Molecular origins of reduced activity and binding commitment of processive cellulases and associated carbohydrate-binding proteins to cellulose III

Shishir P. S. Chundawat,[a]* Bhargava Nemmaru,[a] Markus Hackl,[a] Sonia K. Brady,[b] Mark A. Hilton,[b] Madeline M Johnson,[b] Sungrok Chang,[b] Matthew J. Lang,[b,c] Hyun Huh,[d] Sang-Hyuk Lee,[d] John M. Yarbrough,[e] Cesar A. López,[f] Sandrasegaram Gnanakaran[f]

[a]Department of Chemical & Biochemical Engineering, Rutgers The State University of New Jersey, Piscataway, New Jersey 08854, USA.
[b]Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, USA.
[c]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37235, USA.
[d]Department of Physics and Astronomy, Rutgers The State University of New Jersey, Piscataway, New Jersey 08854, USA.
[e]Biosciences Center, National Renewable Energy Lab, Golden, Colorado 80401, USA.
[f]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

*Corresponding Author: Shishir P. S. Chundawat (ORCID: 0000-0003-3677-6735)
Email: shishir.chundawat@rutgers.edu

Running title: Cellulase and CBM binding interactions with cellulose

Keywords: Carbohydrate-Binding Proteins, Cellulose, Optical Tweezers Force Spectroscopy, Molecular Dynamics, Cellulases, Carbohydrate-Active Enzymes, Lignocellulosic Biofuels

Abstract

Efficient enzymatic saccharification of cellulosic biomass into fermentable sugars can enable production of bioproducts like ethanol. Native crystalline cellulose, or cellulose I, is inefficiently processed via enzymatic hydrolysis, but can be converted into the structurally distinct cellulose III allomorph that is processed via cellulase cocktails derived from *Trichoderma reesei* up to 20-fold faster. However, characterization of individual cellulases from *T. reesei*, like the processive exocellulase Cel7A, show reduced binding and activity at low enzyme loadings towards cellulose III. To clarify this discrepancy, we monitored the single-molecule initial binding commitment and subsequent processive motility of Cel7A enzymes and associated carbohydrate-binding modules (CBM) on cellulose using optical tweezers force spectroscopy. We confirmed a 48% lower initial binding commitment and 32% slower processive motility of Cel7A on cellulose III, which we hypothesized derives from reduced binding affinity of the Cel7A binding domain CBM1. Classical CBM-cellulose pull-down assays, depending on the adsorption model fitted, predicted between 1.2 to 7-fold reduction in CBM1 binding affinity for cellulose III. Force spectroscopy measurements of CBM1-cellulose interactions, along with molecular dynamics simulations, indicated that previous interpretations of classical binding assay results using multi-site adsorption models may have complicated analysis, and instead suggest simpler single-site models should be used. These findings were corroborated by binding analysis of other type-A CBMs (CBM2a, CBM3a, CBM5, CBM10, and CBM64) on both cellulose allomorphs. Finally, we discuss how complementary analytical tools are critical to gain insight into the complex mechanisms of insoluble polysaccharides hydrolysis by cellulolytic enzymes and associated carbohydrate-binding proteins.

Introduction:

Plant biomass, composed of polysaccharides like cellulose, is an ideal feedstock for bioconversion into various bioproducts like ethanol (1, 2). Cellulose is a β-(1→4)-glucose polymer that self-assembles to form crystalline fibrils that are recalcitrant to enzymatic depolymerization (3). Cellulolytic microbes (like *Trichoderma reesei* and *Clostridium thermocellum*) have therefore evolved with enzymes called cellulases that can deconstruct cellulose into fermentable sugars (4–6). Cellulases are comprised of two or more polypeptide domains called catalytic domains (CDs) and CBMs (4). CBMs are characterized by a planar binding motif, that is complementary to crystalline cellulose fibril structure to facilitate CD activity towards insoluble and structurally heterogenous cellulosic substrates (7). Although CBMs facilitate CD activity by proximity based targeting effects, cellulolytic enzymes are inefficient for industrial applications often due to non-productive interactions with the substrate that necessitates high protein loading requirements (4, 8).

Thermochemical pretreatment using acids, bases, or ionic liquids is therefore employed to increase polysaccharide accessibility to enzymes and reduce non-productive cellulase binding (9–11). Pretreatment with anhydrous liquid ammonia results in conversion of native cellulose I to cellulose III allomorph (12), thereby improving hydrolytic activity of several fungal (13) and bacterial cellulase mixtures (14). However, processive exocellulases such as *Tr*Cel7A (or Cel7A from *T. reesei*) and *Tf*Cel6B (or Cel6B from *Thermobifida fusca*), that are workhorse cellulolytic enzymes, often show reduced activity on pretreated cellulose III for reasons poorly understood (14, 15). Although the processive mechanism of Cel7A on native cellulose I has been studied extensively using classical biochemical assays (16–19) and molecular simulations (20, 21), there is limited consensus on how to monitor initial enzyme association with the cellulose chain (18) or dissociation of non-productively bound enzymes (16, 17) to identify rate-limiting steps impacting cellulose hydrolysis. Hence, there is a need for better experimental methods that can track cellulase binding and processive motility in real-time with atomic-scale resolution for distinct substrates.

Single-molecule fluorescence imaging allows estimating exocellulase binding kinetics parameters (e.g., adsorption and desorption rates) (8, 22, 23), whereas high speed atomic force microscopy allows tracking motility of single cellulase molecules (24, 25). However, these methods cannot resolve the slower sub-nanometer translational

rates of processive cellulases relevant to cellulose decrystallization and hydrolysis into cellobiose. We recently reported an optical tweezers force spectroscopy-based cellulase assay technique to track the single-molecule motility of Cel7A on native cellulose with sub-nanometer and millisecond resolution (26). Interestingly, Cel7A CD in the absence of CBM1 showed lower dwell times between catalytic turnover steps suggesting that CBMs could impede full-length cellulase motility on native cellulose I due to non-productive binding. However, we lack a detailed understanding of the mechanistic role of CBMs in full-length processive cellulase binding and motility on cellulose I and other industrially relevant cellulosic substrates like cellulose III.

Here, we have applied our optical tweezer assay to investigate the initial binding stability of Cel7A and its processive motility on cellulose I and cellulose III. To understand the role of CBMs in our observed single-molecule binding instability of Cel7A towards cellulose III, we characterized the binding of CBM1 (from Cel7A) using classical 'pull-down' binding assays and molecular dynamics simulations. We also developed a new optical tweezers based CBM-cellulose bond 'rupture' assay to characterize the binding behavior of single CBM1 proteins alone to distinct cellulose allomorph surfaces under applied force. To generalize these findings further, we characterized CBM3a (another Type-A CBM from *Clostridium thermocellum*) using equilibrium 'pull-down' and kinetic binding assays. We also characterized the binding partition coefficient of several other Type-A CBMs belonging to Family 2a, 5, 10, and 64 to confirm that Type-A CBMs in general showed reduced binding towards cellulose III. Our results highlight some of the challenges associated with the use of overly simplistic Langmuir-type models to analyze classical protein-polysaccharide 'pull-down' assay dataset. In summary, our work highlights how changes in CBM binding to distinct cellulose allomorphs can critically impact processive cellulase motility. Furthermore, our work highlights the necessity of using a multifaceted approach for characterizing the binding heterogeneity and multimodal nature of cellulase-cellulose interactions.

Results

***Trichoderma reesei* cellulase mixture shows improved activity towards cellulose III:** *Cladophora* sp. (*Cladophora glomerata*) derived highly crystalline cellulose I fibers were isolated, as described previously (26), followed by anhydrous liquid ammonia pretreatment to prepare cellulose III (27). Details about cellulose isolation, ammonia pretreatment, spectroscopic characterization and enzymatic hydrolysis methods are provided in the **SI Appendix Experimental Procedures** section. Spectroscopic characterization using X-ray diffraction (XRD) and Fourier Transform Raman spectroscopy (FT-Raman) were conducted to confirm the conversion of cellulose I to cellulose III allomorph following ammonia pretreatment and also measure substrate characteristics like cellulose crystallinity index (CrI) and crystallite size. Similar to previous work (27–30), XRD equatorial reflections for (100), (010), and (110) crystallographic planes for native Cladophora cellulose I were at approximately 14.9°, 17.1°, and 23.0° Bragg angles (2ϴ), respectively (see **Figure 1A**). As previously described (27, 31), equatorial reflections for (010), (002), and (100) crystallographic planes for Cladophora cellulose III were at approximately 11.8°, 17.4°, and 20.9° Bragg angles (2ϴ), respectively. Based on the Segal method, cellulose crystallinity index was estimated to be about 90-95% for both allomorphs. Cellulose crystallite size was about 8.5-9 nm for both allomorphs, estimated using the Scherrer equation based on the full-width half-maximum of the equatorial plane reflection peak. See **SI Appendix Fig. S1** for an atomic force microscopy (AFM) based analysis of individual crystallite fibers which also agrees with the XRD results and expected crystallite shape as reported in previous AFM based analysis of Cladophora derived cellulose (23, 25). Cladophora cellulose based crystallites were at least 2-3 times larger in cross-sectional diameter than previously reported for cellulose microfibrils derived from higher-order plants, such as cotton linters (15). Similar to previous reports (15, 32, 33), Raman spectroscopy also independently confirmed that native Cladophora cellulose I was completely

converted into cellulose III following ammonia treatment (**Figure 1B**).

Next, we performed enzymatic hydrolysis assays to test the activity of a *T. reesei* based commercial cellulase enzyme mixture (i.e., Cellic C.Tec2) towards cellulose I and cellulose III allomorphs. The commercial cellulase enzyme mixture showed ~3-fold improved activity toward cellulose III vs cellulose I at the 24 h and 96 h saccharification time points (see **Figure 1C**). We also confirmed that a purified mixture of *T. reesei* endo and exo-cellulases Cel7B and Cel7A respectively, show up to 10 to 20-fold improved activity toward cellulose III (see **Figure 1D**). These results support our previous observations that improved activity of cellulase mixtures towards cellulose III, arises due to improved endo-exo synergistic activity (14). Cellulose III has a slightly stepped or 'jagged' surface due to the underlying modification of the crystal structure caused by *trans-gauche* to predominantly *gauche-trans* rotameric state of the C6-hydroxymethyl groups (see **Figure 1E**). This jagged cellulose III surface has been shown previously to be more readily hydrated by water molecules, unlike cellulose I (13, 15), and was therefore hypothesized to impact cellulolytic enzyme binding and/or activity. Here, we also characterized the specific activity of purified Cel7A alone towards cellulose I and cellulose III at various enzyme loadings of 0.5, 2.5, and 10 mg/g (see **SI Appendix Table S1**). These results show up to 3-fold improved enzyme activity toward cellulose III at higher enzyme loadings, such as 2.5 and 10 mg/g, however that difference becomes nearly indistinguishable at the lowest enzyme loading of 0.5 mg/g, similar to activity trends previously observed by Gao et al. (2013) and Shibafuji et al. (2014). The underlying molecular origins for decrease in processive bulk activity of Cel7A towards cellulose III at very low enzyme loadings is not clear currently. Previous single-molecule Cel7A motility assays have been conducted at high enzyme loadings where Cel7A 'traffic jams' and poorly understood protein-protein interactions seem to play an important role in cellulose hydrolysis by cellulases (22). However, the activity of Cel7A on cellulose III in the absence of such surface crowding effects at the single-enzyme level has not been characterized using high-resolution optical tweezer based tracking methods.

Single-molecule Cel7A binding and initial substrate engagement is impaired on cellulose III: Single-molecule cellulase motility assays were performed on both cellulose allomorphs to study how subtle differences in cellulose crystal structure impact the binding and processive motility of Cel7A. Details regarding Cel7A motility assay and data analysis rationale are published elsewhere (26). Briefly, Cel7A was attached via sulfo-SMCC (i.e., sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate) cross-linking to a thiol tag on the end of a biotinylated 1010 bp DNA tether and attached to a 1.25 μm streptavidin-coated polystyrene bead (see **Figure 2A**). The Cel7A functionalized bead was positioned directly above a cellulose fiber to initiate binding and the bead position was monitored as the enzyme first bound, hydrolyzed, and processed along the cellulose surface for cellulose I or cellulose III fibril surface. Based on the mechanism for Cel7A action on cellulose (see **Figure 2B**), we propose the term 'motility commitment' or 'binding commitment' to describe the steps prior to initiation of processive motility, i.e. binding, recognition and initial cellulose chain threading within the Cel7A active-site tunnel. During our motility assays, it was possible for us to observe the initial motility commitment of Cel7A for distinct cellulose allomorph surfaces immediately prior to processive motility initiation. To initiate the single-molecule Cel7A motility, a functionalized bead is positioned directly above a surface-affixed cellulose fiber and periodically gently pulled via the piezo stage to test for bound enzymes. Such initial binding is considered stable or committed when the Cel7A-cellulose bond survives, and the enzyme exhibits motility for a period greater than 10 s. Representative traces of binding stability/instability for Cel7A binding to cellulose I and cellulose III are shown in **Figure 2C** and **Figure 2D** respectively (see **SI Appendix Fig. S2** for additional representative traces). In some cases, the full-length Cel7A was seen to bind but not commit to significant motility on the cellulose surface highlighting non-productively engaged cellulases. Alternatively, Cel7A-cellulose bond instability is revealed through initial bead displacement followed by rapid detachment. Given this criteria and observation times of 600s for each trace, Cel7A-cellulose initial bond instability was

determined to be significantly lower for cellulose I (12% of all traces, N=17) than cellulose III (23% of all traces N=13). Although this rapid bead detachment as shown in **Figure 2D** could have been driven in principle either due to improper CBM and/or CD binding/engagement, the large 100 nm spikes in the highlighted region (labeled 'unstable binding') led us to hypothesize that the CBM likely plays a prominent role in this phenomenon owing to its primary function of increasing proximity of CD near cellulose surface (to within a few nm). Furthermore, we analyzed the subsequent processive motility cycles of Cel7A by extracting the enzymes step sizes and dwell time distributions as further discussed below.

Cel7A shows reduced hydrolytic velocity and longer dwell times between catalytic cycles on cellulose III: Representative individual Cel7A processive motility traces and average enzyme velocity on cellulose I and cellulose III are shown in **SI Appendix Fig. S3** and **Figure 3A** respectively**,** which capture the processive motion of single enzymes on the cellulose surface during its deconstruction into soluble sugars (namely cellobiose). The average Cel7A velocity on cellulose I was 0.25 ± 0.35 nm s^{-1} (s.d.; N=68 motility traces), which is marginally higher than that seen on cellulose III, 0.17 ± 0.14 nm s^{-1} (s.d.; N=30 motility traces). The dwell time and step size distributions were then extracted for each individual motility trace as described previously (26) and highlighted in **Figure 3B**. Extraction of the step size distributions from individual motility traces (see **Figure 3C** for step size distributions on cellulose I (red) and cellulose III (black) overlaid) indicated that the mean step size for both cellulose I and cellulose III is close to the 1 nm length of the expected cellobiose product. However, the dwell time for cellulose III was 0.92 s as compared to 0.75 s for cellulose I (**Figure 3D**). The increased dwell time, frequent reverse-stepping or back-motility, and marginally reduced forward enzyme velocity observed on cellulose III vs cellulose I partially explains the lowered or comparable Cel7A bulk saccharification activity observed previously towards cellulose III at very low enzyme loadings (15). In summary, Cel7A shows impaired motility commitment (or initial binding) and slightly reduced processive motility (or hydrolytic velocity) on cellulose III. We hypothesize that processive cellulases like Cel7A show reduced binding/activity towards cellulose III likely due to impaired motility commitment driven by unstable binding to the cellulose surface. As shown in **Figure 2B**, the first step of motility commitment involving enzyme binding to cellulose is primarily driven by the CBM (34). Although the catalytic domain (CD) is responsible for processive motility, the CBM likely also plays a critical role by stepping in tandem with the CD (35). Hence, the rest of this study was aimed towards better understanding the role of CBMs in anomalous motility commitment and processive motility behavior on cellulose III, using a complementary suite of experimental and computational methods.

CBM1 isolated from Cel7A displays lower binding affinity towards cellulose III: Cel7A possesses a CBM from family 1 (called CBM1 hereon), whose structure-function relationships have been well characterized (36–38). However, CBM1 binding towards non-native allomorphs such as cellulose III has not been studied in detail. CBM1 orients and binds to crystalline cellulose I through strong hydrophobic stacking interactions between conserved planar aromatic residues (Y5, Y31, Y32) and axially-oriented hydrogen moieties of individual glucosyl units of the cellulose polymer chain (39), as illustrated in **Figure 4A**. Here, we characterized the equilibrium binding interactions of CBM1 towards cellulose I and cellulose III using solid-state depletion or classical protein-polysaccharide 'pull-down' binding assays (40). CBM1 was tagged with green fluorescent protein (GFP) to allow protein quantitation based on fluorescence as described previously (41). Details regarding gene sequences, cloning, expression, and protein purification strategies for all CBMs tested in this study can be found in the **SI Appendix Experimental Procedures** section (41).

Classical 'pull-down' binding assays employing an extensive range of protein concentrations (0 - 250 μM) resulted in protein-polysaccharide adsorption dataset for CBM1 as shown in **Figure 4B**. Langmuir one-site/two-site and Langmuir-Freundlich based adsorption models (equations displayed in **Figure 4B**) were fitted to the adsorption dataset using non-linear regression, as described previously (7, 13, 40). The model-fitting outputs for all models tested here are shown in **SI**

Appendix Fig. S4. This analysis allowed estimation of the maximum available binding sites ($n_{\max}$) and equilibrium dissociation constant (K_d), in addition to other model-specific parameters (**Table 1**). The total number of binding sites for cellulose I was always higher (~1.2-1.5 fold) compared to cellulose III in all cases, except in the case of high-affinity binding sites ($n_{\max}$) for the two-site model. There was ~1.2 to 7-fold reduction in binding affinity (i.e., inverse of dissociation constant K_d) for cellulose III depending on the exact fitted model. Our analysis indicates that the exact fold reduction in CBM binding affinity for cellulose III is highly dependent on the model used and highlights a potential limitation of classical binding assay methods. To further highlight limitations of the classical assay methods, we also performed data truncation analysis by trimming down our binding dataset for CBM1 to exclude higher protein concentrations (i.e., included maximum concentrations up to 15 μM or 50 μM only) (see **Table 2**). We observed that the number of predicted binding sites for both cellulose I and cellulose III decreased by ~1.3 to 1.8-fold for the truncated datasets. Interestingly, our truncated dataset fitted models predicted a slightly weaker affinity of CBM1 towards cellulose I versus cellulose III, which was contrary to predictions made from model fitting to the full dataset. Hence, to resolve this apparent uncertainty in relative binding affinity trends due to limitations of classical binding assay methods, we resorted to potential of mean force (PMF) calculations to also theoretically estimate the CBM1-cellulose binding affinity using a first-principles approach.

Molecular simulations predict lower CBM1 binding free energy towards cellulose III: Unbiased MD simulations were first performed to obtain the preferred binding orientation of CBM1 on model cellulose I and III crystal surfaces (see **SI Appendix Fig. S5**). As shown in **SI Appendix Fig. S5C-D**, the CBM1 planar binding surface aromatic residues exhibit greater root mean square fluctuation (RMSF) on cellulose III, indicating improper stacking of aromatic residues specifically the Y5 residue. The results from unbiased MD simulations are discussed in detail in the **SI Appendix Results and Discussion** section. A potential of mean force (PMF) was then calculated to estimate the CBM1 binding free energy during adsorption to the hydrophobic surface of both cellulose allomorph models. As shown in **Figure 5**, in the case of cellulose I, only one PMF energy minimum well was observed corresponding to the dominant CBM1-cellulose configuration observed during the unbiased MD simulations whereby the Y31 residue faces the non-reducing end (i.e., the expected canonical orientation based on native Cel7A favored activity from non-reducing end of cellulose). However, in the case of highly crystalline cellulose III, two PMF energy minima wells were observed, one in which Y31 faces the reducing end closer to the surface and another in which it faces the non-reducing end further away from the surface. These configurations are annotated as non-canonical and canonical, respectively, in **Figure 5**. These two configurations are separated by roughly 0.2 nm in the PMF free energy diagram, where the distance is measured normal to cellulose surface, with a marginal energetic barrier of 2 kcal/mol separating the two minima wells. A closer examination of the CBM1 structure revealed that if the protein binds in the so-called 'canonical' orientation to the cellulose III surface at a shorter distance, then the Y5 residue exhibits significant steric clashes with the cellulose III adjacent surface chains (also shown in **SI Appendix Fig. S5D**). The impact of such steric clashes is also captured in the higher RMSF values observed for the key binding motif aromatic residues when CBM1 is weakly bound to cellulose III. This explains why the 'canonical' CBM1 configuration is observed only at slightly longer distances away from the cellulose III surface. Irrespective of the preferred orientation for CBM1 to cellulose III surface and the degree of model cellulose III crystallinity, the calculated free energy of binding for CBM1 was always lower for cellulose III compared to cellulose I. These results support predictions from Langmuir adsorption models where the estimated equilibrium binding affinity for CBM1 was lower for cellulose III than cellulose I.

Family 3a CBM also shows reduction in binding to cellulose III via both equilibrium and kinetic binding assays: To generalize our findings regarding reduced CBM binding affinity towards cellulose III beyond CBM1, we also characterized the equilibrium and kinetic binding behavior of another well-studied type-A CBM from family 3a

(also called CBM3a) from *C. thermocellum* (42, 43). Classical binding assays and adsorption model fitting analysis were performed in a similar way to CBM1 (see **Table 3** for binding parameters and **SI Appendix Fig. S6** for model fitting outputs). Reduced binding affinity (~ 2 to 14-fold higher K_d) was observed for cellulose III depending on the exact model used for data fitting. The total number of binding sites predicted for cellulose I was slightly lower than cellulose III except when the dataset was fitted using a one-site model or in the case of high-affinity sites in two-site model. A closer inspection of the binding assay dataset (**SI Appendix Fig. S6**) suggests that even at the highest CBM3a concentrations tested (~50 μM), proper saturation behavior was not fully observed which might lead to spurious binding parameters as previously discussed during truncation analysis of CBM1 binding data. An alternative approach is to characterize the partition coefficient which is the linear slope of binding isotherm at lower protein loadings (as shown in inset of **Figure 4B**). Here, we also characterized the partition coefficient of a larger library of Type-A CBMs (including CBM1 and CBM3a) and observed a clear reduction in binding towards cellulose III in all cases (see **SI Appendix Fig. S7** for raw data and **SI Appendix Fig. S8** for partition coefficient bar graph). Partition coefficient is the slope of initial linear region of the binding curve between bound protein (μmol/g cellulose) and free protein (μM) as shown in **Fig. 4B.**

We further characterized the binding kinetics of CBM3a using fluorescence recovery after photobleaching (FRAP) and quartz crystal microbalance with dissipation (QCM-D). The raw data from FRAP and QCM-D assays is summarized in **SI Appendix Fig. S9** and **S10,** respectively. Briefly, similar to previous work (7, 44), GFP-CBM3a binding kinetic parameters to cellulose allomorphs were obtained by fitting the FRAP curves to a binding-dominated model ignoring any diffusion relevant contributions. Our FRAP analysis revealed that CBM3a gave a 1.9-fold increase in the desorption rate constant (k_{off}) for cellulose III compared to cellulose I (**Figure 6A**). Similarly, QCM-D also showed ~3-fold increase in k_{off} for Avicel derived cellulose III nanocrystals (**Table 4**). A detailed discussion of FRAP and QCM-D results can be found in the **SI Appendix Results and Discussion** section. We were also able to fit another parameter F_M which represents the fraction of reversibly bound GFP-CBM3a, that was used along with the desorption rate constant to draw conclusions about the relative change in adsorption rate constants as discussed in the SI appendix (**Figure 6B**). A key limitation of these assays is the inability to estimate the true adsorption rate constant (k_{on}), however, the desorption rate constant showed a clear increase for cellulose III, corroborating the reduction in binding affinity as indicated by the solid-depletion assays. Overall, these results indicate that Type-A CBMs like CBM1 and CBM3a show reduced binding affinity towards cellulose III, potentially leading to impaired motility commitment of tethered processive cellulases. Since classical binding assay methods cannot resolve the various binding modes of CBM-cellulose interactions and how these modes differ in the case of cellulose III vs cellulose I, we developed a single-molecule CBM-cellulose bond rupture assay. In addition, results from this single-molecule assay can shed light on the suitability of using multi-site models for analyzing classical 'pull-down' binding assays.

Single-molecule CBM-cellulose bond rupture assay reveals multimodal nature of CBM binding: Here, we designed an optical tweezers-based CBM-cellulose bond rupture assay under applied force to systematically characterize the binding behavior of CBM1 (from Cel7A) towards Cladophora cellulose I and cellulose III. Our tweezer CBM-cellulose assay design is similar to the Cel7A enzyme motility assay as reported in **Figure 2A**. Here, instead of Cel7A, GFP-CBM1 was tethered via a 1,010-bp DNA tether and attached to a 1.09 μm streptavidin-coated polystyrene bead (**Figure 7A**). Cellulose fibers were affixed to a glass coverslip. For each single CBM-cellulose 'rupture' assay run, individual beads were optically trapped and placed in the immediate vicinity of individual cellulose microfibers to facilitate a non-covalent CBM-cellulose bond formation (**Figure 7B**). Upon stable non-covalent bond formation, the stage was moved to a fixed position to pull the DNA-tether taut and exert a force on the CBM-cellulose bond. Total bond lifetime and rupture force were then calculated for each individual CBM-cellulose interaction till bond rupture took place (**Figure 7C**).

Hundreds of rupture events from individual assay runs were pooled and binned at 2.5 pN intervals for cellulose I and cellulose III to generate force-lifetime distribution plots (**Figure 7D**). Raw force-lifetime scatterplots are provided in **SI Appendix Fig. S11A**. Averaging all rupture events, we find that the mean lifetime of CBM1 binding to cellulose I was 1.41 ± 0.20 s (SEM; N=410) and to cellulose III was 1.11 ± 0.12 s (SEM; N=214). Since the bond rupture lifetime under applied force is related to the equilibrium binding off-rate, our rupture assay results are corroborated by the weaker binding affinity of CBM1 estimated by both pull-down assay dataset as well as the PMF calculations. More importantly, our bond rupture mean lifetime results suggest that simple one-site Langmuir adsorption models are more appropriate than complex multi-site adsorption models to estimate the marginal differences in CBM1 binding affinity for distinct cellulosic allomorphs. Note that the standard deviation of lifetimes of the CBM1-cellulose I and CBM1-cellulose III bonds were 4.12 s and 1.82 s, respectively. While marginal differences can be seen at lowest (0-2.5 pN) and highest (17.5-20 pN) rupture force ranges, one-way ANOVA test suggests that the lifetime dataset over the entire rupture force range is not statistically different (**SI Appendix Fig. S11B**).

Furthermore, while the average lifetimes show different profiles, there was also a broad spread in the distribution of observed bond lifetimes with a great deal of overlap between cellulose I and cellulose III indicating that multiple binding states with distinct characteristic bond lifetimes are possible for CBM1 binding to both cellulose I and III. As seen previously for protein-ligand interactions in other single-molecule studies (45), CBM-cellulose binding was expected to show classic slip-bond behavior; i.e., as the rupture force increases, the total bond lifetime decreases. However, fits to the force-lifetime distribution failed to converge to a single exponential decay suggesting that multiple binding modes are likely present for CBM-cellulose. A classical unimodal slip bond would exhibit a single exponential decay (46), therefore it suggests that CBM1 does not follow this simple model when interacting with either cellulose allomorph. Binding of CBM1 on cellulose instead revealed a spread with a more complex multimodal and heterogenous binding behavior. This multimodal distribution was independent of the source of cellulose and similar results were also seen with filter paper derived cellulose fibrils (**SI Appendix Fig. S12**). We also performed controls to test for artifacts associated with full anti-His antibody versus Fab fragment binding but there was no significant difference seen in the multimodal distribution of the force-lifetime results (**SI Appendix Fig. S13**). Interestingly, the multimodal distribution of the force-lifetime was sensitive to the CBM structure as illustrated by the differences in rupture force-lifetime distribution seen for wild-type CBM1 and its Y31A mutant which has a minor modification to the planar aromatic binding residue (**SI Appendix Fig. S14**). Interestingly, while the overall lifetime dataset over all rupture forces tested shows no significant difference based on the one-way ANOVA result, there seems to be significant difference in the bond lifetimes over certain rupture force ranges. The Y31A mutation is known to significantly lower CBM1 bulk-ensemble binding affinity towards native cellulose I (39), but its unknown how this single mutation impacts the processive motility of the full-length Cel7A enzyme. Although these slight differences in CBM bond lifetimes might contribute to the reduced single-molecule velocity or initial binding commitment of Cel7A, the interactions of catalytic domain with this substrate may play an equally important role.

We speculated that the observed multimodal distribution seen for the force-lifetime dataset indicates multiple classes of overlapping binding modes with contributions from different cellulose substructures (47) namely crystalline regions with varying degrees of disorder, different crystal binding faces (48), and varied binding orientation/modes of CBM binding on the hydrophobic face of crystalline cellulose (as summarized in **Figure 8**). However, due to the highly crystalline nature of our Cladophora derived cellulosic substrates (with ~90-95% crystallinity index) and the previous observations that CBM1 likely binds predominantly to one preferred cellulose crystalline face (48), we hypothesize that the multimodal distribution in the force-lifetime dataset could also arise from multiple equilibrium binding modes of CBM1 with distinct orientations on the preferred cellulose binding surfaces (see discussion section below and supporting Monte

Carlo simulation results highlighted in **SI Appendix Figure S16** and **Table S3**). It is also likely that some of these CBM orientations are productive for catalysis whereas some orientations are non-productive. For Cel7A to perform a successful processive step, the CBM needs to step in tandem with the CD along a cellulose chain (35). However, if the CBM orients itself in non-productive orientations (across adjacent cellulose chains, for instance), we speculate that this could lead to increased dwell times for full-length Cel7A as seen on cellulose III. Additional mutant full-length Cel7A assays are necessary to unravel molecular origins of such multimodal binding behavior during cellulase catalytic turnover cycles.

