The cohesin module is a major determinant of cellulosome mechanical stability

Albert Galera-Prat‡, Sarah Moraïs§, Yael Vazana§, Edward A. Bayer§, Mariano Carrión-Vázquez‡

From the ‡Instituto Cajal, IC-CSIC, Avenida Doctor Arce 37, 28002 Madrid, Spain; Instituto Madrileño de Estudios Avanzados en Nanociencia (IMDEA Nanociencia), Cantoblanco, 28049 Madrid, Spain and

the §Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, 76100, Israel.

Running title: Cellulosome Nanomechanics

To whom correspondence should be addressed: Mariano Carrión-Vázquez, Instituto Cajal IC-CSIC, Avenida Doctor Arce 37, 28002, Madrid Spain. Telephone: +34 91 585 4830; FAX: +34 91 585 4754; E-mail: mcarrion@cajal.csic.es

Keywords: Cellulosome, atomic force microscopy (AFM), scaffolding protein, single-molecule biophysics, mechanical stability, nanomechanics, single-molecule force spectroscopy, scaffoldin, linkers, dockerin

ABSTRACT

Cellulosomes are bacterial protein complexes that bind and efficiently degrade lignocellulosic substrates. These are formed by multimodular scaffolding proteins known as scaffoldins, which comprise cohesin modules capable of binding dockerin-bearing enzymes and usually a carbohydrate binding module that anchors the system to substrate. For those cellulosomes bound to the bacterial cell surface, it was suggested that they might be exposed to significant mechanical forces. Accordingly, the mechanical properties of these anchored cellulosomes may be important to understand and improve cellulosome function. Here, we used single-molecule force spectroscopy to study the mechanical properties of selected cohesin modules from scaffoldins of different cellulosomes. We found that cohesins located in the region connecting the cell and the substrate are more robust than those located outside these two anchoring points. This observation applies to cohesins from primary scaffoldins (i.e. those that directly bind dockerin-bearing enzymes) from different cellulosomes despite their sequence differences. Furthermore, we also found that cohesin nanomechanics (specifically, mechanostability and the position of the mechanical clamp of cohesin) is not significantly affected by other cellulosomal components, including linkers between cohesins, multiple cohesin repeats, and dockerin binding. Finally, we also found that cohesins (from both the connecting and external regions) have poor refolding efficiency but similar refolding rates, suggesting that the high mechanostability of connecting cohesins may be an evolutionary conserved trait selected to minimize the occurrence of cohesin unfolding, which could irreversibly damage the cellulosome. We conclude that cohesin mechanostability is a major determinant of the overall mechanical stability of the cellulosome.

Plant cell-wall polysaccharides are a major source of organic carbon in nature (1). Nevertheless, due to their chemical and structural complexity, these polysaccharides are extremely recalcitrant, and thus their applicability in industry as a carbon source for fermentation is severely limited (1). It is widely accepted that a major bottleneck towards the utilization of this biomaterial is the deconstruction of the constituent polysaccharides into fermentable sugars, a process known as saccharification (2-4).

Several microorganisms have evolved systems to degrade these materials (5). In
particular, some anaerobic bacteria produce a highly organized extracellular complex termed the cellulosome (6). In this system, the action of several enzymes is coordinated by their incorporation into protein scaffolds known as scaffoldins. This is achieved by the high-affinity and specific interaction between dockerin modules, present in the different enzymes, and cohesins, β-sandwich-like modules (7,8) usually found in several tandem repeats in the scaffoldins. This strategy allows the coordinated action of several complementary enzymes precisely targeted to their substrate by means of a carbohydrate-binding module (CBM) (5), resulting in synergistic effects that achieve high specific activities (1).

Some scaffoldins also have a Surface Layer Homology (SLH) domain to effectively anchor the cell to the substrate through this cellulosolytic complex (9). As it is the case with many other adhesion systems (10), those cellulosomes that simultaneously bind to the cell and to the substrate might be subject to mechanical stress (11). Mechanical forces may arise due to the relative movement of the cell and its substrate, which would stretch the portion of scaffoldin placed between the two anchoring points (CBM and SLH), referred to herein as the connecting region.

In a hypothetical extreme case, if the mechanical load on cohesins were too great, these modules would unfold and therefore release the enzymes, which are bound via a dockerin module. Since this system’s synergy arises from both targeting the enzymes to the substrate and the proximity of different and complementary enzymes when bound (1,12,13), this would result in a loss of net saccharification activity. Thus, the mechanical properties of cohesin modules within the connecting regions might be important to understand the functioning of the cellulosome and to design new artificial cellulosomes for industrial applications (11).

In agreement with this hypothesis, it was found that the mechanical stability of a cohesin from the connecting region of CipA scaffoldin from Clostridium thermocellum was unusually high (11). On the other hand, a cohesin from the external region (i.e., located outside the two anchoring points) showed a much lower mechanical stability, despite the high sequence identity between these modules.

These exciting results posed many questions that remain open. First, considering the high diversity in cellulosome sequences and architectures, which in some cases involve several intermediate scaffoldins, we should explore if this mechanical hypothesis is generally applicable to other cellulosomes from different species and architectures. Moreover, we must consider that cohesins in a particular cellulosome are not isolated but flanked by linker sequences, that they are present in multiple copies and interact with dockerins. Therefore, it is also important to study how these elements affect the mechanical stability of cohesins so we can obtain an overall and more realistic picture of how cellulosomes respond to mechanical forces.

