

Supplementary 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

^aLehrstuhl für Angewandte Physik and Center for Nanoscience,
Ludwig-Maximilians-Universität, 80799 Munich, Germany

^bTheoretical and Computational Biophysics Group, Beckman Institute
for Advanced Science and Technology, University of Illinois at
Urbana-Champaign, Urbana, Illinois 61801, United States

^cCenter for Integrated Protein Science Munich (CIPSM), University of
Munich, 81377 Munich, Germany

^dDepartment of Biological Chemistry, The Weizmann Institute of
Science, Rehovot 76100, Israel

^eDepartment 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 paper². 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 chloramphenicol containing ZYM-5052 media³ 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 protocols⁴. 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 protocols⁵. 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 $\mu\text{m/s}$. 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 $b = 0.11$ nm connected by a fixed angle $\gamma = 41^\circ$ 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 detail⁸. More specifically molecular dynamics simulations have always been viewed as a general sampling method for the study of conformational changes⁹. 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 model¹⁴ 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 scheme¹¹ 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 simulations¹⁶ 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 VMD¹⁷ and its plugins. Surface contact areas of interacting residues were calculated employing Volarea¹⁸ 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 plugin¹⁹ on VMD¹⁷ 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 $w_{ij} = -\log(|C_{ij}|)$, where C_{ij} is the correlation matrix calculated with Carma²¹. 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 $-\log(0.5) = 0.69$. To calculate the relevance of off-diagonal terms in the correlation matrix we employed Carma to calculate a correlation matrix where *xyz* 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 highly correlated, while XMod is 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 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 Bell-Evans Model

The Bell-Evans model²² 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:

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

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 Bell-Evans picture is given as:

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

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:

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

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 $\Delta G \rightarrow \infty$ independent of ν the phenomenological Bell-Evans theory is recovered.

4.3 Pearson Correlation and covariance matrix

4.3.1 Validation

A $N \times N$ matrix of Pearson correlation coefficients C_{ij} (Supplementary Eq. S4) was calculated from each atoms' xyz 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 Supplementary Eqs. (S5) and (S6)) and Supplementary 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 interactions³⁰. To validate the use of Pearson correlations, we therefore first analyzed independently the contributions from diagonal and off-diagonal elements of each 3×3 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 Supplementary 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:

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

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:

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

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

$$D_{ij}^{mn} = \frac{M_{ij}^{mn}}{\sqrt{M_{ij}^{mm} M_{ij}^{nn}}} \quad (\text{S6})$$

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

4.3.3 Derivation of Main Text Equation 2

Eq. 1 from the main text reads:

$$\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j^T \rangle = k_B T \frac{\partial \mathbf{r}_j}{\partial \mathbf{F}_i} \quad (\text{S7})$$

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

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

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

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

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:

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

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

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

Now we inspect the normalization of C_{ij} :

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

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:

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

Applying the equipartition theorem to this result yields:

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

And therefore:

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

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

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

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

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

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

5 Supplementary Figures

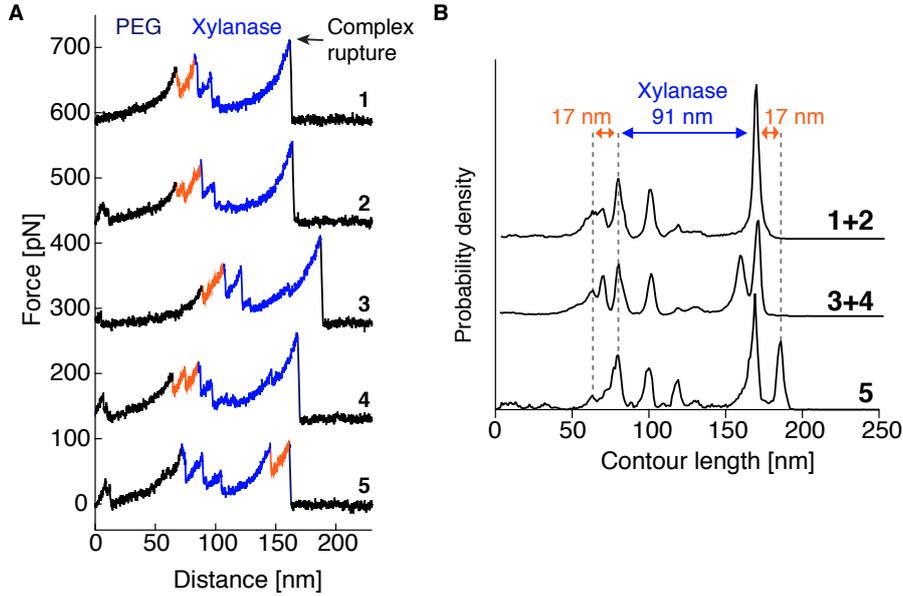


Fig. S1: SMFS of the non-native low force curve class. **A** Typical unfolding fingerprints. All traces showed a characteristic Xyn fingerprint (blue). A 17 – 19 nm increment corresponding to partial N-terminal Coh unfolding (orange) occurs either prior to Xyn unfolding (traces 1-4), or just before complex 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.

