Electronic Supporting Information

Measurements of Single Molecular Affinitive Interactions between Carbohydrate-binding Modules and Crystalline Cellulose Fibrils†

Mengmeng Zhang,[a] Bin Wang[a] and Bingqian Xu*[a]

[a] Single Molecule Study laboratory, College of Engineering and Nanoscale Science and Engineering Center, University of Georgia, Athens, GA 30602, USA. Fax: 01-706-542-3804; Tel: 0- 706-542-0502; E-mail: bxu@engr.uga.edu

ESI-I. Preparation of recombinant CBM 3a and CBM2a.

The sequence encoding CBM3a was amplified by PCR from the genomic DNA of *Clostridium thermocellum* strain ATCC 27405 (GeneBank accession No. CP00568; Nucleotide No. 3620608-3621084), and inserted into the Nco I and Xho I sites of expression vector pET28b [http://www.merck-chemicals.com/life-science-research/vector-table-novagen-pet-vector-table/c_HdSb.s1O77QAAAEhPqsLdcab] to form plasmid pCBM3a.

The translated amino acid sequence of the recombinant CBM3a is:

MGVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKDQTFWCD

HAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKNDW

SNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGLEHHHHHH.

It includes a $(His)_6$ purification tag, and has a predicted M.W. of 18770 and a pI of 6.51.

The plasmid was then transformed into *Escherichia coli* host strain Tuner(DE3) {F^- *ompT* *hsdS*$_B$ (r_B^- m_B^-) *gal dcm lacY1 λ(DE3 [lacI lacUV5-T7gene 1 ind1 sam7 nin5])*}. A colony transformant was grown at 37°C in 1 L Luria broth media supplemented with 50µg/mL of kanamycin (selection drug) until the culture reached a density of 0.5 OD at 600 nm. Induction was performed at 16°C for 10 h in the presence of 0.2 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (inducer). After induction, cells pelleted from the culture were re-suspended in 20 ml of a binding buffer (IBB = 25 mM Tris-Cl, pH 8.0, and 300 mM NaCl) and completely disrupted by sonication. Cell debris was removed by centrifugation at 25,000 ***g*** for 20 min at 4°C. The supernatant was further cleaned up by filtration using a 30-mL syringe with 0.45-µm Acrodisc Supor membrane disc.

Soluble recombinant CBM3a in the cell lysate was purified by Immobilized Metal Affinity Chromatography (IMAC). Specifically, the clear supernatant was applied to a column containing 2 mL bed volume of TALON™ metal-affinity resin (Clontech product, Mountain View, CA). The column was washed with 10 vol. of IBB, followed by successive washes of 5 vol. of IBB-5mM imidazole and IBB-10 mM imidazole, respectively. Pure CBM3a was eluted in 5 mL of IBB-100 mM imidazole, and dialyzed three times against 1 L of cellulose binding buffer (10mM Tris-Cl, pH7.5, 150 mM NaCl). The final protein sample, about 5 mL at a concentration of 20 mg/mL (1.1 mM) was stored in 0.5-mL aliquots at -20°C.

The translated amino acid sequence of the recombinant CBM2a is:

MQTATCSYNITNEWNTGYTGDITITNRGSSAINGWSVNWQYATNRLSSSWNANVSGSNPYS

ASNLSWNGNIQPGQSVSFGFQVNKNGGSAERPSVGGSICSGSVAIEGRHHHHHH.

