# Supporting Information:

# Steered molecular dynamics simulations reveal the role of Ca2+ in regulating mechanostability of cellulose-binding proteins

Melissabye Gunnoo,<sup>a</sup> Pierre-André Cazade,<sup>a</sup> Adam Orlowski,<sup>a</sup> Mateusz Chwastyk,<sup>b,c</sup> Haipei Liu,<sup>d,e</sup> Duy Tien Ta, <sup>d,e</sup> Marek Cieplak,<sup>c</sup> Michael Nash,<sup>d,e</sup> Damien Thompson<sup>a\*</sup>

<sup>a</sup> Department of Physics, Bernal Institute, University of Limerick, V94 T9PX, Ireland

<sup>b</sup> Department of Physics, Arizona State University Tempe, Arizona 85287, USA

<sup>c</sup> Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Warsaw, Poland

<sup>d</sup> Department of Chemistry, University of Basel, 4056 Basel, Switzerland

<sup>e</sup> Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland

\* Correspondence to: damien.thompson@ul.ie

## List of material included:

**Table S1**. List of multiple mutants and their corresponding labels.

Figure S1. F<sub>1</sub> distribution for all point mutants, calculated from at least two repeats.

Figure S2. The difference between contact maps of intermediate states: i1 and i2.

Figure S3. Computed force-extension profiles of CBM3 C-to-N unfolding for two pathways. Both A and B show four peaks.

#### Supplementary note on simulations of pulling from the C-terminus

Figure S4. Force-extension curve for the complete unfolding of CBM3 module for 10 repeats.

**Figure S5**. Force-extension curve of the complete unfolding of native CBM3 domain shows three peaks.

Table S2. List of the main contributors at peaks I, II and III.

**Figure S6**. **A** – Interaction energies between amino acids with hydrophobic side chains in the native CBM3 during the first 500 Å extension.

Figure S7. Lateral shearing mechanism between beta strands 2-7 during ybbR\_N stretching.

**Figure S8**. Longitudinal shearing (unzipping of H-bonds) of beta strands 3-6 during ybbR\_N stretching.

**Figure S9.** A – Histogram showing the unfolding forces of native CBM3, fitted following the Bell-Evans model (red dashed line). A total of 365 unfolding events were recorded and the most probable unfolding force was  $\langle F \rangle = 143.8$  pN. B – Dynamic force spectrum for the unfolding events of CBM3. Plotted points represent unfolding force and loading rate data from SMD simulations at five different pulling velocities (ranging between 0.005 Å ps<sup>-1</sup> and 0.5 Å ps<sup>-1</sup>) and SMFS-experiment at cantilever retraction velocity of 500 nm s<sup>-1</sup>.

**Figure S10.** C-to-N force-extension profiles of native CBM3 at :  $\mathbf{A} - \mathbf{v} = 0.05$  Å/ns  $\mathbf{B} - \mathbf{v} = 0.00005$  Å/ns from coarse-grained simulations. Respective beta strand pairs (corresponding numbering in Fig. 1C) involved in each peak are shown

# Supplementary note on benchmarking against control SMFS measurements and coarse grain models

**Figure S11.** CBM3 showing the positions of single mutations:  $\mathbf{A}$  – for  $\beta$ -sandwich destabilization.  $\mathbf{B}$  – Hydrophobic residues that were changed to more polar amino acids. All computed molecular structures were visualized using VMD software.<sup>1</sup>

**Figure S12**. **A** - Native CBM3 structure after 100 ns of molecular dynamics: the calcium ion is drawn as a yellow sphere and two ASP residues interacting with the ion are shown as pink sticks. **B** - The D46P mutant after 100 ns of dynamics: the calcium ion does not interact with mutant residue P46 but instead migrates along the surface of the protein to coordinate D139, which changes also the native conformation of loop 7-8 (shown as a red line C $\alpha$  trace). **C** – ybbR\_D46P mutant after 100 ns molecular dynamics: the grafted tag allows Asp89 coordinate the calcium, creating a third distinct calcium binding site in which the ion migrates across the binding pocket. **D** – Molecular mechanics combined with Poisson-Boltzmann (MM/PBSA) binding free energy component analysis between the cation and residues (Thr44, Asp46, Thr122, Asn125 and Asp126) involved in the Ca<sup>2+</sup> binding pocket, for native and mutated CBM3. **E** – Bar chart of computed H-bond statistics in loop 7-8 in WT and in mutants where the calcium ion is either destabilized or deleted.

Figure S13. Root mean square deviation of C-alpha atoms of mutated proteins: Beta mutants.

**Figure S14**. Root mean square deviation of C-alpha atoms of mutated proteins: Hydrophobic mutants.

Figure S15. Root mean square deviation of C-alpha atoms of mutated proteins: Calcium interactions mutants.

Figure S16. Root mean square fluctuations of C-alpha atoms of wildtype CBM3.

Figure S17. Root mean square fluctuations of C-alpha atoms of mutated proteins: Beta mutants.

Figure S18. Root mean square fluctuations of C-alpha atoms of mutated proteins: Hydrophobic mutants.

Figure S19. Root mean square fluctuations of C-alpha atoms of mutated proteins: Calcium mutants.