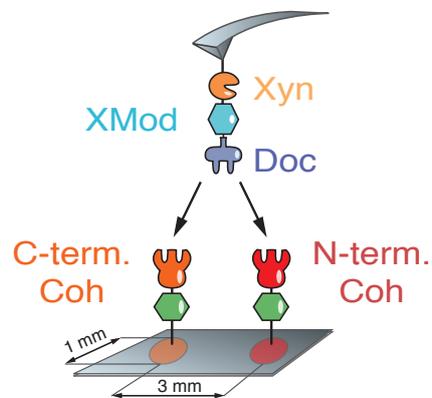


Fig. S2: Comparing the native geometry with the non-native high force class. To exclude uncertainties in cantilever calibration when comparing the native geometry with the non-native HF class, we immobilized both Coh-CBM (native) and CBM-Coh (non-native) on two spatially separated spots on a single cover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

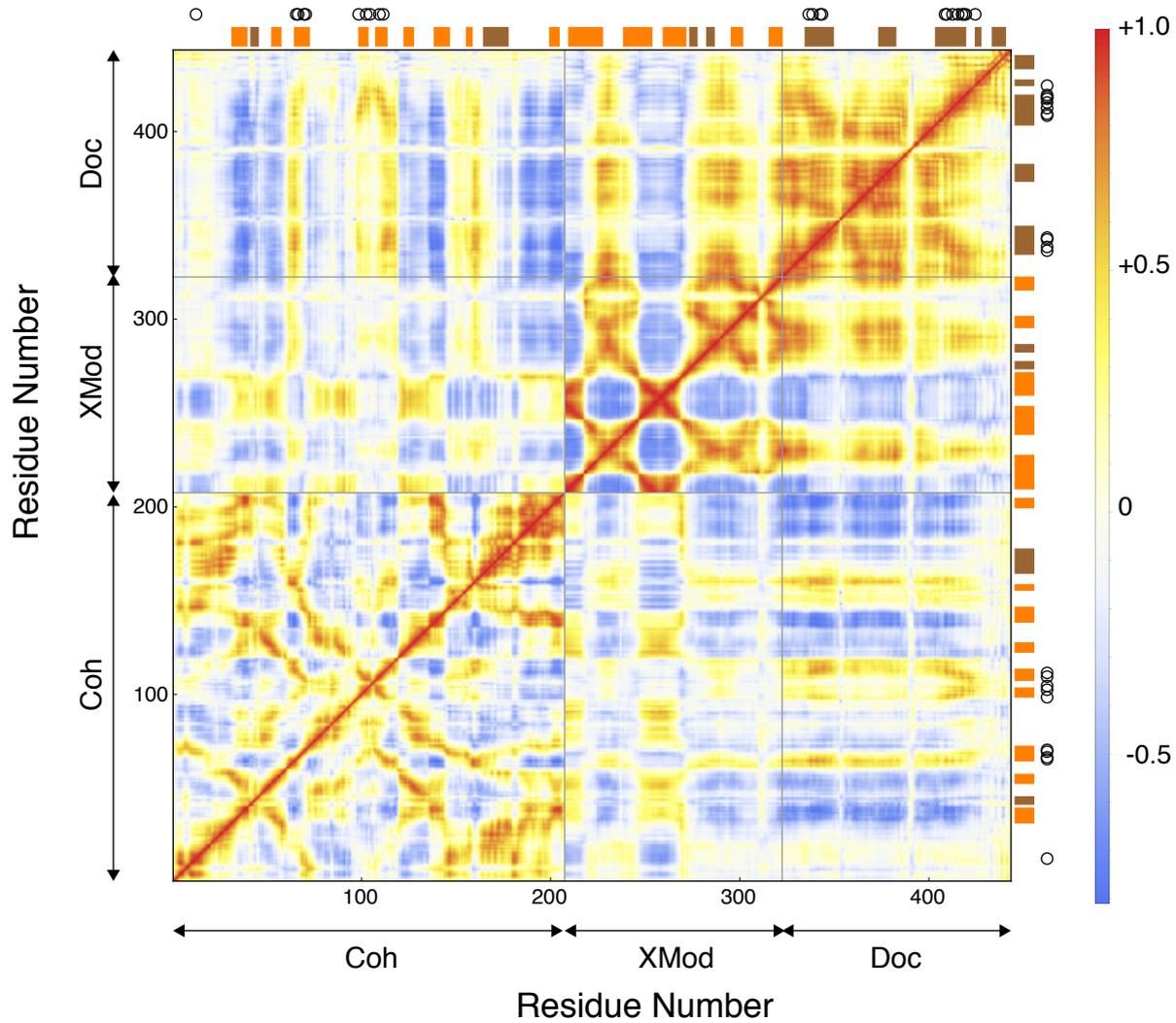


Fig. S3: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the unloaded Xmod-Doc:Coh complex. α -helices and β -strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from the Coh and Doc binding interface.