Discussion

Pretreatments can increase cellulose accessibility to facilitate efficient enzymatic saccharification (49). Extractive ammonia (EA) pretreatment converts cellulose I to cellulose III to reduce biomass recalcitrance towards enzymatic hydrolysis. EA pretreatment achieves cellulosic biomass hydrolysis yields equivalent to its precursor ammonia fiber expansion (or AFEX) pretreatment but with 60% lower enzyme loading requirements (e.g., 18.75 mg enzyme/g cellulose for AFEX versus 7.5 mg/g for EA treated biomass hydrolyzed using commercial enzyme mixture consisting of 50% C.Tec2, 25% H.Tec2 and 25% Multifect Pectinase on a total protein basis) (12). However, there is a need to further reduce total enzyme loading equivalent to the range employed in a commercially viable corn starch liquefaction process using amylases (e.g., less than 1 mg amylase/g starch). One approach to reduce enzyme loading is to identify the potential rate-limiting enzymes in a complex cocktail critical for cellulose III hydrolysis. Endocellulases have been identified to show improved activity towards cellulose III, at various enzyme loadings tested, but concomitantly also showing lower binding to the substrate unlike cellulose I. But surprisingly, exocellulases like Cel7A (*T. reesei*) and Cel6B (*T. fusca*) have mostly shown lower or comparable activity on cellulose III versus native cellulose I, particularly at ultra-low enzyme loadings as reported in this study. While this is not detrimental to the action of cellulase enzyme mixtures, as both fungal and bacterial derived endo- and exo-cellulase mixtures have shown overall improved activity (up to 10-fold as reported here) towards cellulose III versus cellulose I largely due to increased endo-exo cellulase synergy (14, 15), there is clearly room for making improvements in enhacing processive cellulase activity towards cellulose III. Both endo- or exo-cellulases were previsouly reported to exhibit lowered binding towards cellulose III during saccharification. While these results can be explained based on the Sabatier principle recently applied to modeling cellulase action on cellulose (19), since tighter cellulase binding to cellulose need not always correspond to improved activity (13), we still lack a first-principles mechanistic basis for the reduced binding of most full-length cellulases observed to-date towards non-native cellulose III allomorph using advanced optical tweezers based single molecule assays (22).

Here, we developed and applied a single-molecule optical tweezer-based assay which allowed us to distinguish the initial enzyme binding commitment to repeated processive motility cycles that forms the basis of catalytic turnover of processive cellulases like Cel7A. Our results indicate that full-length Cel7A show impaired single-molecule motility commitment towards cellulose III which was hypothesized to arise due to unstable initial binding predominantly driven by the CBM. This hypothesis is also in alignment with reduced overall binding observed previously of full-length Type-A CBM based cellulases to cellulose III (13, 15). Secondly, the Cel7A motility assay showed marginally lower processive velocity on cellulose III than cellulose I, which is consistent with the classical bulk activity assays conducted at very low enzyme loadings that confirmed no significant difference in Cel7A activity on either cellulose allomorph. Although difference between velocities and dwell times between the two allomorphs seems minor, we clearly noticed differences in the binding stability of Cel7A to cellulose III. Previous AFM and super-resolution fluorescence based Cel7A single-molecule motility measurements on cellulose have been conducted at very high enzyme loadings where multiple Cel7A proteins often interact with each other to literally 'push' stuck enzymes out of their way in so-called Cel7A 'traffic jams' on the cellulose surface (22, 24). It is possible that similar protein-protein interactions play an important role in aiding cellulose deconstruction at higher cellulase loadings and could explain why higher

Cel7A loadings results in higher activity towards cellulose III. Furthermore, Igarashi and co-workers have speculated that since cellulose III has a modified crystal surface with a larger exposed protein binding surface than cellulose I, it is possible that multiple bound Cel7A enzymes can simultaneously deconstruct cellulose III to give improved hydrolysis yield but only at higher enzyme loadings (e.g., ~25-50 mg/g). However, in our tweezer-based Cel7A assays we only monitor one bound enzyme molecule at a time, without any interaction effects from other freely diffusing or surface-bound enzymes. Our motility results are therefore representative of single-enzyme behavior that would be expected as we drive down the total protein-to-cellulose loading to extreme enzyme-limiting conditions (e.g., under 0.5 mg/g). It would be interesting to study if addition of exogenous, freely diffusing exocellulases would impact the observed motility of the DNA-bead tethered single-Cel7A enzyme in our assay to unravel the impact of protein-protein interactions on improved catalytic activity on cellulose III. Future studies could also explore the role of possible allosteric effects on processive cellulase catalytic domain interactions with the cellulose allomorph surface to help explain the slightly increased stalling of the enzyme and increased back-stepping seen for Cel7A on cellulose III in particular. Regardless, here we hypothesized that the initial binding commitment of cellulases driven by CBMs is likely a key limiting step to kick-starting efficient cellulose saccharification.

Understanding CBM-polysaccharide binding interactions is critical to gaining mechanistic insights into biomass conversion (50–52) and developing more efficient industrial-grade enzymes (53, 54). Although molecular simulations have been employed to study specific steps of Cel7A cellulase processive cycle such as chain decrystallization (55), glycosylation (20), deglycosylation (21) and dissociation (56), the role of CBMs in initial motility commitment of catalytic domains has not yet been studied in detail (9, 12). From an evolutionary standpoint, Type-A CBMs and cellulase catalytic domains have naturally evolved to breakdown native cellulose I (57) but not cellulose III. Therefore, here we used classical CBM-cellulose pull-down binding assays, molecular dynamics simulations, and optical tweezer-based bond rupture assays to obtain a comprehensive understanding of the binding interactions of a model CBM1 (isolated from Cel7A) towards cellulose I and cellulose III. Classical pull-down bulk ensemble binding assays have been employed extensively to study protein binding to insoluble polysaccharides like cellulose (40). Like previous reports, various adsorption models such as Langmuir one-site/two-site and Langmuir-Freundlich models were applied to extract phenomenological model-based parameters for CBM1 binding towards both cellulose allomorphs. Regardless of the change in binding affinity, we mostly observed a drop in the total available binding sites available for CBM1, which suggests that the surface properties of cellulose allomorph have a significant impact on binding and recognition by CBMs/cellulases (15). A similar reduction in CBM binding was observed for another cellulose allomorph (i.e., cellulose-II) using QCM-D as well (58), suggesting most CBMs likely display subtle differences in binding interactions towards distinct cellulose allomorphs. We further extended our study to other model Type-A CBMs (e.g., CBM3a, CBM64) and confirmed reduced CBM binding partition coefficient observed towards cellulose III for all CBMs tested so far. Reduced mutant CBM3a and CBM64 binding towards distinct cellulose allomorphs further highlights the complex nature of CBM-cellulose binding interactions and its relationship to appended CD activity as shown in another recent study (59). Moreover, Langmuir adsorption models are applicable under some key assumptions (e.g., complete reversibility of protein-ligand binding, absence of bound protein structural deformation or interactions with other bound proteins, absence of overlapping binding sites, and complete surface saturation achieved at the maximum protein loading tested) which can often lead to possibly spurious conclusions resulting from such analyses (60). We also studied a small-molecule CBM-surrogate such as calcofluor white to characterize its binding parameters towards distinct cellulose allomorphs to show that calcofluor also has lower affinity and binding sites available for adsorption to microcrystalline cellulose III versus native cellulose I. But even for a simple stilbene-based derivative like calcofluor, we observed a concave upward behavior in Scatchard plots which is

indicative of overlapping binding sites and/or multiple classes of non-equivalent binding sites (see **SI Appendix Fig. S15** and **SI Appendix** for supporting discussion). These results highlight the complexity of studying CBM-cellulose interactions using simple Langmuir-based adsorption models and the inherent heterogeneity of the substrate binding sites that makes it challenging to gain deeper mechanistic insights from classical protein-polysaccharide pull-down assays alone.

To corroborate our results from CBM1 pull-down binding assay analyses, we performed molecular dynamics (MD) simulations to characterize CBM1 binding. MD simulations have been employed extensively to study cellulolytic enzymes (20, 61–64), and offer detailed atomistic insights into the highly heterogeneous binding interactions of proteins with insoluble polysaccharides. MD simulations have revealed structural and dynamical features of cellulose III such as hydrogen bonding, solvent accessible surface area, and single-cellulose chains decrystallization free energy (15, 55). While few studies have been carried out to understand CBM binding to cellulose, most work has been restricted to native cellulose I (34, 35, 65). Our PMF calculations from molecular simulations revealed that the binding free energy for CBM1 towards highly crystalline cellulose III is marginally (~1.2-fold) lower than cellulose I, confirming our analysis of binding assays based on Langmuir one-site model. The decrease in PMF estimated binding free energy is over 3-fold for more disordered (i.e., less crystalline) forms of cellulose III that are expected to be produced under certain low-temperature ammonia pretreatment conditions (27), but was not the case in this study. Hence, these results further suggest the use of simpler adsorption models like Langmuir one-site model that yield a more representative average binding affinity and available binding sites, instead of using over-parameterized multi-site Langmuir type models which could result in estimation of spurious binding parameters. MD simulations also provided atomistic insight into the molecular origins of reduced CBM1 binding to cellulose III due to improper Y5/Y31/Y32 aromatic residue stacking interactions and steric clashes with the jagged cellulose surface chains. Since C6-hydroxymethyl groups on the surface-exposed cellulose chains in native cellulose I are often highly disordered to adopt additional *gauche-trans* and *gauche-gauche* type rotameric states, it's likely that improper stacking of aromatic residues of CBMs with distinct cellulose crystal faces will impact binding stability to even some forms of native cellulose allomorphs. Interestingly, CBM1 also displayed a preferred non-canonical orientation on the surface of cellulose III which could impair Cel7A catalytic domain motility. This could explain why intact Cel7A displayed impaired motility on cellulose III with longer dwell times than cellulose I, but more experimental work is needed with CDs alone to rule out any allosteric effects that could impact motility as well. Engineering CBMs to reduce steric clashes and enable preferred canonical orientation to aid in efficient cellulase processivity is an area where future advancements can be made using rational structure-guided enzyme engineering strategies.

Although classical pull-down binding assays and MD simulations explain how the impaired cellulase motility commitment on cellulose III could arise from CBM1, the CBM1-cellulose binding/unbinding forces relevant to the processive motility cycles of Cel7A was unclear. Hence, we developed and applied a single-molecule non-covalent bond rupture assay to characterize CBM-cellulose binding interactions under applied force. Single-molecule force spectroscopy has been employed previously to distinguish the nature of protein-ligand bonds (45) and infer multi-modality or conformational transitions involved in protein-ligand binding interactions (66). However, the application of AFM-based force spectroscopy to study CBM-cellulose binding has revealed challenges in distinguishing specific versus non-specific interactions (67). Here, we developed a novel single-molecule optical tweezer-based bond rupture assay with piconewton (pN) force resolution and millisecond (ms) time resolution (66), to understand the heterogeneity of CBM-cellulose unbinding behavior under the application of force. The ultimate goal of bond rupture assay was to understand the role of CBM1 binding in the anomalous processive motility of Cel7A on cellulose III. CBM1 showed multi-modal force-lifetime behavior towards both cellulose I and cellulose III with no statistically significant differences in mean bond lifetimes except under extreme force ranges where the differences were

slightly more pronounced. Interestingly, the rupture assay mean bond lifetime of CBM1 with filter paper derived cellulose I fibrils was significantly higher (by ~2-fold) than Cladophora cellulose I. Overall these results highlight how subtle differences in cellulose fibril ultrastructure can play an important role on impacting CBM binding dynamics at the single-molecule level. Rupture assay bond lifetimes estimated from dynamic force spectroscopy assays can be used predict protein-ligand unbinding off-rates that relate directly to the classical binding affinity constant (68). Considering the mean bond lifetime for CBM1 was only marginally higher for cellulose I versus cellulose III, these results further suggest that a simple one-site Langmuir adsorption model used to fit the pull-down binding assay data would be more appropriate than other multi-site models that predict much larger differences in binding affinity. Our single-molecule CBM-cellulose bond rupture assay suggests that the binding behavior cannot be explained by presence of just one or even two classes of unique and independent binding sites. However, fitting high-quality binding assay dataset to a simple Langmuir one-site model can still yield a global average affinity constant that arises from a combination of binding sites or modes, rather than data overfitting via a two-site or more complex binding models. Our analysis also suggests the use of Langmuir one-site models to obtain binding parameters when studying protein-polysaccharide binding interactions, while also using complementary approaches to cross-validate the molecular-level origins of relative differences in binding behaviors observed for distinct ligands and/or protein mutants. Recent reports on even simpler protein-ligand systems like streptavidin-biotin suggests that ligand unbinding undergoes transition across multiple intermediate states as a function of the loading rate (i.e., applied force), unlike the classical two-state models, to explain the long lifetime of the complexes (69). Therefore, further studies are necessary for the CBM-cellulose system at multiple loading rates. We speculate that the non-productive binding of CBMs with high bond lifetimes could increase CD dwell time and mutant CBMs/cellulases should be analyzed to test this hypothesis further.

Lastly, we were interested to see if it would be theoretically possible to explain the multimodality observed for CBM-cellulose force-lifetime distributions using a simple geometrical probability-based model whereby the CBM is hypothesized to survey multiple binding orientations on the hydrophobic face of cellulose, assuming that different orientations would give a distinct bond lifetime at a given applied force. We were inspired by the classical Buffon needle problem and therefore developed a simple model based on this original problem to predict the probabilistic distribution of the orientation of CBM proteins on the surface of cellulose (70). Here the size of our needle is interpreted as the physical length of the planar binding motif surface (e.g., Y5-Y31-Y32) known to participate in cellulose binding, while the distance between the adjacent cellulose chains on the hydrophobic binding surface are equivalent to the distance between the parallel lines between which the needle can fall on along the line axis or across the axis between crossing multiple lines based on the original Buffon problem. Our Buffon needle model for the wild-type CBM1 predicted that the distribution of CBM1 binding states should mostly align along the cellulose chain axis versus across the chain axis under the assumption that these states are energetically equivalent, as discussed in SI Appendix Results and Discussion section (see **SI Appendix Fig. S16**). Alignment of the CBM needle along the cellulose chain axis is also supported by previous MD simulations (35), lending some credence to this overly simplistic geometrical interpretation of the CBM-cellulose binding problem. Interestingly, 'shortening' of the effective CBM needle length (i.e., by mutation of Y31A for CBM1) increased the likelihood of along the cellulose chain/axis binding events as predicted by the Buffon model. It was interesting to note that our single molecule tweezer based CBM-cellulose rupture assay also indicated a 2.6-fold significantly higher rupture bond lifetimes (in 10-15 pN rupture force range) for the Y31A mutant compared to the wild-type CBM1 on cellulose I, suggesting the intriguing possibility that a subset of the force-lifetime data observed could be representative of specific CBM1 orientations on the cellulose surface. A similar flanking aromatic residue mutation on other Type-A CBMs planar binding sites was recently shown to also enhance engineered endocellulase catalytic towards native cellulose, possibly due to reduced non-productive

mutant enzyme binding driven by particular binding orientations (59). Future work combining site-directed mutagenesis of CBMs, force spectroscopy rupture assays, and MD simulations is necessary to test the impact of specific CBM binding motif mutations on altering certain binding modalities, as analogously illustrated by Jobst *et al.* for the cohesin-dockerin binding system (71).

Binding modules like CBM1 play an oft-neglected synergistic role in the association of Cel7A catalytic domain to cellulose that likely fine-tunes the subtle balance between productive versus non-productive binding (72). Our motility assays have, for the first time in reported literature, captured the early steps of full-length cellulase complexation (also called Cel7A processive cycle motility commitment) to a cellulose reducing end before the catalytic processive cycle begins. Future work will address the role of CBMs in both the association and dissociation processes of full-length cellulases, to obtain a better understanding of the relationship between binding affinity and overall catalytic efficiency for processive cellulases (19). Our work has also shown that while the exact stalling force for halting processive cellulases like Cel7A likely exceeds 30 pN to prevent cellulase motility entirely (26), it is possible that particular CBM binding orientations on the cellulose surface could hinder cellulase motility or processive activity. However, the connection between data collected from single-enzyme motility/rupture assays, enzyme binding/activity, and enzyme-substrate structure dynamics still needs to be more clearly established. In addition, future work should address the interplay of CBM-driven binding affinity and hydrolytic activity of multi-modular cellulases, using biochemical assays similar to those reported in a recent study that applied the Sabatier principle to characterize interfacial cellulose hydrolysis by bound cellulases (19). It is likely that the lower binding and improved activity of endocellulases and exocellulases towards cellulose III at certain enzyme loadings is in accordance with the Sabatier principle.

Experimental Procedures

See SI appendix (Supplementary Text) for all experimental and computational methods used here.

Data availability: All raw data contained within the article is available upon request from the corresponding author (Dr. Shishir Chundawat, Rutgers University, shishir.chundawat@rutgers.edu).

Acknowledgments

The authors acknowledge support from the NSF CBET awards (1604421 and 1846797), ORAU Ralph E. Powe Award, NSF MCB award (1330792), NIH Grant (R01GM101001), Rutgers Global Grant, Rutgers Division of Continuing Studies, Rutgers School of Engineering, DOE Bioimaging Award (Office of Science DE-SC0019313) and the Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494). CL and SG acknowledge support from LANL LDRD ER (XWX2) and LANL institutional program for computational resources. JMY acknowledges support from the Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE) under agreement no. 28598. SPSC would like to particularly thank Professor Brian Fox (UW Madison) and Professor Bruce Dale (MSU) for kindly providing access to their lab's resources at the onset of this project for generation of relevant plasmid DNA and cellulose substrates, respectively. Special thanks to Izak Smith for help with collection of Cladophora algae from the Yahara River-Lake Watershed, Sungsoo & Amy Lim for help with generation of GFP-CBM constructs, Leonardo Sousa for conducting ammonia pretreatment, Shashwat Gupta for XRD data analysis, Ki-Bum Lee & Hyeon-Yeol Cho for access to the AFM, and Umesh Agarwal for access to the FT-Raman spectrometer. BN and JMY would like to thank Dr. Ashutosh Mittal (NREL) for his help with XRD measurements of Avicel cellulose derived nanocrystals. BN would also like to thank Dr. Jeff Linger at NREL for providing access to lab resources to perform QCM-D experiments. Figure 2A was created with BioRender.com.

Author Contributions: SPSC, MJL, and SG designed research; SPSC, BN, MH, SKB, MAH, MMJ, SC, HH, SHL, JMY, and CAL conducted research; SPSC, BN, MH, SKB, MAH, MMJ, SHL, and CAL analyzed the data; and SPSC, BN, MAH, MMJ, and MJL wrote the manuscript.

Competing interests
SPSC declares a competing financial interest(s) having filed two patent applications on pretreatment processes to produce cellulose-III enriched cellulosic biomass for biofuels production (US20130244293A1and WO2011133571A2). All other authors declare that they have no competing interests.

References:

1. Santoro, N., Cantu, S., Tornqvist, C.-E., Falbel, T., Bolivar, J., Patterson, S., Pauly, M., and Walton, J. (2010) A high-throughput platform for screening milligram quantities of plant biomass for lignocellulose digestibility. *BioEnergy Res.* **3**, 93–102

2. Dale, B. E., Bals, B. D., Kim, S., and Eranki, P. (2010) Biofuels done right: Land efficient animal feeds enable large environmental and energy benefits. *Environ. Sci. Technol.* **44**, 8385–8389

3. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Foust, T. D. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science (80-.).* **315**, 804–807

4. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. *Chem. Rev.* **115**, 1308–1448

5. Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002) Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**, 506–577

6. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, 490–495

7. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12229–12234

8. Mudinoor, A. R., Goodwin, P. M., Rao, R. U., Karuna, N., Hitomi, A., Nill, J., and Jeoh, T. (2020) Interfacial molecular interactions of cellobiohydrolase Cel7A and its variants on cellulose. *Biotechnol. Biofuels*. **13**, 1–16

9. Chundawat, S. P. S., Beckham, G. T., Himmel, M., and Dale, B. E. (2011) Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* **2**, 121–145

10. Gao, D., Chundawat, S. P. S., Uppugundla, N., Balan, V., and Dale, B. E. (2011) Binding characteristics of *Trichoderma reesei* cellulases on untreated, ammonia fiber expansion and dilute-acid pretreated lignocellulosic biomass. *Biotech Bioeng.* **108**, 1788–1800

11. Novy, V., Aïssa, K., Nielsen, F., Straus, S. K., Ciesielski, P., Hunt, C. G., and Saddler, J. (2019) Quantifying cellulose accessibility during enzyme-mediated deconstruction using 2 fluorescence-tagged carbohydrate-binding modules. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22545–22551

12. da Costa Sousa, L., Jin, M., Chundawat, S. P. S., Bokade, V., Tang, X., Azarpira, A., Lu, F., Avci, U., Humpula, J., Uppugundla, N., Gunawan, C., Pattathil, S., Cheh, A. M., Kothari, N., Kumar, R.,

Ralph, J., Hahn, M. G., Wyman, C. E., Singh, S., Simmons, B. A., Dale, B. E., and Balan, V. (2016) Next-generation ammonia pretreatment enhances cellulosic biofuel production. *Energy Environ. Sci.* **9**, 1215–1223

13. Gao, D., Chundawat, S. P. S., Sethi, A., Balan, V., Gnanakaran, S., and Dale, B. E. (2013) Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proc. Natl. Acad. Sci.* **110**, 10922–10927

14. Liu, Y., Nemmaru, B., and Chundawat, S. P. S. (2020) *Thermobifida fusca* cellulases exhibit Increased endo–exo synergistic activity, but lower exocellulase activity, on cellulose-III. *ACS Sustain. Chem. Eng.* **8**, 5028–5039

15. Chundawat, S. P. S., Bellesia, G., Uppugundla, N., Sousa, L., Gao, D., Cheh, A., Agarwal, U., Bianchetti, C., Phillips, G., Langan, P., Balan, V., Gnanakaran, S., and Dale, B. E. (2011) Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *J. Am. Chem. Soc.* **133**, 11163–11174

16. Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K., and Westh, P. (2013) Transient kinetics and rate limiting steps for the processive cellobiohydrolase Cel7A: Effects of substrate structure and carbohydrate binding domain. *Biochemistry*. **52**, 8938–8948

17. Kurašin, M., and Väljamäe, P. (2011) Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* **286**, 169–177

18. Fox, J. M., Levine, S. E., Clark, D. S., and Blanch, H. W. (2012) Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry*. **51**, 442–452

19. Kari, J., Olsen, J. P., Jensen, K., Badino, S. F., Krogh, K. B. R. M., Borch, K., and Westh, P. (2018) Sabatier principle for interfacial (heterogeneous) enzyme catalysis. *ACS Catal.* **8**, 11966–11972

20. Knott, B. C., Crowley, M. F., Himmel, M. E., Ståhlberg, J., and Beckham, G. T. (2014) Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity. *J. Am. Chem. Soc.* **136**, 8810–8819

21. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J., and Beckham, G. T. (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. *J. Am. Chem. Soc.* **136**, 321–329

22. Shibafuji, Y., Nakamura, A., Uchihashi, T., Sugimoto, N., Fukuda, S., Watanabe, H., Samejima, M., Ando, T., Noji, H., Koivula, A., Igarashi, K., and Iino, R. (2014) Single-molecule imaging analysis of elementary reaction steps of *Trichoderma reesei* cellobiohydrolase I (Cel7A) hydrolyzing crystalline cellulose Iα and IIII. *J. Biol. Chem.* **289**, 14056–14065

23. Nakamura, A., Tasaki, T., Ishiwata, D., Yamamoto, M., Okuni, Y., Visootsat, A., Maximilien, M., Noji, H., Uchiyama, T., Samejima, M., Igarashi, K., and Iino, R. (2016) 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

24. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttila, M., Ando,

T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science (80-.).* **333**, 1279–1282

25. Gilbert, H. J., Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Penttilä, M., Ando, T., and Samejima, M. (2012) Visualization of cellobiohydrolase I from *Trichoderma reesei* moving on crystalline cellulose using high-speed atomic force microscopy. in *Methods in Enzymology*, pp. 169–182, **510**, 169–182

26. Brady, S. K., Sreelatha, S., Feng, Y., Chundawat, S. P. S., and Lang, M. J. (2015) Cellobiohydrolase 1 from *Trichoderma reesei* degrades cellulose in single cellobiose steps. *Nat. Commun.* **6**, 10149

27. Sousa, L. da C., Humpula, J., Balan, V., Dale, B. E., and Chundawat, S. P. S. (2019) Impact of ammonia pretreatment conditions on the cellulose III allomorph ultrastructure and its enzymatic digestibility. *ACS Sustain. Chem. Eng.* **7**, 14411–14424

28. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal Structure and Hydrogen Bonding System in Cellulose I_α from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Amer. Chem. Soc.* **125**, 14300–14306

29. Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron x-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* **124**, 9074–9082

30. Garvey, C. J., Parker, I. H., and Simon, G. P. (2005) On the interpretation of x-ray diffraction powder patterns in terms of the nanostructure of cellulose I fibres. *Macromol. Chem. Phys.* **206**, 1568–1575

31. Wada, M., Chanzy, H., Nishiyama, Y., and Langan, P. (2004) Cellulose III crystal structure and hydrogen bonding by synchrotron x-ray and neutron fiber diffraction. *Macromolecules*. **37**, 8548–8555

32. Atalla, R. J., and Vanderhart, D. L. (1987) Studies on the structure of cellulose using raman spectroscopy and solid state 13C-NMR. in *Institute for Paper Chemistry (IPC Technical Paper Series 217)*

33. Wiley, J. H., and Atalla, R. H. (1987) Band assignments in the raman spectra of celluloses. *Carbohydr. Res.* **160**, 113–129

34. Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E., and Crowley, M. F. (2012) Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J. Biol. Chem.* **287**, 20603–20612

35. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. *J. Phys. Chem. B*. **114**, 1447–1453

36. Kraulis, P. J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J., and Gronenborn, A. M. (1989) 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

37. Reinikainen, T., Ruohonen, L., Nevanen, T., Laaksonen, L., Kraulis, P., Jones, T. A., Knowles, J.

K. C., and Teeri, T. T. (1992) Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *Proteins-Struct. Func. Genet.* **14**, 475–482

38. Linder, M., and Teeri, T. T. (1996) The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. *Proc. Natl. Acad. Sci.* **93**, 12251–12255

39. Pettersson, Gör., Linder, M., Reinikainen, T., Drakenberg, T., Mattinen, M.-L., Annila, A., Kontteli, M., Lindeberg, G., and Ståhlberg, J. (1995) Identification of functionally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Sci.* **4**, 1056–1064

40. Abbott, D. W., and Boraston, A. B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol.* **510**, 211–31

41. Lim, S., Chundawat, S. P. S., and Fox, B. G. (2014) Expression, purification and characterization of a functional carbohydrate-binding module from *Streptomyces sp. SirexAA-E*. *Protein Expr. Purif.* **98**, 1–9

42. Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., and Steitz, T. A. (1996) Crystal structure of a bacterial family-III cellulose-binding domain: A general mechanism for attachment to cellulose. *EMBO J.* **15**, 5739–5751

43. Walker, J. A., Takasuka, T. E., Deng, K., Bianchetti, C. M., Udell, H. S., Prom, B. M., Kim, H., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules. *Biotechnol. Biofuels*. **8**, 220

44. Moran-Mirabal, J. M., Bolewski, J. C., and Walker, L. P. (2011) Reversibility and binding kinetics of *Thermobifida fusca* cellulases studied through fluorescence recovery after photobleaching microscopy. *Biophys. Chem.* **155**, 20–8

45. Rakshit, S., Zhang, Y., Manibog, K., Shafraz, O., and Sivasankar, S. (2012) Ideal, catch, and slip bonds in cadherin adhesion. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 18815–20

46. Bell, G. I. (1978) Models for the specific adhesion of cells to cells. *Science (80-.).* **200**, 618–627

47. Fox, J. M., Jess, P., Jambusaria, R. B., Moo, G. M., Liphardt, J., Clark, D. S., and Blanch, H. W. (2013) A single-molecule analysis reveals morphological targets for cellulase synergy. *Nat Chem Biol*. **9**, 356–361

48. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 484–489

49. Hendriks, A. T. W. M., and Zeeman, G. (2009) Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresour. Technol.* **100**, 10–18

50. Herve, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc. Natl. Acad. Sci.* **107**, 15293–15298

51. Cuskin, F., Flint, J. E., Gloster, T. M., Morland, C., Basle, A., Henrissat, B., Coutinho, P. M., Strazzulli, A., Solovyova, A. S., Davies, G. J., and Gilbert, H. J. (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc.*

Natl. Acad. Sci. **109**, 20889–20894

52. Gilbert, H. J., Knox, J. P., and Boraston, A. B. (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr. Opin. Struct. Biol.* **23**, 669–677

53. Brunecky, R., Subramanian, V., Yarbrough, J. M., Donohoe, B. S., Vinzant, T. B., Vanderwall, T. A., Knott, B. C., Chaudhari, Y. B., Bomble, Y. J., Himmel, M. E., and Decker, S. R. (2020) Synthetic fungal multifunctional cellulases for enhanced biomass conversion. *Green Chem.* **22**, 478–489

54. Várnai, A., Mäkelä, M. R., Djajadi, D. T., Rahikainen, J., Hatakka, A., and Viikari, L. (2014) Carbohydrate-binding modules of fungal cellulases. Occurrence in nature, function, and relevance in industrial biomass conversion. in *Advances in Applied Microbiology*, pp. 103–165, **88**, 103–165

55. Beckham, G. T., Matthews, J. F., Peters, B., Bomble, Y. J., Himmel, M. E., and Crowley, M. F. (2011) Molecular-level origins of biomass recalcitrance: Decrystallization free energies for four common cellulose polymorphs. *J. Phys. Chem. B*. **115**, 4118–4127

56. Vermaas, J. V, Kont, R., Beckham, G. T., Crowley, M. F., Gudmundsson, M., Sandgren, M., Ståhlberg, J., Väljamäe, P., and Knott, B. C. (2019) The dissociation mechanism of processive cellulases. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23061–23067

57. Blake, A. W., McCartney, L., Flint, J. E., Bolam, D. N., Boraston, A. B., Gilbert, H. J., and Knox, J. P. (2006) Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J. Biol. Chem.* **281**, 29321–29329

58. Hu, F., Zhang, Y., Wang, P., Wu, S., Jin, Y., and Song, J. (2018) Comparison of the interactions between fungal cellulases from different origins and cellulose nanocrystal substrates with different polymorphs. *Cellulose*. **25**, 1185–1195

59. Nemmaru, B., Ramirez, N., Farino, C. J., Yarbrough, J. M., Kravchenko, N., and Chundawat, S. P. S. (2020) Reduced type-A carbohydrate-binding module interactions to cellulose I leads to improved endocellulase activity. *Biotechnol. Bioeng.* 10.1101/2020.07.02.183293

60. Latour, R. A. (2015) The langmuir isotherm: A commonly applied but misleading approach for the analysis of protein adsorption behavior. *J. Biomed. Mater. Res. Part A*. **103**, 949–958

61. Payne, C. M., Jiang, W., Shirts, M. R., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2013) Glycoside hydrolase processivity is directly related to oligosaccharide binding free energy. *J. Am. Chem. Soc.* **135**, 18831–18839

62. Payne, C. M., Bomble, Y. J., Taylor, C. B., McCabe, C., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Multiple functions of aromatic-carbohydrate interactions in a processive cellulase examined with molecular simulation. *J. Biol. Chem.* **286**, 41028–41035

63. Wu, M., Bu, L., Vuong, T. V., Wilson, D. B., Crowley, M. F., Sandgren, M., Ståhlberg, J., Beckham, G. T., and Hansson, H. (2013) Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. *J. Biol. Chem.* **288**, 33107–33117