#### Supplementary note on computed properties of mutant and tagged CBM structures

#### Supplementary note on coordination sphere of calcium ion

Figure S20. N-to-C force-extension profiles of CBM3: A – native protein B – without calcium ion.

Figure S21. Force-extension curve for the first 100 Å N-C distance: Beta mutants. The simulations were repeated at least 3 times.

Figure S22. Force-extension curve for the first 100 Å N-C distance: Hydrophobic mutants.

Figure S23. Force-extension the first 100 Å N-C distance: Calcium interactions mutants.

Figure S24. Force-extension curve for the first 100 Å N-C distance: Ybbr-tag.

**Figure S25**. Force-extension curve for the unfolding of first half of CBM3 (Residue 1 to 82). N terminus is anchored while Ser82 is pulled at constant velocity.

**Figure S26**. Force-extension curve for the unfolding of second half of CBM3 (Residue 82 to 166). C terminus is anchored while Ser82 is pulled at constant velocity.

Figure S27. Maximum force distribution for peak III, for calcium sphere related mutants, calculated from at least two repeats.

**Figure S28**. Force-extension curves of CBM3 steered molecular dynamics simulations at five different constant velocities for C-to-N stretching.

#### Supplementary note on the unfolding landscape of ybbR\_N

Supplementary note on the unfolding landscape of ybbR\_C

#### Supplementary note on Methods

- Molecular dynamics methods
- Experimental methods
- Coarse-grained methods

Movie S1. The most common unfolding pathway of native CBM3.

**Movie S2.** Unfolding pathway of first half of CBM3 (Residue 1 to 82). N terminus is anchored while Ser82 is pulled at constant velocity.

**Movie S3.** Unfolding pathway of second half of CBM3 (Residue 82 to 166). C terminus is anchored while Ser82 is pulled at constant velocity.

Supporting simulation movie files are available at: https://ldrv.ms/f/s!ArX4zU6cjMUQmTCPuuCMjME121pC

	Label	Mutations		
Double	M1	V4N_F22N		
	M2	E5A_D46G		
	M3	E5Q_D46G		
	M4	D46G_F135A		
	M5	K23P_R40P		
	M6	L2G_W151A		
Triple	M7	E3A_E5A_D46G		
	M8	E5A_D46G_L48G		
	M9	E5Q_D46G_K49A		
	M10	F22A_D46G_F135A		
	M11	E5A_D46G_I109A		
Quadruple	M12	E3A_E5A_D46G_F135A		
	M13	E3A_E5A_D46G_F137A		
	M14	E3A_E5A_D46G_V136A		
	M15	E5A_D46G_I109A_F135A		
Multiple	M16	E5A_D46G_I109A_F135G_D139A		
	M17	L2A_V4A_F22A_L34A_W138A		
	M18	L2N_V4N_F22N_L34N_W138N		
	M19	F22G_L37G_L39G_F53G_L92G_V107G_F113G_F135G_L145G		

**Table S1**. List of multiple mutants and their corresponding labels.

**Figure S1**.  $A - F_1$  distribution for all point mutants, calculated from at least two repeats. The red line represents the average  $F_1$  (1023 pN) of native protein and green lines represent the error, **B** – Percentage decrease in  $F_1$  of mutants compared to that of native CBM3.



Figure S2. The difference between contact maps of intermediate states: A & B – contacts loss and gain between native state and i1, C & D - contacts loss and contacts gain between i1 and i2.



**Figure S3**. Computed force-extension profiles of CBM3 C-to-N unfolding for two pathways. Both **A** and **B** show four peaks.



#### **C-terminus simulations**

When the spring is attached to C instead of the N terminus, the order in which strands 1, 2, 8 and 9 unravel is altered and so is the force-extension curve. On visual analysis of our simulations, it can be stated that pulling on the C terminus encourages stretching in the following order;  $\beta$ 9 followed by  $\beta$ 8, then  $\beta$ 1 and  $\beta$ 2. This pathway gives rise to a four peak curve instead of three as previously discussed in N-terminus pulling. However, peak I at 11 Å just like in the N terminus simulations, arises due to hydrophobic interactions combined with peeling off of  $\beta$ 8- $\beta$ 9. The second peak is due to a combination of cation  $\pi$  interactions between Lys23 and Phe135, anion  $\pi$  interactions involving Glu5 and Phe135 as well as resistance against Phe6 on  $\beta$ 1 forming two hydrogen bonds with Val136. At peak 3, a small pair of strands  $\beta 1^{2}-\beta 7^{2}$  along the side of the beta sheets separate and more importantly the longitudinal shearing of  $\beta$ 1- $\beta$ 2 as well as part of hydrogen bond breaking between  $\beta$ 2- $\beta$ 7 contribute to the barrier. The remaining  $\beta 2$ - $\beta 7$  hydrogen bonds split is reflected in the shoulder of peak III. For the last peak, IV, calcium ion and Asp126 interaction are separated. All the other components in the calcium coordination sphere are overcome earlier in the trajectory resulting in a slight decrease in the F<sub>4</sub> associated with this peak (Figure S3). Force distribution in both N and C terminus simulations were compared through analysis of area under the curve. The force integral for N terminus of 4.20 x 10<sup>14</sup> kcal/mol was slightly lower than that of C terminus at  $4.50 \times 10^{14}$  kcal/mol.