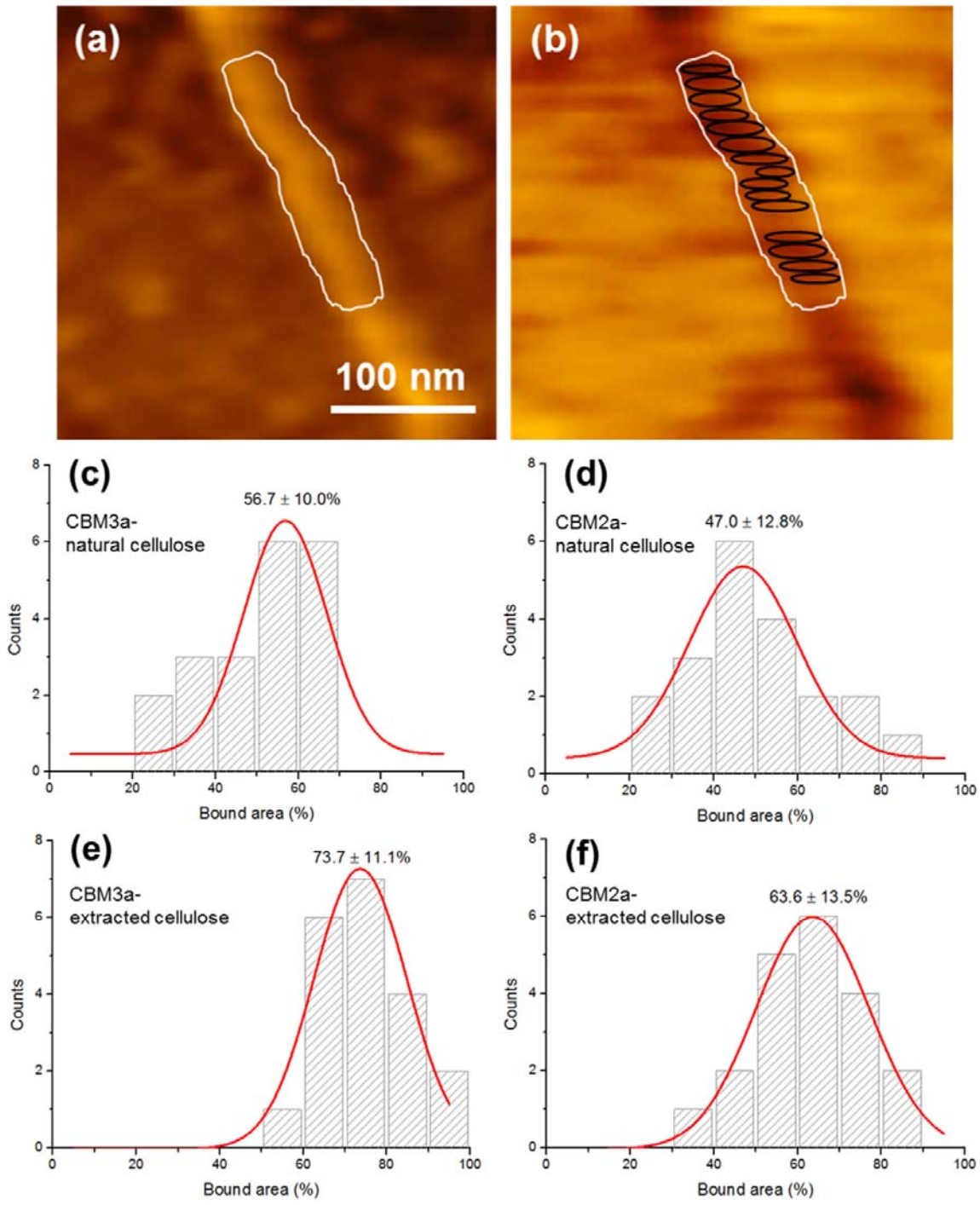
It includes a $(His)_6$ purification tag with a predicted M.W. of 12328 and a pI of 8.08.

ESI-II. Binding area percentage (BAP) calculation of natural cell wall cellulose and extracted single cellulose microfibril bound by CBM3a and CBM2a

We measured the BAP of CBM3a and CBM2a to natural and extracted cellulose microfibril surface in order to quantitatively compare the binding efficiencies of these two CBMs. The bound area generated by CBM3a or CBM2a modified tip during recognition imaging and the apparent surface area of cellulose were calculated by Picoscan software. The BAP was then calculated as shown in Fig. S1 (a-b) using the following equation:

$$BAP(\%) = \frac{\text{Bound area on single cellulose microfibril } (\text{nm}^2)}{\text{Total surface area of single cellulose microfibril } (\text{nm}^2)}$$

Here the bound area is highlighted as the black circle in the recognition image Fig. S1 (b) and the total surface area is highlighted as the white circle in both topographic and recognition image Fig. S1 (a-b). Each single cellulose microfibril is defined as a crystalline cellulose section about 200 nm in length and 25 nm in width. The BAP of 20 above cellulose sections was collected to construct a histogram for each CBM on each substrate. The final BAP value was determined by Gaussian fitting of each histogram and the results are shown in Fig. S1 (c-g).



