

Acoustic Force Spectroscopy Reveals Subtle Differences in Cellulose Unbinding Behavior of Carbohydrate-Binding Modules

Markus Hackl^a, Edward V. Contrada^a, Jonathan E. Ash^a, Atharv Kulkarni^a, Jinho Yoon^b, Hyeon-Yeol Cho^b, Ki-Bum Lee^b, John M. Yarbrough^c, Cesar A. López^d, Sandrasegaram Gnanakaran^d Shishir P. S. Chundawat^{a,*}

^a Department of Chemical and Biochemical Engineering, and ^b Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway NJ, 08854, ^c Biosciences Center, National Renewable Energy Laboratory, Golden, CO, 80401; ^d Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545

*Corresponding author: Shishir P. S. Chundawat (ORCID: 0000-0003-3677-6735). Department of Chemical & Biochemical Engineering, Rutgers, The State University of New Jersey, 98 Brett Road Piscataway, NJ 08854. Phone: +1-848-445-3678

Email: shishir.chundawat@rutgers.edu

Author Contributions: M.H., S.G. and S.P.S.C. designed research; M.H., E.C., J.E.A, A.K., J.Y., H.C., K.L., J.M.Y., and C.A.L. performed experiments; M.H., E.C., J.E.A, and C.A.L. analyzed the data; M.H. and S.P.S.C wrote the manuscript with inputs from all authors.

Competing Interest Statement: No competing interests.

Preprint server: Original version of manuscript was deposited as preprint on Biorxiv.org under <https://doi.org/10.1101/2021.09.20.461102>

Classification: Biological Sciences, Biophysics and Computational Biology.

Keywords: Carbohydrate-Binding Module, Nanocellulose, Biofuels, Single-Molecule Force Spectroscopy, Acoustic Force Spectroscopy

This PDF file includes:

Main Text
Figures 1 to 7
Table 1

Abstract

Protein adsorption to solid carbohydrate interfaces is critical to many biological processes, particularly in biomass deconstruction. To engineer more efficient enzymes for biomass deconstruction into sugars, it is necessary to characterize the complex protein-carbohydrate interfacial interactions. A carbohydrate-binding module (CBM) is often associated with microbial surface-tethered cellulosomes or secreted cellulase enzymes to enhance substrate accessibility. However, it is not well known how CBMs recognize, bind, and dissociate from polysaccharides to facilitate efficient cellulolytic activity due to the lack of mechanistic understanding and a suitable toolkit to study CBM-substrate interactions. Our work outlines a general approach to study the unbinding behavior of CBMs from polysaccharide surfaces using a highly multiplexed single-molecule force spectroscopy assay. Here, we apply acoustic force spectroscopy (AFS) to probe a *Clostridium thermocellum* cellulosomal scaffoldin protein (CBM3a) and measure its dissociation from nanocellulose surfaces at physiologically relevant, low force loading rates. An automated microfluidic setup and method for uniform deposition of insoluble polysaccharides on the AFS chip surfaces are demonstrated. The rupture forces of wild-type CBM3a, and its Y67A mutant, unbinding from nanocellulose surfaces suggests distinct multimodal CBM binding conformations with structural mechanisms further explored using molecular dynamics simulations. Applying classical dynamic force spectroscopy theory, the single-molecule unbinding rate at zero force is extrapolated and found to agree with bulk equilibrium unbinding rates estimated independently using quartz crystal microbalance with dissipation monitoring. However, our results also highlight critical limitations of applying classical theory to explain the highly multivalent binding interactions for cellulose-CBM bond rupture forces exceeding 15 piconewtons.

Significance Statement

Cellulases are multi-modular enzymes produced by microbes that catalyze cellulose hydrolysis into glucose. These enzymes play an important role in global carbon cycling as well as cellulosic biofuel production. CBMs are essential components of cellulolytic enzymes involved in facilitating the hydrolysis of polysaccharides by a tethered catalytic domain (CD). The subtle interplay between CBM binding and CD activity is poorly understood, particularly for heterogeneous reactions at solid-liquid interfaces. Here, we report a highly multiplexed single-molecule force spectroscopy method to study CBM dissociation from cellulose to infer the molecular mechanism governing substrate recognition and dissociation. This approach can be broadly applied to study multivalent protein-polysaccharide binding interactions relevant to other carbohydrates, such as starch, chitin, or hyaluronan, to engineer efficient biocatalysts.

Main Text

Introduction

Carbohydrate-based biopolymers are abundant throughout all forms of life and play a major part in biomolecular recognition processes that have fundamental scientific and applied technological relevance. Protein adsorption to simple or complex carbohydrates at solid interfaces is a critical step in biological processes such as plant growth (1, 2), host infection (3–5), or biofuels production (6, 7). Although carbohydrates display structural diversity, many CBMs that bind to diverse carbohydrates often display structurally similar binding site architecture (8). Therefore, insight into a particular CBM-carbohydrate system would help unravel general principles of protein-carbohydrate binding. Carbohydrate-Active enZymes (CAZymes) are a well-studied domain and suitable model system to further investigate molecular-level interactions. CAZymes, such as processive cellulases often consist of two or more domains called carbohydrate-binding module and catalytic domain, which are responsible for the recognition/binding and breakdown of the substrate, respectively (9). On the other hand, cellulosomes are larger multidomain enzymes where CDs are assembled on a scaffolding domain decorated with CBMs and specific linker domains as shown in **Figure 1-A** (10). Cellulosomes adapt to the substrate topology and display a “sit-and-dig” mechanism where the cellulosome degrades individual cellulose crystals without dissociating from the substrate (11–13). This mode of action contrasts processive cellulases, such as *Trichoderma reesei* Cel7A, which displays a “slide-and-peel” mechanism and frequently dissociates from the substrate (9, 14–16).

Carbohydrate binding modules can be grouped into type-A, B, or C categories based on relevant structural-functional relationships. Both *TrCel7A* and the cellulosome from *Clostridium thermocellum* possess a type-A CBM with a similar overarching architecture of the cellulose binding site. Type-A CBMs preferably bind to insoluble and highly crystalline cellulose, forming a planar flat, platform-like binding surface mostly lined with aromatic residues, complementary to the flat planar structure of the crystalline substrate (17). As such, CBM1 from *TrCel7A* exhibits 3 tyrosine residues at the 5, 31, and 32 positions (18), whereas the 3 aromatic residues on the binding surface of *CtCBM3a* are H57, Y67, and W118 respectively (19) as shown in **Figure 1-B**. Although mutations of the aromatic residues of CBM3a to alanine can reduce the apparent bulk-ensemble binding affinity to native crystalline cellulose, the enzymatic activity of endocellulases fused to those mutants increased by 20-70% compared to the wild type (20). Altering enzyme binding affinity to cellulosic substrates is being explored as a strategy to engineer more efficient cellulases (21, 22). However, engineering highly active cellulases, cellulosomes, and associated cellulolytic microbes still present challenges due to the inadequate understanding of the complex interplay between CD and CBM as well as the multivalent nature of the CBM-carbohydrate interactions.

Traditionally, CBM and cellulase adsorption is characterized by bulk ensemble-based methods such as solid-phase depletion (23, 24), quartz crystal microbalance with dissipation (QCM-D) (25), and isothermal titration calorimetry (ITC) (26, 27). However, these methods rely on simplified models to illustrate binding interactions that do not reflect the underlying molecular mechanism of protein binding to highly multivalent carbohydrate ligands such as cellulose. Techniques like single-molecule fluorescence (16, 28) and force spectroscopy (29) have greatly contributed to our understanding of molecular processes relevant to cellulose degradation. In particular, atomic force microscopy (AFM) has been used previously to characterize CBM desorption from cellulose on the single-molecule level (30, 31). Examples include the identification of binding sites (32), distinguishing specific from non-specific binding (33), and determining the zero-force unbinding rate using Bell’s model (34). Although AFM measures force and distance with pN and nm resolution, the determination of unbinding forces occurs far from equilibrium due to the relatively high loading rates inherent to the conventional AFM technique, potentially obscuring multimodal unbinding behavior seen at physiologically relevant conditions. Alternatively, optical tweezers (OT) have been used to study CBM unbinding at lower loading rates and force clamp mode (35), and

the results suggest a complex unbinding behavior where the bond lifetime data does not follow a single exponential decay function as suggested by AFM studies (34).

In contrast to AFM, acoustic force spectroscopy (AFS) is a new technique that enables the application of low loading rates comparable to optical tweezers while maintaining a higher throughput during single-molecule force spectroscopy (SMFS) assays (36, 37). Similar to OT, the protein of interest is attached to a micrometer-sized bead *via* a double-stranded DNA tether. However, forces on the bead are exerted by acoustic standing waves. Streptavidin-coated beads are commonly used to connect biotinylated DNA tethers (38) due to their high specificity and binding strength (39, 40). On the other end of the tether, the protein of interest is either covalently linked through a thiol-maleimide crosslink (29) or tethered non-covalently *via* the histidine tag to Anti-his antibodies, which in turn are covalently linked to aminated DNA tethers (35). The histidine tag of proteins was used previously to non-covalently link to DNA (41–44) and directly attach proteins to NTA-modified AFM tips (45). It was demonstrated that the His-Ni-NTA bond is stable enough (~120pN at 400pN/s) to facilitate single-molecule force spectroscopy (SMFS) experiments of the tethered protein (46–49) thus allowing AFM-based studies to measure unbinding forces of CBM3a using Ni-NTA (31, 33).

Here, we combine the tethering methods by directly synthesizing a linear dsDNA tether with biotin on one end to attach it to a micron-sized bead, and Ni-NTA on the other end to tether any His-tagged protein. This setup allows SMFS of the His-tagged protein with its polysaccharide ligand deposited onto the AFS microfluidic channel surface to enable high-throughput assays. Such tethers can be used in other tethered bead setups such as optical or magnetic tweezers, highlighting the modularity and versatility of our approach. Furthermore, digoxigenin (DIG) tethers instead of NTA were generated with the same procedure to validate the bead preparation method and analysis of the recorded position and force-distance traces using AFS. A schematic of the protein-ligand systems studied to measure bond rupture forces is shown in **Figures 1- C and D**. Furthermore, an automated method for depositing nanocrystalline cellulose (NCC) inside the AFS chip was developed. The unbinding forces of *C. thermocellum* CBM3a-wt (WT) and its Y67A mutant were measured at fixed, low loading rates. The unbinding forces of the wild type have been previously characterized by AFM (31–34). Furthermore, it was shown previously that the Y67A mutation reduces CBM binding affinity by several orders of magnitude while improving tethered CD activity for reasons not clear (20). The unbinding behavior of the wild-type and mutant CBM3a using SMFS measured at physiologically relevant conditions has not yet been reported. We identified a clear difference in the rupture force distribution pattern observed between WT and Y67A mutant at low loading rates that are often not accessible by alternative SMFS methods like AFM. While the extracted unbinding rate (k_{off}) from our AFS results agrees with bulk ensemble QCM-D results, the classical SMFS model is unable to accurately capture the multivalent protein-polysaccharide binding interactions particularly at higher rupture forces. Lastly, molecular dynamics simulations were performed to provide detailed atomistic structural insights into how a single mutation in the planar binding motif of CBM3a can severely disrupt its multimodal interactions with nanocellulose that were readily probed using our AFS technique.

Results

Deposition and characterization of nanocellulose inside the AFS chip. Sulfuric acid-derived nanocrystalline cellulose (NCC) was used to generate the cellulose model film in this study. The formation of an NCC film inside the AFS chip was accomplished by a multilayer deposition process (50) where poly-L-Lysine (PLL) and NCC were alternately deposited using an automated microfluidic control system. **Figures 2-A and B** show the flowchart and process flow diagram of the process, and a detailed description can be found in the Methods section. Green fluorescent protein (GFP) tagged CBM3a was expressed as described previously (20) and used to characterize the cellulose film deposited on the AFS chip. **Figure 2-C** shows a representative fluorescence image of the NCC-modified AFS chip labeled with GFP-CBM3a. The arrow indicates an area where a

bubble was stuck during the NCC deposition process. Slightly lower amounts of NCC were deposited in that area, resulting in lower fluorescence. The rest of the flow channel displays a uniform fluorescence, indicating that NCC is deposited evenly across the channel. The average fluorescence intensity of the bare glass and PLL treated chips surfaces is 14 ± 1 a.u. (mean $\pm$ SEM) and 12 ± 3 a.u., respectively, whereas the NCC treated chips show a fluorescence intensity of 136 ± 35 a.u. The deposition of a single layer of NCC onto a PLL-treated surface resulted in a fluorescence intensity of 42 ± 24 a.u. Despite significantly higher fluorescence compared to controls, such prepared AFS chips failed to reproducibly provide a consistent response at the single-molecule level even though AFM imaging confirmed the deposition of a uniform layer of NCC (**SI Appendix Fig. S1**). A relationship between the success of a single-molecule experiment and the measured fluorescence intensity was observed, where the likelihood of a successful single-molecule experiments positively correlated with the measured fluorescence. Hence, a multilayer NCC deposition method was used to ensure a consistently high fluorescence signal, which in turn resulted in a reliable rupture force measurement of CBMs. Multilayer NCC-functionalized AFS chips, which were subsequently cleaned and imaged as outlined in the Methods section, showed a fluorescence intensity of 13 ± 0.5 , indicating the removal of nanocrystalline cellulose for reuse of the AFS chips for multiple rounds of experimentation. **Figure 2-D** shows an example surface imaged by AFM, additional AFM images of bare and PLL treated surfaces are found in **SI Appendix Fig. S1**. Similar to spin-coated samples (50, 51), the surface was uniformly covered with nanocrystalline cellulose. AFM image analysis revealed the formation of NCC crystal aggregates at multilayered films. This is reflected by a surface roughness factor (R_a) that is marginally greater than 3 nm compared to less than 2 nm estimated for a single NCC layer.

Observation of shortened DNA tethers on NCC surfaces. The tether preparation method and analysis of traces as outlined in the Methods section, were validated by tethering beads anchored to the AFS chip surface by anti-digoxigenin antibodies (aDIG). The dimensionless contour length (l_{fc}) of DNA tethers bound to aDIG during force calibration was 1.1 ± 0.12 (mean $\pm$ SD, $N=156$). This is in the expected range, given the particle size distribution of the beads. The average rupture force of the DIG-aDIG complex was determined to be 18.8 ± 7.0 pN at a loading rate of 0.14 ± 0.05 pN/s (**SI Appendix Fig. S2**) and is close to the reported value of 16.6 pN at 0.11 pN/s (36). As shown in **Figure 3-A**, overstretching of the DNA tether was observed at ~ 65 pN, thus confirming the formation of single tethers with the bead preparation method outlined in the Methods section. In contrast, the observed dimensionless tether length of DNA for NCC-CBM tethered beads was only 0.83 ± 0.23 , indicating a shortening of the tethers by $\sim 25\%$. However, the force-distance (FD) curves obtained during the linear force ramp follow the extensible worm-like chain (52) or WLC model (**Figure 3-B**), indicating that the tethers are only shortened but not otherwise altered. **Figures 3-C** and **D** show the scatter plots of the rupture force with l_{fc} for DIG-aDIG and NCC-CBM3a-wt at 1 pN/s respectively. The best linear fit (red line) is added as a guide. Additional scatter plots of root-mean-square fluctuation (RMS) and symmetrical motion (Sym) as well as the Pearson and Spearman correlation coefficients can be found in the **SI Appendix Fig. S3** and **Table S1** and **Table S2**. Except for the Pearson coefficient for Sym and rupture force of Y67A at 0.1 pN/s ($p=0.043$), no significant correlation ($p<0.05$) was identified between the measured rupture force and observed length as well as RMS and Sym . We hypothesize that the 1.8 μ m long and flexible DNA tether wraps around and/or binds to exposed and weakly bound NCC crystals during the incubation step or that NCC somehow binds to the DNA, which shortens the apparent contour length. The attached NCC crystals are subsequently detached from the NCC surface when the bead is being pulled away from the surface during force calibration but stay bound to the DNA. Non-specifically tethered beads were observed in control experiments with blank Ni-NTA and GFP tagged beads. However, the number of tethered beads was higher by at least 4x for CBM tethered beads. A schematic describing non-specifically tethered beads can be found in **SI Appendix Fig. S4**. The loss of tethered beads during the flushing step before bead tracking was noted in all cases but was significantly larger for non-CBM tethered beads further indicating weaker, non-specific binding interactions of the DNA to NCC. The rupture force distribution of only tethers close to the expected length and the entire expected single-molecule tethers are identical as it can be seen in

SI Appendix Fig. S5, implying that a single CBM-NCC rupture event was measured even though a shortened tether was observed. Assuming that a single CBM was tethered when the FD curve follows the WLC model, the force calibration and rupture force determination were not affected by the shortening of DNA, and the data were included in all further analyses.