**Figure S4**. Force-extension curve for the complete unfolding of CBM3 module. The simulation was repeated 9 times with all parameters and starting configurations unchanged.

Figure S5. A – Force-extension curve of the complete unfolding of native CBM3 domain shows three peaks.  $\mathbf{B}$  – Change of protein height in the first 11 Å of pulling extension. Protein conformation at 0 Å extension is shown on the left: The beta sheets are tightly stuck to one another due to van der Waals forces between hydrophobic residues. Protein conformation at 11 Å extension is shown on the right: Partial separation of two sheets results in significant loss in van der Waals interactions (Fig. S11). Hydrogen bonding and calcium ion interactions also contribute to this peak. Events include the shearing of pair 68-69, elongating from 11 Å to 30 Å, followed by shearing of the salt bridge Glu5-Lys23 and then breaking of three sets of hydrogen bond pairs between  $\beta 1-\beta 2$  (20 Å to 50 Å). Before any unfolding occurs, in the first 11 Å extension, the average distance between the protein sheets changes from 15 Å to 29 Å which causes a significant loss in van der Waals contacts between residues in the (now stretched) hydrophobic core. C – Protein structure at Peak II: numerous H-bonds (red dotted lines) between  $\beta$ -strands 8-3 and 2-7 are disrupted. H-bonds separation occurs in all three peaks and depending on direction of the force vector relative to hydrogen bonds, either longitudinal or lateral shearing occurs. D – Protein conformation just before the calcium ion pocket is disrupted (Peak III). Yellow sphere represents the calcium ion while Asp46 and Asp126 are shown in red.



Peak	Maximum force (pN)	Contributors	Main β strands involved	No. of H- bonds
I	$1023 \pm 127$	Combination of core hydrophobic interactions, hydrogen bonding and the calcium pocket	β8-β9 and β1-β2	~ 12
II	950 ± 120	Mechanical clamps between beta strands	β3-β8 and part of β2- β7	~ 10
III	$1392 \pm 171$	Ca <sup>2+</sup> binding sphere and H- bonding between β-strands	β3-β6 and β7-β4	~ 15

Table S2. List of the main contributors at peaks I, II and III.

**Figure S6**. A – Interaction energies between amino acids with hydrophobic side chains in the native CBM3 during the first 500 Å extension: Electrostatic energies in black , van der Waals in red and total energies in green, **B** – interaction energies in the multiple mutant M17 with five hydrophobic residues changed to Alanine.



Figure S7. Lateral shearing mechanism between beta strands 2-7 during ybbR\_N stretching. A – Before an external force is exerted. B – All/most H-bonds between the strands break for one of the strands to slide.



**Figure S8**. Longitudinal shearing (unzipping of H-bonds) of beta strands 3-6 during ybbR\_N stretching.



**Figure S9. A** – Histogram showing the unfolding forces of native CBM3, fitted following the Bell-Evans model (red dashed line). A total of 365 unfolding events were recorded and the most probable unfolding force was  $\langle F \rangle = 143.8$  pN. **B** – Dynamic force spectrum for the unfolding events of CBM3. Plotted points represent unfolding force and loading rate data from SMD simulations at five different pulling velocities (ranging between 0.005 Å ps<sup>-1</sup> and 0.5 Å ps<sup>-1</sup>) and SMFS-experiment at cantilever retraction velocity of 500 nm s<sup>-1</sup>.



**Figure S10.** C-to-N force-extension profiles of native CBM3 at :  $\mathbf{A} - \mathbf{v} = 0.05$  Å/ns  $\mathbf{B} - \mathbf{v} = 0.00005$  Å/ns from coarse-grained simulations. Respective beta strand pairs (corresponding numbering in Fig. 1C) involved in each peak are shown.



Benchmarking against control SMFS measurements and coarse grain models

Unfolding of native CBM, experimental SMFS measurements were carried out using automated AFM-based SMFS as described in the Methods section. The most probable unfolding force recorded was  $\langle F \rangle = 143.8$  pN. The histogram showing the unfolding forces of the fingerprint CBM domain is shown in Fig. 6A. When compared to simulations, the absolute forces in AFM experiments are shifted towards lower values due to the much lower loading rate than in simulations. This is as expected and is consistent with literature,<sup>2</sup> given that the SMFS based experiments and SMD simulations are carried out at time scales that differ by more than six orders of magnitude. The force deviation between SMFS-experiments and all-atom simulations can be explained by the Bell-Evans equation, which predicts a linear dependence of the unfolding force with the logarithm of the loading rate.<sup>3</sup> Figure 9B shows the best-fit dynamic force spectrum generated from both SMFS and MD simulation data of native CBM3. Moreover, it should be noted that unlike the first peak barrier measured in

SMD simulations, the unfolding force recorded in SMFS experiments is a measure of combination of all three unfolding force barriers encountered in the unfolding of native CBM3 as contour lengths were  $\sim 500$  Å. Measured values of unfolding rate at zero force,  $k_{off}$  and the unfolding distance  $\Delta x$  were slightly higher in SMD simulations compared to SMFS experiments. The  $k_{off}$  values for best-fit dynamic force spectrum generated from MD simulations and SMFS experiments were  $2.99 \times 10^{-3} s^{-1}$  and  $1.77 \times 10^{-3} s^{-1}$ , respectively while  $\Delta x$  values were 0.73 nm and 0.34 nm, respectively. The difference can be attributed to the difference in time scale between SMD and SMFS as discussed above.

