Chimeric Cellulase Matrix for Investigating Intramolecular Synergism between Non-hydrolytic Disruptive Functions of Carbohydrate-binding Modules and Catalytic Hydrolysis*

Received for publication, November 16, 2011, and in revised form, June 23, 2012. Published, JBC Papers in Press, July 9, 2012, DOI 10.1074/jbc.M111.320358

Yuguo Wang†§, Rentao Tang†, Jin Tao‡, Xiaonan Wang§, Baisong Zheng‡, and Yan Feng†§†

From the †Key Laboratory for Molecular Enzymology and Engineering of Ministry of Education, Jilin University, 2699 Qianjin Street, Changchun 130012, China and ‡State Key Laboratory for Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

Background: Disruptive functions of carbohydrate-binding modules (CBMs) toward crystalline cellulose and their synergism with hydrolyses are important.

Results: Chimeric cellulases derived from the cellulase matrix have activities toward crystalline cellulose.

Conclusion: Analysis of chimeric cellulase activities allows quantification of the disruptive functions of CBMs.

Significance: An efficient strategy was established to investigate non-hydrolytic disruptive functions and their synergism with hydrolysis.

The conversion of renewable cellulosic biomass is of considerable interest for the production of biofuels and materials. The bottleneck in the efficient conversion is the compactness and resistance of crystalline cellulose. Carbohydrate-binding modules (CBMs), which disrupt crystalline cellulose via non-hydrolytic mechanisms, are expected to overcome this bottleneck. However, the lack of convenient methods for quantitative analysis of the disruptive functions of CBMs have hindered systematic studies and molecular modifications. Here we established a practical and systematic platform for quantifying and comparing the non-hydrolytic disruptive activities of CBMs via the synergism of CBMs and a catalytic module within designed chimeric cellulase molecules. Bioinformatics and computational biology were also used to provide a deeper understanding. A convenient vector was constructed to serve as a cellulase matrix into which heterologous CBM sequences can be easily inserted. The resulting chimeric cellulases were suitable for studying disruptive functions, and their activities quantitatively reflected the disruptive functions of CBMs on crystalline cellulose. In addition, this cellulase matrix can be used to construct novel chimeric cellulases with high hydrolytic activities toward crystalline cellulose.

Cellulosic biomass is a renewable resource, which can be converted into fermentable sugars, biofuels, and other materials in an environmentally friendly manner (1–5). Cellulose is the most abundant organic polymer on Earth and is an almost inexhaustible source of raw materials (4, 6). In the depolymerization of cellulose, cellulases can play a key role.

Cellulases can be classified into three major types according to their mode of catalytic action. (i) Endoglucanases (EC 3.2.1.4) cut randomly at internal amorphous sites in the cellulose polysaccharide chains. (ii) Exoglucanases including exo-1,4-β-glucosidases (EC 3.2.1.74) and exocellobiohydrolases (EC 3.2.1.91) act on the ends of cellulose chains. (iii) β-Glucosidases (EC 3.2.1.21) hydrolyze soluble celloolactins and celllobiose (7). Cellulases can also be classified into several glycoside hydrolase (GH) families (8) as shown on the Carbohydrate-Active eNzymes (CAzy) website (9). Multiple forms of cellulases work together to convert cellulose into fermentable sugars, and most commercial cellulase products for cellulose conversion are multicomponent. However, the available technologies for cellulose conversion require high enzyme loads, thus eliminating the cost advantage of converting abundant and low cost cellulosic biomass (5, 10).

Natural cellulose, which is mainly found in plant cell walls, is primarily crystalline cellulose. Its cellulose molecules are long unbranched linear chains linked by β-1,4-glycosidic bonds, and the chains are precisely arranged in parallel. As a result, they form a compact crystalline structure through strong interchain hydrogen-bonding networks, which make crystalline cellulose insoluble, stable, less accessible, and resistant to degradation (3, 11). Cellulases can only hydrolyze the amorphous regions of cellulose or the two ends of the cellulose chains of crystalline cellulose surfaces (7) and cannot penetrate the internal regions that are the main components of crystalline cellulose (supplemental Fig. S1). In other words, no matter how high the cellulase activities are, the enzymatic attack is limited to the cellulosic surfaces.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) JN590045, JN590046, JN590047, JN590048, and JN590049.

1 To whom correspondence should be addressed: State Key Laboratory for Microbial Metabolism and School of Life Science and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, China. Tel.: 86-21-34207189; Fax: 86-21-34207189; E-mail: yfeng2009@mail.sjtu.edu.cn.

2 The abbreviations used are: GH, glycoside hydrolase; CAZy, Carbohydrate-Active eNzymes; CBM, carbohydrate-binding module; GH5, glycoside hydrolase family 5; SMART, simple modular architecture research tool; CMC, carboxymethyl cellulose; MD, molecular dynamics; r.m.s.d., root mean square deviation.
lase load, only a tiny part of crystalline cellulose is exposed to cellulases. The compactness and resistance of crystalline structures are the bottleneck in the efficient degradation of natural cellulose. The looseness and disruption of crystalline structures are therefore more important (12) than simple enzymatic hydrolysis.

Several proteins and modules have disruptive functions toward crystalline cellulose via non-hydrolytic mechanisms (12, 13), including expansins (14), expansin-like proteins (15), swollenins (16), and carbohydrate-binding modules (CBMs) (17–19). They non-hydrolytically loosen, peel, split, or disrupt the packaging of crystalline cellulose and then convert it into its amorphous form, which can be readily hydrolyzed by cellulase. Although they do not directly hydrolyze cellulose, they can significantly increase the efficiency of enzymatic hydrolysis as a moiety within a cellulase molecule, as a subunit within a cellulase (19, 20), or as an independent component (21, 22). They are therefore expected to provide a novel approach to improve the hydrolysis of cellulosic materials. CBMs are the best candidates because of their relatively low molecular weight, greater efficiency toward crystalline cellulose, and greater availability of experimental methods, reported results, and other information (9, 12, 17–19).

