PLOS | ONE

RESEARCH ARTICLE

Structural Insight of a Trimodular Halophilic Cellulase with a Family 46 Carbohydrate-Binding Module

Huaidong Zhang[2,3☯¤], Guimin Zhang[1,4☯], Chaoxiang Yao[2,3], Muhammad Junaid[2,3], Zhenghui Lu[4], Houjin Zhang[2,3]*, Yanhe Ma[1]*

1 State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, **2** Department of Biotechnology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei, China, **3** Key Laboratory of Molecular Biophysics, Ministry of Education, Wuhan, Hubei, China, **4** Hubei Collaborative Innovation Center for Green Transformation of Bio-Resources, College of Life Science, Hubei University, Wuhan, China

☯ These authors contributed equally to this work.
¤ Current address: CAS Key Laboratory of Special Pathogens and Biosafety, Center for Emerging Infectious Diseases, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei, China
* hjzhang@hust.edu.cn (HZ); mayanhe@im.ac.cn (YM)

CrossMark
click for updates

OPEN ACCESS

Citation: Zhang H, Zhang G, Yao C, Junaid M, Lu Z, Zhang H, et al. (2015) Structural Insight of a Trimodular Halophilic Cellulase with a Family 46 Carbohydrate-Binding Module. PLoS ONE 10(11): e0142107. doi:10.1371/journal.pone.0142107

Editor: Petri Kursula, Universitetet i Bergen, NORWAY

Received: May 31, 2015

Accepted: October 16, 2015

Published: November 12, 2015

Copyright: © 2015 Zhang et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the structural data are available from the RCSB Protein Data Bank (www.pdb.org) with the accession numbers 5E0C and 5E09.

Funding: This study was supported by National Key Basic Research Program of China (2013CB933900) and National Natural Science Foundation of China (31170068).

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Cellulases are the key enzymes used in the biofuel industry. A typical cellulase contains a catalytic domain connected to a carbohydrate-binding module (CBM) through a flexible linker. Here we report the structure of an atypical trimodular cellulase which harbors a catalytic domain, a CBM46 domain and a rigid CBM_X domain between them. The catalytic domain shows the features of GH5 family, while the CBM46 domain has a sandwich-like structure. The catalytic domain and the CBM46 domain form an extended substrate binding cleft, within which several tryptophan residues are well exposed. Mutagenesis assays indicate that these residues are essential for the enzymatic activities. Gel affinity electrophoresis shows that these tryptophan residues are involved in the polysaccharide substrate binding. Also, electrostatic potential analysis indicates that almost the entire solvent accessible surface of CelB is negatively charged, which is consistent with the halophilic nature of this enzyme.

Introduction

With the increasing energy cost, dwindling oil fuel reserve and ever-worsening problem of pollution, the search for a replacement for the fossil fuels has become an urgent task. Due to the abundant lignocellulosic substance in the biosphere, the production of biofuel with cellulose has emerged as a promising solution[1, 2]. The building blocks of cellulose are glucose molecules, which is a good raw material for fermentation. However, the cellulose consists of straight chain of glucose polymers. These polymers form rod-like structures which are strengthened by the multiple hydrogen-bonds between or within the polymers. In the cell wall of plant,

microfibers of cellulose crosslink with hemicellulose and lignin to form a resilient biopolymer matrix. The crystalline nature of this matrix makes it difficult to degrade cellulose into glucose units[3]. Various physical and chemical methods have been developed to release the sugar molecules from biomass, however, the bottleneck is the cost-effective and efficient enzymes for industrial-scale conversion of lignocellulose to fermentable sugars[4]. The natural degradation of cellulose is the result of a group of glycoside hydrolases (GHs) working in synergy. Although the exact mechanism of cellulose hydrolysis is difficult to establish due to the complexity of the substrate, several key steps are involved. Firstly, the cellulose fibers are cleaved into fragments by endoglucanases. The shortened cellulose is clipped by the cellobiohydrolases from the end. The resulting cellobiose is then hydrolyzed into glucose by beta-glucosidases[5]. In the industrial setting, similar group of enzymes are used together to digest cellulose. However, the enzymes are put into harsh environments such as high temperature, high salt and acidic/basic conditions. Significant engineering efforts have been made to improve the properties of natural enzymes in order to meet such requirements in the industrial applications[6]. Alternatively, a good source of these industrial enzymes can be found in the microorganisms living in extreme conditions[7]. For example, a thermo-stable cellulase CelDR with optimum temperature of 50 degree centigrade was found in a strain isolated from a hot spring[8]. In addition, the function-based mining of metagenomes from the soil samples in a cold desert resulted in the discovery of an acidic and cold-active cellulase[9]. In our previous study, a halophilic cellulase was identified from the genome library of *Bacillus* sp. BG-CS10, an alkaliphilic and halophilic *Bacillus* strain from a Tibetan salt lake. This cellulase, CelB, is thermo-stable, halophilic and pH-tolerant. CelB can utilize soluble cellulose derivatives, such as carboxylmethyl cellulose and konjac glucomannan, while it can not hydrolyze insoluble cellulose derivatives, such as microcrystalline and cellulose CM-52. Also it has only endoglucanase activity with no exoglucnasase activity detected[10]. Interestingly, its activities are increased 10 fold after addition of 2.5 M NaCl or 3M KCl, which is very rare among cellulases [10].

The two main groups of cellulases are exoglucanases and endoglucanases. A typical exoglucanase has a globular structure with an active site tunnel going across it. The tunnel is surrounded by several anti-parallel beta-sheets. Within the tunnel lies the conserved Trp patches which serve as the anchoring point for the substrates. To accommodate the substrate in the tightly packed tunnel, several loops are located around the tunnel and provide flexibility for the region[11, 12]. In contrast, the structure of a typical endoglucanase has a shallow cleft instead of a tunnel at the active site. Compared to the tunnel found in the exoglucanases, the active site cleft of a endoglucanase provides an easy access for the cellulosic fibers which is still in a tightly packed state[13]. Also in the tunnel are the catalytic residues including two acidic residues such as aspartic acid or glutamic acid[14]. One carboxylate residue serves as the nucleophile and the other serves as the catalytic base. Although the carboxylate base/nucleophile is the classical combination which is found at the active site of most GHs, some diversity can be found in the active site. In some cases, no nucleophile is found at the active site. Instead, the carbonyl oxygen of the substrate acts as the nucleophile and forms oxazoline intermediate[15]. In GH-6 cellulases, a proton transfer network replaces the role of carboxylate base[16]. In the glycosidases from clan GH-E, the nucleophile is a Tyr instead of Glu. In some cases, the catalytic residues consist of an Asp-His dyad. In a beta-N-acetylglucosaminidase from *Bacillus subtilis*, the normal dual-acidic catalytic residues are replaced with an aspartic acid residue and a histidine residue. In such configuration, the histidine functions as a proton donor while aspartic acid is a nucleophile[17]. Also, some cofactors such as phosphate and NAD can act as exogenous base or nucleophile[18, 19].

The catalytic domains of GHs can perform the hydrolysis of glycosidic bond by itself, presumably due to the presence of aromatic residues which can bind to the glucosic moiety

through CH-pi hydrogen bonds. However, many GHs possess carbohydrate-binding modules (CBMs) that connect to the catalytic domains through linker sequences. The CBMs facilitate the GHs to bind to the substrate, therefore, increases the catalytic efficiency. The CBMs are classified into 71 families based on the sequence similarities[20]. High-resolution structures are available for many CBM families. Based on the structural similarity, CBMs can be divided into several fold families[21]. The most common fold is the β-sandwich fold which comprises two β-strands on top of each other. The other fold families include β-trefoil fold, cellulose binding fold and hevein fold families. According to the architecture of the binding sites, the CBMs are classified into three types. The type A CBMs, with a flat binding surface, prefer to bind highly crystalline cellulose and chitin. The type B CBMs have a binding grooves or clefts which are the docking sites for extended carbohydrate polymer chains. In contrast, the type C CBMs bind optimally to the mono-, di- or tri-saccharides[21].

