefaciens* DocScaH (PDB code 8AJY), *Bacteroides cellulosolvens* DocCel48 (PDB code 2Y3N), *Acetivibrio cellulolyticus* DocCel5 (PDB code 5NRM), and *Clostridium thermocellum* DocXyn10 B (PDB code 2CCL) with their binding partners**

| Dockerin | Ka $M^{-1}$ | ΔG° kcal $mol^{-1}$ | ΔH kcal $mol^{-1}$ | −TΔS° kcal $mol^{-1}$ | N |
|---|---|---|---|---|---|
| RfDocWt | 1.35E7 ± 3.68E6 | −89.55 | −49.8 ± 1.062 | 39.75 | 1.05 ± 0.015 |
| RfDocN-term | Nb | Nb | Nb | Nb | Nb |
| RfDocC-term | Nb | Nb | Nb | Nb | Nb |
| BcDocWt | 2.24E7 ± 4.33E6 | −21.95 | −16.16 ± 0.253 | 5.79 | 0.94 ± 0.006 |
| BcDocN-term | Nb | Nb | Nb | Nb | Nb |
| BcDocC-term | 1.08E7 ± 1.26E6 | −50.16 | −30.04 ± 0.244 | 20.12 | 1.00 ± 0.005 |
| AcDocWt | 1.58E6 ± 3.85E5 | −6.7 | −7.72 ± 0.312 | −1.02 | 0.97 ± 0.028 |
| AcDocN-term | 2.42E4 ± 622 | −38.48 | −22.33 ± 0.885 | 16.15 | 1.03 ± 0.035 |
| AcDocC-term | Nb | Nb | Nb | Nb | Nb |
| CtDocWt | 1.78E6 ± 3.23E5 | −82.19 | −47.68 ± 1.335 | 34.51 | 1.05 ± 0.022 |
| CtDocN-term | 2.14E6 ± 2.74E5 | −77.17 | −45.12 ± 0.685 | 32.05 | 1.11 ± 0.012 |
| CtDocC-term | Nb | Nb | Nb | Nb | Nb |

Nb, no binding.