The CBMs related to cellulose, which were originally defined as cellulose-binding domains, can be divided into several families according to the CAZy classification. Because of their binding and disruptive functions toward crystalline cellulose (17, 18), the CBMs may play a vital role in hydrolyzing natural crystalline cellulose, especially in the initial stages of hydrolysis. The binding functions of CBMs can be analyzed quantitatively (22), but these functions cannot characterize the disruptive functions of CBMs. As yet, there is no convenient method available for quantitatively analyzing the disruptive functions of CBMs. Because CBMs only rearrange segments of cellulose chains and do not release any new molecules, the non-hydrolytic disruption of crystalline cellulose by CBMs cannot be sensitively, directly, and conveniently quantified.

Heterologous CBMs and catalytic modules of cellulase can be united into simplified cellulosomes (23, 24) or fused together (25–27). The resulting cellulosomes or chimeric cellulases showed improved activities on crystalline cellulose as a result of the synergism between the CBMs and the catalytic modules. The enhancement of hydrolytic activities on crystalline cellulose may indirectly reflect the disruptive functions of CBMs. However, these studies were sporadic and not systematic, and the disruptive activities of various CBMs cannot be compared with each other. In this work, we developed a strategy for quantitatively analyzing the disruptive activities of CBMs utilizing the synergism of CBMs and catalytic modules. The endoglucanase FnCel5A, which can rapidly hydrolyze non-crystalline cellulose and does not hydrolyze crystalline cellulose (28), was used; the complementary combination of FnCel5A and CBMs can hydrolyze crystalline cellulose to release detectable sugars in a synergistic manner. We constructed a cellulase matrix containing FnCel5A in which heterologous CBMs could be inserted to form a series of chimeric cellulases. By investigating the activities of these chimeric cellulases, we could investigate the characteristics of CBMs.

### EXPERIMENTAL PROCEDURES

**Phylogenetic Analysis**—We assembled the protein sequence information from the CAZy database (9); the information was classified, annotated, and linked by CAZy. From the CAZy webpage of glycoside hydrolase family 5 (GH5) in August 2010, we chose the proteins that were annotated with appropriate EC (Enzyme Commission) numbers and then collected the non-redundant protein sequences of those proteins via the CAZy web page links.

We used these protein sequences to request the simple modular architecture research tool (SMART) (29, 30) service and processed these sequences according to the annotations of the domains and regions in SMART. We only reserved the sequences of enzymatic catalytic modules (catalytic domains) and CBMs present in GH5 enzymes. If there was more than one module in a protein sequence, we divided the sequence into several new independent sequences with only one module. These edited polypeptide sequences were used for the next phylogenetic analysis. ClustalX 2.0.12 (31) was used for alignments with default parameters. The phylogenetic trees of catalytic modules and CBMs were individually reconstructed.

The maximum likelihood phylogenetic trees were reconstructed using the PhyML v3.0 (32) program under the LG + $\Gamma$ (four rate categories; estimated $\alpha$) model of amino acid substitution and with 100 bootstrap replicates and were displayed using MEGA4 software (33). Bayesian Markov chain Monte Carlo phylogenetic analyses were performed using the BEAST (34) program under the JTT + $\Gamma$ + I + strict model with four $\gamma$ rate categories. The analysis of GH5 catalytic modules was run for 47 million Markov chain Monte Carlo steps, and the analysis of CBMs in GH5 enzymes was run for 100 million Markov chain Monte Carlo steps. All Markov chain Monte Carlo output results were analyzed using the program Tracer 1.5. The Bayesian phylogenetic trees were displayed using FigTree 1.3.1.

**Production of Endoglucanase FnCel5A**—The gene fnCel5A coding the endoglucanase was cloned from *Fervidobacterium nodosum* Rt17-B1 (DSMZ 5306; ATCC 35602) and was amplified by PCR. The amplified gene was digested with NdeI and BamHI and inserted into the similarly restricted vector pET-11a (Novagen) to yield the expression plasmid pFnCel5A. The endoglucanase cellulaseFnCel5A was solubly overproduced in the *Escherichia coli* BL21-CodonPlus (DE3)-RIL strain (Stratagene) harboring pFnCel5A. The experimental procedures of cloning, production, purification, and characterization were the same as those described previously (28).

**Preparation of Recombinant pFnCel5A-linker Vector**—Based on the characteristics of the Pro/Thr-rich linker in endoglucanase H from *Clostridium thermocellum* (P16218), which is closely related to *FnCel5A* (supplemental Fig. S2), we designed a linker sequence (Fig. 2b) and added it to the downstream sequence of FnCel5A in plasmid pFnCel5A using the rapid PCR site-directed mutagenesis method (35). The primers 5’-TGA-AGGTTTTCGAGAATAAGACGGATATGTCCCGGGTCC-3’ and 5’-CAAGACCGACCAAAACCGGCCGTAGCTGATAGAAGATCCGGCTGCTAAC-3’ were used to amplify the whole plasmid of pFnCel5A. PCR
amplification, used to create an insertion of the designed linker sequence, was performed using Pfu DNA polymerase (Fermentas). After 18 cycles of PCR, the sample was incubated with the restriction enzyme DpnI (Fermentas) for 3 h at 37 °C to remove the methylated template and concentrated by ethanol precipitation. The resulting product was transformed into E. coli XL-blue (Stratagene), and plasmid preparations of individual colonies were screened by restriction enzyme digestion for the insertion mutation, i.e. pFnCel5A-linker. The DNA sequences of the candidate plasmid were sequenced to verify identity with the anticipated sequence of pFnCel5A-linker. At the end of the linker sequence in pFnCel5A-linker, there are NheI and BamHI restriction sites (Fig. 2b) for inserting a heterologous CBM to form a novel chimeric cellulase.

