Single versus dual-binding conformations in cellulosomal cohesin–dockerin complexes

Michael A Nash1,2,3, Steven P Smith4, Carlos MGA Fontes5 and Edward A Bayer6

Cohesins and dockerins are complementary interacting protein modules that form stable and highly specific receptor–ligand complexes. They play a crucial role in the assembly of cellulose-degrading multi-enzyme complexes called cellulosomes and have potential applicability in several technology areas, including biomass conversion processes. Here, we describe several exceptional properties of cohesin–dockerin complexes, including their tenacious biochemical affinity, remarkably high mechanostability and a dual-binding mode of recognition that is contrary to the conventional lock-and-key model of receptor-ligand interactions. We focus on structural aspects of the dual mode of cohesin–dockerin binding, highlighting recent single-molecule analysis techniques for its explicit characterization.

Addresses
1 Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
2 Department of Chemistry, University of Basel, 4056 Basel, Switzerland
3 Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule (ETH-Zürich), 4058 Basel, Switzerland
4 Department of Biomedical and Molecular Sciences, Queen’s University, Kingston, Ontario K7L 3N6, Canada
5 CIISA-Faculdade de Medicina Veterinária, ULisboa, Pólo Universitário do Alto da Ajuda, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal
6 Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding author: Nash, Michael A (michael.nash@unibas.ch)

Current Opinion in Structural Biology 2016, 40:89-96
This review comes from a themed issue on Carbohydrate-protein interactions and glycosylation
Edited by Nagasuma Chandra and Harry Gilbert

http://dx.doi.org/10.1016/j.sbi.2016.08.002
0959-440/© 2016 Elsevier Ltd. All rights reserved.

Introduction
The structural integrity of a plant is defined largely by the cell wall, which comprises the polysaccharides cellulose and hemicellulose and the complex cross-linked phenolic polymer lignin [1]. The decomposition of lignocellulosic biomass is a crucial process for carbon cycling and energy production in Nature, and is achieved primarily by microorganisms. In particular, the bacteria and fungi produce an elaborate set of enzymes, including glycoside hydrolases, polysaccharide lyases and carbohydrate esterases, and lytic polysaccharide monooxygenases, to access this rich energy source by catalyzing the cleavage of the glycosidic bonds and associated carbohydrate esters of the cellulosic and hemicellulosic plant cell wall polysaccharides [2]. In the case of aerobic bacteria and fungi, these hydrolytic enzymes are produced and released in substantial quantities into the extracellular environment, where they diffuse independently to their lignocellulosic substrates [3]. In addition to a catalytic module, these enzymes typically also comprise a non-catalytic carbohydrate-binding module (CBM) that targets the enzyme to its substrate [4,5]. In contrast, select lignocellulose-degrading anaerobic bacteria organize their secreted complement of enzymes into a multi-enzyme complex, called the cellulosome, where substrate targeting by associated CBMs and enhanced catalytic synergism, owing to the proximity among the resident enzymes is optimized for degradation of this recalcitrant substrate [6].

Emergent architectural complexity of bacterial cellulosomes
First discovered over three decades ago in the thermophilic, anaerobic bacterium Clostridium (Ruminiclostridium) thermocellum, cellulosomal complexes have since been identified in several other cellulytic bacteria, including Acetivbrio cellulolyticus, Bacteroides (Pseudobacteroides) cellulosolvens, Clostridium (Ruminiclostridium) cellulolyticum, Clostridium (Ruminiclostridium) cellulocrans, Clostridium (Ruminiclostridium) clarifaciens, Clostridium (Ruminiclostridium) papyrosolvens, Ruminococcus champanellensis, and Ruminococcus flavefaciens [7]. The basic cellulosome structure, of which that from C. thermocellum continues to serve as the prototype, constitutes three primary multimodular subunits: (1) enzymatic subunits responsible for the lignocellulose-degrading properties of the cellulosome; (2) a large non-catalytic scaffoldin protein, referred to as CipA in C. thermocellum, onto which the enzyme subunits are arranged; and (3) a cell-surface anchoring subunit that tethers the entire complex to the bacterial cell surface (Figure 1) [8]. However, recent genome sequencing, bioinformatics, and biochemical analyses have revealed more structurally elaborate cellulosomal complexes produced by several of the abovementioned bacterial species. In addition to the larger array of hydrolytic enzymes identified in these cellulosomes...
Regardless of the bacterial species, the single most essential element in assembly of celluloses and their attachment to the bacterial cell envelope via the cell surface anchoring subunits involves highly specific, high-affinity (i.e., $K_d < 10^{-7}$ M) interactions between tandemly linked cohesins and dockerin modules. These cohesin–dockerin interactions govern the integration of the enzymatic subunits as well as the assembly of the primary, adaptor and anchoring scaffoldins. Given the indispensable role of cohesin and dockerin modules in the assembly of a functional cellulose and the type and interspecies specificity observed [11], a particular focus has centered on the structural and mechanistic basis of these interactions.