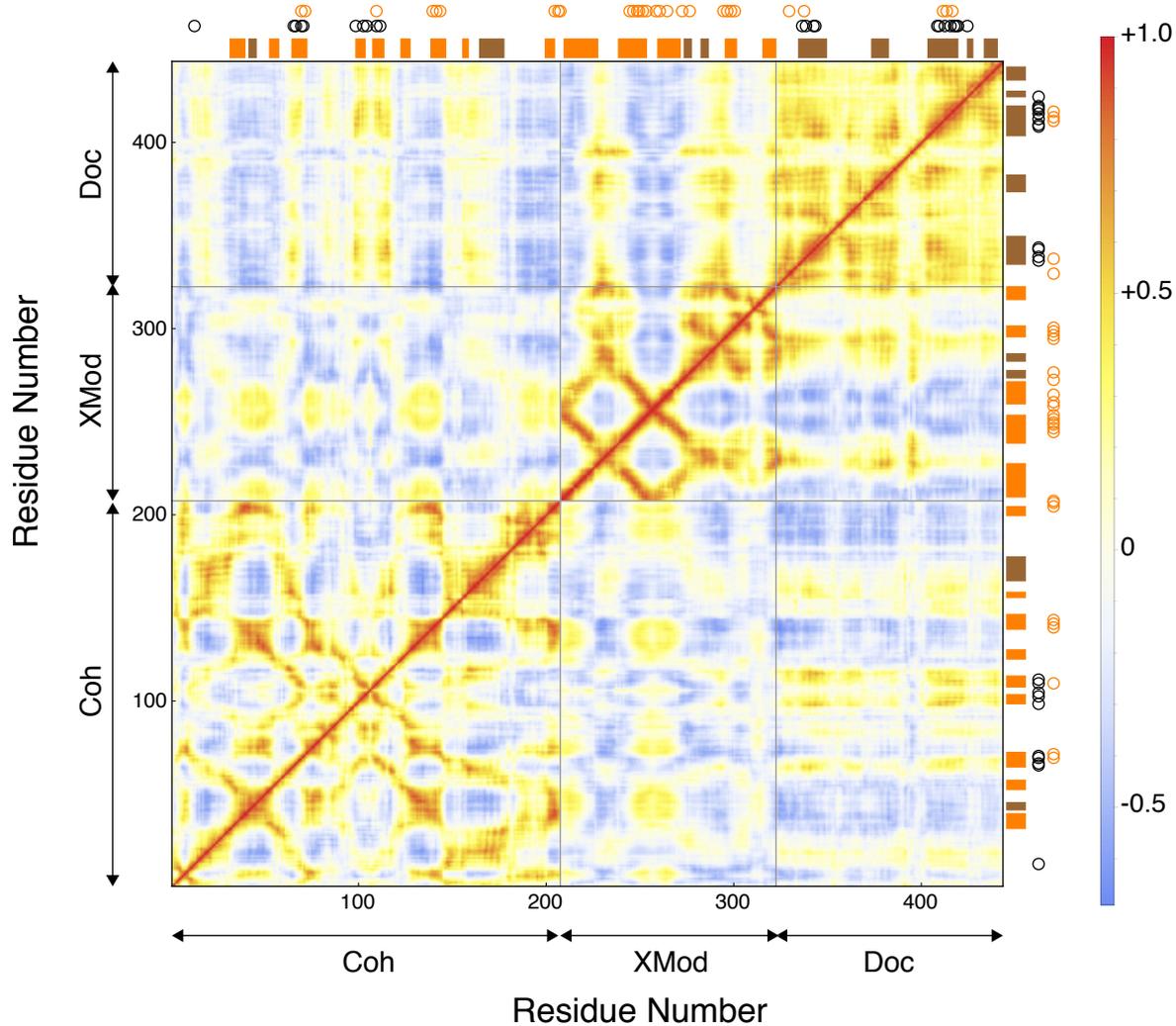


Fig. S4: Heat maps of the Pearson Correlation coefficient (C_{ij}) of the Xmod-Doc:Coh complex loaded with force in the native pulling geometry. α -helices and β -strands are highlighted with brown and orange rectangles, respectively. Black circles indicate binding residues from Coh and Doc binding interfaces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