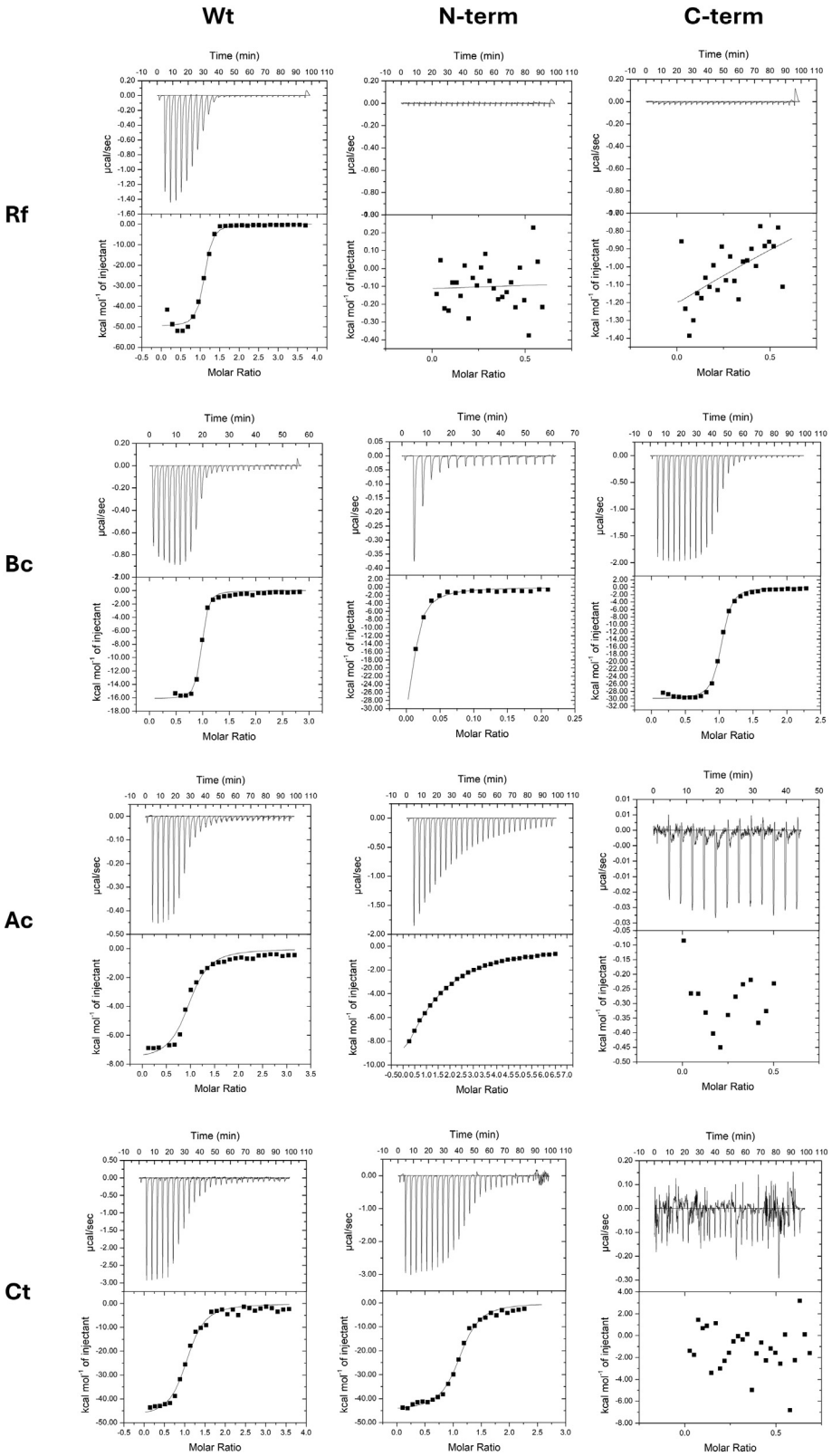


**Figure 5. Binding isotherms for the interaction between the truncated dockerin variants of *R. flavefaciens* DocScaH (PDB code 8AJY), *Bacteroides cellulosolvens* DocCel48 (PDB code 2Y3N), *Acetivibrio cellulolyticus* DocCel5 (PDB code 5NRM), and *Clostridium thermocellum* DocXyn10 B (PDB code 2CCL) with their binding partners**. The *upper* part of each panel shows the raw heats of binding, whereas the *lower* parts comprise the integrated heats after correction for heat of dilution. The curve represents the best fit to a single-site binding model.

contributions differ. Specifically, residues Arg41 and Thr44 dominate the interaction at the first Coh binding site, whereas residues Tyr50 and Ser54 mediate binding at the second site. This suggests that the Docs utilize distinct residues at different binding sites to enable the simultaneous binding of two Doc molecules. Mutating either Arg41 or Ser54 independently reduced binding affinity, likely due to disruption at one of the two Coh binding platforms. However, complete loss of binding was observed when both residues were mutated, as interactions with both Coh binding sites were compromised.

The resolution of two distinct Coh-Doc peaks in size exclusion chromatography indicates the presence of two differently sized complexes in the purified sample. We propose that the smaller peak represents a Coh-Doc complex with a single bound Doc, in contrast to the larger complex where two Docs are bound to a single Coh. The heterogeneity of the 1:1 complex likely results from the Doc binding to either of the two Coh binding sites, which may also explain its crystallization challenges. This observation suggests a tripartite binding mechanism for *R. flavefaciens* group-2 Docs, where the configurations include two Docs bound to a single Coh, a single Doc bound at the first binding site, or a single Doc bound at the second binding site (Fig. 6*A*). This tripartite binding mode differs from the classic dual-binding mode in that the Doc must switch binding sites to adopt a 180°-opposing orientation, enabling two separate Docs to bind simultaneously to the same Coh. The introduction of the tripartite binding mechanism could play a significant role in enhancing the efficiency of the cellulosome system, resulting in a more compact arrangement of enzymes within the complex and a more efficient use of the same amount of scaffoldin, reducing the need for additional protein production, and making the system more cost-effective. In practice, this means that the same amount of scaffoldin with a fixed number of Coh units can bind double the number of Docs and, consequently, double the number of enzymes compared to the classic Doc system.