Coarse-grained (CG) simulations of CBM unfolding were performed using a Go-like model<sup>4-</sup> <sup>8</sup>, as described in the Methods section. In this model, only C $\alpha$  are taken into account for the dynamics and their interactions are governed by a contact map of atoms. Force-extension profiles of coarse-grained unfolding of native CBM3 at two velocities are presented in Figure S10. The complete unfolding lengths of the protein is the same in the unfolding profile of CBM in SMD (Fig. S5) and CG simulations (Fig. S10). However, the number of peaks and position of each peak in the CG profiles differ from those in the atomistic models (Fig. S10). This discrepancy is due to calcium sphere interactions which were not included explicitly in the CG model used. As a result, the sharp peak at 250 Å is missing in unfolding profiles from CG compared to MD simulations. In addition, the unfolding pathways (Table S2 and Fig. S10B) from the two simulation methods diverge slightly. Nevertheless, CG approach permits extensive sampling of CBM3 unfolding at multiple pulling velocities and more importantly at velocities similar to those used in experiments. The general agreement between the detailed atomic scale MD dataset and the experimental measurements and coarse grain models strengthens the predictive power of the MD data to, for example, identify promising leads for future site-directed mutagenesis experiments.

**Figure S11.** CBM3 showing the positions of single mutations:  $A - for \beta$ -sandwich destabilization. B – Hydrophobic residues that were changed to more polar amino acids. All computed molecular structures were visualized using VMD software.<sup>1</sup>



**Figure S12**. **A** - Native CBM3 structure after 100 ns of molecular dynamics: the calcium ion is drawn as a yellow sphere and two ASP residues interacting with the ion are shown as pink sticks. **B** - The D46P mutant after 100 ns of dynamics: the calcium ion does not interact with mutant residue P46 but instead migrates along the surface of the protein to coordinate D139, which changes also the native conformation of loop 7-8 (shown as a red line C $\alpha$  trace). **C** – ybbR\_D46P mutant after 100 ns molecular dynamics: the grafted tag allows Asp89 coordinate the calcium, creating a third distinct calcium binding site in which the ion migrates across the binding pocket. **D** – Molecular mechanics combined with Poisson-Boltzmann (MM/PBSA) binding free energy component analysis between the cation and residues (Thr44, Asp46, Thr122, Asn125 and Asp126) involved in the Ca<sup>2+</sup> binding pocket, for native and mutated CBM3. **E** – Bar chart of computed H-bond statistics in loop 7-8 in WT and in mutants where the calcium ion is either destabilized or deleted.





Figure S13. Root mean square deviation of C-alpha atoms of mutated proteins: Beta mutants.



**Figure S14**. Root mean square deviation of C-alpha atoms of mutated proteins: Hydrophobic mutants.

**Figure S15**. Root mean square deviation of C-alpha atoms of mutated proteins: Calcium interactions mutants.









**Figure S17**. Root mean square fluctuations of C-alpha atoms of mutated proteins: Beta mutants.



**Figure S18**. Root mean square fluctuations of C-alpha atoms of mutated proteins: Hydrophobic mutants.

**Figure S19**. Root mean square fluctuations of C-alpha atoms of mutated proteins: Calcium mutants.



#### Computed properties of mutant and tagged CBM structures

The structural stability of each mutant was monitored in explicit water during 0.1 microseconds of molecular dynamics. All mutants, with the exception of D46P and D46G, had deviated to a similar extent from their starting structures after 40 ns, resulting in RMSD values (root mean square deviation of the backbone C $\alpha$  atoms) of 0.8 ± 0.1 Å (Fig. S13-S15). D46P and D46G showed higher deviations (RMSD = 1.4 Å and 1.0 Å). Figures S16-S19 show that all beta and hydrophobic mutants displayed per-residue flexibilities (RMSF) similar to that of the native protein with sharp increases between residues 80 to 86 reflecting the flexibility of strands 5-6. D46P (Fig. S19) showed the strongest disruption with high RMSF values in the regions spanning residues 6-11 and 126-131. Both RMSD and RMSF results suggest that the majority of mutations conserved the structure of the native protein with only localized effects at the site of mutation, except in the case of D46P where loop 7-8 shifts to a new position in the protein (Fig. S12B). By contrast, in D46G loop 7-8 remains unaffected, but switching Asp46 to a small uncharged residue makes strand 3-4 (residues 43-50, Fig. 1C) more flexible, hence the high RMSD.