To understand the spacial arrangement and the functions of individual domains in CelB, we have solved the structure of the trimodular cellulase CelB. The structural analysis indicates that the three domains (catalytic, CBM_X, CBM46) form a tightly-packed L-shape structure. The catalytic domain and the CBM46 are located next to each other and an extended substrate-binding cleft is found between them. Inside of the cleft are several tryptophan residues which are proven to be important for substrate binding. Also, almost the entirely surface of CelB is negatively charged, which is a signature of halophilic proteins.

Materials and Methods

Construction of Expression Vectors

The coding sequence for mature CelB was amplified from the genome of *Bacillus* sp. BG-CS10, ligated into pET-28a(+) vector and sequenced for correction [10]. The pET-28a(+) vector with the *celB* gene was transformed into *Escherichia coli* strain BL21(DE3) Rosetta for protein expression. The single-point mutations were introduced by PCR-based site-directed mutagenesis method. The primers used are listed in S1 Table. Eighteen mutants were constructed including W37A, H103Q, H104Q, W107A, E149Q, W156A, H224Q, W229A, E271Q, W310A, H349A, H364A, H366A, W425A, H426A, W476A, Y484A, Y490A (S1 Table). The resulting plasmids were amplified in DH5 α. The mutations were confirmed by DNA sequencing.

Expression and Purification of CelB

The bacteria were grown overnight at 37°C in Luria Bertani (LB) broth with kanamycin (50 μg/mL) and chloramphenicol (100 μg/mL), and inoculated into LB broth with kanamycin (50 μg/mL) and chloramphenicol (100 μg/mL). The culture was incubated at 37°C, 200 $r \cdot min^{-1}$ in shaker flasks. When the optical density at 600 nm (OD_{600}) reached 0.6, IPTG was added into the medium up to 0.1mM. The protein CelB expression was induced at 16°C, 200 $r \cdot min^{-1}$ for 16h. The culture was collected by centrifugation at 6,000 g at 4°C for 10 min after induction and the pellet was resuspended in 20 mM Tris-HCl, pH 8.3, 500 mM KCl. The cells were lysed by ultrasonication. And the cell debris and the supernatant were separated by centrifugation at 4°C, 13,800 g for 30 min.

The supernatant was applied to Ni-NTA Resins (GE Healthcare) pre-equilibrated with 20 mM Tris-HCl, pH 8.3, 500 mM KCl and then washed with 20 mM Tris, pH 8.3, 500 mM KCl, 30 mM imidazole. After elution with 20 mM Tris-HCl, pH 8.3, 500 mM KCl, 200 mM imidazole, the His-tag was removed with thrombin (1 unit/mg, Sigma). The protein was dialyzed with 20 mM Tris-HCl, pH 8.3, 200 mM KCl and loaded to a Superdex 75 column (16/60, GE Healthcare) equilibrated with 20mM Tris-HCl, pH 8.3, 100 mM KCl [22]. Then 2 M ammonium sulfate was added into the protein solution. The mixture was loaded onto a hydrophobic

chromatography (Phenyl Sepharose6 Fast Flow, GE Healthcare) and eluted with 50 mM Tris-HCl, pH 7.0. The purity and molecular weight of the protein sample were analyzed by SDS-PAGE. The protein was concentrated by 10K, Amicon® Ultra-4 Centrifugal Filter Units, and the protein concentration was estimated spectrometerically with OD_{280} [23]. The selenomethionine-labeled CelB was expressed as described in the literature and purified as native protein[24]. In brief, 2 ml culture was used to inoculate 2L M9 medium supplemented with 50 μg/mL kanamycin, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and 5 g/L dextrose. The culture was grown at 37°C until OD_{600} reached 0.8. Then lysine, phenylalanine and threonine were added at 100 mg/L, leucine, isoleucine and valine were added at 50 mg/L, then L-selenomethionine was added at 40 mg/L. After incubation for 15 min, 0.1 mM IPTG was added to induce the protein expression and the incubation was continued for another 12 hours at 16°C.

Crystallization and Data Collection

Crystals of native CelB were grown by the sitting-drop vapor diffusion method. Briefly, 1 μL of protein (10 mg/mL) was mixed with the precipitant solution containing 2.0 M ammonium sulfate, 5% (v/v) 2-propanol. The crystals of CelB appeared after 7 days incubation at 289.15 K and reached their maximum size 10 days later. Selenomethionine-labeled CelB crystals were grown in the same manner, and the precipitant solution was optimized by adding 0.1 M potassium sodium tartrate tetrahydrate. A cryo-protectant solution was made by supplementing precipitant solution with 25% (v/v) glycerol (Sigma). The crystals were immersed in the cryo-protectant briefly before frozen in liquid nitrogen.

Diffraction data was collected at 100K from selenomethionine-labeled CelB crystals at the selenium absorption edge with a Quantum 315 CCD detector (ADSC) at a wavelength of 0.979197 Å on beam-line BL17U, Shanghai Synchrotron Radiation Facility (SSRF), Shanghai, China. The data set was indexed and integrated with XDS package [25] and scaled with Aimless [26]. Initial phases were obtained by single-wavelength anomalous diffraction (SAD) using the anomalous scattering from the selenomethionine incorporated at the methionine sites. All 13 selenium sites in the asymmetric unit were located with PHENIX package and the phase was refined to a figure of merit of 0.284[27]. The phase was further improved with DM program in the CCP4 package (figure of merit = 0.782)[26]. The resulting electron density map was of very good quality and side-chains of 80% residues could be automatically built with PHENIX[27]. The manual building was done with Coot[28] and the structure was refined with PHENIX after each building cycle[29]. Portion of the data (5%) was set aside to calculate free R factor, which was used to monitor the bias throughout the model building process[30]. The stereochemistry of the model was validated at the late stage of manual building with MolProbity[31]. Native data set was collected at 100 K with a Rigaku R-AXIS IV++ detector at Public Technology Service Platform, Wuhan Institute of Biotechnology, Wuhan 430074, China. The native structure was solved by molecular replacement with Phaser, using selenomethionine-labeled CelB structure as the search model[32]. Data parameters and refinement statistics are summarized in Table 1.

Enzymatic Activity Assay and Gel Affinity Electrophoresis

The enzymatic activity was evaluated by measuring the reducing sugars released from substrate carboxymethylcellulose (CMC). Purified enzymes (native or mutant) were added to a reaction mixture containing 1% CMC, 2.5M NaCl, phosphate buffer, pH 5.0. The mixture was incubated at 55°C for 30 min. And the amount of reducing sugars was measured by the dinitrosalicylic acid reagent (DNS) method.

Table 1. Data collection and refinement statistics.

	Native CelB	SeMet CelB
Data collection		
Space group	I 4 2 2	I 4 2 2
Cell dimensions		
a, b, c (Å)	120.14, 120.14, 205.33	120.99, 120.99, 205.00
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	65.45–2.35(2.43–2.35)	85.55–2.34(2.43–2.34)
Wavelengths(Å)	1.5418	0.9791
Rmerge[1]	0.10(0.66)	0.09(0.87)
Mean I/σ(I)	19.40(2.50)	11.60(2.02)
Completeness (%)	98.20 (95.50)	99.70(100.00)
Redundancy	7.70(5.60)	6.30(6.50)
Wilson B-factors ($Å^2$)	30.70	43.88
Anomalous completeness(%)		99.20(100.00)
Anomalous Redundancy		3.30(3.30)
FOM/DM FOM[2]		0.28/0.78
Refinement		
Resolution (Å)	65.45–2.35(2.43–2.35)	43.96–2.34(2.43–2.34)
No. of total reflections	262046(18271)	103130(6616)
No. of unique reflections	30948(2956)	32069(940)
R_{work} / R_{free}[3]	0.18/0.23	0.19/0.23
Number of atoms		
Protein	4312	4312
Water	461	196
Average B-factor ($Å^2$)		
Protein	31.00	48.40
Water	36.50	51.40
R.m.s deviations		
Bond lengths (Å)	0.007	0.006
Bond angles (°)	1.03	0.99
Ramachandran plot[4]		
Allowed region (%)	3.60	4.35
Favored region (%)	96.40	95.65
Outliers (%)	0.00	0.00
PDB ID	5E0C	5E09

[1]. Data in the parenthesis was calculated based on the highest resolution shell.

[2]. Mean figure of merit with or without density modification.