Rupture force analysis and application of the Dudko-Hummer-Szabo (DHS) model. The rupture force distributions for CBM3a-wt and its Y67A mutant at a loading rate of 1 pN/s and 0.1 pN/s are shown in **Figure 4**. The histogram bin width was chosen based on the Freedman-Diaconis rule (53) since the data deviates from a single normal distribution. To capture the apparent multimodal distribution, a double normal distribution was fit to the histogram. The means and standard deviations are summarized in **SI Appendix Table S3**. Although the first mean is similar for wild type (8.5 pN) and Y67A (7.9 pN) at 1 pN/s, there is a clear single rupture force peak observed for the Y67A mutant, but not for the wild type. This difference is even more pronounced when comparing the rupture force distributions at 0.1 pN/s. Two distinct rupture force peaks were observed for the wild type at 3.5 pN and 7.1 pN, respectively, whereas Y67A showed only one peak at 4.5 pN. All histograms show a “tail” towards larger rupture forces, which is defined by the second normal fit. At 1 pN/s, CBM3a-wt shows a distinct peak at 17.5 pN, followed by a long tail up to 35 pN, whereas no clear second peak but only a tail until 25 pN was observed for Y67A.

Figures 5-A and B show the transformation of rupture force histograms to force-dependent bond lifetime data using Equation 1 (circles) and the fit of Equation 2 (solid lines) of the Dudko-Hummer-Szabo (DHS) model (54, 55) described in the Methods section, for WT and Y67A respectively. Data from rupture force histograms obtained at different loading rates should fall on the same master curve for force-dependent bond lifetimes as predicted by Equation 2 if the unbinding kinetics at constant force follow a single-exponential function (55). Although there is some overlap of bond lifetimes obtained at 0.1 pN/s and 1 pN/s for both WT and Y67A, the fit of Equation 2 inadequately describes the data for both shape factors ν . A similar observation of bond lifetime data not following classical models was made recently for another type-A CBM1 and its Y31A mutant using optical tweezers (35), although the force-dependent bond lifetime was obtained in force-clamp mode. Surprisingly, no significant difference in the force-dependent bond lifetime was found between CBM1 and its Y31A mutant for most rupture forces. **Figures 5-C to F** show the rupture force histograms of WT and Y67A and the predicted probability density according to Equation 3. Both shape factors produce a qualitatively similar probability distribution but insufficiently replicate the measured rupture forces. The main reason for discrepancies in bond lifetime and rupture probability distribution fit is the shape of the underlying rupture force histogram. Both the WT and Y67A rupture force histograms show tailing towards higher rupture forces with no clear peak, which results in almost force-independent bond lifetimes at higher rupture forces. The multimodal distribution observed for WT at both loading rates results in bond lifetime data not exactly following a single exponential decay function. As shown in **Table 1**, only $\nu=2/3$ yields unity for the numerical approximation of $\int p(f)df$ over the modelled force range, despite qualitatively similar fits of the bond lifetime data and probability density for both shape factors. The extrapolated unbinding rates (k_{off}) at zero force and $\nu=2/3$ for the WT is 0.0091s^{-1} and approximately twice as high as the k_{off} for Y67A at 0.0044s^{-1} . The unbinding rate of CBM3a-wt from sulfuric acid-derived microfibrils isolated from poplar and extracted from AFM-SMFS rupture force data using the Bell’s model was estimated to be 0.0089s^{-1} (34) and is close to the value obtained in our study.

Table 1 summarizes the fit parameters from Equation 2 as well as the numerical approximation of $\int p(f)df$ for both loading rates. The transition state distance, $x^\ddagger$, is 0.88 nm and 1.12 nm for WT and Y67A, respectively, and agrees with a transition state distance based on Bell’s model of 0.63 nm for CBM3a-wt (34). The apparent free energy of activation, $\Delta G^\ddagger$, is $5.4 k_B T$ and $8.1 k_B T$ for CBM3a-wt and Y67A respectively, and contrasts $45.3 k_B T$ previously reported (34). Both $x^\ddagger$ and $\Delta G^\ddagger$ are similar for the wild type and mutant, indicating a similar unbinding pathway. The Bell and DHS models assume a one-dimensional unbinding pathway, which may not represent the underlying molecular interactions based on the multimodal rupture force distributions measured in

this study, as well as evidence of different CBM binding orientations to crystalline cellulose that give rise to multiple non-equivalent binding sites (35, 56).

Bulk ensemble CBM3a-nanocellulose off-rate qualitatively agrees with AFS results. QCM-D experiments using hydrochloric acid-derived NCC as substrate reported a 1.4-fold increase in the off-rate for the Y67A mutant compared to the WT (20). However, using sulfuric acid-derived NCC, our QCM-D analysis using a classical one-site binding site adsorption model yielded a k_{off} of $26.8 \pm 2.4 \times 10^{-5} \text{ s}^{-1}$ (mean $\pm$ SD, $n=2$) and $19.7 \pm 1.2 \times 10^{-5} \text{ s}^{-1}$ for WT and Y67A, respectively. This result supports the findings from AFS experiments that the WT unbinds more frequently, although the absolute values differ between AFS and QCM-D. In contrast, the number of available binding sites determined by QCM-D reduced from $306 \pm 41 \times 10^{12}$ molecules to $177 \pm 43 \times 10^{12}$ molecules between WT and Y67A, respectively.

Molecular simulations reveal stabilizing role of Y67 on CBM3a binding mechanism. All-atom molecular dynamics (MD) simulations were performed with CBM3a-wt and Y67A mutant aligned on a cellulose-I crystal such that the vector formed between W118 and Y67 points to the reducing end of the glucan chain (**Figure 6-A**). An equilibrated configuration is provided in **Figure 6-B** highlighting the amino acids of the CBM3a-wt planar binding motif with larger probability of interaction with cellulose (using a cutoff of 0.35 nm). In addition to the preserved $CH - \pi$ forces arising from residues W118, H57, and Y67, several H-bonds are formed between adjacent polar residues and the substrate, as identified previously (19). **Figure 6-C** summarizes the time-averaged H-bond formation between residues and the cellulose surface for wild type and mutant. Changes in H-bond formation are observed with CBM residues S9, N10, N16 and D56, which is indicative of the alternation of the H-bond network due to the Y67A mutation.

Further analysis suggests that the observed alteration of the H-bond network between wild type and mutant could be associated with a change in the orientation of the CBM with respect to cellulose. To examine the overall association of the CBM to cellulose, we computed the orientation of the CBM with respect to the cellulose surface. We defined an orientational angle (θ) as the vector formed between alpha carbons of residues W118 and Y67 and the normal vector of the cellulose plane (**Figure 6-A**). Notably, the average orientational angle for the WT is close to an orthogonal configuration ($93.0 \pm 2.8^\circ$, mean $\pm$ SD) indicative that W118 and Y67 are perfectly aligned with respect to the water-cellulose interface. However, the Y67A mutation leads (on average) to an imperfect alignment ($\theta = 100.6 \pm 3.3^\circ$), with episodes of spontaneous partial detachment, indicated by angles as large as 114° (**SI Appendix Fig. S7**).

Interestingly, the Y67A mutation not only affects the association with cellulose, but also leads to intramolecular rearrangement of residues within the CBM binding motif. Specific intramolecular H-bonds in the aromatic binding motif of CBM3a are summarized in **Figure 6-D**. The Y67 mutation results in the total disruption of the H-bond between residues 57 and 67, enabling the imidazole group of H57 to flip $\sim 180^\circ$ with respect to the surface more frequently (**SI Appendix Fig. S6**). This alternative configuration decreases the formation of H-bonds between H57 and Q110, contrasting with the observed behavior seen in the wild type. The H-bond formation of Q110 with the substrate remains similar for both proteins, indicating that the Y67 mutation does not have an impact on this residue. The overall loss in hydrogen bonding involving H57 for the mutant, likely facilitates neighboring residue D56 to play a more prominent role, both in terms of its H-bonding to cellulose as well as its intramolecular H-bonding with R112. Thus, MD simulations can capture the intricate rebalancing of intra-molecular interactions within the CBM binding motif, due to the substitution at Y67 position, which also correlates with altered mutant CBM binding interactions with cellulose.

Discussion

We established a layer-by-layer deposition method for immobilizing nanocrystalline cellulose onto microfluidic chip surfaces and determined single-molecule CBM-cellulose unbinding forces at varying loading rates using AFS. Any soluble or insoluble polysaccharide substrate that can be

spin-coated on glass surfaces and is small enough not to clog the flow channel, can be readily immobilized within the AFS microfluidic chip using our proposed approach. Examples include the immobilization of regenerated cellulose, cellulose microfibrils, or chitin nanocrystals (30, 57, 58). Cellulose nanocrystals offer an especially promising platform for further chemical modifications (59, 60) either pre- or post- immobilization to fine-tune protein adsorption (61), allowing the application of SMFS to a wider range of applications. Furthermore, a robust method for preparing tethered beads based on the well-known biotin-streptavidin and His-Ni-NTA interactions is presented here. Histidine tags are widely used to purify heterologously expressed proteins. Therefore, our proposed one-step tether synthesis *via* PCR with biotin and NTA modified primers is a convenient method to characterize most heterologous proteins for SMFS without further modifications.

Both, the CBM3a-wt and Y67A mutant are fused to GFP on their N-termini, which is located on the opposite side of the aromatic binding motif. The force loading geometry (pulling from C- vs. N-terminus) can influence the measured rupture forces. For example, the biotin-streptavidin unbinding forces are reduced ~50% when pulling from the N-terminus compared to the C-terminus due to the partial unfolding of the N-terminus (62). In the same study, it was shown that the attachment configuration influences the width of the rupture force distribution due to non-specifically bound streptavidin. Although the unfolding force of CBM3a was not determined at the loading rates used in our experiments, it is unlikely that the CBM unfolds at forces <30 pN given its high mechanical stability (63). Even if partial unfolding would occur, it is expected that the protein should detach at lower forces, which was not observed in our study. In our tether attachment configuration, the (non-specific) binding probability due to DNA or GFP binding to NCC is low compared to CBM tethers. Double-stranded DNA only weakly binds to cellulose (64) in solution, however certain single-stranded DNA fragments engineered as binding aptamers have been shown to specifically bind to cellulose (65). Since we use double-stranded DNA in our assay, we can rule out any significant interference of DNA-cellulose interaction on the observed rupture forces. Furthermore, CBM tethers binding to weakly bound NCC crystals are most likely removed during the flushing step or the AFS force calibration step. Thus, weakly bound NCC crystals can be eliminated as a potential source for the observation of larger rupture forces seen during our study.

Analysis of the rupture force distribution reveals distinct differences between CBM3a-wt and its Y67A mutant. The fact that no rupture forces greater than 25 pN were measured for Y67A at 1 pN/s could be related to the difference in sample size (N=259 vs. N=138 for WT and Y67A, respectively) as the tail of larger rupture forces at 0.1 pN/s is similar for WT and Y67A (N=161 vs N=159 for WT and Y67A, respectively). A similar shape of rupture force distributions was observed in previous AFM-based studies for CBM3a (33) and CBM1 (30, 66) but previous AFM analysis also found a more Gaussian-like distribution for CBM3a (31, 34). King *et al.* (33) showed that specific binding of CBM3a can be blocked with the addition of NCC and restored by washing the CBM-functionalized AFM tip with an excess of water. In that previous study, both the initial and restored rupture force distributions displayed tailing, suggesting that non-specific binding was likely not the reason for the observation of higher rupture forces as seen in our case.

The tailing of the rupture force distributions towards larger rupture forces may also be correlated with the naturally evolved role of the CBM for proper functioning of the cellulosome. As cellulosomal microbes colonize cellulosic substrates, they are subjected to high interfacial shear forces, for example in the gut-intestine of higher organisms (67). The main cellulosomal scaffold protein cohesin, is relatively stable and unfolds only under forces greater than 140 pN (68, 69), leaving the inter-domain CBM mostly intact (63). Depending on the pulling speed and complex-stabilizing neighboring modules, the cohesin-dockerin interaction, which tethers cellulases to the scaffold, shows multiple unbinding modes with catch-bond behavior (63, 70, 71). A similar response to external forces is hypothesized to be found in other scaffold units such as the CBM. Thus, CBMs may have also evolved to remain bound to cellulose during elevated levels of mechanical stress, but remain flexible enough for the cellulosome to adopt to different bound conformations on the cellulose surface to facilitate substrate hydrolysis (11, 72). This flexibility could be reflected in our

observed broad and bimodal rupture force distribution and the failure of the DHS model to predict the unbinding rate at zero force. Nevertheless, multivalency which can also take the form of multiple CDs interacting with the substrate, may be as important in withstanding mechanical stress but this has not yet been adequately characterized in the literature.

Surface diffusion of cellulases on crystalline cellulose was experimentally verified, although the extent of surface diffusion was minor compared to dynamic CBM-driven binding and unbinding of cellulases to the substrate (73, 74). To date, no motility or processive motion has been observed experimentally for CBMs without being tethered to a CD. However, a computational study of CBM1 from *T. reesei* revealed that CBM1 can diffuse from the hydrophilic to the hydrophobic surface of a cellulose I crystal during which multiple local energy minima with distinct orientations were sampled (56). Similarly, it has been shown that CBM1 can bind in a non-canonical orientation to cellulose-III (35), which indicates that type-A CBMs potentially display a much larger range of binding orientations on crystalline cellulose surfaces. Single-molecule imaging found that CBM1 exhibits distinct surface binding events (28), which could be correlated to distinct regions of crystalline cellulose, and such binding modes may potentially also be found in other type-A CBMs such as CBM3a. When fused to an endoglucanase, CBM3a occupies more binding sites on crystalline cellulose compared to CBM1 fused to the same CD, further suggesting the presence of specific binding sites accessible to different type-A CBMs (75).

The Y67A mutation is located at the edge of the binding site of CBM3a, thus reducing and disrupting the effective total planar binding motif available for multimodal binding interactions with cellulose. Our MD simulations showed that the absence of Y67 “tilts” the whole CBM towards W118 resulting in more frequent H-bond formation between polar residues and the substrate. While MD simulations were carried out using an ideal native cellulose crystal, these results indicate a somehow compensatory effect due to enhanced H-bond formation. In our experiments, in addition to the non-ideal crystallinity, the sulfuric acid- derived NCC also displays sulfate groups on the surface (0.3 μmol sulfur/mg dry NCC). Thus, the mutant could additionally engage with these charged sulfate groups, compensating for the loss of π -stacking forces. These interactions may recover the affinity, resulting in similar off-rates and unbinding forces of WT and mutant, contrasting a previously reported reduction in off-rate for the same mutant using hydrochloric acid- derived NCC (20). However, binding orientations which were determined or stabilized by the Y67-substrate interaction may no longer be favorable in the absence of this residue as we observed a reduction in total available binding sites by 1.8-fold, despite similar unbinding rates determined by QCM-D.