64. Jana, S., Hamre, A. G., Wildberger, P., Holen, M. M., Eijsink, V. G. H., Beckham, G. T., Sørlie, M., and Payne, C. M. (2016) Aromatic-mediated carbohydrate recognition in processive *Serratia*

marcescens chitinases. *J. Phys. Chem. B*. **120**, 1236–1249

65. Alekozai, E. M., GhattyVenkataKrishna, P. K., Uberbacher, E. C., Crowley, M. F., Smith, J. C., and Cheng, X. (2014) Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose I_β. *Cellulose*. **21**, 951–971

66. Das, D. K., Mallis, R. J., Duke-Cohan, J. S., Hussey, R. E., Tetteh, P. W., Hilton, M., Wagner, G., Lang, M. J., and Reinherz, E. L. (2016) Pre-T cell receptors (Pre-TCRs) leverage Vβ complementarity determining regions (CDRs) and hydrophobic patch in mechanosensing thymic self-ligands. *J. Biol. Chem.* **291**, 25292–25305

67. King, J. R., Bowers, C. M., and Toone, E. J. (2015) Specific binding at the cellulose binding module-cellulose interface observed by force spectroscopy. *Langmuir*. **31**, 3431–40

68. Dudko, O. K., Hummer, G., and Szabo, A. (2008) Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15755–60

69. Rico, F., Russek, A., González, L., Grubmüller, H., and Scheuring, S. (2019) Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 6594–6601

70. Buffon, G. (1777) Essai d'arithmétique morale. in *Histoire naturelle, générale er particulière, Supplément 4*, pp. 46–123

71. Jobst, M. A., Milles, L. F., Schoeler, C., Ott, W., Fried, D. B., Bayer, E. A., Gaub, H. E., and Nash, M. A. (2015) Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *Elife*. **4**, e10319

72. Kont, R., Kari, J., Borch, K., Westh, P., and Väljamäe, P. (2016) Inter-domain synergism Is required for efficient feeding of cellulose chain into active site of cellobiohydrolase Cel7A. *J. Biol. Chem.* **291**, 26013–26023

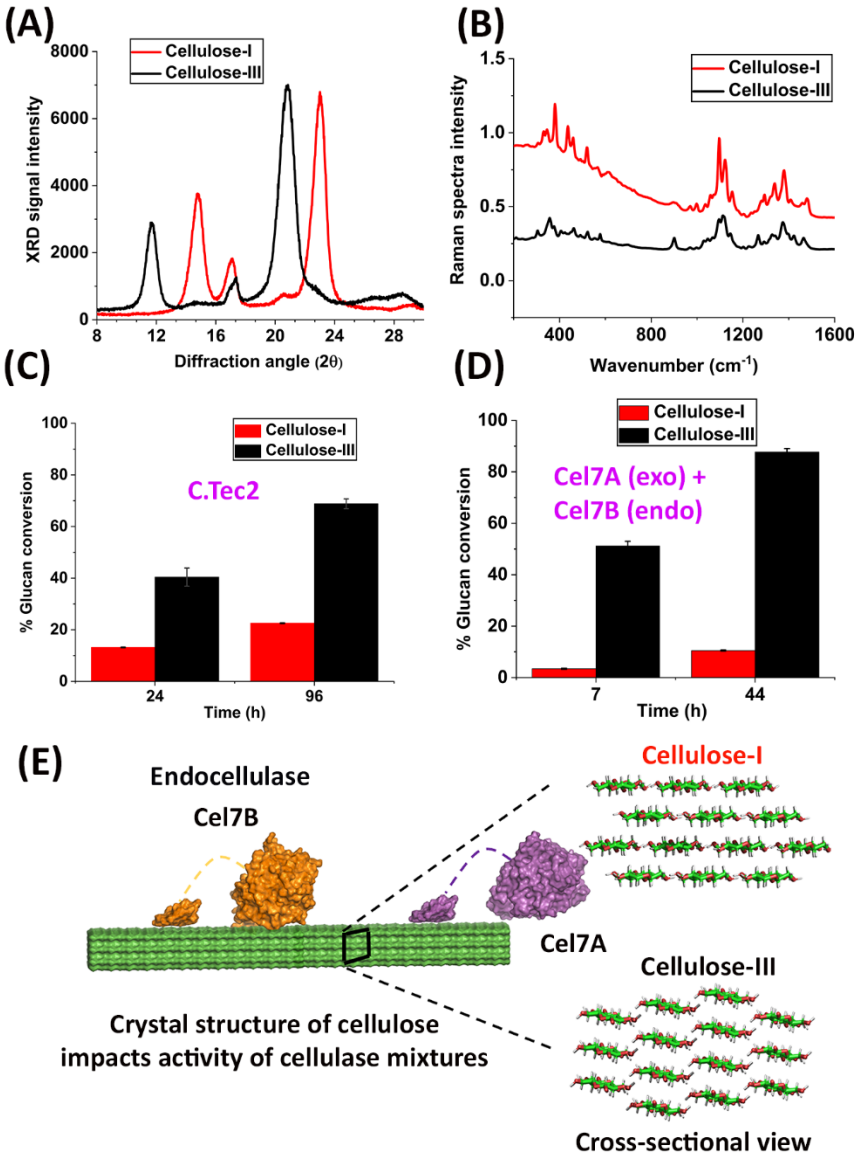


Figure 1. **Cladophora derived highly crystalline cellulose III allomorph is more readily hydrolyzed by synergistic combinations of cellulases.** (A) XRD and (B) FT-Raman spectra for cellulose I and cellulose III derived from Cladophora confirms respective allomorphic states. (C) Hydrolytic activity of Cellic C.Tec2 (Novozymes) cellulase cocktail towards cellulose I and cellulose III for varying hydrolysis times. (D) Hydrolytic activity of an equimolar mixture of *T. reesei* Cel7A exo- and Cel7B endo-cellulases, respectively, supplemented with 10% β-glucosidase, towards cellulose I and III. Specific activity for Cel7A alone can be found in **SI Appendix Table S1**. (E) Cross-sectional view of model cellulose I and III allomorphs depicting key morphological differences in fibril shape that impact endo-exo cellulase synergism toward cellulose III, as also reported previously (15). Published crystal structures of Cel7A (PDB code: 1CEL) and Cel7B (PDB code: 1EG1) were used to generate this figure. Here, hydrolytic activity is reported as mean value for replicate assays with error bars depicting one standard deviation.

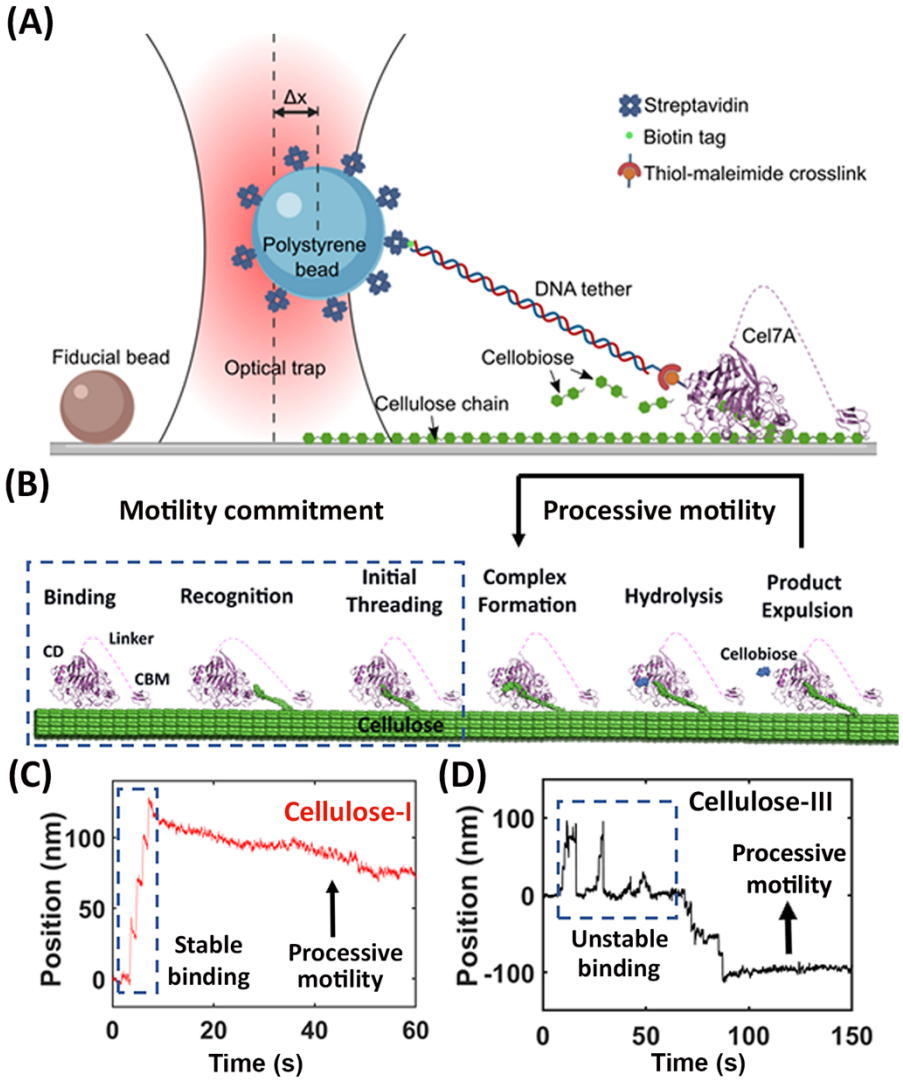


Figure 2. **Processive cellulase Cel7A shows unstable single-molecule binding on cellulose III.** (A) Schematic of cellulase motility assay setup (not to scale) is shown where a streptavidin coated bead is tethered to a single Cel7A molecule via a thiol-maleimide crosslink to a DNA linker containing a biotin tag on the opposite end and bound to cellulose to initiate Cel7A motility to produce cellobiose. Here, Δx represents the distance bead is displaced from trap center. (B) Processive cellulase Cel7A degrades cellulose via a multi-step mechanism involving (i) enzyme binding to cellulose, (ii) recognition of cellulose reducing end by catalytic domain (CD), (iii) threading of cellulose chain through active site, (iv) formation of a catalytically active complex by nucleophilic attack, (v) glycosidic bond hydrolysis, and (vi) cellobiose product expulsion from active site and forward stepping of the enzyme. Steps (iv), (v) and (vi) are repeated multiple times, leading to processive motion until the enzyme desorbs from the surface. Steps (i), (ii) and (iii) precede the processive motion of enzyme and hence determine enzyme commitment to motility (collectively called here as 'motility commitment'). (C) Single bead position trace representing initial stable binding to cellulose I followed by Cel7A motility. (D) Initial unstable binding to cellulose III followed by eventual Cel7A motility. The position of bead fluctuates significantly in the case of cellulose III to about 100 nm, indicating that the enzyme desorbs from cellulose surface multiple times before initiation of processive motion. Additional representative traces showcasing unstable protein binding prior to Cel7A motility initiation can be found in **SI Appendix Fig. S2**. Published crystal structure of Cel7A (PDB code: 1CEL) and Cel7B (PDB code: 1EG1) were used to generate this figure.

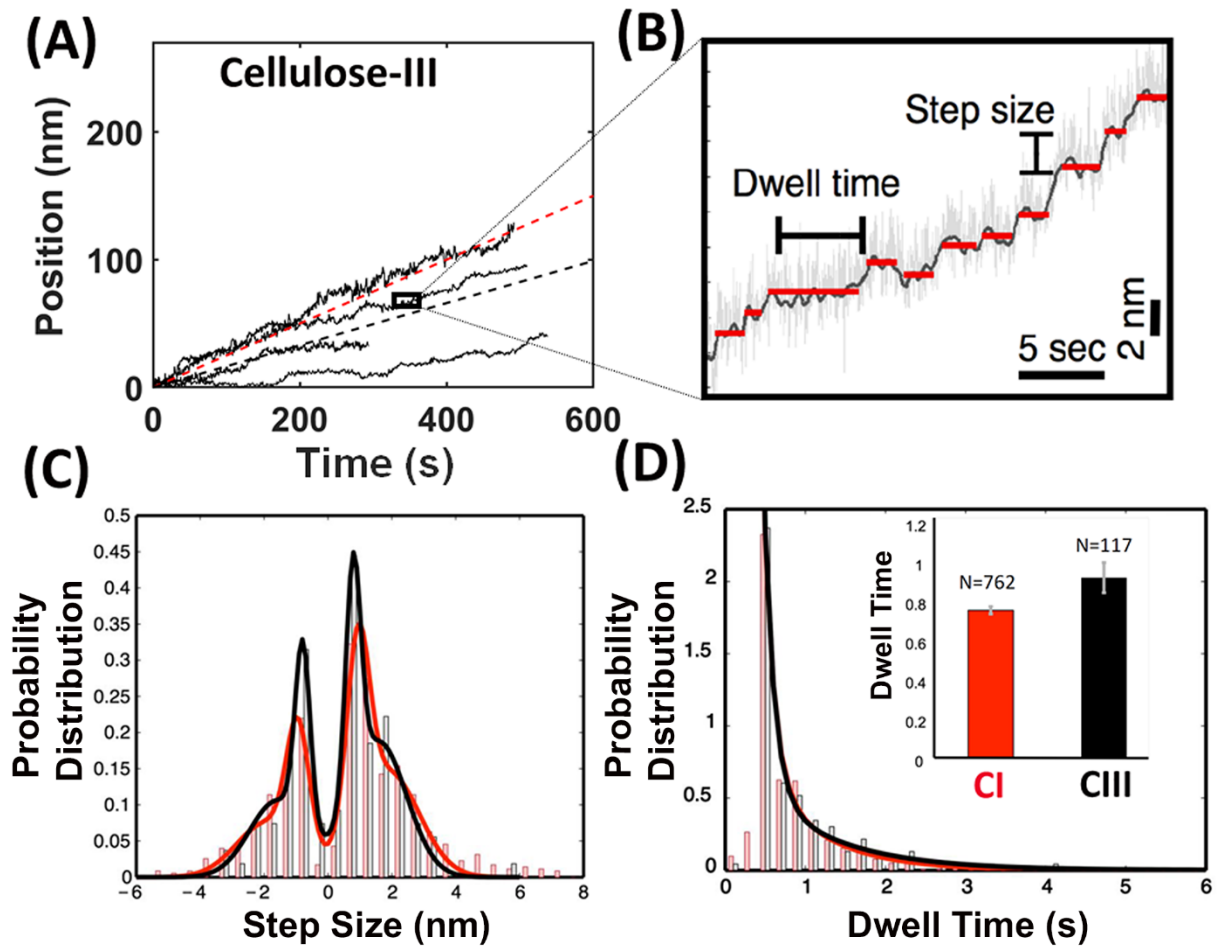


Figure 3. **Processive cellulase Cel7A shows reduced single molecule processive velocity on cellulose III.** (A) Representative traces of Cel7A enzyme motility on Cladophora derived cellulose III (in black) are shown here. Dashed lines in (A) indicate average velocities, 0.25 ± 0.35 nm s^{-1} (s.d.; cellulose I; N=68; in red) and 0.17 ± 0.14 nm s^{-1} (s.d.; cellulose III; N=30; in black). Representative traces for cellulose I can be found in **SI Appendix Fig. S3.** (B) Magnified view of individual motility cycle of enzyme that is made up of several dwells and steps. Dwell time and step size distributions are obtained as previously discussed by Brady et al. (2015). (C) All individual motility traces were analyzed to determine step-size distributions (as bars) fitted to Gaussian curves based on the fundamental (~1 nm) and 2x fundamental steps expected for Cel7A cellodextrin products (i.e., cellobiose) profile on cellulose I (in red) and III (in black). Slightly increased back-stepping of Cel7A on cellulose III (39% reverse steps) versus cellulose I (35% reverse steps) is seen here. (D) Dwell time distributions (as bars) were fitted to single-exponential decay curves to estimate characteristic dwell time constant (see inset) and were found to be higher for cellulose III (in black) versus cellulose I (in red). The average dwell times for cellulose I and cellulose III are 0.75 s and 0.92 s, respectively.

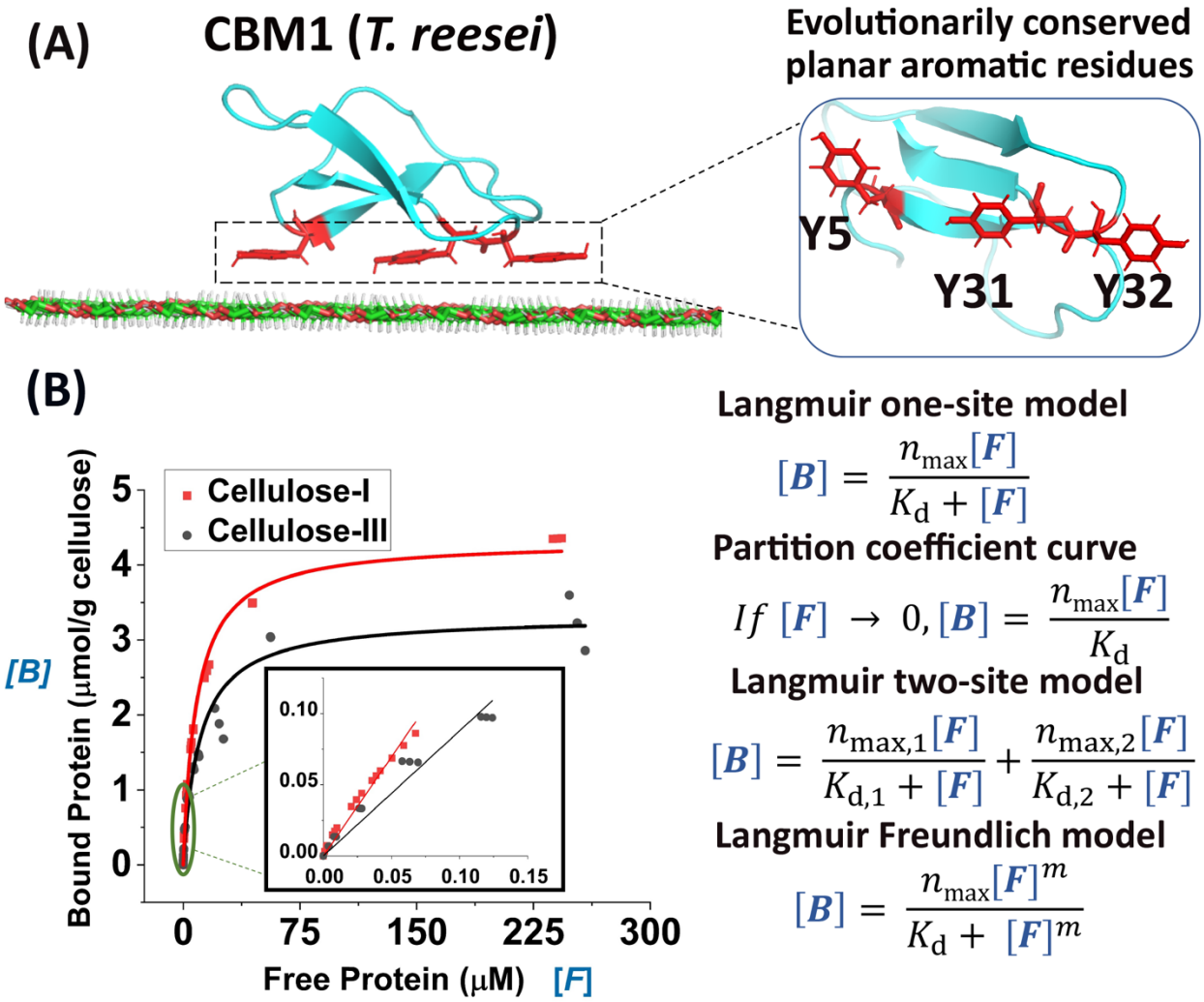


Figure 4**. CBM1-cellulose solid-depletion binding assay data and fitted Langmuir-type model equations.** (A) CBM1 (PDB code: 1CBH) from Cel7A docked on the hydrophobic face of crystalline cellulose I with axial hydrogens shown in white outlines (left). The planar binding motif comprised of aromatic residues highlighted in red (Y5, Y31 and Y32) is shown in inset (right). (B) GFP-CBM1 (*T. reesei*) equilibrium binding data for Cladophora cellulose I and cellulose III to estimate equilibrium adsorption constants are shown here. Non-linear relationship between bound and free GFP-CBM1 concentration for cellulose I (in red dots) and cellulose III (in black dots) is shown here for replicate assays. Fitted line depicts a Langmuir one-site model. Inset graph shows the linear region of this model to estimate partition coefficient. Relationship between bound and free protein for various adsorption models tested such as Langmuir one-site, two-site, and Freundlich models are shown here. Representative model fits for original CBM1-cellulose binding data are shown in **SI Appendix Fig. S4**, with results summarized in **Table 1**.

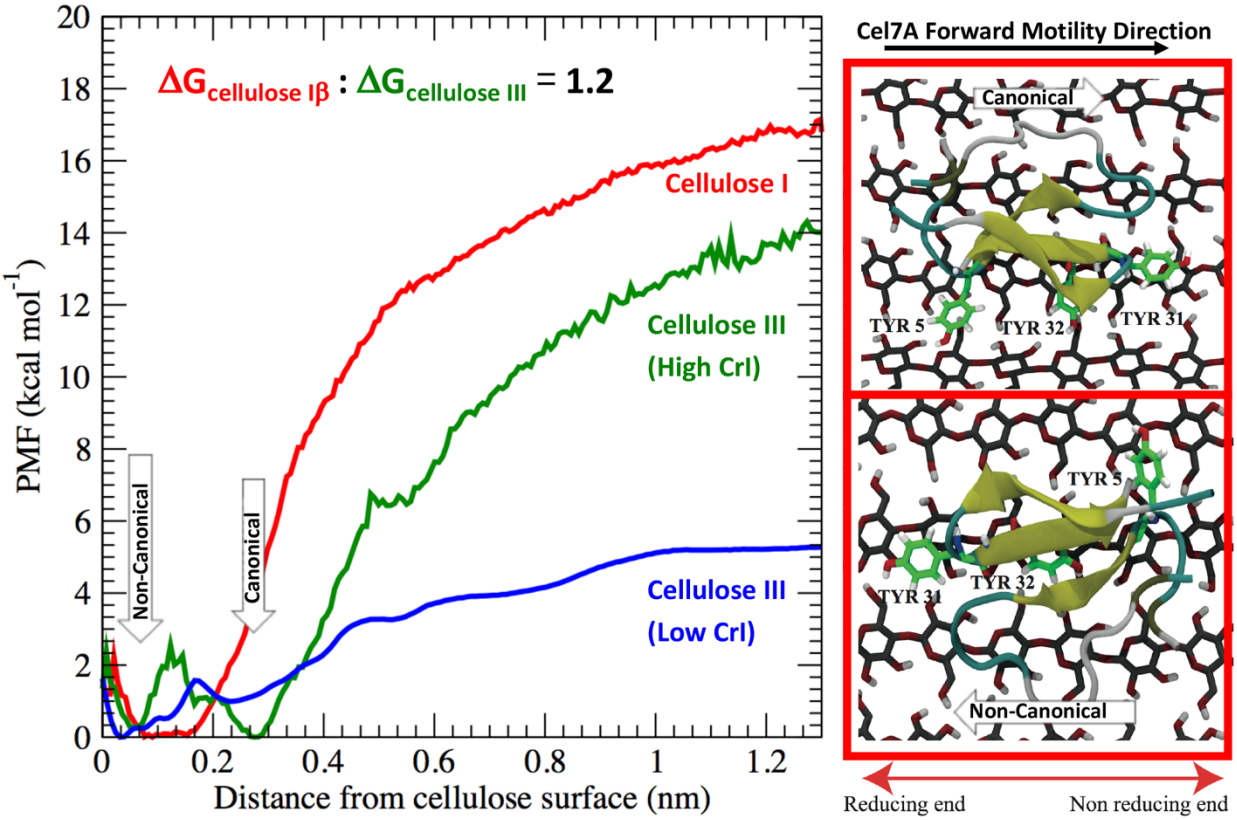


Figure 5. **Molecular dynamics (MD) simulations provide an atomistic basis for reduced affinity and distinct multimodal binding interactions of CBM1 to cellulose III allomorph surfaces.** Potential mean force (PMF) calculations were carried out to estimate the binding free energy of CBM1 with cellulose allomorphs to show that binding free energy is at least 1.2-fold higher for cellulose I (in red) versus high crystallinity cellulose III. High crystallinity index (CrI; in green) and low CrI (in blue) models of cellulose III were studied here for sake of comparison (see SI appendix for details). Note that the two energy wells for cellulose III correspond to the canonical and non-canonical orientations observed for bound CBM1. Canonical orientation refers to Y5 residue facing the reducing end, as it favors the processive motility of Cel7A from reducing to non-reducing end of cellulose chain. Figure inset here shows canonical (top) and non-canonical (bottom) orientations of CBM1 on high CrI cellulose III, along with the preferred direction of the processive Cel7A motility during cellulose saccharification. Additional details about the MD analysis and root mean square fluctuations of critical binding motif aromatic residues due to improper CBM1 stacking to cellulose III surface are highlighted in **SI Appendix Fig. S5**.

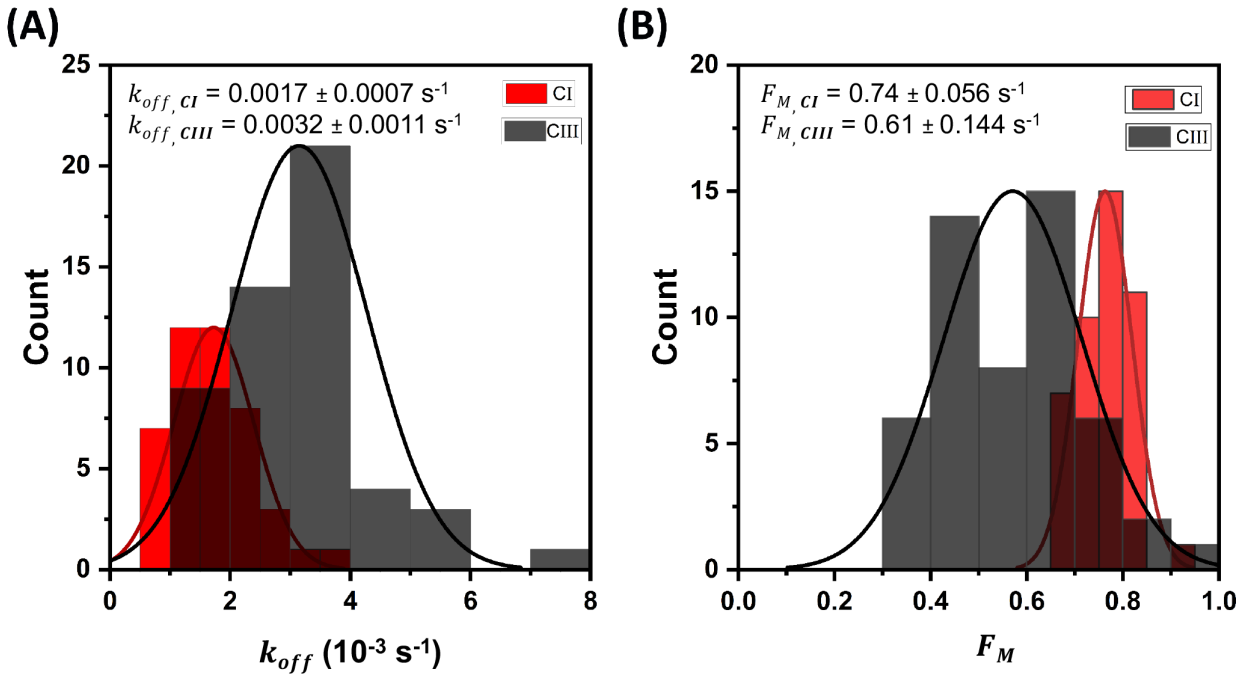


Figure 6. **Fluorescence Recovery After Photobleaching (FRAP) based analysis of GFP-CBM3a binding kinetics to cellulose allomorphs indicates an increase in the dissociation off-rate (k_{off}) for CBM3a for cellulose III versus cellulose I.** Panels A and B compare the dissociation off-rate and mobile fraction of reversibly bound proteins (or F_M) histograms for GFP-CBM3a on Cladophora cellulose I (CI) vs. cellulose III (CIII) with the Gaussian fit parameters (mean ± s.d.) as insets. Here, $\boldsymbol{F_M}$ represents the fraction of reversibly bound GFP-CBM3a and based on the model used to analyze the FRAP data is dependent on the pseudo-adsorption rate, desorption off-rate, concentration of protein in solution, and the slope of the calibration curve between fluorescence intensity and protein concentration. Details about the FRAP model parameters and data analysis approach is provided in the SI Appendix Methods. Representative FRAP recovery curves are shown in **SI Appendix Fig. S9**.