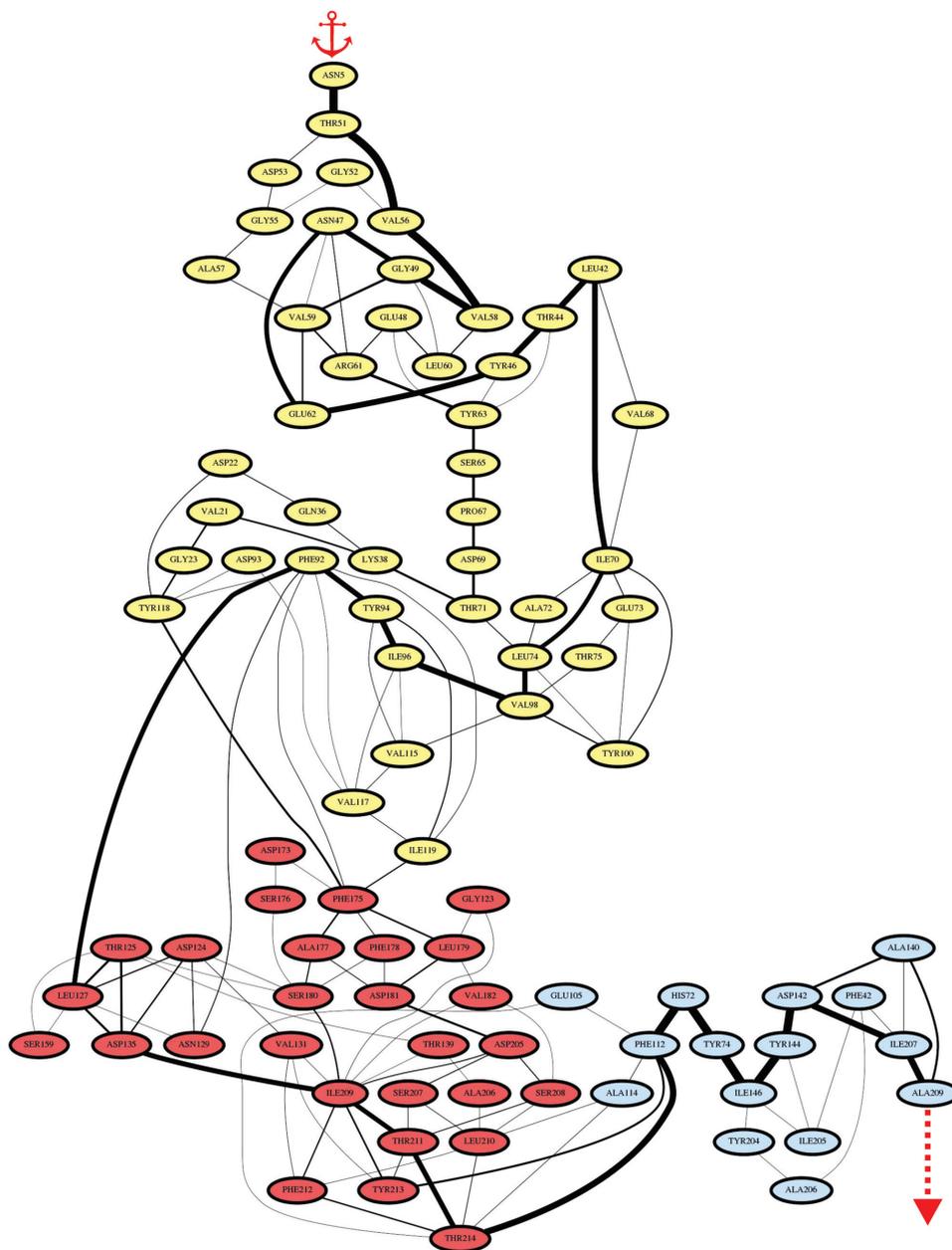


Fig. S5: Force propagation pathway through the loaded XMod-Doc:Coh complex in the native pulling geometry (N-terminal pulling of Xmod-Doc, C-terminal pulling of Coh) obtained from dynamical network analysis. Residues belonging to Xmod, Doc and Coh are colored in yellow, red and blue, respectively. Connecting 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.