[3]. R-factor = (Σhkl||Fo|-|Fc||)/Σhkl|Fo| where Fo and Fc are the observed and calculated structure factors respectively. R_{free} was calculated with a randomly-selected 5% subset which was excluded from the refinement process.

[4]. Statistics of Ramachandran plot were calculated with MolProbity.

doi:10.1371/journal.pone.0142107.t001

Affinity gel electrophoresis was used to investigate the binding affinity between CelB mutants and soluble polysaccharides. Proteins were resolved on nondenaturating polyacrylamide gels 10% (w/v) containing 0.3 mg/ml of hydroxyethyl cellulose (HEC). Electrophoresis was carried out for 5h at room temperature.

Molecular Dynamics Simulation

Molecular dynamics simulation of CelB was carried out in high salt concentration using *pmemd.cuda* [33] module of Amber14 [34]. The protein was solvated in a rectangular box filled with TIP3P water molecules using tleap module of Amber14 [35]. A buffer distance of 12 Å was set between the protein edge and the box boundary in all directions. NaCl salt molecules were added to reach 1M concentration. Amber ff14SB force field was used to generate coordinate and topology files for the protein [36]. In order to remove bad contacts between solvent and protein, energy minimization was carried out in two steps. Firstly, the system was minimized keeping the protein fixed with harmonic constraint of a strength of 500 $kcal \cdot mol^{-1} \cdot Å^{-2}$. Secondly, the whole system was minimized without any constraint. The above each step was performed with the steepest descent minimization of 1000 steps followed by a conjugate gradient minimization of 1000 steps. The system was then heated to 300K in 2000 steps. Finally the system was simulated for 50 ns and the trajectory was saved after each 20 ps. The SHAKE algorithm was used for the covalent bonds involving hydrogen [37]. The Particle Mesh Ewald (PME) method was adopted to treat the long-range electrostatic interactions [38].

Results

The Overall Structure of CelB

The CelB was expressed in *E. coli* BL21(DE3) Rosetta, purified and subjected to crystallization screening. The crystals of CelB appeared after 7 days incubation at 289.15 K in the sitting-drop crystallization plates and reached their maximum size 10 days later. The CelB crystal belongs to the body-centered tetragonal system (I 4 2 2), with unit-cell parameters a = b = 120.14 Å, c = 205.33 Å. The data-collection and processing statistics are presented in Table 1.

The overall structure of CelB consists of a typical $(\beta/\alpha)_8$ TIM barrel catalytic domain, a CBM_X domain and a CBM46 domain (Fig 1). The substrate-binding sites are composed of a deep cleft across the catalytic domain. The cleft is formed by several loops and small α helices on top of the barrel. Deep inside of the cleft are the catalytic residues including Glu149 and Glu271. Also in the active sites are several well-exposed aromatic residues serving as the anchoring points for the polysaccharide molecules. The CBM_X domain is located right next to the catalytic domain. The connection between these two domains is established through hydrogen bonds between Arg240/Glu344, Lys219/Glu344, Gln285/Asn346, and the N-H/pi interaction between Arg339/His366. Unlike most CelB with CBM attached, there is no linker present between CBM_X domain and catalytic domain. There is a loop (Gly338-Leu347) at the N-terminus of CBM_X domain. However, this loop is tightly tethered to the CBM_X domain as well as the catalytic domain. Therefore, it does not function as a conventional loop. Also tightly connected is the CBM46 in the molecule. There is no linker between the CBM_X and CBM46 domains either. The CBM46 domain forms hydrogen bonds with both the CBM_X domain and catalytic domain. As a result, instead of forming an extended chain, these three domains form a tightly packed L-shaped structure. More importantly, the aromatic residues on the CBM46 (Phe470, Trp476) are close to the catalytic cleft and form an extended binding site for the substrate.

The Catalytic Center and Substrate-binding Cleft

The catalytic domain of CelB (residues 1–340) adopts a typical TIM barrel fold that is found in many GHs of GH5 family. In addition to the canonical $(\beta/\alpha)_8$ fold, some additional structural elements are also found in the catalytic domain, such as two 3/10 helices (residues 275–278 and 281–283) as well as an extended loop (residues 145-159) between beta strand 4 and alpha

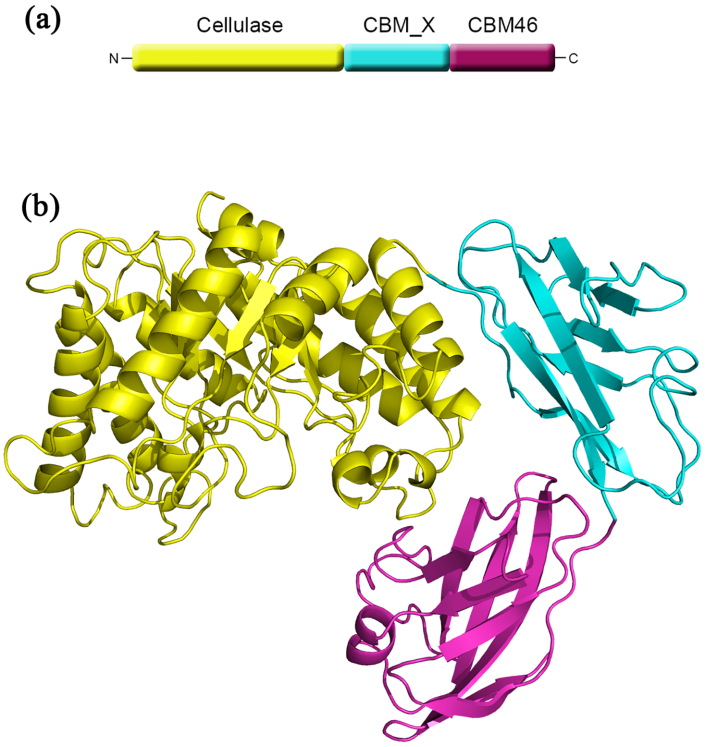

Fig 1. Overall structure of CelB. (a) The domain arrangement in the primary structure of CelB. The cellulase catalytic domain, colored in yellow, is located at the N-terminus, the CBM46 domain, colored in red, is located at the C-terminus and the CBM_X domain is located in the middle portion of the protein. (b) The cartoon representation of CelB overall structure. The catalytic domain is in yellow, the CBM46 domain is in red and the CBM_X domain is in cyan. The catalytic domain and the CBM46 domain form an extended substrate-binding cleft. CBM_X domain serves as the rigid connection between the first and the third domains.

doi:10.1371/journal.pone.0142107.g001

helix 6 (S1 Fig). These elements are located at the entrance of the TIM barrel and significantly deepen the substrate-binding cleft. The Dali secondary structure comparison performed with the catalytic domain shows that it is homologous to various endoglucanases[39]. The closest structural homolog is the endoglucanase D (engD) from *Clostridium cellulovorans*, with a RMSD of 2.1 Å. Several tryptophan residues (Trp37, Trp107, Trp156, Trp229, Trp310 and Trp476) are located on the surface of the substrate-binding cleft (Fig 2). These aromatic residues presumably serve as the attaching points for the polysaccharide substrate. Except for Trp107 and Trp476, these aromatic residues occupy similar positions as those in the active site of engD, which implies similar substrate binding patterns of these two cellulases. Actually, Trp107 is conserved in other GH5s [40]. In engD, Trp162 and Tyr232 form a clamp that encloses the substrate-binding cleft [41]. However, in CelB, due to the shift of a long loop, such clamp is not present, which results in a more open active site. The superposition of CelB and engD structures indicates that the Trp37 and Trp310 of CelB overlap with the corresponding residues in engD. These two tryptophan residues form the -3 and -2 sub-sites for the substrate. The His103 and Asn148 of CelB also overlap with their counterparts in engD, which form the -1 sub-site.