(g)

	Natural	extracted
CBM3a	56.7±10.0%	73.7±11.1%
CBM2a	47.0±12.8%	63.6±13.5%

Fig. S1 The calculation of BAP. (a) and (b) Representative AFM imgaes used to calculated the BAP of CBM3a-bound single cellulose microfibril. Black circle: the bound area; white circle: the

total surface area. (c) and (d) BAP calculation of CBM3a and CBM2a (respectively) to natural cell wall cellulose. (e) and (f) BAP calculation of CBM3a and CBM2a (respectively) to single cellulose microfibril. (g): Summary of the above binding area percentage of the CBM-cellulose interaction. Each histogram was constructed from the BAP values calculated on 20 single cellulose microfibrils (approximately 200 nm in length and 25 nm in width). The binding percentage is the peak value followed with standard deviation that determined by Gaussian fitting of the histogram.

Based on the above binding area analysis, we found that under the same experimental conditions and scanning parameters (scan rate: 1.20 line/s, servo gain: 0.3/1.0), CBM3a was capable of binding about 9.7% more area than that of CBM2a on the natural cell wall cellulose and about 11.1% more area than that of CBM2a on the extracted single cellulose microfibrils.[1, 2]

ESI-III. Topographic and recognition imaging of extracted single cellulose micrfibrils by CBM2a-functionalized AFM tip

As shown in Fig. S2, the extracted cellulose microfibrils in topographic image (a) show clear recognition signals at the corresponding positions in recognition image (b). The cross-section profiles in (c) and (d) are the comparison of two representative single cellulose microfibrils marked as a-a', b-b' in (a) and c-c', d-d' in (b). Each binding event on these two microfibrils is highlighted by vertical red, dashed lines.