Here we use single-molecule force spectroscopy (SMFS), based on atomic force microscopy (AFM), in combination with steered molecular dynamics (SMD) simulations to tackle these important questions. To this end, we study the mechanical properties of a variety of cohesins from different scaffoldin regions of cellulosomes with distinct complexity. Furthermore, we analyze the effect of linker sequences, the presence of multiple cohesin modules in a single scaffoldin and the binding of dockerin on the mechanical stability of cohesins in order to get a more complete picture of the nanomechanics of cellulosomal complexes and test the mechanical hypothesis of the cellulosome.

RESULTS

Cohesin nanomechanics in primary scaffoldins - In order to test the general applicability of the mechanical hypothesis of cellulosomes, we first studied the mechanical stability of additional cohesin modules from the connecting and external regions of primary scaffoldins, i.e. scaffoldins with cohesins that directly bind dockerin-bearing enzymes. To this end, cohesins from C. thermocellum and Acetivibrio cellulolyticus cellulosomes were chosen since these are well-characterized complexes with different architectures.

In particular, we analyzed the mechanical stability of cohesins 1 and 9 from C.
thermocellum CipA scaffoldin (CtA1 and CtA9, respectively), as well as cohesins 3 and 4 from *A. cellulolyticus* ScaA (AcA3 and AcA4, respectively). CtA1 and AcA3 are found in the external region of their respective scaffoldin, while CtA9 and AcA4 form part of the connecting region. CtA1 and CtA9 show the lowest sequence identity among cohesins from this scaffoldin and also among cohesins inside their respective connecting/external regions (see Fig. S1), while cohesins AcA3 and AcA4 are part of a more complex cellulosome system, which contains additional intermediate scaffoldins.

The nanomechanical analysis of these modules (Fig. 1) showed that the mechanical stability, measured as the average unfolding force, was much higher for the cohesins located in the connecting regions ($F_u = 460 \pm 110$ pN for CtA9 and $F_u = 439 \pm 78$ pN for AcA4) than for those from the external regions ($F_u = 124 \pm 25$ pN for CtA1 and $F_u = 209 \pm 88$ pN for AcA3). Interestingly, despite the sequence differences the values that we found for these two new cohesins were very similar to those of the previously studied cohesins 2 (CtA2) and 7 (CtA7) from *C. thermocellum* CipA (214 ± 8 pN for CtA2 and 480 ± 14 pN for CtA7 (11); note that values in this paper were reported as mean ± sem. The corresponding values of standard deviation are 41 and 77, respectively).

The histogram of the $\Delta L_C$ (the contour length release upon each unfolding event) shows a narrow distribution of values which is similar for all cohesins studied and is independent of their relative position in the scaffoldin (49.2 ± 1.3 nm for CtA1, 49.0 ± 0.9 nm CtA9, 48.0 ± 0.9 nm for AcA3 and 50.5 ± 0.6 nm for AcA4, respectively). This result indicates that their mechanical clamp, *i.e.*, the region mainly responsible for the mechanical stability, is likely located in an equivalent position. Furthermore, since the measured $\Delta L_C$ is very close to the expected length for a protein of this size (~140[aa]*0.4[nm/aa]-5nm=51nm), this indicates that the mechanical clamp is located near the protein’s N- and C-termini. SMD simulations of the mechanical unfolding of cohesins performed using the published structure of CtA9 (Fig. S2) and a model of CtA1 allowed to confirm that the mechanical clamp implies the first and last $\beta$-strands of cohesin modules (Fig. S3), as was the case for the three previously studied cohesins.

It must be considered that the magnitude of the mechanical stability of a protein is dependent on the loading rate, *i.e.* the rate at which force is applied to the system. Like in the previous work on cohesin nanomechanics, the current experiments were performed at the relatively high loading rates accessible experimentally using an AFM; however, it must be noted that the loading rates that the cellulosome may be bearing during the degradation of the plant cell wall might be somewhat different.

To further investigate the mechanical properties of cohesins and obtain the kinetic parameters of the unfolding process, we studied the mechanical stability of CtA1 and CtA9 at different pulling speeds (Fig. 1C). This was used to calculate the mechanical unfolding kinetic parameters (unfolding rate without force, $\alpha_0$, and distance to the transition state, $\Delta x$) of both cohesins.

These two cohesins showed a clearly different behavior. CtA1 showed lower mechanical stabilities at all pulling speeds and much lower dependency on the pulling velocity as compared to CtA9. This translates to a much shorter $\Delta x$ and smaller $\alpha_0$ for CtA9 (0.094 ± 0.002 nm and (5.3 ± 0.6)·10^{-3} s^{-1}) compared to CtA1 (0.336 ± 0.008 nm and (9 ± 1)·10^{-2} s^{-1}). The $\Delta x$ value found for CtA9 is comparable to that of CtA7 (0.110 nm (11)), a cohesin also found in the connecting region, while that of CtA1 is even larger than that of CtA2 (0.17 nm), the other cohesin from the external region. This remarkably short $\Delta x$ and the lower unfolding rates explain the fundamental features responsible for the high mechanical stabilities observed for cohesins in the connecting region, which make these modules more brittle than those located in the external portion.

**Mechanical stability of cohesins from secondary scaffoldins** - To obtain further insight into the mechanical stability of cohesins according to their location in the cellulosome we performed a similar analysis using cohesins from secondary scaffoldins of the cellulosome from *A. cellulolyticus*. In particular, we stretched cohesin 4 of ScaB (AcB4) and cohesin 3 from
ScaC (AcC3). These modules are located between the two anchoring points of the system (the CBM and the SLH) although they do not directly bind dockerin-bearing enzymes. Instead, ScaB cohesins bind to dockerin modules from ScaA scaffoldins, while ScaC cohesins bind to ScaB dockerins (14) (Fig. 2).