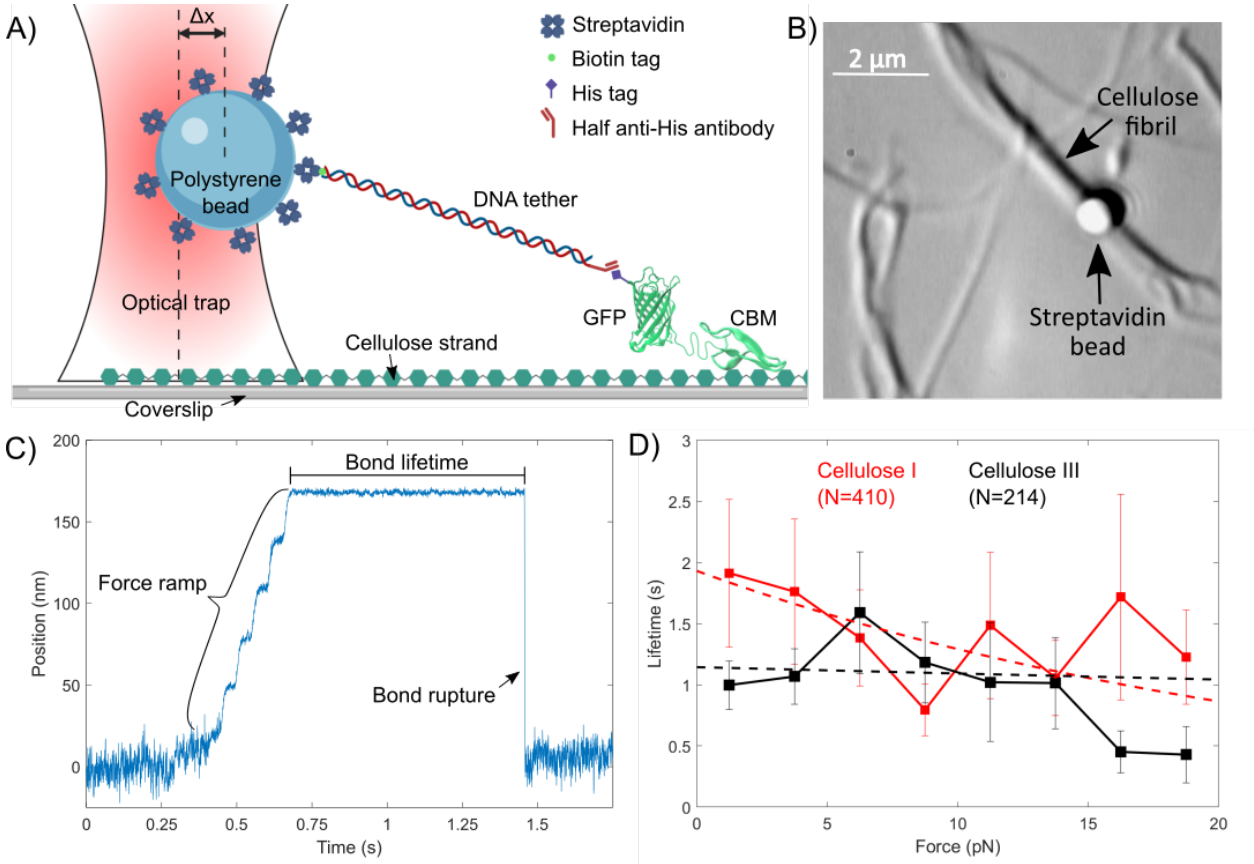


Figure 7. **Optical tweezers based single-molecule bond 'rupture' assay reveals multimodal nature of CBM1-cellulose binding interactions.** (A) Schematic of rupture assay setup (not to scale) is shown here where a streptavidin coated bead is tethered to a single His-GFP-labeled CBM1 via a DNA linker containing an anti-His antibody Fab and a biotin tag on opposite ends. The biotin end specifically binds to streptavidin whereas the ani-His antibody Fab specifically binds to the histidine tag of the GFP-labeled CBM1. Here, Δx represents the distance bead is displaced from the trap center. Figure was created with BioRender.com. Published structures of CBM1 (PDB code: 1CBH) and GFP (PDB code: 2B3P) were used in this rendering (B) Brightfield image of rupture assay showing Cladophora based cellulose microfibrils localized on the glass cover slip. CBM-cellulose binding is facilitated by moving the optically trapped bead close to the fiber. Bead position is tracked by a detection laser as force is loaded across the bond. (C) Representative position trace for a single CBM-cellulose rupture event showing bond lifetime and a single rupture is shown here. (D) Force vs Lifetime relationship for the CBM1-cellulose interaction on Cladophora cellulose I (black) and cellulose III (red) is shown. Lifetimes were binned at 2.5 pN intervals. Weighted single exponential fits are shown as dashed lines. Error bars depict standard error from the reported mean for each bin. N represents the total number of CBM-cellulose bond rupture events measured for each substrate. Additional supporting raw data scatterplots can be found in **SI Appendix Fig. S11-S14**.

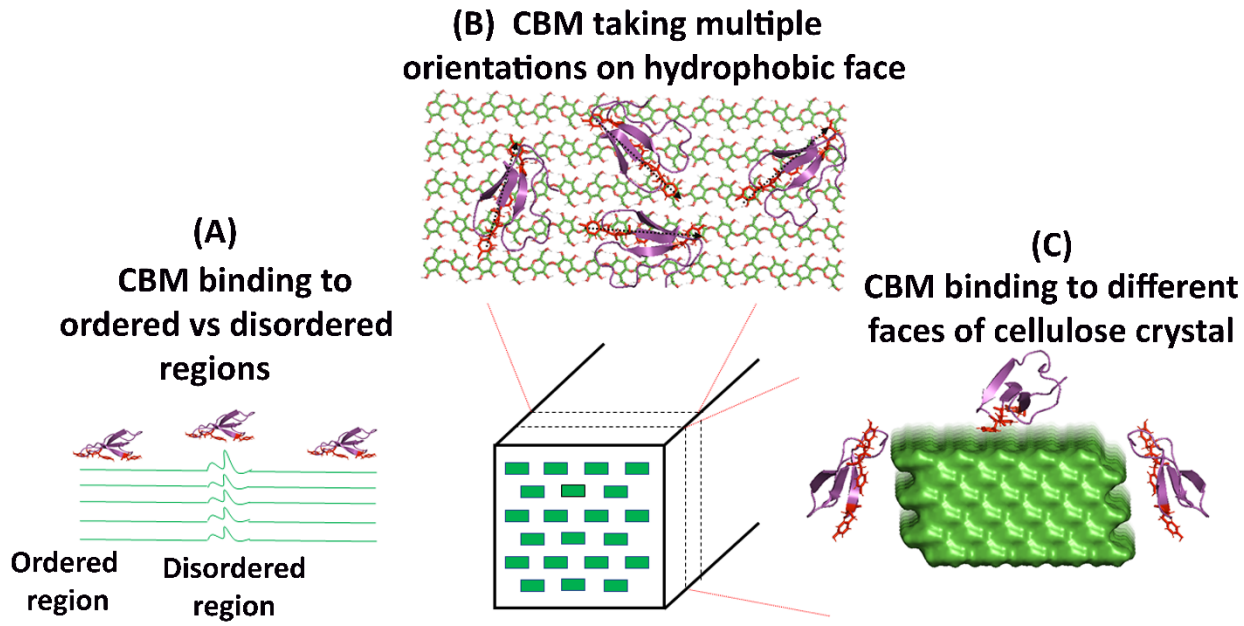


Figure 8. **Schematic outlining the three possible classes of binding sites theoretically accessible by CBM1 on native cellulose I fibers or crystal surface**. (A) Cellulose chains are considered to possess local regions of disorder (also termed amorphous regions) and it is likely that the binding free energy to more ordered (or crystalline) regions is slightly different from that of highly disordered regions and hence be regarded as different classes of binding sites. (B) Previous molecular simulation studies show that the hydrophobic face of cellulose crystal is the preferred binding face for type-A CBMs such as CBM1 (34). However, it is likely that the CBM possesses multiple binding orientations with respect to a cellulose chain due to non-specific hydrophobic interactions which drive CBM-cellulose binding. (C) In addition, although molecular simulations predict that hydrophobic face is the 'preferred' crystal face for CBM binding on microsecond time scales, transmission electron microscopy (TEM) studies have shown the possibility of CBM binding to various other faces of the cellulose crystal (48). Overall, it is likely that the combination of all these potential binding sites, depending on cellulose source and overall ultrastructure, leads to the heterogeneity observed in binding of CBMs to distinct cellulose allomorphs. Here, CBM1 (PDB code: 1CBH) from Cel7A was used to generate the figure.

Langmuir One-Site Binding Model		
	Cellulose-I	**Cellulose-III**
$n_{\max}$	4.34 ± 0.05	3.32 ± 0.09
K_{d}	8.69 ± 0.31	10.55 ± 0.91
RMSE	0.17	0.17
Langmuir Two-Site Binding Model		
$n_{\max,1}$	4.14 ± 0.00	2.81 ± 0.03
$K_{\mathrm{d},1}$	13.68 ± 0.13	25.06 ± 0.95
$n_{\max,1}$	0.42 ± 0.01	0.75 ± 0.03
$K_{\mathrm{d},2}$	0.13 ± 0.00	0.92 ± 0.07
RMSE	0.11	0.12
Langmuir Freundlich Binding Model		
$n_{\max}$	4.80 ± 0.02	3.82 ± 0.02
K_{d}	6.90 ± 0.04	7.77 ± 0.07
m	0.77 ± 0.00	0.73 ± 0.00
RMSE	0.14	0.14

Table 1. **Langmuir-based binding model parameters for GFP-CBM1 adsorption to Cladophora-based cellulose I and III**. Here, binding dissociation constant (K_{d}; μM), maximum available binding sites ($n_{\max}$; μmol/g cellulose), and Freundlich power constant (m) fitted parameters are shown. Model fitting details for all Langmuir-based adsorption models are provided in the SI appendix. The errors reported were standard errors to parameter fits obtained. Representative model fits for original CBM1-cellulose binding data are shown in **SI Appendix Fig. S4**.

	Cellulose-I	**Cellulose-III**
	Original data set (without truncation)	
n_{max}	4.34 ± 0.05	3.32 ± 0.09
K_d	8.69 ± 0.31	10.55 ± 0.91
RMSE	0.17	0.17
	Dataset truncated to 50 μM	
n_{max}	3.95 ± 0.05	3.23 ± 0.14
K_d	7.05 ± 0.24	9.88 ± 1.07
RMSE	0.008	0.16
	Dataset truncated to 15 μM	
n_{max}	3.18 ± 0.07	1.82 ± 0.05
K_d	4.68 ± 0.21	2.91 ± 0.21
RMSE	0.008	0.05

Table 2. **Langmuir one-site binding model fitting analysis to truncated CBM1-Cellulose I pull-down binding assay data.** Data for CBM1 binding to Cladophora based cellulose-I was truncated to maximum 50 μM and 15 μM free protein concentration and fitted again using a Langmuir one-site binding model. Here, the model parameters binding dissociation constant (K_d; μM), maximum available binding sites (n_{max}; μmol/g cellulose) are reported in addition to the root mean square error (RMSE) for model fitting. Standard error from the mean for each parameter are reported here.

Langmuir One-Site Binding Model		
	Cellulose-I	**Cellulose-III**
$n_{\max}$	2.36 ± 0.10	1.48 ± 0.11
K_{d}	2.15 ± 0.46	6.15 ± 1.75
RMSE	0.22	0.17
Langmuir Two-Site Binding Model		
$n_{\max,1}$	1.81 ± 0.03	3.66 ± 0.22
$K_{\mathrm{d},1}$	16.64 ± 1.35	227.30 ± 17.8
$n_{\max,2}$	0.99 ± 0.03	0.67 ± 0.02
$K_{\mathrm{d},2}$	0.28 ± 0.02	0.64 ± 0.05
RMSE	0.17	0.12
Langmuir Freundlich Binding Model		
$n_{\max}$	3.21 ± 0.05	7.58 ± 0.50
K_{d}	2.61 ± 0.06	19.10 ± 1.27
m	0.53 ± 0.01	0.42 ± 0.01
RMSE	0.17	0.13

Table 3. **Langmuir-based binding model parameters for GFP-CBM3a adsorption to Cladophora-based cellulose I and III**. Here, binding dissociation constant (K_{d}; μM), maximum available binding sites ($n_{\max}$; μmol/g cellulose), and Freundlich power constant (m) fitted parameters are shown. Model fitting details for all Langmuir-based adsorption models are provided in the SI appendix. The errors reported were standard errors to parameter fits obtained. Representative model fits for original CBM3a-cellulose binding data are shown in **SI Appendix Fig. S6**.

	A (x 10^{-12} bound molecules)	k_{on}^{*} (s^{-1})	k_{off} (s^{-1})
Cellulose I	145.55 ± 0.40	0.13 ± 0.02	4.60 ± 0.21
Cellulose III	97.71 ± 1.62	0.14 ± 0.01	11.30 ± 0.04

Table 4. Kinetic rate constants for GFP-CBM3a adsorption and desorption towards nanocrystalline cellulose allomorphs estimated using Quartz Crystal Microbalance (QCM-D) based binding assay data. Here, k_{on}^{*} is a pseudo-adsorption rate constant which is the product of true K_{on} and the free protein concentration, while k_{off} is the true desorption rate constant. Fitted parameter means and standard deviations from two replicate assays are reported here. Sauerbrey equation was used to obtain the mass of adsorbed protein on cellulose film (or as total number of bound molecules upon achieving full binding saturation as represented by '*A*' here) using the frequency change at third overtone. The equations used for raw QCM-D data fitting are shown in **SI Appendix** Experimental Procedures section. Representative QCM-D sensorgrams are shown in **SI Appendix Fig. S10.**

Supplementary Information (SI) for

Molecular origins of reduced activity and binding commitment of processive cellulases and associated carbohydrate-binding proteins to cellulose III

Shishir P. S. Chundawat,[a*] Bhargava Nemmaru,[a] Markus Hackl,[a] Sonia K. Brady,[b] Mark A. Hilton,[b] Madeline M Johnson,[b] Sungrok Chang,[b] Matthew J. Lang,[b,c] Hyun Huh,[d] Sang-Hyuk Lee,[d] John M. Yarbrough,[e] Cesar A. López,[f] Sandrasegaram Gnanakaran[f]

[a]Department of Chemical & Biochemical Engineering, Rutgers The State University of New Jersey, Piscataway, New Jersey 08854, USA.

[b]Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, Tennessee 37235, USA.

[c]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37235, USA.

[d]Department of Physics and Astronomy, Rutgers The State University of New Jersey, Piscataway, New Jersey 08854, USA.

[e]Biosciences Center, National Renewable Energy Lab, Golden, Colorado 80401, USA.

[f]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA.

Corresponding Author Name: Shishir P. S. Chundawat* (ORCID: 0000-0003-3677-6735)

Corresponding Author Email: shishir.chundawat@rutgers.edu

This PDF file includes:

Supplementary text (Experimental Procedures and Results)
Figures S1 to S16
Tables S1 to S3
Legends for Movies S1 to S3
SI References

Other supplementary materials for this manuscript include the following:

Movies S1 to S3

SI Appendix Experimental Procedures:

Crystalline Cellulose Isolation and Anhydrous Liquid Ammonia Pretreatment: High crystallinity cellulose I (called native Cladophora cellulose I) from *Cladophora* sp. (*Cladophora glomerata*) was isolated and characterized as described previously (1). High purity (>98% cellulose content, dry weight mass basis or dwb) plant-derived microcrystalline cellulose I (called Avicel cellulose I) was purchased from Sigma-Aldrich (Avicel PH-101, Lot No. BCBD6923V). These native cellulose samples were used to generate respective Avicel or Cladophora derived crystalline cellulose III using a suitable anhydrous liquid ammonia based pretreatment process (2, 3). All cellulose III samples were kindly treated by Dr. Leonardo Sousa using a typical anhydrous liquid ammonia treatment protocol as highlighted elsewhere (3). Briefly, cellulose III was prepared in a high-pressure stirred batch reactor at 90 °C for 30 min (for Avicel) or 4 h (for Cladophora) residence time using at least a minimum 6:1 anhydrous liquid ammonia-to-cellulose loading ratio (dwb). The reactor pressure was maintained constant at 1000 psi using nitrogen gas during the pretreatment, and ammonia was slowly evaporated from the reactor through a venting valve after the desired residence time. During this evaporation process, the temperature of the reactor was slowly decreased and kept stabilized at 25 °C. The treated cellulose sample was then removed from the reactor and placed overnight in the fume hood to evaporate any residual ammonia. All treated cellulose samples were stored at 4 °C in a zip sealed bag prior and were used directly without any further drying.

Cellulose Characterization using XRD & FT-Raman Spectroscopy: Details regarding the X-ray diffraction (XRD) method and data analysis methods/results are provided elsewhere (2, 3). Briefly, XRD was performed on an X-ray diffractometer with beam parallelized by a Gobel mirror (D8 Advance with Lynxeye detector; Bruker, Bruker AXS Inc., Madison, WI, USA). CuKα radiation (wavelength = 1.5418 Å) was generated at 40 kV with 40 mA current and the detector slit was set to 2.000 mm. Samples were analyzed using a coupled 2θ/θ scan type with a continuous PSD fast scan mode. The 2θ started at 8.000° and ended at 30.0277° with increments of 0.02151°, while θ started at 4.0000° and ended at 15.0138° with increments of 0.01075°. Step time was 1.000 s (i.e., 1025 total steps, effective total time 1157 s per run). Dry cellulose samples (approximately 0.5 g) were placed in a specimen holder ring made of PMMA with 25 mm diameter and 8.5 mm height, rotating at 5 degrees per minute during analysis. Cellulose crystallinity was estimated based on the Segal peak height (for Cladophora derived samples) and amorphous peak deconvolution based methods (4, 5). Please note that Miller indices used in this paper for each contributing predominant diffraction peak/s conform to the convention with 'c' as the fiber axis, a right-handed relationship among the axes and the length of a<b, as recommended recently by Alfred French (6), to avoid confusion with other naming conventions. Briefly, for the XRD Segal peak height method, cellulose crystallinity index was calculated from the ratio of the height of the (110) or (200) plane equatorial reflection peak and the height of the minimum between the (110) or (200) and (010) or (110) plane equatorial reflection peaks for Cladophora or Avicel PH-101 cellulose I, respectively. For cellulose III, cellulose crystallinity index was calculated from the ratio of the height of the (100) plane equatorial reflection peak and the height of the minimum between the (100) and (002) plane equatorial reflection peaks. Note that, the three main peaks for native Cladophora cellulose I one-chain triclinic unit cell have Miller indices of (100), (010) and (110), which are the counterparts to the (1-10), (110) and (200) peaks of Avicel PH101 cellulose I pattern. Peak deconvolution methods have been used extensively to calculate cellulose crystallinity index (5, 7–9). Avicel derived cellulose I and III samples were recently analyzed using the amorphous peak deconvolution method (10). XRD peak deconvolutions were carried out using PeakFIT (Version 4.12, Systat Software Inc, San Jose, CA) as described elsewhere (2, 5). For all peak deconvolutions F values are always > 30,000 while R-squares > 0.999.

Additional supporting details regarding the FT (Fourier Transform) Raman based spectroscopic characterization methods/results are provided here as well. Briefly, a MultiRam FT-Raman spectrometer (Bruker) was used to collect Raman spectra for cellulose samples. The FT-Raman spectrometer was equipped with a 1064-nm 1000-mW Nd:YAG laser. For Raman analysis, cellulose pellets were first prepared from either air-dried or lyophilized samples prior to analysis. In most cases, spectra with high signal-to-noise (S/N) ratios was obtained upon using a 660 mW laser power setting and collecting over 512 scans per sample. The spectra were converted to ASCII format and

exported to Microsoft Excel for direct plotting/analysis. The interconversion of cellulose I to III was confirmed based on previously published reports using Cladophora or cotton linters derived cellulose allomorphs (3, 11–14). Peak assignments of the vibrational spectrum of cellulose I and III have been described elsewhere (3, 13, 15). Briefly, 250-550 cm^{-1} region for cellulose has predominant group motions attributed to skeletal-bending modes involving C-C-C, C-O-C, O-C-C, and O-C-O internal bond coordinates. The 550-750 cm^{-1} region corresponds to mostly out-of-plane bending modes involving C-C-C, C-O-C, O-C-O, C-C-O, and O-H internal bond coordinates. The peaks around 900 cm^{-1} are shown to involve bending of H-C-C and H-C-O bonds localized at C-6 atoms of the hydroxymethyl group. The 950-1200 cm^{-1} region corresponds to mostly stretching motions involving C-C and C-O internal bond coordinates. The 1200-1500 cm^{-1} region corresponds to mostly bending motions involving H-C-C, H-C-O, H-C-H, and C-O-H internal bond coordinates. The region of 1400-1500 cm^{-1} for cellulose has been shown to be particularly sensitive to the CH_2 scissor bending modes that are sensitive to the *Trans-Gauche* or TG (1480 cm^{-1}) and *Gauche-Trans* or GT (1460 cm^{-1}) conformations of the C6-hydroxymethyl group (13).

***Enzymatic hydrolysis of cellulose allomorphs*:** All Cladophora cellulose samples were subjected to enzymatic hydrolysis using a commercial cellulase cocktail at 0.5% glucan loading in a 1 ml reaction volume. All hydrolysis assays were carried out in 2 mL flat-bottomed microcentrifuge tubes using 5 mg of pre-weighed cellulose suspended in 500 µL of 50 mM Na-Acetate buffer, at pH 5.0 (for Cellic C.Tec2 (Novozymes A/S, Denmark)), along with suitably diluted stock enzyme solution to achieve desired enzyme loadings (i.e., mg total enzyme loaded per gram of added cellulose per well). The total Cellic C.Tec2 enzymes (Novozymes A/S, Denmark) loadings used during enzymatic hydrolysis was fixed at 5 mg/g glucan loading, unless specified otherwise. The protein concentration (193 mg/ml) for the C.Tec2 enzyme stock solutions was determined using the Kjeldahl method (16). Sodium azide was added to prevent any microbial growth (0.1% w/v final concentration). All tubes were incubated at 50 ˚C in an orbital shaking ThermoMixer (Eppendorf) incubator set at 1000 RPM for the desired saccharification time (0-96 h). A similar procedure was used for enzymatic hydrolysis using an equimolar mixture of purified Cel7A and Cel7B enzyme or Cel7A enzyme alone. For the Cel7A-Cel7B mixture experiment, the enzyme loading for Cel7A was kept at 10 mg enzyme per g cellulose and β-glucosidase was added to prevent inhibition by cellobiose. For the Cel7A activity assay experiment alone, three enzyme loadings were tested (0.5, 2.5, and 10 mg enzyme per g cellulose). Cel7A, Cel7B, and β-glucosidase were purified and isolated from commercial enzyme sources as outlined previously (17). The hydrolyzate supernatants were analyzed for total reducing sugar concentrations using the standard dinitrosalicylic acid (DNS) colorimetric assay as reported earlier (18). Briefly, 30 µl of the hydrolysate supernatant (w/wo 2-fold dilution) was incubated with 60 µl of DNS stock reagent in PCR tubes/plates at 95 °C for 5 min in an Eppendorf thermal cycler. After the PCR plates cooled down to room temperature, the DNS reaction mixture was transferred and diluted in DI water using a clear, flat-bottom microplate for finally measuring solution absorbance at 540 nm. Suitable reducing sugar standards (e.g., glucose standards ranging from 0.1–5 g/l) were included for the DNS assay. All hydrolysis experiments were carried out in duplicates. Error bars reported represent one standard deviation (±1σ) from mean values for replicate assays.

Sample preparation for AFM imaging of Cladophora CI and CIII: Approximately 5 mg of dried cellulose I (CI) and cellulose III (CIII) fibers derived from *Cladophora glomerata* were each added to a microtube and suspended in 1 ml of DI water. At first, the fibers were manually dispersed through pipetting the suspension up and down using a wide opening 1 ml pipette. Subsequently, the suspensions were sonicated for 1 minute (model FB705 Fisher Scientific, USA, settings 10% amplitude, 2 s on, 5 s off), then pipetted up and down until a segregation of the fibers was observed. Two hundred microliters (200 µl) of the resulting suspensions were transferred to new microtubes and filled up to 1 ml with DI water. The fibers were further fragmented by pipetting through a 1 ml pipette until all large aggregates were dispersed. The suspensions were stored at 4°C until use and resuspended prior to usage.

AFM imaging of Cladophora CI and CIII microfibrils: The microscope cover glasses (No. 1.5, 22x22mm, VWR, USA) were rinsed in the following order, DI water, acetone (NF/FCC grade, Fisher Scientific, USA) and DI water and then dried with a stream of nitrogen. Twenty microliters (20 µl) of cellulose I and III samples were each placed in the middle of the glass slide and dried over night at 50°C. Non-contact mode AFM measurements were carried

out with a Park systems NX10 AFM using non-contact cantilever (SSS-NCHR, Park Systems, South Korea) with a force constant of 42 N/m (specific range: 10-130) and a resonant frequency of 330 kHz (specific range: 204-497). For each of the substrates, 2.5 x 5 μm^2 sized areas were chosen at random places of the sample. Data were analyzed using the XEI software (Park Systems, South Korea).

Functionalization of beads for tweezer binding-rupture and motility assays: For the motility assays, purified Cel7A enzymes were tethered to polystyrene beads and assayed for their motility on Cladophora derived cellulose I and cellulose III based on identical methods reported earlier in our Cel7A tweezer motility study (1). For the rupture assays, CBM1 was tethered to polystyrene beads via the His_8-tag on the N-terminus of our purified GFP-CBM1 construct (details on GFP-CBM construct design and purification are provided below), with minor modifications from our previously published work (1). Using PCR, 1,010-bp DNA linkers were created from the M13mp18 plasmid template with a biotin tag on one end and an amine group on the other. The anti-His antibody was crosslinked to the amine group using a sulfo-SMCC (sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate) intermediate. In the cases of using the anti-His Fab, the anti-His antibody was cleaved using 3-MEA (2-Mercaptoethylamine) before crosslinking. To functionalize the beads with GFP-CBM1, 1.09 μm streptavidin beads (Spherotech), biotin/anti-His functionalized DNA linkers, and His_8-tagged GFP-CBM1 constructs were incubated together in PBS at 4°C for 45 minutes on a rotator. After incubation, the beads were washed by spinning down at 9000 rpm for 3.5 minutes, removing the unreacted components in the supernatant, resuspending in 50 mM acetate buffer (pH 5.0), and sonicating for 2 minutes at 20% amplitude. This process was repeated two more times. Beads were functionalized such that, statistically, zero or one GFP-CBM1 molecule is bound to each bead. This was determined through serial dilution until a maximum of half the beads bound to cellulose fibers during the experiment.

Cellulose solution and slide preparation for tweezer binding-rupture and motility assays: Purified and dried cellulose samples (*Cladophora* based cellulose I or III) were used to create a heterogeneous cellulose mixture by first mixing the desired cellulose sample to deionized water in a 1 mg/mL ratio. The mixture was then sonicated for 2 minutes at 50% in a cup sonicator and vortexed for 15 s on high setting. The cellulose, still clumped at this point, was pulled up and down in solution with a 16-gauge syringe for 1-2 minutes before going back on the vortex for 15 s. These steps were repeated three times. The resulting mixture was then diluted in a 1:20 ratio by mixing 500 μL of the prepared solution with 500 μL deionized water. This slurry suspension was then stored at 4 °C. Whatman Grade 1 Filter Paper based cellulose stock suspension slurry was prepared as described previously (1), to be used for some control GFP-CBM1 binding-rupture assays. When preparing to load a slide, a small sample (~100 μL) of the stored cellulose mixture is removed from the stock and the cellulose pulled apart by sonicating for 2 minutes at 50% in a cup sonicated. This solution was directly loaded onto the glass slide. Slides are prepared by creating a 10-15 μL volume flowcell using a KOH etched coverslip and double-sided sticky tape. The stock cellulose solution (*Cladophora* based cellulose I or III) was then added to the flowcell and allowed to dry out in an oven at ~95 °C for an hour, allowing cellulose fibrils to non-specifically bind to the slide surface. The surface was then blocked with 10 mg/mL BSA in acetate buffer (pH 5.0) for 15 minutes to prevent non-specific sticking of the beads to the glass surface. Finally, the GFP-CBM1 functionalized beads solution was loaded onto the slide and the slide sealed shut. For the Cel7A motility assays, 0.75 μm non-functionalized polystyrene beads (Spherotech—PP-08-10) were allowed to nonspecifically adhere to the coverslip surface, in an incubation step before BSA blocking, to serve as fiducial markers allowing for instrumental drift tracking during data acquisition.

Single molecule tweezer binding-rupture assay data acquisition and analysis: CBM1 functionalized beads were trapped using a 1064-nm laser setup as described before (1), and placed alongside a surface-bound stationary fiber. Experiments were conducted at a fixed room temperature (21 °C). After position calibration and trap stiffness measurements, the bead was actively placed on a cellulose fiber roughly running along the axis of the microscope stage. Upon binding, the bead was centered, acquisition started, and a force applied to the tethered bead by stepping the piezo stage along the axis of the fiber. With force applied, the position of the bead is held until rupture. Once a tether is ruptured, it is sometimes possible to tether the bead to the fiber again, in which case, the same method of

force application is applied while data acquisition continues. Data were collected at a 3-kHz sampling frequency and then filtered with a 10-point exponential moving average before analysis. Custom Matlab codes were then used determine the rupture forces and the bond lifetimes of full ruptures. The force-lifetime data was binned every 2.5 pN and then we tried to fit the data to a single or a double exponential decay characteristic of a slip bond.

GFP-CBMs gene synthesis and cloning: *E. coli* codon optimized genes encoding CBMs, with additional flanking *AflII* and *BamHI* restriction sites, inserted into a standard pUC57-Kan vector were ordered from Genscript USA Inc (Piscataway, NJ). DNA sequences for all CBMs are provided in the table below. An *E. coli* expression vector pEC-GFP-CBM3a was kindly provided by the Fox lab (UW Madison). Sequence information regarding the family 3a CBM from *Clostridium thermocellum* expressed using this pEC vector have been published already (19). The pEC vector sequence map and strategies for primer design and CBM genes sub-cloning have been reported already (19, 20). Briefly, polymerase incomplete primer extension (PIPE) based ligation independent cloning approach was used to transfer the CBM nucleotide gene sequences from the respective pUC57 to pEC vector (21). For the creation of CBM1 Y31A mutant used for bond rupture assay (reported in **Fig. S14**), site-directed mutagenesis with complementary forward and reverse primers was used. Polymerase chain reaction (PCR) was catalyzed by Herculase II Fusion DNA polymerase (Agilent Technologies, Santa Clara, CA). Destination pEC vector and CBM insert gene amplification was carried out using suitably designed PIPE primer pairs. The PCR amplification of pUC57 and pEC vectors using corresponding vector/insert PIPE reactions primer pairs were carried out in separate tubes. After PCR, respective CBM PIPE reaction product aliquots (2 μL) were mixed together and immediately transformed into competent *E. coli* E. cloni 10G cells (Lucigen, Madison, WI). If the PIPE cloning strategy was not successful, the pUC57 and pEC vectors were digested using *AflII* (New England Biolabs, Ipswich, MA) and *BamHI* (Promega, Madison, WI) restriction enzymes. The restriction enzyme products were ligated using T4 DNA ligase (New England Biolabs) and the ligation mixture was instead transformed into competent *E. cloni* 10G cells. Individual transformant colonies were next screened by PCR amplification and transformants containing inserts with the approximate correct size were identified by agarose gel electrophoresis. Plasmids isolated from positive colonies were sequenced to confirm nucleotide identity at the UW Biotechnology Center (and/or at Genscript, Piscataway, NJ). Transformed strains were stored as 20% glycerol stocks were maintained at -80 ˚C, while all relevant pEC-GFP-CBM plasmids were also maintained at -80 ˚C for long-term storage.