The type-I cohesin–dockerin interaction—the dockerin fold revisited

Dockerm modules have been classified into three types, types I, II and III, according to their primary sequences. Type-I dockerin modules are approximately 70 residues in length and comprise two duplicated 22-amino acid residue segments (Figure 1a) [12]. The N-terminus of each of these segments possesses a 12-residue calcium-binding loop very similar to the EF-hand consensus motif that mediates calcium-induced folding of this module, and has been referred to as a F-hand motif [12]. Unlike the type-I dockerin, some type-II and type-III dockerins occur in tandem with adjacent X-modules that appear to play a structural stabilizing role [13,14*].

The first structure of an isolated type-I dockerin module, determined by NMR spectroscopy, showed the two F-hand motifs arranged in an antiparallel conformation with amino acid residues from the α-helices of the two F-motifs forming a modest hydrophobic core. As such, the dockerin module appeared to possess an apparent loosely packed overall planar topology that when compared to the dockerin structure in complex with its cognate cohesin module suggested it underwent a substantial cohesin–induced conformational change [15]. The structure of this type-I dockerin module was recently reassessed, again by NMR spectroscopy, where a larger number and more diverse set of distance restraints were included in the structure calculation [16*]. While the helices of the F-hand motifs adopted an antiparallel arrangement similar to the initial structure, the revisited structure was more compact due to a more extensive hydrophobic core and displayed an internal two-fold structural symmetry, which coincided with the sequence symmetry of the two F-hand motifs (Figure 1a). The resultant structure was very similar to the dockerin conformation when in complex with its cohesin partner (Figure 1b). Thus, it appears that the type-I dockerin module adopts an inherent calcium-dependent cohesin–primed conformation [16*].

Structural basis of cohesin–dockerin dual-binding mode

The X-ray crystal structure of a representative type-I cohesin–dockerin complex from the C. thermocellum cellulose was a transformative study that revealed the dockerin module bound via extensive hydrophobic and hydrogen-bonding contacts to a planar region of the β-sheet formed by strands 3, 5, 6, and 8 within the cohesin jellyroll fold (Figure 1b) [17]. Despite the sequential and structural symmetry of the type-I dockerin module involving the two F-hand motifs, the dockerin contacted the cohesin surface in an asymmetric manner. The helix of the second F-hand motif formed contacts with the

Figure 1

Structural basis of type-I cohesin–dockerin complex in cellulose assembly. (a) A backbone ribbon representation of the isolated type-I dockerin module from the Cel48S enzymatic subunit of C. thermocellum illustrates the 22-residue tandem F-hand motifs packed in a globular antiparallel orientation [16*]. (b) The backbone ribbon representation of a C. thermocellum type-I cohesin–dockerin complex comprising the xylanase 10B dockerin module (red) and the second cohesin module from the C. thermocellum CipA scaffoldin subunit (yellow) [17]. The calcium ions are depicted as bronze spheres and cohesin N-terminus and C-terminus are labeled accordingly.
cohesin module over its entire length, including a Ser/Thr pair (Ser46/Ser47 in xylanase 10B) while the symmetrical helix of the dockerin structure (helix 1 of the first F-hand motif) only minimally contacted the cohesin module at its C-terminus (Figure 2a). These findings along with previous bioinformatics and mutagenesis studies [12,18] indicated that the conserved Ser/Thr pairs of both F-hand motifs in the type-I dockerin structure (e.g., Ser10/Thr11; Ser46/Ser47 in xylanase 10B) were critical for binding and are referred to as ‘recognition residues’. Consequently, the authors proposed that this high-affinity protein-protein interaction could occur with the type-I dockerin module in two orientations; one exemplified by the crystal structure and the other whereby the dockerin module is rotated 180° such that the complementary F-hand motif forms essentially identical contacts with the cohesin module [17].