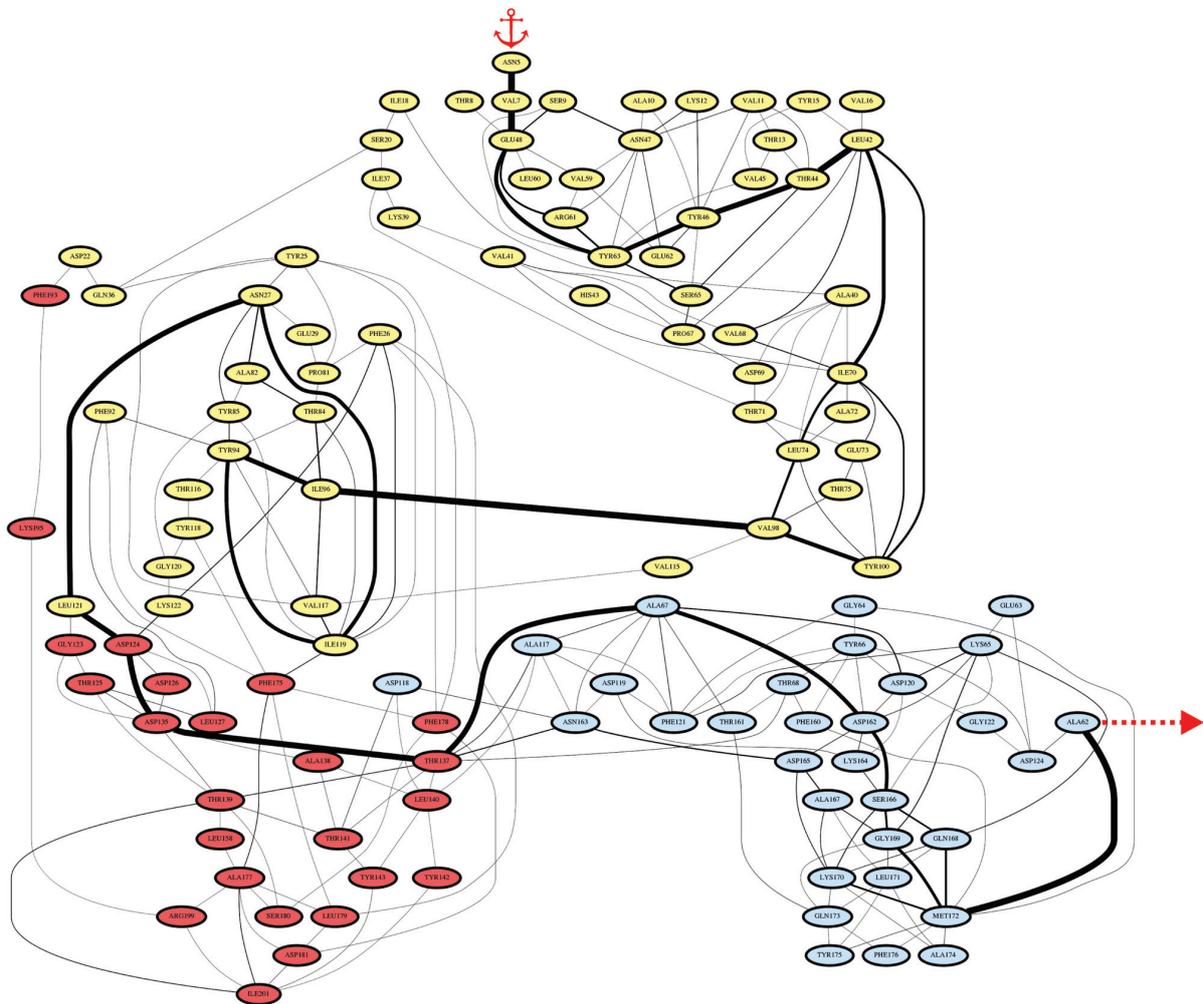


Fig. S7: Force propagation pathway through the loaded XMod-Doc:Coh complex in the non-native pulling geometry (N-terminal pulling of Xmod-Doc, N-terminal pulling of Coh) showing low-force unbinding characteristics and partial N-terminal Coh unfolding. Residues belonging to Xmod, Doc and Coh are colored 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.