In summary, based on insights from AFS assays and MD simulations, we hypothesize that the planar aromatic binding motif of CBM3a can be grouped into two regions, as highlighted in **Figure 7** as blue and red regions. The first region is dominated by the interactions of W118 and R112/D56 with the substrate, whereas the second region is established by H57 and Y67 interactions with the substrate. For the wild-type, both regions are intact. Thus, pulling on the protein results in a bimodal distribution, depending on which region is first dissociated from the cellulose surface. In contrast, the interaction with the substrate of the second region for the mutant (highlighted in red in **Figure 7**) is greatly reduced due to the significantly higher tilt angle, thus resulting in a unimodal rupture force distribution. Alternatively, the CBM may also take on additional binding orientations on the crystalline surface, where binding residues may span across multiple glucan chains, in addition to binding along a single glucan chain. Even in such a case with more complex binding modes considered (i.e., analogous to Buffon’s needle model for multiple CBM-cellulose binding modes), the unbinding force may differ between on-chain and across-chain binding events, which still may be differentiated using SMFS (35). To further understand the role of each binding residue in the recognizing and dissociating from the substrate, rupture force measurements and MD simulations of other mutants (such as H57 or W118 mutated to alanine), are suggested in future studies. While the application of the DHS model for CBM3a-wt yielded an unbinding rate comparable to previous SMFS results, such classical models still failed to accurately predict the unbinding rate of the Y67A mutant as well as describe the broad rupture force distribution with the obtained fit parameters for

both proteins. This issue might be resolved if bond lifetime measurements are carried out in force-clamp mode rather than transforming rupture force histograms to bond lifetime data. Nevertheless, rupture force histograms could be used further to evaluate the existence of a catch-bond behavior for CBMs. Understanding the influence of each binding residue on the binding and unbinding rate will pave the way for rational engineering approaches to fine-tune CBM-substrate interactions for optimized catalytic activity of cellulases and cellosomes. This will also open up new avenues for CBM utilization, such as nanomaterials (76) or as interfacial anchors for cell immobilization (77). Similar studies using multiplexed SMFS will deepen our fundamental understanding of the complex multimodal interactions between a wide range of proteins with carbohydrates at interfaces, and the functional role of such biophysical interactions in biology.

Materials and Methods

Chemicals and substrates. Unless otherwise mentioned, all reagents were either purchased from VWR International, USA, Fisher Scientific, USA, or Sigma-Aldrich, USA. Streptavidin-coated polystyrene particles (SVP30) with a nominal diameter of 3.11 μm were purchased from Spherotech Inc, USA. Amino-functionalized beads (01-01-503) with a nominal diameter of 5 μm were purchased from Micromod Partikeltechnologie GmbH, Germany, and used as fiducial beads to account for drift during AFS assays. Sulfuric acid-hydrolyzed nanocrystalline cellulose was kindly donated by Richard Reiner at the USDA Forest Product Laboratory (78).

DNA tethers. Linear double-stranded DNA tethers were synthesized in one step by PCR using the pEC-GFP-CBM3a plasmid as a template and 5' modified primers. The biotin-modified primer (forward primer, 5'-biotin-C6-GGCGATCGCCTGGAAGTA) was purchased from Integrated DNA Technologies, Inc. USA. The NTA modified primer (backward primer, 5'-NTA-SS-C6-TCCAAAGGTGAAGAACTGTTACC) was purchased from Gene Link, Inc. USA. The whole plasmid (5.4 kb) was amplified, then purified using the PCR Clean-up kit (IBI Scientific USA) resulting in a linear DNA tether of ~1.8 μm length with one modification on each end of the DNA. Amplification and product purity was verified by gel electrophoresis. In addition, a linear DNA tether of the same length was amplified using a digoxigenin-modified primer instead of NTA (5'-DIG-NHS-TCCAAAGGTGAAGAACTGTTACC, Integrated DNA Technologies, Inc. USA) to bind to anti-digoxigenin Fab fragment antibodies (11214667001, Roche).

Proteins. His₈-GFP-CBM3a wild type and its Y67A mutant were expressed and purified as described previously (20).

Buffers. All AFS experiments were carried out in working buffer (WB) containing 10 mM phosphate buffer at pH 7.4 supplemented with 0.31 mg/ml BSA and casein and 0.19 mg/ml Pluronic F-127, respectively. In addition, two blocking buffers were used to passivate the surface before the experiment. Buffer B1 consists of 10 mM phosphate buffer supplemented with 2.5 mg/ml BSA and casein. Buffer B2 consists of 10 mM phosphate buffer supplemented with 2.2 mg/ml BSA and casein and 5.6 mg/ml Pluronic F-127 respectively. All buffers were degassed in a vacuum (-90 kPa) for 30 minutes.

QCM-D experiments. Quartz Crystal Microbalance with dissipation experiments were carried out and analyzed as described previously (20) except for using 10 mM phosphate buffer at pH 7.4 and sulfuric acid-derived NCC.

Atomistic MD simulations of CBM3a interacting with cellulose. Initial coordinates of cellulose-I were built using a cellulose builder script (79). The fiber was constructed to generate a parallelepiped geometry with crystalline parameters 4 4 5. Initial coordinates of CBM3a were downloaded from the Protein Data Bank (ID: 4JO5) and missing sidechains were re-constructed using the Chimera molecular viewer with the most populated rotameric configurations based on the Dunbrack database (80). All simulations were carried with the Amber16 molecular dynamics package (81) and spatial coordinates were collected every 100 ps for analysis. A detailed

description of the system set-up as well as MD protocol is provided in the **SI Appendix text**. Initial and final MD simulation configurations of WT and Y67A mutant interactions with cellulose surface are provided in txt format as supplementary information.

Cellulose film preparation and AFS chip cleaning. The microfluidic chips used in the AFS are custom designed by LUMICKS B.V., the Netherlands for re-use. Therefore, a reliable protocol for the immobilization and removal of NCC needed to be established. A multilayer deposition process (50) using an automated microfluidic control system was employed from Elvexsys S.A.S, France to obtain a stable cellulose film. The system consists of a microfluidic controller (OB1, driven by compressed nitrogen), a 10-port distribution valve (MUX-D), pressurized fluid reservoirs (2-50 ml), and a manifold. To avoid potential damage to the 10-port valve when in contact with NCC, the valve was used to direct the pressurized nitrogen to the correct reservoir instead of directly controlling the liquid streams. Due to this configuration, installing check-valves on each line was necessary prior to entering the manifold to avoid backflow and cross-contamination between reservoirs. The flowsheet of the setup is shown in Figure 2 (panel A and B) and the detailed part list can be found in **SI Appendix Table S4**. The microfluidic resistance of the setup including the AFS chip was determined to be $3 \mu\text{l} / (\text{min} * \text{mbar})$ and the volume flown through the chip was calculated based on the set pressure and duration. First, the cleaned chip was rinsed with 2 ml DI water, followed by flushing through 200 μl 0.05% (w/v) PLL and incubation for 1 minute. Next, the chip was rinsed with 1ml DI water and blow-dried for 1 minute. 200 μl of NCC at a concentration of 0.5% (w/v) was incubated for 1 minute, followed by 1ml water rinse and drying for 1 minute. The deposition of PLL and NCC was repeated four more times. Following the final NCC layer deposition, the chip was blow-dried for 20 minutes. Finally, the chip was disassembled and the bottom part including the flow cell was placed in an oven at 50°C to dry up overnight.

To confirm cellulose deposition using AFM, flow cells of the same channel geometry as the AFS chips were prepared by cutting the channel from *Parafilm*® and fixing it between two microscope slides. Holes were drilled in one slide to connect 1/16" OD (1/32" ID) PTFE tubing. After assembly, the multilayer deposition process described above was employed manually. The slides were taken apart and dried up overnight at 50°C and stored in a desiccator until AFM imaging. The deposited NCC samples were visualized from the randomly selected area by an AFM (NX-10, Park systems). The AFM was used in non-contact mode operation with a scan size between 2x2 μm and 5x5 μm , 0.3 Hz scan rate, and 11.1 nm set point with the non-contact mode AFM tip (SSS-NCHR, Park systems). The AFM images were analyzed using XEI software (Park systems).

To directly verify the deposition of NCC inside the AFS chip, the fluorescence intensity of GFP-CBM3a-wt bound to NCC was measured. All experiments were carried in at least triplicates. The chip was first rinsed with 500 μl DI water and 500 μl phosphate buffer followed by 15 minutes of passivation of the surface in B1 and B2 buffer, respectively. GFP-CBM3a-wt was diluted in WB to a concentration of 1 μM and incubated for 5 minutes, followed by rinsing 1 ml of WB. The fluorescence images were taken with a CMOS camera (Kiralux, Thorlabs Inc. USA) using $\mu\text{Manager}$ (82) on an inverted fluorescence microscope (Olympus IX 71) equipped with the necessary filters to enable GFP fluorescence. Control experiments on bare glass and PLL treated chip surfaces were performed to estimate the degree of non-specific binding of GFP-CBM3a. All images were corrected for background and shading (83).

The NCC was removed by incubating piranha solution (7:3 concentrated H_2SO_4 :30% H_2O_2 , v/v) two times for 15-30 minutes at 50°C with 500 μl DI water rinses in between. The next step in the cleaning procedure involved incubation of 1 M NaOH for 1-12 hours at room temperature followed by incubation of piranha solution for 15-30 minutes at 50°C, rinsing with 5 ml DI water, and drying. If the AFS chips were used for single-molecule experiments, 5 μm NH_2 -functionalized beads (to serve as fiducial beads) were diluted ~1:1000 in 0.01 M HCl and dried up inside the chip overnight at 50°C before the chip was functionalized with NCC.

Tethered bead preparation for single-molecule force spectroscopy. Single-molecule experiments were carried out on a G1 AFS instrument with G2 AFS chips provided by LUMICKS B.V. After immobilizing NCC, the AFS chip was rinsed with 500 μ l DI water and 500 μ l phosphate buffer. Next, the surface was passivated with B1 and B2 buffer for 15 minutes each and rinsed with WB. The NTA-DNA tethers were diluted to 6 pM in WB containing 6 nM NiCl₂. The bead-DNA-CBM construct was prepared in a two-step procedure. First, 15 μ l streptavidin-coated beads and Ni-NTA-DNA tethers were mixed to yield less than 1 DNA tether per bead and incubated on a rotisserie for 30 minutes. Details about the specificity of the Ni-NTA moiety for His-tagged CBMs can be found in **SI Appendix Fig. S8** and the determination of the overall binding efficiency of DNA tethers to the beads is described in **SI Appendix**. The functionalized beads were washed twice by spinning down, removing the supernatant, and resuspending in 100 μ l WB. GFP-CBM3a-wt or Y67A mutant were diluted to 2 nM in WB. The washed and DNA functionalized bead pellet is resuspended in 20 μ l of either WT or Y67A solution (resulting in a >1000x molar excess of CBM with respect to DNA) and placed on the rotisserie for 30 minutes. Next, the beads were washed twice in WB to remove any unbound CBM and resuspended in 20 μ l WB or B2 if a high non-specific bead binding was observed during SFMS experiments. There was no significant difference in the partition coefficient between WB and B2 for WT ($p=0.68$, $df=7$) and Y67A ($p=0.49$, $df=7$) mutant. Refer to **SI Appendix** for information about the experimental setup and **SI Appendix Fig. S9** for binding data. The CBM-DNA-bead construct was flushed through the AFS chip and incubated for 30 minutes. Non-bound beads were subsequently washed out with WB at a flow rate of 2 μ l/min using a syringe pump (New Era Pump Systems Inc., USA). A small force of ~0.2-0.5 pN was applied to speed up the flushing step. For illustration, a schematic of the single-molecule setup is shown in **Figure 1-D**. After measuring the rupture forces, the chip was rinsed with 100 μ l WB, and the next CBM-DNA-bead sample was inserted.

To verify that the amplified DNA tethers are 1.8 μ m in length, anti-digoxigenin fab fragments dissolved in PBS (20 μ g/ml) were non-specifically bound to the AFS glass surface for 20 minutes, followed by the same passivation procedure as outlined above. The DNA tethers in this experiment were functionalized with digoxigenin instead of NTA (see **Figure 1-C**). The DNA-to-bead ratio was between 5-10 to ensure a sufficient yield of single-molecule tethers. DNA-functionalized beads were incubated on the surface for 10-30 minutes, and the flushing process, bead tracking, and analysis procedure were identical to CBM-tethered bead experiments.

Bead tracking, force ramp application, and determination of rupture forces. Tracking and analysis of the beads were accomplished using the software package provided by LUMICKS, with slight modifications to allow efficient export of rupture forces and associated tethers statistics as well as force-distance curves to a spreadsheet. The procedure for identifying a single-molecule tether, force calibration, and rupture force determination is described in detail elsewhere (36). The beads were tracked at 20 Hz using a 4x magnification objective. The trajectory of the beads without applied force was monitored for 8-10 minutes to determine the point of surface attachment (anchor point). Next, the force on each bead was calibrated by applying a constant amplitude for 2-4 minutes. Typically, 2-3 different amplitude values were used to build the calibration curve between the applied amplitude and effective force on each bead. Single-molecule tethers were identified by the root-mean-square fluctuation (*RMS*) and symmetrical motion (*Sym*) of the bead around the anchor point during the time frame for anchor point determination. For the CBM-cellulose experiment, values of single-molecule tethers for *RMS* and *Sym* are in the range between 850-1200 nm and 1.0-1.3 respectively. During force calibration, the diffusion coefficient of the bead and the force were used as fit parameters. This diffusion coefficient was compared to the diffusion coefficient determined by the Stokes-Einstein relation and was in the range between 0.8-1.2 for single tethers. The force obtained during force calibration was used to estimate the theoretical extension of DNA using the extensible WLC model (52). This extension was compared to the measured length during that force calibration point to yield the dimensionless length l_{fc} and was expected to be close to 1 for single tethers. Next, a linear force ramp of either 0.1 or 1 pN/s was applied. Rupture forces were determined through the software by finding the time frame at which

the z-position of the bead was outside the interval covered by the lookup-table (LUT) value (36). An example time trace of a typical rupture force measurement is shown in **SI Appendix Fig. S10**. Each trace and force-extension curve (FD) during force ramp application was inspected manually to determine the rupture force accurately.

Analysis of rupture forces. Further evaluation of traces as well as data analysis was carried out by a custom-written MATLAB® script as briefly described below. For each known single-molecule trace, several indicators such as *RMS*, *Sym*, l_{fc} , rupture force and loading rate, along with the obtained force-distance (FD) curve during force ramp application, were imported into MATLAB®. To each FD curve, the dimensionless contour length l_c of the WLC model based on the expected contour length of 1800 nm was fitted using a persistence length of $l_p=42$ nm and stretch modulus $S=1300$ pN (84). This fitted length (determined during the force ramp) was compared to the dimensionless length during force calibration l_{fc} , and only traces close to 1 were further analyzed. The script also identified traces in which the rupture force or loading rate was 3 standard deviations away from the sample mean. Those traces were examined manually and discarded if the FD curve or any other mentioned statistics indicated that the trace did not originate from a single-tethered bead. To ensure that no bias was introduced by removing traces, the remaining data was subjected to a Pearson and Spearman correlation coefficient test between the obtained rupture force and *RMS*, *Sym* and l_{fc} respectively. Finally, the obtained rupture force histograms were converted to force- dependent bond lifetime data and analyzed using the procedure outlined by Dudko *et al.* (55) to obtain the bond lifetime in the absence of force. The rupture force histograms were converted to force- dependent bond lifetime data using Equation 1:

$$\tau(F_k) = \frac{\Delta F (h_k / 2 + \sum_{i=k+1}^N h_i)}{h_k \dot{F}(F_k)} \quad (1)$$

Where $\tau(F_k)$ and $\dot{F}(F_k)$ are the average bond lifetime and loading rate at the k^{th} bin and $F_k = F_0 + (k - 1/2)\Delta F$. The rupture force histogram is composed of N bins of width ΔF starting from F_0 and ending at $F_0 + N\Delta F$. The number of counts in the i^{th} bin is C_i and the height of each bin can be calculated as $h_i = C_i / N_{\text{tot}}\Delta F$ where N_{tot} is the total number of counts.

The force-dependent bond lifetime $\tau(F)$ is described using Equation 2:

$$\tau(F) = \tau_0 \left(1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger}\right)^{1-1/\nu} e^{-\beta \Delta G^\ddagger \left[1 - \left(1 - \frac{\nu F x^\ddagger}{\Delta G^\ddagger}\right)^{1/\nu}\right]} \quad (2)$$

Where $\beta = 1/k_B T$, $\tau_0 = 1/k_{\text{off}}$ is the bond lifetime (or inverse of the unbinding rate k_{off}), $x^\ddagger$ is the transition state distance and $\Delta G^\ddagger$ the apparent free energy of activation in the absence of the external force. The shape factor $\nu=1/2$ or $2/3$ describes the underlying free-energy profile as cusp or linear-cubic, respectively.

