NANO LETTERS

November 2015
Volume 15, Number 11
pubs.acs.org/NanoLett

Mechanics of an ultrastable protein complex

Slippery to sticky transition of hydrophobic nanochannels

Nanoscale and single-dot patterning of colloidal quantum dots

ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Mapping Mechanical Force Propagation through Biomolecular Complexes

Constantin Schoeler,[†,#] Rafael C. Bernardi,[‡,#] Klara H. Malinowska,[†] Ellis Durner,[†] Wolfgang Ott,[†,§] Edward A. Bayer,[∥] Klaus Schulten,[‡,⊥] Michael A. Nash,[*,†] and Hermann E. Gaub[†]

[†]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany

[‡]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

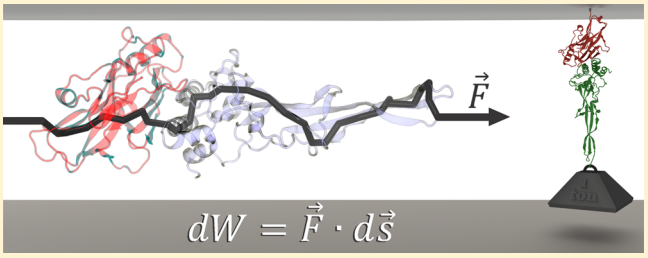
[§]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany

[∥]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel

[⊥]Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States

Supporting Information

ABSTRACT: Here we employ single-molecule force spectroscopy with an atomic force microscope (AFM) and steered molecular dynamics (SMD) simulations to reveal force propagation pathways through a mechanically ultrastable multidomain cellulosome protein complex. We demonstrate a new combination of network-based correlation analysis supported by AFM directional pulling experiments, which allowed us to visualize stiff paths through the protein complex along which force is transmitted. The results implicate specific force-propagation routes nonparallel to the pulling axis that are advantageous for achieving high dissociation forces.



KEYWORDS: *Force propagation, single molecule force spectroscopy, steered molecular dynamics, network analysis, cohesin−dockerin*

Mechanical forces play a fundamental role in biological systems. Cells are able to sense and respond to mechanical cues in their environment by, for example, modulating gene expression patterns,[1] reshaping the extracellular matrix,[2] or exhibiting differential biochemical activities.[3] At the molecular level, these behaviors are governed by mechanically active proteins. Such proteins are able to sense and respond to force by undergoing conformational changes,[4] exposing cryptic binding sequences,[5] acting synergistically with ion channels,[6] or modulating their function in a variety of ways.[7−9]

Experimental methods including AFM single-molecule force spectroscopy (SMFS) allow direct measurement of molecular mechanical properties. These studies have demonstrated the importance of the shear topology involving parallel breakage of hydrogen bonds in providing mechanical stability to protein folds.[10,11] Many globular domains and protein complexes also exhibit a directional dependence in unfolding mechanics, consisting of stiff and soft axes.[12−18] Pulling geometry can be defined by controlling the positions of the chemical linkages between protein monomer units through a variety of bioconjugate techniques.

Primary sequences of mechanically active proteins are extremely diverse, essentially rendering them undetectable by conventional bioinformatics approaches. Yet, another computational approach, namely, molecular dynamics (MD), allows sampling of structural conformations of large and frequently mechanostable protein complexes.[19,20] Analysis of these conformations from MD trajectories have recently led to the development of network-based correlation methods for investigating signal transmission and allosteric regulation in proteins.[21−23] In network models, local correlations of positional fluctuations in a protein are represented as a web of inter-residue connections. Within such a network, the behavior of nodes that are highly correlated and within close physical proximity can be analyzed to obtain the shortest path between two network nodes (i.e., amino acids). This analysis helps to identify which connecting residues are most important for intramolecular communication.[23−25] Examination of multiple pathways, also known as suboptimal paths, within an acceptable deviation from the optimal path helps to detect the web of nodes critical for transmission of information.

Among MD methods, steered molecular dynamics (SMD) simulations in which external forces are used to explore the response and function of proteins have become a powerful tool especially when combined with SMFS.[6] SMD has been successfully employed in a wide range of biological systems, from the investigation of protein mechanotransduction,[5,26] to

Received: July 9, 2015
Revised: August 5, 2015
Published: August 11, 2015

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

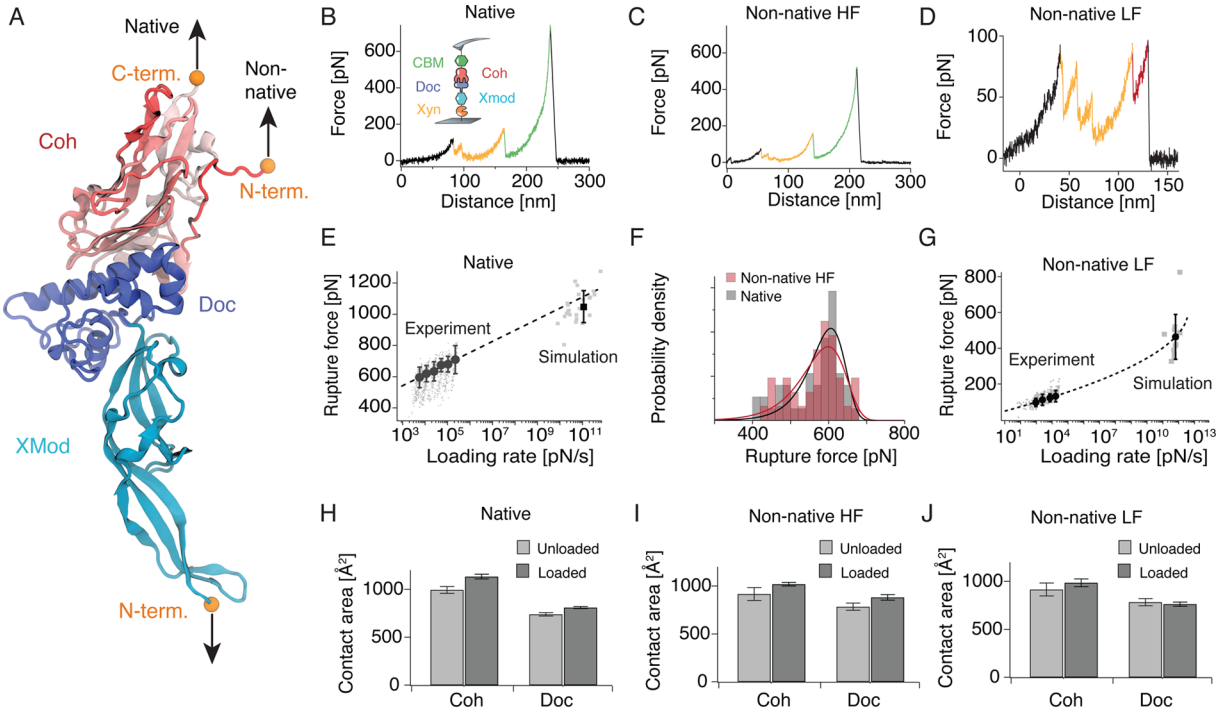


Figure 1. Single molecule force spectroscopy and steered molecular dynamics of XMod-Doc:Coh in two pulling configurations. (A) Crystal structure of the XMod-Doc:Coh complex (PDB 4IU3) with orange spheres marking the termini where force was applied. (B) Experimental unfolding trace for the native pulling configuration at a pulling speed of 1600 nm s^{-1}. The inset shows a schematic of the pulling geometry. Unfolding signatures of the Xyn and CBM marker domains are marked in orange and green, respectively. (C) Experimental unfolding trace for the non-native high force class obtained at a pulling speed of 700 nm s^{-1}. (D) Experimental unfolding trace for the non-native low force class obtained at a pulling speed of 700 nm s^{-1}. The additional 17−19 nm contour length increment attributed to N-terminal Coh unfolding is shown in red. (E) Dynamic force spectrum for XMod-Doc:Coh unbinding in the native geometry obtained from experiment and simulations. Gray points and squares represent the rupture force/loading rate pairs obtained from experiment and simulation, respectively. Black circles represent the most probable rupture force/loading rate obtained from Gaussian fits to the experimental data at six pulling speeds. The black square shows the mean rupture force and loading rate for the simulated rupture events. (F) Rupture force histograms obtained at a pulling speed of 800 nm s^{-1} for the native (gray, $n = 46$) and non-native high force class (red, $n = 48$). Fitted probability densities $p(F)$ are shown as solid black and red lines. Data for both pulling configurations were obtained with the same cantilever to minimize calibration errors. (G) Dynamic force spectrum for XMod-Doc:Coh unbinding in the non-native low force class obtained from experiments and simulation. The same representation as in (E) is used. (H,I,J) Unloaded and loaded surface contact areas for the different pulling geometries ((H) native, (I) non-native high force class, and (J) non-native low force class).

permeability of membrane channels,[27,28] and the characterization of protein−receptor interactions.[29] SMD simulations have also been used to study force propagation through proteins by employing force distribution analysis (FDA).[30,31] In FDA, all pairwise forces, which are usually calculated in MD simulations, are stored in $N \times N$ matrices, where N is the number of atoms.[32] These pairwise forces can then be used to assess a protein's response to a mechanical or allosteric signal.[33] In the FDA approach, atoms under mechanical strain are identified by subtracting forces of both loaded and unloaded states for each pair of interacting atoms.[31] However, to achieve a sufficient signal-to-noise ratio, FDA will often require exhaustive sampling of the conformational space.[32,34] FDA, therefore, requires more computational resources than usual SMD studies, which are frequently already computationally demanding. There is therefore a clear need for new analysis methods that enable visualization of force propagation pathways from a single SMD trajectory.

Here we implemented a novel combination of SMD, network-based correlation analysis, and thermodynamic fluctuation theory, supported by AFM-SMFS experiments to study force propagation through a protein complex subjected to different pulling geometries. We chose an ultrastable receptor−ligand interaction as a model system because of its remarkably high mechanical stability,[29] which effectively improves the signal-to-noise ratio. This complex consists of two interacting protein domains called cohesin (Coh) and dockerin (Doc) that maintain bacterial adhesion of *Ruminococcus flavefaciens* to cellulosic substrates. Doc is found within the same polypeptide chain as a stabilizing ancillary domain called X-module (XMod), located N-terminally of Doc. Based on its position with the *R. flavefaciens* cellulosomal network, Coh is mechanically anchored *in vivo* at its C-terminal end to the cell surface. Our prior work demonstrated that, when force is applied to the complex in the native configuration (i.e., C-terminal Coh, N-terminal XMod-Doc anchor points), the complex is extremely stable, exhibiting high rupture forces of 600−750 pN at loading rates from 1−100 nN s^{-1}.[29] Since the bulk equilibrium affinity of the complex is an unremarkable 20 nM,[35] we hypothesized that the high mechanostability is explained by a catch bond mechanism. AFM rupture force data and SMD simulations supported this prediction, where it was observed that the contact surface area of the two proteins increased as mechanical force was applied.

To characterize the mechanisms behind Coh:Doc high stability, here we additionally pulled the complex apart in a non-native configuration (i.e., N-terminal Coh, N-terminal XMod-Doc anchor points). In the non-native pulling

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

configuration, we found that the complex dissociated along two competing pathways with very different mechanical characteristics.

Our new dynamic network analysis protocol reveals how different mechanical behaviors are attributable to differences in the direction of force transmission across the binding interface. Together, the experiments and simulations depict a simple physical mechanism for achieving high complex rupture forces: the complex directs force along pathways orthogonal to the pulling axis.

Single-Molecule Pulling Experiments and SMD. For SMFS experiments, XMod-Doc was produced as a fusion protein with an N-terminal Xylanase (Xyn) domain. Coh was produced as either an N- or C-terminal fusion domain with a carbohydrate binding module (CBM). These fusion domains were used for site specific immobilization to a glass surface and AFM cantilever to achieve the two loading configurations shown in Figure 1A and further served as marker domains with known unfolding length increments to validate single-molecule interactions and sort SMFS data traces.[36]

For the native pulling configuration found *in vivo*, Coh-CBM and XMod-Doc are loaded from their C- and N-termini, respectively (Figure 1A). A representative unbinding trace for the native pulling configuration is shown in Figure 1B. We measured the loading rate dependence of complex rupture using both experimental and SMD data sets (unbinding trace from SMD shown in Figure 3A) and plotted them on a combined dynamic force spectrum (Figure 1E). The linear Bell model produced fit parameters for the effective distance to the transition state $\Delta x = 0.13$ nm, and the zero-force off rate $k_{off} = 4.7 \times 10^{-4}$ s^{-1}. Both experimental and simulation data are well described by a single Bell expression, despite the differences in loading rates between experiments and simulation. The observation suggests that the application of force does not significantly change Δx for this particular configuration.

To test the influence of pulling geometry on mechanical stability, we performed SMFS and SMD on the system where Coh was pulled from the opposite terminus (i.e., non-native N-terminus, cf. Figure 1A). Unlike the native pulling geometry, this geometry exhibited two clearly distinct unbinding pathways that are characterized by different force ranges (high or low) at which the complex dissociated. We refer to these pathways as non-native high force (HF) (Figure 1C) and non-native low force (LF) (Figure 1D).