At the bottom of the active site are the catalytic Glu149 and Glu271 that are present in most of GHs. In the vicinity of these two catalytic residues is the His224 that is conserved in several GH5 family GHs[13, 41] (Fig 3). Other conserved residues close to the catalytic residues are Arg59, His103 and Asn148. The His103 and Asn148 residues are likely to function as binding

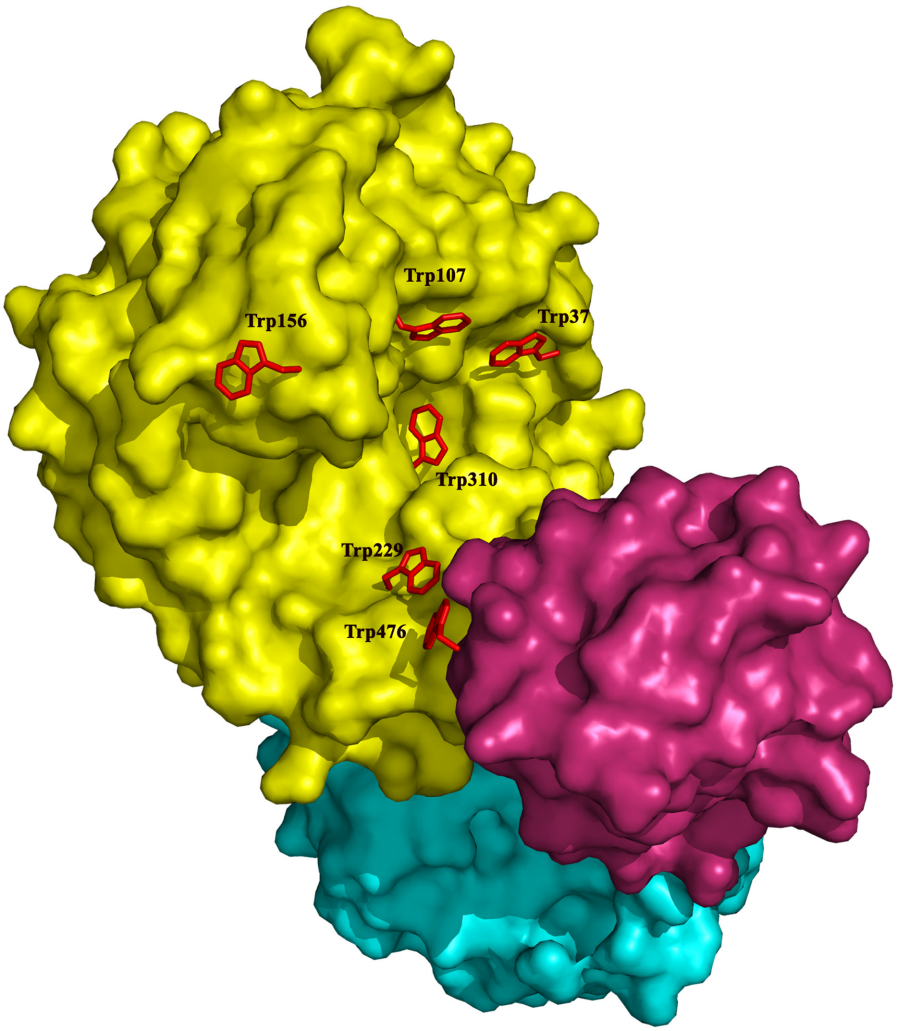

Fig 2. Surface representation of CelB. The catalytic domain is colored in yellow, the CBM_X domain is colored in cyan and the CBM46 domain is colored in red. The tryptophan residues around the substrate binding cleft are shown in sticks and colored in red. The rest of molecule is shown as surface.

doi:10.1371/journal.pone.0142107.g002

sites for the substrate through H-bonds. The Arg59 residue forms several H-bonds with the surrounding residues, including Asn22, Glu145, Asn148, and Glu271. It is possible that Arg59 serves as a pillar to maintain the active site structure.

The Carbohydrate Binding Modules

A distinctive structural feature of CelB is the CBM_X domain and the CBM46 domain at the C-terminus (Fig 1). The CBMs usually play supplementary roles in the cellulase functions. The absence of CBM only has modest effect on the enzyme activities. However, the truncation experiments indicate that these CBMs are essential for CelB's function (Fig 4). The deletion of either CBM abolishes the enzyme activities. Sequence analysis shows that the first putative CBM, CBM_X, belongs to an uncharacterized CBM family, while the second putative CBM belongs to CBM family 46 [20]. Unlike the typical CBM, the CBM_X does not form a beta-sheet sandwich. Instead, it has a continuous beta-sheet and forms a barrel-like structure. The

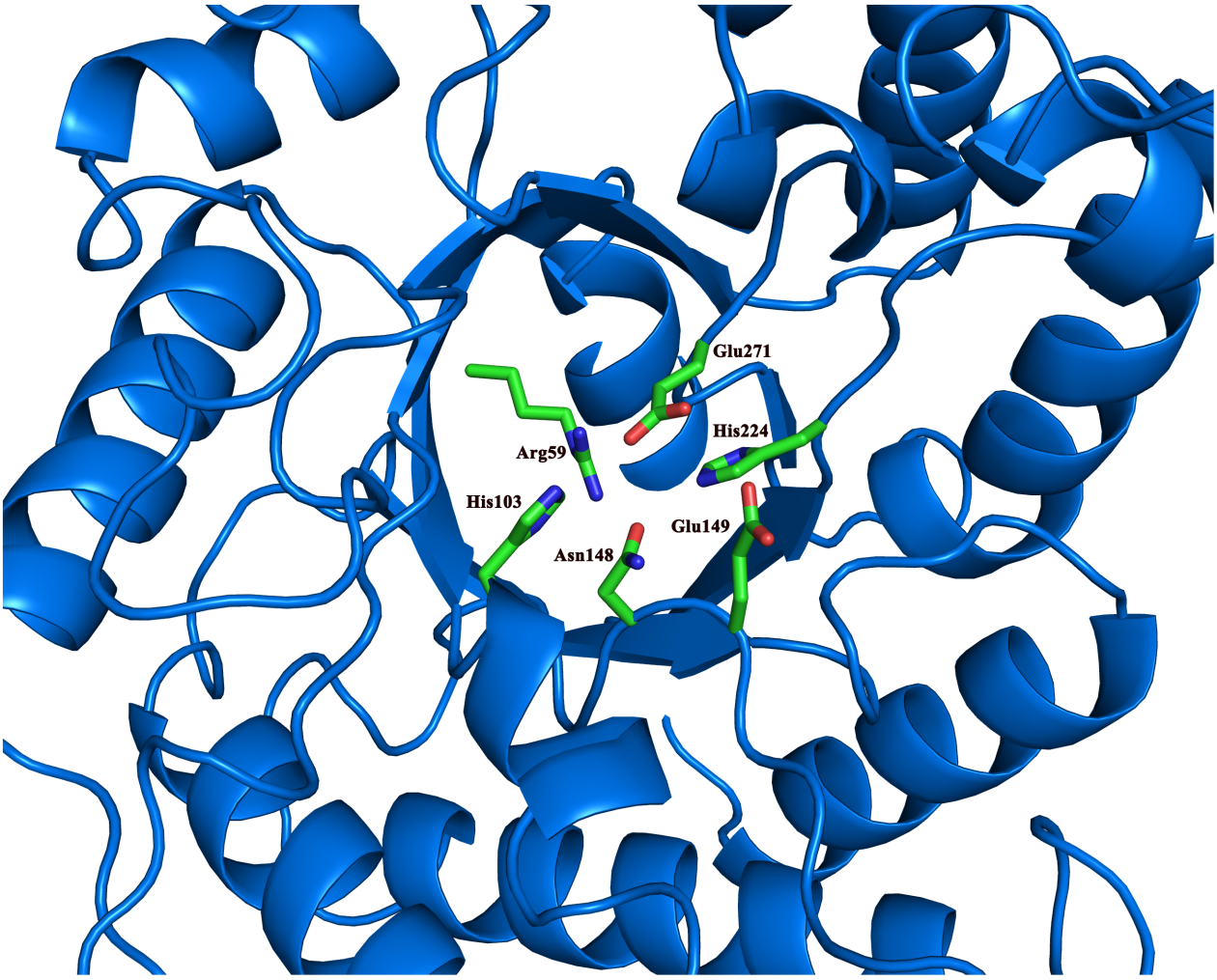

Fig 3. The conserved residues in the catalytic domain. The conserved residues around the two catalytic Glu residues are shown in sticks and colored by elements. The rest of the catalytic domain is shown in ribbons and colored in blue.

doi:10.1371/journal.pone.0142107.g003

Dali secondary structure comparison shows that its closest structural homolog is the neural cell adhesion molecule 2 which has an Ig-like fold[39].