The computed global RMSD of the ybbR tagged CBM structure was  $1.6 \pm 0.2$  Å. When the ybbR-tag was excluded from the RMSD calculation, the protein RMSD was  $0.8 \pm 0.1$  Å, which was the same as that of native CBM3. The RMSDs of ybbR-D126A and ybbR-D46P were also low at  $0.8 \pm 0.1$  Å and  $0.9 \pm 0.1$  Å, respectively. This result suggests that the introduction of the ybbR-tag (with or without a single mutation) does not significantly perturb the CBM structure. By contrast, we computed minimal structural changes in ybbR-D46P compared to D46P alone. The addition of the ybbR-tag allows more flexibility around the insertion region and as a result promotes interactions between the calcium ion and an alternative aspartic acid residue (Asp89) which sits in closer proximity in ybbR-D46P than in D46P, hence loop 7-8 remains unchanged in ybbR-D46P (Fig. S12C).

#### Coordination sphere of calcium ion

The calcium ion is held in a buried cavity by loops connecting  $\beta$ 3- $\beta$ 4 and  $\beta$ 7- $\beta$ 8 (Fig. 1C), and forms a coordination sphere through electrostatic interactions with seven oxygen atoms. On loop 3-4 (which links  $\beta$ -strands 3 and 4), oxygen atoms from residues Thr44 and Asp46 form part of the calcium coordination sphere and the other oxygen atoms are from residues Thr122, Asn125 and Asp126 on loop 7-8. While Asp126 only uses one of its carboxylate oxygen atoms for binding, Asp46 provides a bidentate ligand,<sup>9</sup> and these Asp46 and Asp126 carboxylate oxygen atoms were the last to separate from the calcium ion during the forced unfolding simulations. An ordered water molecule completes the calcium pentagonal bipyramidal binding sphere, remaining within 3.8 ± 0.2 Å of Ca<sup>2+</sup> during at the 100ns equilibrium molecular dynamics. The lifespan of a specific water residue in the binding sphere varied in the range of 3-7 ns in the simulation trajectory. When the water residue leaves the sphere, it is replaced by another ordered water residue.

Figure S20. N-to-C force-extension profiles of CBM3: A - native protein B - without calcium ion.





**Figure S21**. Force-extension curve for the first 100 Å N-C distance: Beta mutants. The simulations were repeated at least 3 times.





Figure S22. Force-extension curve for the first 100 Å N-C distance: Hydrophobic mutants.



Figure S23. Force-extension the first 100 Å N-C distance: Calcium interactions mutants.



Figure S24. Force-extension curve for the first 100 Å N-C extension: Ybbr-tag.



**Figure S25**. Force-extension curve for the unfolding of first half of CBM3 (Residue 1 to 82). N terminus is anchored while Ser82 is pulled at constant velocity.



**Figure S26**. Force-extension curve for the unfolding of second half of CBM3 (Residue 82 to 166). C terminus is anchored while Ser82 is pulled at constant velocity.



#### The unfolding landscape of ybbR\_N

In ybbR\_N SMD simulations, beta strands 5-6 were the first to separate, followed by simultaneous separation of beta strand pairs 6-3 and 1-2 (Movie S2). These all follow an unzipping mechanism that occurs when the applied force is orthogonal to the target  $\beta$ -strands.<sup>10</sup> By contrast, in the next event,  $\beta$ 2 and  $\beta$ 7 detach by longitudinal shearing. Based on molecular models<sup>10</sup> of simplified systems, when longitudinal shearing of a number of hydrogen bonds (approximately eight hydrogen bonds between  $\beta$ 2-7) takes place, the shearing force required to rupture one hydrogen bond is equal to or slightly less than the total force-extension curves in Figure S24, in which a sharper peak is observed at extension ~150 Å which corresponds to the unbinding of  $\beta$ -strands 2 and 7. Note the ybbR\_N unfolding does not require dismantling of the calcium ion binding site.

#### The unfolding landscape of ybbR\_C

In the case of ybbR\_C simulations, just like for ybbR\_N, beta strands 5 and 6 are the first to delaminate by lateral shearing. This occurs in combination with a realignment of the protein which is followed by the unzipping of numerous beta strand pairs in the following order:  $\beta$ 8-9,  $\beta$ 8-3,  $\beta$ 6-3 and finally the calcium ion separates from residue 46 and  $\beta$ 4-7 ruptures concurrently (Movie S3). Due to mostly lateral shearing of beta strands in ybbR\_C simulations, the rupture force required prior to breaking the calcium sphere is below 1000 pN (Fig. S25).

Figure S27. Maximum force distribution for peak III, for calcium sphere related mutants, calculated from at least two repeats.



**Figure S28**. Force-extension curves of CBM3 steered molecular dynamics simulations at five different constant velocities for C-to-N stretching.