$$M_{ij} = \begin{pmatrix} M_{11x} & M_{11y} & M_{11z} & M_{12x} & M_{12y} & M_{12z} & M_{13x} & M_{13y} & M_{13z} & M_{14x} & M_{14y} & M_{14z} & M_{15x} & M_{15y} & M_{15z} \\ M_{12x} & M_{12y} & M_{12z} & M_{13x} & M_{13y} & M_{13z} & M_{14x} & M_{14y} & M_{14z} & M_{15x} & M_{15y} & M_{15z} & M_{21x} & M_{21y} & M_{21z} \\ M_{13x} & M_{13y} & M_{13z} & M_{14x} & M_{14y} & M_{14z} & M_{15x} & M_{15y} & M_{15z} & M_{21x} & M_{21y} & M_{21z} & M_{22x} & M_{22y} & M_{22z} \\ M_{21x} & M_{21y} & M_{21z} & M_{22x} & M_{22y} & M_{22z} & M_{31x} & M_{31y} & M_{31z} & M_{32x} & M_{32y} & M_{32z} & M_{33x} & M_{33y} & M_{33z} \\ M_{22x} & M_{22y} & M_{22z} & M_{31x} & M_{31y} & M_{31z} & M_{32x} & M_{32y} & M_{32z} & M_{33x} & M_{33y} & M_{33z} & M_{41x} & M_{41y} & M_{41z} \\ M_{31x} & M_{31y} & M_{31z} & M_{32x} & M_{32y} & M_{32z} & M_{33x} & M_{33y} & M_{33z} & M_{41x} & M_{41y} & M_{41z} & M_{42x} & M_{42y} & M_{42z} \\ M_{41x} & M_{41y} & M_{41z} & M_{42x} & M_{42y} & M_{42z} & M_{43x} & M_{43y} & M_{43z} & M_{44x} & M_{44y} & M_{44z} & M_{45x} & M_{45y} & M_{45z} \\ M_{32x} & M_{32y} & M_{32z} & M_{41x} & M_{41y} & M_{41z} & M_{42x} & M_{42y} & M_{42z} & M_{43x} & M_{43y} & M_{43z} & M_{44x} & M_{44y} & M_{44z} \\ M_{33x} & M_{33y} & M_{33z} & M_{42x} & M_{42y} & M_{42z} & M_{43x} & M_{43y} & M_{43z} & M_{44x} & M_{44y} & M_{44z} & M_{45x} & M_{45y} & M_{45z} \\ M_{44x} & M_{44y} & M_{44z} & M_{45x} & M_{45y} & M_{45z} & M_{51x} & M_{51y} & M_{51z} & M_{52x} & M_{52y} & M_{52z} & M_{53x} & M_{53y} & M_{53z} \\ M_{51x} & M_{51y} & M_{51z} & M_{52x} & M_{52y} & M_{52z} & M_{53x} & M_{53y} & M_{53z} & M_{54x} & M_{54y} & M_{54z} & M_{55x} & M_{55y} & M_{55z} \\ M_{52x} & M_{52y} & M_{52z} & M_{53x} & M_{53y} & M_{53z} & M_{54x} & M_{54y} & M_{54z} & M_{55x} & M_{55y} & M_{55z} & M_{55x} & M_{55y} & M_{55z} \\ M_{53x} & M_{53y} & M_{53z} & M_{54x} & M_{44y} & M_{44z} & M_{55x} & M_{55y} & M_{55z} & M_{55x} & M_{55y} & M_{55z} & M_{55x} & M_{55y} & M_{55z} \end{pmatrix}$$

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.

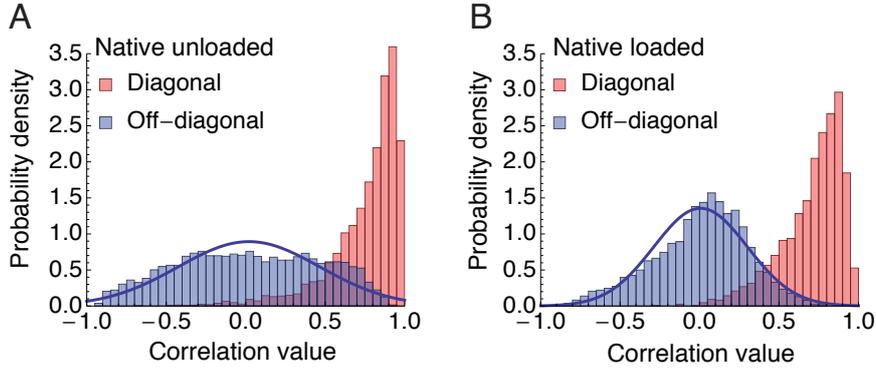


Fig. S9: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A** the native unloaded and **B** the native loaded scenario.