### *Truncation of classical Docs can generate functional single-repeat Doc versions*

We investigated whether the novel binding mechanism observed in group-2 Docs is exclusive to these inherently short Docs or if it can be generalized to other Doc sequences. To explore this, we conducted a structure-guided cleavage of

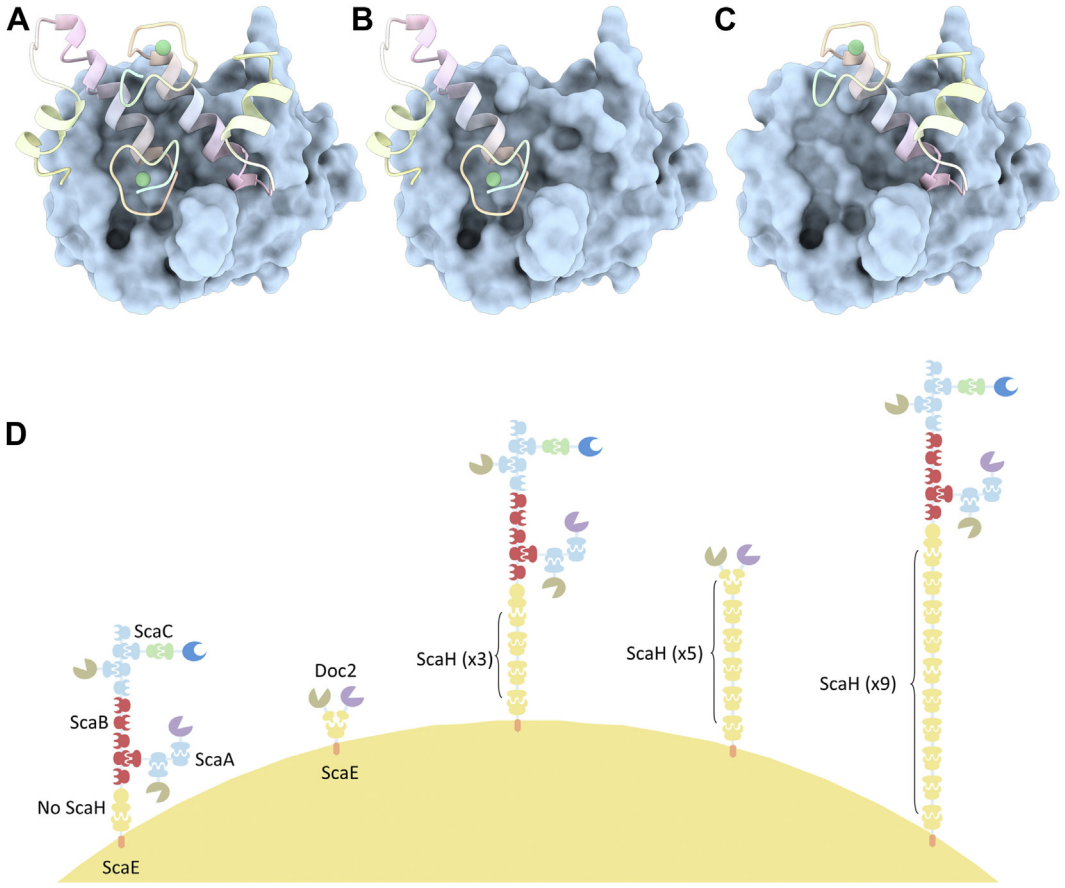


**Figure 6. Unusual assembly strategies of the *R. flavefaciens* cellulosome.** Present results suggest that the complex formed between *Rf*Doc2 and CohScaE can adopt 3 distinct spatial interactions: one with two Doc copies binding simultaneously to the Coh (*A*), whose crystal structure is described in this work; one with a single Doc binding to one of the Coh binding platforms (*B*); and one with one Doc binding to the other Coh binding platform (*C*). Panel (*D*) summarizes the unique strategies of the *R. flavefaciens* cellulosome assembly, including the specificity switching adaptor scaffoldin ScaC, the selective introduction of the DBM by ScaH, which also functions as a variable length adaptor, positioning different cellulosome copies at different distances from the cell wall, and finally, the newly described double dockerin interaction involving group-2 dockerins.

dual-binding mode Docs from four different cellulosome-producing species into two distinct fragments corresponding to their N-terminal and C-terminal regions. The Docs analysed in this study included *C. thermocellum* DocCel48S (*Ct*DocCel48) (22), *B. cellulosolvens* DocCel48 (*Bc*DocCel48) (21), *A. cellulolyticus* DocCel5 (*Ac*DocCel5) (34), and *R. flavefaciens's* DocScaH (*Rf*DocScaH) (26). This approach allowed us to assess whether the truncated fragments retained the capacity to bind their respective Coh partners, akin to group-2 Docs.

Interestingly, with the exception of *R. flavefaciens'* group 4 DocScaH, it was possible to generate functional truncated versions for all the Docs. Two of those, namely the N-terminal portion of *Ct*DocCel48 and the C-terminal portion of *Bc*DocCel48, were able to match the affinity of their wild-type counterparts, while the N-terminal portion of *Ac*DocCel5 interacted with its cognate Coh with a 100-fold decrease in affinity. While ITC also suggests that the N-terminal portion of *Bc*DocCel48 might retain some Coh binding capability, the error values are very large, likely due to structural instability, as suggested by the considerable aggregation observed in the Doc solution.