In contrast, the second CBM domain, CBM46, possesses multiple tryptophan, tyrosine and phenylalanine residues on the surface, which is a common feature for CBMs. The mutagenesis experiments indicate that the removal of the aromatic side-chains diminishes the enzymatic activities (Fig 4). Given the locations of these residues, it suggests that these aromatic residues are the docking sites for the carbohydrate substrates. The structure of CBM46 consists of a large anti-parallel beta-sheet rolled into a barrel. It also possesses one alpha helix and several long loops connecting the beta-strands. The structure is stabilized by a network of hydrogen bonds. Also stacking interaction is found between Phe453 and Tyr490, which contributes to the stability of the structure. Although the Blast search does not assign any specific function for this domain, in the carbohydrate-active enzymes database (CAZY), this region of the CelB is classified as the CBM46 group member [20]. As indicated by CAZY, the CBM46 domain is often found in multiple-domain cellulases, which matches the characteristics of CelB. It noteworthy that neither CBM_X domain nor CBM46 has the DxDxDG calcium-binding motif and no

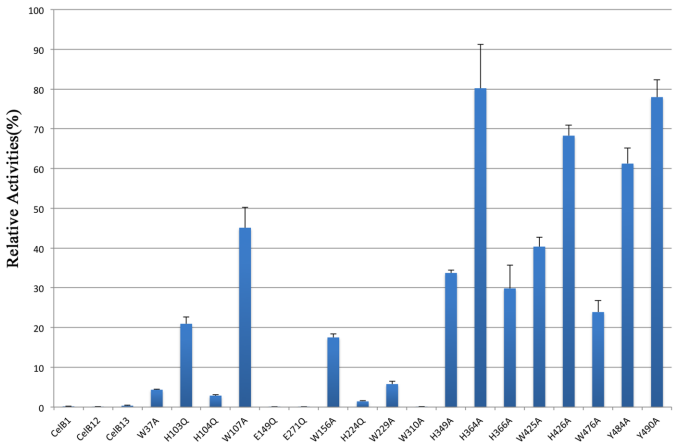

Fig 4. Histogram representation of relative activities of CelB mutants. The activities of mutants were represented as the percentage of native enzyme activity. The data is the average of at least three individual experiments. The error bars show the standard deviation (S.D.) of different repeats of the same assay. CelB1 designates the catalytic domain alone. The CelB12 designates the truncated protein without the CBM46 domain. The CelB13 designates the fusion protein containing the catalytic domain and the CBM46 domain.

doi:10.1371/journal.pone.0142107.g004

calcium is found in the structure of CelB. This is also different from many classic CBMs among which calcium plays an important role in structural stability and substrate recognition[42–44].

Structural Basis for the Salt Resistance of CelB

Electrostatic potential analysis indicates that almost the entire solvent accessible surface of CelB is negatively charged (Fig 5). Some small positive patches are found at two groups of arginine residues including Arg41/Arg44/Arg83, Arg536/Arg537 and a group of charged residues including Lys21/His208. The structural analysis of a close homolog, BhCel5B from *Bacillus Haloduran* C-125, shows that its surface is also negatively charged, indicating it is a common feature for this group of cellulase extracted from high-salt environment[45]. The predominantly negative charge is attributed to the large amount of acidic residues present on the protein surface. Sequence analysis shows that the acidic residues (Asp and Glu) account for 16.3% of the total residues of CelB. A direct consequence of exposed acidic residues is the low isoelectric point (pI) of CelB. The predicted pI of CelB is only 4.53 as predicted by ProtParam[46]. This is lower than other GH5 cellulases of the similar size and modular composition. The negatively charged residues on the surface and low pI should be the result of high-concentration of salt in the environment it exists[10]. Actually, the highly acidic surface is the hallmark of many salt-resistant proteins[11]. Like CelB, some of these GHs are not only resistant to salt, but also activated by the increased salt concentrations, which makes them promising candidates for biocatalyst used in extreme conditions [11, 47, 48].

Site-directed Mutagenesis and Truncation of CelB

Site-directed mutagenesis and truncation were used to investigate the importance of key residues in the catalytic domain as well as two CBM domains. The residues close to the catalytic Glu residues were mutated to Gln instead of Ala, in order to keep the length of the side-chains approximately the same. The mutants were used to carry out enzymatic assays in the same experimental setting as the native enzyme. The relative activities of the mutants were calculated as the percentage of the native enzyme activity (Fig 4). The CelB1 mutant, which only contains the catalytic domain, has no activity. The CelB12 mutant, which is the truncated protein

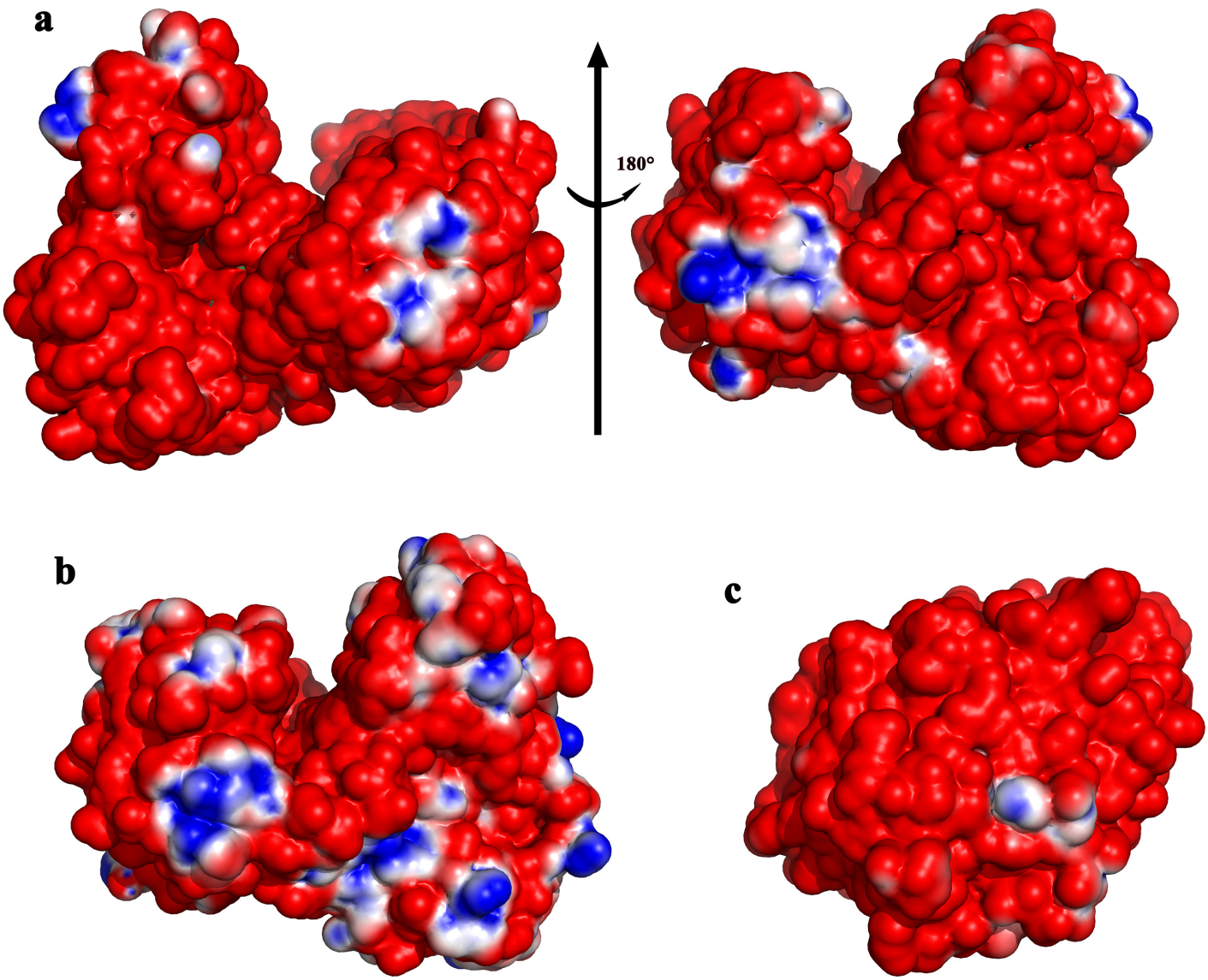

Fig 5. Surface electrostatic potential of CelB and other haloduric cellulases. The electrostatic potential between -1 kT/e and 1 kT/e is shown as a colored gradient from red (acidic) to blue (basic). (a) The left view is rotated 180 degrees from the right view. Almost all the surface of CelB has negative electrostatic potential. Only small patches of positive electrostatic potential can be found. (b) The surface electrostatic potential of Cel5B from *Bacillus halodurans*. (c) The surface electrostatic potential of LqCel7B from *Limnoria quadripunctata*.