CBM Family	CBM Gene Sequence	Organism Source
1	CCGGGTCCGACCCAGAGCCATTATGGCCAGTGCGGTGGTATTGGTTA TAGCGGTCCGACCGTGTGCGCAAGCGGTACCACCTGCCAGGTGCTG AACCCGTATTATAGCCAGTGCCTG	*Trichoderma reesei*
2a	GCGGCGAGCGGCGCACGTTGCACCGCAAGTTATCAAGTGAATAGCG ATTGGGGGAACGGTTTCACGGTTACCGTCGCAGTTACCAACTCAGGT TCTGTTGCTACCAAAACCTGGACGGTGTCGTGGACCTTCGGCGGTAA TCAGACTATCACCAACAGCTGGAACGCGGCGGTCACACAGAACGGC CAGAGTGTGACTGCACGTAACATGAGCTACAATAATGTTATTCAACC AGGCCAAAATACGACCTTTGGTTTTCAAGCCTCGTACACGGGCAGTA ACGCAGCACCGACCGTTGCGTGCGCGGCGAGT	*Acidothermus cellulolyticus*
2a	Refer to Lim et al. 2014 (20)	*Streptomyces sp. SirexAA-E*

3a	Refer to Whitehead et al. 2017 (19)	*Clostridium thermocellum*
5	ATGGGTGATTGTGCTAACGCAAATGTCTATCCGAACTGGGTGTCTAA AGATTGGGCGGGTGGTCAACCGACGCATAACGAAGCGGGTCAGAGC ATTGTGTATAAAGGCAACCTGTACACCGCGAATTGGTACACCGCATC AGTGCCGGGTTCAGACTCATCGTGGACGCAGGTTGGTAGTTGTAATT GA	*Erwinia chrysanthemi*
10	ATGGGCAATCAACAATGTAACTGGTATGGCACCCTGTATCCGCTGTG TGTGACGACGACGAATGGCTGGGGCTGGGAAGATCAACGCAGCTGC ATCGCCCGTAGCACCTGCGCGGCTCAACCGGCACCGTTTGGCATCGT GGGTAGCGGCTGA	*Cellvibrio japonicus*
64	CCGACCCCGTCTGGCGAATATACGGCGATTGCCCTGCCGTTTACCTAC GATGGCGCCGGTGAATATTACTGGAAAACCGACCAATTCAGCACCGA TCCGAATGACTGGTCACGTTATGTCAACTCGTGGAATCTGGATCTGCT GGAAATTAACGGTACCGACTACACGAATGTGTGGGTTGCACAGCATC AAATCACGCCGGCTAGTGATGGCTACTGGTATATTCACTACAAAGGC TCGTATCCGTGGTCGCATGTGGAAATCAAA	*Spirochaeta thermophila DSM 6192*

The sequence-verified plasmid, named pEC-GFP-CBM, encoded a 5' His_8-tag, followed by GFP, a linker sequence, and the relevant CBM. The linker sequence encoded for a 42 amino acid linker peptide reported previously by Takasuka and co-workers (22). Here, we have characterized in detail two distinct but structurally homologous family 2a CBMs from *Streptomyces* sp. SirexAA-E and *Acidothermus cellulolyticus*. *Streptomyces* sp. SirexAA-E was recently identified by GLBRC researchers for its high cellulolytic ability (23). In addition, we have also characterized two distinct but structurally homologous family 1 CBMs from *Trichoderma reesei* belonging to Cel7A (cellobiohydrolase I or CBHI) and Cel6A (cellobiohydrolase II or CBHII) processive cellulases that gave similar results to CBM1 from Cel7A. However, here we exclusively report results for CBM1 from Cel7A. Previous studies have shown that the structure and function relationship of homologous CBMs belonging to the same family are often similar, as also seen in our current work when comparing binding results between the two family 2a CBMs.

Small-scale expression testing of GFP-CBMs: pEC-GFP-CBM plasmids were first transformed into BL21-CodonPlus-RIPL [λDE3] (Stratagene, Santa Clara, CA) or Rosettagami 2 [DE3] (Novagen, Santa Clara, CA) *E. coli* competent strains for small-scale protein induction/expression optimization screening. After cells were grown to the optical density of 0.5-0.7 (mid-exponential phase) in a non-inducing medium (24), expression was induced in either an auto-induction medium (24) at 25 ˚C or in LB medium with varying concentrations of IPTG (0.1-1mM) during incubation at 16 ˚C, 25 ˚C, or 37 ˚C. Apart from GFP-CBM2a (ActE), as also reported previously (20), soluble cytoplasmic protein production was observed for all other GFP-CBMs in nearly all of the expression conditions tested. Superfolder enhanced GFP (or eGFP) tag has been reported to increase soluble fusion protein production yields by likely preventing aggregation of hydrophobic proteins like CBMs (25). Nevertheless, small-scale immobilized metal affinity chromatography with Ni^{2+}-NTA (immobilized metal affinity chromatography or IMAC) purification and cellulose binding assays were conducted for all soluble protein fractions to confirm

cellulose-binding activities and hence identify optimum cell culture conditions for large-scale protein production/purification to run all reported bulk and single-molecule CBM binding assays.

Large-scale expression of GFP-CBMs: *E. coli* BL21-CodonPlus-RIPL [λDE3] (Stratagene, Santa Clara, CA) or Rosettagami 2 [DE3] (Novagen, Santa Clara, CA) competent strains were transformed with the relevant pEC-GFP-CBM plasmid based on the small-scale expression results. Suitable transformants were inoculated into 50 mL of chemically defined non-inducing medium (24), in the presence of 50 μg/mL kanamycin and 25 μg/mL chloramphenicol selection antibiotics. The non-inducing medium contained 2 mM $MgSO_4$, a 1:1000 dilution of trace metal salts mixture (equivalent to 50 mM Fe^{3+}, 20 mM Ca^{2+}, 10 mM Mn^{2+}, 10 mM Zn^{2+}, 2 mM Co^{2+}, 2 mM Cu^{2+}, 2 mM Ni^{2+}, 2 mM Mo^{6+}, 2 mM Se^{4+}, 2 mM H_3BO_3) into the medium, 0.5% glucose, 0.25% aspartate, 50 mM NH_4Cl, 25 mM KH_2PO_4, 25 mM Na_2HPO_4, 5 mM Na_2SO_4, 0.01% methionine, 1% of 17 amino acids (except cysteine, tyrosine, and methionine) each, and a vitamin cocktail (200 nM of vitamin B12, nicotinic acid, pyridoxine, thiamine, p-aminobenzoic acid, and pantothenate; 5 nM folic acid, and riboflavin). The culture was incubated overnight at 25 ˚C and then used to inoculate 2 liters of auto-induction medium (24). The auto-induction medium contained 1.2% tryptone, 2.4% yeast extract, 2.3% KH_2PO_4, 12.5% K_2HPO_4, 0.375% aspartate, 2 mM MgSO4, 0.8% glycerol, 0.015% glucose, and 0.5% α-lactose. The cultures were grown at 25 ˚C for ~27 h. The cells were harvested by centrifugation at 8000xg for 10 minutes at 4˚C and the cell pellet was stored at -80˚C until further use. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO).

Purification of GFP-CBMs: The recovered cell pellet was thawed and re-suspended in 150 mL of ice cold 20 mM phosphate, pH 7.4, containing 500 mM NaCl, 20% v/v glycerol, 10 μg/ml lysozyme, and a protease inhibitor cocktail (containing benzamidine, EDTA and E-64 protease inhibitor from Sigma-Aldrich). The cells were sonicated with an ultrasound sonicator (550 Sonic Dismembrator, Fisher Scientific, Pittsburgh, PA) fitted with a microprobe (1-inch probe diameter) at 4 ˚C for 5 min with 10-s on-bursts and 30-s off periods. The cell debris containing the inclusion bodies was pelleted at 21,000 rpm at 4 ˚C (30 min) and the supernatant was collected in all cases except for GFP-CBM2a (ActE). Details regarding GFP-CBM2a (ActE) expression and purification are provided elsewhere (20). Briefly, due to the insolubility of the expressed GFP-CBM2a (ActE) under all conditions tested, this protein construct was first isolated from inclusion bodies, refolded, and then purified using IMAC as described previously (20). For all other GFP-CBMs, IMAC using Ni^{2+}-NTA based columns/media (GE Healthcare) was first used to isolate and purify His_8-tagged proteins from the *E. coli* cell lysate. All column-based protein purifications were carried out on a ÄKTA-FPLC system (GE Healthcare, Pittsburgh, PA). The cell lysate supernatant was first loaded onto the IMAC column at a medium flow rate of 1 – 2 ml/min. The column was then washed with buffer A (100 mM MOPS, pH 7.4, containing 10 mM imidazole and 100 mM NaCl), followed by additional washing using 95% IMAC buffer A spiked with 5% IMAC buffer B (100 mM MOPS, pH 7.4, containing 500 mM imidazole and 100 mM NaCl), and last followed by elution in 100% IMAC buffer B at a flow rate of 5 ml/min. Protein purity and molecular weight at each stage of the protein purification process was examined by SDS-PAGE (Criterion XT Bis-Tris Precast Gels, Bio-Rad). The presence of partially cleaved GFP-CBMs was identified in the IMAC-B eluents for some protein constructs (namely CBM1, CBM2a, CBM5, CBM10), which necessitated further purification using an amorphous cellulose or hydrophobic interaction affinity-based purification method, as already outlined elsewhere (20, 26), to isolate the intact protein fractions. Briefly, for cellulose affinity-based purification method, IMAC-B protein eluents were directly applied to a phosphoric acid swollen amorphous cellulose (PASC) media at the recommended loading (~200 mg crude protein added per gram dry weight cellulose) for preparative-scale purification (26). The amorphous cellulose slurry was prepared ahead of time and preequilibrated in a 50 mM pH 6.5 MES buffer (equilibration buffer or buffer A) at the desired solids concentration (10 g/L), prior to addition of the IMAC-B protein eluent. The crude protein-cellulose slurry was then intermittently and gently mixed at room temperature for a total incubation time of 0.5 h. The protein bound to PASC was then separated from the unbound protein in the supernatant by gentle centrifugation at 3500xg for 10 min at 25˚C. The recovered PASC pellet was then resuspended in a wash buffer (i.e., equilibration buffer+1M NaCl), using a 4:1 buffer to PASC pellet ratio (v/v), and gently mixed at room temperature for 10 mins to remove non-specifically bound proteins. The recovered PASC pellet containing the adsorbed GFP-CBMs was then finally suspended in

100% ethylene glycol elution solution, using a 4:1 glycol to PASC pellet ratio (v/v). The final ethylene glycol concentration of ~80% (v/v) was sufficient to elute a significant fraction of reversibly bound GFP-CBMs into the supernatant. The eluted protein rich supernatant was separated from PASC pellet and stored in 80% glycol solution at −20 ˚C for short term storage or immediately concentrated using IMAC columns prior to buffer exchange into 10 mM pH 6.5 MES (or pH 5.5) buffer and storage at −80 ˚C for long term storage in 0.5-1 ml aliquots. The molecular weight of the intact purified GFP-CBM monomers was confirmed by SDS-PAGE to match with the predicted translation products. Protein concentrations were estimated spectrophotometrically at 280 nm using the extinction coefficients calculated from the amino acid sequences for each construct. The histidine tags were not removed and have been reported to not influence CBM binding to cellulose (20, 27).

GFP-CBM 'pull-down' binding assays with cellulose allomorphs: Cladophora based cellulose allomorphs, prepared as described above, were used to estimate binding affinity and partition coefficients for GFP-CBMs. All cellulose binding assays were carried out in 2 mL microcentrifuge tubes using 5 mg of pre-weighed cellulose suspended in 500 µL of 10 mM MES buffer, at pH 5.5 (for CBM1) or 6.5 (for all other CBMs), containing 2.5 mg/mL of bovine serum albumin (BSA) to minimize non-productive protein binding to the tube wall and/or denaturation at the air-buffer interface during mixing. Purified GFP-CBMs aliquots were thawed and buffer-exchanged (PD-10 desalting column, GE Healthcare) into 10 mM MES, at desired pH, and added to each 2 mL tube to achieve a final solution concentration. The final concentration typically ranged from 1-500 µg/mL for estimating partition coefficients or from 1 µg/mL to 10 mg/mL for full scale binding affinity measurements. The tubes were mixed in an orbital mixer set at 1000 rpm (ThermoMixer, Eppendorf) at 25˚C for 1.5-2 h to allow binding equilibrium to be reached. Controls with only proteins but no substrate were mixed to track total added protein concentration, while never-mixed controls with only proteins/buffer were used as controls to account for possible protein loss during mixing. Insoluble cellulose was recovered by centrifugation at 17000xg for 2 min at 25˚C and 200 µL of the supernatant (w/wo suitable dilution in identical buffer) containing unbound protein was assayed for GFP fluorescence (488 nm excitation and 509 nm emission; with 495 nm cut-off) to quantify the total fraction of bound protein versus the original added amount (2). All assays were performed in at least duplicate for each protein loading condition (~22 protein loadings). Error bars reported for all bulk binding assays represent one standard deviation (±1σ) from mean values of replicates.

Calcofluor White dye 'pull-down' binding assays with cellulose allomorphs: Calcofluor White dye (Sigma-Aldrich) binding assays on Avicel PH-101 based cellulose I and III were performed in 1.5 ml Eppendorf tubes at varying dye concentrations, similar to previously reported cellulose-solute binding assay methods (2, 17). Briefly, 5 mg of Avicel cellulose-I or cellulose-III was added to each well by pipetting suitable aliquots of a uniformly suspended cellulose slurry prepared in deionized water. Suitable calcofluor dilutions were prepared in deionized water (at pH 7.0) to achieve the desired concentration in each tube. Note that 30 mM NaCl was added to each tube. For cellulose blanks, deionized water alone instead of cellulose was added to make up the volume instead to the final desired 620 µl level. All binding assays were performed with three replicates per assay condition. The tubes were then shaken in Thermomixer at 1000 rpm for 1 h at room temperature. After 1 hour, all tubes were centrifuged at 17000xg for 5 minutes at 25˚C and supernatant was removed from each well to be transferred into opaque microplates for reading calcofluor fluorescence. The plates were read using a Molecular Devices M5e spectrophotometer at the following settings: 365 nm excitation, 450 nm emission. Langmuir based models were also fitted to the binding dataset as described below. Error bars reported for all binding assays represent one standard deviation (±1σ) from mean values from three replicates.

Langmuir-type adsorption model fitting to pull-down adsorption assay data: The bound (µmol protein/g cellulose) and free (µM) protein concentrations from the binding assays were fit to a Langmuir single-site, two-site, and Langmuir-Freundlich adsorption models to determine maximum binding capacity ($n_{\max}$) and equilibrium dissociation constant (K_{d}) for each cellulose allomorph and GFP-CBM combination. The model equations are displayed in **Figure 4B**. The linear range of the binding curve was used to estimate the partition coefficient (α) as described previously (20). The data from pull-down binding assays is processed to yield free protein and bound

protein concentrations (or calcofluor concentrations) in μM and μmol/g cellulose respectively in Excel™. This data set was then exported into OriginPro 2019. OriginPro software already has several non-linear curve fitting options available. However, separate functions were created for Langmuir single-site, two-site, and Langmuir-Freundlich models using the custom function builder tool. Similar Langmuir-type models have been used previously for Cel7A-cellulose binding analysis (2, 17). These functions were then used to fit binding assay data with the following settings for curve fitting: Levenberg-Marquardt algorithm with a tolerance of $1e^{-9}$. If the fitting analysis did not converge in Origin for more complex models such as Langmuir two-site and Freundlich models, Excel Solver™ was used for curve fitting in those cases. The sum of errors was specified as objective function and parameters were estimated using the GRG non-linear method. Parameter uncertainties were then calculated using an elaborate process involving Monte Carlo simulations as detailed elsewhere (28).

Reversible binding of GFP-CBMs to cellulose*:* The application of Langmuir-type adsorption models is dependent on reversible binding of protein to substrate. Hence, reversible binding properties of the GFP-CBMs to Cladophora cellulose I and cellulose III, was determined as described previously (20, 29). This approach was used to confirm the reversible binding of GFP-CBM1 and GFP-CBM3a to both Cladophora cellulose allomorphs in the pull-down binding assay setup (data not shown). Briefly, after the binding equilibrium was re-established after dilution of the mixture, the newly estimated bound/free protein concentration was confirmed to lie along the adsorption isotherm curve as already shown by us previously for CBM2a (ActE) (20). Since the partition coefficient was determined regardless of the dilution ratio, this result confirmed that the protein showed indeed reversible adsorption to cellulose. The presence of excess BSA prevented non-specific binding of CBMs to various surfaces like plastic tube walls (30), which along with reduced possible GFP-CBM denaturation at air-liquid interfaces during extensive mixing at low protein concentrations due to presence of sacrificial BSA (31), also minimized bias in reversibility binding measurements of GFP-CBMs particularly to cellulose III.

Molecular dynamics simulation of CBM1 and cellulose allomorphs interactions: All molecular dynamics (MD) simulations were conducted using the Amber force field for CBM1 protein (1cbh pdb code) (32), TIP3P explicit water model (33), and the Glycam force field for representing the cellulose fiber (34). As described previously (35), X-ray and neutron diffraction coordinates were used to generate one rhomboid cellulose Iβ fibril and one rhomboid cellulose III fibril (36, 37). Each fibril consisted of 36 glucan chains each. Two different types of crystalline cellulose III were used for the simulations, one with lower crystallinity and another with high crystallinity (based on the available crystalline core cellulose crystal structure). However, since no detailed structural information is available for lower crystallinity cellulose III (3), an idealized model structure was obtained by increasing the simulation temperature to 400 K starting with the high crystallinity cellulose III structure to increase overall disorder within the crystalline fiber. The cellulose Iβ rhomboid shape was chosen as the base case control because of its wide hydrophobic surfaces accessibility which has shown to be preferred binding site for Type-A CBMs like CBM1 (38, 39). Furthermore, this hydrophobic binding surface plane of cellulose Iβ is also identical to cellulose Iα (40), from a crystallographic point of view, which would aid in drawing similar conclusions when assessing CBM1 binding to hydrophobic surfaces of native cellulose I fibrils from Cladophora, which is mostly enriched in cellulose Iα unlike cellulose I fibrils from Avicel derived cellulose nanocrystals that are enriched in cellulose Iβ.

Two types of simulations were run: (a) First, unbiased simulations were used to probe the binding dynamics of CBM1 starting from the hydrophobic surfaces of cellulose I and III microfibrils on microsecond timescales. MD simulations were conducted in order to populate the most preferred orientations of CBM1 on either cellulose allomorph surface. CBM1 was first aligned in the canonical direction for the reducing end specific action of Cel7A to processively hydrolyze cellulose chain (i.e., the Y31 residue end of CBM1 pointing towards the nonreducing end of cellulose). If no stable binding was seen, as was the case with cellulose III, the CBM1 was aligned in the most stable conformation identified in the unbiased simulations (i.e., Y31 residue end of CBM1 pointing towards the reducing end of cellulose). From these unbiased simulations, we were able to calculate the distribution frequency of the most preferred rotameric states for the closely cellulose-contacting aromatic residues in CBM1. (b) Second, the preferred orientation inferred from unbiased simulations for CBM1 on cellulose I and III allomorphs were used

in order to estimate their respective association free energy with the cellulose fiber. Thus, a potential of mean force (PMF) was calculated to estimate the CBM1 binding free energy during adsorption to the hydrophobic surface of all cellulose allomorphs. We used the coordinates of the last frame from the unbiased CBM1-cellulose systems at 310 K. A total of 14 independent windows per system were used, which were spaced apart by 1 Å. A restraining potential of 1000 kJ mol^{-1} nm^{-2} was applied to the center of mass (COM) of CBM1 with respect of the COM of either cellulose I or cellulose III along the normal (z) coordinate. For each window, 1.5 μs long simulations were performed. The desorption free energy was reconstructed using the weighted histogram approach and convergence was assessed via block averaging by dividing the trajectory in three independent blocks. Umbrella sampling-based simulations were conducted along a path where the CBM1 translated from the hydrophobic surface of the microfibril to the bulk solvent. The position of the potential changes every 1 Å along the pulling reaction coordinate (in this case the z vector). Harmonic potential applied to the center of mass of the protein. Along the trajectory, the two most populated configurations are shown here in the main text figure, together with the protein residues in close contact with the external cellulose surface. These configurations also correspond to the minimum wells seen in the energy plots.

All simulations were performed using a 2-fs time step using GROMACS software. The LINCS algorithm was applied to constrain all bond lengths with a relative geometric tolerance of 10^{-4}. Non-bonded interactions were handled using a twin-range cutoff scheme. Within a short-range cutoff of 0.9 nm, the interactions were evaluated every time step based on a pair list recalculated every five-time steps. The intermediate-range interactions up to a long-range cutoff radius of 1.4 nm were evaluated simultaneously with each pair list update and were assumed constant in between. A PME method was used to account for electrostatic interactions with a grid spacing set to 0.15 nm. During the equilibration (0.1 μs), systems were coupled using a Berendsen barostat to 1.0 bar via an isotropic pressure approach, with relaxation time of 1.0 ps. Afterwards, system was coupled to a Parrinello barostat algorithm and constant temperature was maintained by weak coupling of the solvent and solute separately to a velocity-rescaling scheme with a relaxation time of 1.0 ps.

Preparation of Avicel cellulose nanocrystals through acid hydrolysis for QCM-D: Avicel cellulose III (or CIII) was prepared from Avicel PH-101 (also referred to as Avicel cellulose-I (or CI) (Sigma Aldrich Lot#BCBG9043V) as described previously (3). Nanocrystals were prepared from both Avicel cellulose-I and Avicel cellulose III using the same procedure. Briefly, 2 g of Avicel was added to 70 ml 4 N HCl in a glass beaker and placed in a pre-heated water bath at a temperature of 80^{o} C. The slurry was stirred every half hour using a spatula to ensure cellulose is well suspended. After 4 hours of reaction time, the acid hydrolysis mixture was diluted with 50 ml deionized water. The slurry was then aliquoted into 50 ml centrifuge tubes, with 40 ml slurry in each tube and centrifuged at 1600xg for 10 minutes at 25 ^{0}C. The supernatant was decanted, and the cellulose pellet was washed with 10 ml deionized water. The wash steps were repeated, and the supernatants were discarded until they turned hazy around pH 3.3. The haziness of supernatant indicates evolution of cellulose nanocrystals and hence these supernatants were collected into a separate bottle for future usage.

Preparation of cellulose thin films for QCM-D: This procedure was developed based on similar previous studies investigating cellulose hydrolysis and binding of cellulases using QCM-D (41). Briefly, 4.95 MHz quartz crystal sensors (0.55" diameter), with SiO_2 coating were purchased from Filtech (product code QSX0303). The sensors were first rinsed in water, followed by ethanol and then blow-dried. The sensors were then immersed in 0.02% PDADDMAC (Sigma Aldrich 409022), which serves as an anchoring layer for cellulose, for 1 hour at 25^{o}C with orbital mixing. This was followed by washing with deionized water for 1 hour with orbital mixing at 25^{o}C. The sensors were then blow-dried and spin-coated using a pre-cycle spin for 3 s at 1500 rpm, followed by a spin cycle for 60 s at 3000 rpm. This spin coating step was repeated 10-20 times to obtain a uniform cellulose film thickness of ~20 nm (as measured using the QSoft software using Sauerbrey model). In a previous study (41), a lesser number of spin-coating steps was used to achieve the same thickness, however, this discrepancy may be related to the nanocrystal slurry concentration and hence needs to be optimized accordingly.

Fluorescence recovery after photobleaching (FRAP) to study CBM-cellulose binding kinetics: We assembled a simple fluidic chamber using a glass coverslip (24 mm X 30 mm, No. 1.5, Thermo Scientific) and a holed glass slide (470150-480, Ward's Science) that was pre-cleaned with acetone (A949, Fisher chemical), methanol (A454, Fisher chemical), and distilled water. To improve the cellulose sample adhesion, the glass substrates were treated with plasma (PDC-001, Harrick Plasma) for 10 minutes. A drop of 20 µL cellulose solution was deposited on a coverslip, which was dried in a hybridization oven (VWR) at 50°C overnight. A sample chamber was assembled by attaching the coverslip to a glass slide with multiple pieces of double-sided 3M tape; the gap between the edges of two adjacent tapes formed a channel. The open edges were sealed with 5 minutes epoxy glue (Devcon). Next, 100 µL of 5 µM GFP-CBM3a buffered solution (10 mM MES pH 5.5; no BSA included) was finally injected into the sample chamber through the holes on the glass slide, which then was incubated for 10 minutes before running FRAP experiment.

For FRAP experiment, we used a Total Internal Reflection Fluorescence (TIRF) microscope, which was homebuilt with an inverted microscope (Ti-E, Nikon), high NA objective lens (CFI-apo 100X, NA 1.49, Nikon), 488 nm laser (Coherent) and EMCCD camera (iXon Ultra-888, Andor) (42). We scanned through the cellulose sample to find an area to be imaged and acquired a reference wide-field fluorescence image at a low 488 nm excitation power (2-3 mW) prior to photobleaching. Guided by the pre-bleach image, we decided on multiple locations to be photobleached, and created focused laser spots at the corresponding positions using a spatial light modulator (X13138-01, Hamamatsu). Typically, a maximum of 25 spots on the cellulose fibers along with 2 spots on the background area were chosen. We locally photobleached the selected spots with illumination of a strong 488 nm laser power (50-120 mW) for 1 minute, and then time-lapse wide-field fluorescence images were acquired for 40 minutes to monitor fluorescence recovery. Focus drift was actively compensated with Perfect Focus System (Nikon) while XY-drift was corrected using ImageJ software (NIH).

Custom written MATLAB scripts were employed for the extraction of the recovery curves and data analysis. The photobleached segments were automatically identified by subtracting pre-bleach from post-bleach images. The recovery curves were corrected for non-specific GFP-CBM binding to the glass slide within a segment by setting all pixel values below a segment-specific threshold to zero. The average pixel intensity value of each segment before photobleaching was used to normalize the recovery curve. Additionally, the FRAP signal was baselined to zero relative intensity based on the first FRAP data point to account for the non-zero dark current of the camera and inhomogeneous illumination across the FOV as well as inefficient photobleaching. Each FRAP curve was fitted to the model given in equation below developed by Moran-Mirabal (43) using a Levenberg–Marquardt curve fitting approach built-in MATLAB. The baseline correction is then added to $F_{M,fit}$ to obtain the true F_M value.

$$I(t) = F_M * \left(1 - e^{k_{\text{off}} * t}\right)$$

where

$$F_M = \frac{k^*_{\text{on}} C_F \alpha}{k_{\text{off}}}$$

With $I(t)$ being the normalized intensity value at time t, F_M being the fraction of reversibly bound GFP-CBM3a, k_{off} the desorption rate constant for GFP-CBM3a, k^*_{on} the pseudo-adsorption rate constant assuming a first order binding reaction, C_F the concentration of protein in solution and α the slope of calibration curve between fluorescence intensity and protein concentration. Fits with an $R^2 \leq 0.85$, $F_M \leq 0$ or $F_M \geq 1$ were excluded from further analysis. The collection of fit parameters for cellulose allomorph (F_M, k_{off}) were then fitted to a Gaussian distribution to extract the mean and standard deviation of each parameter. Under the assumption that C_F and α are similar in the case of cellulose I (CI) and cellulose III (CIII), the above equation can be reduced to the following:

$$\frac{F_{M,CI}}{F_{M,CIII}} = \frac{k^*_{\text{on,CI}}}{k^*_{\text{on,CIII}}} \times \frac{k_{\text{off,CIII}}}{k_{\text{off,CI}}}$$

This equation can then be used to infer the ratio of $k^*_{\mathrm{on,CI}}$ and $k^*_{\mathrm{on,CIII}}$ and hence comment on how binding rate constant on one allomorph compared to another (see ref. (43) for a detailed overview of this model equation).

Quartz crystal microbalance with dissipation (QCM-D) based CBM-cellulose binding assay and data analysis: Binding assays were performed using QSense E4 instrument (NanoScience Instruments). Quartz sensors with cellulose thin films were mounted and equilibrated with buffer (10 mM MES pH 6.5; no BSA included) at a flow rate of 100 µl/min for 10 minutes using a peristaltic pump. The flow of buffer was then stopped, and the cellulose films were left to swell overnight in buffer. The frequency and dissipation changes were tracked for all harmonics and the cellulose films were considered amenable to binding studies if the third harmonic stabilized after overnight equilibration in the buffer. Proteins of interest (i.e., GFP-CBM3a) were diluted to a concentration of 1 µM beforehand and flown over the sensors at a flow rate of 100 µl/min for 10 minutes. All proteins tested, attained saturation within 10 minutes as noticed from frequency and dissipation traces. The CBM-cellulose system was left to equilibrate for at least 30 minutes. Unbinding of proteins was tracked by flowing 10 mM MES pH 6.5 buffer at 100 µl/min for at least 30 minutes. The sensors were finally treated with 5% contrad solution followed by deionized water at 100 µl/min for 10 minutes, to remove any traces of protein left in the tubing. The frequency data was converted to mass deposited on sensor using Sauerbrey equation, which was in turn converted to the number of molecules of protein deposited on cellulose surface. The QCM data was then used to obtain pseudo-association rate constant (k^*_{on}) and dissociation rate constant (k_{off}) using the equations provided below:

Binding rate equation:

$$[EC] = A(1 - e^{-(k^*_{\mathrm{on}})t})$$

Unbinding rate equation:

$$[EC] = Ae^{-(k_{\mathrm{off}})t}$$

Where $[EC]$ = Number of molecules; t = Time (minutes)

Buffon needle model analysis of CBM orientation probability distribution on cellulose surface: Buffon's needle model arose from a problem first posed by the eponymous French mathematician Buffon (44, 45). The original model is an analytical solution to the question of the probability of a short needle crossing the parallel lines on a ruled paper or wooden floor. An analytical problem to this solution does exist and the numerical solution to this problem would involve the Monte Carlo method. Here, we have adopted the Monte Carlo approach to simulate the CBM-cellulose system with some simplifications which render the binding process completely from a simply geometric probability point-of-view and without any energetic constraints taken into consideration. The CBM was assumed to be a needle with length of 2.08 nm representing the distance between centers of flanking planar Tyrosine rings (Y5 and Y31) based on the published NMR structure (PDB ID: 1cbh) (see **Fig. S16A**). Cellulose-I binding surface was assumed to be an array of parallel lines, each line representing a cellulose chain, with an inter-chain distance of 0.8 nm (see **Fig. S16B**). In the energetically favorable configuration of this system as predicted by MD simulations, CBM1 aligns with two aromatic residues (Y31, Y32) on one cellulose chain and another aromatic residue (Y5) on an adjoining cellulose chain. Hence, any event where the CBM needle lands on a chain perfectly (C_0), crosses only one chain (C_1) were clustered into a category called events 'along the chain'. Similarly, events where the CBM needle crosses two chains (C_2) and three chains (C_3) were clustered into a category called events 'across the chain'. The configuration of this system at any time can be defined by the angle Θ between the cellulose chains and the CBM needle. A Monte Carlo simulation was then conducted with 100,000 trials whereby a random number generator was used to sample the angle Θ (see **Fig. S16C-D**). Configuration from a given trial can then be classified into one of the five categories mentioned above ($C_0 - C_3$) and eventually, the probability of CBM crossing two chains or more. See **Fig. S16** for additional methodology related details. The fraction of events along and across the chain are computed based on the following formulae.

$$Fraction\ of\ events\ along\ the\ chain = \frac{C_0 + C_1}{C_0 + C_1 + C_2 + C_3}$$

$$Fraction\ of\ events\ across\ the\ chain = \frac{C_2 + C_3}{C_0 + C_1 + C_2 + C_3}$$

SI Appendix Results and Discussion:

AFM imaging of Cladophora cellulose allomorphs: Previous studies on cellulase or CBM binding have often employed two-site binding models (46), with the assumption of two distinct classes of binding sites arising from any of the following phenomena: (i) binding to amorphous/crystalline regions or (ii) binding to two different faces of crystalline cellulose fiber, and/or (iii) due to different orientations of CBM on hydrophobic face of cellulose. Since Cladophora derived cellulose is a highly crystalline substrate and Type-A CBMs bind to crystalline regions predominantly, a small number of high affinity amorphous binding sites seems unlikely. However, the availability of different crystal faces for binding to non-native allomorph cellulose III needed to be examined and hence we performed AFM imaging on both Cladophora cellulose I and cellulose III microfibrils to verify the general morphological differences between two substrates (see **Fig. S1**). AFM imaging indeed suggests that the Cladophora cellulose I elementary fibrils have two distinct crystalline planar surfaces likely available for CBM binding unlike cellulose III, consistent with predicted crystal structure of algal cellulose and morphological changes undertaken following ammonia treatment (37, 40, 47). This change in cellulose I fibril shape is expected to take place during ammonia pretreatment due to a solid-state polymorphic transformation that alters the cellulose III unit cell crystal structure (37, 48–50).