AFM data traces classified as non-native HF showed similar characteristics as those in the native pulling configuration (cf. Figure 1B,C,F). The non-native LF traces, however, exhibited a markedly different unfolding behavior (Figure 1D). Xyn unfolding (highlighted in orange) was regularly observed, but CBM unfolding was only very rarely observed. The complex usually did not withstand forces high enough to unfold CBM when rupturing along the non-native LF path. Among non-native LF curves, we regularly found an additional contour length increment of 17−19 nm consistent with unfolding of ∼60 amino acids located at the N-terminus of Coh. This unfolding occurred immediately following Xyn unfolding (Figure 1D, red), or alternatively prior to Xyn unfolding, or with a substep (Supplementary Figure S1). Taken together, it appears that partial Coh unfolding from the N-terminus destabilizes the complex, causing lower rupture forces (Figure 1G).

The experimental rupture forces from the non-native HF class were indistinguishable from those arising in the native configuration. To confirm this, we performed additional measurements where both Coh configurations were alternately probed with the same Xyn-XMod-Doc functionalized cantilever (Supplementary Figure S2), eliminating inaccuracies introduced through multiple cantilever calibration. Most probable rupture forces at a pulling speed of 800 nm s^{-1} of 606 and 597 pN for the native configuration and non-native HF class, respectively, were determined in the Bell Evans model (Figure 1F, Supplementary eq S2), demonstrating that the native and non-native HF classes are experimentally indistinguishable.

For the LF class, we analyzed the final complex rupture event and plotted the combined dynamic force spectrum (Figure 1G). Here, simulated and experimentally observed data were not well described by a single Bell expression. In such cases nonlinear models have been developed to obtain kinetic and energetic information from dynamic force spectra.[37,38] To fit the combined data, we used the nonlinear Dudko−Hummer−Szabo (DHS) model (Supplementary eq S3) and obtained values of $\Delta x = 0.42$ nm and $k_{off} = 0.005$ s^{-1}. The DHS model further provides the free energy difference ΔG between the bound state and the transition state as a fit parameter, which was found to be $\Delta G = 129\ k_B T$. The model fit produced a distance to transition that was much longer than observed for the native configuration. Independent SMD simulations for the non-native pulling configuration were found to also lead to HF and LF unbinding scenarios (see below, Figure 4A,D, respectively).

The differential solvent contact area was calculated from SMD simulations to estimate the intermolecular contact area in the Doc:Coh complex. In the native configuration, the simulated Doc:Coh contact area increased by 14% and 9% for Coh and Doc, respectively (Figure 1H). For the non-native HF class, the contact area increased by 11% and 12% for Coh and Doc, respectively (Figure 1I). In the non-native LF class, the contact area increased by only 7% for Coh and decreased by 3% for Doc (Figure 1J). Evidently, an increased surface contact area for Doc in the native and non-native HF pathways correlated with high mechanostability of the system.

Force Propagation Theory: A Simple Model. To further understand the observed unbinding pathways, we sought to identify paths through the molecule along which the externally applied load propagates. From thermodynamic fluctuation theory,[39,40] it is known that the correlation of fluctuations of atoms i and j and the force $\mathbf{F}_i$ on atom i are related through

$$\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{1}$$

where $\Delta \mathbf{r}_i = \mathbf{r}_i(t) - \langle \mathbf{r}_i(t) \rangle$ and $\mathbf{r}_i$ is the position of atom i. The derivative on the right-hand side of eq 1 states that neighboring atoms i and j will move with high correlation due to an external force $\mathbf{F}_i$ acting on atom i if the coupling between them is strong. Hence, a given element of a correlation matrix $M_{ij} = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle$ will be large in the case of a strong interaction potential between i and j. When force is propagated through a molecule, soft degrees of freedom will be stretched out along the path of force propagation, while stiff degrees become more important for the dynamics of the system.

Consequently, paths with high correlation of motion describe the paths along which force propagates through the system. To illustrate this behavior for a toy system, we employed the NAMD[41] SMD[42] constant velocity protocol to a test pattern of identical spheres connected with harmonic springs of different

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

stiffness (Figure 2A). The position of one sphere was fixed during the simulation, while another sphere on the opposite

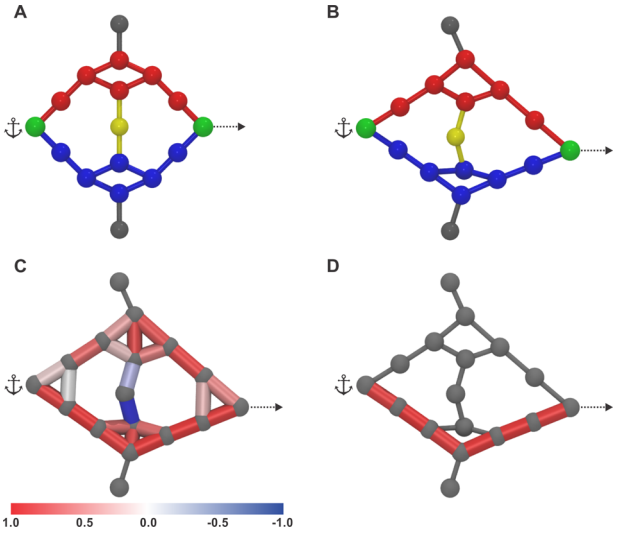


Figure 2. Network analysis test simulation. (A) Simulated pattern of atoms depicted by spheres. Connecting lines between atoms represent harmonic springs with different stiffnesses (red, k; blue, $5k$; yellow, $7.5k$; black, $10k$). The green atom was fixed (anchor), while a second green atom was withdrawn at constant speed (arrow). Black and yellow atoms and their adjacent springs were introduced to maintain the general shape of the pattern. (B) Deformed sphere pattern at the end of the simulation. (C) Edges between nodes are weighted by the corresponding correlation matrix elements. (D) The path with highest correlation of motion is shown in red.

side of the structure was withdrawn at constant velocity. The strained structure at the end of the simulation is shown in Figure 2B. We assigned weights to the lines between spheres according to the Pearson correlation coefficient C_{ij} (Supplementary eq S4) between those network nodes (Figure 2C). The Pearson correlation coefficient differs from the left-hand side of eq 1 by a normalization factor $(\langle \Delta \mathbf{r}_i^2(t)\rangle \langle \Delta \mathbf{r}_j^2(t)\rangle)^{-1/2}$ and was chosen to make our analysis mathematically more tractable. For a detailed discussion on this choice of correlation measure, see Supporting Information. In a harmonic potential approximation, the equipartition theorem can be applied to this normalization factor resulting in the following expression for C_{ij}:

$$C_{ij} = \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \sqrt{k_{i,\mathrm{eff}} k_{j,\mathrm{eff}}} \tag{2}$$

where $k_{i,\mathrm{eff}} = \left(\frac{1}{k_{x_i}} + \frac{1}{k_{y_i}} + \frac{1}{k_{z_i}} \right)^{-1}$ and k_{x_i} is the curvature of the potential on atom i in the x direction. For a full derivation, see Supporting Information. Equation 2 illustrates how Pearson correlation is a suitable measure to identify the stiff paths in our simple model. We then used dynamical network analysis implemented in VMD[49] to find the path of highest correlation (Figure 2D). As expected from eq 1, we found this path to be the one connected by the stiff springs.

Force Propagation through XMod-Doc:Coh Complex. The simple pattern of spheres validated our general approach of using local correlations to identify load-bearing pathways through networks. We next employed dynamical network analysis to understand force propagation through the XMod-Doc:Coh complex.

The dynamic networks for the native configuration (unloaded and loaded) are shown in Figure 3B,C, respectively. While the network shows multiple suboptimal paths in the unloaded scenario, the loaded case exhibits a well-defined main path along which force propagates through the system. Interestingly, in the loaded configuration, force propagates through both binding helices of Doc, which results in a force path with large normal components to the unbinding axis close to the binding interface as illustrated in Figure 3D. It had been shown for another ultrastable protein, namely, silk crystalline units, that curving force paths distribute tension through the entire system.[31] A strategy that assumes an indirect path would therefore allow the system to have more time to absorb the tension from the applied force. The result here supports the view that directing the force along a path with significant perpendicular components to the pulling axis leads to high mechanical stability. In a simple mechanical picture, a certain amount of mechanical work, namely $dW = \mathbf{F}\cdot d\mathbf{s}$, is required to separate the two binding interfaces by a distance Δz and break the interaction. In this simplified picture, $d\mathbf{s}$ points along the unbinding axis, whereas the force $\mathbf{F}$ is locally largely perpendicular to this direction. Consequently, a larger force is required to break the interaction than in a scenario where the force path would point along the unbinding axis.

To validate this picture, we repeated the same analysis for the non-native HF and non-native LF pathways. The HF simulation (Figure 4A) exhibited only a small stretching of the flexible N-terminal region of Coh and complex dissociation at approximately 800 pN and a pulling distance around 10 nm. However, the LF case shown in Figure 4D exhibited a stepwise N-terminal Coh unfolding, dissociating at a force of about 480 pN at a pulling distance of about 25 nm. This behavior confirmed our assignment of the experimentally observed 17–19 nm contour length increment to Coh unfolding up to residue 62 in PDB 4IU3.

While the experimental data did not show a detectable difference between the native configuration and the non-native HF class, the propagation of force takes place along a different pathway (Figure 4B). For N-terminal Coh pulling, helix 3 of Doc is not involved in the propagation of force as it is for the native geometry. In the native configuration, force propagates through the center of Coh, while for non-native HF the path is shifted toward the side of the molecule. Despite these differences, there is a common feature between the native and non-native HF pathways. At the binding interface, the pathway again shows pronounced components perpendicular to the unbinding axis (cf. Figure 4C), suggesting that this feature is indeed responsible for the exceptional mechanical strength observed for these two unbinding pathways.

Figure 4E shows the force propagation pathway for the non-native LF class prior to rupture. Due to the unfolding of the N-terminal Coh segment, the propagation of force is shifted even further away from the central portion of Coh than for the non-native HF class. Interestingly, force is propagated through the small helical segment of Coh (ALA167-GLN179), a portion of the molecule that is not involved in force propagation for any of the other analyzed trajectories. Unlike in the aforementioned scenarios, there is no pronounced tendency for perpendicular force components at the binding interface for the non-native LF class. In fact, the force is propagated along a path largely parallel to the pulling axis (cf. Figure 4F). In cases where force

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370–7376

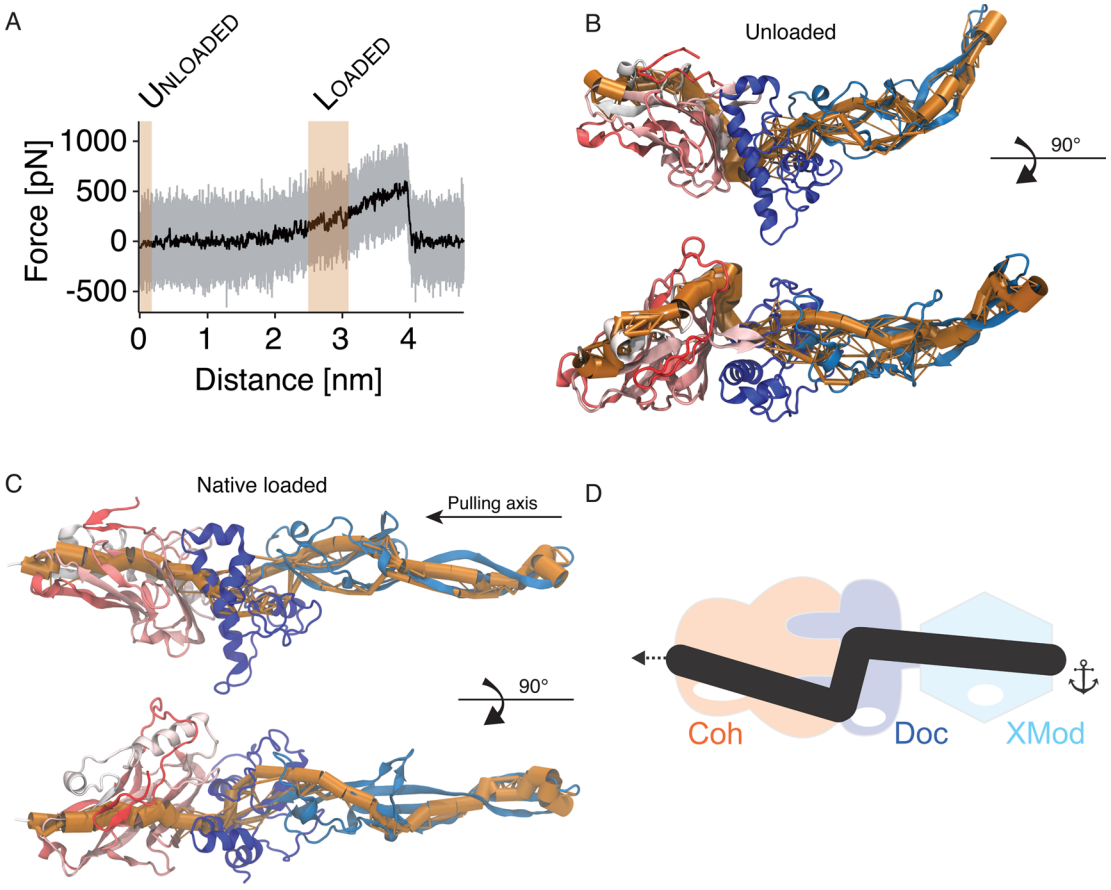


Figure 3. Force propagation through XMod-Doc:Coh in the native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray. The black line represents a moving average with a box size of 500 steps. The highlighted red areas denote the windows where dynamic networks and contact areas were calculated. (B) Network paths for the unloaded system. The thickness of the orange tube represents the number of suboptimal correlation paths passing between two nodes. (C) Network paths for the loaded system. A detailed 2D representation of the pathway, highlighting the amino acids present in the pathway, is shown in Supplementary Figure S5. (D) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis.

propagation occurs parallel to the pulling axis, as in Figure 4E, low mechanical stability was observed.