While these results confirm that the functional nature of Group 2 Docs can be replicated in other cellulosomal systems, the range of different results shows that it cannot be generalized. While some truncations yielded completely functional short dockerins, it only worked either on the N-terminal portion or the C-terminal portion, despite the original Docs having a dual binding mode, in which binding can be mediated either by the N-terminal half or the C-terminal half. Likewise, most truncated variants resulted in partial or complete loss of function, which suggests that there is a biological advantage to having both repeats. The larger format likely results in increased structural stability while also being more resilient to mutation events. Also, the higher number of Coh contacting residues may have implications in specificity, which is crucial to manage cellulosomal architecture and composition.

Unexpectedly, despite structural similarities with group-2 Docs, the C-terminal fragment of *R. flavefaciens* DocScaH failed to retain binding when isolated from the rest of the module. This suggests that group-2 Docs may have evolved unique adaptations to maintain their binding capabilities in their truncated form. While these Docs may exhibit extensive structural similarities, subtle sequence variations, such as the Phe29 residue in *Rf*Doc2, which interacts with the C-terminal helix of the opposing Doc unit and potentially stabilizes the structure, appear to significantly influence their binding functionality. Further studies are required to pinpoint these critical structural motifs and assess whether they can be engineered into other cellulosomal systems.

Nonetheless, from a bioengineering point of view, these results open a new avenue for the design of tailor-made cellulosomes. The possibility to generate artificially truncated Docs that are fully functional means that we are no longer limited to the identification of naturally truncated Docs to design more cost-efficient cellulosomes based on short Docs with multiple Coh-Doc specificities that allow organizing the enzymes in a precise order within the artificial scaffoldin.

### *Concluding remarks*

The intricate complexity of the *R. flavefaciens* cellulosome is well recognized, and the unique binding mechanism exclusive to group-2 Docs, as described here, further emphasizes this sophistication (Fig. 6*B*). While the classical dual-binding mode enables a single Doc to interact with a Coh in two distinct orientations, the newly identified double-Doc tripartite binding mode offers enhanced flexibility and potentially increases the catalytic capacity of the cellulosome complex, underscoring the remarkable structural and functional adaptability of the *R. flavefaciens* cellulosome.

As far as the authors' knowledge, the only other naturally truncated Doc ever reported is that of *A. cellulolyticus'* ScaI scaffoldin (36). However, no functional or structural studies have been conducted on this Doc, which shares very low homology with *Rf*Doc2 (16.3% identity). As a result, its role within the *A. cellulolyticus* cellulosome remains unclear. Furthermore, ScaI lacks an N-terminal signal peptide, suggesting it might be the product of incomplete gene sequencing. Consequently, group-2 Docs of *R. flavefaciens* remain the only known fully functional, naturally truncated Docs.

From an evolutionary standpoint, the symmetrical conformation observed in classical Docs is thought to have originated from gene duplication events (27, 37, 38). However, whether group-2 Docs represent ancestral precursors of modern Docs or are products of gene truncation remains unclear. The fact that group-2 Docs are functional suggests that such duplications may not be strictly necessary for binding. The inability of the engineered truncated DocScaH to bind CohScaE further underscores the specialized nature of the naturally truncated group-2 Docs, implying that these proteins have evolved specific adaptations to maintain function despite their shorter architecture. But while gene duplication and symmetry may not be essential for binding, the retention of duplicated regions may provide advantages, such as enhanced binding resilience in the event of mutations, increased structural stability, or even increased specificity.

The high-affinity and high-specificity interaction between Coh and Doc modules has gathered significant interest for its potential use in affinity-based molecular biology and diagnostic techniques, as well as for the development of tailor-made cellulosomes for biomass conversion. The short dockerin format described here provides a new model for the development of these biotechnological tools. For example, previous efforts have explored Doc-based fusion tags for the purification of recombinant proteins using immobilized Cohs. Using shorter Docs would be advantageous as they are less likely to disturb proper folding of the recombinant protein and may be easier to elute from the immobilized Coh support. The naturally truncated Group 2 Docs from *R. flavefaciens* could be particularly interesting for this application, as artificial truncation can, in some cases, reduce stability or impair binding, as

observed here. Furthermore, the ability of group-2 Docs to bind Cohs in pairs would enhance the efficiency of the protein purification processes. Likewise, the double dockerin format can be used to promote signal amplification in detection or diagnostic systems: a cohesin module can be fused to a detection antibody, and a fluorescently tagged dockerin would allow a two-fold increase in signal in the presence of the target antigen. The design of tailor-made cellulosomes can also benefit from the short dockerin format as it would allow more compact and cost-efficient formats. The fact that these short dockerins can be engineered from Docs of different origins would also allow the integration of multiple specificities into the designer scaffolds to promote a specific enzyme arrangement within the cellulosomes.