# Methods

## **Molecular dynamics methods**

#### Structure preparation and equilibrium molecular dynamics

The native structure of the CBM3 module was obtained from the RCSB Protein Data Bank (PDB code 1NBC)<sup>9</sup> and solvated in a box of TIP3P<sup>11</sup> water molecules measuring approximately 90 Å x 70 Å x 70 Å. The simulation cell was neutralized by adding an appropriate number of Na<sup>+</sup> and Cl<sup>-</sup> ions. The mutant structures were prepared by simply substituting the natural residue in the starting crystal structure. Following energy minimization using the conjugate gradient method, each simulation cell was equilibrated and thermalized for 600 picoseconds then maintained at 300K and 1 atm using Langevin dynamics with an integration timestep of 2 femtoseconds. The final structure obtained following 100 nanoseconds of free dynamics was used as the initial protein conformation for the pulling SMD simulations. Molecular dynamics simulations were carried out using the NAMD 2.11 code<sup>12</sup> with the Charmm36 force field.<sup>13-15</sup> All bonds involving hydrogen atoms were constrained using the SHAKE algorithm in NAMD.<sup>16</sup> Non-bonded interactions were calculated using a cut-off of 12 Å for both van der Waals and Coulomb forces.

#### Steered molecular dynamics (SMD)

SMD simulations were performed on CBM in both constant velocity and constant force modes. The computational protocols and parameters were kept the same as those used in the equilibrium simulations except for the size of the water box. In the pulling simulations, the box was made large enough to ensure that the protein is surrounded by a thick layer of water molecules throughout the complete unfolding (length 600 Å, width 60 Å, height 60 Å). In total, the water and protein system contained approximately 200,000 atoms. SMD simulations were carried out by anchoring the C atom of the C-terminus (fixed atom) and applying an external force to a virtual spring attached to the N atom of the N-terminus (SMD atom). Similarly in C-terminus simulations, the protein was stretched by anchoring the N atom on the N-terminus while pulling the C-atom on the C-terminus. The stiffness *k* of the spring and the constant pulling velocity (Fig. S28) were optimized to give the best balance between accuracy and computational cost. A spring constant k = 8 kcalmol<sup>-1</sup>Å<sup>-2</sup> and a pulling velocity of 0.05 Å/ps were used in all SMD simulations. To gather sufficient statistics, we performed at least two repeat simulations on each of the WT CBM3 and all mutant CBM3 proteins.

In the constant force molecular dynamics simulations, a linear function of the distance between the N and C terminus atoms (same atoms as were selected in SMD simulations) is added to the Hamiltonian of the system. If the magnitude of the force is sufficiently large, a relaxation trajectory from the native to the complete unfolded state can be observed. Constant forces of 200-450 pN were tested. At sub-300 pN forces, no unfolding was observed during the 50 ns simulation time scale.

To predict the rupture forces occurring along an alternative pulling pathways, the 11residue ybbR-tag with the sequence DSLEFIASKLA was inserted into native CBM between residues 84 and 85. Mechanical unfolding of the ybbR-tagged CBM proteins was simulated in two ways, which we labelled ybbR\_N and ybbR\_C. In the ybbR\_N simulations the N atom of the N-terminus was held fixed while residue Ser86 was pulled. For the ybbR\_C configuration, the protein was stretched by anchoring the C atom on the C-terminus while pulling Ser86. We also modelled the effect of ybbR-tag insertion in two point mutants, D46P and D126G.

To further probe the origin of the computed MD rupture forces, we extracted also binding energy estimates from the equilibrium MD simulations. The MM/PBSA method which combines three energy terms to account for a change in energy upon binding (Eqn 1), was used to compute the binding free energies of calcium ion and residues in its binding sphere, using GROMACS tool g\_mmpbsa.<sup>17</sup> The first term corresponds to standard MM energy terms from bonded (bond, angle and dihedral), electrostatic and van der Waals interactions. The second term takes into account the free energy of solvation of the different species and the third term accounts for the entropy associated with the complex. The solvent is treated as a homogenous dielectric medium with dielectric constant of bulk water (80) while the protein and calcium ion are treated as another (buried) dielectric medium with a dielectric constant of 4.

$$\Delta G_{bind} = \Delta E_{gas} + \Delta G_{solv} - T\Delta S \tag{1}$$

In order to implement g\_mmpbsa, 50 ns MD calculations were performed using the GROMACS 5.1 code<sup>18</sup> with the CHARMM36 protein force field<sup>13-15, 19, 20</sup> and TIP3P water model.<sup>11</sup> The Lennard-Jones interactions were truncated at 1.0 nm and the particle mesh Ewald method<sup>21</sup> was used to calculate electrostatic interactions with a real space cut-off at 1.2 nm. The length of each covalent bond to hydrogen was constrained using the LINCS algorithm<sup>22</sup> which allowed an integration timestep of 2 femtoseconds. Simulations were carried out at 300 K to simulate room temperature using the Velocity-rescale thermostat.<sup>23</sup> The reference pressure was set to 1 bar and compressibility to 4.5 x 10-5 per bar using the Parrinello-Rahman algorithm.<sup>24</sup> The coupling to the isotropic pressure reservoir was updated every 5 ps.

In followup studies of the mechanical stability of CBM3 protein, it would be beneficial to compute precise free energies of the unfolding intermediates identified in this work, using Jarzynski's equality (also known as the nonequilibrium work relation) to directly compare potentials of mean force (PMF) for the native protein and promising mutants.

#### **Experimental methods**

#### Expression and purification of proteins for single-molecule force spectroscopy

The two proteins used here are denoted ybbR-HIS-Top7-XDoc and Coh-CBM-HIS-ybbR. Here, ybbR denotes the ybbR peptide tag, HIS denotes a 6x histidine tag, Coh and XDoc refer to the Cohesin and Dockerin derived from *Ruminococcus flavefaciens* which form a stable complex<sup>25</sup>, Top7 is a synthetic domain<sup>26</sup>, and CBM refers to the wild type CBM studied in this work.