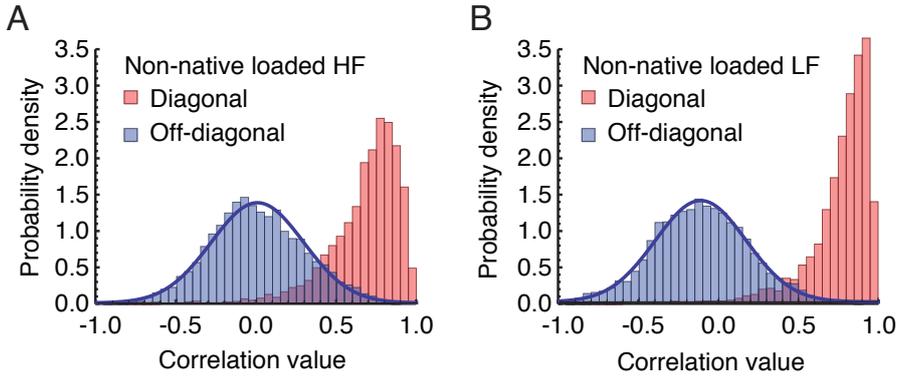


Fig. S10: Histograms showing contributions of diagonal and off-diagonal terms of the full covariance matrix elements fulfilling proximity criteria for **A** the non-native HF and **B** the non-native LF scenario.