The nanomechanical analysis of these modules showed that the two cohesins have very different mechanical stabilities. Cohesin AcC3 from the cell-wall-anchored scaffoldin showed a high mechanical stability ($F_U = 458 \pm 53$ pN), comparable to that of other cohesins from the connecting region. On the other hand, despite being located between the two anchoring points of the system cohesin AcB4’s mechanical stability is remarkably lower ($F_U = 247 \pm 54$ pN), being indeed similar to the low stabilities of cohesins from the external region.

**Mechanical stability of cohesins into a scaffoldin: the effect of natural linkers** - Most scaffoldins are composed of several cohesin modules as well as others with complementary functions (5) joined by linker sequences of varying length (15), which are usually glycosylated (16,17). The presence of linker sequences in a module may affect its mechanical properties in different ways. On the one hand, linkers may establish additional contacts with the module (18), which could lead to its stabilization. On the other, long sequences near the mechanical clamp of a protein might enhance fluctuations of the nearby regions of the module thus lowering its mechanical stability (19).

To test the effect of the presence of natural linker sequences on cohesin mechanical stability we stretched both CcC1Link, the first cohesin of *C. cellulolyticum* CipC including the nine residues linkers on each side in the pAFM vector. Since our proteins are recombinantly expressed, they are not glycosylated as it occurs in the native system. This affords the characterization of the specific elastic properties of these amino acid sequences and makes our study comparable to previous research. Unfolding force histograms (Fig. 3) showed that the unfolding force of CcC1Link (398 ± 47 pN) was very similar to that of the same cohesin alone, without linker sequences (425 ± 9 pN or 403 ± 64 pN (11,20), at the same pulling speed). Similarly, the $\Delta L_C$ also appears to remain unchanged (48.4 ± 1.1 nm for CcC1Link and 48.5 ± 0.2 nm for CcC1) (11)). Taken together, these results indicate that the presence of linker sequences neither affects the mechanical stability of cohesins, as reflected by their conserved mechanical stability, nor the location of the mechanical clamp, according to the unchanged $\Delta L_C$.

The mechanical stability of CcC1 was found to be lower than that of a fragment containing cohesins 1 to 7 from the same scaffoldin (CcC1-7) (11). Since linker sequences do not appear to affect the mechanical stability of cohesins, this difference was hypothesized to arise either from additional interactions between adjacent cohesin modules in the same scaffoldin or to the presence of cohesin modules of higher mechanical stability. To test these two possibilities, we studied the mechanical behavior of two constructs, one carrying a fragment of the scaffoldin containing the first three cohesins from *C. cellulolyticum* CipC (CcC123) and another carrying the third cohesin from the same scaffoldin (CcC3).

The unfolding force distribution of the CcC123 construction (Fig. 3) clearly shows two distinct populations: one with a mechanical stability of 404 ± 64 pN and another one centered at 627 ± 79 pN. The former can be attributed to the unfolding of CcC1, according to the similar average unfolding force (both isolated and with linkers), while the high mechanical stability peak shows a similar unfolding force to that of the individual CcC3 cohesin (592 ± 125 pN) as well as that found for the CcC1-7 (592 ± 111 pN (11)). Furthermore, the contour length distributions for both proteins (48.5 ± 1.3 nm for CcC123 and 48.5 ± 1.3 nm for CcC3) are similar to those found for CcC1, CcC1Link and CcC1-7, which indicates that the mechanical clamp is unaffected by the presence of several cohesins in the construct. The analysis by SMD of a structural model of CcC3 cohesin also yielded similar results to those found experimentally (Fig. S4), thus supporting our findings.

**Docking interaction does not alter cohesin mechanical stability** - Protein-ligand interactions often modify significantly the thermodynamic stability of proteins, as is the case for cohesin and dockerin (21). Although
thermodynamic and mechanical stabilities are not generally correlated, including the case of cohesins (11), it has also been observed that the interaction of a protein with its ligand (even with small molecules) can also modify the mechanical stability of a protein (22,23). The cohesin-dockerin interaction takes place in a region far away from the cohesin mechanical clamp (11,24) (Fig. 4, Fig. S5). Considering that in other systems long range interactions provide a mechanical stabilization (25) it is important to test this possibility.

To this end, we stretched CtA2, from the external region, and CtA7, from the connecting region, in the presence of 5-fold molar excess of the Cel8A cellulase from *C. thermocellum*, which carries a dockerin capable of interacting with both cohesins (Fig. S5).

The mechanical stability of both cohesins alone (179 ± 34 pN and 515 ± 69 pN for CtA2 and CtA7, respectively) were found to be very similar to that obtained when Cel8A was present during the stretching experiment (195 ± 42 pN and 506 ± 87 pN for CtA2 and CtA7 in the presence of dockerin, respectively) (Fig. 4B). Moreover, the Δ$L_c$ for both cohesins at each condition was identical (Fig. 4C-D), which indicates that upon binding to dockerin, the mechanical stability and the mechanical clamp of cohesins (from both the external and connecting regions) remain unchanged.

Nevertheless, since the unfolding traces show no features to guarantee that Cel8A dockerin is interacting with the cohesin molecule being stretched, we performed SMD simulations of the mechanical unfolding of the same cohesins in the presence of a dockerin. Since the initial structure is known, SMD simulations would also allow us to study the two different binding modes of dockerin and cohesin (26). In this context, the mechanical stability obtained by SMD for CtA2 interacting with a dockerin in the two modes (489 ± 51 pN for PDB code 1OHZ (24), and 476 ± 91 pN for PDB code 2CCL (26)) was found to be similar to those obtained for CtA2 alone (470 ± 80 pN for PDB code 1ANU, (8,11)) in agreement with our experiments (Fig. 4E-F). A similar result was obtained for a CtA7 structure interacting with the same dockerin obtained by modeling (621 ± 60 pN for the model interaction and 710 ± 120 pN for CtA7, PDB code 1AOH (7,11)) (Fig. 4).