Cloning was performed using standard methods. The pET28a expression plasmids were transformed into *E. coli* BL21(*DE3*) cells using chemical transformation and selected on LB plates supplemented with 50  $\mu$ g/mL kanamycin (LB<sup>kan</sup>). One single colony was inoculated into 5 mL LB<sup>kan</sup> and the culture was grown at 37 °C overnight with shaking. On the following day, 2 mL of this preculture was inoculated into 200 mL LB<sup>kan</sup> and the culture was grown at 37 °C with shaking until the OD<sub>600</sub> reached 0.5. Expression was then induced with 0.5 mM IPTG. The expression was performed at 25°C for 6 hrs while shaking. The cells were

harvested by centrifugation at 4000g for 10 min and re-suspended in 20 mL lysis buffer (50 mM Tris, 50 mM NaCl pH 8.0 containing 2 mg lysozyme and 40 Units of *DNase*I), and incubated on ice for 15 min before sonication (35% amplitude, 2 second pulse on, 2 second pulse off) for 10 min. The cell lysate was then centrifuged at 18,000 g (4 °C) for 30 min. The clear supernatant was loaded onto a 5 mL His-Trap FF column (pre-equilibrated with PBS) using an AKTA Pure system. The column was washed with 30 column volumes (CVs) of PBS followed by 5 CVs of PBS supplemented with 25 mM imidazole. After that, the His<sub>6</sub>-tagged proteins were eluted from the column with 3 CVs of PBS supplemented with 500 mM imidazole. The eluates were dialyzed 3 times against PBS and concentrated to ~9.4 mg/ml (224  $\mu$ M for ybbR-HIS-Top7-XDoc or 202  $\mu$ M for Coh-CBM\_HIS-ybbR). Purity was analyzed on 12% SDS-PAGE gel and concentration was determined by using BCA assay.

### **AFM sample preparation**

ybbR-HIS-Top7-XDocIII and CohIII-CBM-HIS-ybbR were immobilized on AFM cantilevers and coverglasses respectively, using the ybbR-tagged surface chemistry according to previously published protocol<sup>25</sup>. In brief, levers and coverglasses were first cleaned by UV-ozone treatment and piranha solution, respectively. 3-Aminopropyl (diethoxy)methylsilane (APDMES, ABCR GmbH, Karlsruhe, Germany) was used to modify the surfaces of the cantilevers and coverglasses with amine groups. The amine groups were subsequently conjugated to a 5 kDa NHS-PEG-Mal linker (Rapp Polymere, Tübingen, Germany) in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5. PEGylated levers or coverglasses were next incubated with Coenzyme A (CoA, 20 mM) in buffer comprising 50mM sodium phosphate, 50mM NaCl, 10mM EDTA, pH 7.2 for 2 h at room temperature. SFP-catalyzed coupling of the fingerprint fusion proteins (ybbR-HIS-Top7-XDocIII and CohIII-CBM-HIS-ybbR) to the CoA levers or surfaces was done in Ca-TBS (25 mM Tris, 72 mM NaCl, 1mM CaCl2, pH7.2) at room temperature for 2 h.

#### Single-molecule force spectroscopy measurement

To characterize the unfolding of wild-type CBM, SMFS measurements were carried out using automated AFM-based SMFS on a Force Robot 300 system from JPK Instruments. The measurement buffer was Ca-TBS (25 mM Tris, 72 mM NaCl, 1mM CaCl2, pH7.2), at room temperature. The AFM-SMFS measurement was carried out in constant speed mode with a cantilever retraction velocity of 500 nm.s<sup>-1</sup>. Several thousand force-distance curves were obtained and analyzed using previously published protocols.<sup>25</sup> Curves were selected for analysis using a filtering algorithm that searched for contour length increments that matched the lengths of the specific protein fingerprint domains: Top7(~29 nm) and CBM (~56 nm). The histogram of unfolding forces of the CBM fingerprint was fitted with a Bell-Evans distribution to determine the values of the most probable unfolding force (<F>), effective distance to the transition state ( $\Delta x$ ), and the natural off-rate ( $k_{off}$ ), which yielded values of <F>= 143.8 pN,  $\Delta x$ =0.343 nm and  $k_{off}$  = 1.769 × 10<sup>-6</sup> s<sup>-1</sup> for wild-type CBM. A line fit to the force vs. time trace in the immediate vicinity before the CBM unfolding event was used to estimate the loading rate, which for the reported unfolding histogram was found to be 3.525 ± 0.986 nN s<sup>-1</sup>.