The distribution of rupture forces is described by Equation 3:

$$p(F) = \frac{1}{\dot{F}(F)\tau(F)} e^{-\int_0^F \frac{1}{\dot{F}(f)\tau(f)} df} \quad (3)$$

Acknowledgments

S.P.S.C. acknowledges support from the NSF (CBET CAREER Award 1846797), Rutgers Aresty Research Center, and Rutgers SOE Startup Funds. J.M.Y was supported by the DOE, Office of

Energy Efficiency and Renewable Energy under Agreement No. 28598. K.-B.L. acknowledges partial support from the NSF (CBET-1803517), the New Jersey Commission on Spinal Cord (CSCR17IRG010; CSCR16ERG019), and NIH R01 (1R01DC016612, 3R01DC016612-01S1, and 5R01DC016612-02S1). C.A.L. and S.G. acknowledge support from LANL LDRD ER (XWX2) and LANL institutional program for computational resources. M.H. would like to thank Dr. Bhargava Nemmaru for his help with the QCM-D data analysis. Figures 1-A,C,D, Figure S4, Figure S8-A and Figure S10-A were created using BioRender.com.

References

1. N. Georgelis, N. H. Yennawar, D. J. Cosgrove, Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14830–5 (2012).
2. C. D.J., Loosening of plant cell walls by expansins. *Nature* **407**, 321–326 (2000).
3. L. Qin, *et al.*, A nematode expansin acting on plants. *Nature* **427**, 30–30 (2004).
4. M. D. L. Suits, *et al.*, Conformational analysis of the *Streptococcus pneumoniae* hyaluronate lyase and characterization of its hyaluronanspecific carbohydrate-binding module. *J. Biol. Chem.* **289**, 27264–27277 (2014).
5. M. J. Jedrzejewski, Pneumococcal Virulence Factors: Structure and Function. *Microbiol. Mol. Biol. Rev.* **65**, 187–207 (2001).
6. S. P. S. Chundawat, G. T. Beckham, M. E. Himmel, B. E. Dale, Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals. *Annu. Rev. Chem. Biomol. Eng.* **2**, 121–145 (2011).
7. M. E. Himmel, *et al.*, Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science (80-.)*. **315**, 804–807 (2007).
8. A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769–781 (2004).
9. C. M. Payne, *et al.*, Fungal cellulases. *Chem. Rev.* **115**, 1308–1448 (2015).
10. C. M. G. A. Fontes, H. J. Gilbert, Cellulosomes: Highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* **79**, 655–681 (2010).
11. M. Eibinger, T. Ganner, H. Plank, B. Nidetzky, A Biological Nanomachine at Work: Watching the Cellulosome Degrade Crystalline Cellulose. *ACS Cent. Sci.* **6**, 739–746 (2020).
12. E. A. Bayer, H. Chanzy, R. Lamed, Y. Shoham, Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**, 548–557 (1998).
13. R. H. Doi, A. Kosugi, Cellulosomes: Plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.* **2**, 541–551 (2004).
14. K. Igarashi, *et al.*, Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science (80-.)*. **333**, 1279–1282 (2011).
15. Y. Zhang, M. Zhang, R. Alexander Reese, H. Zhang, B. Xu, Real-time single molecular study of a pretreated cellulose hydrolysis mode and individual enzyme movement. *Biotechnol. Biofuels* **9**, 85 (2016).
16. Y. Shibafuji, *et al.*, Single-molecule imaging analysis of elementary reaction steps of *trichoderma reesei* cellobiohydrolase i (Cel7A) hydrolyzing crystalline cellulose Ia and III. *J. Biol. Chem.* **289**, 14056–14065 (2014).
17. B. W. McLean, *et al.*, Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10a to cellulose: Specificity and identification of functionally important amino acid residues. *Protein Eng.* **13**, 801–809 (2000).

18. P. J. Kraulis, *et al.*, Determination of the Three-Dimensional Solution Structure of the C-Terminal Domain of Cellobiohydrolase I from *Trichoderma reesei*. A Study Using Nuclear Magnetic Resonance and Hybrid Distance Geometry-Dynamical Simulated Annealing. *Biochemistry* **28**, 7241–7257 (1989).
19. J. Tormo, *et al.*, Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751 (1996).
20. B. Nemmaru, *et al.*, Reduced type-A carbohydrate-binding module interactions to cellulose I leads to improved endocellulase activity. *Biotechnol. Bioeng.* **118**, 1141–1151 (2021).
21. D. Gao, *et al.*, Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 10922–10927 (2013).
22. J. Kari, *et al.*, Physical constraints and functional plasticity of cellulases. *Nat. Commun.* **12**, 1–10 (2021).
23. D. W. Abbott, A. B. Boraston, Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol.* **510**, 211–231 (2012).
24. A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, C. A. Haynes, Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci.* **93**, 12229–12234 (1996).
25. Y. Zhang, *et al.*, Interactions between type A carbohydrate binding modules and cellulose studied with a quartz crystal microbalance with dissipation monitoring. *Cellulose* **4** (2020).
26. J. Guo, J. M. Catchmark, Binding specificity and thermodynamics of cellulose-binding modules from *trichoderma reesei* Cel7A and Cel6A. *Biomacromolecules* **14**, 1268–1277 (2013).
27. S. J. Charnock, *et al.*, Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello-and mannohexaose. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14077–14082 (2002).
28. A. Nakamura, *et al.*, Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase *trichoderma reesei* Cel6A and its domains on crystalline cellulose. *J. Biol. Chem.* **291**, 22404–22413 (2016).
29. S. K. Brady, S. Sreelatha, Y. Feng, S. P. S. Chundawat, M. J. Lang, Cellobiohydrolase 1 from *Trichoderma reesei* degrades cellulose in single cellobiose steps. *Nat. Commun.* **6**, 10149 (2015).
30. A. Griffo, *et al.*, Binding Forces of Cellulose Binding Modules on Cellulosic Nanomaterials. *Biomacromolecules* **20**, 769–777 (2019).
31. M. Zhang, S.-C. Wu, W. Zhou, B. Xu, Imaging and Measuring Single-Molecule Interaction between a Carbohydrate-Binding Module and Natural Plant Cell Wall Cellulose. *J. Phys. Chem. B* **116**, 9949–9956 (2012).
32. M. Zhang, B. Wang, B. Xu, Mapping single molecular binding kinetics of carbohydrate-binding module with crystalline cellulose by atomic force microscopy recognition imaging. *J. Phys. Chem. B* **118**, 6714–6720 (2014).
33. J. R. King, C. M. Bowers, E. J. Toone, Specific binding at the cellulose binding module-cellulose interface observed by force spectroscopy. *Langmuir* **31**, 3431–3440 (2015).

34. M. Zhang, B. Wang, B. Xu, Measurements of single molecular affinity interactions between carbohydrate-binding modules and crystalline cellulose fibrils. *Phys. Chem. Chem. Phys.* **15**, 6508–6515 (2013).
35. S. P. S. Chundawat, *et al.*, Molecular origins of reduced activity and binding commitment of processive cellulases and associated carbohydrate-binding proteins to cellulose III. *J. Biol. Chem.* **296**, 100431 (2021).
36. G. Sitters, *et al.*, Acoustic force spectroscopy. *Nat. Methods* **12**, 47–50 (2014).
37. D. Kamsma, R. Creighton, G. Sitters, G. J. L. Wuite, E. J. G. Peterman, Tuning the Music: Acoustic Force Spectroscopy (AFS) 2.0. *Methods* **105**, 26–33 (2016).
38. F. Ritort, Single-molecule experiments in biological physics: Methods and applications. *J. Phys. Condens. Matter* **18** (2006).
39. E. Evans, Energy landscapes of biomolecular adhesion and receptor anchoring at interfaces explored with dynamic force spectroscopy. *Faraday Discuss.*, 1–16 (1998).
40. F. Rico, V. T. Moy, Energy landscape roughness of the streptavidin–biotin interaction. *J. Mol. Recognit.* **20**, 495–501 (2007).
41. R. P. Goodman, *et al.*, A facile method for reversibly linking a recombinant protein to DNA. *ChemBioChem* **10**, 1551–1557 (2009).
42. G. D. Meredith, H. Y. Wu, N. L. Allbritton, Targeted protein functionalization using his-tags. *Bioconjug. Chem.* **15**, 969–982 (2004).
43. J. Shimada, T. Maruyama, M. Kitaoka, N. Kamiya, M. Goto, DNA-enzyme conjugate with a weak inhibitor that can specifically detect thrombin in a homogeneous medium. *Anal. Biochem.* **414**, 103–108 (2011).
44. J. Shimada, *et al.*, Conjugation of DNA with protein using His-tag chemistry and its application to the aptamer-based detection system. *Biotechnol. Lett.* **30**, 2001–2006 (2008).
45. L. Schmitt, M. Ludwig, H. E. Gaub, R. Tampé, A metal-chelating microscopy tip as a new toolbox for single-molecule experiments by atomic force microscopy. *Biophys. J.* **78**, 3275–3285 (2000).
46. C. Verbelen, H. J. Gruber, Y. F. Dufrêne, The NTA-His6 bond is strong enough for AFM single-molecular recognition studies. *J. Mol. Recognit.* **20**, 490–494 (2007).
47. F. Kienberger, *et al.*, Recognition Force Spectroscopy Studies of the NTA-His6 Bond. *Single Mol.* **1**, 59–65 (2000).
48. M. Conti, G. Falini, B. Samorì, How strong is the coordination bond between a histidine tag and Ni- nitrilotriacetate? An experiment of mechanochemistry on single molecules. *Angew. Chemie - Int. Ed.* **39**, 215–218 (2000).
49. R. W. Friddle, A. Noy, J. J. De Yoreo, Interpreting the widespread nonlinear force spectra of intermolecular bonds. *Proc. Natl. Acad. Sci.* **109**, 13573–13578 (2012).
50. E. D. Cranston, D. G. Gray, Morphological and optical characterization of polyelectrolyte multilayers incorporating nanocrystalline cellulose. *Biomacromolecules* **7**, 2522–2530 (2006).

51. E. Kontturi, *et al.*, Cellulose nanocrystal submonolayers by spin coating. *Langmuir* **23**, 9674–9680 (2007).
52. T. Odijk, Stiff Chains and Filaments under Tension. *Macromolecules* **28**, 7016–7018 (1995).
53. D. Freedman, P. Diaconis, On the histogram as a density estimator:L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**, 453–476 (1981).
54. O. K. Dudko, G. Hummer, A. Szabo, Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.* **96**, 1–4 (2006).
55. O. K. Dudko, G. Hummer, A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15755–15760 (2008).
56. M. R. Nimlos, *et al.*, Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J. Biol. Chem.* **287**, 20603–20612 (2012).
57. C. Aulin, *et al.*, Nanoscale Cellulose Films with Different Crystallinities and Mesostuctures—Their Surface Properties and Interaction with Water. *Langmuir* **25**, 7675–7685 (2009).
58. S. Ahola, J. Salmi, L.-S. Johansson, J. Laine, M. Österberg, Model Films from Native Cellulose Nanofibrils. Preparation, Swelling, and Surface Interactions. *Biomacromolecules* **9**, 1273–1282 (2008).
59. R. J. Moon, A. Martini, J. Nairn, J. Simonsen, J. Youngblood, Cellulose nanomaterials review: structure, properties and nanocomposites. *Chem. Soc. Rev.* **40**, 3941 (2011).
60. T. Abitbol, A. Palermo, J. M. Moran-Mirabal, E. D. Cranston, Fluorescent labeling and characterization of cellulose nanocrystals with varying charge contents. *Biomacromolecules* **14**, 3278–3284 (2013).
61. Z. Huang, V. S. Raghuwanshi, G. Garnier, Functionality of immunoglobulin G and immunoglobulin M antibody physisorbed on cellulosic films. *Front. Bioeng. Biotechnol.* **5**, 1–10 (2017).
62. S. M. Sedlak, *et al.*, Direction Matters: Monovalent Streptavidin/Biotin Complex under Load. *Nano Lett.* **19**, 3415–3421 (2019).
63. S. W. Stahl, *et al.*, Single-molecule dissection of the high-affinity cohesin-dockerin complex. *Proc. Natl. Acad. Sci.* **109**, 20431–20436 (2012).
64. B. Alberts, G. Herrick, “[11] DNA-cellulose chromatography” in *Methods in Enzymology*, (Academic Press, 1971), pp. 198–217.
65. B. J. Boese, R. R. Breaker, In vitro selection and characterization of cellulose-binding DNA aptamers. *Nucleic Acids Res.* **35**, 6378–6388 (2007).
66. B. Arslan, *et al.*, The Effects of Noncellulosic Compounds on the Nanoscale Interaction Forces Measured between Carbohydrate-Binding Module and Lignocellulosic Biomass. *Biomacromolecules* **17**, 1705–1715 (2016).
67. L. Artzi, E. A. Bayer, S. Moraïs, Cellulosomes: Bacterial nanomachines for dismantling

- plant polysaccharides. *Nat. Rev. Microbiol.* **15**, 83–95 (2017).
68. T. Verdorfer, *et al.*, Combining in Vitro and in Silico Single-Molecule Force Spectroscopy to Characterize and Tune Cellulosomal Scaffoldin Mechanics. *J. Am. Chem. Soc.* **139**, 17841–17852 (2017).
 69. A. Galera-Prat, S. Morais, Y. Vazana, E. A. Bayer, M. Carrión-Vázquez, The cohesin module is a major determinant of cellulosome mechanical stability. *J. Biol. Chem.* **293**, 7139–7147 (2018).
 70. M. A. Jobst, *et al.*, Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *Elife* **4**, 1–19 (2015).
 71. Z. Liu, *et al.*, High force catch bond mechanism of bacterial adhesion in the human gut. *Nat. Commun.* **11**, 1–12 (2020).
 72. J. J. Adams, *et al.*, Insights into higher-order organization of the cellulosome revealed by a dissect-and-build approach: Crystal structure of interacting *Clostridium thermocellum* multimodular components. *J. Mol. Biol.* **396**, 833–839 (2010).
 73. J. Moran-Mirabal, J. Bolewski, L. P. Walker, Thermobifida fusca Cellulases Exhibit Limited Surface Diffusion on Bacterial Micro-Crystalline Cellulose. *Biotech Bioeng* **110**, 47–56 (2013).
 74. E. J. Jervis, C. A. Haynes, D. G. Kilburn, Surface diffusion of cellulases and their isolated binding domains on cellulose. *J. Biol. Chem.* **272**, 24016–24023 (1997).
 75. G. Carrard, A. Koivula, H. Söderlund, P. Béguin, Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 10342–10347 (2000).
 76. J. M. Malho, *et al.*, Enhanced plastic deformations of nanofibrillated cellulose film by adsorbed moisture and protein-mediated interactions. *Biomacromolecules* **16**, 311–318 (2015).
 77. A. Wierzbza, U. Reichl, R. F. B. Turner, R. A. J. Warren, D. G. Kilburn, Adhesion of mammalian cells to a recombinant attachment factor, CBD/RGD, analyzed by image analysis. *Biotechnol. Bioeng.* **46**, 185–193 (1995).
 78. R. S. Reiner, A. W. Rudie, “Process scale-up of cellulose nanocrystal production to 25 kg per batch at the Forest Products Laboratory” in *Production and Applications of Cellulose Nanomaterials*, (TAPPI Press, 2013), pp. 21–24.
 79. T. C. F. Gomes, M. S. Skaf, Cellulose-builder: A toolkit for building crystalline structures of cellulose. *J. Comput. Chem.* **33**, 1338–1346 (2012).
 80. M. V. Shapovalov, R. L. Dunbrack, A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844–858 (2011).
 81. S. Le Grand, A. W. Götz, R. C. Walker, SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).
 82. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, N. Stuurman, Computer control of microscopes using manager. *Curr. Protoc. Mol. Biol.*, 1–17 (2010).

83. M. A. Model, J. K. Burkhardt, A standard for calibration and shading correction of a fluorescence microscope. *Commun. Clin. Cytom.* **46**, 309–316 (2001).
84. M. D. Wang, H. Yin, R. Landick, J. Gelles, S. M. Block, Stretching DNA with optical tweezers. *Biophys. J.* **72**, 1335–1346 (1997).