These results allow us to conclude that the mechanical stability of cohesins (from both external and connecting regions) is not affected by dockerin binding, irrespective of the binding mode and, in agreement with the experimental results described above (those from CtA2 and CtA7 in the presence of Cel8A cellulase), is only determined by the intrinsic mechanical properties of the cohesin module itself.

Refolding kinetics of cohesin modules - The high mechanical stability of cohesins reduces the probability of their mechanical unfolding. Nevertheless, when unfolding occurs, rapid refolding may provide a way to restore a functional system. Therefore, refolding may represent a complementary mechanism to maintain an active cellulosome in the presence of mechanical stress. To examine this aspect, we studied the kinetics of cohesin refolding after mechanical unfolding as well as how this parameter relates to the mechanical hypothesis of the cellulosome.

A three-pulse force-clamp protocol was used to study cohesin refolding. First, the protein was unfolded by applying a high force thus allowing to directly monitor the number of unfolded modules. This was followed by a relaxation step where the applied force was quenched. Finally, a second unfolding force pulse was applied to count the number of refolded modules. A homopolyprotein consisting of eight tandem repeats of either CtA1 or CtA7 was used in these experiments. The fraction of refolded modules was studied as a function of the relaxation time for each protein (Fig. 5) and an exponential function was fitted to the data to obtain the observed rate constants. The rate constants for the refolding of both proteins (3.8 ± 0.9 s⁻¹ for CtA1 and 2.8 ± 0.3 s⁻¹ for CtA7) was rather similar and comparable to those of other proteins subjected to mechanical stress (27). The factor $A_0$, the expected maximum refolding fraction, is slightly higher for CtA1 (42 ± 4 %) than that of CtA7 (33.8 ± 1.4 %). In both cases, this parameter is clearly below 100% which suggests that the refolding efficiency of mechanically unfolded cohesins is low.
DISCUSSION

We have used here protein nanomechanics to comprehensively explore the mechanical hypothesis of the cellulosome and testing critical components of the system.

The nanomechanical characterization in this study has been performed using constructions that ensure that force is applied in the same geometry as the one present in the cohesins from natural cellulosomes. Thus, the results obtained directly reflect the mechanical behaviour of these elements in their physiological environment.

The nanomechanical analysis of cohesins from *C. thermocellum* and *A. cellulolyticus* primary scaffoldins revealed that their mechanical stability is clearly related to their position in the corresponding scaffoldin: cohesins from the connecting region show much higher mechanical stability than those from the external region. This is in agreement with our working hypothesis, despite the sequence diversity observed among cohesins from different microorganisms and even among cohesins from the same scaffoldin.

It is interesting to note that sequence identity and mechanical stability in the studied cohesins show two clearly different trends. According to sequence identity, cohesins in the same scaffoldin are more similar among them than to any of the cohesins from the other scaffoldin. On the other hand, according to the nanomechanical analysis reported here and in previous studies (11) the mechanical stability of cohesins in the connecting region of both scaffoldins is higher and more similar between them than that of the cohesins from the external region, despite the sequence differences. This finding reinforces the idea that mechanical stability of cohesins is associated to the region of the scaffoldin where they appear and that this seems to be of general applicability to cellulosomes.

When analysing the mechanical stability of *A. cellulolyticus* secondary scaffoldins, we observed that AcB4 showed a mechanical stability comparable to that of cohesins in the external region. At a first glance, this seems to contradict the mechanical hypothesis. However, two important considerations must be taken into account: first, this cohesin binds not to a dockerin-bearing enzyme but to an enzyme-laden scaffoldin ScaA, and second, that the cellulose activity mainly arises from the targeting effect of the CBM and the proximity among the enzymes (12,13). Thus, the release of an enzyme-loaded ScaA is not expected to strongly affect the activity of the system as long as no strong synergies exist among the various ScaA molecules bound to the same ScaB scaffoldin.

Taken together, these results indicate that the mechanical hypothesis of the cellulosome only applies to primary scaffoldins. In fact, even those microorganisms that produce cell surface-anchored cellulosomes were found to produce free cellulosomes (14,28-30).

Our analysis of the effect of the presence of linkers and multiple cohesins on the mechanical stability of a cohesin module indicates that the mechanical stability of cohesins in a scaffoldin is mainly determined by its intrinsic mechanical stability, which is independent of the presence of other elements. Accordingly, the results obtained for CcC123 closely resemble the sum of these from the CcC1 and CcC3 cohesins. This behaviour is different from that observed previously when stretching CcC1-7 (31). In that case, a unimodal distribution was observed as opposed to the bimodal distribution found here with CcC123. Nevertheless, it should be taken into account that, since in CcC1-7 no single-molecule markers were present in the fusion protein, CcC1 was there a terminal module which reduces the probability of being observed in AFM-SMFS (in fact the CcC1-7 recording shown in (11) contains 5 cohesin modules). This is due to uncontrolled point of attachment (*i.e.*, picking location) inherent to the AFM experiments (CcC1 would not be stretched in those cases). Alternatively, it may have been either misfolded, due to surface interactions, or stretched with a different geometry (hampering the identification and, as a result, its inclusion in the analysis). All these situations would result in an underrepresentation of this module in the results, thus leading to a unimodal distribution.

Finally, we observed that cohesins show a reduced refolding capability after mechanical unfolding. Since the refolding efficiency seems to be independent of the cohesin position in the
scaffoldin protein, this may well be a general feature of cohesins.