#### **Coarse-grained methods**

#### CG models

Simplified molecular dynamics simulations of protein stretching within a coarse-grained structure-based (equivalently, Go-like<sup>4</sup>) model<sup>5-8</sup> were performed. Instead of full all-atom resolution, only C $\alpha$  atoms were taken into account for the dynamics and their interactions were governed by a contact map. The contact map constitutes the list of amino acids for which heavy atoms, represented by enlarged van der Waals spheres, overlap in the native conformation.<sup>27</sup> Such pairs are called native contacts. The interaction between two C $\alpha$  atoms creating the native contact was described by the Lennard-Jones potential in the form

$$V(r_{ij}) = 4\varepsilon \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right)$$
(1)

where  $r_{ii}$  is the actual distance between atoms i and j. The parameter  $\sigma_{ii}$  was chosen to obtain a minimum of the structure's potential energy at its native state. The value of the binding energy parameter  $\varepsilon \sim 110 \text{ pN} \cdot \text{Å}^{-1}$ , which is also close to the energy of the O–H–N hydrogen bond of 1.65 kcal/mol, was determined to agree with protein stretching experiments.<sup>28</sup> Implicit solvent was used and the presence of the water molecules was mimicked by Langevin noise and damping terms.<sup>29</sup> The interactions between all of the other amino acids were purely repulsive and were described by the truncated Lennard-Jones potential. The interactions along the backbone were described by a harmonic potential with a spring constant of 50  $\epsilon \cdot A^{-2}$ . The characteristic time-scale of the simulations was of the order of 1 ns. The stretching of the protein was performed at  $T = 0.3\epsilon/kB$  (kB is the Boltzmann constant).<sup>30</sup> The protein termini were each attached to a spring with the spring constant equal to 0.06 ɛ/Å.<sup>6, 7</sup> One of the springs was anchored and another one was pulled with a constant velocity of 0.005 Å. $\tau^{-1}$  and simulations were also done at a different velocity, v = 0.00005 Å  $\tau^{-1}$ . For each velocity at least 10 trajectories were recorded. The particular native contacts in CBM3 were considered broken when the extension length exceeded 1.5  $\sigma_{ij}$ . The maximum force peak was measured from the force-displacement graphs.

# References

- 1. W. Humphrey, A. Dalke and K. Schulten, *J Mol Graph Model*, 1996, **14**, 33-38.
- 2. T. Verdorfer, R. C. Bernardi, A. Meinhold, W. Ott, Z. Luthey-Schulten, M. A. Nash and H. E. Gaub, *J Am Chem Soc*, 2017, DOI: 10.1021/jacs.7b07574.
- 3. E. Evans and K. Ritchie, *Biophysical Journal*, 1997, **72**, 1541-1555.
- 4. N. Go, Annual review of biophysics and bioengineering, 1983, **12**, 183-210.
- 5. M. Cieplak and T. X. Hoang, *Biophysical Journal*, 2003, **84**, 475-488.
- 6. M. Cieplak, T. X. Hoang and M. O. Robbins, *Proteins: Structure, Function, and Bioinformatics*, 2004, **56**, 285-297.

- 7. I. S. J and C. Marek, *Journal of Physics: Condensed Matter*, 2007, **19**, 283201.
- 8. J. I. Sułkowska and M. Cieplak, *Biophysical Journal*, 2008, **95**, 3174-3191.
- 9. J. Tormo, R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham and T. A. Steitz, *Embo J.*, 1996, **15**, 5739-5751.
- 10. R. Rohs, C. Etchebest and R. Lavery, *Biophys J*, 1999, **76**, 2760-2768.
- 11. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *Journal of Chemical Physics*, 1983, **79**, 926-935.
- 12. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781-1802.
- 13. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, *J Chem Theory Comput*, 2012, **8**, 3257-3273.
- 14. A. D. MacKerell, M. Feig and C. L. Brooks, *Journal of the American Chemical Society*, 2004, **126**, 698-699.
- A. D. MacKerell, D. Bashford, M. 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. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J Phys Chem B*, 1998, **102**, 3586-3616.
- 16. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.
- 17. R. Kumari, R. Kumar and A. Lynn, *Journal of Chemical Information and Modeling*, 2014, **54**, 1951-1962.
- 18. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.
- 19. J. Huang and A. D. MacKerell, *Journal of Computational Chemistry*, 2013, **34**, 2135-2145.
- 20. O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell, *Journal of Computational Chemistry*, 2008, **29**, 2543-2564.
- 21. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577-8593.
- 22. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *Journal of Computational Chemistry*, 1997, **18**, 1463-1472.
- 23. G. Bussi, D. Donadio and M. Parrinello, *The Journal of chemical physics*, 2007, **126**, 014101.
- 24. M. Parrinello and A. Rahman, *Journal of Applied Physics*, 1981, **52**, 7182-7190.
- 25. C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash and H. E. Gaub, *Nano letters*, 2015, **15**, 7370-7376.
- 26. B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker, *Science (New York, N.Y.)*, 2003, **302**, 1364-1368.
- 27. J. Tsai, R. Taylor, C. Chothia and M. Gerstein, *Journal of molecular biology*, 1999, **290**, 253-266.
- 28. M. Sikora, J. I. Sułkowska and M. Cieplak, *PLOS Computational Biology*, 2009, **5**, e1000547.
- 29. G. S. Grest and K. Kremer, *Physical Review A*, 1986, **33**, 3628-3631.
- 30. P. Szymczak and M. Cieplak, *J Chem Phys*, 2006, **125**, 164903.