The aforementioned force propagation architecture along with the effect of increasing contact surface area upon mechanical loading combine for elevated mechanostability of the system.[29] In cases where we observed an N-terminal Coh unfolding of 62 amino acids in the non-native geometry, the system was no longer able to summon this mechanism, causing dissociation at much lower forces.

Previously, our groups have reported on a family of mechanically stable protein ligand receptor complexes that are key building blocks of cellulosomes,[29,44−46] the multienzyme complexes used by select anaerobic bacteria to digest lignocellulose. However, the molecular origins of the stability of these complexes remained largely unclear. An initial clue was obtained when, in a previous work, we were able to show that contact surface area of the two proteins increased as mechanical force was applied.[29] In a different study,[47] coarse-grained MD simulations showed much smaller rupture forces at similar loading rates both for native and non-native pulling than we report here. This disagreement is likely due to the inability of the coarse-grained model to capture the rearrangement of amino acid side chains observed here. As we demonstrated, force propagation calculation from network-based correlation analysis helped in investigating the dramatic effect on the mechanical stability of the Doc:Coh interaction when different pulling geometries are applied. Our methodological approach, to the best of our knowledge, has never been applied even though network analysis of SMD trajectories was performed before to probe the mechanism of allosteric regulation in imidazole glycerol phosphate synthase.[48]

In summary, for both unbinding cases where we observed high mechanostability, we found that across the binding interface, force propagated along paths with strong normal components to the pulling direction. Such a behavior was not observed for the non-native LF class, where, presumably due to N-terminal Coh unfolding, the system was no longer able to direct the force across the binding interface at high angles. From these findings, we conclude that the ultrastable complex formed by Coh and Doc achieves its remarkable mechanostability by actively directing an externally applied force toward an unfavorable angle of attack at the binding interface, consequently requiring more force to achieve a given amount of separation along the pulling direction. Our results show that this mechanically stable complex uses an architecture that exploits simple geometrical and physical concepts from Newtonian mechanics to achieve high stability against external forces. The analytical framework derived here provides a basis for developing a deeper understanding of the functioning of various mechanoactive proteins that are crucial for physiolog-

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

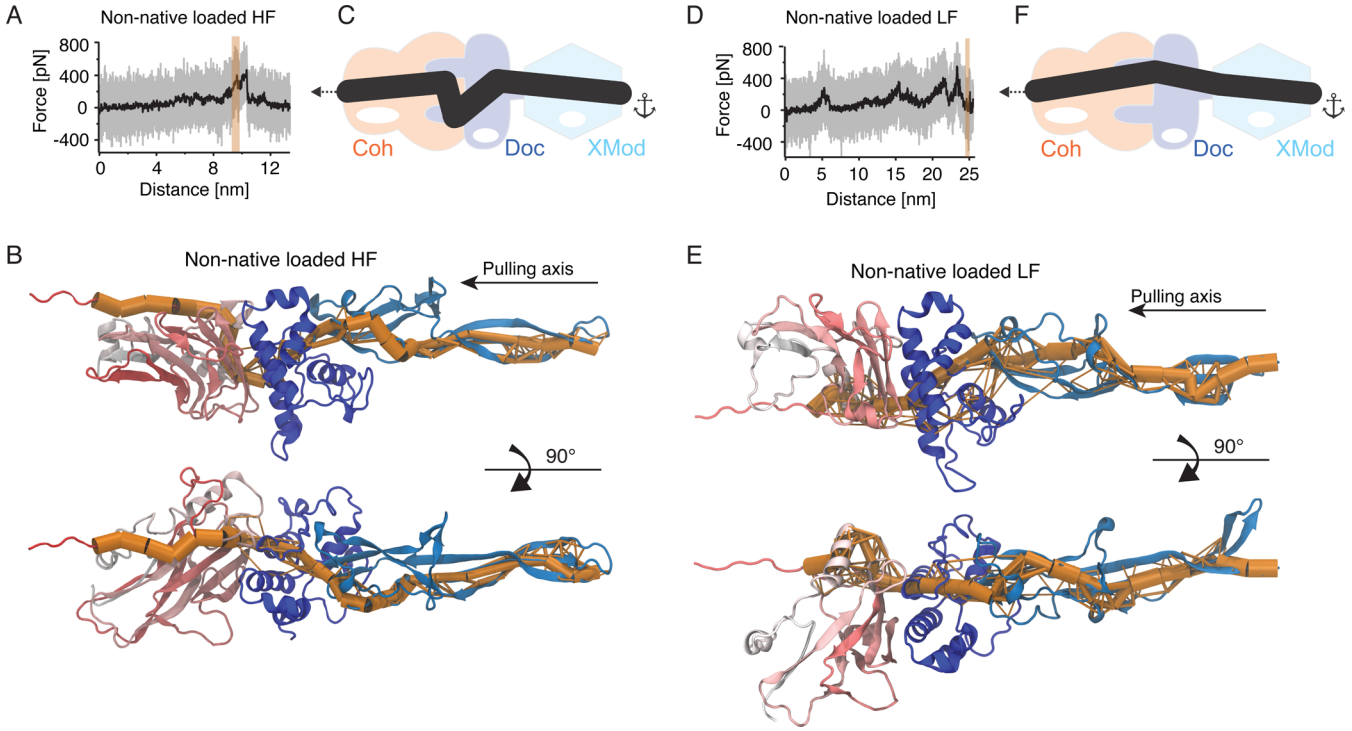


Figure 4. Force propagation through XMod-Doc:Coh in the non-native pulling configuration. (A) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500 steps. Note that this computational pulling experiment revealed a high-force behavior. (B) Network pathways calculated from dynamical network analysis for the non-native HF trajectory. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S6. (C) Schematic model of force propagation across the Coh:Doc binding interface. Force takes a path across the binding interface with large components perpendicular to the unbinding axis. (D) Unbinding trace of XMod-Doc:Coh in the non-native pulling configuration obtained from SMD at a pulling speed of 0.25 Å ns^{-1}. The full trajectory is shown in gray; the black line represents a moving average with a box size of 500. This computational pulling experiment revealed partial Coh unfolding that led to LF behavior. (E) Network pathways for the non-native LF scenario. A detailed 2D representation of the pathway, presenting amino acid identification, is shown in Supplementary Figure S7. (F) Schematic model of force propagation across the Coh:Doc binding interface. Unlike in both HF scenarios, force propagates across the binding interface mostly along the unbinding axis.

ically relevant processes such as mechanotransduction, cellular mechanosensing, and pathogenesis. Additionally, it could provide a design platform for development of artificial mechanoactive systems with applications as tissue engineering scaffolds or components in engineered nanomaterials.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.nanolett.5b02727.

Details of experimental materials and methods, steered molecular dynamics simulations, and additional theoretical discussion (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.nash@lmu.de.

Author Contributions

#These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Zaida Luthey-Schulten, Helmut Grubmüller, Marcelo Melo, and Jonathan Lai for helpful discussions. The authors gratefully acknowledge funding from an advanced grant to HEG from the European Research Council (Cellufuel Grant 294438). This work was supported by grants from the National Institutes of Health (NIH, 9P41GM104601 to K.S.) and from the Energy Biosciences Institute (UCB BP 2014OOO4J01 to K.S.). Simulations made use of the Texas Advanced Computing Center (TACC) as part of the Extreme Science and Engineering Discovery Environment (XSEDE, MCA93S028 to K.S.) and NERSC/Edison supercomputer as part of the DoE ALCC program. Resources of the National Energy Research Scientific Computing Center (NERSC) are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. A grant to E.A.B., H.E.G., and M.A.N. from GIF, the German-Israeli Foundation for Scientific Research and Development is also noted. The authors appreciate the support of the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number: 604530 and an ERA-IB Consortium (EIB.12.022), acronym FiberFuel. Additional support was obtained from grants (No. 1349) to E.A.B. from the Israel Science Foundation (ISF) and the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. is the incumbent of The Maynard I. and Elaine

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Wishner Chair of Bio-organic Chemistry. M.A.N. acknowledges funding from Society in Science—The Branco Weiss Fellowship program administered by ETH Zürich, Switzerland, and from the LMU Excellence Initiative.

REFERENCES

(1) Bao, G.; Suresh, S. *Nat. Mater.* **2003**, *2*, 715−725.
(2) Grodzinsky, A. J.; Levenston, M. E.; Jin, M.; Frank, E. H. *Annu. Rev. Biomed. Eng.* **2000**, *2*, 691−713.
(3) Kumar, S.; Weaver, V. M. *Cancer Metastasis Rev.* **2009**, *28*, 113−127.
(4) Vogel, V.; Thomas, W. E.; Craig, D. W.; Krammer, A.; Baneyx, G. *Trends Biotechnol.* **2001**, *19*, 416−423.
(5) del Rio, A.; Perez-Jimenez, R.; Liu, R.; Roca-Cusachs, P.; Fernandez, J. M.; Sheetz, M. P. *Science* **2009**, *323*, 638−641.
(6) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−1148.
(7) Engel, A.; Muller, D. J. *Nat. Struct. Biol.* **2000**, *7*, 715−718.
(8) Noy, A.; Friddle, R. W. *Methods* **2013**, *60*, 142−150.
(9) De Yoreo, J. J.; Chung, S.; Friddle, R. W. *Adv. Funct. Mater.* **2013**, *23*, 2525−2538.
(10) Rief, M.; Gautel, M.; Schemmel, A.; Gaub, H. E. *Biophys. J.* **1998**, *75*, 3008−3014.
(11) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109−1112.
(12) Li, Y. D.; Lamour, G.; Gsponer, J.; Zheng, P.; Li, H. *Biophys. J.* **2012**, *103*, 2361−2368.
(13) Lee, W.; Zeng, X.; Rotolo, K.; Yang, M.; Schofield, C. J.; Bennett, V.; Yang, W.; Marszalek, P. E. *Biophys. J.* **2012**, *102*, 1118−1126.
(14) Jagannathan, B.; Elms, P. J.; Bustamante, C.; Marqusee, S. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17820−17825.
(15) Dietz, H.; Berkemeier, F.; Bertz, M.; Rief, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 12724−12728.
(16) Carrión-Vázquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernandez, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738−743.
(17) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731−737.
(18) Popa, I.; Berkovich, R.; Alegre-Cebollada, J.; Badilla, C. L.; Rivas-Pardo, J. A.; Taniguchi, Y.; Kawakami, M.; Fernandez, J. M. *J. Am. Chem. Soc.* **2013**, *135*, 12762−12771.
(19) Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K. *Curr. Opin. Struct. Biol.* **2015**, *31*, 64−74.
(20) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872−877.
(21) Ribeiro, A. S. T.; Ortiz, V. *J. Phys. Chem. B* **2015**, *119*, 1835−1846.
(22) Van Wart, A. T.; Durrant, J.; Votapka, L.; Amaro, R. E. *J. Chem. Theory Comput.* **2014**, *10*, 511−517.
(23) Van Wart, A. T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R. E. *J. Chem. Theory Comput.* **2012**, *8*, 2949−2961.
(24) Alexander, R. W.; Eargle, J.; Luthey-Schulten, Z. *FEBS Lett.* **2010**, *584*, 376−386.
(25) Chennubhotla, C.; Bahar, I. *Mol. Syst. Biol.* **2006**, *2*, 1−13.
(26) Hytönen, V. P.; Vogel, V. *PLoS Comput. Biol.* **2008**, *4*, e24.
(27) Khalili-Araghi, F.; Gumbart, J.; Wen, P.-C.; Sotomayor, M.; Tajkhorshid, E.; Schulten, K. *Curr. Opin. Struct. Biol.* **2009**, *19*, 128−37.
(28) Li, W.; Shen, J.; Liu, G.; Tang, Y.; Hoshino, T. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 271−281.
(29) 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.; Nash, M. A. *Nat. Commun.* **2014**, *5*, 5635.
(30) Stacklies, W.; Vega, M. C.; Wilmanns, M.; Gräter, F. *PLoS Comput. Biol.* **2009**, *5*, e1000306.
(31) Xiao, S.; Stacklies, W.; Cetinkaya, M.; Markert, B.; Gräter, F. *Biophys. J.* **2009**, *96*, 3997−4005.
(32) Stacklies, W.; Seifert, C.; Graeter, F. *BMC Bioinf.* **2011**, *12*, 101.
(33) Seifert, C.; Gräter, F. *Biophys. J.* **2012**, *103*, 2195−2202.
(34) Palmai, Z.; Seifert, C.; Gräter, F.; Balog, E. *PLoS Comput. Biol.* **2014**, *10*, e1003444.
(35) Salama-Alber, O.; Jobby, M. K.; Chitayat, S.; Smith, S. P.; White, B. A.; Shimon, L. J. W.; Lamed, R.; Frolow, F.; Bayer, E. A. *J. Biol. Chem.* **2013**, *288*, 16827−16838.
(36) Puchner, E. M.; Franzen, G.; Gautel, M.; Gaub, H. E. *Biophys. J.* **2008**, *95*, 426−434.
(37) Friddle, R. W.; Noy, A.; De Yoreo, J. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 13573−13578.
(38) Dudko, O. K.; Hummer, G.; Szabo, A. *Phys. Rev. Lett.* **2006**, *96*, 108101.
(39) Greene, R. F.; Callen, H. B. *Phys. Rev.* **1951**, *83*, 1231−1235.
(40) Erman, B. *Phys. Biol.* **2011**, *8*, 056003.
(41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(42) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. *Biophys. J.* **1997**, *72*, 1568−1581.
(43) Sethi, A.; Eargle, J.; Black, A. A.; Luthey-Schulten, Z. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6620−6625.
(44) Otten, M.; Ott, W.; Jobst, M. A.; Milles, L. F.; Verdorfer, T.; Pippig, D. A.; Nash, M. A.; Gaub, H. E. *Nat. Methods* **2014**, *11*, 1127−1130.
(45) Stahl, S. W.; Nash, M. A.; Fried, D. B.; Slutzki, M.; Barak, Y.; Bayer, E. A.; Gaub, H. E. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 20431−20436.
(46) Jobst, M. A.; Schoeler, C.; Malinowska, K. H.; Nash, M. A. *J. Visualized Exp.* **2013**, *82*, e50950.
(47) Wojciechowski, M.; Thompson, D.; Cieplak, M. *J. Chem. Phys.* **2014**, *141*, 245103.
(48) Amaro, R. E.; Sethi, A.; Myers, R. S.; Davisson, V. J.; Luthey-Schulten, Z. A. *Biochemistry* **2007**, *46*, 2156−2173.
(49) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