Construction and Production of Chimeric Cellulases—The CBMs used in this work are summarized in supplemental Table S1. pGCIp2, pGCCIp4, and pGCCIp6 (23) were a kind gift from Pierre Bégün (Institute Pasteur, Paris, France), pSZ143 (36) and the plasmid containing the gene of Cel9A-90 wild type in the pBluescript plasmid (37) were from Diana C. Irwin (Cornell University, Ithaca, NY), The DNA fragments encoding TrCBM1-1, TrCBM1-2, CfCBM2, TjCBM2a, and TjCBM2 were amplified by PCR using the corresponding primers (supplemental Table S2), which supplied NheI and BamHI restriction sites, and respectively cloned into a suitable digested pFnCel5A-linker to generate pFnCel5A-TrCBM1-1, pFnCel5A-TrCBM1-2, pFnCel5A-CfCBM2, pFnCel5A-TjCBM2a, and pFnCel5A-TjCBM2 (Fig. 2c). These recombinant plasmids were transformed into the E. coli BL21-CodonPlus (DE3)-RIL strain (Stratagene) to produce the chimeric cellulases FnCel5A-TrCBM1-1, FnCel5A-TrCBM1-2, FnCel5A-CfCBM2, FnCel5A-TjCBM2a, and FnCel5A-TjCBM2, respectively.

The E. coli cells were cultured in 2 × YT medium (1.6% tryptone, 1% yeast extract, and 0.5% NaCl) and induced to produce the chimeric cellulases by isopropyl β-d-thiogalactoside. The cells were harvested, suspended, and disintegrated. After centrifugation, the chimeric cellulases in the supernatant were purified using a Q Sepharose Fast Flow (anion exchange) column (GE Healthcare). Enzyme purity was determined by SDS-PAGE analysis. Protein concentrations were determined according to the Bradford method, and bovine serum albumin was used as the standard. The chimeric cellulases were produced, purified, and determined as described for FnCel5A (28) with slight changes. In this study, the chimeric cellulase FnCel5A-CBMs include FnCel5A-TrCBM1-1, FnCel5A-TrCBM1-2, FnCel5A-CfCBM2, FnCel5A-TjCBM2a, and FnCel5A-TjCBM2, and their DNA sequences have been deposited in GenBank under accession numbers JN590045, JN590046, JN590047, JN590048, and JN590049, respectively.

Assay of Hydrolytic Activity—The hydrolytic activities of the FnCel5A and chimeric cellulases were measured after 5 or 30 min of incubation at a certain temperature in 50 mM phosphate, citrate buffer in the presence of 1% (w/v) substrate (carboxymethyl cellulose (CMC) or Avicel). We used CMC (sodium salt, medium viscosity; Fluka) as the amorphous cellulose. We used Avicel (PH-101; Fluka) as the crystalline cellulose in our laboratory experiments because it has been treated to remove hemi-celluloses and more extensive amorphous regions of cellulose fibers (7). The reducing sugars released from the substrates were determined using 3,5-dinitrosalicylic acid reagent (38). If the amount of reducing sugars was too small to be determined, we prolonged the incubation time or increased the enzyme concentration as appropriate. One unit (IU) of enzyme activity was defined as the amount of enzyme that releases 1 μmol of reducing sugars/min.

Effects of pH and Temperature on Activity—With CMC as the substrate, the pH and temperature optima of FnCel5A and the chimeric cellulases were determined by measuring the release of reducing sugars from CMC. The pH optimum was determined by measuring the enzyme activity at 83 °C in various pH buffers. The temperature optimum was determined in 50 mM phosphate, citrate buffer (pH 5.2) at different temperatures.

The pH dependence of chimeric cellulase activities on Avicel was examined by measuring the hydrolytic activities in various pH buffers at 50 °C. The temperature dependence of their activities on Avicel was investigated by determining the hydrolytic activities at different temperatures and in 50 mM phosphate, citrate buffer at their own optimal pH, which had been determined in the aforementioned analysis of their pH dependence.

Temperature Stability—Thermal stability studies were performed by incubating about 2 mg/ml purified enzymes for various lengths of time at different batches of temperatures in 50 mM Tris-HCl buffer (pH 7.0). The residual activities of the chimeric cellulases against Avicel were determined at 50 °C and pH 5.2.

Molecular Dynamics Simulations—The initial coordinates of the endoglucanase FnCel5A were taken from its crystal structure (Protein Data Bank code 3NCO). Among the CBMs used in this study (supplemental Table S1), the NMR structures of TrCBM1-1 and CfCBM2 are available, and their initial coordinates were obtained from the Protein Data Bank (codes 1CBH and 1EXG, respectively). Using the linker sequence (shown in Fig. 2b), we built and optimized an initial conformation for the linker using a previously described method (39). Finally, we joined the structures of FnCel5A (optimized by 12-ns molecular dynamics (MD) simulation in explicit solvent at 323 K), the linker, and the CBMs together to build the initial structures of the chimeric cellulase FnCel5A-CBMs using the InsightII package (Accelrys, San Diego, CA).

All MD simulations were unrestrained and carried out in the canonical ensemble using the sander module of AMBER version 11 (40). The f003 force field (41) was used, and the simulation system was enveloped in an octahedral box of TIP3P (42) water molecules. MD simulations were performed at 353 (80 °C) or 323 K (50 °C), and data analyses were performed using the packages of AmberTools 1.4 (40). The protein structures were visualized using the program PyMOL (Schrödinger, LLC) and VMD (43).

RESULTS
Phylogenetic Relationships of GHS Catalytic Modules and Associated CBMs—We aligned 327 polypeptide sequences of GH5 catalytic modules and constructed an unrooted maximum likelihood phylogenetic tree (supplemental Fig. S2a, left) and a Bayesian phylogenetic tree (supplemental Fig. S2b, left). We gathered and aligned 167 CBM sequences in GH5 enzymes and
Synergism between Non-hydrolytic Disruption and Hydrolysis

reconstructed their phylogenetic trees using the maximum likelihood method and the Bayesian method (supplemental Fig. S2, a and b, right, respectively). The sequence names of the catalytic module and the CBM within the same enzyme are linked in supplemental Fig. S2. These two figures were simplified and are shown as Fig. 1, a and b, respectively.