doi:10.1371/journal.pone.0142107.g005

without CBM46 domain, does not have any activity. The fusion protein of catalytic domain and the CBM46 (designated as CelB13) does not have any activity either. Therefore, the tri-modular CelB possess a long active-site cleft that is formed by all three domains. The absence of any domain abolishes the enzyme activity. It is thermodynamically unfavorable to have the bulky hydrophobic tryptophan residues exposed to the solvent. In the catalytic domains of cellulases, such exposed tryptophan residues are often observed, which are proven to serve as the binding sites for the polysaccharide. The tryptophan residues in the catalytic domain were systematically mutated to alanine to investigate the functions of their bulky side chains (W37A, W107A, W156A, and W229A). As expected, the subsequent enzymatic assay indicates that these tryptophan residues are essential to the enzyme activities. The activities of these mutants

are significantly reduced, compared to the native enzyme. In particular, the W37A and W229A mutants only have less than 10% of the original activity (Fig 4).

The histidine residues around the catalytic residues are also mutated (H103Q, H104Q, H224Q). In the canonical cellulase catalytic mechanism, the histidine residues are not included. However, a recent study on another GH5 cellulase has shown that the mutations on His residues close to the catalytic residues can deactivate the enzyme[49]. Histidine residues may serve as a part of proton transfer network, or part of the hydrogen bonding network which maintains the stability of the active site. In GH subfamily GH5_36, arginine instead of histidine is conserved at this position. The replacement of arginine with histidine preserves part of the enzyme activity, suggesting arginine and histidine serve similar roles in the proton transfer network [50]. His103 and His104 residues are well exposed in the active site. They may bind to the substrate through hydrogen bonds. They could also shuffle the protons into the surround solvent. However, H224 is partially shielded. Early research regarding hyperthermophilic endo-β-1,4-glucanase has suggested the two glutamic acid residues and the histidine residue right next to it may form a catalytic triad. But the exact role of this histidine in CelB is yet to be determined.

CBM46 domain is a member of CBM family 46. Like the other CBM members, this domain has several aromatic residues well exposed to the solvent. Mutagenesis assays were used to investigate functions of these aromatic residues (W476A, Y484A, Y490A). The W476A and Y484A mutations, which are in the grove between the catalytic and CBM46 domains, have larger impact on the enzymatic activity than the Y490A mutation which is outside of the grove. To confirm the importance of tryptophan residues in the binding of polysaccharide substrate, the CelB mutants (H103Q, W156A, W229A, W476A) were used to perform the affinity gel electrophoresis assay. The results clearly indicate that the affinity of W476A and W229A mutants to HEC is much lower than the native protein. As a result, these two mutants migrate much faster in the affinity gel, which indicates W229A and W476A play important roles in substrate binding. Likewise, the W156A mutant also migrates faster than the native protein, but to a much less extent. In contrast, the H103Q mutation does not change the migration speed of CelB (Fig 6). To investigate the importance of CBM_X domain, several aromatic residues in this domain were mutated. The subsequent enzymatic assay indicates that H349A, H366A and W425A have some impact on the enzyme activities. His349 and His366 are exposed on the surface of the protein and they may be involved in the binding of the substrate. Although they are located far from catalytic residues, the polymer nature of the substrate CMC makes it possible to interact with the residues distal from the active site. The reason why W425A mutation is detrimental is unclear.

Molecular Dynamics Simulation of CelB

In the classic cellulase system, the CBM domain is connected the catalytic domain by a flexible linker. The CBM domain binds to the cellulosic substrate first. Then the catalytic domain adjusts its position with the flexible linker and find the best orientation on the substrate. In the case of CelB, the CBM46 and catalytic domains are not only bridged by a rigid CBM_X domain but also contact each other directly through hydrogen bonds. To investigate the flexibility of the CelB, we performed molecular dynamics (MD) simulation. The result indicates that the overall folding of CelB is very stable throughout the simulation, indicating the CelB is unlikely to bind the substrate through the classic model (Fig 7). However, we notice that the aromatic residues, which are proven to be involved in the substrate binding, are mostly located in the loop regions. These regions undergo structural change during the simulation, which may provide the flexibility required for substrate binding and subsequent product release (Fig 7).

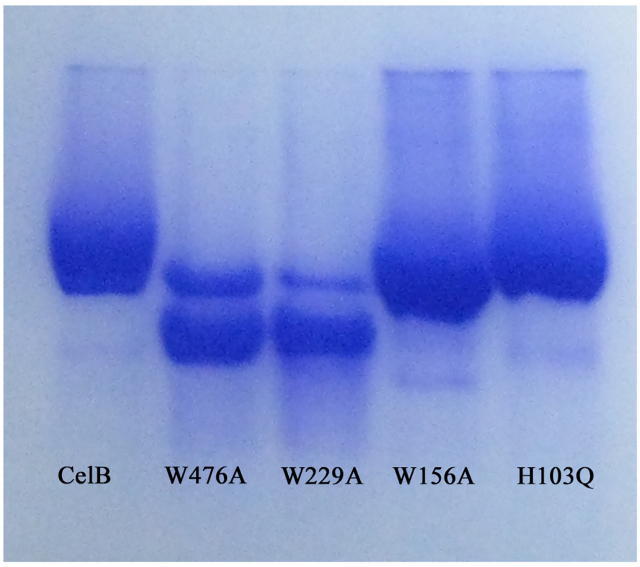

Fig 6. Affinity gel electrophoresis of CelB mutants. The CelB mutants were electrophoresed on a 10% (w/v) nondenaturating polyacrylamide gel containing 0.3 mg/ml of HEC. The native CelB protein was used as the control.

doi:10.1371/journal.pone.0142107.g006

Discussion

In many cellulases, extended linker sequences are found between the catalytic domains and the CBMs[40, 51]. It was suggested that the linkers help to recruit catalytic domains to the cellulosic substrates after the CBMs bind to the substrates. The presence of CBMs and the unstructured linker sequences help to anchor the catalytic domains on the substrates, therefore, boosts the catalytic rates[52]. The mechanism of the linkers implies that the linker sequences are most useful when the cellulases are used to digest insoluble substrate such as microcrystalline cellulose[53]. The rigid nature of the insoluble substrates requires the cellulases to stay flexible in order to fit the surfaces of the substrates. However, for CelB, all its substrates are soluble[10]. The soluble substrates have the abilities to adjust their positions in solution so that they can fit in the active site of CelB. As indicated by the MD simulation, the overall structure of CelB stays unchanged over the 50ns simulation process. The rigid nature of CelB indicates that it uses a different mechanism to anchor the enzyme to the substrate molecule. In order to catch the substrate with higher efficiency, CelB expands its substrate-binding site beyond the catalytic domain. The binding cleft spanning two domains provides much bigger binding surface and more aromatic residues for the substrate to recognize. This mechanism helps CelB to catch the soluble substrate even if it does not have the flexible linkers.

The CelB structure represents a rare example in cellulases that the catalytic domain, a CBM_X domain and a CBM46 domain form a tightly-packed L-shaped structure. Although rare in cellulases, such tight arrangement could be found in other GHs such as dextrinase and cyclodextrin glycosyltransferase, which are both starch-converting enzymes[54, 55]. Like the structure of CelB, the CBMs in these GHs are arranged around the catalytic domain. It is possible that the putative CBM(CBM_X) serves both as a spacer as well as the substrate-binding site. As a result, the CBM at the C-terminus is positioned at the opening of catalytic domain and forms part of the long substrate-binding cleft. CelB and the starch-converting enzymes may share the similar mechanism as far as substrate recruitment.