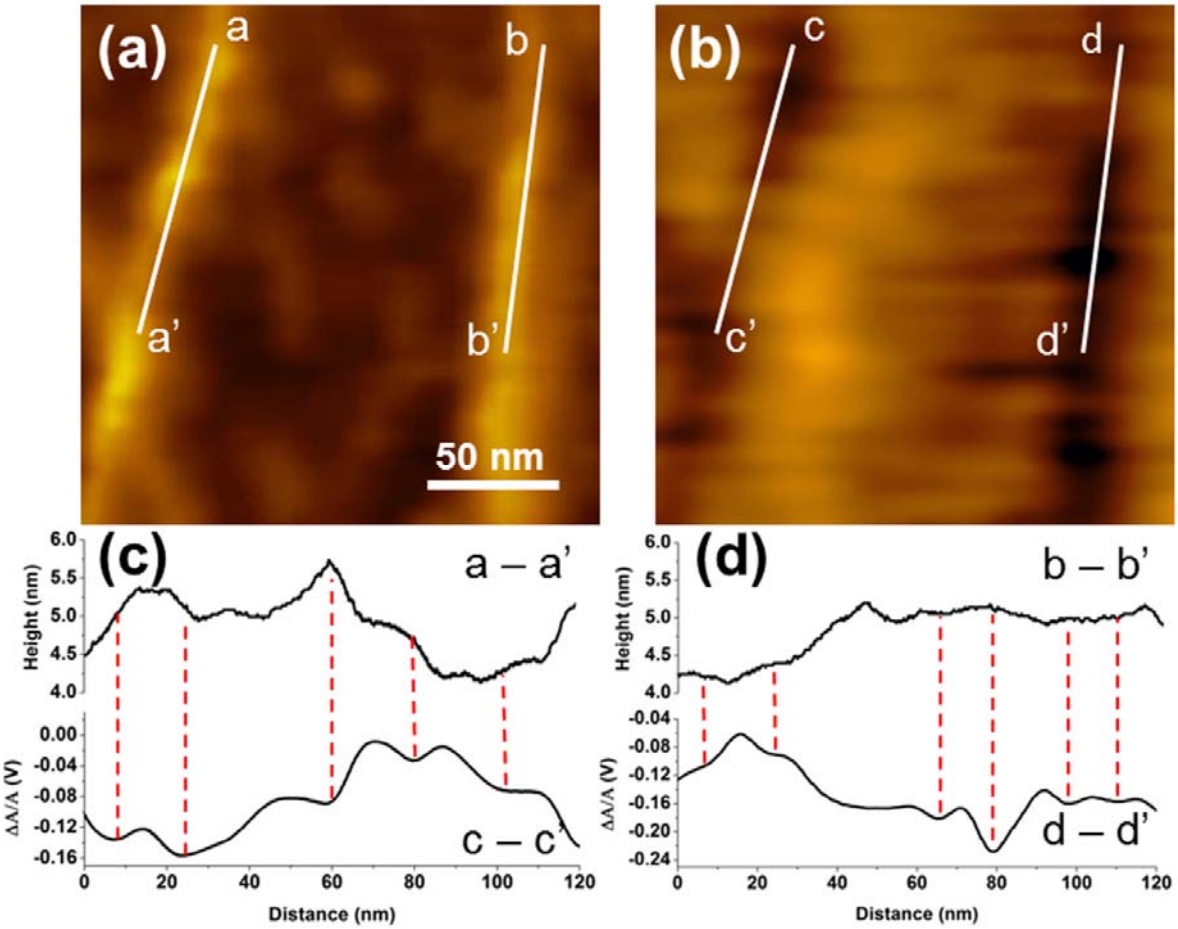


Fig. S2 Topographic (a) and recognition (b) images of extracted single cellulose microfibril with cross-section analysis (c) and (d) on a smaller area. (a) and (b) are imaged by CBM2a-functionalized AFM tip.

The cross-section analysis of CBM2a-celluose interaction shows less binding events on the extracted single cellulose microfibril surface with the same length than those between CBM3a and the extracted single cellulose microfibril surface. The binding intervals on the cellulose microfibril surface vary from 10 to 30 nm, which is larger than 5 to 20 nm as observed on the CBM3a-bound cellulose surface, indicating a less efficient binding of CBM2a to the crystalline cellulose surface.

ESI-IV. Representative force-distance curves under selected loading rates of CBM-extracted single cellulose microfibril interactions

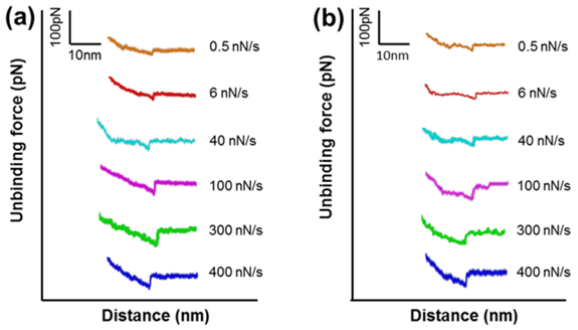


Fig. S3 Representative force-distance under 6 loading rates of the CBM3a- extracted single cellulose microfibril (a) and CBM2a- extracted single cellulose microfibril (b) unbinding interactions.

Fig. S3 shows the representative force-distance curves under selected loading rates. Each specific interaction curve is shown in the length of approximately 30-40 nm, and no obvious difference was observed between the shapes of the curves measured from CBM3a-extracted single cellulose microfibril and CBM2a- extracted single cellulose microfibril interaction.

ESI-V. The unbinding force and t_F vs. loading rate plots, the average force-distance curve and the reconstructed free energy for the interactions of CBM2a-extracted single cellulose microfibril, CBM3a-natural cell wall cellulose microfibril and CBM2a-natural cell wall cellulose microfibril.