These four figures represent the same principle despite their slight differences due to the diversity of phylogenetic methods. Both catalytic modules and associated CBMs cluster into several distinct groups. The CBMs in the GH5 enzymes belong to 13 CBM families (CBM1, -2, -3, -4, -5, -6, -9, -10, -11, -17, -27, -28, and -33), which are distantly related or unrelated as revealed by the branch lengths within the phylogenetic trees of the CBMs. The catalytic modules of the GH5 enzymes tend to cluster into several groups with similar catalytic activities. However, the combinations of catalytic modules and CBMs in GH5 enzymes are irregular; i.e., these modules could be randomly present within an enzyme to some extent, suggesting inherent flexibility of the connections between catalytic modules and CBMs.

In addition, the lengths of the catalytic modules and CBMs in the GH5 enzymes were determined according to the annotations in SMART, corresponding to the spans of these modules shown in supplemental Fig. S2. The distributions of their polypeptide lengths were statistically analyzed and are shown in supplemental Fig. S3. The peptide lengths of catalytic modules were between 233 and 359 residues, and most of them (93%) were between 250 and 350 residues. On the other hand, the peptide lengths of CBMs varied considerably and were between 29 and 221 residues. The size of the biggest CBM was 7.6 times that of the smallest CBM. The majority (78%) of these CBMs were distributed in two clusters: 42% had small sizes (29–47 residues), and 36% had medium sizes (76–102 residues).

Construction and Production of Chimeric Cellulases—First, based on soluble overproduction and easy purification of FnCel5A (28), the vector pFnCel5A-linker was constructed to add a linker sequence with three restriction sites to the 3' terminus of the gene encoding FnCel5A (Fig. 2). The linker can be replaced with other linker sequences using Smal and Nhel restriction sites or Smal and BamHI. Heterologous CBM sequences can be inserted between Nhel and BamHI restriction sites to form new plasmids that could produce chimeric cellulases containing the catalytic module of FnCel5A, heterologous CBMs, and the linker between these two modules.

We constructed five recombinant plasmids by inserting five different heterologous CBM sequences (supplemental Table S1) in the vector pFnCel5A-linker. The five chimeric cellulases were solubly overproduced in E. coli cells harboring the recombinant plasmids. The production and purification results for the chimeric cellulases are similar to those for FnCel5A except that the purified proteins have various molecular weights. The construction and properties of these chimeric cellulases are summarized in Table 1.

Enzymatic Properties of Chimeric Cellulases—For consistency in comparisons, we measured the hydrolytic activities of FnCel5A and the five chimeric cellulases against CMC uniformly at 83 °C in 50 mM phosphate, citrate buffer (pH 5.2) and measured their hydrolytic activities on Avicel at 50 °C and pH 5.2. Their hydrolytic activities are listed in Table 1. Although the hydrolytic activities of the chimeric cellulases toward CMC were slightly lower than that of FnCel5A, their activities against Avicel were evident. Moreover, different chimeric cellulases had different activities against Avicel depending on the particular linked CBMs,
and the highest activity was ~5 IU/mg. In contrast, FnCel5A without a CBM did not obviously hydrolyze Avicel. It is just linking a CBM with the FnCel5A molecule that could make the resulting chimeric cellulase effectively hydrolyze crystalline cellulose.

The effects of pH and temperature on the activities of chimeric cellulases were measured using CMC and Avicel as the substrates. We found that their optimum pH values for activities on CMC were around 5 (Fig. 3a), and the optimum temperatures
were 80–85 °C (Fig. 3b). The results (Fig. 3, a and b) show that the pH and temperature profiles of the five chimeric cellulases for activities on CMC are roughly similar to each other and to those of FnCel5A (28). The enzymatic properties of chimeric cellulases on CMC may be primarily determined by the characteristics of the FnCel5A moiety, which shows high hydrolytic activities on CMC, because CBMs are the non-catalytic modules (17). On the other hand, the pH and temperature dependences of the activities of the five chimeric cellulases against Avicel are not only different from those of FnCel5A activity on CMC but are also different from each other (Fig. 3, c and d). The most striking change is that the optimum temperatures of the chimeric cellulases are significantly lower than that of FnCel5A. Because FnCel5A has no detectable activity on Avicel alone, the properties of the CBMs in the five chimeric cellulases are critical in hydrolyzing Avicel. The hydrolytic activities of the chimeric cellulases FnCel5A-TrCBM1-1, FnCel5A-TrCBM1-2, FnCel5A-CfCBM2, FnCel5A-TfCBM2a, and FnCel5A-TfCBM2, which were measured against Avicel at their own optimal pH values and temperatures (Fig. 3, c and d), were 1.55, 2.18, 3.16, 5.75, and 0.46 IU/mg, respectively.

Temperature Stabilities of Chimeric Cellulases—The activities of the five chimeric cellulases against Avicel quickly decreased at high temperatures (data not shown). We therefore determined the temperature stabilities of the chimeric cellulases at lower temperatures than those for FnCel5A. The chimeric cellulase solutions were incubated at 45–55 °C, and then the residual activities were measured against Avicel. Fig. 4 shows that the temperature stabilities of these chimeric cellulases are obviously inferior to that of FnCel5A, although the
The eight conserved amino acid residues of FnCel5A are Arg-59, His-103, Asn-143, Glu-144, His-203, Tyr-205, Glu-260, and Trp-293, as suggested previously (28); these residues may form the bottom surface of the catalytic cleft, which is shown in Fig. 6. Among these eight residues, the catalytic residues Glu-144 and Glu-260 act as the catalytic proton donor and nucleophile, respectively, according to the alignment (28) and crystal structure analysis of the endoglucanase Cel5A from Thermostoga maritima (44), which is the endoglucanase most closely related to FnCel5A to date (as shown in supplemental Figs. S2 and S5). On the other hand, we identified the aromatic residues (Tyr-354, Tyr-380, and Tyr-381 of FnCel5A-TrCBM1-1 and Trp-362, Trp-399, and Trp-417 of FnCel5A-CfCBM2) as the functionally important residues in accordance with the information on TrCBM1-1 (45) and CfCBM2 (46) in the literature; these conserved aromatic residues contribute to stacking interactions with the glucose rings of crystalline cellulose surfaces (17, 47). These aromatic residues are highlighted in Fig. 6.