Truncation and sensitivity analysis for CBM1 binding data set: We varied the free protein concentrations over nearly four orders of magnitude (0.02-250 μM for CBM1 and 0.02-50 μM for CBM3a) to accurately capture binding interactions at both low (<10-fold of min K_d) and high (>10-fold of max K_d) protein-to-substrate saturation concentrations. This broad range of protein concentrations is often recommended to properly characterize CBM adsorption to polysaccharides (51), but is not always followed by researchers, which could explain some of the discrepancy in the literature for reported n_{max}/K_d values for various CBMs. Langmuir two-site models are often interpreted as consisting of two different classes of binding sites (46), one with high affinity and another with low affinity (a higher value of K_d corresponds to a lower affinity). Langmuir-Freundlich model, on the other hand, also provides information regarding the binding cooperativity for a given substrate (52). However, since most studies do not validate all model assumptions, it is very challenging to draw any molecular-level understanding of protein adsorption process using such traditional approaches alone. Furthermore, since the root-mean square errors from the model fits to the experimental data are nearly comparable in most cases, it is difficult to choose an appropriate model based on goodness of model fit without any observer bias. Details regarding sensitivity of results to truncation of original CBM1 data set is discussed already in the main manuscript (see **Table 2**). Furthermore, we also performed sensitivity analysis for the original CBM1 to cellulose I binding data set by studying the variation in RMSE when the binding parameters (n_{max}, K_d) are varied individually by 10% and 20% each (see **Table S2**). This sensitivity analysis revealed that the n_{max} parameter is quite sensitive to error in data acquired at higher protein concentrations, emphasizing the need for enough replicates while probing higher protein concentrations.

MD simulations reveal improper stacking interactions of aromatic residues with cellulose III: MD simulations were conducted to identify the preferred conformations of CBM1 on each cellulose allomorph surface and estimate the theoretical binding affinity of CBM1 for the most preferred orientation. Two different types of crystalline cellulose III models were used for the simulations, one with lower crystallinity and another with high crystallinity (based on the available crystalline core cellulose crystal structure). A recent study (3) has shown that CBM3a binding to cellulose-III closely depends on the cellulose III fibril crystallinity. Therefore, we were interested in modeling the binding of CBM1 to two types of cellulose III crystals with varying degree of crystallinity for this study. However, since no detailed structural information is available for lower crystallinity cellulose III, a modeled structure was obtained by increasing the simulation temperature to 400 K starting with the high crystallinity cellulose III structure to increase overall disorder within the crystalline fiber. Previous studies employing MD simulations (39) and CBM-cellulose EM-immunolabeling experiments (53) have already shown that the preferred face for binding of CBM1 to cellulose I is the hydrophobic face. Therefore, the hydrophobic faces of cellulose I and cellulose III were initially chosen to study the binding preferences of CBM1 in this study as well.

As shown in **Fig. S5A**, the unbiased MD simulations revealed that the most populated configurations for CBM1 on cellulose I have the CBM1 Y31 residue facing towards the non-reducing end of the cellulose chain (as also reported previously (38)). This chain-end preference could have evolved from the reducing end specificity of the Cel7A catalytic domain to which the native CBM1 is attached at the C-terminus. Furthermore, it was also observed that CBM1 remained stably bound to cellulose I surface regardless of whether the Y31 residue was facing the reducing or non-reducing end. Interestingly, CBM1 was stably bound to the cellulose III surface (for both low and high crystallinity forms of cellulose III) only when the Y31 residue was facing the reducing end of cellulose. See **Movies S1-S3** for representative visualization of CBM1-cellulose allomorphs interactions from unbiased MD simulations. The relative diffusivity of CBM1 and the RMSF for key aromatic residues is provided here as well (see **Fig. S5B-C**). The improper stacking of Y5 residue on cellulose III specifically is highlighted in **Fig. S5D**.

CBM3a exhibits increased desorption rate constant towards cellulose III: In addition to equilibrium binding assay methods, we used FRAP to study GFP-CBM3a binding to Cladophora derived cellulose allomorphs (with native cellulose enriched in Iα) and QCM-D to study GFP-CBM3a binding to plant-derived Avicel microcrystalline cellulose allomorphs (with native cellulose enriched in Iβ) (54) to obtain a more comprehensive picture of CBM3a binding kinetics towards various cellulose allomorphs.

Procedures for preparation of cellulose nanocrystals, nanocellulose thin film deposition, QCM-D binding assay, and data analysis procedures are outlined in the SI appendix materials and methods section. Sauerbrey equation was used to obtain the mass of adsorbed protein on cellulose film using the frequency change at third overtone (55). The number of protein molecules was then estimated from the mass of adsorbed protein and this resulted in QCM sensorgrams as shown in **Fig. S10A-B.** This data was then analyzed using exponential fitting routines (see exponential fits to acquired QCM raw data in **Fig. S10C-D**), to obtain a pseudo association rate constant (k_{on}^{*}) and dissociation rate constant (k_{off}). While k_{on}^{*} did not show a significant difference between the two allomorphs, CBM3a gave a nearly 3-fold increase in $k_{\mathrm{off}}$for cellulose III (**Table 4**). In addition, the maximum amount of protein bound to cellulose III was 1.5-fold less. These observations align well with the reduction in binding of CBM3a to cellulose III observed using the classical pull-down binding assay method. A similar reduction in CBM binding was observed to another non-native allomorph (i.e., cellulose-II) using QCM-D (56).

Next, we performed FRAP experiments, with details regarding the assay method and data analysis outlined in the SI appendix experimental procedures section. Briefly, similar to previous work (43, 46), GFP-CBM3a binding kinetic parameters to cellulose allomorphs were obtained by fitting the FRAP curves to a binding-dominated model ignoring any diffusion relevant contributions (see supplementary information **SI Appendix Fig. S9** for representative FRAP images and recovery traces). Our FRAP analysis revealed that CBM3a gave a 1.9-fold increase in k_{off} for cellulose III compared to cellulose I (**Figure 6**). F_M and k_{off} were together used to extrapolate the ratio of adsorption rate constants for cellulose I and cellulose III (see **Experimental Procedures** for model equations) which was calculated to be 0.64 ± 0.29. Hence, the difference in adsorption rate constants between the two allomorphs could only be marginal whereas the desorption rate constant alone was deduced to be significantly lower for cellulose III. In summary, solid state depletion equilibrium binding assays as well as both QCM and FRAP binding assays together provide complementary information on the differences in binding affinity and kinetic rate constants for CBM binding to cellulose allomorphs. However, neither of these methods alone provide a detailed molecular-level understanding of Type-A CBM binding, particularly for weaker binders like CBM1 that impact cellulase binding and activity towards distinct allomorphs.

Reduced partition coefficients of Type-A CBMs towards cellulose III: Since it was challenging to pick a suitable Langmuir model to identify a clear molecular-basis for reduced CBM1 or CBM3a binding to cellulose III, we also estimated the partition coefficients (i.e., $n_{\mathrm{max}}/K_{\mathrm{d}}$) for a larger library of several Type-A CBMs for both cellulose allomorphs at room temperature under non-saturating protein loadings (raw data provided in SI Appendix **Fig. S7** and **Fig. S8** for bar graph). The relative binding order of CBM families based on estimated partition coefficient

value was roughly similar between both allomorphs, where Family 1~Family 2a << Family 64 < Family 3a < Family 5~Family 10. However, the partition coefficients for all Type-A CBMs tested were always significantly lower for cellulose III. The decrease in partition coefficients ranged from 2-fold to 13-fold depending on the CBM family, with CBM64 and CBM1 showing the largest (~13.3-fold) and smallest (~2.3-fold) fold change between the two allomorphs, respectively. Note that the GFP domain alone had insignificant binding affinity towards either cellulose allomorph (~0.1 L/g partition coefficient), suggesting that the CBM domains alone were largely responsible for binding of GFP-CBM fusion proteins to either cellulose allomorph.

***Reduced calcofluor dye adsorption to cellulose allomorphs**:* Calcofluor is a trans-stilbene based fluorescent dye that has a low molecular weight (917 g/mol) and a core structure analogous to the planar interface of Type-A CBMs involving both aromatic and polar functional groups that provides the thermodynamic driving force for binding to complementary cellulose surfaces. Calcofluor is thus an ideal inorganic CBM-like surrogate probe useful for understanding the multivalent binding interactions. Interestingly, Calcofluor dye was also found to show reduced partition coefficient towards microcrystalline cellulose III by 2.6-fold versus native cellulose I (see **Fig. S15A**). Calcofluor is expected to bind cellulose with high affinity, based on reports on how it impacts *in vivo* cellulose synthesis (57) and histological staining analysis of cell wall polysaccharides (58). However, to the best of our knowledge, a direct measurement of calcofluor affinity to microcrystalline cellulose has not been reported. Based on our two site Langmuir model fitting analysis, we found that calcofluor white dye binds to microcrystalline cellulose I with a lower K_{d1} of 38.8 μM to 81% of total available sites and K_{d2} of 1.2 μM to the remaining 19% of total available sites (n_{max} total was 21.5 μmol/g cellulose). For cellulose III, we found that calcofluor white dye binds with a higher K_{d1} of 79.3 μM to 70% of total available sites and K_{d2} of 1.8 μM to the remaining 30% of total available sites (n_{max} total was 9.2 μmol/g cellulose). Based on single site model fitting analysis, we found that calcofluor white dye binds to microcrystalline cellulose I with a K_d of 18±3 μM and n_{max} of 19.9±1.0 μmol/g cellulose, while for cellulose III a K_d of 15±4.5 μM and n_{max} of 7.7±0.5 μmol/g was estimated. However, the two-site model gave a much better fit for the data unlike the one-site model. Three-site model gave no further improvement to model fitness for either cellulose I or III. While, these values are likely dependent on the buffer ionic strength conditions, calcofluor has lower affinity and binding sites available for adsorption to microcrystalline cellulose III versus native cellulose I. It is known that strong non-covalent interaction forces stabilize interaction of calcofluor along the repeating ~1 nm cellobiosyl-unit of cellulose along the chain axis (58). This could explain why calcofluor dye binds in a well-defined orientation parallel along the fiber axis to both chitin and cellulose (59). However, altering the crystal structure of native cellulose I to cellulose III could destabilize calcofluor binding due to steric clashes with adjacent cellulose chains due to changed crystal morphology and/or increased fibril hydrophilicity. These findings shed further light into the thermodynamic mechanism driving reduced binding interactions of critical planar CBM binding surface associated amino acid residues to cellulose III. Interestingly, as reported previously for CBM1 binding to crystalline cellulose (52), Scatchard plot analysis for calcofluor binding to cellulose was also non-linear and concave-upward (see SI appendix **Fig. S15B**). Calcofluor dye binding to multiple classes of non-equivalent binding sites could provide a classical interpretation of concave-upward Scatchard plots (60). Nevertheless, when characterizing bulk ensemble binding interactions to highly heterogenous crystalline cellulose microfibril surfaces, even for simple CBM-analogues like calcofluor, our work brings to the light the severe challenges associated with choice of multi-site adsorption models and possible misinterpretation of results.

Buffon needle model for predicting CBM1 orientations: The model formulation and Monte Carlo simulation procedure used here are described in detail in **SI Appendix Experimental Procedures** section. Our simple Buffon needle model predicted that the geometric probability of a CBM1 wild type 'needle' to bind along a single cellulose chain is ~42%, while the remaining ~58% of events would include binding across multiple cellulose chains (ignoring any energetic barriers to binding orientation) (see **Table S3**). Interestingly, if a mutation on CBM1 (Y31A for instance) is considered as having reduced needle length, that would increase the percentage of events along the chain to 90%. Hence, performing these planar aromatic residue mutations and testing the impact of these mutations on the heterogeneity of CBM binding to cellulose studied using bond rupture assays could give us some insight into

the role played by CBM binding orientations on observed heterogeneity. Preliminary data from our CBM-cellulose bond rupture assay indeed shows that there are some distinct differences in the force-lifetime plots for CBM1 wild-type and Y31A mutant on cellulose I (e.g., see 0-2.5 pN and 10-15 pN ranges in **Fig. S14**). However, it needs to be emphasized that this is an over-simplistic model and an understanding of the energetic constraints could better simulate the complex reality of CBM-cellulose interactions. Moreover, we currently lack the ability to theoretically relate this simple model's predictions back directly to the bond rupture assay results. Most previous CBM binding focused studies (38, 39, 61) have not emphasized the possible orientations of CBM1 on the surface of cellulose I. Beckham *et al.* (38) previously showed that although CBM1 prefers to bind along the cellulose chain as well, slightly rotated (by ~10-15°) CBM1 orientations across multiple cellulose chains are energetically feasible as well on individual fiber surfaces. More detailed MD simulations and corresponding rupture assays need to be conducted to check how mutations of CBMs impact along the chain axis versus across multiple chain axis binding and its potential impact on non-productive processive cellulase binding.

SI Appendix Figures & Tables

Fig. S1. AFM imaging of Cladophora-derived cellulose I or CI (A, D) and cellulose III or CIII (B, E) elementary fibrils, the scale bar is 500nm. The black line in panel A and B corresponds to the height signal (forward and backward) shown in panel D and E. As is can be seen in panel D and E, there is no difference in the forward and backward signals, hence only the forward signal was used to analyze the height profile dimensions which are depicted in panel C. Cellulose I fiber showed a peak with a clear shoulder while no such shoulder was seen for the cellulose III fibril, which is consistent with the modification in fiber shape after ammonia pretreatment.

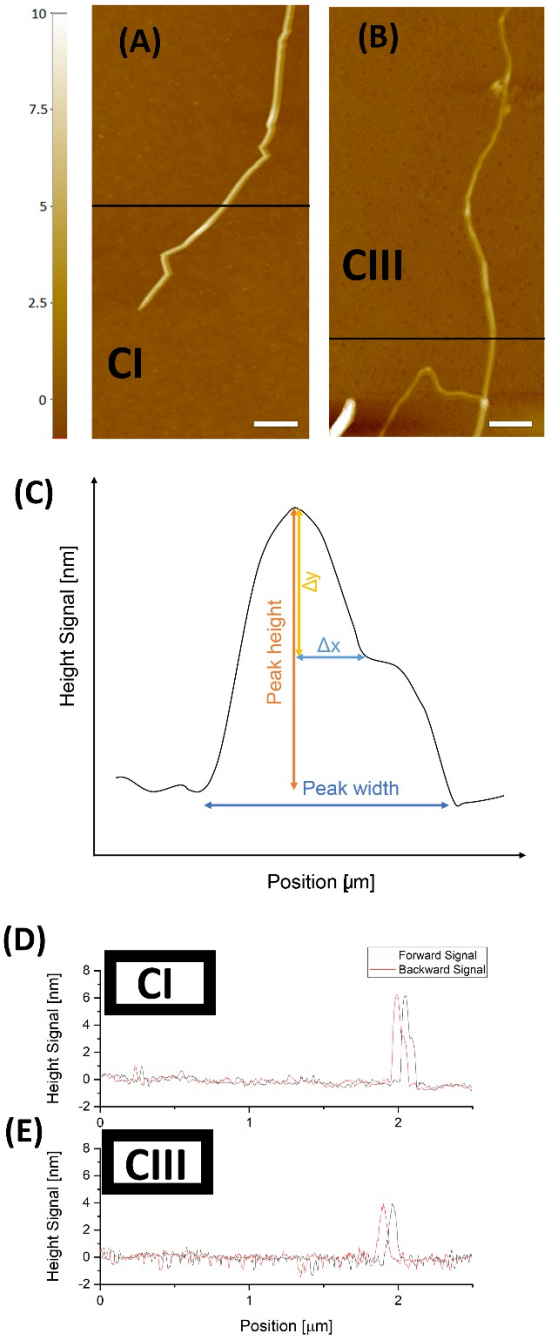


Fig. S2. Single-molecule optical tweezers-based verification of binding stability and instability for full length Cel7A cellulases processively hydrolyzing Cladophora based cellulose I and III. Three representative traces each are shown for the conditions of stability on cellulose I (red), stability on cellulose III (blue) and instability on cellulose III (black). Cel7A was found to be more often stably bound to cellulose I than to cellulose III before initiating stable processive motility accordingly. Of the traces showing instability, Cel7A was likely to have more rupture events during initiation of the processive catalytic cycle on cellulose III than on cellulose I.

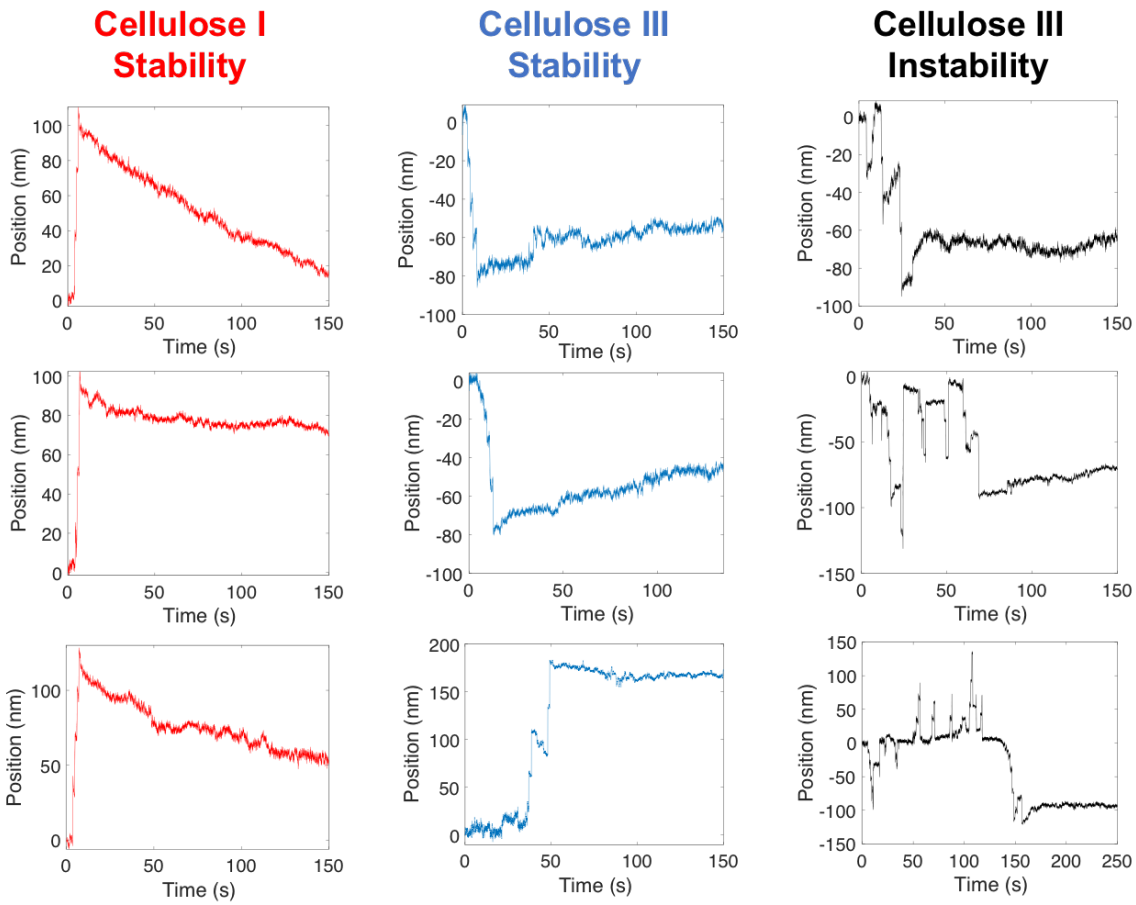


Fig. S3. Representative motility traces for Cel7A on Cladophora cellulose I. Dashed lines indicate average Cel7A velocities during processive hydrolysis of cellulose as 0.25 ± 0.35 nm s^{-1} (s.d.; for cellulose I; N=68; in red) and 0.17 ± 0.14 nm s^{-1} (s.d.; for cellulose III; N=30; in black).

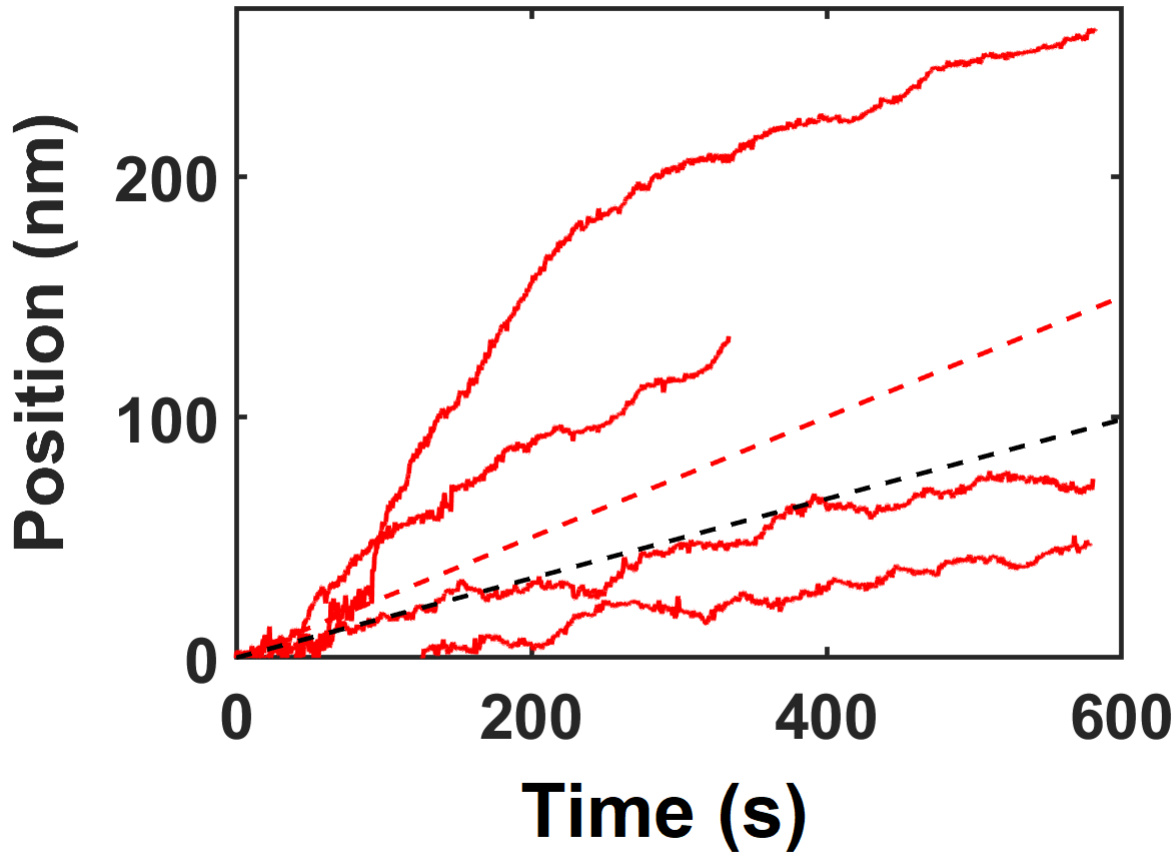


Fig. S4. Langmuir-type adsorption model fits (in red) for CBM1 binding data (in black) to Cladophora-derived Cellulose I (A, C, E) and Cellulose III (B, D, F).

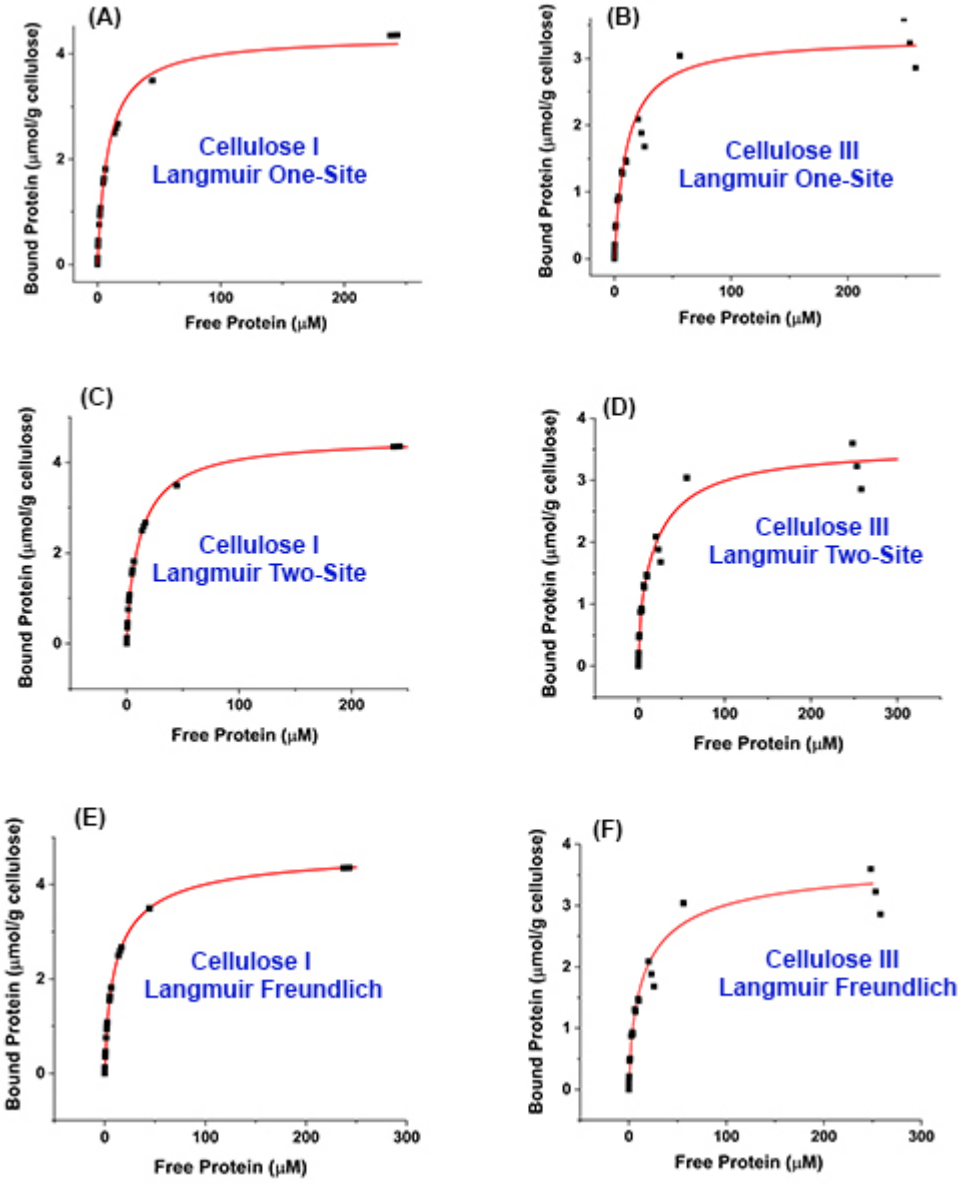
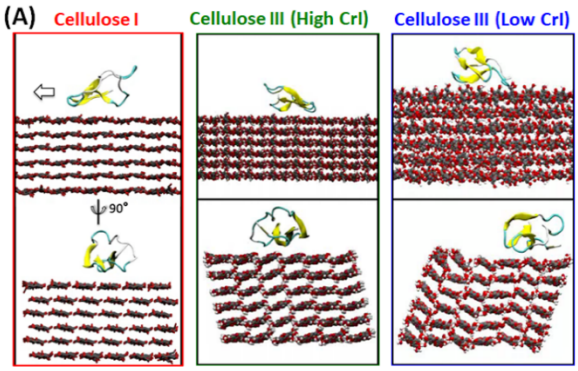


Fig. S5. (A) Side and cross-sectional snapshot view of CBM1-cellulose unbiased MD simulations conducted for distinct cellulose ultrastructures; cellulose I and cellulose III (with varying degrees of surface chain order or crystallinity index-CrI). Arrow direction indicates the preferred canonical binding orientation of CBM1 for cellulose I. (B) Average diffusivity of CBM1 on cellulose allomorph surfaces was estimated from the trajectories of the unbiased MD simulations to show significantly higher values for cellulose III versus cellulose I, again suggestive of weaker CBM1 binding interactions with the former substrate. (C) Steric clashes of planar CBM1 binding surface aromatic residues with cellulose III surface provides an atomistic basis for reduced binding towards cellulose-III. Root mean square fluctuations (RMSF) values seen for CBM1 planar binding surface tyrosines (Y5, Y31, and Y32) was significantly higher for cellulose III of decreasing crystallinity. (D) Representative image from MD simulations highlighting the improper stacking of Y5 residue on cellulose III is highlighted here.



(B)

CBM1 Diffusivity (cm^2/s)	Cellulose I	Cellulose III (High CrI)	Cellulose III (High CrI)
ρ	1.4×10^{-11}	6×10^{-9}	2.3×10^{-5}

(C) Root mean square fluctuation (RMSF) of key residues

CBM1 Residue #	Cellulose I	Cellulose III (Low CrI)	Cellulose III (High CrI)
Y5	0.44	0.65	0.9
Y31	0.47	0.54	0.9
Y32	0.35	0.4	0.6

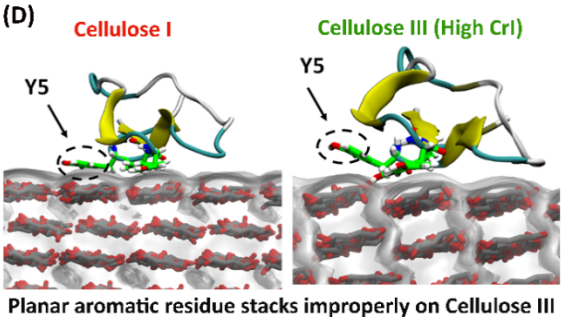


Fig. S6. Langmuir-type adsorption model fits (in red) for CBM3a binding data (in black) to Cladophora-derived Cellulose I (A, C, E) and Cellulose III (B, D, F).