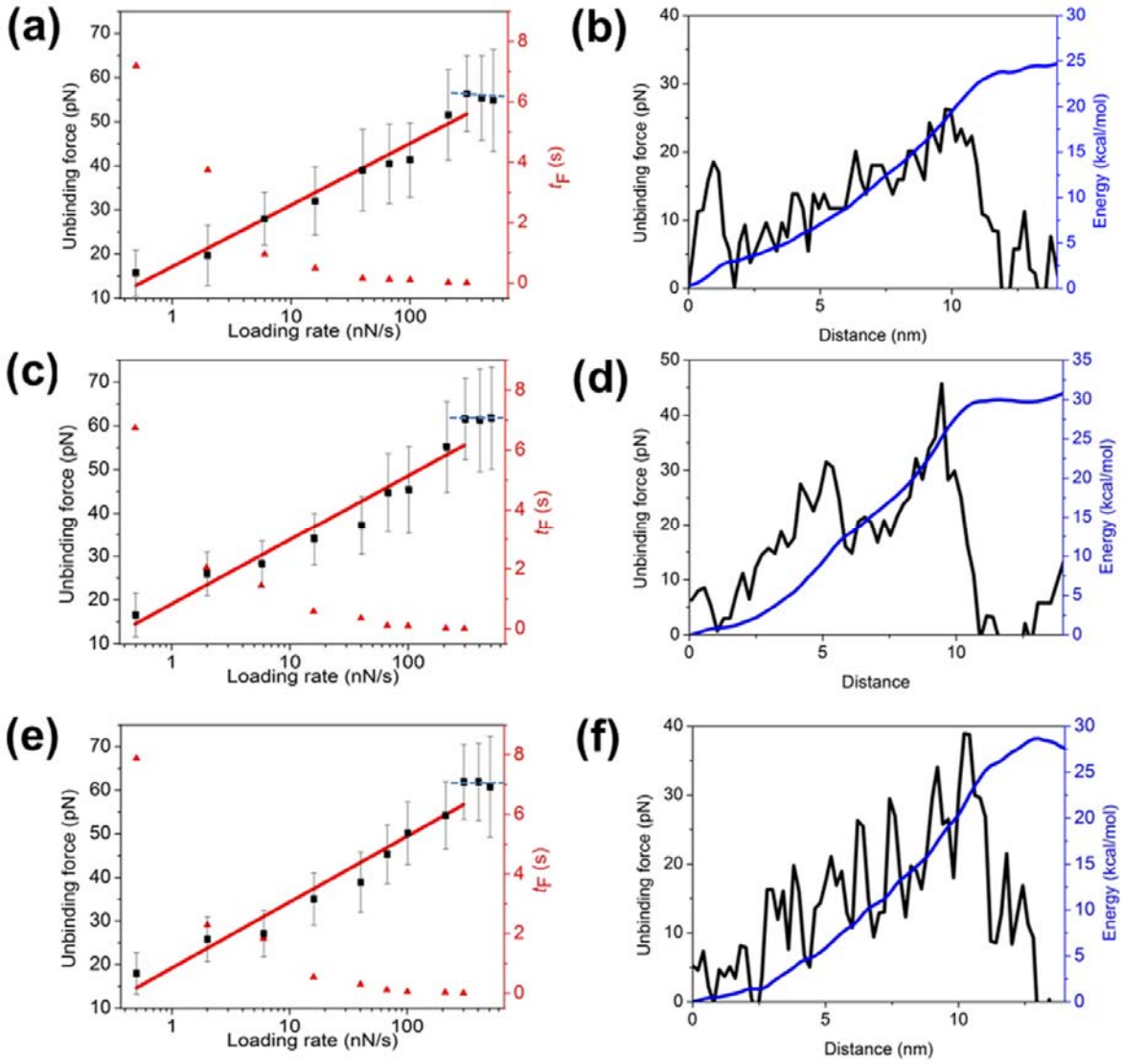


Fig. S4. (a, c, e) The unbinding force vs. loading rate and t_F plots for CBM2a- extracted single cellulose microfibril, CBM3a-natural cellulose microfibril and CBM2a-natural cellulose microfibril interactions, respectively. The data points of unbinding force are shown in black squares and the data points of t_F are shown in red triangles. The fitting plots for unbinding force are colored in red. The last three data points for the unbinding forces under the largest loading rates where the energy barrier is reached are marked by a dashed, blue line. (b, d, f) Average *F-D* curves (colored in blue) obtained by the weighted average work of individual *F-D* curves, and the reconstructed free energy (colored in black) obtained by the integral of the average

Each of the free energy profile is rebuilt based on 20 mean force-distance curves. The reconstructed free energy of CBM2a-extracted single cellulose microfibril interaction is 24.08 kcal/mol or 40.29 k_BT, for CBM3a-natural cellulose microfibril interaction, the value is 29.86 kcal/mol or 50.02 k_BT, and for CBM2a-natural cellulose microfibril interaction, the value is 28.61 kcal/mol or 48.02 k_BT. The differences among these CBM-cellulose complexes are not significant.

ESI-VI. The calculated and experimental t_F values plotted with the 9 loading rates before the saturated state

As indicated by Fig. S5, the experimental bond lifetime under external force t_F is in good consistency with the calculated values.