One possible reason for this low refolding efficiency might be domain swapping events occurring in these nearly-identical domains. Our experimental results do not rule out the possibility that the multiple cohesin copies found in close contact may affect their refolding capabilities. It has been shown that it is possible for a polyprotein to form an aberrant fold (a kind of "superfold" formed by the sequence corresponding to two consecutive modules plus the linker in between) although with low probability (27). It must be noted that, in natural scaffoldins, cohesin modules are separated by linkers of various lengths that are expected to be glycosylated. These may act as spacers that effectively separate the sequence of one cohesin from the adjacent one and therefore may avoid the formation of such aberrant folds. In our experiments, we did not observe any unfolding step after refolding that might correspond to the length of 2 or more modules. During the second unfolding step, the presence of steps of very short duration and highly variable length was commonly observed. This feature has been observed in similar experiments with I27 polyprotein and were attributed to minimum energy compact structures formed during in the folding pathway (32).

Upon the formation of the substrate-cellulosome-bacteria complex, other non-covalent interactions are established by the CBM and other cohesin-dockerin pairs. The mechanical stability of the CipA XDock-type II cohesin interaction that drives the C. thermocellum cellulosome assembly on its surface has not been characterized by SMFS. Still, the Ruminococcus flavefaciens type III coh-Dock interaction, used by the cell to anchor to the cellulosic substrate, was shown to resist extremely high forces, even (higher than many cohesins, at a wide range of loading rates (33) while in a computational study (34) it was shown that mechanical cohesin-dockerin unbinding is highly anisotropic and that when pulling from the ends that would be loaded in physiological conditions, unfolding forces were higher and unfolding patterns were more defined.

On the other hand, the mechanical stability of the CBM-cellulose interaction has also been experimentally addressed (35) and found relatively low unbinding forces. Nevertheless, since multiple binding sites are available on the surface of cellulose, it is reasonable to expect that the cellulosome would be effectively anchored to the cellulose for extended periods of time, thus exposing cohesins to forces.

Taken together, our results show that, despite the wide sequence diversity and cellulosome complexity, cohesin modules from the connecting region of primary scaffoldins show high mechanical stability. Furthermore, this mechanical stability is intrinsically determined by each cohesin module alone and is not affected by the presence of linkers or other cohesin modules, as in a scaffoldin, or by the interaction with dockers as it occurs in the cellulosome complex. Thus, the mechanical stability of cohesin modules in cellulosomal scaffoldins is determined entirely by its intrinsic properties of each individual cohesin module. Considering their low efficiency of refolding, the high mechanical stability of connecting cohesins appears to provide a means to prevent cohesin unfolding under mechanical stress so that enzyme release would be prevented and cellulosome activity would be preserved.

**EXPERIMENTAL PROCEDURES**

**Protein engineering** - The heteropolyproteins used for AFM-SMFS were similar to those reported previously (11) where the selected cohesin is cloned between several repeats of the single-molecule marker I27 from human cardiac titin in the pAFM vector (36). The homopolyproteins contained 8 repeats of the cohesin of interest as well as an N-terminal His tag and two C-terminal cysteine residues.

**Protein expression and purification** - Proteins were expressed in E. coli BL21 (DE3), BLR (DE3) or C41 (DE3) (37) in LB media, and IPTG was used for induction of expression. Protein purification was performed as described previously (11) by chromatography using
immobilized metal-ion affinity chromatography on a nickel-nitrioltriacetic acid (Ni-NTA) column (GE Healthcare). When needed, further purification was performed by size exclusion or anion exchange chromatography.

**AFM-SMFS** - The AFM-SMFS experiments were carried out in either a custom-made AFM described previously (38) or in an Atomic Force Spectrometer (Luigs & Neumann, GmbH). The spring constant of each cantilever (BL-RC cantilevers, Olympus, Tokyo, Japan) was calibrated in buffer using the thermal method (39). Most experiments were carried out on Ni-NTA coverslips like those previously described (31,40). Alternatively, experiments to study the dockerin-binding effect on cohesin were done on gold surfaces (Arrandee, Werther, Germany).

Purified protein samples were incubated in PBS or for the experiments examining the dockerin binding effect in 50 mM Tris, 300 mM NaCl, 1 mM CaCl$_2$, pH 7.5 for 15 min for attachment. Most experiments were performed at a constant pulling velocity of 400 nm/s. Data were collected and analyzed in Igor Pro 6 (Wavemetrics Inc., Oswego, OR) using described protocols (11). The set of criteria for single-molecule identification was as described previously (11); briefly: first, only traces showing the minimum number of I$_{27}$ unfolding events (3 for CcC123 and 4 in the other cases) were selected for the analysis to guarantee that the cohesin module was stretched plus the unfolding event of a cohesin. Once the mechanical fingerprint ($F_U$ and $\Delta L_C$) of the cohesin was well characterized, clean recordings with at least 2 but not more than 5 (4 in the case of CcC123) I$_{27}$ unfolding events and with total length compatible with that of the protein were included in the analysis. The obtained values were plotted in histograms and fitted to Gaussian curves. The values shown represent the average ± standard deviation of the Gaussian fitting.

Refolding experiments were performed in the force-clamp mode using a three-pulse protocol. Only those traces showing 5 to 8 unfolding events during the first unfolding pulse and that unfolded to the same total length in the second unfolding pulse were considered for this analysis. As the exponential equation was fit to the fraction of refolded modules as a function of relaxation time and the associated error was calculated by bootstrapping from the data obtained.

**Structural modeling** - The template and query sequences were aligned by using Clustal Omega (41) and used as input for MODELLER to generate the structural model (42).