Experimental visualization of the dual-binding mode involved mutation of the Ser/Thr pair in the second F-hand motif of the type-I dockerin module (Ser46/Ser47) to alanine residues thereby forcing the opposite motif to assume the primary contact site with the cohesin module. The ensuing crystal structure of the type-I dockerin Ser45Ala/Thr46Ala mutant in complex with the cohesin confirmed the existence of the reverse mode of binding with respect to the initial structure (Figure 2b) [19,20]. Complementary binding studies indicated that wild-type and mutated type-I dockerin constructs all displayed similar affinities ($K_d \sim 10^{-8}$ M) for the type-I cohesin module. In this context, mutagenesis studies have portended that dual binding may occur in vivo [21]. Despite these studies, some of which have relied on mutated versions of the dockerin module to observe alternative binding orientations, there has yet to be a definitive, direct structural observation of the type-I cohesin–dockerin dual-binding mode freely diffusing in solution.

A distinguishing sign of the dual-binding mode is the strict sequence identity or strong similarity of the two duplicated segments of the dockerin module [19,20,22]. While the dual-binding mode is characteristic of the type-I dockerins [22], not all type-I dockerins adhere to this trait [23]. Conversely, although type-II and type-III dockerin modules are typically asymmetric in their sequences and presumably exhibit single modes of binding [13,14,24], several have been identified which are decidedly symmetric which would signify dual-binding modes [25,26].

The dual binding properties of cohesin–dockerin complexes appear to be entirely unique in the molecular recognition field. To our knowledge, no other system with a comparable plasticity of receptor–ligand complex assembly can be identified in the literature. The term ‘dual-binding mode’ has, however, been reported elsewhere in the literature, although they describe interactions between small molecules and proteins, rather than interactions between two separate proteins. Protein
kinase inhibitors, for example, have been reported to bind their targets using ‘dual modes’ [27]. Additionally, computational modeling showed that G protein-coupled estrogen receptors interact with ligands through a ‘dual-binding mode’ [28]. In both of these systems, small molecules bind to much larger proteins, resulting in small shifts in orientation of the small molecule in the binding pocket between the two modes. To our knowledge the cohesin–dockerin dual-binding mode paradigm is exceptional both in terms of the size of the molecules involved (i.e., globular protein modules) and the magnitude of the differences between the two bound conformations.

**Single-molecule characterization methods**

Until recently, functional studies on cellulose protein components had only been performed using conventional biochemical techniques (e.g., ELISA-based assays, chromatography, reducing sugar assays) and structural methods (e.g., X-ray crystallography, NMR spectroscopy). While these techniques have provided detailed insights into the cellululosomal protein structure/function relationship, there remains a void in our understanding of how these proteins have evolved in response to mechanical selection pressures in vivo.

It is now widely appreciated that directed mechanical forces play an important role in cellular systems, for example, by orchestrating cellular differentiation and gene expression patterns, as well as propagating signals through the extracellular matrix [29]. Protein mechanical properties are also known to play a role in several bacterial adhesion systems [30,31], suggesting they might represent an important parameter for cellulose-producing bacteria which are required to adhere to and digest crystalline lignocellulosic substrates within a variety of hydrodynamic environments (e.g., human and herbivore digestive tracts, acidic, basic and thermophilic ecosystems, as well as hydrothermal vents).

The advent of single-molecule force spectroscopy using nanoscale force transducers (e.g., atomic force microscope (AFM) cantilevers, magnetic tweezers), has allowed the routine application of piconewton forces to biomolecules [32–38,39,40–44] and biomolecular complexes [45–55]. By characterizing the mechanical responses of cellulose protein modules and associated complexes [32,56,57,58,59,60,61,62], we are beginning to appreciate that cellulose components have evolved to comprise among the most mechanically robust families of protein modules ever analyzed. Indeed, celluloses appear to be evolutionarily selected for performing the arduous task of biomass deconstruction in some of the most extreme environments (e.g., elevated temperatures, acidic and basic pH), and, as a direct consequence, their protein modules display exceptional mechanical stability.

**Mechanical stability versus thermodynamic affinity**

The difference between mechanical stability (i.e., bound lifetime under force) and its relationship to the dissociation constant (i.e., \( K_d = \frac{k_{off}}{k_{on}} \)) is important to recognize, especially for biomolecular systems involved in adhesion to solid surfaces, as is the case for cellulosome complexes adhering to solid biomass. The kinetic off-rate measured using conventional biochemical characterization methods (e.g., surface plasmon resonance, thermophoresis, enzyme-linked immunosorbent assays) is an average of all the thermally accessible unbinding pathways of a molecular complex out of the bound state. Under the application of non-equilibrium mechanical force, however, biomolecular complexes are conceptually being pulled out of the energy well along a constrained pathway. The off-rate measured under mechanical perturbation is, therefore, not necessarily the same as in the equilibrium condition. This discrepancy is significant, especially when characterizing systems that are exposed to mechanical stress in vivo.