In conclusion, the discovery of this novel Coh-Doc binding mechanism not only deepens our understanding of cellulosome assembly but also opens new avenues for advancing Coh-Doc technology. These findings pave the way for innovative applications in molecular biology, diagnostics, and biotechnology, offering exciting opportunities for future research and development.

## Experimental procedures

### Gene synthesis and DNA cloning

Dockerins are small unstable protein modules when expressed individually in heterologous hosts. To promote stability, *R. flavefaciens* FD-1 *Rf*Doc2 (WP_009984933.1) 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 (Merk Millipore, Germany) under the control of separate T7 promoters. The *Rf*Doc2-encoding gene was positioned at the 5′ end and the *RfCohScaE-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. We use the polyhistidine tag at the Doc N-terminal, because the expression levels of both Coh and Doc are higher, contrary to Coh tagging which results in the accumulation of large levels of unbound Coh in the purification product.

To produce recombinant *Rf*Doc2 and CohScaE individually, the recombinant complex was digested with BglII to excise the Doc-encoding amplicon. This strategy gave a pET28a derivative encoding the recombinant CohScaE fused to a C-terminal hexahistidine tag, maintaining plasmid integrity by re-ligating. The *Rf*Doc2-encoding gene was subcloned into the pETG-20A plasmid by Gateway recombination (Life Technologies, California, United States) following the manufacturer's protocol. The resulting expressed product consisted of a His-tagged *Rf*Doc2 fused to Thioredoxin (TrxA) for increased solubility and stability.

To design truncated Docs for the binding experiments, a structural alignment was performed between *Rf*Doc2 and each wild-type dockerin. Doc two was used as a reference to determine where to split the other dockerins into two truncated derivatives (N- and C- terminal fragments). Genes encoding the truncated dockerins were then synthesized as described above. *Rf*Doc2 variants (Table S2) were produced through site-directed mutagenesis, using the primers listed in Table S3. Each of the newly generated genes was fully sequenced to verify that only the desired mutation accumulated in the nucleic acid chain.

### 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 OD600 of 0.4 to 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), initial purification was carried out by IMAC in an FPLC system (GE Healthcare) using conventional protocols with a 35 mM imidazole wash and a 35 to 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). Gel-filtration chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) 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) and washed three times with molecular biology grade water (Sigma Chem. Co) 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). SDS–PAGE gels (14% w/v) were used to confirm the purity and molecular mass of the recombinant complexes. CohScaE, *Rf*Doc2, and its mutant derivatives, as well as DocScaH and its truncated versions, used in ITC experiments, were expressed as described above and purified with His GraviTrap gravity-flow nickel-charged Sepharose columns (GE Healthcare). 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).

### Nondenaturing gel electrophoresis

For the nondenaturing gel electrophoresis (NGE) experiments, each wild-type Doc and the truncated variants, at

a concentration of 50 μM, were incubated in the presence and absence of 50 μM of the respective Coh for 1 h at room temperature and separated on a 10% native (nondenaturing) 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. Results were confirmed by ITC.

### *Crystallization of CohScaE-DocScaH*

Several crystallization conditions were tested by using the sitting-drop vapor-diffusion method, with the aid of an Oryx8 robotic nanodrop dispensing system (Douglas Instruments). The commercial kits JCSG+, Crystal Screen, PEG/Ion (Hampton Research), and in-house prepared 80 factorial solutions were used for the screening. Drops (1 μl) of 10 and 20 $mg{\cdot}mL^{-1}$ *Rf*CohScaE-Doc2 were mixed with 1 μl reservoir solution at room temperature. The resulting plates were then stored at 293 K. Crystal formation was observed under only one condition after approximately 30 days from setting up the plates. All other apparent crystals were determined to be salt crystals. The successful condition was further optimized to produce high-quality crystals suitable for structural analysis.