We performed 25-ns MD simulations for FnCel5A-TrCBM1-1 and FnCel5A-CfCBM2 at 323 K (50 °C) in explicit solvent. We show their backbone r.m.s.d. values in supplemental Fig. S6. The r.m.s.d. values for the FnCel5A moieties of chimeric cellulases fluctuated narrowly similar to that for the individual FnCel5A protein (supplemental Fig. S4), and the r.m.s.d. fluctuations for the main structures of the CBM moieties were narrow; we therefore consider that their structures could remain independent and stable in the chimeric cellulases. In contrast, the r.m.s.d. values for the entire molecules of the chimeric cellulases fluctuated wildly and never converged perhaps because the two modules linked by the flexible linker could move with relatively little restriction in the same molecule. These results suggest that the moieties of FnCel5A (catalytic module) and CBMs could retain their respective independent structures, and their relative spatial positions could be shifted relatively freely in the same chimeric cellulase molecule. We used Glu-144 and Glu-260 to represent the active sites for catalysis and represented the binding sites of the two CBMs using the functional aromatic residues aligned on the surfaces (shown in Fig. 6). We showed that the distances between the active sites and the binding sites represented by these key residues fluctuated around 2–8 nm during the MD simulation of FnCel5A-

chimeric cellulases can maintain activity against Avicel for a sufficient length of time, i.e. a few hours, at moderately high temperatures. FnCel5A is very stable at high temperatures (28), and the activities of FnCel5A and these chimeric cellulases on CMC remained approximately constant at 45—55 °C. The temperature stabilities of the chimeric cellulases on Avicel were therefore mainly affected by the CBM moieties in their molecules.

**Computational Simulations of Cellulases**—Two MD simulations of FnCel5A were performed in explicit solvent at 323 K (50 °C) for 12 ns and at 353 K (80 °C) for 8 ns, respectively. We show the backbone root mean square deviation (r.m.s.d.) values between the trajectory structures and the starting structures of the MD simulations in supplemental Fig. S4. The narrow r.m.s.d. fluctuations suggest that the overall structure of FnCel5A is stable at 50 and 80 °C. Furthermore, we superposed the crystal structure of FnCel5A and its structures at 50 and 80 °C in Fig. 5. As a member of GH5, FnCel5A shows a common (βα)_{8}-barrel fold. Comparative analysis of these three structures shows that the “stability face” of the (βα)_{8}-barrel enzyme had conformational stability; in contrast, its “catalytic face” had more flexibility in different states, even for the same enzyme, i.e. FnCel5A.

![FIGURE 5. Superposition of FnCel5A tertiary structures in the crystalline state (blue; Protein Data Bank code 3NCO), at 50 °C (yellow), and at 80 °C (red). The structures of FnCel5A at 50 and 80 °C are the average structures of snapshots over their final 2-ns MD simulation intervals, respectively. The three-dimensional structure status of GH5 enzymes is a (βα)_{8}-barrel structure, so the FnCel5A molecule could be divided into a catalytic face and a stability face according to the suggestion of Höcker et al. (49).](https://example.com/figure5)

The linker peptide is shown in yellow, a., structure of FnCel5A-TrCBM1-1; the FnCel5A moiety is viewed from the catalytic face. b., structure of FnCel5A-CfCBM2; the FnCel5A moiety is viewed from the side.

![FIGURE 6. Tertiary structures of chimeric cellulases. The FnCel5A moiety is shown in blue. Its conserved residues (Arg-59, His-103, Asn-143, Glu-144, His-203, Tyr-205, Glu-260, and Trp-293) form the bottom surface of the catalytic cleft (shown in “red”). The catalytic residues Glu-144 and Glu-260 are indicated by red sticks. CBM moieties are shown in green, and their aromatic residues with functional significance are indicated by red sticks. The linker peptide is shown in yellow. a., structure of FnCel5A-TrCBM1-1; the FnCel5A moiety is viewed from the catalytic face. b., structure of FnCel5A-CfCBM2; the FnCel5A moiety is viewed from the side.](https://example.com/figure6)
TrCBM1-1 (supplemental Fig. S7) and around 4–10 nm during the MD simulation of FnCel5A-CfCBM2 (supplemental Fig. S8).

Furthermore, we took simulated snapshots from the MD simulations at 1-ns intervals after 5.0 ns and superimposed the structures from the snapshots in Fig. 7. We could summarize the relative motions of the FnCel5A moiety, linker, and CBM within a chimeric cellulase molecule and display the changes and probabilities of their relative spatial positions during the MD simulations. These figures show that the CBM moieties (or the FnCel5A moieties) tended to move around the FnCel5A moieties (or the CBM moieties), and their trajectories half-encircled the FnCel5A moieties (or the CBM moieties) like umbrellas when we centered and restricted the FnCel5A moieties (or the CBM moieties) and vice versa. This kind of image for superimposed tertiary structures could visualize the probable spatial positions and motion trajectories of one module relative to the other, which was artificially restricted at the center. In other words, the probable positions and motion of one module were pictured as a probability “cloud” around the other. Moreover, probability cloud density could represent the frequency with which one module or its peptide fragment was likely to be in a specific region. Fig. 7 shows that the FnCel5A moieties and the CBM moieties could, respectively, be closely half-encircled by the motion trajectories of the other moieties and the linker peptides like umbrellas, and their active sites for catalysis and binding were fully exposed on the protein surfaces without any obstruction.