Figures and Tables

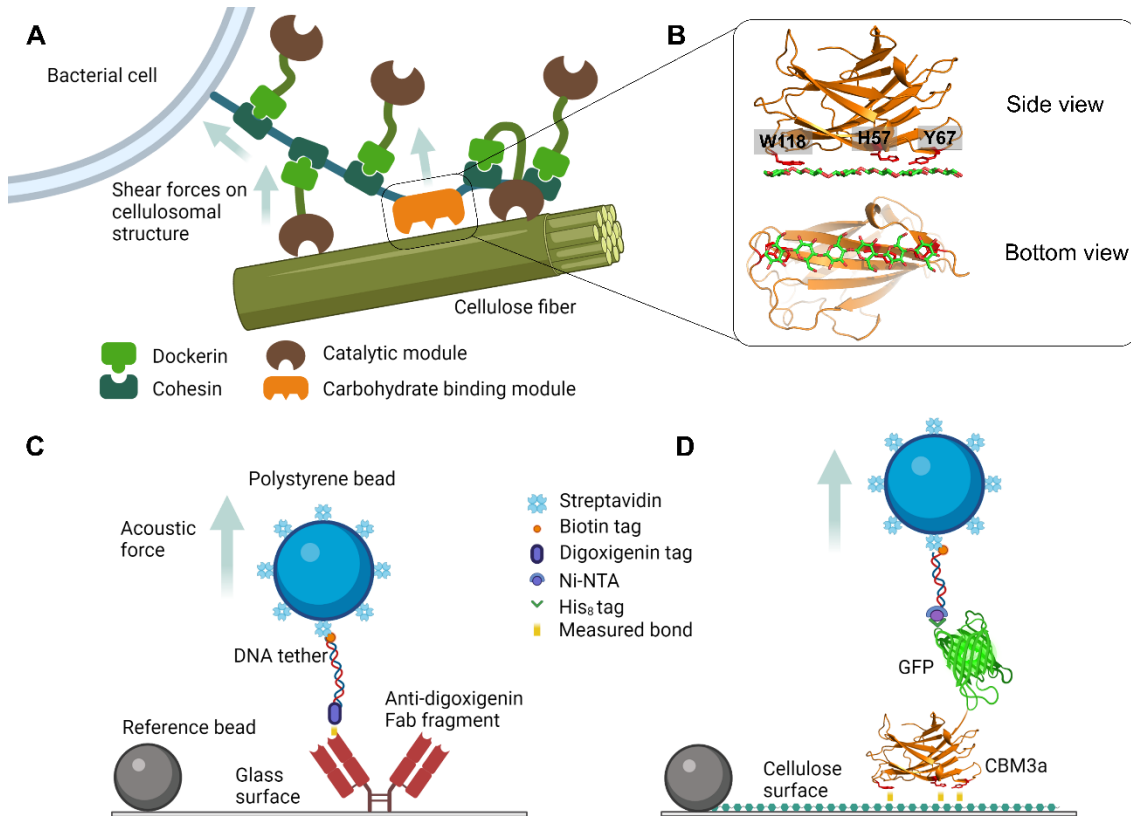


Figure 1. Schematic of a generic cellulosome and acoustic force spectroscopy experimental setup to characterize single-molecule model protein-ligand and CBM3a-polysaccharide unbinding forces (not to scale). A) Generic bacterial cell surface anchored cellulosome is shown adhering to a single cellulose fiber. The carbohydrate-binding module (CBM) binds to cellulose and directs the catalytic domains to the cellulose surface. Shear forces due to the substrate or cell movement are exerted on the cellulosome scaffold. B) Side and bottom view of CBM3a structure with key aromatic residues involved in binding to cellulose (PDB: 1NBC). The aromatic residues W118, H57, and Y67 form a flat binding surface complementary to the cellulose surface. C) Schematic outlining the measurement of the unbinding force of model digoxigenin (DIG) ligand from surface-bound anti-DIG antibody to validate the bead preparation method as well as analysis procedure of AFS traces. D) Schematic outlining the measurement of the unbinding force of His-GFP tagged CBM3a from a nanocrystalline cellulose surface using the AFS assay.

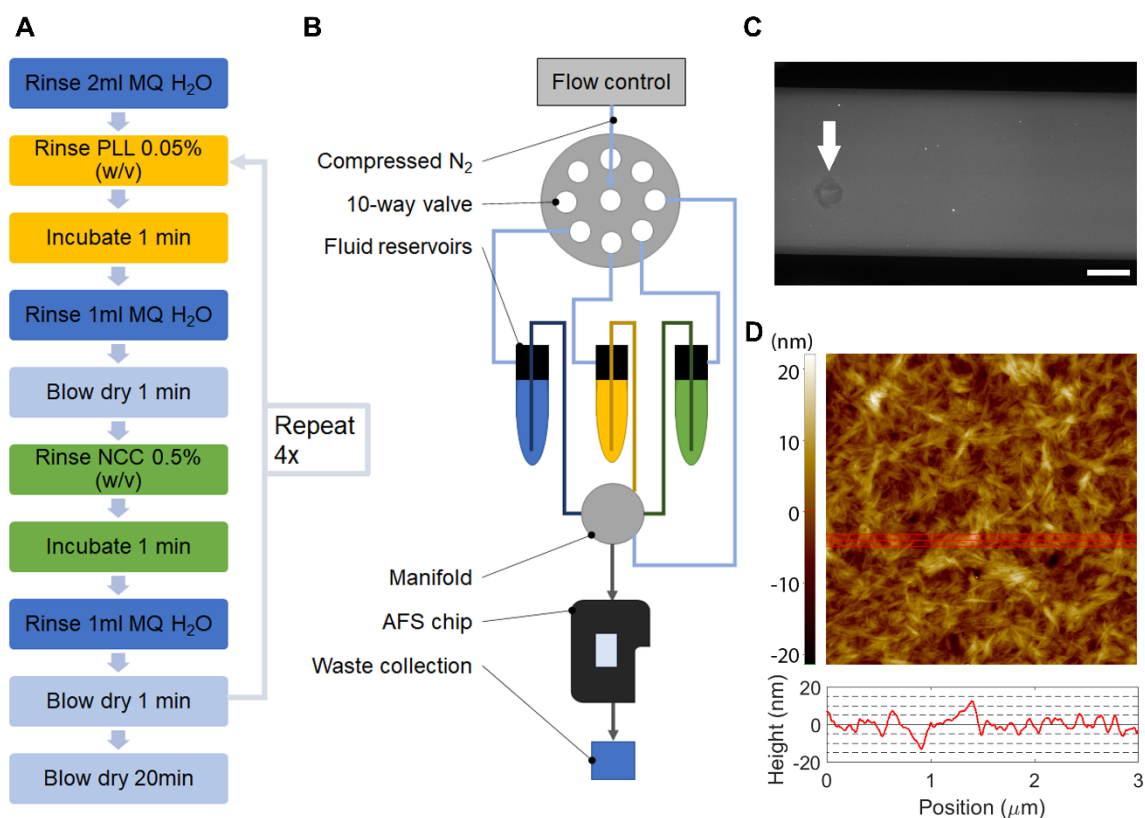


Figure 2. Multilayer deposition of nanocrystalline cellulose (NCC) within the AFS chip enables the characterization of a uniform and reproducible surface. A) Flowchart and B) Process flow diagram of the NCC deposition method, C) Fluorescence image of the NCC modified AFS chip. GFP-CBM3a-wt was used to bind to and visualize the deposited NCC film. The arrow indicates a representative area where a bubble was stuck at some point during the NCC deposition, resulting in less NCC bound and hence a lesser amount of GFP-CBM3a bound to that area as well. The scale bar is 500 μm. D) AFM image (3x3 μm) of the NCC film deposited on a glass slide showing a densely covered surface. The red line represents the area used to obtain the average height profile trace shown below. Despite minor aggregation of NCC crystals during layer-by-layer deposition, height differences are less than 20 nm.

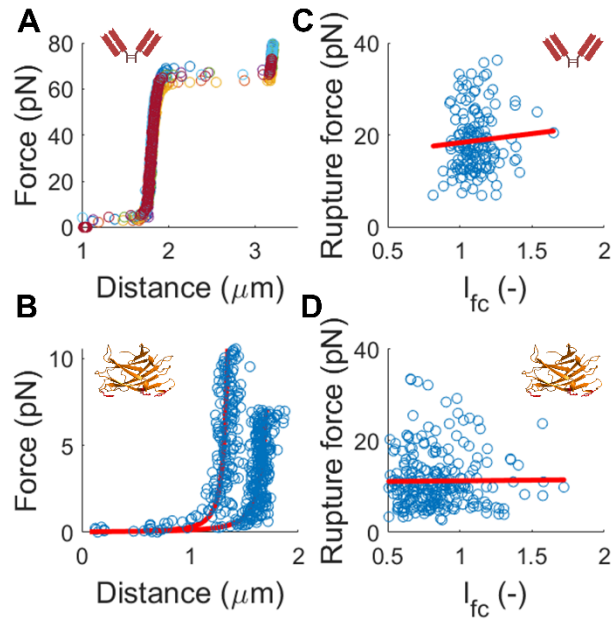


Figure 3. No correlation was observed between tether length and rupture force for DIG-aDIG and CBM3a-cellulose interactions. A) Force-distance (FD) curves of DNA anchored to the chip surface by the DIG-aDIG bond (N=7). The extension at ~65 pN is characteristic for a single DNA tether and indicates overstretching of DNA. B) Example of FD curves for DNA anchored to the chip surface by the NCC-CBM-bond. The red line shows the WLC fit with $l_p=42$ nm and $S=1300$ pN. Despite following the WLC model, the tethers show a reduction in length of 25% on average. No overstretching was observed since CBMs detach from the surface well below 65 pN. C) Scatterplot and linear fit (red line) of rupture force and dimensionless length during force calibration (l_{fc}) for DIG-aDIG (N=156). D) Scatterplot and linear fit (red line) of rupture force and l_{fc} for CBM3a-wt at 1 pN/s (N=259). No significant correlation is found between the measured rupture force and l_{fc} (See SI Appendix Table S1 and Table S2).

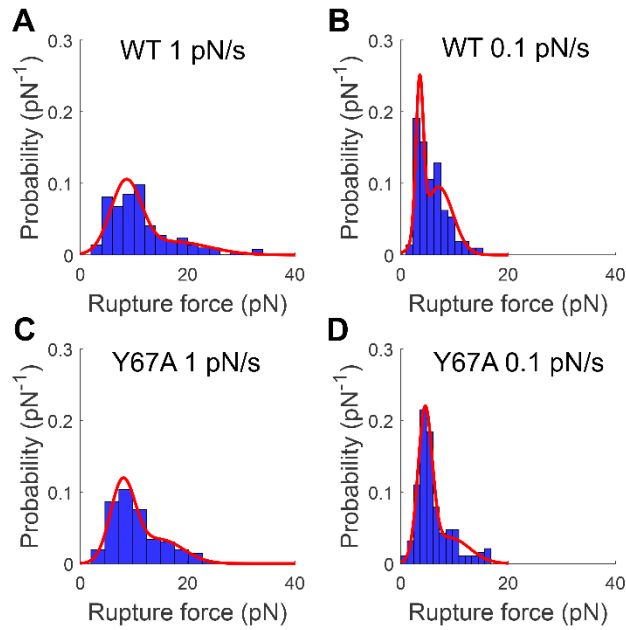


Figure 4. AFS reveals distinct multimodal CBM-cellulose rupture force distribution at lower loading rates. A-B) Obtained rupture force histograms and fit to a double normal distribution for CBM3a-wt at a loading rate of 1 pN/s (N=259) and 0.1 pN/s (N=161) respectively. C-D) Rupture force histograms and fit to a double normal distribution for CBM3a Y67A at a loading rate of 1 pN/s (N=138) and 0.1 pN/s (N=159). The fit parameters are summarized in **SI Appendix Table S3**. The tail towards higher rupture forces is observed in all cases, however, the Y67A mutant displays only a single peak at both loading rates, whereas CBM3a-wt shows no clear single peak, but rather 2 or more rupture force peaks.

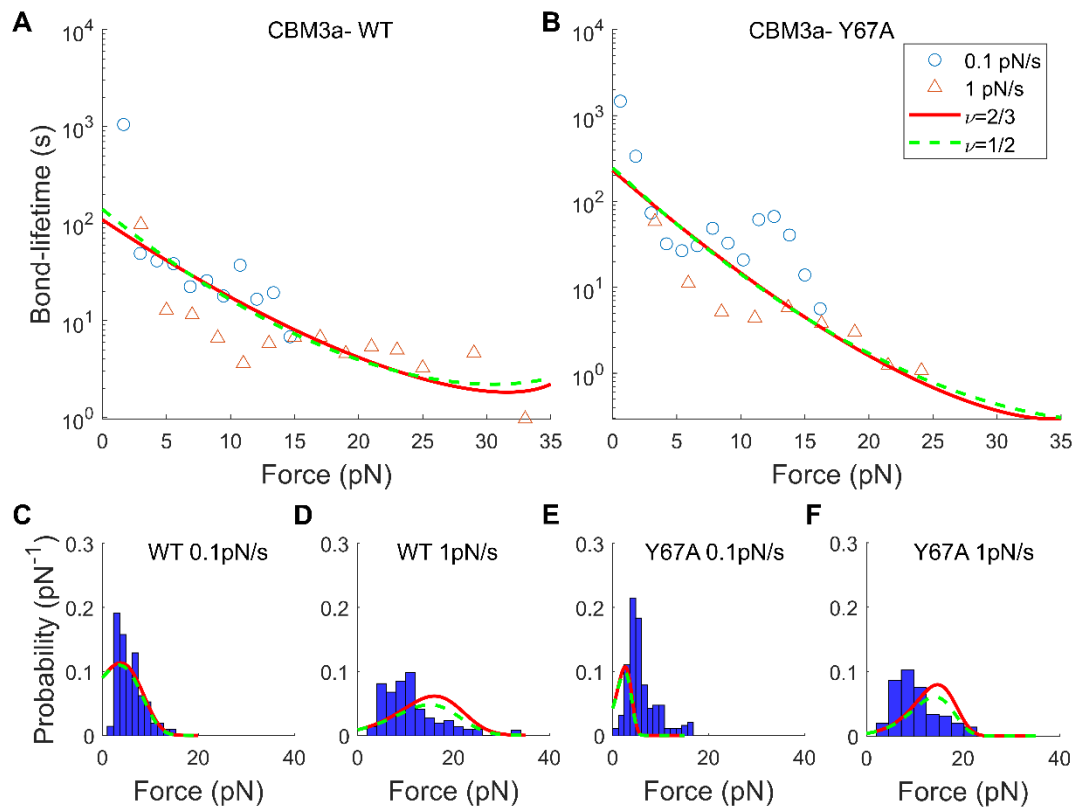


Figure 5. Application of the DHS model to obtained CBM3a-cellulose rupture forces highlights limitations of classical theory to study multivalent protein-polysaccharide unbinding interactions. A, B) Force-dependent bond lifetime obtained from transforming rupture force distributions at 0.1 pN/s (o) and 1 pN/s (Δ) using Equation 1 for WT and Y67A, respectively. The fit of Equation 2 is shown for $\nu=2/3$ (red, solid line) and $\nu=1/2$ (green, dashed line). C-F) Rupture force distributions at 0.1 pN/s and 1 pN/s with the fit of Equation 3 for WT and Y67A respectively, using the parameters obtained from fitting Equation. 2 to data in A) and B) for $\nu=2/3$ (red, solid line) and $\nu=1/2$ (green, dashed line). While both shape factors yield a qualitatively similar fit, only $\nu=2/3$ results in $\int p(f)df = 1$.