A total of four crystals were cryoprotected in a 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 II detector) revealed that the best diffracting crystals were formed in a solution of 0.2 M Lithium acetate dihydrate pH 7.9 and, 20% w/v Polyethylene glycol 3350.

### *3D structure determination and refinement*

The crystal structure of *Rf*CohScaE-Doc2 was determined by molecular replacement. X-ray diffraction data were collected on beamline ID23-1 at the ESRF, Grenoble, using a Eiger 2 16M CdTe (Dectris Ltd). A systematic grid search was carried out to select the best-diffracting part of each crystal. iMosflm (39) was used for strategy calculation during data collection. All data sets were processed using the Grenoble Automatic Data ProcEssing (GrenADES) pipeline, which uses XDS (40), POINTLESS, SCALA, and AIMLESS software. Data-collection statistics are given in Table 1. The best-diffracting crystal diffracted to a resolution of 2.6 Å and belonged to the hexagonal spacegroup $P6_4$. Phaser MR was used to carry out molecular replacement, using the structure of CohScaE in complex with the Doc from ScaH (8AJY) (31). One copy of the *Rf*CohScaE-Doc2 heterodimer was present in the asymmetric unit. The partially obtained model was completed with Buccaneer (41) and with manual modelling in COOT (42). It was then refined using REFMAC5 (43) and PDB REDO (44) interspersed with model adjustment in COOT. The final round of refinement was performed using the TLS/restrained refinement procedure, using each chain as a single group, giving the final model to 2.9 Å resolution (Protein Data Bank code 9GVB). The rmsd of bond lengths, bond angles, torsion angles and other indicators were continuously monitored using validation tools in COOT and MOLPROBITY (45). 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 (46).

### *Isothermal titration calorimetry*

All ITC experiments were carried out in a Microcal VP-ITC system (Malvern Panalytical). *R. flavefaciens*, *B. cellulosolvens* and *A. cellulolyticus* experiments took place at 308 K, while experiments with *C. thermocellum* modules were conducted at 338K. The purified wild-type *Rf*Doc2 and mutant variants were diluted to 80 μM, while CohScaE was diluted to 150 μM, in a buffer containing 50 mM HEPES pH 7.5, 0.5 mM $CaCl_2$ and 0.5 mM TCEP. The artificially truncated Docs were diluted to 90 μM, while their Coh counterparts were diluted to 200 μM, in a similar buffer. 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 cohesin 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 Figures S3 and S5.

### *Accession numbers*

PDB: 9GVB.

## Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code PDB 9GVB [https://www.rcsb.org/structure/9GVB. All further data supporting the findings of this study are available from the corresponding author, upon reasonable request.

*Supporting information*—This article contains supporting information (26).

*Acknowledgments*—We acknowledge the European Synchrotron Radiation Facility (ESRF) for access to beamline ID23-EH1 of the synchrotron facility through BAG-Portugal (proposal MX-2276). 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 (46).

*Author contributions*—M. D., C. M. G. A. F., and P. B. conceptualization; M. D., M. C. F., B. C., J. A. M. P., E. A. B., C. M. G. A.

F., and P. B. methodology; M. D., A. L. C., M. C. F., B. C., S. N., and P. B. investigation; M. D. and P. B. writing-original draft; M. D. and P. B. visualization; A. L. C., M. C. F., J. A. M. P., S. N., and E. A. B. formal analysis; A. L. C., M. J. R., J. A. M. P., S. N., E. A. B., and C. M. G. A. F. writing–review & editing; P. B. and M. J. R. funding acquisition; P. B. supervision; P. B. project administration.

*Funding and additional information*—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 interests*—The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marlene Duarte reports financial support was provided by Foundation for Science and Technology. Carlos M.G.A. Fontes reports a relationship with NZYTech Lda that includes: board membership and employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

*Abbreviations*—The abbreviations used are: CAZymes, carbohydrate active enzymes; Coh, cohesin; Doc, dockerin; LRR, leucine-rich repeat; SBM, single-binding mode; Sca, scaffoldins.