**SMD Simulations** - Molecular dynamics (MD) simulations were carried out using implicit solvent according to the Generalized Born Surface Area approach (43) as described previously (11). For SMD, a restraint with force constant of 5 kcal/mol·Å$^2$ was applied to the first and last $C_\alpha$ and their distance was increased at a rate of 1 Å/ps. The trajectories obtained were analyzed in VMD (44).

**Monte Carlo simulations** - The kinetic mechanical unfolding parameters were estimated according to a method previously described (45,46). The $\Delta L_C$ was set to 49 nm and the persistence length was fixed to 0.4 nm. For best comparison with the heteropolyprotein constructions, CtA9 was simulated with an initial molecule length of 100 nm (since the cohesin typically unfolds last, after several I$_{27}$ modules are unfolded), while in the case of CtA1 an initial length of 4 nm was used. Best parameters were chosen as those that more closely reproduced the unfolding force at all pulling velocities according to its lower $\chi^2$. The values shown in the text correspond to the average ± standard deviation of all best fitting parameters obtained by bootstrapping.
Acknowledgements: We thank D.V. Laurents and the members of MCV lab for the critical reading of the manuscript. We also appreciate the technical assistance of M Shamshoum and the scientific contributions of I. Noach and Y. Barak (Chemical Research Support, Weizmann Institute).

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: MCV conceived the study; MCV, AGP, EAB, SM, YV designed the experiments; AGP performed the experiments and processed the raw data; AGP and MCV wrote the manuscript with contributions from co-authors. Authors gave approval to final manuscript version.

REFERENCES

1. Bayer, E. A., Lamed, R., and Himmel, M. E. (2007) The potential of cellulases and cellulosomes for cellulosic waste management. Curr Opin Biotechnol 18, 237-245
2. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66, 506-577, table of contents
3. White, B. A., Lamed, R., Bayer, E. A., and Flint, H. J. (2014) Biomass utilization by gut microbiomes. Annu Rev Microbiol 68, 279-296
4. Demain, A. L., Newcomb, M., and Wu, J. H. (2005) Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 69, 124-154
5. 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
6. Artzi, L., Bayer, E. A., and Morais, S. (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 15, 83-95
7. Tavares, G. A., Beguin, P., and Alzari, P. M. (1997) The crystal structure of a type I cohesin domain at 1.7 A resolution. J Mol Biol 273, 701-713
8. Shimon, L. J., Bayer, E. A., Morag, E., Lamed, R., Yaron, S., Shoham, Y., and Frolow, F. (1997) A cohesin domain from Clostridium thermocellum: the crystal structure provides new insights into cellulosome assembly. Structure 5, 381-390
9. Leibovitz, E., and Beguin, P. (1996) A new type of cohesin domain that specifically binds the dockerin domain of the Clostridium thermocellum cellulosome-integrating protein CipA. J Bacteriol 178, 3077-3084
10. Bustamante, C., Chemla, Y. R., Forde, N. R., and Izhaky, D. (2004) Mechanical processes in biochemistry. Annu Rev Biochem 73, 705-748
11. Valbuena, A., Oroz, J., Hervas, R., Vera, A. M., Rodriguez, D., Menendez, M., Sukowska, J. I., Cieplak, M., and Carrion-Vazquez, M. (2009) On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proc Natl Acad Sci U S A 106, 13791-13796
12. Fierobe, H. P., Mechaly, A., Tardif, C., Belaich, A., Lamed, R., Shoham, Y., Belaich, J. P., and Bayer, E. A. (2001) Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes. J Biol Chem 276, 21257-21261
13. Fierobe, H. P., Bayer, E. A., Tardif, C., Czjek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., and Belaich, J. P. (2002) Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J Biol Chem 277, 49621-49630
14. Hamberg, Y., Ruimy-Israeli, V., Dassa, B., Barak, Y., Lamed, R., Cameron, K., Fontes, C. M., Bayer, E. A., and Fried, D. B. (2014) Elaborate cellulosome architecture of Acetivibrio cellulolyticus revealed by selective screening of cohesin-dockerin interactions. PeerJ 2, e636
15. Bayer, E., Smith, S., Noach, I., Alber, O., Adams, J., Lamed, R., Shimon, L., and Frolow, F. (2009) Can we crystallize a cellulosome. Biotechnology of lignocellulose degradation and biomass utilization, 183-205

16. Gerwig, G. J., de Waard, P., Kamerling, J. P., Vliegenthart, J. F., Morgenstern, E., Lamed, R., and Bayer, E. A. (1989) Novel O-linked carbohydrate chains in the cellulase complex (cellulosome) of Clostridium thermocellum. 3-O-Methyl-N-acetylglucosamine as a constituent of a glycoprotein. J Biol Chem 264, 1027-1035

17. Gerwig, G. J., Kamerling, J. P., Vliegenthart, J. F., Morgenstern, E., Lamed, R., and Bayer, E. A. (1993) The nature of the carbohydrate-peptide linkage region in glycoproteins from the cellulosomes of Clostridium thermocellum and Bacteroides cellulosolvens. J Biol Chem 268, 26956-26960

18. Noach, I., Levy-Assaraf, M., Lamed, R., Shimon, L. J., Frolow, F., and Bayer, E. A. (2010) Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad. J Mol Biol 399, 294-305

19. Li, H., Wang, H. C., Cao, Y., Sharma, D., and Wang, M. (2008) Configurational entropy modulates the mechanical stability of protein GB1. J Mol Biol 379, 871-880

20. Chwastyk, M., Galera-Prat, A., Sikora, M., Gomez-Sicilia, A., Carrion-Vazquez, M., and Cieplak, M. (2014) Theoretical tests of the mechanical protection strategy in protein nanomechanics. Proteins 82, 717-726