DISCUSSION

The relatively low activities of cellulases have conventionally been considered to restrict the efficient degradation and conversion of cellulosic biomass. However, there are a few cellulases that can hydrolyze amorphous cellulose polysaccharide chains with high activities, but their hydrolytic activities toward crystalline cellulose are very low because of difficulties penetrating crystalline cellulose. The key to effective hydrolysis of crystalline cellulose is to disrupt the crystalline structure and expose the cellulose polysaccharide chains. The disruptive functions of CBMs toward crystalline cellulose (17) may play a key role and be even more important than the hydrolytic activities of hydrolases in the conversion of cellulosic biomass.

The need for a convenient and efficient method of quantitatively analyzing the non-hydrolytic disruptive functions of CBMs is evident. The non-hydrolytic disruptive functions may only rearrange segments of polysaccharide chains in crystalline cellulose; these slight physical changes have hardly been observed and quantified. In contrast, the measurement of hydrolase activities can be easily performed by determining the amounts of sugars released. CBMs and hydrolases can be combined, and the disruptive functions can be quantitatively investigated via their synergism. A CBM can disrupt and convert a well ordered crystalline region of cellulose into an amorphous region with exposed easily hydrolyzable amorphous sites of cellulose but cannot actually hydrolyze cellulose. After an amorphous cellulose region comes in contact with a hydrolase and is subsequently hydrolyzed by the latter, a synergistic reaction is completed with the release of detectable sugars. The non-hy-

FIGURE 7. Superimposed tertiary structures of chimeric cellulases FnCel5A-TrCBM1-1 (a and b) and FnCel5A-CfCBM2 (c and d) obtained from MD simulations at 323 K (50 °C) at 1-ns intervals after 5.0 ns. With the FnCel5A moiety (residues 8–320; shown in blue) at the centers, the simulated structures of FnCel5A-TrCBM1-1 (a) and FnCel5A-CfCBM2 (c) were superimposed using the r.m.s.d. tool of VMD. With the CBM moieties (residues 342–385 for TrCBM1-1 and residues 342–455 for CfCBM2; shown in green) at the centers, the simulated structures of FnCel5A-TrCBM1-1 (b) and FnCel5A-CfCBM2 (d) were superimposed by the same method. The chimeric cellulase linkers are shown in yellow. The protein molecules are shown in schematic mode. The catalytic residues (Glu-144 and Glu-260 of the FnCel5A moiety) and the functional aromatic residues (Tyr-354, Tyr-380, and Tyr-381 of FnCel5A-TrCBM1-1 and Trp-362, Trp-399, and Trp-417 of FnCel5A-CfCBM2) are shown by red sticks.
Synergism between Non-hydrolytic Disruption and Hydrolysis

drolytic disruption is the rate-determining step in a synergistic reaction. Therefore, the amount of released sugars may indirectly reflect the disruptive function if other processes in a synergistic reaction are fast enough.

At present, these synergistic reactions can be achieved in three ways. (i) Independent CBMs and hydrolases are simply mixed in a reaction system. However, CBMs and hydrolases do not interact closely with each other, and the results can be influenced by many factors (e.g. the proportions of the ingredients). In addition, many CBM peptides are too small to be produced in general production systems. (ii) CBMs and hydrolases are united into simplified cellulosomes (23, 24). However, this mode of connection may weaken the synergism between CBMs and hydrolases because the relationship between them might become distant. (iii) CBMs and catalytic modules of hydrolases can be fused together by direct connection or using linker peptides. CBMs are usually independently folding units and can function autonomously in chimeric proteins (19). The resultant chimeric cellulases showed improved activities on crystalline cellulose (25–27). In our strategy, an investigated CBM is connected to a cellulase with known enzymatic properties (e.g. FnCel5A) by a designed flexible linker, and the quantitative activities of the constructed chimeric cellulase should reflect the properties of this CBM.

We first investigated the feasibility of this strategy via phylogenetic analysis, statistics, and computational simulations. Based on the results for the GH5 enzymes (Fig. 1 and supplemental Figs. S2 and S3), it is likely that catalytic modules can be assembled with various sizes and structures of CBMs from various CBM families and distant organisms with few restrictions. The flexible combinations of catalytic modules and CBMs in nature improve the ability of microorganisms to exploit new environments. We speculate that a possible process of the combinations occurs through horizontal (or lateral) gene transfer between heterologous organisms (48), the fusion and recombination of genes, and the modification of combined genes. The natural combinations of catalytic modules and CBMs are nevertheless finite and are rarely suitable for applications; more combinations can be specially designed and constructed in a laboratory. We used the cellulase Fncel5A, which is a (βα)₁₁-barrel enzyme and contains only one catalytic module (Fig. 5), without any CBM or linker peptide. Its extended N and C termini back onto its catalytic cleft, suggesting that an additional peptide linked to its termini may not cause an obstruction. We designed a flexible linker peptide to connect Fncel5A and CBMs based on phylogenetic information and experience. Finally, we regarded TrCBM1-1 and CjCBM2 as two typical CBMs because their tertiary structures are available, and their peptide lengths are representative of CBMs in GH5 enzymes (supplemental Fig. S3b). We performed computational simulations of the chimeric cellulases containing these two target CBMs.

The simulation results indicate that the catalytic module and the CBMs within the same cellulase molecule could not only retain their independent and unobstructed functional structures but also adopt appropriate close spatial positions to achieve synergistic reactions via this strategy. Furthermore, the motion trajectories of one module relative to the other formed umbrella-like half-encircled configurations (Fig. 7). When the CBM disorders well aligned cellulose molecules of crystalline cellulose and creates a patch of amorphous cellulose region, the short diffusion distances between this amorphous region and the intramolecular catalytic module speed up their approach and encounter, and the umbrella-like half-encircling motion provides the intramolecular catalytic module with more opportunities to promptly make contact with this amorphous region and hydrolyze it. These characteristics make it easy and fast for a patch of amorphous cellulose created by the CBM to contact the intramolecular catalytic module and be hydrolyzed.

To establish a practical and systematic experimental method, we constructed an expression vector containing a suitable cellulase sequence and a designed linker sequence that serves as a vector of the cellulase matrix for chimeric cellulases. Heterologous CBM sequences could be inserted between the restriction sites at the end of the linker sequence and connect with the catalytic module to form a series of chimeric cellulases (shown schematically in Fig. 2). By investigating the activities of chimeric cellulases on cellulose substrates, we could quantitatively analyze and compare the characteristics of CBMs.