In the equilibrium case, when we catalog the \( K_d \) values and \( k_{off} \) values of cohesin–dockerin complexes from the literature, the typical observed affinities are in the range of 0.1–100 nM [63–68] with off-rates in a range between \( 10^{-3} \) and \( 10^{-5} \) s\(^{-1}\). On the basis of several published studies [32,56,57,59,60,61,69], we find that off-rates under force obtained through single-molecule mechanical unbinding experiments tend to be lower by 10–100 fold, falling across a range of \( 10^{-4}–10^{-7} \) s\(^{-1}\). This suggests that cohesin–dockerin complexes may have slightly lower off-rates under force than at equilibrium, providing evidence for an alternative mechanical unbinding pathway with enhanced stability. Such a pathway could have evolved based on hydrodynamic selection pressure, adhesion selection pressures, or other physiological processes that require celluloses to adhere at solid–liquid interfaces (i.e., on cellulose nanocrystals). Further supporting this mechanical picture of cellulosomal components is the fact that the forces required to unfold cohesin modules have large differences based on the location of cohesin within the cellulosomal networks. ‘Bridging’ cohesin modules that are mechanically stressed within the scaffolds are more stable by a factor of 2 or more than ‘hanging’ cohesin modules that do not experience mechanical load [62].

Given the remarkable mechanical properties of cohesin–dockerin complexes, the molecular mechanisms by which these complexes achieve their high mechanical stability represent together a topic of significant interest. Recent results from all-atom steered molecular dynamics simulations reported by Schoeler et al. support two primary mechanisms through which the type-III cohesin–dockerin complex from *R. flavefaciens* achieves remarkable mechanical stability: (1) rearrangement of side-chain
Cohesin–dockerin complexes under mechanical force.
(a) Visualizations from steered molecular dynamics simulations of a type-III cohesin–dockerin complex as mechanical load is applied. Upon loading with force, rearrangement of binding residues of cohesin (blue) and dockerin (red) cause the side chains interdigitate, resembling teeth of a zipper. (b) The total contact surface area of cohesin and dockerin in unloaded and loaded conformation was calculated from molecular dynamics simulations. Upon application of force, the contact area increases. (c) Dynamic network analysis was used to visualize how force propagates through the cohesin–dockerin complex. The thickness of the orange tube represents the number of suboptimal correlation paths passing between alpha carbons, and indicates the most probable path for force transmission. Force was found to propagate perpendicular to the pulling axis at the binding interface, providing a mechanism for mechanical stability.
Figures were adapted with permissions from [60] (a and b) and [59] (c).

residues upon mechanical loading (Figure 3a) and (2) directing force through the binding interface along vectors with high normal component with respect to the pulling axis (Figure 3b) [59**].

Mechanical unbinding for interrogating cohesin–dockerin dual-binding mode
As detailed above, mutagenesis studies and X-ray crystallography have been used previously to identify several cohesin–dockerin complexes that exhibit the dual-binding mode. Observing both bound conformations for the wild-type complex, however, has not been possible using bulk ensemble methods. The two binding modes have very similar thermodynamic parameters, relying on an identical network of hydrogen bond interactions and structurally homologous binding residues to assemble [17,19]. To address this limitation, Jobst et al. recently used a single-molecule approach in which the two binding modes could be differentiated based on differences in mechanical stability (Figure 4) [57**]. Although both dockerin binding modes involve analogous hydrogen bonded interaction pairs, binding free energy, and equilibrium dissociation constants, their mechanical properties were found to vary significantly. This difference was exploited by looking at the shift in an unfolding force distribution of an adjacent iLov domain in single-molecule pulling experiments [57**]. The authors first validated the approach on mutant dockerin modules that showed a preference for one binding mode, and noted specific unbinding pathways that differed for the mutants. They subsequently confirmed the presence of the dual-binding mode in the wild-type dockerin, thus confirming that the wild-type sequence can bind in both orientations. This work represents a new general approach to analyzing biomolecular complexes under mechanical force, and could be utilized in other cohesin–dockerin systems to interrogate dual-binding modes.

Conclusions
The presence of dual-binding conformations has been confirmed in several cohesin–dockerin systems. However, several questions still remain. For example, it is not clear what role the dual-binding mode plays in the catalytic activity of cellulosomes. It has been suggested that dual-binding modes would confer a robust mechanism against mutation or potentially allow cautious communalism within mixed microbial communities. The
dual-binding mode was also hypothesized to add an additional degree of conformational plasticity to the process of cellulose assembly to limit potential steric hindrance of neighboring enzymatic subunits. Another possible consequence would be to provide enhanced conformational freedom of the cellulosomal enzymes to act on their preferred substrate during deconstruction of the plant cell wall. However, these remain hypotheses and further experimental exploration in this area is required.