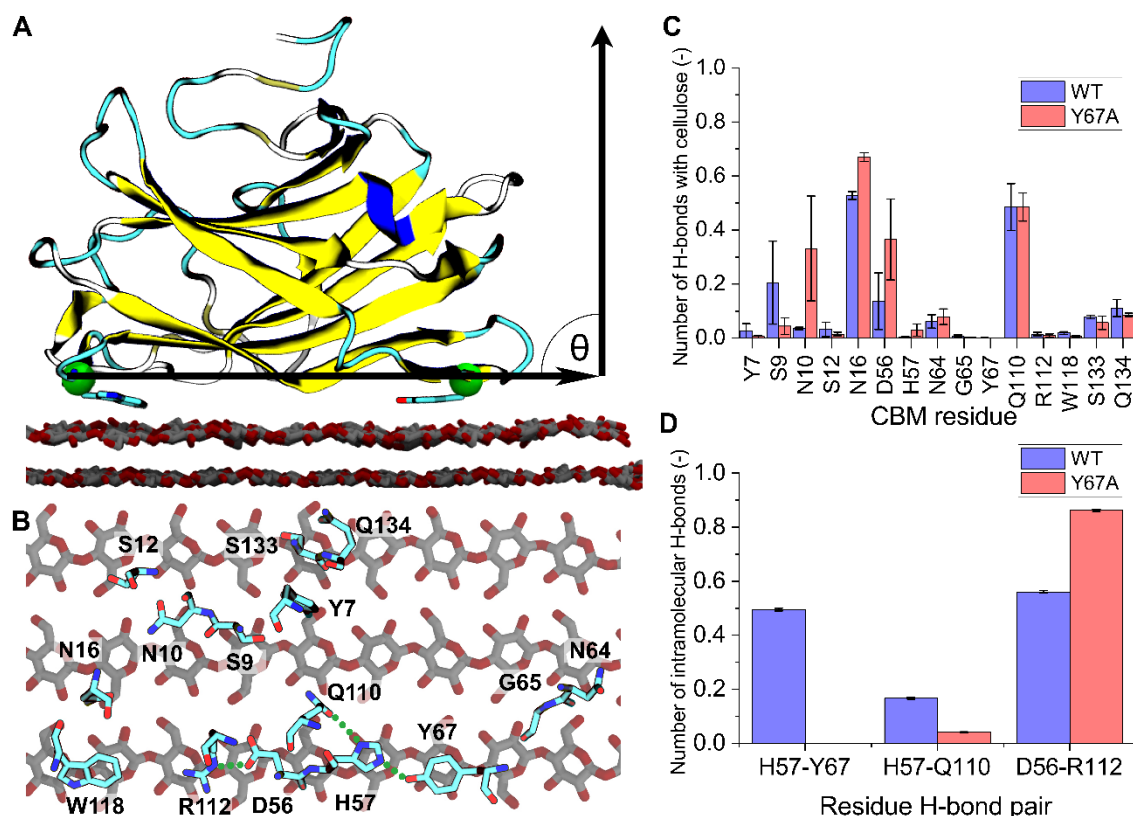


Figure 6: MD simulations provide structural insights into the multiple interactions of CBM3a residues with the cellulose surface. A) Representative configuration of CBM3a-wt interacting with the cellulose crystal as obtained from unbiased MD simulations. The vector formed between the C α of W118 (left green sphere) and C α of Y67 (right green sphere) indicates the horizontal alignment of the CBM towards the reducing end of the crystal. The angle between this vector and the normal vector of the surface is defined as θ . B) Close-up top view of the planar binding motif residues of CBM3a-wt identified to be in close contact with cellulose during MD simulations. Backbone and hydrogen atoms were omitted for clarity. Select inter-residual H-bonds are indicated by the dotted green line. C) Comparison of average number of H-bonds with cellulose for CBM3a-wt and Y67A mutant residues, respectively. Reduction in S9 seems to be compensated by increased stabilization for N10, N16 and D56 in the mutant. D) Average intramolecular H-bond formation between pairs of amino acids in both the WT and Y67A mutant. The Y67A mutation leads to the total bond rupture between the H57 and A67 pair, however significantly greater interactions are observed between D56 and R112. Error bars in C) represent the average deviation of all trajectories of two independent simulations and error bars in D) are SEM.

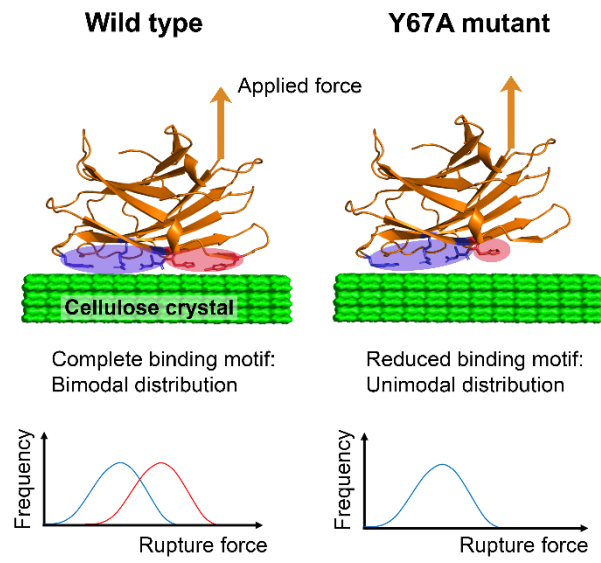


Figure 7: Summary of hypothesized origin of multimodal rupture force distribution observed for CBM3a-wt bound to cellulose. The planar binding motif may be grouped into two regions as highlighted in blue and red oval regions shown here. For CBM3a-wt, both regions are intact and interacting with the surface in a multimodal manner. Hence, pulling on the protein yields a bimodal rupture force distribution, depending on which region ruptures from the surface first. The Y67A mutant binds to the cellulose crystal slightly tilted, reducing the interactions between the red region and the substrate. Applying a force on the mutant may therefore result in a unimodal rupture force distribution since only the blue binding region highlighted is fully engaged with the substrate at any given point in time.

Table 1. Fit parameters of force-dependent bond lifetimes for CBM3a-wt and Y67A mutant, respectively. The integral column refers to the numerical integration of the rupture force probability as described by Equation 3 in the Methods section for both loading rates.

	ν (-)	k_{off} (s ⁻¹)	$x^\ddagger$ (nm)	$\Delta G^\ddagger$ (k _B T)	$\int p(f)df$ at 0.1 pN/s [1 pN/s] (-)
WT	2/3	0.0091	0.88	5.4	1.0 [1.0]
	1/2	0.0071	1.24	5.9	0.95 [0.80]
Y67A	2/3	0.0044	1.12	8.1	1.0 [1.0]
	1/2	0.0041	1.22	8.7	0.93 [0.79]

Supporting Information for

Acoustic Force Spectroscopy Reveals Subtle Differences in Cellulose-Unbinding Behavior of Carbohydrate Binding Modules

Markus Hackl^a, Edward V. Contrada^a, Jonathan E. Ash^a, Atharv Kulkarni^a, Jinho Yoon^b, Hyeon-Yeol Cho^b, Ki-Bum Lee^b, John M. Yarbrough^c, Cesar A. López^d, Sandrasegaram Gnanakaran^d, Shishir P. S. Chundawat^{a,*}

^a Department of Chemical and Biochemical Engineering, and ^b Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway NJ, 08854, ^c Biosciences Center, National Renewable Energy Laboratory, Golden, CO, 80401; ^d Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, 87545

*Corresponding author Shishir P. S. Chundawat
Email: shishir.chundawat@rutgers.edu

This PDF file includes:

Supplementary text
Figures S1 to S10
Tables S1 to S4
Legend for MD simulation files in rtf format
SI References

Other supporting materials for this manuscript include the following (available online at <https://www.pnas.org/doi/abs/10.1073/pnas.2117467119#supplementary-materials>):

Start_WT.pdb.rtf
Final_WT.pdb.rtf
Start_Y67A.pdb.rtf
Final_Y67A.pdb.rtf

Supplementary text

Biotinylated DNA binding efficiency to streptavidin-coated beads. The binding efficiency was determined using a supernatant assay (1) in which the concentration of unbound DNA was measured using the Quant-it PicoGreen dsDNA Assay kit (Thermo Fisher Scientific). For each experiment, 70 μ l of 3.11 μ m diameter streptavidin-coated beads were washed in Tris-EDTA (TE) buffer at pH 7 containing 1mg/ml BSA and mixed in a PCR tube with 1.8 μ m long biotin/digoxigenin-DNA tethers at a ratio of 100 tethers per bead. Control experiments with the same concentration of DNA but without beads were used to estimate the amount of non-specific binding of DNA to the PCR tube. The bead-DNA mixture was incubated on a rotisserie for 30 minutes and subsequently spun down to separate beads and supernatant. The supernatant (50 μ l) was transferred to a clear-bottom 96-well plate and mixed with the fluorescent dye following the manufacturer's protocol. The fluorescence was measured at 480 nm excitation, 520 nm emission with a cut-off of 495 nm. The concentration of DNA was determined using a standard curve of λ -DNA supplied with the kit. The binding efficiency was calculated as $BE = 1 - \frac{c_S}{c_A}$, where c_S is the concentration of DNA in the supernatant of the bead containing the sample and c_A the concentration of DNA in the supernatant of the control sample. The degree of non-specific binding of DNA was calculated as $NS = 1 - \frac{c_A}{c_T}$ where c_T is the concentration of DNA initially added. The binding efficiency of 1.8 μ m long biotin/digoxigenin-DNA tethers was determined to be $8.8 \pm 0.5\%$ (mean $\pm$ SD, n=7), whereas the degree of non-specific binding was significantly higher with $20.2 \pm 0.87\%$.

NTA-Tether specificity for His-tagged proteins. To confirm the specificity of the NTA-DNA tether for His-tagged proteins, a similar approach as described in (2) was performed. Instead of attaching the NTA-DNA via the biotin handle to streptavidin-coated microplates, streptavidin-coated beads (20 μ l) were used with a DNA-to-bead ratio of 1000, 10,000, and 20,000. The respective amount of DNA was added to the beads with 100-1000x molar excess of NiCl₂ and incubated in WB on a rotisserie for 30 minutes. Next, the beads were washed twice with WB followed by resuspension in 20 μ l WB containing His-tagged CelE-CBM3a (3) at a concentration of 285 nM followed by incubation for 15 minutes. This resulted in a molar excess of protein between 10-200 with respect to DNA. To test whether the protein is immobilized to the beads via the His-tag, half of the bead samples were washed three times with WB, whereas the other half was washed three times with WB containing 500 mM imidazole to elute the protein from the Ni-NTA tethers. All samples were resuspended in 60 μ l working buffer containing 2mM pNP-cellobiose (4-Nitrophenyl- β -D cellobioside, Carbosynth Ltd) and incubated at 50°C for 24 hours with overhead mixing. A schematic of the experimental setup is shown in **Fig. S8-A**. The concentration of released pNP was determined by measuring the absorbance at 405nm and comparison to a standard curve. **Fig. S8-B** shows the hydrolysis of pNP-cellobiose as a function of the molar DNA-to-bead ratio (amount of CelE-CBM3a immobilized per bead). The conversion increases with increasing CelE-CBM3a density on the beads. Control experiments in which CelE-CBM3a-functionalized beads have been washed with imidazole show no significant conversion, thus verifying that the prepared DNA tether specifically binds to streptavidin-coated beads and His-tagged proteins.

Pull-down (solid-state depletion) assay for CBM3a-wt and Y67A mutant on NCC. Pull-down assays of CBM3a-wt and Y67A were performed in WB and B2 to investigate the effect of blocking agents on the binding affinity. The assays were carried out in clear 96-round bottom well plates in triplicates. To each well, 10 μ l of 1 mg/ml NCC for WT and 10 μ l of 10mg/ml of NCC for Y67A were added, followed by the addition of 90 μ l of protein at a concentration between 0.25 μ M and 4 μ M. Control experiments for non-specific binding to the wells without substrate at the same protein concentrations were prepared on the same plate. The plates were covered with parafilm and incubated on a thermomixer at room temperature for 3 hours and shaking at 500 rpm. After binding, 7 μ l of 5 M NaCl were added to precipitate the NCC, and the plates were centrifuged for 15min at 3200 g. The supernatant was transferred to an opaque 96-well flat bottom plate and the fluorescence was measured at 488 nm excitation with 509 nm emission and a 495 nm cut-off. The fluorescence intensity was converted to protein concentration based on a standard curve which

was prepared parallel to the experiment. The concentration data were converted to bound protein per g of NCC and fitted to the partition coefficient described in Equation S1.

$$B = \frac{n_{max}}{K_d} * F \quad (S1)$$

Where B represents the amount of bound protein per gram of NCC and F stands for the free protein concentration. The ratio of the maximum number of binding sites, n_{max} , and dissociation constant, K_d is referred to as the apparent partition coefficient and describes the distribution of protein between substrate and solution in equilibrium **Fig. S9** shows the binding data of CBM3a-wt and Y67A mutant in WB and B2 along with the fit Equation S1. The partition coefficient for CBM3a-wt is 6.5 ± 0.4 (mean $\pm$ SE) L/g and 6.4 ± 0.7 L/g for WB and B2, respectively and 0.73 ± 0.05 L/g and 0.69 ± 0.11 L/g for the Y67A mutant in WB and B2 respectively.

MD simulation system set-up. The protein was placed to position residues Y67, H57, R112 and W118 close ($3 \text{ \AA}$) to the surface of the cellulose fiber, with the Y67 pointing towards the reducing end of the fiber (-OH term of cellobiose). Atoms C4, C3 and C1 of terminal glucoses were positionally restrained to mimic an infinitely long fiber. The system was represented using the CHARMM36 force field (4), generated with the CHARMM-gui web (5) interface. The GLYCAM (6) force field was used for representing cellulose. The system was solvated using a CHARMM modified version of the TIP3 (7) water model and excess of charges were neutralized and overall ionic strength of 150 mM was set into the system using K+/Cl- ions. After energy minimization, the system was thermalized to 310 K for a short time (20 ns) using the GROMACS 2018.6 MD engine (8). Four snapshots from the equilibration procedure were later transformed into AMBER-formatted topology using the gromber tool of ParmEd from AmberTools 16 (9) for posterior MD production. MD production runs were carried using the AMBER 16 software (10). Water molecules were kept rigid with SETTLE (11), while other covalent bond lengths involving hydrogen were constrained with SHAKE (11) (tolerance $1/4 \cdot 10^{-6}$ nm) algorithm. Non-bonded Lennard-Jones (LJ) interactions were evaluated using an atom-based cut-off with forces switched smoothly to zero between 1.0 and 1.2 nm. Non-bonded Coulomb potentials were calculated using the smooth particle-mesh Ewald method (12) with Fourier grid spacing of 0.08–0.10 nm and fourth-order interpolation. Simulation in the canonical NpT ensemble was achieved by applying an isotropic coupling using a Monte Carlo barostat (10) method fixed at 1.01325 bars. Simulations were carried out at 310 K using a velocity Langevin dynamics algorithm with a coupling constant of 1 ps. Forces were integrated using a time step of 4 fs, which was enabled by hydrogen mass repartitioning. Nonbonded neighbor lists were built up to a distance cutoff of 1.4 nm and updated heuristically. Four independent replicates were run with a collective simulation time of 10 μ s. Trajectories were saved each 100 ps for analysis.

Supplementary Figures

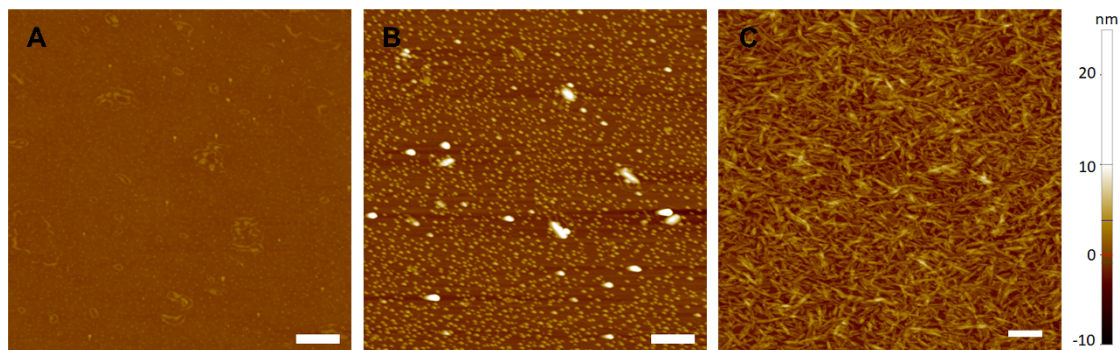


Fig. S1. AFM images of glass slides at different stages of the multilayer deposition process. A) bare surface after piranha treatment, B) poly-L-lysine (PLL) treated surface, C) 1x NCC layer on a PLL layer. The surface roughness factor (R_a) along randomly selected lines is 0.19 nm for the bare surface, 1.19 nm for the PLL treated surface, and 1.30 nm for one NCC layer on PLL. The scale bar is 500 nm.