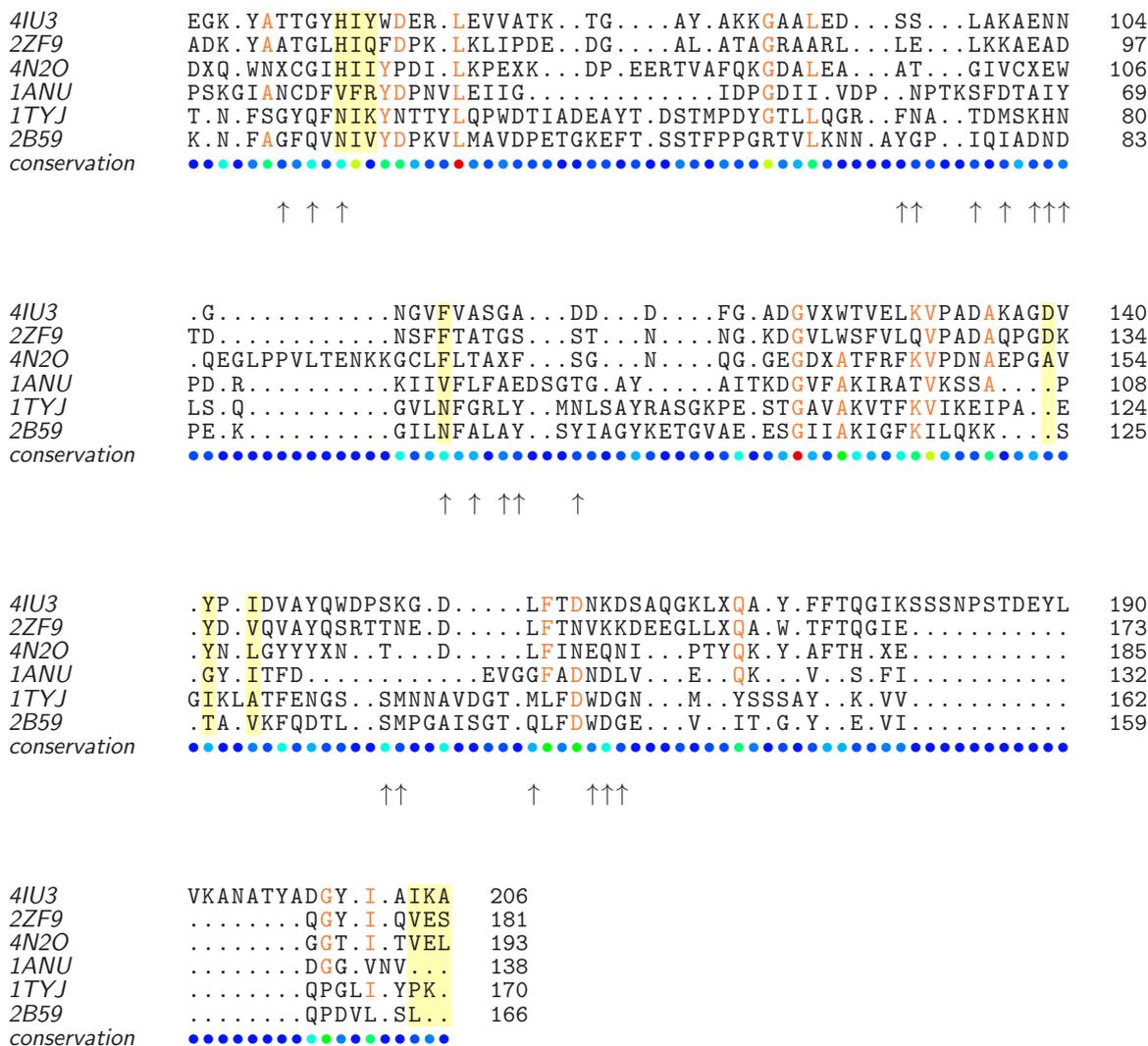


Fig. S11: Structure-aligned sequences of six crystallized cohesins. Residues on the force propagation path are highlighted in yellow. Arrows indicate binding residues. Residue conservation is color coded from blue - lack of conservation, to red - residue fully conserved. Crystal structures used: 4IU3 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*.

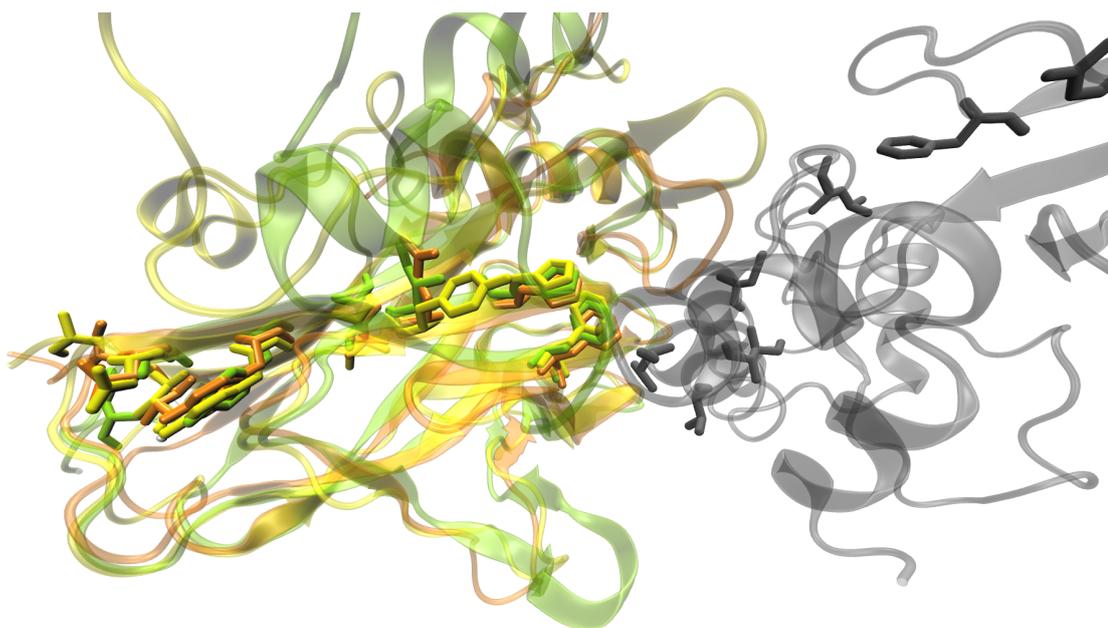


Fig. S12: Structure and sequence conservation of the force propagation pathway residues in Coh. CohE from the ScaE cell anchoring protein, *Rf* FD-1 used in this work (PDB 4IU3) is highlighted in green. Highly homologous structures of CohE from *Rf* strain 17 (PDB 2ZF9) and Coh G from *Rf* FD-1 (PDB 4A2O) are colored in orange and yellow, respectively. Residues lying in force propagation path are showed as sticks. XDoc from 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 1 and χ 2 Dihedral 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, Jeffrey 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] R Merkel, P Nassoy, A Leung, K Ritchie, and Evan Evans. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, 1999.
- [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.