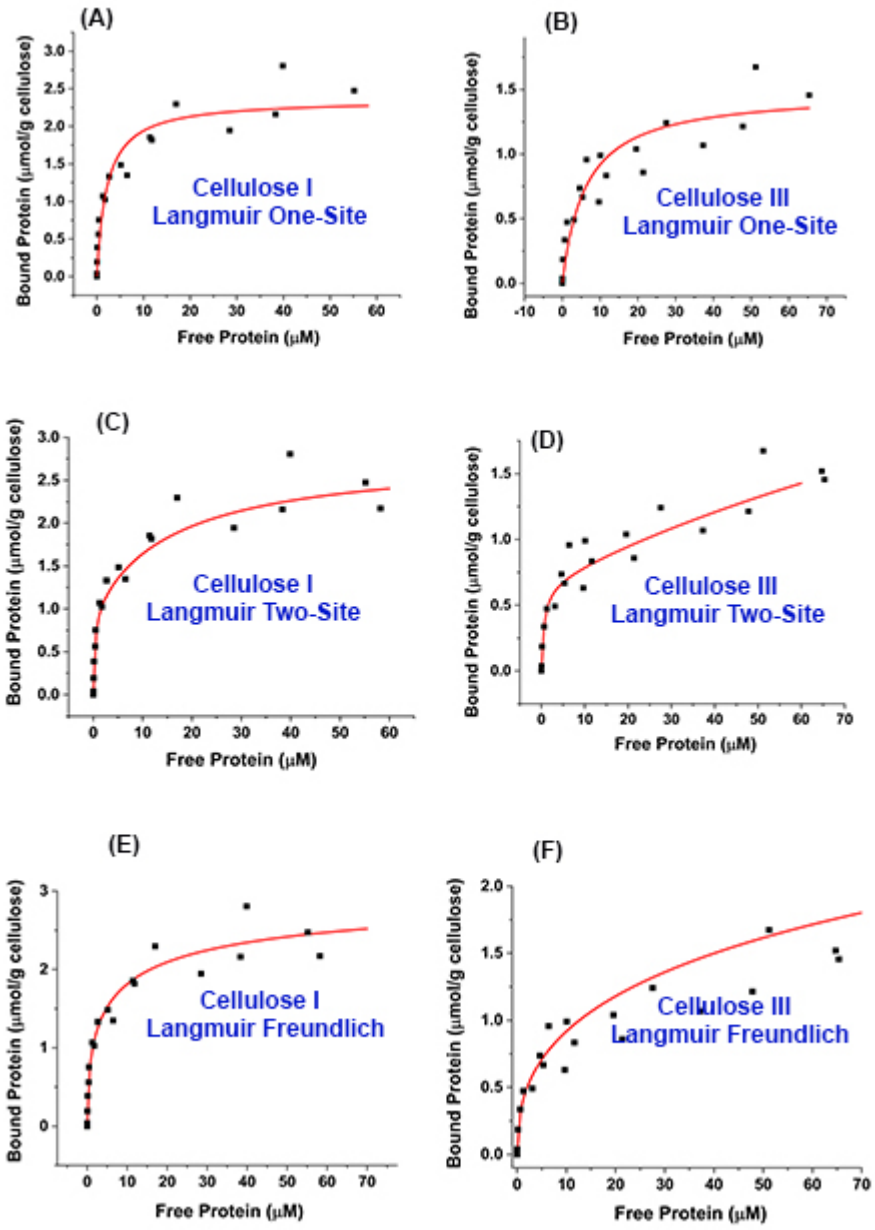


Fig. S7. Protein adsorption data and fitted partition coefficient slopes for various Type-A CBMs to Cladophora-derived Cellulose I (in black) and Cellulose III (in red).

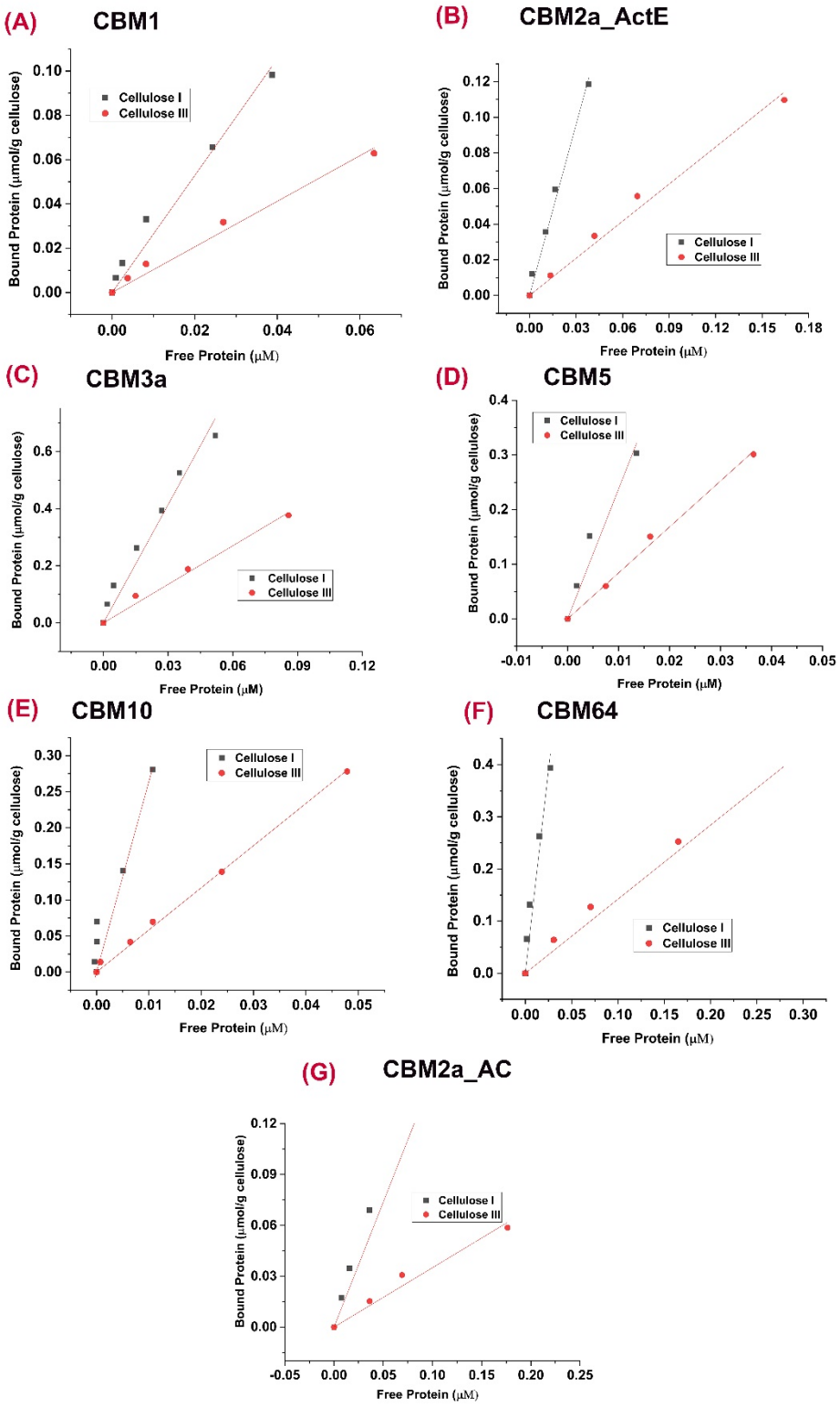


Fig. S8. Comparison of partition coefficients (liters/gram) estimated for various Type-A CBMs towards Cladophora based cellulose I (in red) and cellulose III (in black) are shown here. Error bars are standard deviations for reported mean values. The partition coefficient bars for GFP only control protein are not clearly visible since it exhibits very low binding to either cellulose allomorph.

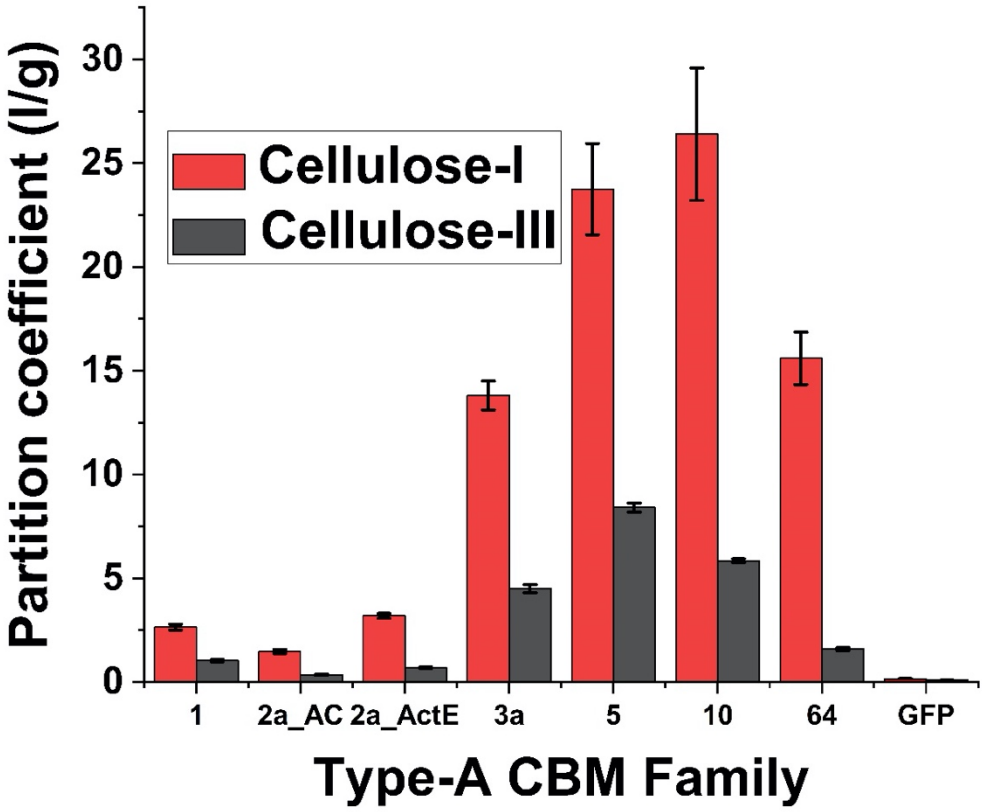


Fig. S9. Characterization of GFP-CBM3a binding kinetics to Cladophora cellulose I (CI) and cellulose III (CIII) using FRAP. Panel A-C show representative snapshots of the FRAP image acquisition with panel A as the image before photobleaching, B first frame after local photobleaching, and panel C at the end of the image acquisition, the scale bar is 10µm. Panel D and E show example recovery curves for cellulose I and III, respectively. Note that the shown recovery curves are baselined adjusted to zero as mentioned in the experimental procedures section.

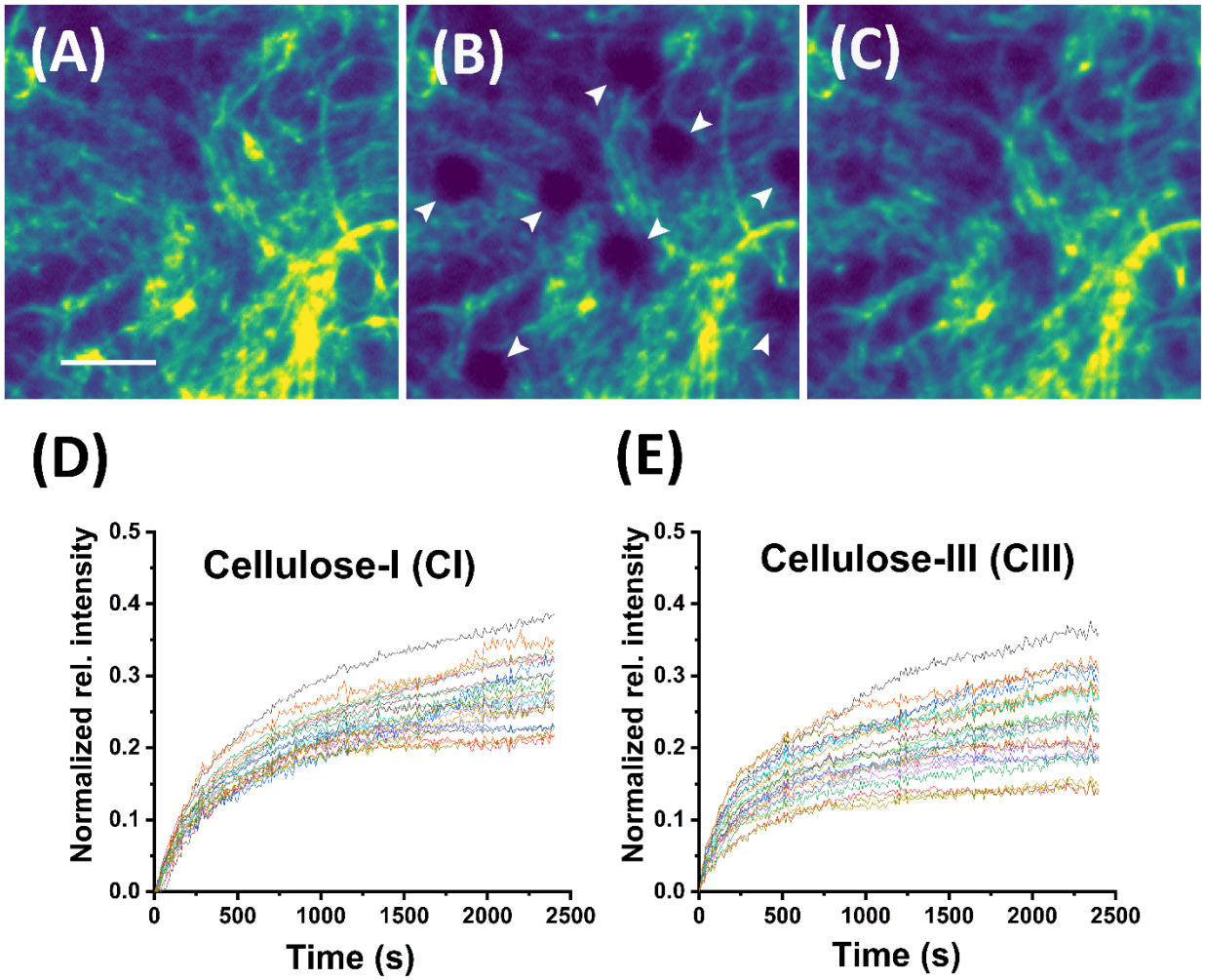


Fig. S10. Quartz crystal microbalance (QCM) based analysis of GFP-CBM3a binding parameters towards Avicel based cellulose I (CI) and cellulose III (CIII) nanofibrils. Representative QCM traces for binding/unbinding dynamics of GFP-CBM3a towards cellulose I (A) and cellulose III (B) are shown below. Representative binding model fits (in green) to raw QCM data (in black) is shown for binding (C) and unbinding (D) regimes for the sensorgram reported in (A).

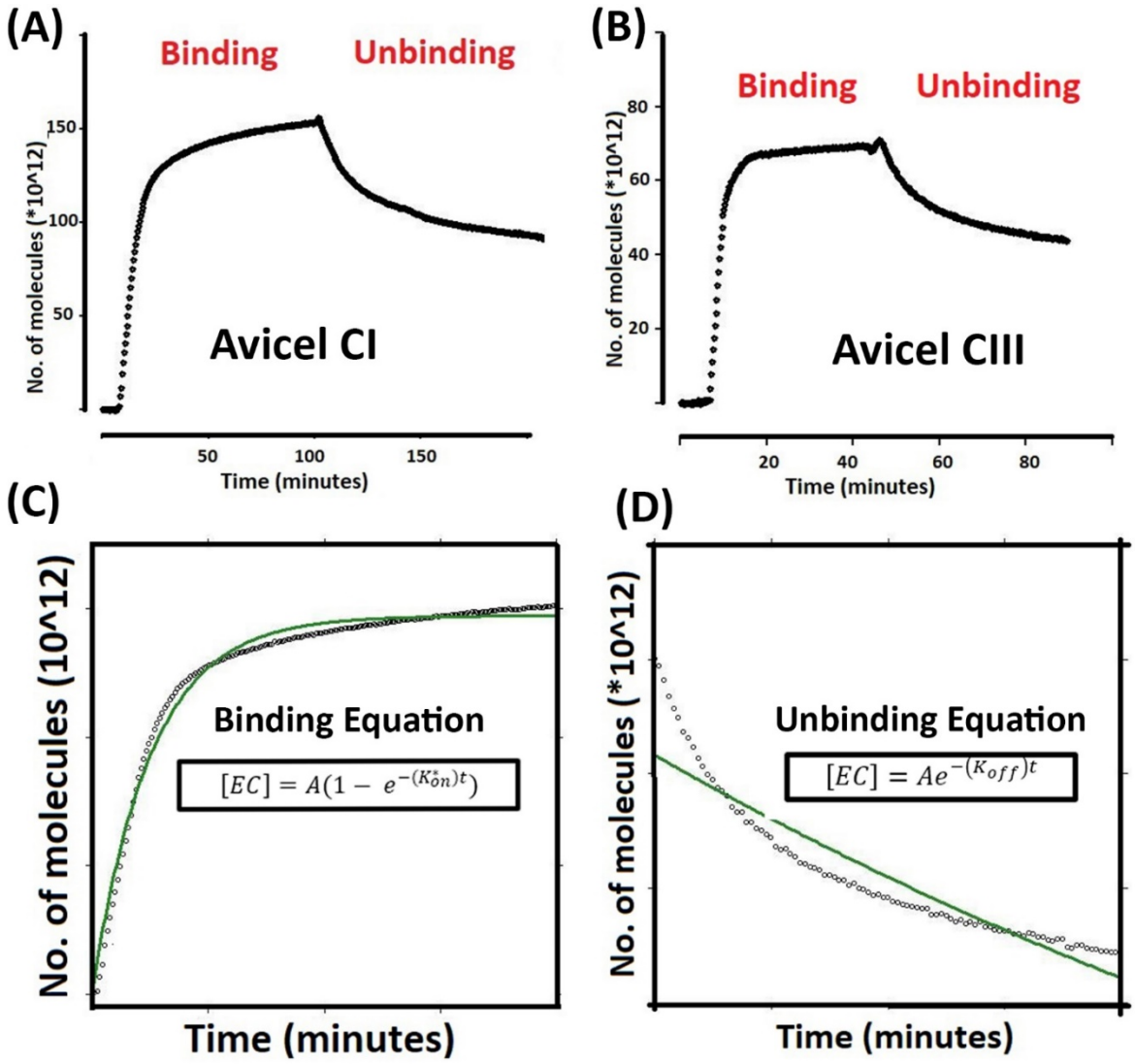
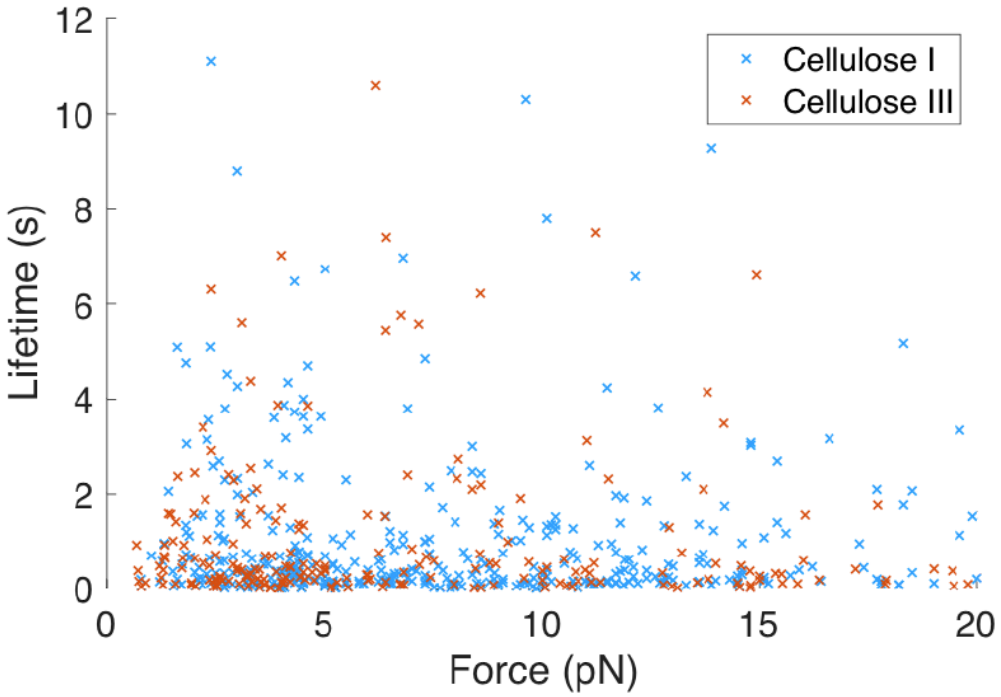


Fig. S11. (A) Force vs lifetime raw data scatterplot for the CBM1 non-covalent bonds to Cladophora celluloses I (blue) and III (brown) are shown here. Total number of individual rupture events measured (N) on cellulose I is 410 and on cellulose III is 214. For visual clarity, we omitted data points above 20 pN or 12 s from the scatterplot (31 for cellulose I; and 5 for cellulose III). But we did not exclude any data from our report or analysis. Our one-way ANOVA test (B) concluded that there was no significant difference (p=0.20) between the two entire datasets or at 2.5 pN intervals. Such a wide variance in lifetimes further supports our claim of multiple binding regimes.

(A)



(B)

Force Range	Mean Lifetime Cellulose I (s)	Mean Lifetime Cellulose III (s)	ANOVA Result
All Forces	1.46 ± 0.22	1.01 ± 0.07	F(1,536) = 1.67, p = 0.20
0-2.5 pN	1.91 ± 0.60	1.00 ± 0.20	F(1,78) = 1.91, p = 0.17
2.5-5 pN	1.76 ± 0.59	1.07 ± 0.23	F(1,178) = 0.84, p = 0.36
5-7.5 pN	1.76 ± 0.70	1.59 ± 0.50	F(1,60) = 0.04, p = 0.85
7.5-10 pN	0.79 ± 0.21	1.18 ± 0.33	F(1,71) = 0.96, p = 0.33
10-12.5 pN	1.49 ± 0.60	1.02 ± 0.48	F(1,70) = 0.16, p = 0.69
12.5-15 pN	1.06 ± 0.31	1.01 ± 0.37	F(1,52) = 0.02, p = 0.93
15-17.5 pN	1.72 ± 0.84	0.45 ± 0.17	F(1,21) = 1.17, p =0.29
17.5-20 pN	1.23 ± 0.38	0.43 ± 0.23	F(1,20) = 1.82, p =0.19

Fig. S12. Force vs lifetime relationships for the CBM1 non-covalent bonds to Cladophora derived cellulose I (blue) and filter paper (green) derived cellulose microfibrils. Both force-lifetime distributions failed to converge to the classical slip bond model and revealed that the CBM1-cellulose interaction is multimodal across different native cellulosic substrates. Interestingly, the reported mean lifetime of the CBM1-filter paper cellulose bond (3.03 ± 0.37 SEM) is higher than that of the CBM1-cladophora cellulose bond (1.41 ± 0.20 SEM).

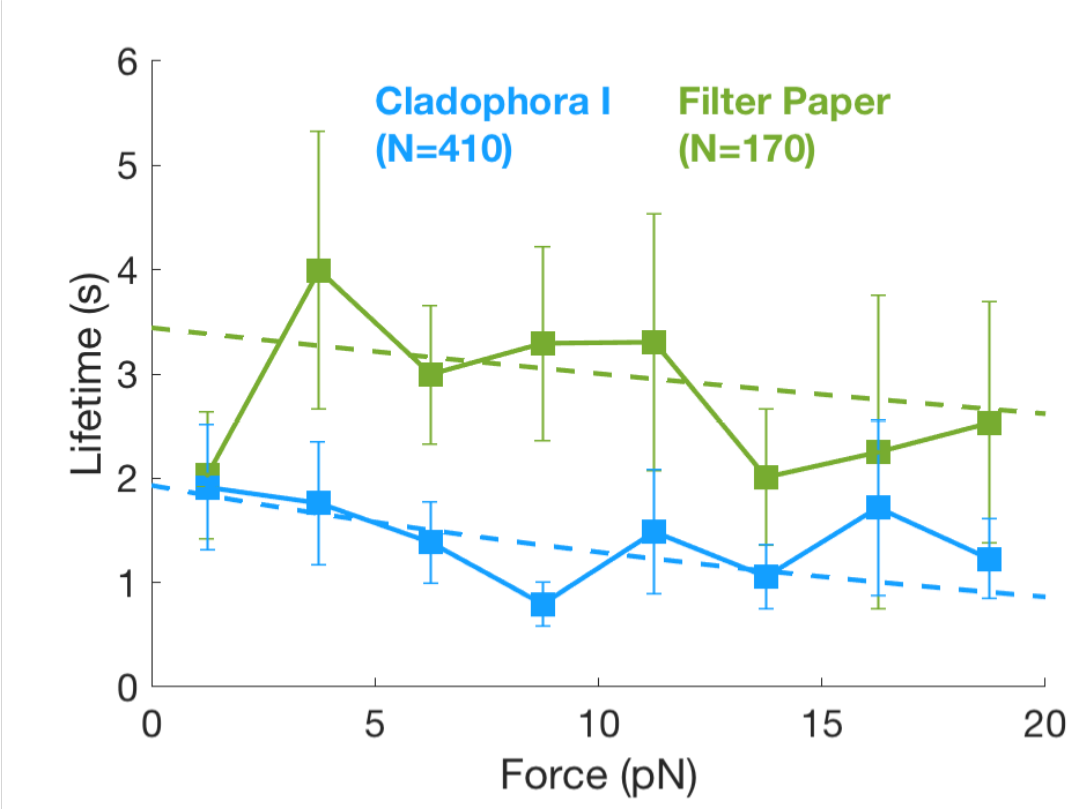
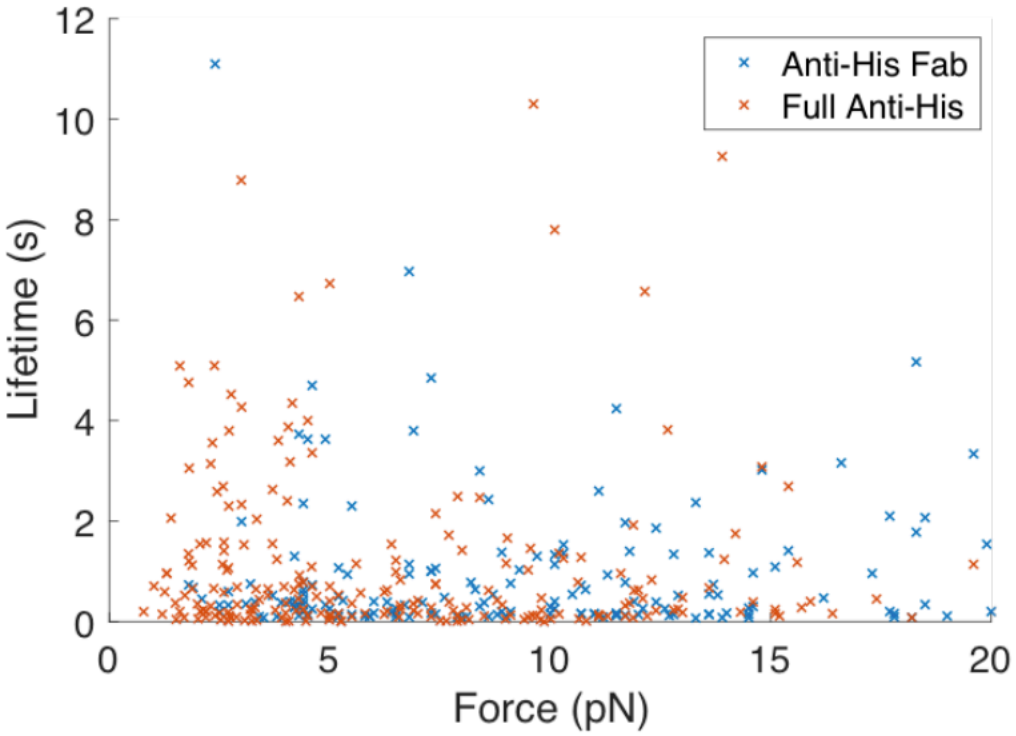


Fig. S13. (A) Force vs lifetime raw data scatterplot for the CBM1 non-covalent bond on *Cladophora* cellulose I using an anti-His Fab in the assay construct (blue) and using full anti-His antibody (brown). Total number of events measured (N) using the full antibody is 233 and using the Fab is 187. For visual clarity, we omitted data points above 20 pN or 12 s from the scatterplot (8 for full antibody; 23 for Fab). We did not exclude any data from our report or analysis. Our one-way ANOVA test (B) concluded that there was no significant difference between the two entire datasets or at 5 pN intervals. Because of the statistical similarity, we combined both datasets to represent our CBM1-cellulose I data.

(A)

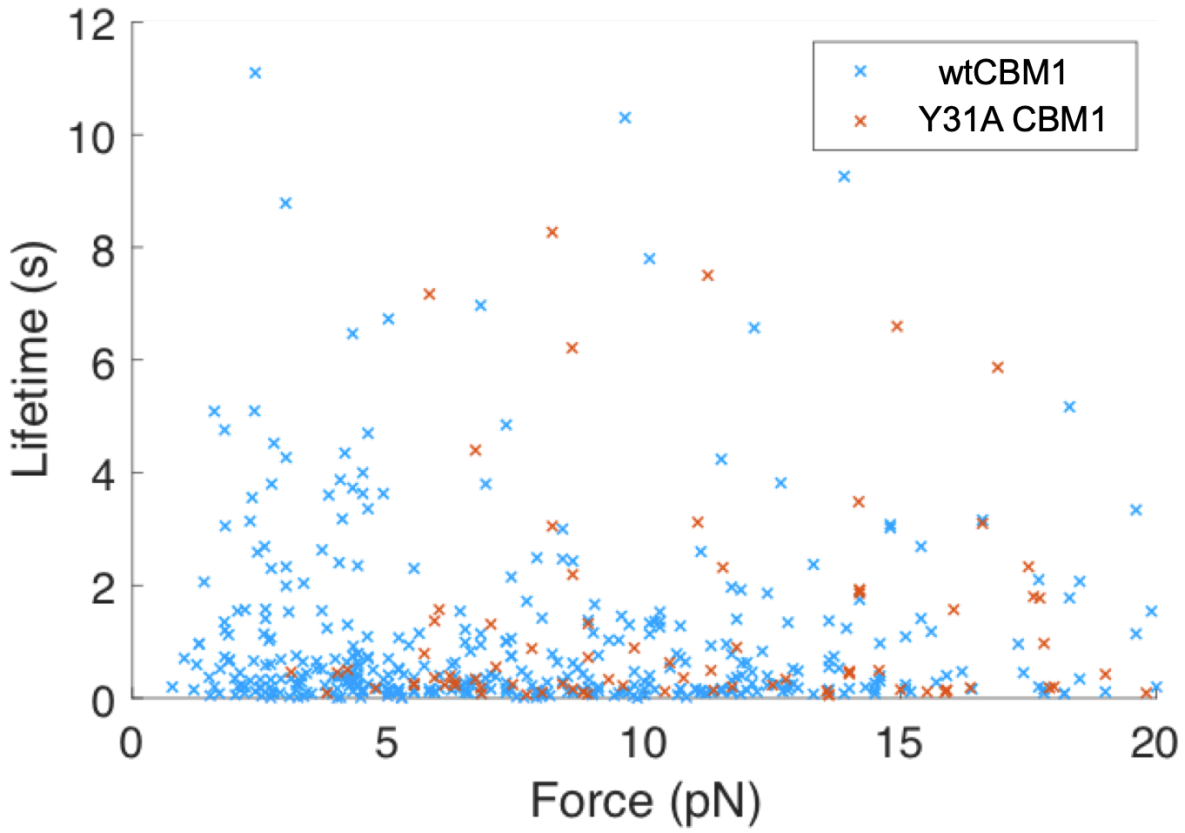


(B)

Force Range	Mean Lifetime Anti-His Fab (s)	Mean Lifetime Full Anti-His (s)	ANOVA Result
All Forces	1.78 ± 0.37	1.04 ± 0.18	$F(1,370) = 2.74$, $p = 0.10$
0-5 pN	1.91 ± 0.60	1.55 ± 0.57	$F(1,149) = 0.13$, $p = 0.72$
5-10 pN	1.29 ± 0.41	0.83 ± 0.18	$F(1,104) = 1.10$, $p = 0.30$
10-15 pN	2.19 ± 0.90	0.75 ± 0.12	$F(1,86) = 3.00$, $p = 0.09$
15-20 pN	1.66 ± 1.06	1.28 ± 0.40	$F(1,25) = 0.13$, $p = 0.72$

Fig. S14. (A) Force vs lifetime raw data scatterplot for the CBM1 non-covalent bond on *Cladophora* cellulose I using the wild-type CBM1 protein (blue) and using the Y31A CBM1 mutant (brown). Total number of events measured (N) using CBM1 is 410 and using the mutant is 93. For visual clarity, we omitted data points above 20 pN or 12 s from the scatterplot (31 for CBM1; 11 for Y31A-CBM1). We did not exclude any data from our report or analysis. Our one-way ANOVA test (B) concluded that there was no significant difference between the two entire datasets or at the 0-5 pN, 5-10 pN, and 15-20 pN ranges. However, there was a significant difference observed at the 10-15 pN range indicating that structural changes on CBM does indeed affect the CBM1-cellulose interactions measured using our single-molecule rupture assay method. Future single molecule studies could explore the effects of other protein structural changes on binding to cellulose.