Single-molecule mechanical experiments are shedding light on the dual-binding mode and onto mechanical properties of cellulose components in general. Given that the thermodynamic parameters of each binding mode are not discernible, a single-molecule approach is logical where each instance of the complex is interrogated individually. It is clear that the mechanical properties of cellulose networks are exceptional within the protein world, and we currently explain this in terms of mechanical selection pressures. The cohesin–dockerin interaction is among the most stable molecular complexes under force. Further characterization of mechanical properties of cellulose components could open up improved understanding of biomass conversion systems, as well as new technological applications relying on stable protein mechanics.

Acknowledgments
European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number: 604530 and a grant from GIF, the German-Israeli Foundation for Scientific Research and Development (Grant No. G-1191-147-9/2012) to MAN and EAB are gratefully acknowledged. Additional support was obtained from grants to EAB from the Israel Science Foundation (ISF Grant No. 1349), to SPS from the National Science and Engineering Research Council of Canada Discovery Grant RGPIN 398305, and to CMGAF from the Fundação para a Ciência e a Tecnologia (Lisbon, Portugal) Grants PTDC/BIA-PRO101990/2008 and EXP/ BIA-MIC/1176/2012. EAB is the incumbent of The Maynard L. and Elaine Wishner Chair of Bio-organic Chemistry. MAN acknowledges funding from the Society in Science — The Branco Weiss Fellowship program administered by ETH Zurich, Switzerland.

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

15. Chen C, Cui Z, Xiao Y, Cui Q, Smith SP, Lamed R, Bayer EA, Feng Y: Revisiting the NMR solution structure of the Cel48S type-I dockerin module from Clostridium thermocellum reveals a cohesin–primed conformation. J Struct Biol 2014, 188:188-193. A NMR spectroscopy study revisits the folding properties of the C. thermocellum type-I dockerin module. With the incorporation of 13C and 15N-based structural restraints the Cel48S dockerin module was shown to adopt a conformation very similar to that observed for the dockerin module in its cognate CipA scaffold type-I cohesin module, suggesting that in the presence of calcium these dockerin exist in a cohesin–ready state.


Using Acetobacter cellulolyticum type I cohesin-dockerin complexes, the authors demonstrate that cellulose cell surface attachment involves a cohesin-dockerin dual binding interaction. This is the first study reporting that the dual binding mode operates not only for cellulose assembly but also for the binding of the cellulosome to the cell envelope.


24. Voronov-Goldman M, Yaniv O, Gul O, Yoffe H, Salama-Alber O, Slutzki M, Levy-Assaraf M, Jindou S, Shimon LJW, Borovok I et al.: Standalone cohesin as a molecular scaffold in cellulose assembly. \textit{FEBS Lett} 2015, 589:1569-1576. The standalone \textit{Ruminococcus flavefaciens} cohesin, CohG, was crystalized together with the scaffoldin (CtfA)-borne dockerin, which was demonstrated to assume the opposite binding orientation compared to its previously described interaction with the adaptor scaffoldin, ScaE. The availability of the second dockerin site of the complex to bind ScaE suggested that CohG may serve as a molecular scaffold for delivery of scaffoldins to the cell surface.


57. Jobst MA, Milles LF, Schoeler C, Ott W, Fried DB, Bayer EA, Gaub HE: Resolving dual binding conformations of cellulose cohesin–dockerin complexes using single-molecule force spectroscopy. Differences in mechanical properties were used to discriminate the two binding conformations.


60. Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, Ott W, Fried DB, Bayer EA, Schulten K et al.: Ultrastable cellulosome–adhesion complex tightens under load. Nat Commun 2014, 5:1-35. This work was the first to show that cohesin–dockerin type-III complexes are ultra-stable when stressed under mechanical force, yet maintain only nM thermodynamic affinity. The exceptional rupture forces of this complex (~500–800 pN) highlight the orthogonality of thermodynamic versus mechanical stability.

61. Otten M, Ott W, Jobst MA, Milles LF, Verderfer T, Pippig DA, Gaub HE: From genes to protein mechanics on a chip. Nat Methods 2014, 11:1127-1130. The mechanical properties and small size of dockerin modules were advantageous for providing a universal handle domain to mechanically interrogate proteins of interest in a microarray format. This approach shows how microfluidic methods combined with mechanically stable dockerins allow for parallelization and improved throughput in single-molecule mechanical measurements on proteins.