## References


1. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic. Acids. Res.* **42**, D490–D495
2. Paës, G., Navarro, D., Benoit, Y., Blanquet, S., Chabbert, B., Chaussepied, B., *et al.* (2019) Tracking of enzymatic biomass deconstruction by fungal secretomes highlights markers of lignocellulose recalcitrance. *Biotechnol. Biofuels* **12**, 76
3. Fontes, C. M. G. A., and Gilbert, H. J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681
4. 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
5. Bensoussan, L., Moraïs, S., Dassa, B., Friedman, N., Henrissat, B., Lombard, V., *et al.* (2017) Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. *Environ. Microbiol.* **19**, 185–197
6. Hamberg, Y., Ruimy-Israeli, V., Dassa, B., Barak, Y., Lamed, R., Cameron, K., *et al.* (2014) Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions. *PeerJ* **2**, e636
7. Dassa, B., Utturkar, S., Hurt, R. A., Klingeman, D. M., Keller, M., Xu, J., *et al.* (2015) Near-complete genome sequence of the cellulolytic bacterium Bacteroides (pseudobacteroides) cellulosolvens ATCC 35603. *Genome. Announc.* **3**, e01022
8. Cann, I., Bernardi, R. C., and Mackie, R. I. (2016) Cellulose degradation in the human gut: *Ruminococcus champanellensis* expands the cellulosome paradigm: *Ruminococcus champanellensis* celulosome. *Environ. Microbiol.* **18**, 307–310
9. Artzi, L., Dassa, B., Borovok, I., Shamshoum, M., Lamed, R., and Bayer, E. A. (2014) Cellulosomics of the cellulolytic thermophile Clostridium clariflavum. *Biotechnol. Biofuels.* **7**, 100
10. Adams, J. J., Currie, M. A., Ali, S., Bayer, E. A., Jia, Z., and Smith, S. P. (2010) 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**, 833–839
11. Bule, P., Alves, V. D., Israeli-Ruimy, V., Carvalho, A. L., Ferreira, L. M. A., Smith, S. P., *et al.* (2017) Assembly of Ruminococcus flavefaciens cellulosome revealed by structures of two cohesin-dockerin complexes. *Sci. Rep.* **7**, 759
12. Jindou, S., Brulc, J. M., Levy-Assaraf, M., Rincon, M. T., Flint, H. J., Berg, M. E., *et al.* (2008) Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium Ruminococcus flavefaciens. *FEMS Microbiol. Lett.* **285**, 188–194
13. Berg Miller, M. E., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., *et al.* (2009) Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of Ruminococcus flavefaciens FD-1. *PLoS One* **4**, e6650
14. Jindou, S., Borovok, I., Rincon, M. T., Flint, H. J., Antonopoulos, D. A., Berg, M. E., *et al.* (2006) Conservation and divergence in cellulosome architecture between two strains of Ruminococcus flavefaciens. *J. Bacteriol.* **188**, 7971–7976
15. Rincon, M. T., Cepeljnik, T., Martin, J. C., Barak, Y., Lamed, R., Bayer, E. A., *et al.* (2007) A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of Ruminococcus flavefaciens. *J. Bacteriol.* **189**, 4774–4783
16. Israeli-Ruimy, V., Bule, P., Jindou, S., Dassa, B., Moraïs, S., Borovok, I., *et al.* (2017) Complexity of the Ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. *Sci. Rep.* **7**, 42355
17. Rincon, M. T., Dassa, B., Flint, H. J., Travis, A. J., Jindou, S., Borovok, I., *et al.* (2010) Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, Ruminococcus flavefaciens FD-1. *PLoS One* **5**, e12476
18. Rincón, M. T., Martin, J. C., Aurilia, V., McCrae, S. I., Rucklidge, G. J., Reid, M. D., *et al.* (2004) ScaC, an adaptor protein carrying a novel cohesin that expands the dockerin-binding repertoire of the Ruminococcus flavefaciens 17 cellulosome. *J. Bacteriol.* **186**, 2576–2585
19. Bule, P., Alves, V. D., Leitão, A., Ferreira, L. M. A., Bayer, E. A., Smith, S. P., *et al.* (2016) Single binding mode integration of hemicellulose-degrading enzymes via adaptor scaffoldins in Ruminococcus flavefaciens cellulosome. *J. Biol. Chem.* **291**, 26658–26669
20. Xu, Q., Resch, M. G., Podkaminer, K., Yang, S., Baker, J. O., Donohoe, B. S., *et al.* (2016) Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities. *Sci. Adv.* **2**, e1501254
21. Duarte, M., Viegas, A., Alves, V. D., Prates, J. A. M., Ferreira, L. M. A., Najmudin, S., *et al.* (2021) A dual cohesin–dockerin complex binding mode in Bacteroides cellulosolvens contributes to the size and complexity of its cellulosome. *J. Biol. Chem.* **296**, 100552
22. Carvalho, A. L., Dias, F. M. V., Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., *et al.* (2007) Evidence for a dual binding mode of dockerin modules to cohesins. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3089–3094
23. Bule, P., Pires, V. M. R., Alves, V. D., Carvalho, A. L., Prates, J. A. M., Ferreira, L. M. A., *et al.* (2018) Higher order scaffoldin assembly in Ruminococcus flavefaciens cellulosome is coordinated by a discrete cohesin-dockerin interaction. *Sci. Rep.* **8**, 6987
24. Bayer, E. A., Lamed, R., White, B. A., and Flint, H. J. (2008) From cellulosomes to cellulosomics. *Chem. Rec.* **8**, 364–377
25. Brás, J. L. A., Pinheiro, B. A., Cameron, K., Cuskin, F., Viegas, A., Najmudin, S., *et al.* (2016) Diverse specificity of cellulosome attachment to the bacterial cell surface. *Sci. Rep.* **6**, 38292
26. Duarte, M., Alves, V. D., Correia, M., Caseiro, C., Ferreira, L. M. A., Romão, M. J., *et al.* (2023) Structure-function studies can improve