DOI: 10.1021/acs.nanolett.5b02727
Nano Lett. 2015, 15, 7370−7376

Supporting Information

Mapping mechanical force propagation through biomolecular complexes

Constantin Schoeler[a], Rafael C. Bernardi[b], Klara H. Malinowska[a], Ellis Durner[a], Wolfgang Ott[a,c], Edward A. Bayer[d], Klaus Schulten[b,e], Michael A. Nash[*a], and Hermann E. Gaub[a]

[a]Lehrstuhl für Angewandte Physik and Center for Nanoscience, Ludwig-Maximilians-Universität, 80799 Munich, Germany
[b]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States
[c]Center for Integrated Protein Science Munich (CIPSM), University of Munich, 81377 Munich, Germany
[d]Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
[e]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

[*] michael.nash@lmu.de

1 Materials and Methods

1.1 Site Directed Mutagenesis

We performed site-directed mutagenesis of *Ruminococcus flavefaciens* strain FD1 chimeric cellulosomal proteins. A pET28a vector containing the previously cloned *R. flavefaciens* CohE from ScaE fused to cellulose-binding module 3a (CBM3a) from *C. thermocellum*, and a pET28a vector containing the previously cloned *R. flavefaciens* XMod-Doc from the CttA scaffoldin fused to the XynT6 xylanase from *Geobacillus stearothermophilus*[1] were subjected to QuikChange mutagenesis to install the mutations described in the prior paper[2]. All mutagenesis products were confirmed by DNA sequencing analysis.

1.2 Expression and Purification of Cysteine-Mutated Xyn-XMod-Doc

The Xyn(T129C)-XMod-Doc protein was expressed in *E. coli* BL21 cells in kanamycin-containing media that also contained 2 mM calcium chloride, overnight at 16°C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer containing 20 mM imidazole and 2mM calcium chloride. The bound protein was eluted using TBS buffer containing 250 mM imidazole and 2 mM calcium chloride. The solution was dialyzed with TBS to remove the imidazole, and then concentrated using an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at ~ 20°C. The concentrations of the protein stock solutions were determined to be ~ 5 mg/mL by absorption spectrophotometry.

1.3 Expression and Purification of Coh-CBM and mutated Coh-CBM C63S

The Coh-CBM (C63S) fusion protein was expressed in *E. coli* BL21(DE3) RIPL in kanamycin and chloramphenicole containing ZYM-5052 media[3] overnight at 22°C. After harvesting, cells were lysed using sonication. The lysate was then pelleted, and the supernatant fluids were applied to a Ni-NTA column and washed with TBS buffer. The bound protein was eluted using TBS buffer containing 200 mM imidazole. Imidazole was removed with a polyacrylamide gravity flow column. The protein solution was concentrated with an Amicon centrifugal filter device and stored in 50% (v/v) glycerol at -80°C. The concentrations of the protein stock solutions were determined to be ~ 5 mg/mL by absorption spectrophotometry.

1.4 Sample Preparation

Cantilevers and cover glasses were functionalized according to previously published protocols[4]. Briefly, cantilevers and cover glasses were cleaned by UV-ozone treatment and piranha solution, respectively. Levers and glasses were silanized using (3-aminopropyl)-dimethyl-ethoxysilane (APDMES) to introduce surface amine groups. Amine groups on the cantilevers and cover glasses were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker in sodium borate buffer. Disulfide-linked dimers of the Xyl-XMod-Doc proteins were reduced for 2 hours at room temperature using a TCEP disulfide reducing bead slurry. The protein/bead mixture was rinsed with TBS measurement buffer, centrifuged at 850 rcf for 3 minutes, and the supernatant was collected with a micropipette. Reduced proteins were diluted with measurement buffer (1:3 (v/v) for cantilevers, and 1:1 (v/v) for cover glasses), and applied to PEGylated cantilevers and cover glasses for 1 h. Both cantilevers and cover glasses were then rinsed with TBS to remove

unbound proteins, and stored under TBS prior to force spectroscopy measurements. Site specific immobilization of the Coh-CBM-ybbR fusion proteins to PEGylated cantilevers or coverglasses was carried out according to previously published protocols[5]. Briefly, PEGylated cantilevers or coverglasses were incubated with Coenzyme A (CoA) (20 mM) stored in coupling buffer for 1h at room temperature. Levers or surfaces were then rinsed with TBS to remove unbound CoA. Coh-CBM-ybbR fusion proteins were then covalently linked to the CoA surfaces or levers by incubating with Sfp phosphopantetheinyl transferase for 2 hours at room 37°. Finally, surfaces or levers were subjected to a final rinse with TBS and stored under TBS prior to measurement.

1.5 Single Molecule Force Spectroscopy Measurements

SMFS measurements were performed on a custom built AFM controlled by an MFP-3D controller from Asylum Research running custom written Igor Pro (Wavemetrics) software. Cantilever spring constants were calibrated using the thermal noise / equipartition method. The cantilever was brought into contact with the surface and withdrawn at constant speed ranging from 0.2–6.4 μm/s. An x-y stage was actuated after each force-extension trace to expose the molecules on the cantilever to a new molecule at a different surface location with each trace. Typically 20,000–50,000 force-extension curves were obtained with a single cantilever in an experimental run of 18-24 hours. A low molecular density on the surface was used to avoid formation of multiple bonds. While the raw datasets contained a majority of unusable curves due to lack of interactions or nonspecific adhesion of molecules to the cantilever tip, select curves showed single molecule interactions with CBM and Xyn unfolding length increments. We sorted the data using a combination of automated data processing and manual classification by searching for contour length increments that matched the lengths of our specific protein fingerprint domains: the xylanase (∼89 nm) and the CBM (∼56 nm). After identifying these specific traces, we measured the loading rate dependency of the final Doc:Coh ruptures based on bond history.

1.6 Data Analysis

Data were analyzed using slight modifications to previously published protocols[4;6;7]. Force extension traces were transformed into contour length space using the QM-FRC model with bonds of length $b = 0.11\,\mathrm{nm}$ connected by a fixed angle $\gamma = 41^\circ$ and and assembled into barrier position histograms using cross-correlation. For the loading rate analysis, the loading rate at the point of rupture was extracted by applying a line fit to the force vs. time trace in the immediate vicinity prior to the rupture peak. The loading rate was determined from the slope of the fit. The most probable rupture forces and loading rates were determined by applying probability density fits to histograms of rupture forces and loading rates at each pulling speed.

1.7 Molecular Dynamics Simulations

Connecting dynamics to structural data from diverse experimental sources, molecular dynamics simulations allow one to explore off-equilibrium properties of protein structure complexes in unparalleled detail[8]. More specifically, molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes[9]. The structure of the XMod-Doc:Coh complex had been solved by means of X-ray crystallography at 1.97Å resolution and is available at the protein data bank (PDB:4IU3). The system was then solvated and the net charge of the protein and the calcium ions was neutralized using sodium atoms as counter-ions,

which were randomly arranged in the solvent. Total system size was approximately 580k atoms. The MD simulations in the present study were performed employing the molecular dynamics package NAMD[10;11]. The CHARMM36 force field[12;13] along with the TIP3 water model[14] was used to describe all systems. The simulations were carried out assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for pressure, kept at 1 bar, and temperature coupling. A distance cut-off of 11.0 Å was applied to short-range, non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME)[15] method. The equations of motion were integrated using the r-RESPA multiple time step scheme[11] to update the van der Waals interactions every two steps and electrostatic interactions every four steps. The time step of integration was chosen to be 2 fs for all simulations performed. The first two nanoseconds of the simulations served to equilibrate systems before the production runs, which varied from 200 ns to 1.3 μs in the different simulations. To characterize the coupling between dockerin and cohesin, we performed SMD simulations[16] of constant velocity stretching (SMD-CV protocol) with pulling speed of 0.25 Å/ns. In all simulations, SMD was employed by restraining the position of one end of the XMod-Doc domain harmonically, and moving a second restraint point, at the end of the Coh domain, with constant velocity in the desired direction. The procedure is equivalent to attaching one end of a harmonic spring to the end of a domain and pulling on the other end of the spring. The force applied to the harmonic pulling spring is then monitored during the time of the molecular dynamics simulation. All analyses of MD trajectories were carried out employing VMD[17] and its plugins. Surface contact areas of interacting residues were calculated employing Volarea[18] implemented in VMD. The area is calculated using a probe radius defined as an *in silico* rolling sphere that is scanned around the area of the dockerin exposed to the cohesin and also the cohesin area exposed to the dockerin. The Network View plugin[19] on VMD[17] was employed to perform dynamical network analysis. A network was defined as a set of nodes, all α-carbons, with connecting edges. Edges connect pairs of nodes if corresponding monomers are in contact, and 2 nonconsecutive monomers are said to be in contact if they fulfill a proximity criterion, namely any heavy atoms (nonhydrogen) from the 2 monomers are within 4.5 Å of each other for at least 75% of the frames analyzed. As suggested by Sethi et al.[20], nearest neighbors in sequence are not considered to be in contact as they lead to a number of trivial suboptimal paths. The dynamical networks were constructed from 20 ns windows of the total trajectories sampled every 400 ps. The probability of information transfer across an edge is set as $w_{ij} = -log\left(\mid C_{ij} \mid\right)$, where C_{ij} is the correlation matrix calculated with Carma[21]. Using the Floyd-Warshall algorithm, the suboptimal paths were then calculated. The tolerance value used for any path to be included in the suboptimal path was $-log\left(0.5\right) = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where x, y, z components of each atom were considered independently.

2 Protein Sequences

Sequences of protein constructs used in this work are listed here. Domains as well as engineered tags and residues are color-coded.