(A)



(B)

Force Range	Mean Lifetime wtCBM1 (s)	Mean Lifetime Y31A CBM (s)	ANOVA Result
All Forces	1.46 ± 0.22	2.25 ± 0.48	F(1,455) = 2.36, p = 0.13
0-5 pN	1.80 ± 0.46	0.33 ± 0.08	F(1,154) = 0.34, p=0.56
5-10 pN	1.06 ± 0.22	1.94 ± 0.71	F(1,139) = 2.44, p = 0.12
10-15 pN	1.34 ± 0.40	3.55 ± 1.08	F(1,115) = 5.66, p = 0.02
15-20 pN	1.45 ± 0.51	1.19 ± 0.39	F(1,41) = 0.13, p = 0.71

Fig. S15. Langmuir-type adsorption data (symbols) or model (dotted lines) fits (A) and Scatchard-plot representation of adsorption data (B) for Calcofluor White dye binding data to Cladophora-derived Cellulose I (in red) and Cellulose III (in grey). Here, the dotted lines represent Langmuir-two site model fits. Error bars depict standard deviation from the reported means for replicate measurements (often hidden due to size of symbol).

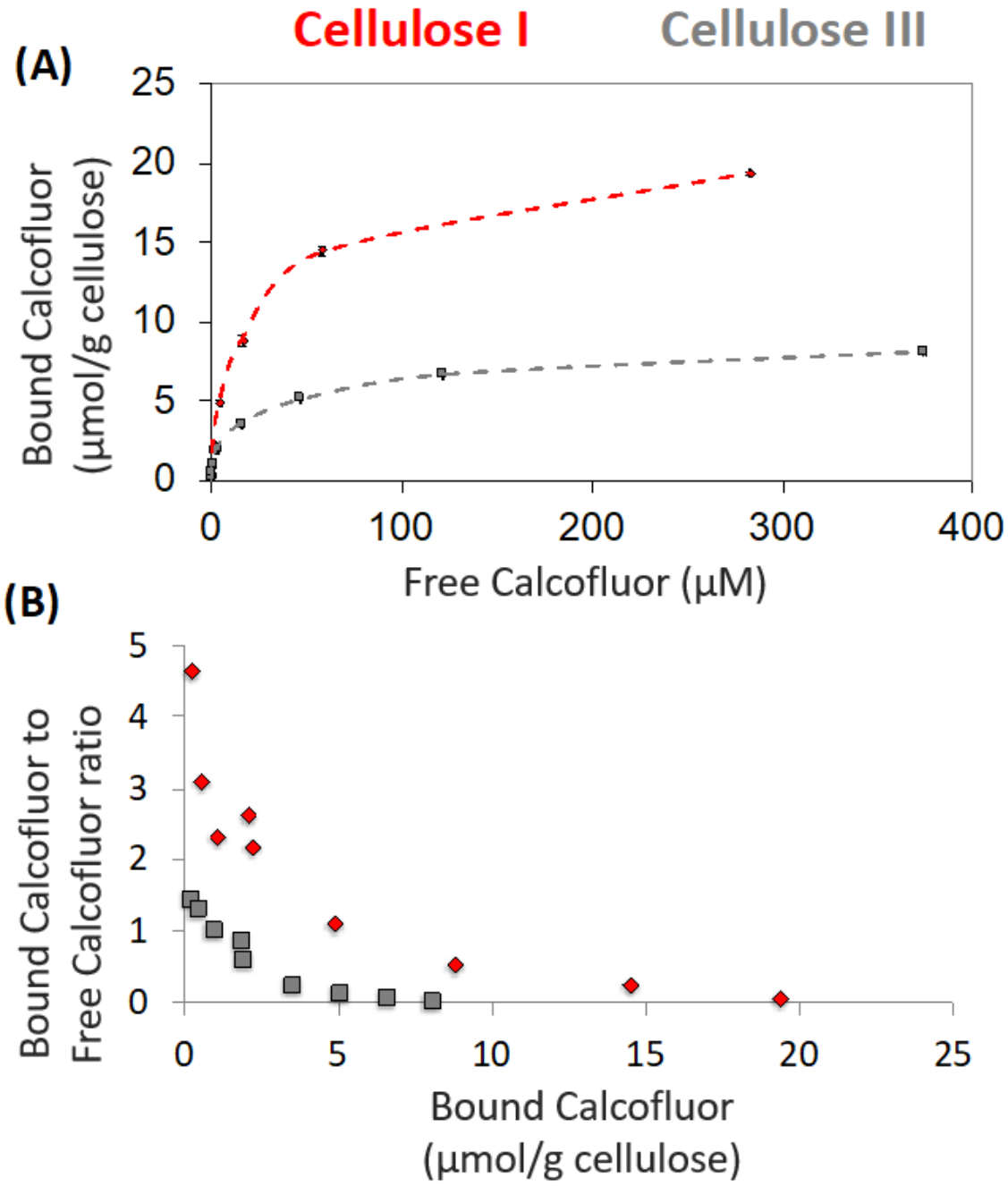


Fig. S16. Buffon needle problem inspired geometric probability model to determine likely orientations of CBM protein on flat cellulose binding surface. (A) Crystal structure of CBM1 (PDB code: 1CBH) represented as a needle (solid red line) of length 2.08 nm, (B) Hydrophobic face of cellulose I represented as an array of parallel lines (dotted black line) with a spacing of 0.8 nm, (C) Buffon needle inspired CBM-cellulose model schematic to determine geometric probability of all possible orientations of CBM1 on cellulose surface. The original Buffon model formulation is discussed in his classical 1777 paper (44). (D) Monte Carlo simulation methodology flowchart (see SI Appendix Experimental Procedures section).

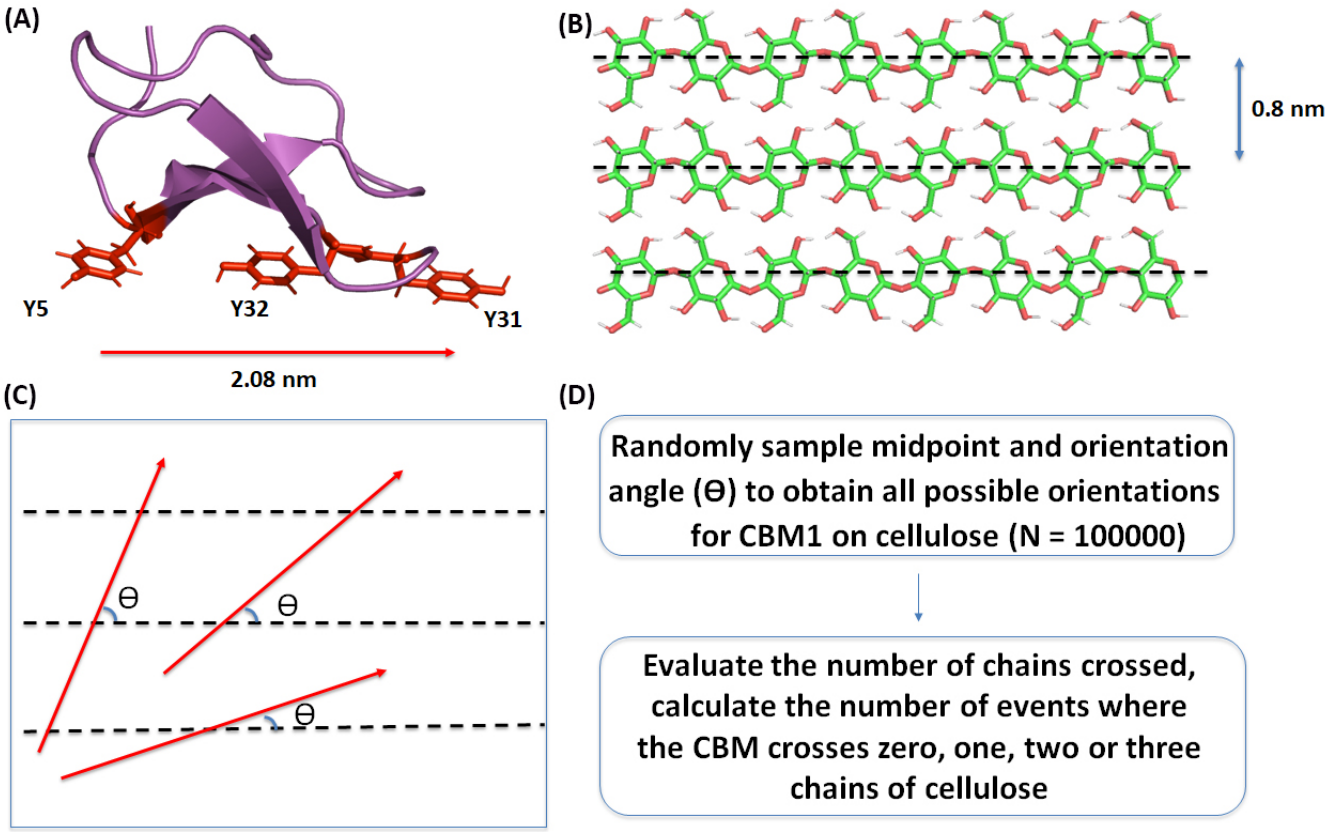


Table S1. Specific activity of Cel7A cellulase alone towards Cladophora cellulose I and cellulose III at 0.5, 2.5, and 10 mg enzyme per g cellulose loading. Specific activity (mean ± s.d.) is reported here in terms of nmol cellobiose released per μmol enzyme per minute of cellulose hydrolysis reaction.

	Cellulose - I	Cellulose-III
Enzyme loading (mg/g)	Specific Activity (nmol cellobiose per μmol enzyme per min)	
0.5	783.5 ± 324	909.3 ± 285
2.5	225.8 ± 23.6	787.8 ± 74.8
10	146.0 ± 73.0	426.0 ± 123

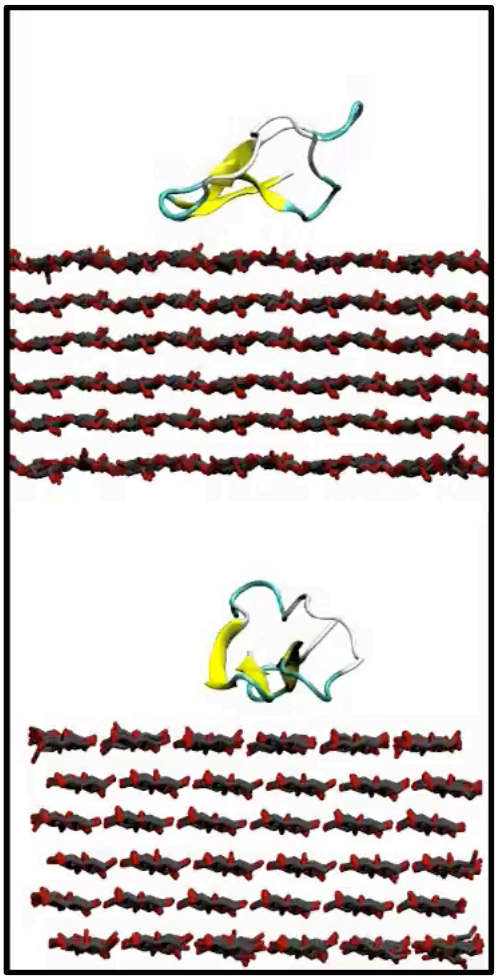
Table S2. Sensitivity analysis was performed by changing the predicted Langmuir one-site model parameters by 10% or 20% and checking for goodness of fit (i.e., root-mean square error or RMSE) of the original CBM1 binding dataset.

Sensitivity analysis for one-site model		
		RMSE
Original data predicted $n_{\max}$ & K_{d}		0.17
10% change in original model parameters	$0.90n_{\max}$	0.23
	$1.10n_{\max}$	0.17
	$0.90K_{\mathrm{d}}$	0.17
	$1.10K_{\mathrm{d}}$	0.17
20% change in original model parameters	$0.80n_{\max}$	0.35
	$1.20n_{\max}$	0.35
	$0.80K_{\mathrm{d}}$	0.20
	$1.20K_{\mathrm{d}}$	0.19

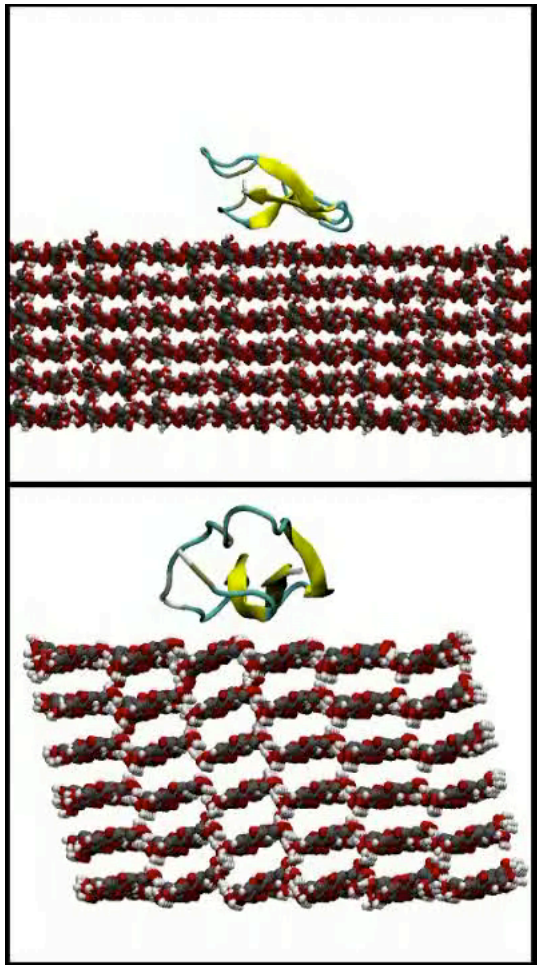
Table S3. CBM-cellulose Buffon needle model simulation predicts that wild-type (WT) CBM1 is equally likely to align and bind with its aromatic binding residues aligned both along a single (58% probability) or across multiple (42% probability) cellulose chains. However, the mutant CBM1 (Y31A) with a 'shortened' needle is more likely to align and bind along a single cellulose chain (90% probability).

Protein Type	Along the chain events	Across the chain events
CBM1 WT	58%	42%
CBM1 Y31A	90%	10%

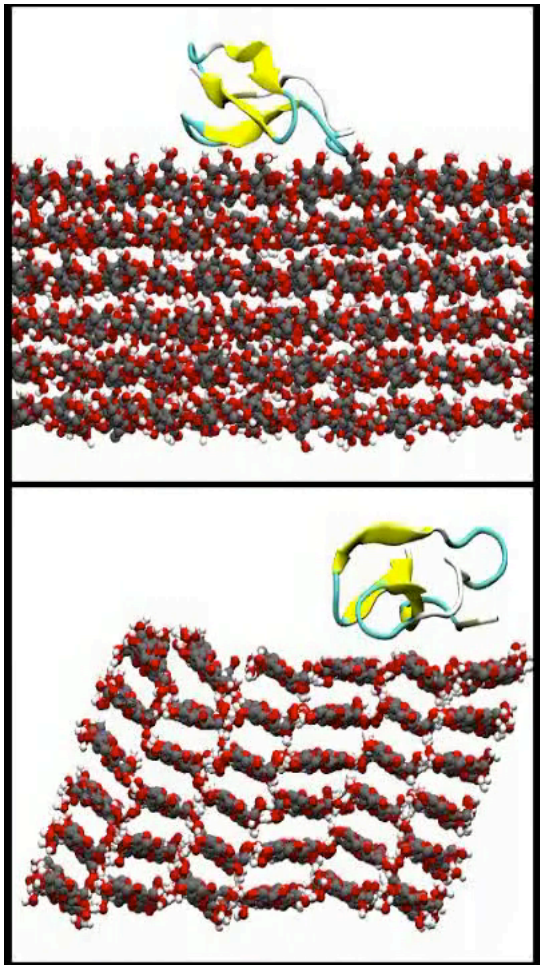
Movie S1. Snapshot of side-view (top) and cross-sectional (bottom) view of CBM1-cellulose I unbiased MD simulations.



Movie S2. Snapshot of side-view (top) and cross-sectional (bottom) view of CBM1-cellulose III (high crystallinity) unbiased MD simulations.



Movie S3. Snapshot of side-view (top) and cross-sectional (bottom) view of CBM1-cellulose III (low crystallinity) unbiased MD simulations.



SI Appendix References:

1. Brady, S. K., Sreelatha, S., Feng, Y., Chundawat, S. P. S., and Lang, M. J. (2015) Cellobiohydrolase 1 from *Trichoderma reesei* degrades cellulose in single cellobiose steps. *Nat. Commun.* **6**, 10149

2. Chundawat, S. P. S., Bellesia, G., Uppugundla, N., Sousa, L., Gao, D., Cheh, A., Agarwal, U., Bianchetti, C., Phillips, G., Langan, P., Balan, V., Gnanakaran, S., and Dale, B. E. (2011) Restructuring the crystalline cellulose hydrogen bond network enhances its depolymerization rate. *J. Am. Chem. Soc.* **133**, 11163–11174

3. Sousa, L. da C., Humpula, J., Balan, V., Dale, B. E., and Chundawat, S. P. S. (2019) Impact of ammonia pretreatment conditions on the cellulose III allomorph ultrastructure and its enzymatic digestibility. *ACS Sustain. Chem. Eng.* **7**, 14411–14424

4. Ruland, W. (1961) X-ray determination of crystallinity and diffuse disorder scattering. *Acta Cryst*. **14**, 1180–1185

5. Park, S., Baker, J., Himmel, M., Parilla, P., and Johnson, D. (2010) Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels*. **3**, 10

6. French, A. D. (2013) Idealized powder diffraction patterns for cellulose polymorphs. *Cellulose*. **21**, 885–896

7. Hult, L., Iversen, T., and Sugiyama, J. (2003) Characterization of the supramolecular structure of cellulose in wood pulp fibres. *Cellulose*. **10**, 103–110

8. He, J., Cui, S., and Wang, S.-Y. (2008) Preparation and crystalline analysis of high-grade bamboo dissolving pulp for cellulose acetate. *J Appl Polym Sci*. **107**, 1029–1038

9. Garvey, C. J., Parker, I. H., and Simon, G. P. (2005) On the interpretation of x-ray diffraction powder patterns in terms of the nanostructure of cellulose I fibres. *Macromol. Chem. Phys.* **206**, 1568–1575

10. Chundawat, S. P. S., Sousa, L. daCosta, Roy, S., Yang, Z., Gupta, S., Pal, R., Zhao, C., Liu, S.-H., Petridis, L., O'Neill, H., and Pingali, S. V. (2020) Ammonia-salt solvent promotes cellulosic biomass deconstruction under ambient pretreatment conditions to enable rapid soluble sugar production at ultra-low enzyme loadings. *Green Chem.* **22**, 204–218

11. Atalla, R. H., and Vanderhart, D. L. (1984) Native cellulose: a composite of two distinct crystalline forms. *Science (80-.).* **223**, 283–285

12. Agarwal, U., Reiner, R., and Ralph, S. (2010) Cellulose I crystallinity determination using FT–Raman spectroscopy: univariate and multivariate methods. *Cellulose*. **17**, 721–733

13. Agarwal, U. P. (2014) 1064 nm FT-Raman spectroscopy for investigations of plant cell walls and other biomass materials. *Front. Plant Sci.* **5**, 490

14. Atalla, R. J., and Vanderhart, D. L. (1987) Studies on the structure of cellulose using raman spectroscopy and solid state 13C-NMR. in *Institute for Paper Chemistry (IPC Technical Paper Series 217)*

15. Wiley, J. H., and Atalla, R. H. (1987) Band assignments in the raman spectra of celluloses. *Carbohydr. Res.* **160**, 113–129

16. Chundawat, S. P. S., Lipton, M. S., Purvine, S. O., Uppugundla, N., Gao, D., Balan, V., and Dale, B. E. (2011) Proteomics based compositional analysis of complex cellulase-hemicellulase mixtures. *J. Proteome Res.* **10**, 4365–4372

17. Gao, D., Chundawat, S. P. S., Sethi, A., Balan, V., Gnanakaran, S., and Dale, B. E. (2013) Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proc. Natl. Acad. Sci.* **110**, 10922–10927

18. Gao, D., Chundawat, S. P. S., Krishnan, C., Balan, V., and Dale, B. E. (2010) Mixture optimization of six core glycosyl hydrolases for maximizing saccharification of ammonia fiber expansion (AFEX) pretreated corn stover. *Bioresour. Technol.* **101**, 2770–2781

19. Whitehead, T. A., Bandi, C. K., Berger, M., Park, J., and Chundawat, S. P. S. (2017) Negatively supercharging cellulases render them lignin-resistant. *ACS Sustain. Chem. Eng.* **5**, 6247–6252

20. Lim, S., Chundawat, S. P. S., and Fox, B. G. (2014) Expression, purification and characterization of a functional carbohydrate-binding module from *Streptomyces sp. SirexAA-E*. *Protein Expr. Purif.* **98**, 1–9

21. Klock, H. E., Lesley, S. A., and Doyle, S. A. (2009) The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis High Throughput Protein Expression and Purification. *Methods Mol. Biol.* **498**, 91–103

22. Takasuka, T. E. ., Walker, J. A. ., Bergeman, L. F. ., Vander Meulen, K. A. ., Makino, S. I. ., Elsen, N. L. ., and Fox, B. G. (2014) Cell-free translation of biofuels enzymes. *Methods Mol. Biol.* **1118**, 71–95

23. Takasuka, T. E., Book, A. J., Lewin, G. R., Currie, C. R., and Fox, B. G. (2013) Aerobic deconstruction of cellulosic biomass by an insect-associated Streptomyces. *Sci. Rep.* **3**, 1030

24. Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234

25. Kavoosi, M., Meijer, J., Kwan, E., Creagh, A. L., Kilburn, D. G., and Haynes, C. A. (2004) Inexpensive one-step purification of polypeptides expressed in Escherichia coli as fusions with the family 9 carbohydrate-binding module of xylanase 10A from T. maritima. *J. Chromatogr. B*. **807**, 87–94

26. Hong, J., Ye, X., Wang, Y., and Zhang, Y. H. P. (2008) Bioseparation of recombinant cellulose-binding module-proteins by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. *Anal. Chim. Acta*. **621**, 193–199

27. McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A., and Kilburn, D. G. (2000) 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

28. Hu, W., Xie, J., Chau, H. W., and Si, B. C. (2015) Evaluation of parameter uncertainties in nonlinear regression using Microsoft Excel Spreadsheet. *Environ. Syst. Res.* 10.1186/s40068-015-0031-4

29. Linder, M., and Teeri, T. T. (1996) The cellulose-binding domain of the major cellobiohydrolase of *Trichoderma reesei* exhibits true reversibility and a high exchange rate on crystalline cellulose. *Proc. Natl. Acad. Sci.* **93**, 12251–12255

30. Goldstein, M. A., Takagi, M., Hashida, S., Shoseyov, O., Doi, R. H., and Segel, I. H. (1993) Characterization of the cellulose-binding domain of the Clostridium cellulovorans cellulose-binding protein A. *J. Bacteriol.* **175**, 5762–5768

31. Bhagia, S., Dhir, R., Kumar, R., and Wyman, C. E. (2018) Deactivation of cellulase at the air-liquid interface is the main cause of incomplete cellulose conversion at low enzyme loadings. *Sci. Rep.* **8**, 1350

32. Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958

33. Boonstra, S., Onck, P. R., and van der Giessen, E. (2016) CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state. *J. Phys. Chem. B*. **120**, 3692–3698

34. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008) GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655

35. López, C. A., Bellesia, G., Redondo, A., Langan, P., Chundawat, S. P. S., Dale, B. E., Marrink, S. J., and Gnanakaran, S. (2015) MARTINI coarse-grained model for crystalline cellulose microfibers. *J. Phys. Chem. B*. **119**, 465–473

36. Nishiyama, Y., Langan, P., and Chanzy, H. (2002) Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron x-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* **124**, 9074–9082

37. Wada, M., Chanzy, H., Nishiyama, Y., and Langan, P. (2004) Cellulose III crystal structure and hydrogen bonding by synchrotron x-ray and neutron fiber diffraction. *Macromolecules*. **37**, 8548–8555

38. Beckham, G. T., Matthews, J. F., Bomble, Y. J., Bu, L., Adney, W. S., Himmel, M. E., Nimlos, M. R., and Crowley, M. F. (2010) Identification of amino acids responsible for processivity in a family 1 carbohydrate-binding module from a fungal cellulase. *J. Phys. Chem. B*. **114**, 1447–1453

39. Nimlos, M. R., Beckham, G. T., Matthews, J. F., Bu, L., Himmel, M. E., and Crowley, M. F. (2012) Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J. Biol. Chem.* **287**, 20603–20612

40. Nishiyama, Y., Sugiyama, J., Chanzy, H., and Langan, P. (2003) Crystal structure and hydrogen bonding system in cellulose Iα from synchrotron x-ray and neutron fiber diffraction. *J. Amer. Chem. Soc.* **125**, 14300–14306

41. Brunecky, R., Subramanian, V., Yarbrough, J. M., Donohoe, B. S., Vinzant, T. B., Vanderwall, T. A., Knott, B. C., Chaudhari, Y. B., Bomble, Y. J., Himmel, M. E., and Decker, S. R. (2020) Synthetic fungal multifunctional cellulases for enhanced biomass conversion. *Green Chem.* **22**, 478–489

42. Lee, S.-H. (2018) Optimal integration of wide field illumination and holographic optical tweezers

for multimodal microscopy with ultimate flexibility and versatility. *Opt. Express*. **26**, 8049

43. Moran-Mirabal, J. M., Bolewski, J. C., and Walker, L. P. (2011) Reversibility and binding kinetics of *Thermobifida fusca* cellulases studied through fluorescence recovery after photobleaching microscopy. *Biophys. Chem.* **155**, 20–8

44. Buffon, G. (1777) Essai d'arithmétique morale. in *Histoire naturelle, générale er particulière, Supplément 4*, pp. 46–123

45. Wood, G. R., and Robertson, J. M. (1998) Buffon got it straight. *Stat. Probab. Lett.* **37**, 415–421

46. Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G., and Haynes, C. A. (1996) Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12229–12234

47. Sugiyama, J., Harada, H., and Saiki, H. (1987) Crystalline morphology of Valonia macrophysa cellulose III_I revealed by direct lattice imaging. *Int. J. Biol. Macromol.* **9**, 122–130

48. Bellesia, G., Chundawat, S. P. S., Langan, P., Dale, B. E., and Gnanakaran, S. (2011) Probing the early events associated with liquid ammonia pretreatment of native crystalline cellulose. *J. Phys. Chem. B*. **115**, 9782–9788

49. Sawada, D., Hanson, L., Wada, M., Nishiyama, Y., and Langan, P. (2014) The initial structure of cellulose during ammonia pretreatment. *Cellulose*. **21**, 1117–1126

50. Wada, M., Nishiyama, Y., Bellesia, G., Forsyth, T., Gnanakaran, S., and Langan, P. (2011) Neutron crystallographic and molecular dynamics studies of the structure of ammonia-cellulose I: rearrangement of hydrogen bonding during the treatment of cellulose with ammonia. *Cellulose*. **18**, 191–206

51. Abbott, D. W., and Boraston, A. B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol.* **510**, 211–31

52. Sugimoto, N., Igarashi, K., Wada, M., and Samejima, M. (2012) Adsorption characteristics of fungal family 1 cellulose-binding domain from *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose: Negative cooperative adsorption via a steric exclusion effect. *Langmuir*. **28**, 14323–9

53. Lehtio, J., Sugiyama, J., Gustavsson, M., Fransson, L., Linder, M., and Teeri, T. T. (2003) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 484–489

54. Wada, M., Heux, L., and Sugiyama, J. (2004) Polymorphism of cellulose I family: Reinvestigation of cellulose IV_I. *Biomacromolecules*. **5**, 1385–1391

55. Kankare, J. (2002) Sauerbrey equation of quartz crystal microbalance in liquid medium. *Langmuir*. **18**, 7092–7094

56. Hu, F., Zhang, Y., Wang, P., Wu, S., Jin, Y., and Song, J. (2018) Comparison of the interactions between fungal cellulases from different origins and cellulose nanocrystal substrates with different polymorphs. *Cellulose*. **25**, 1185–1195

57. Benziman, M., Haigler, C. H., Brown, R. M., White, A. R., and Cooper, K. M. (1980) Cellulose

biogenesis: Polymerization and crystallization are coupled processes in Acetobacter xylinum. *Proc. Natl. Acad. Sci.* **77**, 6678–6682

58. Wood, P. J. (1980) Specificity in the interaction of direct dyes with polysaccharides. *Carbohydr. Res.* **85**, 271–287

59. Herth, W., and Schnepf, E. (1980) The fluorochrome, calcofluor white, binds oriented to structural polysaccharide fibrils. *Protoplasma*. **105**, 129–133

60. Scatchard, G. (1949) The attractions of proteins for small molecules and ions. *Ann. N. Y. Acad. Sci.* **51**, 660–672

61. Alekozai, E. M., GhattyVenkataKrishna, P. K., Uberbacher, E. C., Crowley, M. F., Smith, J. C., and Cheng, X. (2014) Simulation analysis of the cellulase Cel7A carbohydrate binding module on the surface of the cellulose I_β. *Cellulose*. **21**, 951–971