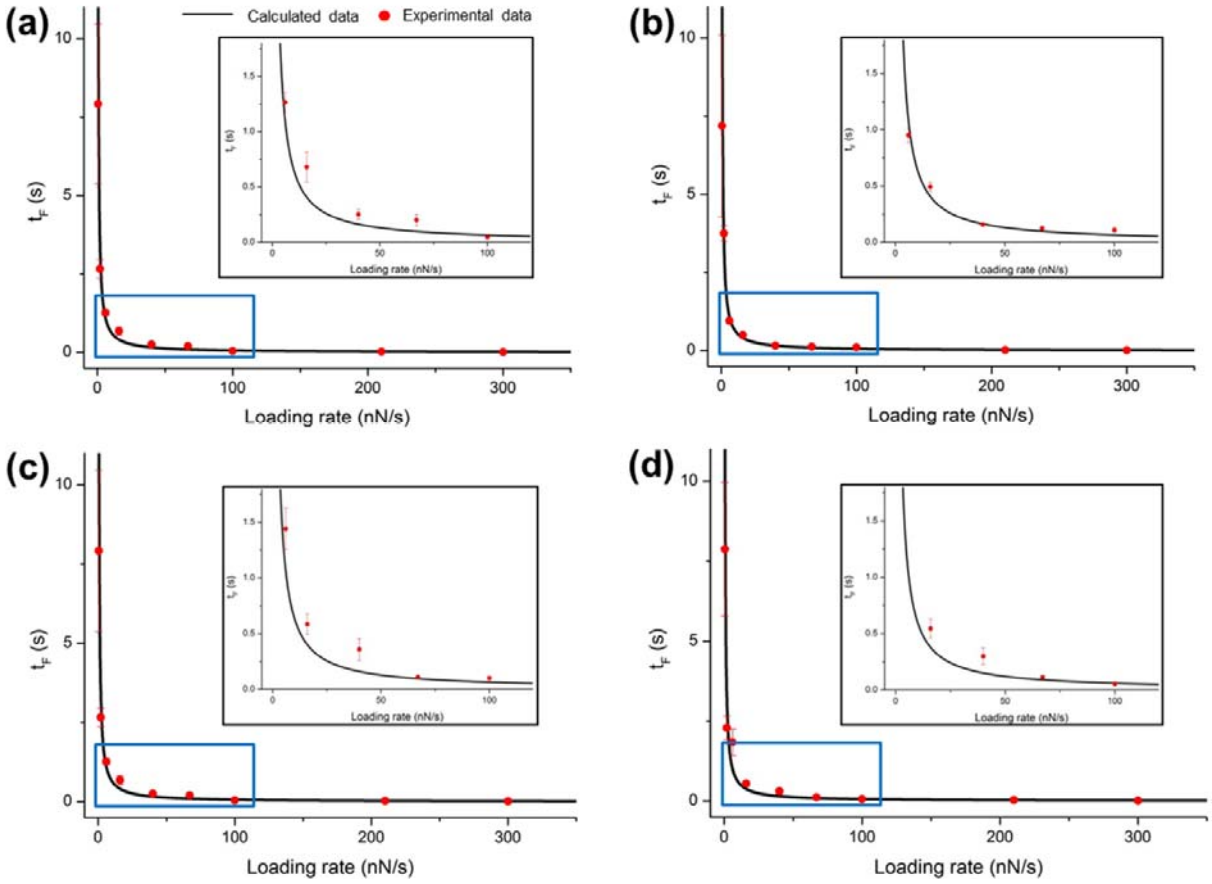


Fig. S5 Comparisons of calculated (black line) and experimental (red dots) t_F value for (a) CBM3a-extracted single cellulose microfibril complex, (b) CBM2a-extracted single cellulose microfibril complex, (c) CBM3a-natural cellulose microfibril complex, and (d) CBM2a-natural cellulose microfibril complex.

The t_F of CBM2a- extracted single cellulose microfibril interaction decreased from 7.19 s under 0.5 nN/s to 0.01 s under 300 nN/s). The t_F of CBM3a-natural cellulose microfibril interaction decreased from 6.74 s under 0.5 nN/s to 0.01 s under 300 nN/s, while that of CBM2a-natural cellulose microfibril interaction decreased from 7.88 s under 0.5 nN/s to 0.01 s under 300 nN/s. Similarly, CBM3a had a shorter bond lifetime on natural cellulose microfibril and longer bond lifetime on extracted single cellulose microfibril. However, CBM2a had a shorter bond lifetime on extracted single cellulose microfibril and longer bond lifetime on natural cellulose. The differences of bond lifetime among the four complexes are not very pronounced. The probable reason is that, compared to CBM2a, the highly conserved binding residues on the CBM3a planar strip are more sensitive to the impurities on the natural cell wall which can block the binding sites along the cellulose microfibrils. For CBM2a, the disruption of the crystalline structure on the cellulose microfibril surface probably has more influence to the planer strip and anchor residues for binding.[2,3]

References:

1. B. W. McLean, M. R. Bray, A. B. Boraston, N. R. Gilkes, C. A. Haynes and D. G. Kilburn, *Protein Eng.*, 2000, **13**, 801-809.
2. J. Tormo, R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham and T. A. Steitz, *Embo. J.*, 1996, **15**, 5739-5751.
3. A. W. Blake, L. McCartney, J. E. Flint, D. N. Bolam, A. B. Boraston, H. J. Gilbert and J. P. Knox, *J. Biol. Chem.*, 2006, **281**, 29321-29329.