2.1 HIS-Xyn(T128C)-XDoc

X-module
Dockerin type III

Xylanase
Linker or extra residues

M S H H H H H H K N A D S Y A K K P H I S A L N A P Q L D Q R Y K N E F T I G A
A V E P Y Q L Q N E K D V Q M L K R H F N S I V A E N V M K P I S I Q P E E G K
F N F E Q A D R I V K F A K A N G M D I R F H T L V W H S Q V P Q W F F L D K E
G K P M V N E C D P V K R E Q N K Q L L L K R L E T H I K T I V E R Y K D D I K
Y W D V V N E V V G D D G K L R N S P W Y Q I A G I D Y I K V A F Q A A R K Y G
G D N I K L Y M N D Y N T E V E P K R T A L Y N L V K Q L K E E G V P I D G I G
H Q S H I Q I G W P S E A E I E K T I N M F A A L G L D N Q I T E L D V S M Y G
W P P R A Y P T Y D A I P K Q K F L D Q A A R Y D R L F K L Y E K L S D K I S N
V T F W G I A D N H T W L D S R A D V Y Y D A N G N V V V D P N A P Y A K V E K
G K G K D A P F V F G P D Y K V K P A Y W A I I D H K V V P N T V T S A V K T Q
Y V E I E S V D G F Y F N T E D K F D T A Q I K K A V L H T V Y N E G Y T G D D
G V A V V L R E Y E S E P V D I T A E L T F G D A T P A N T Y K A V E N K F D Y
E I P V Y Y N N A T L K D A E G N D A T V T V Y I G L K G D T D L N N I V D G R
D A T A T L T Y Y A A T S T D G K D A T T V A L S P S T L V G G N P E S V Y D D
F S A F L S D V K V D A G K E L T R F A K K A E R L I D G R D A S S I L T F Y T
K S S V D Q Y K D M A A N E P N K L W D I V T G D A E E E

2.2 Coh-CBM(C2A,C63S)-HIS-ybbR

CohIII
CBM (C2A, C63S)
ybbR-Tag
Linker or extra residues

M G T A L T D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D
A A G Q T V T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G
A Y A K K G A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M
W T V E L K V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A
Q G K L M Q A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I
K A G E P G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A N
T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I D L S
K L T L R Y Y Y T V D G Q K D Q T F W S D H A A I I G S N G S Y N G I T S N V K
G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R F A K
N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G K E P
G E L K L P R S R H H H H H H G S L E V L F Q G P D S L E F I A S K L A

2.3 CBM(T2C)-Coh-HIS

CBM (T2C)
CohIII
Linker or extra residues

M C N T P V S G N L K V E F Y N S N P S D T T N S I N P Q F K V T N T G S S A I
D L S K L T L R Y Y Y T V D G Q K D Q T F W C D H A A I I G S N G S Y N G I T S

N V K G T F V K M S S S T N N A D T Y L E I S F T G G T L E P G A H V Q I Q G R
F A K N D W S N Y T Q S N D Y S F K S A S Q F V E W D Q V T A Y L N G V L V W G
K E P G G S V V P S T Q P V T T P P A T T K P P A T T I P P S D D P N A M A L T
D R G M T Y D L D P K D G S S A A T K P V L E V T K K V F D T A A D A A G Q T V
T V E F K V S G A E G K Y A T T G Y H I Y W D E R L E V V A T K T G A Y A K K G
A A L E D S S L A K A E N N G N G V F V A S G A D D D F G A D G V M W T V E L K
V P A D A K A G D V Y P I D V A Y Q W D P S K G D L F T D N K D S A Q G K L M Q
A Y F F T Q G I K S S S N P S T D E Y L V K A N A T Y A D G Y I A I K A G E P L
E H H H H H H

3 Supplementary Discussion

The Pearson correlation matrices of the Xmod-Doc:Coh complex before and after applying force in the native pulling configuration are presented in Supplementary Figure S3 and S4, respectively. For the unloaded complex, movements within Doc domain are seen to be highly correlated, while XMod is seen to be divided into two anti-correlated sub-domains, one comprising the β-sheet fragment close to the N-terminus (residues 5-15 and 45-66) and the other constituting the rest of the domain. Intra-domain correlations of Coh exhibit more a complex pattern to which both secondary (anti-parallel β-strands and β-sheet at the binding interface) and tertiary structure (vicinity of C- and N-termini) contribute. Some of the inter-domain correlations in the complex originate from spatial vicinity and direct interactions, specifically at the Doc:Coh binding interface and at XMod contacts with Doc inserts. However, coupling between distant parts of the complex is also present. For example, fluctuations of the non-binding part of Coh are correlated with the N-terminal part of XMod and strongly anti-correlated with Doc domain.

4 Supplementary Notes

4.1 Constant Barrier Distance Model

The constant barrier distance model[16], also referred to as the Bell-Evans model[22], is commonly used to estimate the distance to the transition state Δx and the natural off-rate k_0 of mechanically induced receptor ligand dissociation from single-molecule force spectroscopy experiments. It predicts that the most probable rupture force $\langle F \rangle$ is linearly dependent on the logarithm of the force loading rate[16]:

$$\langle F(r) \rangle = \frac{k_B T}{\Delta x} \ln \frac{\Delta x \cdot r}{k_0 k_B T} \tag{S1}$$

where k_B is Boltzmann's constant, T is the temperature and r is the loading rate at the point of rupture.

The probability density distribution of rupture forces at given loading rate r in this model is given as[16]:

$$p\left(F\right) = \frac{k_0}{r} \exp\left[\frac{\Delta x}{k_B T} F - \frac{k_0 \cdot k_B T}{\Delta x \cdot r}\left(e^{\frac{\Delta x}{k_B T} F} - 1\right)\right] \tag{S2}$$

4.2 Dudko-Hummer-Szabo Model

The Dudko-Hummer-Szabo (DHS)[23;24] model describes a non-linear dependence for the most probable rupture force on loading rate:

$$\langle F(r)\rangle = \frac{\Delta G}{\nu \Delta x}\left\{1 - \left[\frac{k_B T}{\Delta G}\ln\left(\frac{k_B T k_0}{\Delta x r}e^{\frac{\Delta G}{k_B T}+\gamma}\right)\right]^{\nu}\right\} \tag{S3}$$

where ΔG is the free energy of activation and $\gamma = 0.577$ is the Euler-Mascheroni constant. The model parameter ν defines the single-well free-energy surface model used ($\nu = \frac{2}{3}$ for linear-cubic and $\frac{1}{2}$ for cusp free-energy. For $\nu = 1$ and $\Delta G \to \infty$ independent of ν the Eqs. (S1) and (S2) are recovered.

4.3 Pearson Correlation and covariance matrix

4.3.1 Validation

An $N \times N$ matrix of Pearson correlation coefficients C_{ij} (Supporting Eq. S4) was calculated from each atom's x, y, z position throughout the simulation trajectory, which inherently ignores off-diagonal elements of the atomic 3×3 submatrices D_{ij}^{mn} from the full normalized $3N \times 3N$ covariance matrix (*i.e.*, correlations along orthogonal axes are neglected, see Supporting Eqs. (S5) and (S6)) and Supporting Fig S8.

Although this quasi-harmonic approximation is commonly employed in correlation analysis[19;25–29], it is not *a priori* justified for complicated biomolecular interactions[30]. To validate the use of Pearson correlations, we therefore first analyzed independently the contributions from diagonal and off-diagonal elements of each 3 x 3 covariance submatrix for each pair of α-carbons within the structure (Fig. S9A and B). Both with and without applied force, the off-diagonal elements roughly follow Gaussian distributions centered around a correlation value of 0. Interestingly, as force was applied, the standard deviation of the distribution of off-diagonal correlation values decreased from $\sigma_{unloaded} = 0.45$ to $\sigma_{loaded} = 0.29$. This indicated a lesser influence of off-diagonal elements on the highly (anti-)correlated motion within the system under force (see Supporting Discussion 3). The diagonal elements of the sub-matrices that are used for calculating the Pearson correlation values showed a dramatically different behavior. Both in the unloaded and loaded state, the resulting distributions were strongly shifted towards highly correlated motion, and the shape of the distribution remained mostly unchanged after application of force. Since our analysis relies on the identification of paths of highest correlation through proximate residues, the quasi-harmonic approximation implied by the use of Pearson correlation is justified, especially for suboptimal pathway analysis. The resulting distributions of on- and off-diagonal matrix elements of each covariance submatrix for the loaded configuration HF class (Fig. S10A) and LF class (Fig. S10B) exhibited the same characteristics as previously described for the native configuration, with off-diagonal elements showing symmetric correlations around zero and diagonal elements showing highly correlated motions.

4.3.2 Supplementary Equations

The Pearson correlation coefficient C_{ij} used in our dynamical network analysis protocol is given by:

$$C_{ij} = \frac{\langle \Delta \mathbf{r}_i(t) \cdot \Delta \mathbf{r}_j(t)\rangle}{\left(\left\langle \Delta \mathbf{r}_i(t)^2\right\rangle \left\langle \Delta \mathbf{r}_j(t)^2\right\rangle\right)^{\frac{1}{2}}} \tag{S4}$$

where $\Delta\mathbf{r}_i(t) = \mathbf{r}_i(t) - \langle\mathbf{r}_i(t)\rangle$.

The full $3N \times 3N$ covariance matrix M_{ij} for atoms i and j consists of 3×3 submatrices of the form:

$$\left\langle \Delta\mathbf{r}_i(t)\,\Delta\mathbf{r}_j(t)^T \right\rangle = M_{ij} = \begin{pmatrix} M_{ij}^{xx} & M_{ij}^{xy} & M_{ij}^{xz} \\ M_{ij}^{yx} & M_{ij}^{yy} & M_{ij}^{yz} \\ M_{ij}^{zx} & M_{ij}^{zy} & M_{ij}^{zz} \end{pmatrix} \tag{S5}$$

The full normalized correlation matrix is calculated from M_{ij}:

$$D_{ij}^{mn} = \frac{M_{ij}^{mn}}{\sqrt{M_{ij}^{mm} M_{ij}^{nn}}} \tag{S6}$$

Consequently, the Pearson correlation coefficient is calculated as the trace of the normalized 3×3 submatrices ($C_{ij} = \mathrm{Tr}\, D_{ij}$).

4.3.3 Derivation of Main Text Equation 2

Eq. 1 from the main text reads:

$$\left\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j^T \right\rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \tag{S7}$$

Combining Eqs. (S7) and (S4) yields:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\left\langle \Delta\mathbf{r}_i^2(t) \right\rangle \left\langle \Delta\mathbf{r}_j^2(t) \right\rangle \right)^{-\frac{1}{2}} \tag{S8}$$

For an arbitrary potential $U_i(\mathbf{r})$ of atom i, a Taylor expansion around the potential minimum (set to be at 0) yields:

$$U_i(\mathbf{r}) = 0 + \underbrace{\mathbf{r}_i^T \nabla U(0)}_{=0} + \frac{1}{2}\mathbf{r}_i^T \underline{H}(0)\,\mathbf{r}_i + ... \tag{S9}$$

where $\underline{H}(0)$ is the Hessian matrix evaluated at the potential minimum. Assuming Schwarz' theorem holds for $U_i(\mathbf{r})$, $\underline{H}(0)$ is a symmetric matrix and therefore has real eigenvalues and orthonormal eigenvectors. Hence, a change to the eigenbasis of $\underline{H}(0)$ is a rotation of the coordinate system. In this new basis the Hessian is diagonal:

$$\underline{H}(0) \rightarrow \underline{H}'(0) = \begin{pmatrix} k_{x'} & 0 & 0 \\ 0 & k_{y'} & 0 \\ 0 & 0 & k_{z'} \end{pmatrix} \tag{S10}$$

This yields a simple expression for the second order term in Eq. (S9):

$$U_i(\mathbf{r}') = \frac{1}{2}\mathbf{r}'^T \underline{H}'(0)\,\mathbf{r}' = \frac{1}{2}\left(k_{x'} x'^2 + k_{y'} y'^2 + k_{z'} z'^2 \right) \tag{S11}$$

Now we inspect the normalization of C_{ij}:

$$\langle \Delta\mathbf{r}_i^2(t) \rangle = \langle \mathbf{r}_i^2(t) - 2\mathbf{r}_i(t)\langle\mathbf{r}_i(t)\rangle + \langle\mathbf{r}_i(t)\rangle^2 \rangle \tag{S12}$$

In the harmonic approximation of the potential of atom i, $\langle \mathbf{r}_i(t) \rangle = 0$, and therefore $\langle \Delta \mathbf{r}_i(t)^2 \rangle = \langle \mathbf{r}_i^2(t) \rangle$. In the basis of $\underline{H'}(0)$ this becomes:

$$\langle \mathbf{r'}_i^2(t) \rangle = \langle x'_i(t)^2 + y'_i(t)^2 + z'_i(t)^2 \rangle = \langle x'_i(t)^2 \rangle + \langle y'_i(t)^2 \rangle + \langle z'_i(t)^2 \rangle \tag{S13}$$

Applying the equipartition theorem to this result yields:

$$\langle x'_i(t)^2 \rangle = \frac{k_B T}{k'_{x_i}} \tag{S14}$$

And therefore:

$$\langle \Delta \mathbf{r}'_i(t)^2 \rangle = k_B T \left(\frac{1}{k'_{x_i}} + \frac{1}{k'_{y_i}} + \frac{1}{k'_{z_i}} \right) = \frac{k_B T}{k'_{i,eff}} \tag{S15}$$