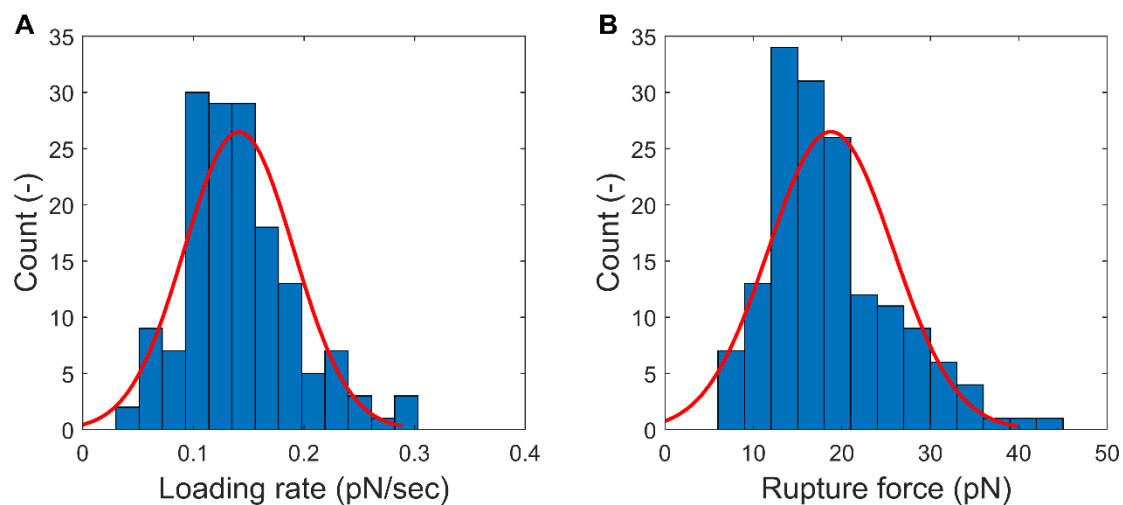


Fig. S2. Digoxigenin-anti-digoxigenin (DIG-aDIG) unbinding forces measured using acoustic force spectroscopy. Loading rate (A) and rupture force (B) histogram of DIG-aDIG (N=156). The fitted values (mean $\pm$ SD) are 0.14 ± 0.05 pN/s and 18.8 ± 7.0 pN respectively.

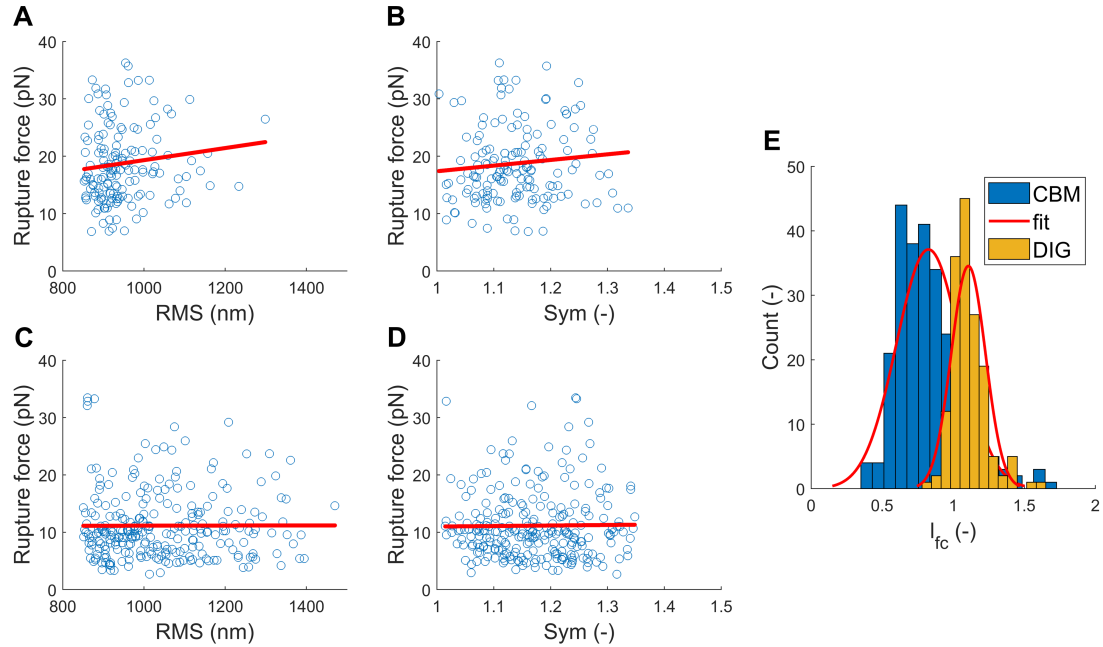


Fig. S3. No correlation is found between the rupture force and *RMS* or *Sym*, even though shorter tethers are observed on NCC surfaces. A-B) Scatter plots of rupture force with *RMS* and *Sym* of DIG-aDIG rupture measurements and C-D) Scatter plots of rupture force with *RMS* and *Sym* of CBM3a-wt- NCC rupture measurements at 1 pN/s. The red line indicates the best linear fit to the data. No significant correlation between rupture force and *RMS* or *Sym* is observed. Refer to **Table S3** and **Table S4** for correlation coefficients. E) Histogram of dimensionless length during force calibration (l_{fc}) for DIG-aDIG (yellow, N=156) and CBM3a-wt- NCC (blue, N=259). The red line shows the normal distribution fit to each data set and mean $\pm$ SD are 0.83 ± 0.23 and 1.1 ± 0.12 for CBM3a and DIG-aDIG, respectively. On average, a 25% reduction in measured length was observed for tethers on NCC surfaces.

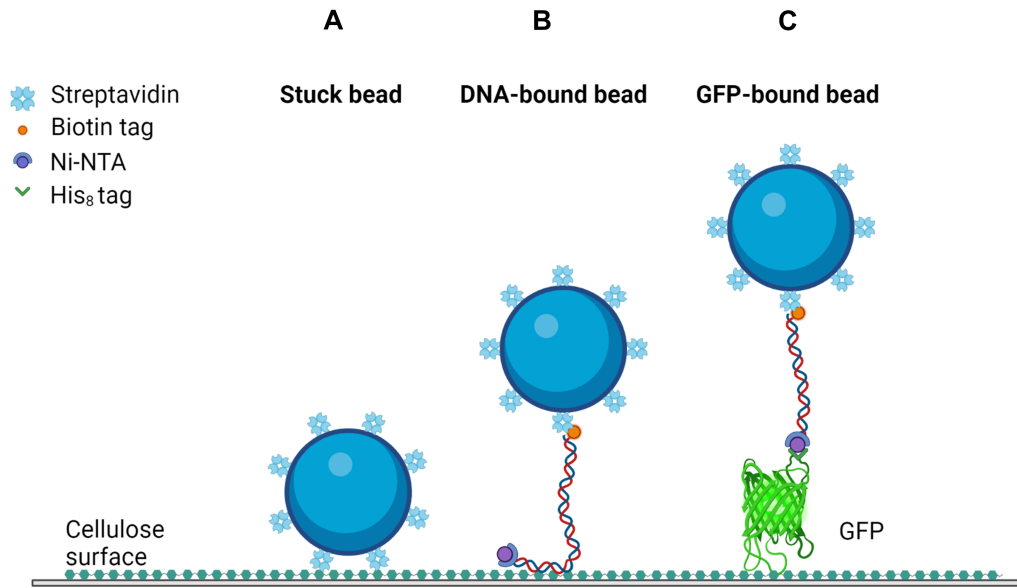


Fig. S4. Schematic of possible mechanisms for non-specifically bound beads to cellulose. A) Bead stuck directly to the surface usually shows a small RMS ($< 500\text{nm}$) and can easily distinguished from DNA-tethered beads. B) DNA non-specifically bound to the surface. Such beads show 4x lower binding probability on the NCC surface compared to CBM- tethered beads and the binding is weak such that most beads are removed during the flushing step prior to the experiment. C) GFP non-specifically bound to the surface. Similar to B) the binding probability is lower compared to CBM-tethered beads. A combination of all three situations is possible in the actual experiment.

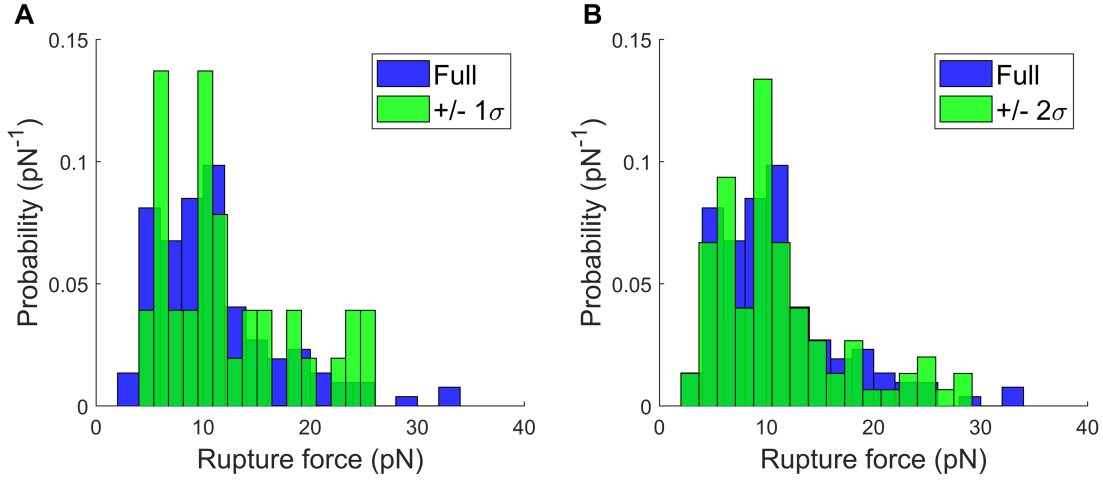


Fig. S5. Comparison of entire rupture force histogram of CBM3a-wt obtained at 1 pN/s (N=259) with reduced data sets. The criteria for inclusion in the reduced data set is an expected length during force calibration (l_{fc}) identical to l_{fc} obtained from DIG-aDIG experiments ($\mu \pm \sigma = 1.1 \pm 0.12$). The blue histogram represents the full data set, whereas the green histogram shows reduced data for A) $\mu \pm \sigma$ (N=37) and B) $\mu \pm 2\sigma$ (N=88). Qualitatively, the same shape of the histogram is obtained in both cases, thus indicating that shorter than expected DNA tethers are not affecting the rupture force measurement.

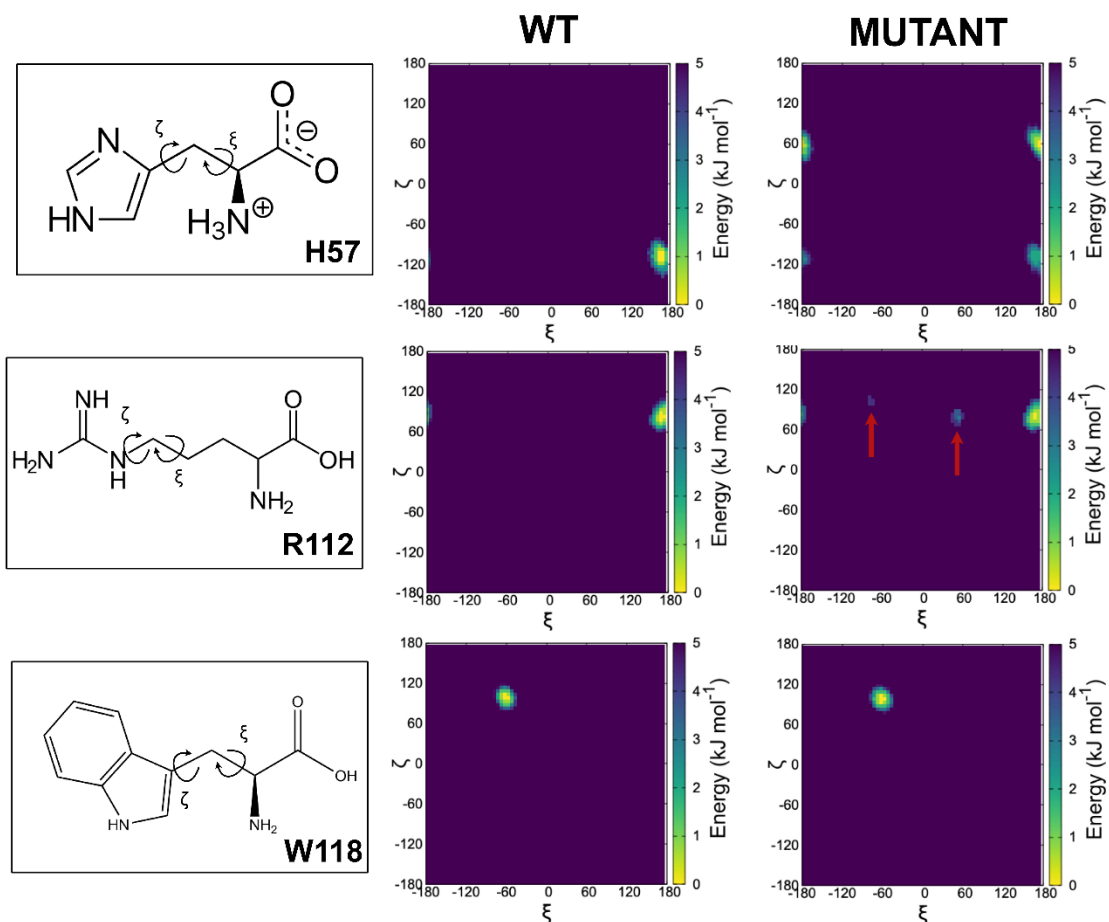


Fig. S6. Comparison of ζ and ξ torsion angles for H57, R112, and W118 between wild-type and Y67A mutant of CBM3a. While the ζ and ξ torsion angles for CBM3a-wt remain around one local minimum, some deviation from that minimum is observed for the Y67A mutant. The rotation of the histidine of $\sim 180^\circ$ (as defined by ζ) results in a lower energy minimum in the mutant due to the absence of the stabilizing Y67 residue. In addition, a greater variation in ξ -angles is observed for the mutant protein. R112 observes two more energy minima along the ξ -rotation axis compared to the wild-type (indicated by red arrows). W118 is not affected by the Y67A mutation as both torsion angles are similar for the wild-type and mutant proteins.