To establish a cellulase matrix of chimeric cellulases to quantify non-hydrolytic disruptive activities of CBMs, it is crucial to select a suitable cellulase with known enzymatic properties as the catalytic module. A suitable cellulase must (i) have a high hydrolase activity against non-crystalline cellulose, (ii) show no activity toward crystalline cellulose, and (iii) have excellent stability. When a suitable cellulase is linked with target CBMs and works in a synergistic manner, the target CBMs disrupt and convert crystalline cellulose into an easily hydrolyzable form, which can be immediately hydrolyzed to reducing sugars by this cellulase moiety. Thus, the amounts of reducing sugars released should indirectly represent the non-hydrolytic disruptive activities toward crystalline cellulose. Meanwhile, the catalytic module of the selected cellulase without activities on crystalline cellulose can only hydrolyze the easily hydrolyzable regions created by the CBMs and cannot directly hydrolyze crystalline cellulose to release extra detectable sugars. The non-hydrolytic disruption of crystalline cellulose by CBMs is the rate-determining step and determines the measurable activities against crystalline cellulose. In this work, we used the thermostable endoglucanase Fncel5A to construct the cellulase matrix Fncel5A-linker, which corresponds to the vector pFncel5A-linker (Fig. 2), because the characteristics of Fncel5A (28) can fulfill the criteria described above. Other suitable hydrolyses that satisfy these three criteria may also be used to construct this kind of cellulase matrix. Furthermore, this kind of strategy can also be developed to investigate non-hydrolytic disruptions of other compounds (e.g. insoluble hemicellulose and crystalline chitin) as long as suitable hydrolyases and enzyme matrices are selected.

Once the special cellulase matrix has been established, it will be a unified platform for quantitatively analyzing the disruptive activities of CBMs. For example, the cellulase matrix is Fncel5A-linker in our case. Five heterologous CBMs were conveniently joined in the cellulase matrix to form the chimeric cellulases, which have obvious activities toward crystalline cellulose. These chimeric cellulases are derived from the same
cellulase matrix, *i.e.* one unified platform, and differ only in the joined CBMs. Therefore, the differences among the hydrolytic activities of the whole cellulases toward crystalline cellulose reflect the disruptive characteristics of the CBMs.

The enzymatic properties of the five chimeric cellulases and *Fn*Cel5A toward CMC are roughly similar to each other, and the perturbations of the CBM moieties are slight (Fig. 3). Nevertheless, the activities of the chimeric cellulases against CMC were a little lower than that of *Fn*Cel5A. Initially, we thought that the random motions might result in the flexible linker and CBMs being too close to the entrance of the catalytic cleft within a specified probability; this would impede the free diffusion of polysaccharide chains in and out of the catalytic cleft. Moreover, the larger the CBM, the easier it is for the CBM surface to move close to the catalytic cleft, thus reducing the chances of hydrolyzing the amorphous cellulose chains. However, our MD simulations show that this kind of random motion near the catalytic cleft is very rare, and Table 1 shows that the activities toward CMC are not associated with the CBM sizes. We suggest that a CBM with a cellulose-binding module and the enzymatic properties of these chimeric cellulases against crystalline cellulose are significantly different from each other and are likely primarily determined by the CBM moieties. The non-hydrolytic disruption of crystalline cellulose by CBMs is the rate-determining step, which determines the measurable activities against Avicel. As these chimeric cellulases were derived from the unified cellulase matrix, the different profiles of the chimeric activities against Avicel reflect the properties of the linked CBMs, which have different sources and characteristics. If we use these chimeric cellulase activities to represent the disruptive functions of the CBMs and define 1 disruptive function unit as the amount of CBM that synergistically releases 1 μmol of reducing sugars/min via the synergism, the disruptive activities of the CBMs (listed in Table 1) are 1.55, 2.18, 1.82, 5.00, and 0.24 disruptive function unit, respectively. Quantitative investigations and comparisons of various CBMs with respect to their disruptive functions toward crystalline cellulose would therefore become feasible. Using this quantitative strategy, we can focus on studies of the disruptive functions of CBMs and design and identify potent CBMs to effectively convert crystalline cellulose by capitalizing on the synergism between CBMs and cellulases.

This type of cellulase matrix can be further developed and designed for more applications. For instance, various linker sequences can be substituted via pairs of restriction sites (Fig. 2b), or various linkers and CBMs can be connected in series or linked to the N terminus of the catalytic module. The cellulase matrix can also be used for the construction of novel chimeric cellulases for hydrolyzing crystalline cellulose and should prove valuable for enzyme engineering and design. Work is in progress to design a new type of cellulase that is efficient at destroying and hydrolyzing natural crystalline cellulose in biomass. Besides, these cellulases with high disruptive activities can also be used as additives or synergists to enhance the activities of existing industrial cellulase preparations or to decrease enzyme loads.