Plugging this result into Eq. (S8), one finds:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\langle \Delta \mathbf{r}_j(t)^2 \rangle \right)^{-\frac{1}{2}} \tag{S16}$$

Repeating the above steps for atom j yields the final result:

$$C_{ij} = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \left(\frac{k_B T}{k'_{i,eff}} \right)^{-\frac{1}{2}} \left(\frac{k_B T}{k'_{j,eff}} \right)^{-\frac{1}{2}} \tag{S17}$$

$$= \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \cdot \sqrt{k'_{i,eff} \cdot k'_{j,eff}} \tag{S18}$$

5 Supplementary Figures

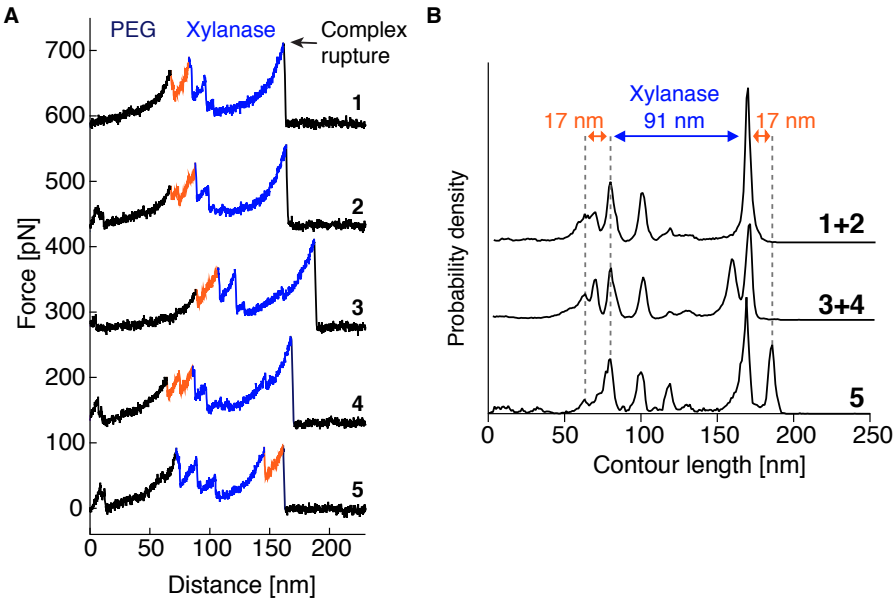


Fig. S1: SMFS of the non-native low force curve class. **A** Typical unfolding fingerprints. All traces showed a characteristic Xyn fingerprint (blue). A $17-19$ nm increment corresponding to partial N-terminal Coh unfolding (orange) occurs either prior to Xyn unfolding (traces 1-4), or just before complex rupture (trace 5). It was observed as a single event (traces 1,3 and 5) or showed substructure (traces 2 and 4). **B** Traces were grouped and assembled into contour length histograms. One or more of the unassigned increments combined into a $17-19$ nm increment.

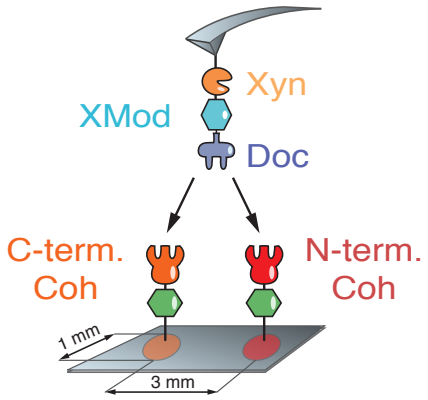


Fig. S2: Comparing the native geometry with the non-native high force class. Two exclude uncertainties in cantilever calibration when comparing the native geometry with the non-native HF class, we immobilized both Coh-CBM (native) and CBM-Coh (non-native) on two spatially separated spots on a single cover glass. These spots where then alternately probed with the same Xyn-XMod-Doc functionalized cantilever.

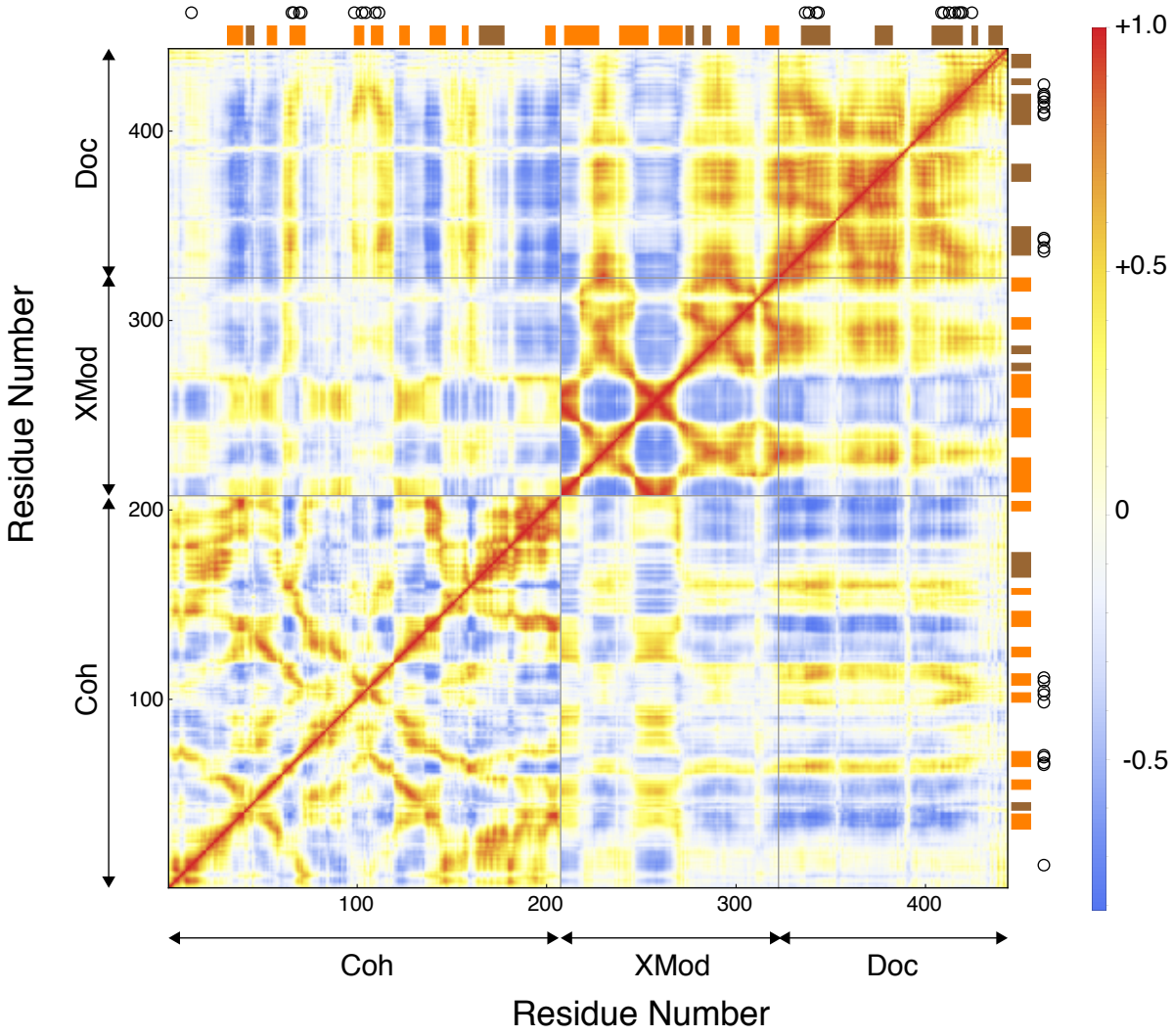


Fig. S3: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the unloaded Xmod-Doc:Coh complex. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from the Coh and Doc binding interface.

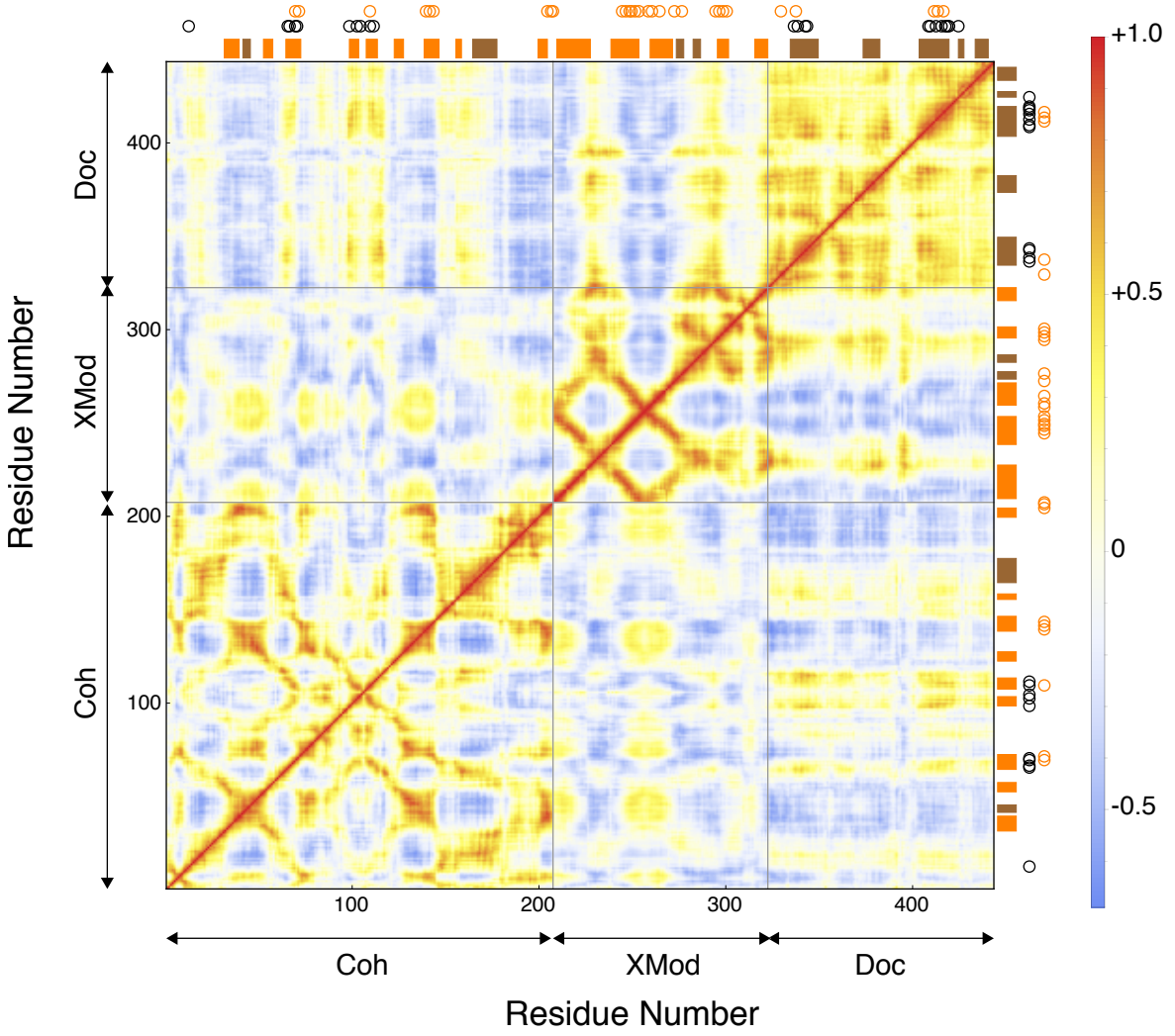


Fig. S4: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the Xmod-Doc:Coh complex loaded with force in the native pulling geometry. α-helices and β-strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from Coh and Doc binding interfaces and orange circles represent residues on the force propagation path.