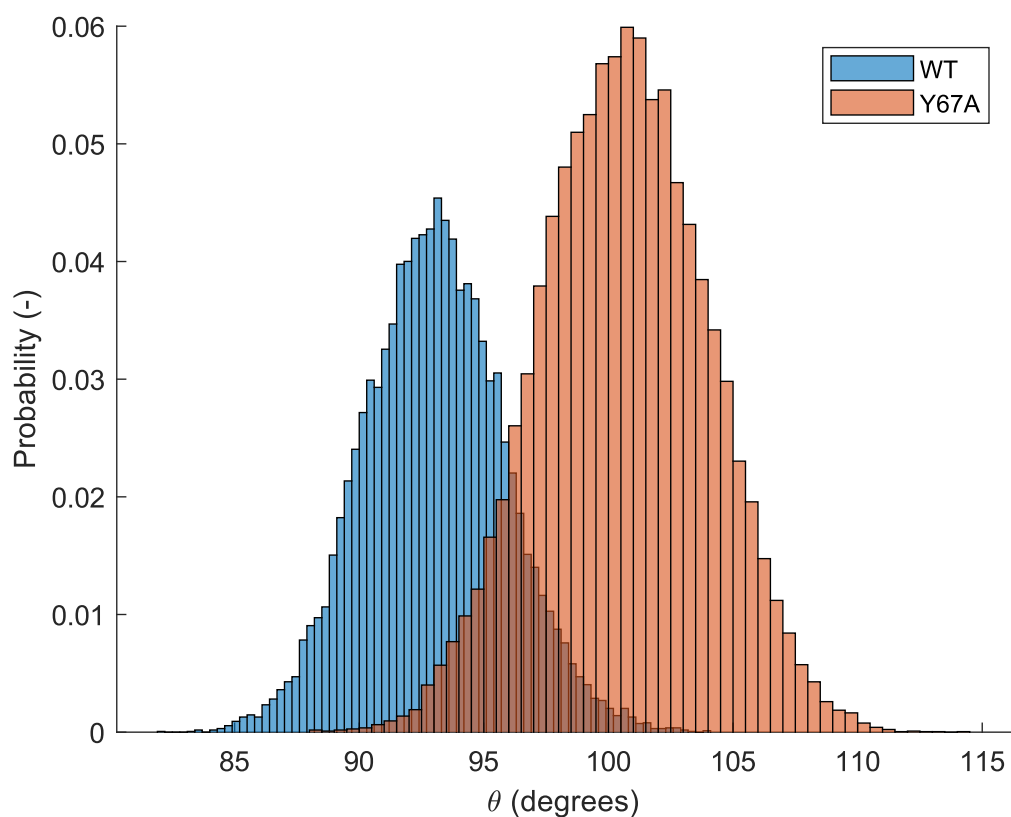


Fig. S7. Histogram of CBM orientations with respect to the cellulose surface sampled during MD simulations. The orientational angle (θ), as defined in the Main text Fig 6, increases for the Y67A mutant compared to the WT. N=16,346 for the WT and N=21,968 for Y67A respectively.

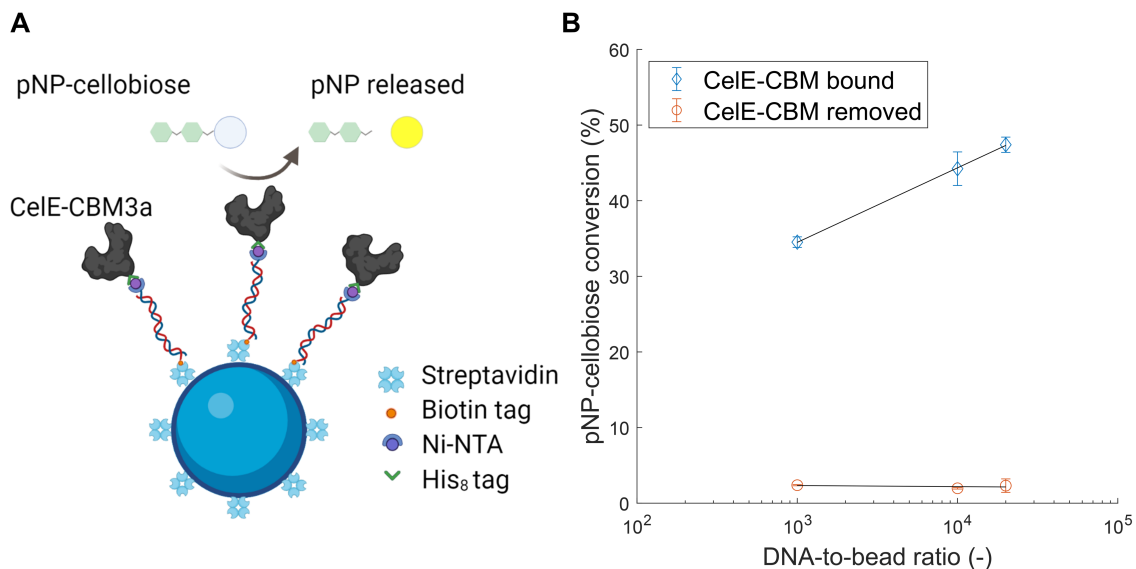


Fig. S8. Validation of binding specificity of Ni-NTA modified DNA tethers to His-tagged proteins. A) Schematic of the experimental setup. The DNA is connected to the bead by the streptavidin-biotin bond, whereas CelE-CBM3a is bound to the DNA by the Ni-NTA-Histidine bond. CelE-CBM3a hydrolyses the pNP-cellobiose to release pNP, which can be quantified by measuring the absorbance at 405 nm. B) Conversion of pNP-cellobiose as a function of molar DNA-to-bead ratio shows that samples washed with imidazole lose all the immobilized CelE and result in no significant conversion. Error bars represent the standard deviation of 3 experimental replicates. The line is the best logarithmic fit as a guide.

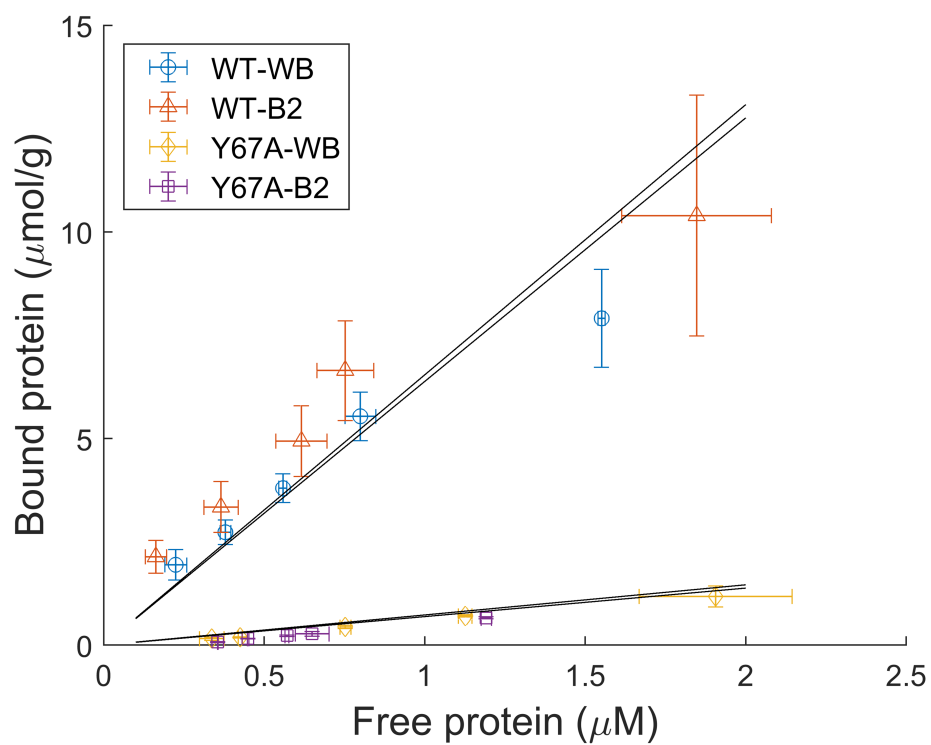


Fig. S9. Bulk ensemble binding data of CBM3a-wt and Y67A mutant to nanocrystalline cellulose. Binding data of CBM3a-wt in WB (o), B2 (Δ) and Y67A mutant in WB ($\diamond$) and B2 ($\square$). Error bars indicate the standard deviation of three experiments. Solid lines represent the fit of equation S1 to the data to obtain the partition coefficient.

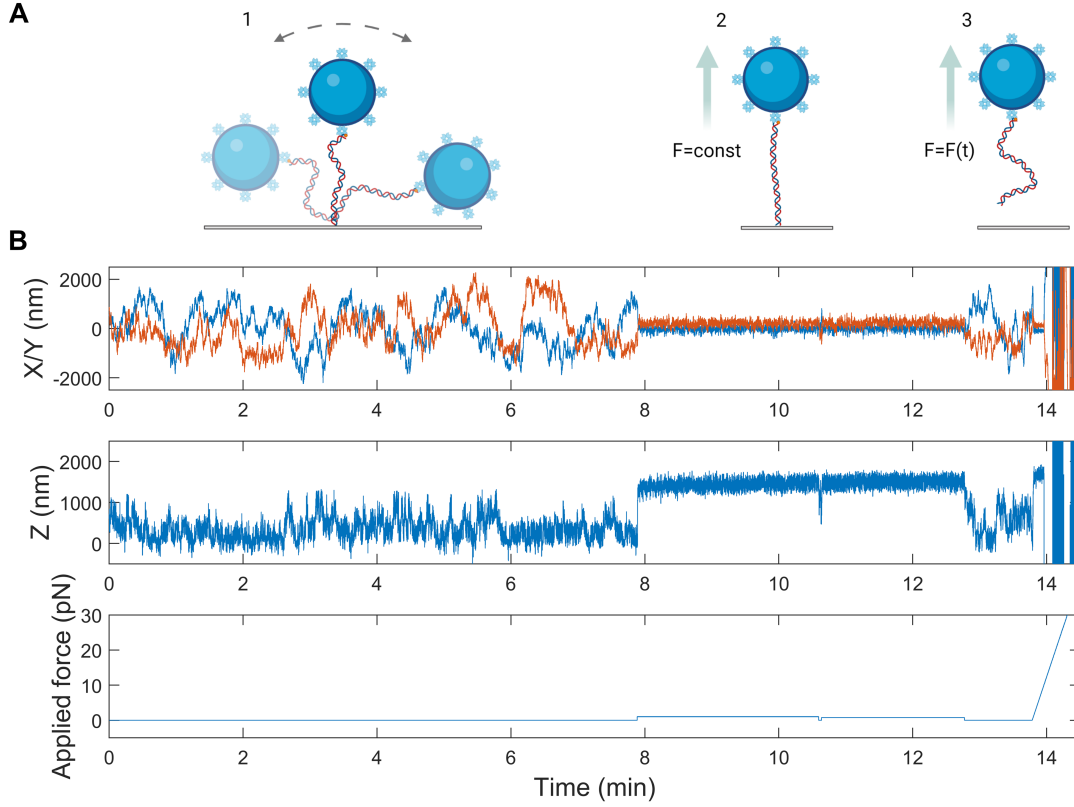


Fig. S10. Overview of a typical AFS single-molecule rupture force assay. A) Sketch of a tethered particle during anchor point determination (1), force calibration (2), and rupture during the force ramp (3). B) Example of a single-molecule trace recorded with the AFS. The anchor point was determined during the first 8 minutes at which the *RMS* and *Sym* values were calculated. Between minutes 8 and 12.5, the force was calibrated, and the measured extension was compared to the theoretical extension following the WLC model to yield l_{fc} . Finally, a linear force ramp was applied until the bond ruptured at approximately 14 minutes.

Supplementary Tables

Table S1. Pearson correlation coefficient with p-values in parenthesis tested between the measured rupture force and *RMS*, *Sym* and *l_{fc}* respectively. No significant correlation is observed except between rupture force and *Sym* for Y67A at 0.1pN/s (p=0.043).

	<i>RMS</i>	<i>Sym</i>	<i>l_{fc}</i>
DIG-aDIG	0.1163 (0.148)	0.0968 (0.229)	0.0672 (0.404)
WT 1pN/s	0.0024 (0.969)	0.0126 (0.840)	0.0109 (0.861)
WT 0.1 pN/s	0.0480 (0.545)	0.0382 (0.630)	-0.0128 (0.872)
Y67A 1pN/s	-0.083 (0.334)	-0.0741 (0.388)	-0.0268 (0.755)
Y67A 0.1pN/s	-0.0498 (0.533)	0.1609 (0.043)	-0.0214 (0.789)

Table S2. Spearman correlation coefficient with p-value in parenthesis tested between the measured rupture force and *RMS*, *Sym* and *l_{fc}* respectively. No significant correlation is observed.

	<i>RMS</i>	<i>Sym</i>	<i>l_{fc}</i>
DIG-aDIG	0.1081 (0.179)	0.1102 (0.171)	0.0582 (0.471)
WT 1pN/s	0.0217 (0.728)	-0.0158 (0.780)	-0.0004 (0.995)
WT 0.1 pN/s	0.0329 (0.679)	0.0274 (0.730)	-0.0185 (0.816)
Y67A 1pN/s	-0.1073 (0.210)	-0.0547 (0.524)	-0.0664 (0.439)
Y67A 0.1pN/s	-0.0229 (.0774)	0.1497 (0.060)	-0.0074 (0.926)

Table S3. Mean and standard deviation of double normal fit of the rupture force distribution for CBM3a-wt and Y67A mutant shown in Fig 4 in the main manuscript. $p(\mu_1)$ denotes the fraction of the first peak.

		$p(\mu_1)$ (-)	μ_1 (pN)	σ_1 (pN)	μ_2 (pN)	σ_2 (pN)
WT	1 pN/s	0.71	8.52	2.86	17.54	6.44
	0.1 pN/s	0.37	3.50	0.69	7.06	2.66
Y67A	1 pN/s	0.64	7.90	2.31	14.90	4.29
	0.1 pN/s	0.67	4.52	1.13	9.08	3.67

Table S4. Parts list for the microfluidic setup shown in Fig 2-B in the Main text.

Item Description	QTY	Part Number	Vendor
Pressure Controller	1	OB1-MKIII+	Elvesys S.A.S.
10-way Valve	1	MUX-D	Elvesys S.A.S.
50ml Fluid Reservoirs	1	KRM1	Elvesys S.A.S.
2ml Fluid Reservoirs	2	KRXS-V2	Elvesys S.A.S.
PTFE Tubing 1/16" OD (1/32" ID)	8m	KFSPPI	Elvesys S.A.S.
Low -Pressure Manifold Assembly, 6 port	1	P-152	IDEX Corporation
Adapter, Male Luer Lock x Female 1/4-28 Flat Bottom	4	P-675	IDEX Corporation
Adapter, Female Luer x Male 1/4-28 Flat Bottom	4	P-624	IDEX Corporation
Check Valve, Female Luer x Male Luer Lock	4	30505-92	Masterflex

Legend for MD simulation files in rtf format

File name	Description
Start_WT.pdb.rtf	Example of MD simulation configuration file of CBM3a wild type binding to a cellulose-I crystal at the beginning of the simulation
Final_WT.pdb.rtf	Example of MD simulation configuration file of CBM3a wild type binding to a cellulose-I crystal at the end of the simulation
Start_Y67A.pdb.rtf	Example of MD simulation configuration file of CBM3a-Y67A mutant binding to a cellulose-I crystal at the beginning of the simulation
Final_Y67A.pdb.rtf	Example of MD simulation configuration file of CBM3a-Y67A mutant binding to a cellulose-I crystal at the end of the simulation

Note: To convert files to pdb format, the file extension rtf must be removed from the file name.

SI References

1. E. W. A. Visser, L. J. Van Ijzendoorn, M. W. J. Prins, Particle Motion Analysis Reveals Nanoscale Bond Characteristics and Enhances Dynamic Range for Biosensing. *ACS Nano* **10**, 3093–3101 (2016).
2. J. Shimada, *et al.*, Conjugation of DNA with protein using His-tag chemistry and its application to the aptamer-based detection system. *Biotechnol. Lett.* **30**, 2001–2006 (2008).
3. C. K. Bandi, A. Goncalves, S. V. Pingali, S. P. S. Chundawat, Carbohydrate-binding domains facilitate efficient oligosaccharides synthesis by enhancing mutant catalytic domain transglycosylation activity. *Biotechnol. Bioeng.* **117**, 2944–2956 (2020).
4. J. Huang, *et al.*, CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2016).
5. S. Jo, *et al.*, CHARMM-GUI 10 years for biomolecular modeling and simulation. *J. Comput. Chem.* **38**, 1114–1124 (2017).
6. S. S. Mallajosyula, S. Jo, W. Im, A. D. MacKerell, “Molecular Dynamics Simulations of Glycoproteins Using CHARMM” in *Glycoinformatics*, T. Lütke, M. Frank, Eds. (Springer New York, 2015), pp. 407–429.
7. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
8. S. Páll, M. J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS in *Solving Software Challenges for Exascale*, S. Markidis, E. Laure, Eds. (Springer International Publishing, 2015), pp. 3–27.
9. R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).
10. S. Le Grand, A. W. Götz, R. C. Walker, SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations. *Comput. Phys. Commun.* **184**, 374–380 (2013).
11. S. Miyamoto, P. A. Kollman, Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
12. T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).