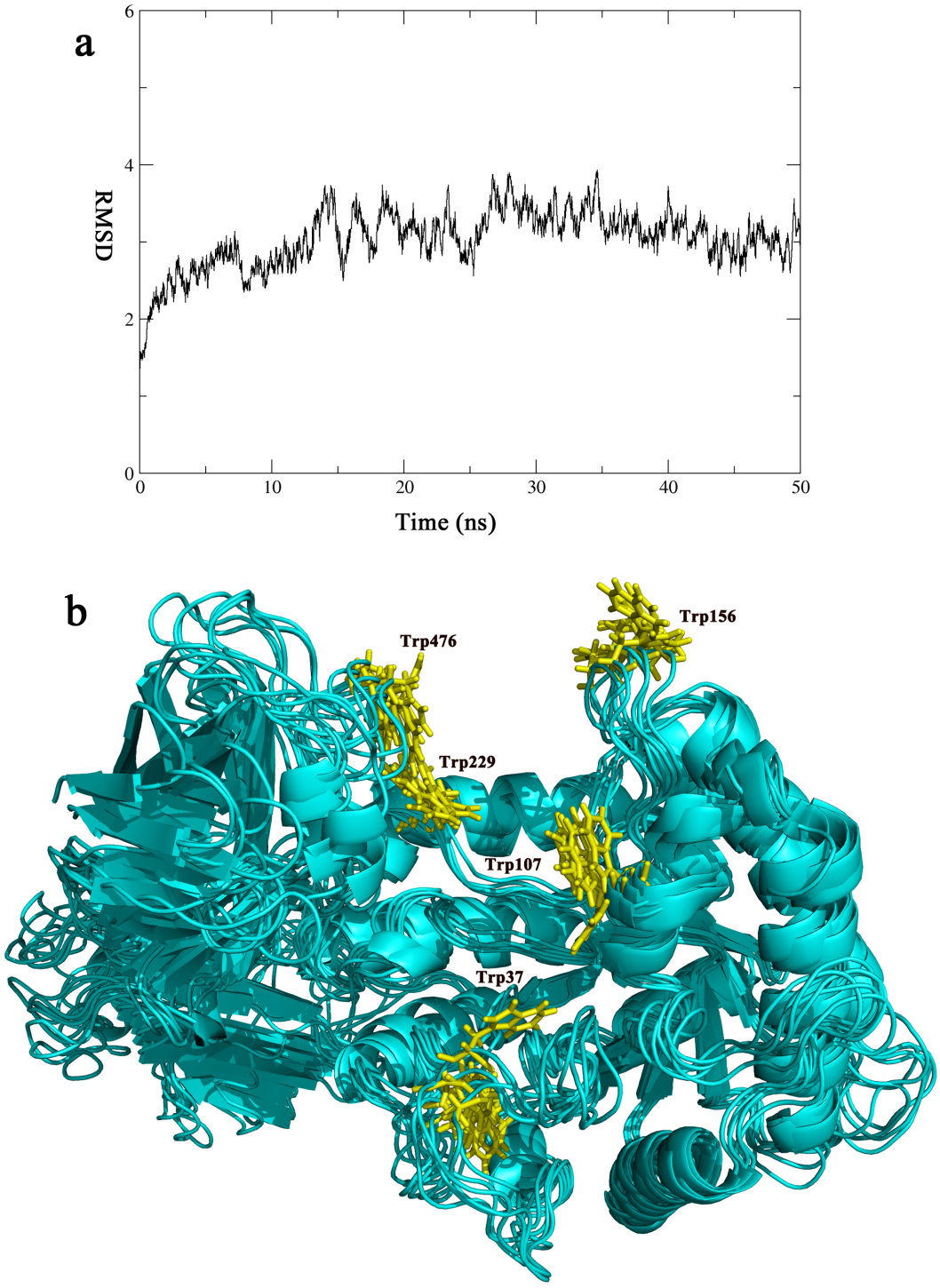

Fig 7. The molecular dynamics simulation of CelB. The CelB structure is simulated over a period of 50 ns. (a) The RMSD Plot. The RMSD of the simulation was plotted against time, according to the simulation trajectories. (b) The cluster representation of CelB structure. The snapshots of the trajectory were taken every 10 ns along the simulation process. The snapshots were overlapped with each other. Only small fluctuation is found and they are mainly concentrated in the loop areas. The tryptophan residues (Trp37, Trp107, Trp156, Trp229, Trp476) are colored in yellow and shown in sticks.

doi:10.1371/journal.pone.0142107.g007

The CelB is a halophilic cellulase. Its activities increase 10 fold in the presence of 2.5M NaCl [10]. Such property has been observed with other endoglucanases, but to a much less extent [56, 57]. No significant structure change was observed when salt concentration was increased, which rules out the possibility of refolding of the protein under high salt[10]. The surface of CelB is highly negatively charged (Fig 5). However, there are several small positive charged patches present. It is possible that under low salt conditions, these positively charged patches serve as an attaching point for the neighboring highly negatively charged CelB molecules. And these molecules may form oligomer through electrostatic interactions. In fact, as observed by dynamic light scattering, the diameters of particles in CelB solution increase significantly under the low salt conditions (data not shown). The formation of such oligomers would prevent the substrate binding and, therefore, hinder the catalytic activities.

Supporting Information

S1 Fig. The sequence alignment of several cellulases with a GH5 catalytic domain and a CBM46 domain.
(DOCX)

S2 Fig. The limited proteolysis of CelB mutants.
(DOCX)

S1 Table. Primers for *celB* gene amplification and site-directed mutagenesis.
(DOCX)

Acknowledgments

We would like to thank Dr. Defeng Li at Institute of Biophysics, Chinese Academy of Sciences for his help in the data processing. Also we would like to thank Dr. Zhongzheng Yang at Public Technology Service Platform, Wuhan Institute of Biotechnology, for his assistance in data collection. We are grateful for Dr. Junjun Liu for his helpful discussion regarding molecular dynamics simulation.

Author Contributions

Conceived and designed the experiments: Houjin Zhang YM. Performed the experiments: Huaidong Zhang GZ CY MJ. Analyzed the data: Huaidong Zhang Houjin Zhang GZ. Contributed reagents/materials/analysis tools: ZL. Wrote the paper: Huaidong Zhang Houjin Zhang MJ.

References

1. Carroll A, Somerville C. Cellulosic biofuels. Annual review of plant biology. 2009; 60:165–82. Epub 2008/11/19. doi: 10.1146/annurev.arplant.043008.092125 PMID: 19014348.
2. Jung SK, Parisutham V, Jeong SH, Lee SK. Heterologous expression of plant cell wall degrading enzymes for effective production of cellulosic biofuels. Journal of biomedicine & biotechnology. 2012; 2012:405842. Epub 2012/08/23. doi: 10.1155/2012/405842 PMID: 22911272; PubMed Central PMCID: PMC3403577.
3. Harris D, DeBolt S. Synthesis, regulation and utilization of lignocellulosic biomass. Plant biotechnology journal. 2010; 8(3):244–62. Epub 2010/01/15. doi: 10.1111/j.1467-7652.2009.00481.x PMID: 20070874.
4. Hendriks AT, Zeeman G. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresource technology. 2009; 100(1):10–8. Epub 2008/07/05. doi: 10.1016/j.biortech.2008.05.027 PMID: 18599291.
5. Bornscheuer U, Buchholz K, Seibel J. Enzymatic Degradation of (Ligno)cellulose. Angewandte Chemie. 2014. doi: 10.1002/anie.201309953 PMID: 25136976.