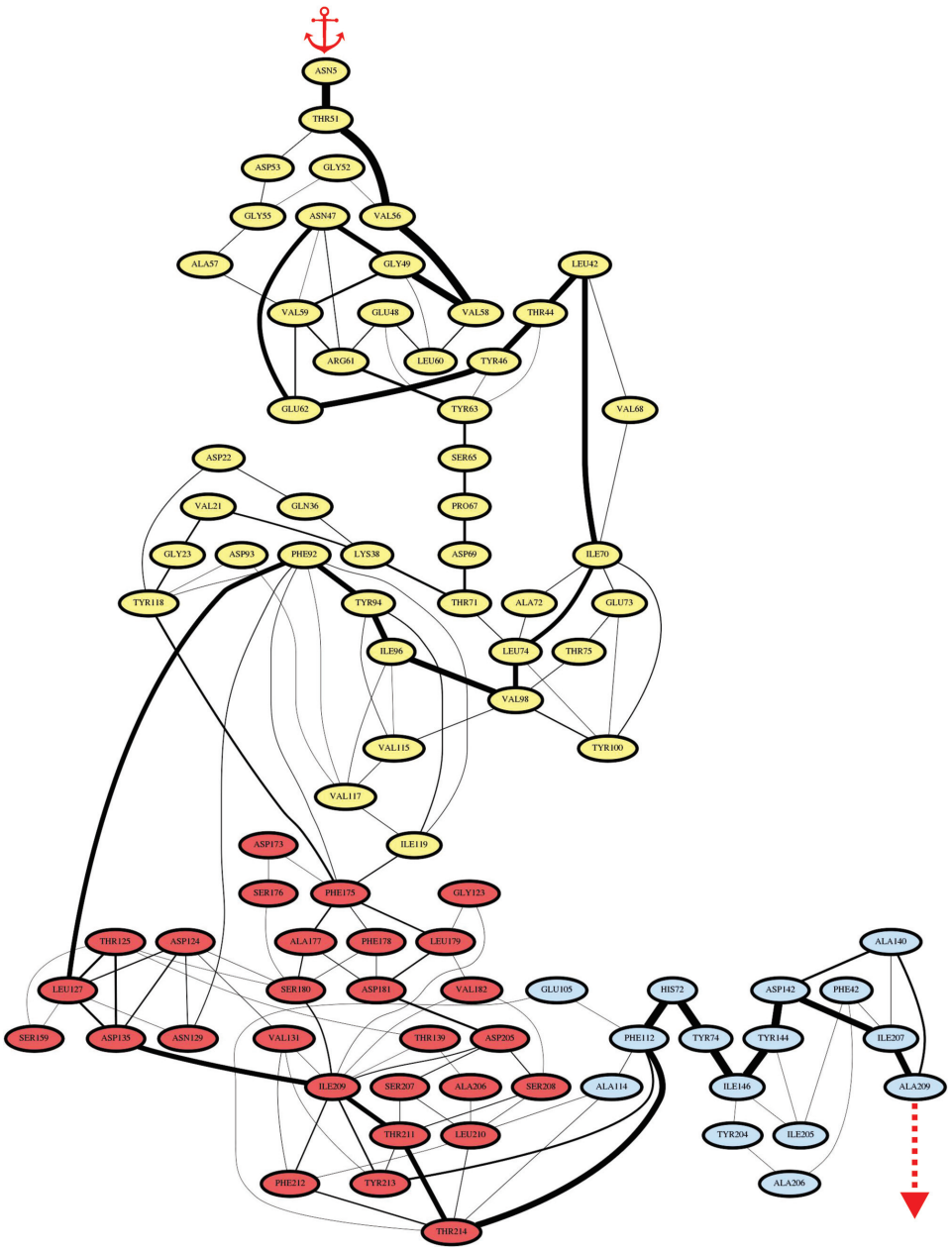


Fig. S5: Force propagation pathway through the loaded XMod-Doc:Coh complex in the native pulling geometry (N-terminal pulling of Xmod-Doc, C-terminal pulling of Coh) obtained from dynamical network analysis. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

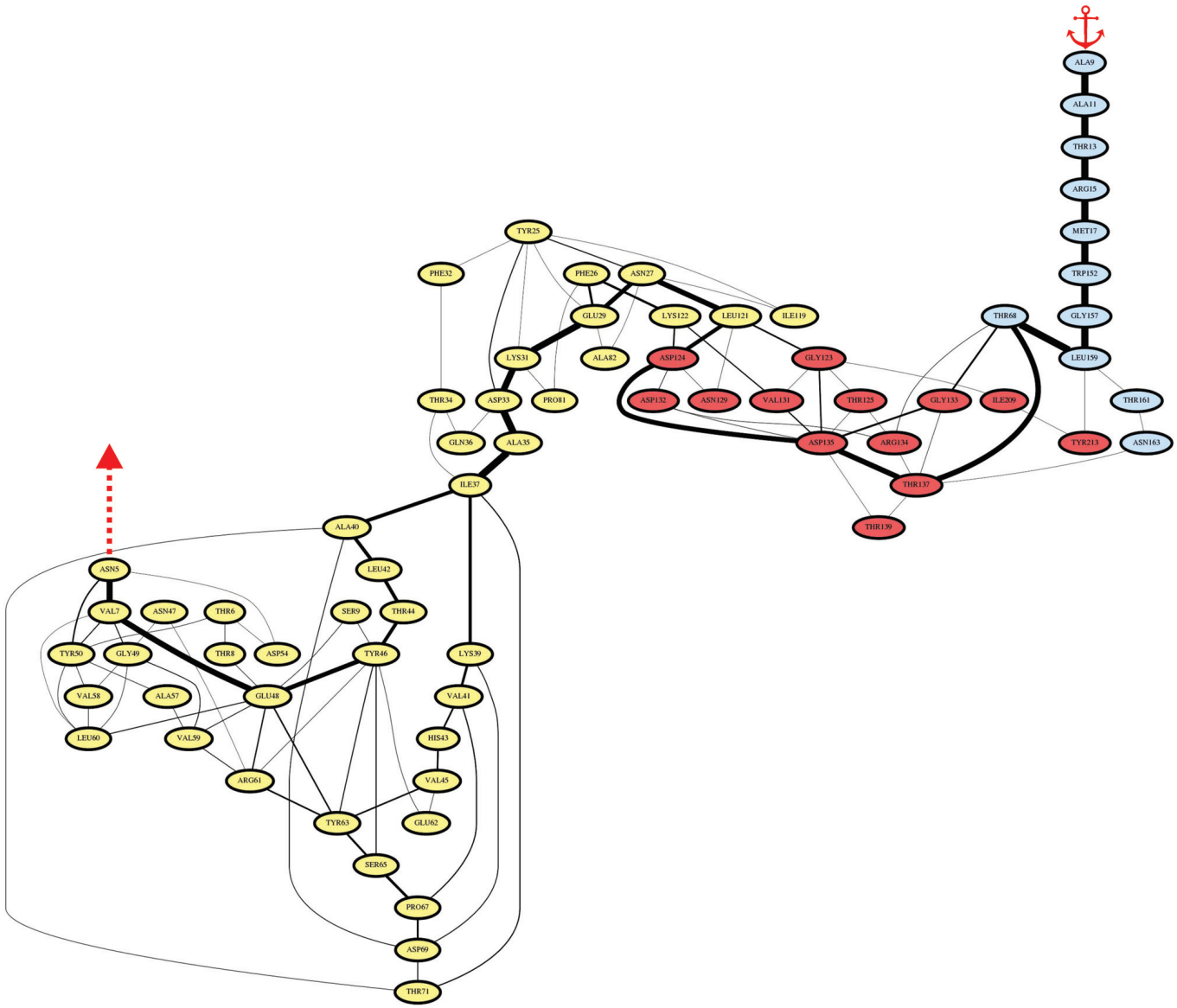


Fig. S6: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing high-force unbinding characteristics and no C-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

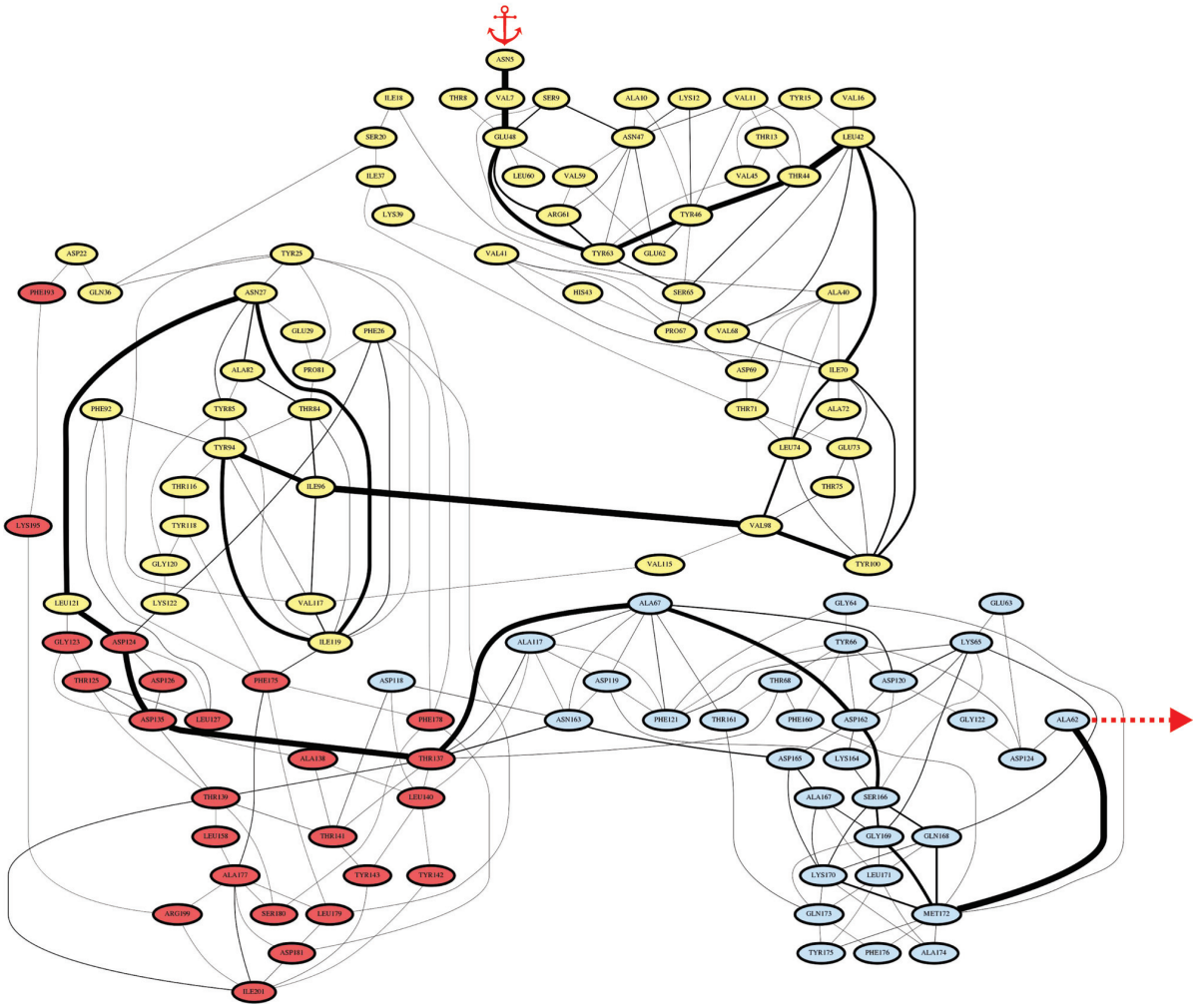


Fig. S7: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing low-force unbinding characteristics and partial N-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting lines between residues represent edges identified in our Network Analysis protocol and constitute the suboptimal paths between the pulling points. Edge thickness represents the number of suboptimal paths going through the edge.

Fig. S8: Full unnormalized covariance Matrix M_{ij} for a five atom system from which the full normalized covariance matrix is calculated according to Eq. (S6). On- and off-diagonal elements from one of the atomic submatrices are highlighted in yellow and blue, respectively.

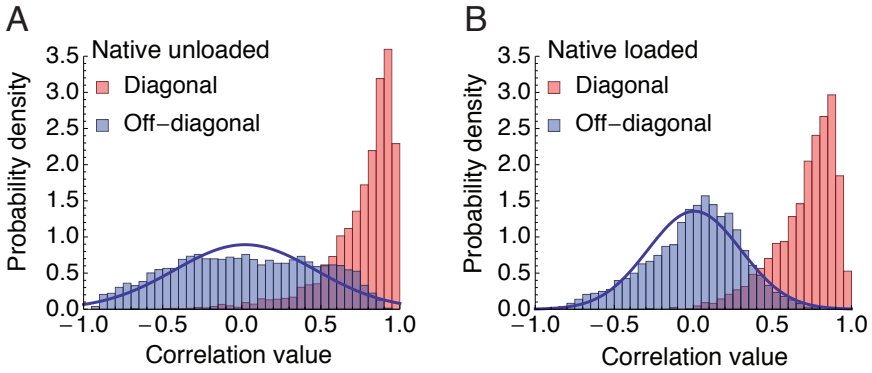


Fig. S9: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A**, the native unloaded, and **B** the native loaded, scenario.

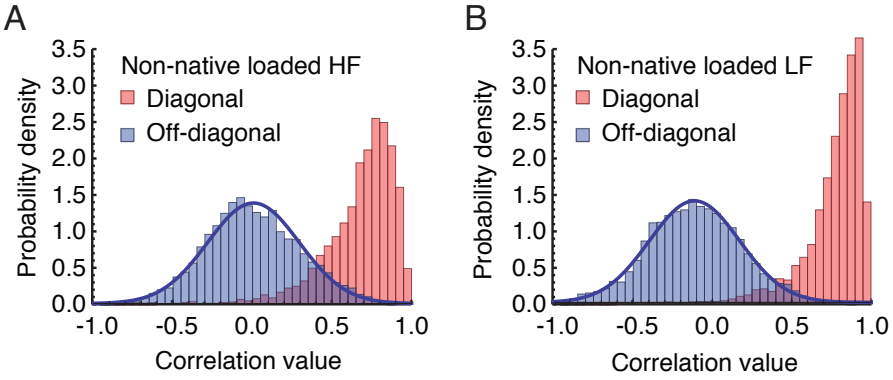


Fig. S10: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A,** the non-native HF, and **B** the non-native LF, scenario.