21. Salama-Alber, O., Jobby, M. K., Chitayat, S., Smith, S. P., White, B. A., Shimon, L. J., Lamed, R., Frolow, F., and Bayer, E. A. (2013) Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural butresses. J Biol Chem 288, 16827-16838

22. Junker, J. P., Ziegler, F., and Rief, M. (2009) Ligand-dependent equilibrium fluctuations of single calmodulin molecules. Science 323, 633-637

23. Oroz, J., Valbuena, A., Vera, A. M., Mendieta, J., Gomez-Puertas, P., and Carrion-Vazquez, M. (2011) Nanomechanics of the cadherin ectodomain: "canalization" by Ca2+ binding results in a new mechanical element. J Biol Chem 286, 9405-9418

24. Carvalho, A. L., Dias, F. M., Prates, J. A., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M., Romao, M. J., and Fontes, C. M. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex. Proc Natl Acad Sci U S A 100, 13809-13814

25. Cao, Y., Yoo, T., Zhuang, S., and Li, H. (2008) Protein-protein interaction regulates proteins' mechanical stability. J Mol Biol 378, 1132-1141

26. Carvalho, A. L., Dias, F. M., Nagy, T., Prates, J. A., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M., Romao, M. J., Fontes, C. M., and Gilbert, H. J. (2007) Evidence for a dual binding mode of dockerin modules to cohesins. Proc Natl Acad Sci U S A 104, 3089-3094

27. Carrion-Vazquez, M., Oberhauser, A. F., Fowler, S. B., Marszalek, P. E., Broedel, S. E., Clarke, J., and Fernandez, J. M. (1999) Mechanical and chemical unfolding of a single protein: a comparison. Proc Natl Acad Sci U S A 96, 3694-3699

28. Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., Jr., McKeown, C. K., Lankford, P. K., Samatova, N. F., and Mielenz, J. R. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis. PloS One 4, e5271

29. Xu, Q., Resch, M. G., Podkaminer, K., Yang, S., Baker, J. O., Donohoe, B. S., Wilson, C., Klingeman, D. M., Olson, D. G., Decker, S. R., Giannone, R. J., Hettich, R. L., Brown, S. D., Lynd, L. R., Bayer, E. A., Himmel, M. E., and Bomble, Y. J. (2016) Dramatic performance of
Clostridium thermocellum explained by its wide range of cellulase modalities. Sci Adv 2, e1501254
30. Artzi, L., Dassa, B., Borovok, I., Shamshoum, M., Lamed, R., and Bayer, E. A. (2014) Cellulosomics of the cellulytic thermophile Clostridium clariflavum. Biotechnol Biofuels 7, 100
31. Hervas, R., Oroz, J., Galera-Prat, A., Goni, O., Valbuena, A., Vera, A. M., Gomez-Sicilia, A., Losada-Urzaiz, F., Uversky, V. N., Menendez, M., Laurens, D. V., Bruix, M., and Carrion-Vazquez, M. (2012) Common features at the start of the neurodegeneration cascade. PLoS Biol 10, e1001335
32. Garcia-Manyes, S., Dougan, L., Badilla, C. L., Brujic, J., and Fernandez, J. M. (2009) Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. Proc Natl Acad Sci U S A 106, 10534-10539
33. Schoeler, C., Malinowska, K. H., Bernardi, R. C., Milles, L. F., Jobst, M. A., Durner, E., Ott, W., Fried, D. B., Bayer, E. A., Schulten, K., Gaub, H. E., and Nash, M. A. (2014) Ultrastable cellulosome-adhesion complex tightens under load. Nat Commun 5, 5635
34. Wojciechowski, M., Thompson, D., and Cieplak, M. (2014) Mechanostability of cohesin-dockerin complexes in a structure-based model: anisotropy and lack of universality in the force profiles. J Chem Phys 141, 245103
35. Zhang, M., Wang, B., and Xu, B. (2013) Measurements of single molecular affinity interactions between carbohydrate-binding modules and crystalline cellulose fibrils. Phys Chem Chem Phys 15, 6508-6515
36. Steward, A., Toca-Herrera, J. L., and Clarke, J. (2002) Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. Protein Sci 11, 2179-2183
37. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol 260, 289-298
38. Valbuena, A., Oroz, J., Vera, A. M., Gimeno, A., Gomez-Herrero, J., and Carrion-Vazquez, M. (2007) Quasi-simultaneous imaging/pulling analysis of single polyprotein molecules by atomic force microscopy. Rev Sci Instrum 78, 113707
39. Hutter, J. L., and Bechhoefer, J. (1993) Calibration of atomic-force microscope tips. Review of Scientific Instruments 64, 1868-1873
40. Vera, A. M., and Carrion-Vazquez, M. (2016) Direct Identification of Protein-Protein Interactions by Single-Molecule Force Spectroscopy. Angew Chem Int Ed Engl 55, 13970-13973
41. Sievers, F., and Higgins, D. G. (2014) Clustal omega. Curr Protoc Bioinformatics 48, 3 13 11-16
42. Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234, 779-815
43. Tsui, V., and Case, D. A. (2000) Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 56, 275-291
44. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. J Mol Graph 14, 33-38, 27-38
45. Rief, M., Fernandez, J. M., and Gaub, H. E. (1998) Elastically coupled two-level systems as a model for biopolymer extensibility. Physical Review Letters 81, 4764
46. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997) Reversible unfolding of individual titin immunoglobulin domains by AFM. Science 276, 1109-1112
This work was supported by a Seventh Framework Programme in Nanosciences, Nanotechnologies, Materials & New Production Technologies (7PM-NMP 2013-17, 604530-2) and the ERA-IB-ERANET-2013-16 (EIB.12.022) through the Spanish MINECO (PCIN-2013-011-C02-01). AGP acknowledges financial support from an FPU fellowship from Spanish MECD.