**REFERENCES**

1. Demain, A. L., Newcomb, M., and Wu, J. H. (2005) Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124–154
2. Hahn-Hägerdal, B., Galbe, M., Gorwa-Grauslund, M. F., Lidén, G., and Zacchi, G. (2006) Bio-ethanol—the fuel of tomorrow from the residues of today. *Trends Biotechnol.* 24, 549–556
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* 315, 804–807
4. Rubin, E. M. (2008) Genomics of cellulolytic biofuels. *Nature* 454, 841–845
5. United States Department of Energy (2006) *Breaking the Biological Barrier to Cellulosic Ethanol: A Joint Research Agenda*, Publication DOE/SC-0095, United States Department of Energy Office of Science, Washington, D. C.
6. Klemm, D., Heubel, B., Fink, H. P., and Bohn, A. (2005) Cellulose: fascinating biopolymer and sustainable raw material. *Angew. Chem. Int. Ed. Engl.* 44, 3358–3393
7. 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
8. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochim. J.* 280, 309–316
9. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active Enzymes Database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238
10. Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., Himmel, M., Keller, M., McMillan, J. D., Sheehan, J., and Wyman, C. E. (2008) How biotech can transform biofuels. *Nat. Biotechnol.* 26, 169–172
11. Zhang, Y. H., Himmel, M. E., and Mielzen, J. R. (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnol. Adv.* 24, 452–481
12. Arantes, V., and Saddler, J. N. (2010) Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels* 3, 4
13. Wilson, D. B. (2009) Cellulases and biofuels. *Curr. Opin. Biotechnol.* 20, 295–299
14. Cosgrove, D. J. (2000) Loosening of plant cell walls by expansins. *Nature* 407, 321–326
15. Kim, E. S., Lee, H. J., Bang, W. G., Choi, I. G., and Kim, K. H. (2009) Functional characterization of a bacterial expansin from *Bacillus subtilis* for enhanced enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.* 102, 1342–1353
16. Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyssönen, E., Bhatia, A., Ward, M., and Penttilä, M. (2002) Swollenin, a *Trichoderma reesei* protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulose materials. *Eur. J. Biochem.* 269, 4202–4211
17. Boraston, A. A., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781
18. Hashimoto, H. (2006) Recent structural studies of carbohydrate-binding modules. *Cell. Mol. Life Sci.* 63, 2954–2967
19. Shoeyev, O., Shani, Z., and Levy, I. (2006) Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol. Rev.* 70, 283–295
20. Doi, R. H., Kosugi, A., Murashima, K., Tamaru, Y., and Han, S. O. (2003) Cellulosomes from mesophilic bacteria. *J. Bacteriol.* 185, 5907–5914
21. Chen, X. A., Ishida, N., Todaka, N., Nakamura, R., Maruyama, J., Takahashi, H., and Kitamoto, K. (2010) Promotion of efficient saccharification of crystalline cellulose by *Aspergillus fumigatus* Sw1. *Appl. Environ. Microbiol.* 76, 2556–2561
22. Moser, F., Irwin, D., Chen, S., and Wilson, D. B. (2008) Regulation and characterization of *Thermobifida fusca* carbohydrate-binding module
Synergism between Non-hydrolytic Disruption and Hydrolysis

proteins E7 and E8. *Biotechnol. Bioeng.* **100**, 1066–1077

23. Carrard, G., Koivula, A., Söderlund, H., and Béguin, P. (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10342–10347

24. Fierobe, H. P., Bayer, E. A., Tardif, C., Czijek, M., Mechaly, A., Béláich, A., Lamed, R., Shoham, Y., and Béláich, J. P. (2002) Degradation of cellulose substrates by celluloseosome chimeras. Substrate targeting versus proximity of enzyme components. *J. Biol. Chem.* **277**, 49621–49630

25. Chhabra, S. R., and Kelly, R. M. (2002) Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carboxydrate binding module (CBM). *FEBS Lett.* **531**, 375–380

26. Bae, H. J., Turcotte, G., Chamberland, H., Karita, S., and Vézina, L. P. (2003) A comparative study between an endoglucanase IV and its fused protein complex Cel5-CBM6. *FEBS Microbiol. Lett.* **287**, 205–211

27. Mahadevan, S. A., Wi, S. G., Lee, D. S., and Bae, H. J. (2008) Site-directed mutagenesis and CBM4 engineering of Cel5A (*Thermotoga maritima*). *FEBS Microbiol. Lett.* **287**, 205–211

28. Wang, Y., Wang, X., Tang, R., Yu, S., Zheng, B., and Feng, Y. (2010) A novel thermostable cellulase from *Fervidobacterium nodosum*. *J. Mol. Catal. B Enzym.* **66**, 294–301

29. Schultz, J., Milpeltz, F., Bork, P., and Ponting, C. P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857–5864

30. Letunic, I., Doerks, T., and Bork, P. (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.* **37**, D229–D232

31. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGgettigan, P. A., McWilliam, H., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948

32. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704

33. Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinformatics* **9**, 299–306

34. Drummond, A. J., and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214

35. Weiner, M. P., and Costa, G. L. (1994) Rapid PCR site-directed mutagenesis. *PCR Methods Appl.* **4**, S131–S136

36. Zhang, S., Irwin, D. C., and Wilson, D. B. (2000) Site-directed mutation of *Thermobifida fusca* exocellulase Cel6B. *Eur. J. Biochem.* **267**, 3101–3115

37. Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by *Thermobifida fusca* Cel9A. *Appl. Environ. Microbiol.* **73**, 3165–3172

38. Miller, G. L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal. Chem.* **31**, 426–428

39. Simmerling, C., Stockbine, B., and Roitberg, A. E. (2002) All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* **124**, 11258–11259

40. Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R. C., Zhang, W., Merz, K. M., Roberts, B., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossvary, I., Wong, K. F., Paesani, F., Vanicek, J., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., and Kollman, P. A. (2010) AMBER 11. University of California, San Francisco

41. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, Z., Zhang, W., Yang, R., Cieplak, P., Luo, R., and Lee, T. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012

42. Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935

43. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28

44. Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A., and Adams, P. D. (2010) Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic *Thermotoga maritima*. *J. Struct. Biol.* **172**, 372–379

45. Mattinen, M. L., Kontteli, M., Kerouvo, J., Linder, M., Annila, A., Lindeberg, G., Reinikainen, T., and Drakenberg, T. (1997) Three-dimensional structures of three engineered cellulose-binding domains of celllobiohydrolase I from *Trichoderma reesei*. *Protein Sci.* **6**, 294–303

46. Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., and Harvey, T. S. (1995) Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993–7009

47. Guillon, D., Sanchez, S., and Rodriguez-Sanoja, R. (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* **85**, 1241–1249

48. Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304

49. Höcker, B., Jürgens, C., Wilmanns, M., and Sterner, R. (2001) Stability, catalytic versatility and evolution of the (βα)_8-barrel fold. *Curr. Opin. Biotechnol.* **12**, 376–381