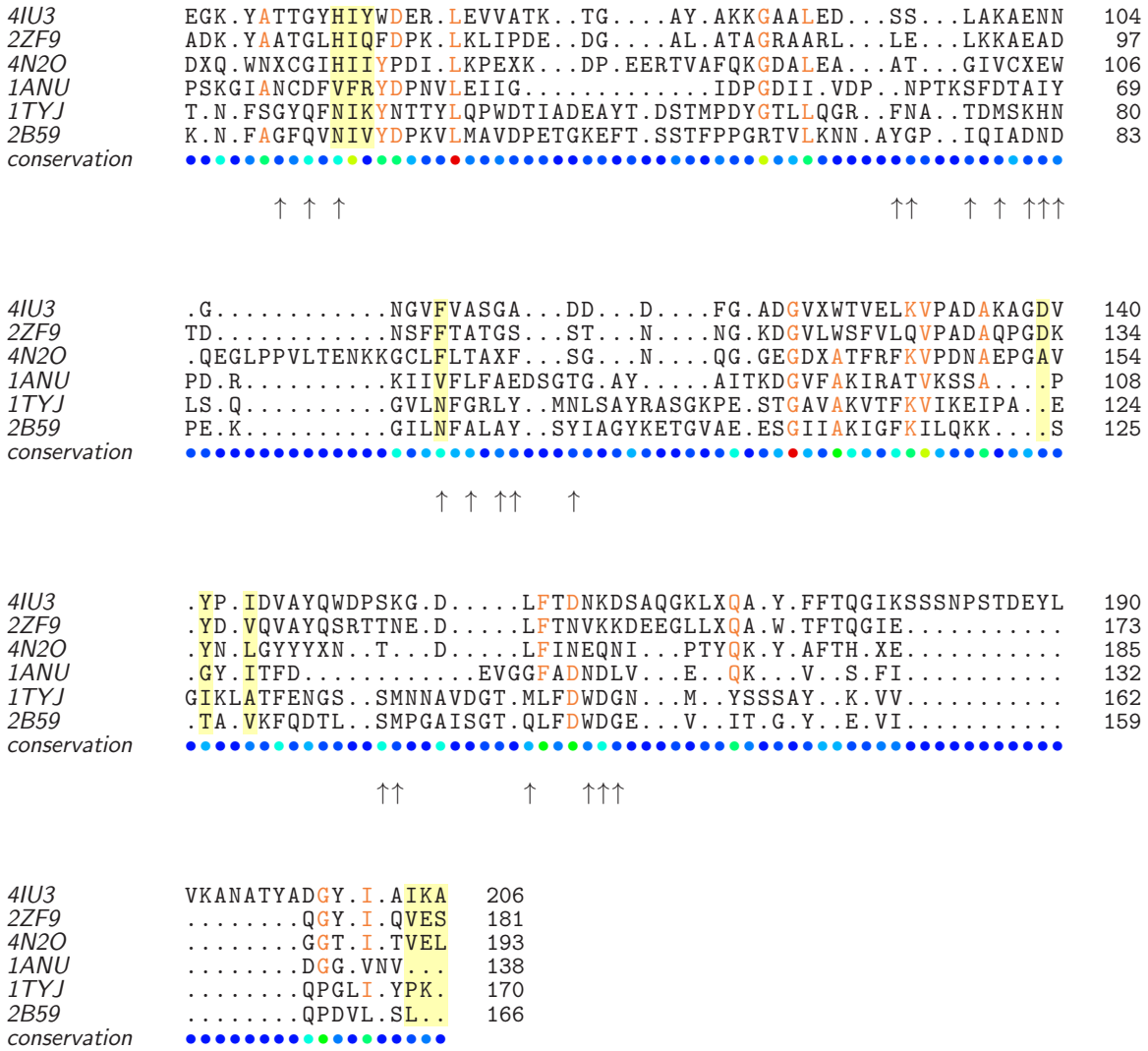


Fig. S11: Structure-aligned sequences of six crystallized cohesins. Residues on the force propagation path are highlighted in yellow. Arrows indicate binding residues. Residue conservation is color-coded from blue - lack of conservation, to red - residue fully conserved. Crystal structures used: 4IU3 ScaE *Rf* FD-1, 2ZF9 ScaE *Rf* strain 17, 4N2O CohG *Rf* FD-1, 1ANU CohC2 CipC *Ct*, 1TYJ CohA11 ScaA *Bc*, 2B59 SdbA *Ct*.

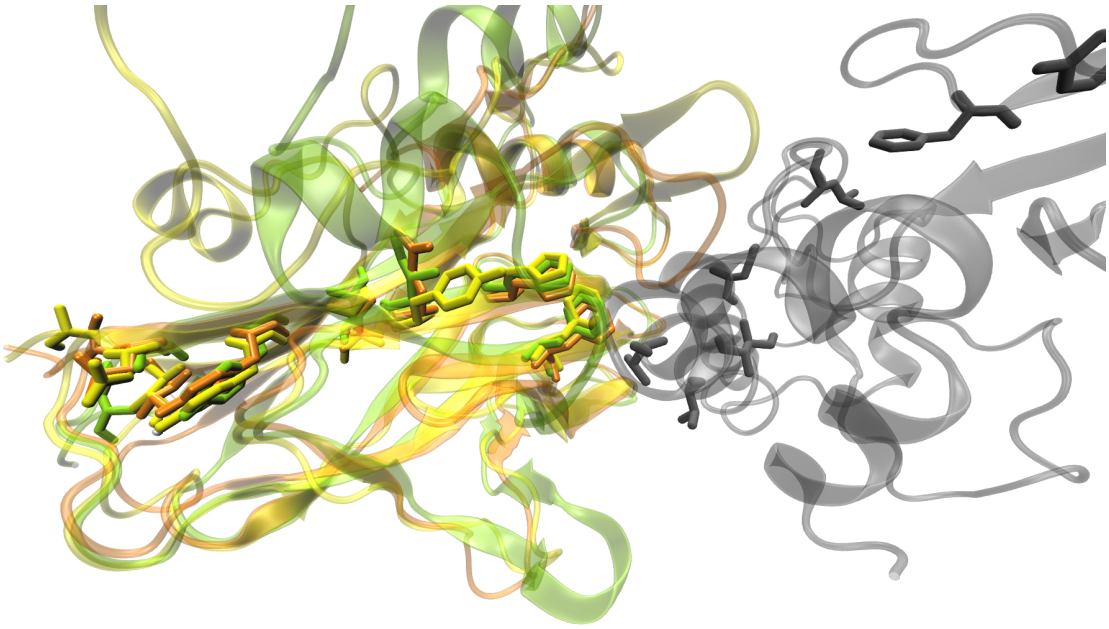


Fig. S12: Structure and sequence conservation of the force propagation pathway residues in Coh. CohE from the ScaE cell anchoring protein, *Rf* FD-1 used in this work (PDB 4IU3) is highlighted in green. Highly homologous structures of CohE from *Rf* strain 17 (PDB 2ZF9) and Coh G from *Rf* FD-1 (PDB 4A2O) are colored in orange and yellow, respectively. Residues lying in the force propagation path are shown as sticks. XDoc from the CttA *Rf* FD-1 scaffold used in this work is shown in gray.

References

[1] Orly Salama-Alber, Maroor K Jobby, Seth Chitayat, Steven P Smith, Bryan A White, Linda J W Shimon, Raphael Lamed, Felix Frolow, and Edward A Bayer. Atypical cohesin-dockerin complex responsible for cell-surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J. Biol. Chem.*, 288(23):16827–16838, April 2013.

[2] Constantin Schoeler, Klara H Malinowska, Rafael C Bernardi, Lukas F Milles, Markus A Jobst, Ellis Durner, Wolfgang Ott, Daniel B Fried, Edward A Bayer, Klaus Schulten, Hermann E Gaub, and Michael A Nash. Ultrastable cellulosome-adhesion complex tightens under load. *Nat. Commun.*, 5:1–8, December 2014.

[3] F William Studier. Protein production by auto-induction in high-density shaking cultures. *Protein Expres. Purif.*, 41(1):207–234, May 2005.

[4] Markus A Jobst, Constantin Schoeler, and Michael A Nash. Investigating receptor-ligand systems of the cellulosome with AFM-based single-molecule force spectroscopy. *J. Vis. Exp.*, 82(82):e50950, 2013.

[5] Jun Yin, Alison J Lin, David E Golan, and Christopher T Walsh. Site-specific protein labeling by Sfp phosphopantetheinyl transferase. *Nat. Protoc.*, 1(1):280–285, June 2006.

[6] Stefan W Stahl, Michael A Nash, Daniel B Fried, Michal Slutzki, Yoav Barak, Edward A Bayer, and Hermann E Gaub. Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci. U.S.A.*, 109(50):20431–20436, December 2012.

[7] Elias M Puchner, Gereon Franzen, Mathias Gautel, and Hermann E Gaub. Comparing proteins by their unfolding pattern. *Biophys. J.*, 95(1):426–434, July 2008.

[8] Juan R Perilla, Boon Chong Goh, C Keith Cassidy, Bo Liu, Rafael C Bernardi, Till Rudack, Hang Yu, Zhe Wu, and Klaus Schulten. Molecular dynamics simulations of large macromolecular complexes. *Curr. Opin. Struct. Biol.*, 31:64–74, 2015.

[9] R C Bernardi, M C R Melo, and K Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta*, 1850(5):872–877, 2015.

[10] Laxmikant Kalé, Robert Skeel, Milind Bhandarkar, Robert Brunner, Attila Gursoy, Neal Krawetz, James Phillips, Aritomo Shinozaki, Krishnan Varadarajan, and Klaus Schulten. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.*, 151(1): 283–312, 1999.

[11] James C Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D Skeel, Laxmikant Kalé, and Klaus Schulten. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26(16):1781–1802, 2005.

[12] Robert B Best, Xiao Zhu, Jihyun Shim, Pedro E M Lopes, Jeetain Mittal, Michael Feig, and Alexander D MacKerell, Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1and χ 2Dihedral Angles. *J. Chem. Theory Comput.*, 8(9):3257–3273, 2012.

[13] A D MacKerell, D Bashford, Bellott, R L Dunbrack, J D Evanseck, M J Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, F T K Lau, C Mattos, S Michnick, T Ngo, D T Nguyen, B Prodhom, W E Reiher, B Roux, M Schlenkrich, J C Smith, R Stote, J Straub, M Watanabe, J Wiórkiewicz-Kuczera, D Yin, and M Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[14] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926, 1983.

[15] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[16] S Izrailev, S Stepaniants, M Balsera, Y Oono, and K Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72(4):1568–1581, 1997.

[17] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J. Mol. Graphics*, 14(1):33–38, 1996.

[18] João V Ribeiro, Juan A C Tamames, Nuno M F S A Cerqueira, Pedro A Fernandes, and Maria J Ramos. Volarea - a bioinformatics tool to calculate the surface area and the volume of molecular systems. *Chem. Biol. Drug. Des.*, 82(6):743–755, December 2013.

[19] J Eargle and Z A Luthey-Schulten. NetworkView: 3D display and analysis of protein·RNA interaction networks. *Bioinformatics*, 28(22):3000–3001, 2012.

[20] Anurag Sethi, John Eargle, Alexis A Black, and Zaida Luthey-Schulten. Dynamical networks in tRNA:protein complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 106(16):6620–6625, April 2009.

[21] N Glykos. Software news and updates carma: A molecular dynamics analysis program. *J. Comput. Chem.*, 27(14):1765–1768, 2006.

[22] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85(1):5–15, July 2003.

[23] Olga K Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, March 2006.

[24] O K Dudko, G Hummer, and A Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *PNAS*, 105(41):15755–15760, October 2008.

[25] M Karplus and J N Kushick. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14:325–332, 1981.

[26] R M Levy, M Karplus, J Kushick, and D Perahia. Evaluation of the Configurational entropy for Proteins: Application to Molecular Dynamics Simulations of an α-Helix. *Macromolecules*, 17(7):1370–1374, 1984.

[27] Rommie E Amaro, Anurag Sethi, Rebecca S Myers, V Jo Davisson, and Zaida A Luthey-Schulten. A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. *Biochemistry*, 46(8):2156–2173, February 2007.

[28] Rebecca W Alexander, John Eargle, and Zaida Luthey-Schulten. Experimental and computational determination of tRNA dynamics. *FEBS Lett.*, 584(2):376–386, January 2010.

[29] Rafael C Bernardi, Isaac Cann, and Klaus Schulten. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnol. Biofuels*, 7(83):1–8, 2014.

[30] Oliver F Lange and Helmut J Grubmüller. Generalized correlation for biomolecular dynamics. *Proteins*, 62(4):1053–1061, March 2006.