The abbreviations used are: AFM, atomic force microscope; SMFS, single-molecule force spectroscopy; SMD, steered molecular dynamics; CBM, carbohydrate binding module; SLH, surface layer homology;

FIGURE LEGENDS

FIGURE 1. Cohesins on the connecting region of primary scaffoldins show high mechanical stability. A) Representation of C. thermocellum scaffoldin CipA where the modules studied in this work are highlighted. The inset describes the symbols used. Unfolding force (left) and $\Delta L_C$ (right) histograms for CtA1 (in green, n=70) and CtA9 (blue, n=59). A representative force-extension recording of each protein is plotted at the right. A schematic drawing of the protein used for AFM-SMFS analysis is shown at the right of the recordings where five I27 repeats are used as single-molecule markers (black in force-extension traces). B) Schematic representation of A. cellulolyticus ScaA scaffoldin. The unfolding force (left) and $\Delta L_C$ (right) histograms of AcA3 (yellow, n=187) and AcA4 (red, n=37). C) Dynamic force spectra of CtA1, CtA9 and the single-molecule marker I27 (black). Open symbols represent unfolding forces calculated using Monte Carlo simulations.

FIGURE 2. Mechanical properties of representative cohesins from secondary scaffoldins of A. cellulolyticus. A) Representation of two A. cellulolyticus secondary scaffoldins (ScaB and ScaC), where the selected cohesins studied are highlighted and enumerated. The inset describes the symbols used. B-C) Unfolding force (B) and $\Delta L_C$ (C) histograms of AcB4 (yellow, n = 59) and AcC3 (blue, n = 58) fitted to Gaussian curves. D) Schematic representation of the constructions used for this analysis and force-extension recordings of their unfolding. Cohesin unfolding is highlighted in the corresponding color while black peaks show the unfolding of the single-molecule markers. Above the recordings, schematics of constructions used are shown, similar to Fig. 1A.

FIGURE 3. The intermodular linkers and the presence of multiple cohesins do not affect the mechanical stability of cohesin modules. A) Schematic representation of C. cellulolyticum CipC scaffoldin. The modules studied are highlighted, and lines indicate the fragment of the scaffoldin included in each analysis. B-C) Unfolding force (left) and $\Delta L_C$ (right) histograms of CcC1Link (n = 105; B), CcC123 (n = 586; C) and CcC3 (n = 110). A schematic representation of the construction used in each case is also shown. Data for CcC1 and CcC1-7 correspond to those reported in (11), which were plotted here for comparison.

FIGURE 4. Dockerin binding does not affect the mechanical stability of cohesins. A) Schematic drawing of the C. thermocellum CipA scaffoldin where CtA2 (yellow) and CtA7 (red) are highlighted. B-D) Unfolding force (B) and $\Delta L_C$ histograms of CtA2 (C) and CtA7 (D) in the absence (top, n = 69 and n = 102, respectively) or presence (bottom, n = 185 for CtA2 and n= 40 for CtA7) of Cel8A (a C. thermocellum enzyme containing the appropriate dockerin for interaction with the test cohesin). E-G) Superposition of 1OHZ (E), 2CCL (F) and CtA7-Dockerin model (G) force-extension traces calculated using SMD. Red triangles indicate the first frame of the trajectory where cohesin-dockerin interaction was lost. The structure of each complex is shown above the trajectories where the N-terminus of each component has been marked with a blue sphere.
FIGURE 5. Refolding kinetics of cohesin modules. A) Simultaneous extension-time (top) and force-time (bottom) traces for CtA1 (red) and CtA7 (blue). For CtA1, a 90 pN for 2 s force pulse was applied to unfold the protein (shown in higher magnification in the inset), while a ramp from 150 to 550 pN in 2s was used to unfold CtA7. Then force was reduced during the relaxation time and a second unfolding step was applied to probe the number of refolded modules. B) Fraction of refolded modules as a function of the relaxation time.
Figure 1

A) Connecting region

External region

Connecting region

Cell

CipA

Type I cohesin

Type II dockerin

Catalytic unit

CBM

I27

B) External region

Connecting region

ScaA

C)

Force (pN)

Pulling velocity (nm/s)

CtA1

CtA9

AcA3

AcA4

I27

Force (pN)

Pulling velocity (nm/s)

10

40

400

1000

4000

100

300

500

100

Figure 1
Figure 2

A

Cell

ScaC

ScaB

ScaA

Type I cohesin

Type I dockerin

Type II cohesin

SLH

I27

B

C

D

Force (pN)

100 200 300 400 500

Norm. Freq.

AcB4

AcC3

Norm. Freq.

100 200 300 400 500

ΔL_c (nm)

200 pN

50 nm
Figure 3

A

B

C

**Figure 3**

A diagram showing the structure of CipC with labeled regions CcC1, CcC1-7, CcC123, CcC12, CcC3, and CcC1Link.

B

A graph showing the normalized frequency distribution of force (pN) and displacement (ΔLc nm) for CcC1Link, CcC1, and CcC1-7.

C

Another graph showing the normalized frequency distribution of force (pN) and displacement (ΔLc nm) for CcC123 and CcC3.
Figure 4

A

B

C

D

E

F

G

1OHZ

2CCL

10 nm

500 pN

10 nm

500 pN

10 nm

500 pN

10 nm

500 pN

10 nm
Figure 5

A

Protein extension

100 nm

Relaxation time (s)

Fraction refolded (%)

CtA1

CtA7

2 seconds

Force (pN)

0

100 nm

100 nm

2 seconds

2 seconds

90

0

550

150

B

Fraction refolded (%)

Relaxation time (s)

CtA1

CtA7

0

10

20

30

40