binding affinity of cohesin-dockerin interactions for multi-protein assemblies. *Int. J. Biol. Macromolecules.* **224**, 55–67
27. Lytle, B. L., Volkman, B. F., Westler, W. M., Heckman, M. P., and Wu, J. H. D. (2001) Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding Domain11Edited by P. E. Wright. *J. Mol. Biol.* **307**, 745–753
28. Slutzki, M., Jobby, M. K., Chitayat, S., Karpol, A., Dassa, B., Barak, Y., *et al.* (2013) Intramolecular clasp of the cellulosomal Ruminococcus flavefaciens ScaA dockerin module confers structural stability. *FEBS Open. Bio.* **3**, 398–405
29. Jones, S., and Thornton, J. M. (2004) Searching for functional sites in protein structures. *Curr. Opin. Chem. Biol.* **8**, 3–7
30. Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000) An overview of the structures of protein-DNA complexes. *Genome. Biol.* **1**, reviews001.1
31. Salama-Alber, O., Jobby, M. K., Chitayat, S., Smith, S. P., White, B. A., Shimon, L. J. W., *et al.* (2013) Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.* **288**, 16827–16838
32. Voronov-Goldman, M., Levy-Assaraf, M., Yaniv, O., Wisserman, G., Jindou, S., Borovok, I., *et al.* (2014) Structural characterization of a novel autonomous cohesin from Ruminococcus flavefaciens. *Acta. Crystallogr. F. Struct. Biol. Commun.* **70**, 450–456
33. Brás, J. L. A., Alves, V. D., Carvalho, A. L., Najmudin, S., Prates, J. A. M., Ferreira, L. M. A., *et al.* (2012) Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode*. *J. Biol. Chem.* **287**, 44394–44405
34. Bule, P., Cameron, K., Prates, J. A. M., Ferreira, L. M. A., Smith, S. P., Gilbert, H. J., *et al.* (2018) Structure–function analyses generate novel specificities to assemble the components of multienzyme bacterial cellulosome complexes. *J. Biol. Chem.* **293**, 4201–4212
35. Artzi, L., Bayer, E. A., and Moraïs, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95
36. Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C. L., *et al.* (2012) Genome-wide analysis of acetivibrio cellulolyticus provides a blueprint of an elaborate cellulosome system. *BMC Genomics.* **13**, 210
37. Chauvaux, S., Beguin, P., Aubert, J. P., Bhat, K. M., Gow, L. A., Wood, T. M., *et al.* (1990) Calcium-binding affinity and calcium-enhanced activity of Clostridium thermocellum endoglucanase D. *Biochem. J.* **265**, 261–265
38. 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
39. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R., and Leslie, A. G. W. (2011) *iMOSFLM* : a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallogr. Section. D. Biol. Crystallogr.* **67**, 271–281
40. Kabsch, W. (2010) Xds. *Acta. Crystallogr. Section. D. Biol. Crystallogr.* **66**, 125–132
41. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta. Crystallogr. D. Biol. Crystallogr.* **62**, 1002–1011
42. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *coot*. *Acta. Crystallogr. Section. D. Biol. Crystallogr.* **66**, 486–501
43. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., *et al.* (2011) *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta. Crystallogr. Section. D. Biol. Crystallogr.* **67**, 355–367
44. Joosten, R. P., Long, F., Murshudov, G. N., and Perrakis, A. (2014) The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220
45. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 12–21
46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., *et al.* (2021) UCSF ChimeraX : structure visualization for researchers, educators, and developers. *Protein. Sci.* **30**, 70–82