6. Bommarius AS, Sohn M, Kang Y, Lee JH, Realff MJ. Protein engineering of cellulases. Current opinion in biotechnology. 2014; 29C:139–45. doi: 10.1016/j.copbio.2014.04.007 PMID: 24794535.
7. van den Burg B. Extremophiles as a source for novel enzymes. Current opinion in microbiology. 2003; 6(3):213–8. PMID: 12831896.
8. Li W, Zhang WW, Yang MM, Chen YL. Cloning of the thermostable cellulase gene from newly isolated *Bacillus subtilis* and its expression in *Escherichia coli*. Molecular biotechnology. 2008; 40(2):195–201. Epub 2008/06/26. doi: 10.1007/s12033-008-9079-y PMID: 18576142.
9. Bhat A, Riyaz-Ul-Hassan S, Ahmad N, Srivastava N, Johri S. Isolation of cold-active, acidic endocellulase from Ladakh soil by functional metagenomics. Extremophiles: life under extreme conditions. 2013; 17(2):229–39. Epub 2013/01/29. doi: 10.1007/s00792-012-0510-8 PMID: 23354361.
10. Zhang G, Li S, Xue Y, Mao L, Ma Y. Effects of salts on activity of halophilic cellulase with glucomannanase activity isolated from alkaliphilic and halophilic *Bacillus* sp. BG-CS10. Extremophiles: life under extreme conditions. 2012; 16(1):35–43. doi: 10.1007/s00792-011-0403-2 PMID: 22012583.
11. Kern M, McGeehan JE, Streeter SD, Martin RN, Besser K, Elias L, et al. Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(25):10189–94. Epub 2013/06/05. doi: 10.1073/pnas.1301502110 PMID: 23733951; PubMed Central PMCID: PMC3690837.
12. Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engstrom A, et al. Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus *Heterobasidion irregulare*. The Journal of biological chemistry. 2013; 288 (8):5861–72. Epub 2013/01/11. doi: 10.1074/jbc.M112.440891 PMID: 23303184; PubMed Central PMCID: PMC3581431.
13. Alvarez TM, Paiva JH, Ruiz DM, Cairo JP, Pereira IO, Paixao DA, et al. Structure and function of a novel cellulase 5 from sugarcane soil metagenome. PloS one. 2013; 8(12):e83635. doi: 10.1371/journal.pone.0083635 PMID: 24358302; PubMed Central PMCID: PMC3866126.
14. Fushinobu S, Alves VD, Coutinho PM. Multiple rewards from a treasure trove of novel glycoside hydrolase and polysaccharide lyase structures: new folds, mechanistic details, and evolutionary relationships. Current opinion in structural biology. 2013; 23(5):652–9. Epub 2013/07/03. doi: 10.1016/j.sbi.2013.06.001 PMID: 23816329.
15. Vuong TV, Wilson DB. Glycoside hydrolases: catalytic base/nucleophile diversity. Biotechnology and bioengineering. 2010; 107(2):195–205. Epub 2010/06/17. doi: 10.1002/bit.22838 PMID: 20552664.
16. Vuong TV, Wilson DB. The absence of an identifiable single catalytic base residue in *Thermobifida fusca* exocellulase Cel6B. The FEBS journal. 2009; 276(14):3837–45. doi: 10.1111/j.1742-4658.2009.07097.x PMID: 19523117.
17. Litzinger S, Fischer S, Polzer P, Diederichs K, Welte W, Mayer C. Structural and kinetic analysis of *Bacillus subtilis* N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism. The Journal of biological chemistry. 2010; 285(46):35675–84. doi: 10.1074/jbc.M110.131037 PMID: 20826810; PubMed Central PMCID: PMC2975192.
18. Hidaka M, Kitaoka M, Hayashi K, Wakagi T, Shoun H, Fushinobu S. Structural dissection of the reaction mechanism of cellobiose phosphorylase. The Biochemical journal. 2006; 398(1):37–43. doi: 10.1042/BJ20060274 PMID: 16646954; PubMed Central PMCID: PMC1525018.
19. Yip VL, Thompson J, Withers SG. Mechanism of GlvA from *Bacillus subtilis*: a detailed kinetic analysis of a 6-phospho-alpha-glucosidase from glycoside hydrolase family 4. Biochemistry. 2007; 46 (34):9840–52. doi: 10.1021/bi700536p PMID: 17676871.
20. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2014; 42(Database issue):D490–5. doi: 10.1093/nar/gkt1178 PMID: 24270786; PubMed Central PMCID: PMC3965031.
21. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. The Biochemical journal. 2004; 382(Pt 3):769–81. Epub 2004/06/25. doi: 10.1042/bj20040892 PMID: 15214846; PubMed Central PMCID: PMC1133952.
22. Zhang H, Chen J, Wang H, Xie Y, Ju J, Yan Y, et al. Structural analysis of HmtT and HmtN involved in the tailoring steps of himastatin biosynthesis. FEBS letters. 2013; 587(11):1675–80. Epub 2013/04/25. doi: 10.1016/j.febslet.2013.04.013 PMID: 23611984.
23. Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. Analytical biochemistry. 1989; 182(2):319–26. Epub 1989/11/01. PMID: 2610349.
24. Doublie S. Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems. Methods in molecular biology (Clifton, NJ). 2007; 363:91–108. Epub 2007/02/03. doi: 10.1007/978-1-59745-209-0_5 PMID: 17272838.

25. Wolfgang K. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–32. doi: 10.1107/S0907444909047337 PMID: 20124692

26. Project CC. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50(Pt 5):760–3. Epub 1994/09/01. doi: 10.1107/s0907444994003112 PMID: 15299374.

27. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66(Pt 2):213–21. Epub 2010/02/04. doi: 10.1107/s0907444909052925 PMID: 20124702; PubMed Central PMCID: PMC2815670.

28. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60(Pt 12 Pt 1):2126–32. Epub 2004/12/02. doi: 10.1107/s0907444904019158 PMID: 15572765.

29. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012; 68(Pt 4):352–67. Epub 2012/04/17. doi: 10.1107/s0907444912001308 PMID: 22505256; PubMed Central PMCID: PMC3322595.

30. Kleywegt GJ, Brunger AT. Checking your imagination: applications of the free R value. Structure (London, England: 1993). 1996; 4(8):897–904. Epub 1996/08/15. PMID: 8805582.

31. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr. 2010; 66(Pt 1):12–21. Epub 2010/01/09. doi: 10.1107/s0907444909042073 PMID: 20057044; PubMed Central PMCID: PMC2803126.

32. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Echols N, Headd JJ, et al. The Phenix software for automated determination of macromolecular structures. Methods (San Diego, Calif). 2011; 55(1):94–106. Epub 2011/08/09. doi: 10.1016/j.ymeth.2011.07.005 PMID: 21821126; PubMed Central PMCID: PMC3193589.

33. Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. Journal of chemical theory and computation. 2012; 8(5):1542–55. doi: 10.1021/ct200909j PMID: 22582031; PubMed Central PMCID: PMC3348677.

34. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr., et al. The Amber biomolecular simulation programs. Journal of computational chemistry. 2005; 26(16):1668–88. doi: 10.1002/jcc.20290 PMID: 16200636; PubMed Central PMCID: PMC1989667.

35. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. The Journal of chemical physics. 1983; 79(2):926–35.

36. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Structure, Function, and Bioinformatics. 2006; 65(3):712–25.

37. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977; 23 (3):327–41.

38. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics. 1993; 98(12):10089–92.

39. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic acids research. 2010; 38 (Web Server issue):W545–9. doi: 10.1093/nar/gkq366 PMID: 20457744; PubMed Central PMCID: PMC2896194.

40. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, et al. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. Nature. 2014; 506 (7489):498–502. doi: 10.1038/nature12907 PMID: 24463512.

41. Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, et al. Structure, dynamics, and specificity of endoglucanase D from *Clostridium cellulovorans*. Journal of molecular biology. 2013; 425 (22):4267–85. doi: 10.1016/j.jmb.2013.05.030 PMID: 23751954; PubMed Central PMCID: PMC4039632.

42. Kim HW, Kataoka M, Ishikawa K. Atomic resolution of the crystal structure of the hyperthermophilic family 12 endocellulase and stabilizing role of the DxDxDG calcium-binding motif in *Pyrococcus furiosus*. FEBS letters. 2012; 586(7):1009–13. Epub 2012/05/10. doi: 10.1016/j.febslet.2012.02.029 PMID: 22569255.

43. Abou-Hachem M, Karlsson EN, Simpson PJ, Linse S, Sellers P, Williamson MP, et al. Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*. Biochemistry. 2002; 41(18):5720–9. PMID: 11980476.

44. Jamal-Talabani S, Boraston AB, Turkenburg JP, Tarbouriech N, Ducros VM, Davies GJ. Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. Structure (London, England: 1993). 2004; 12(7):1177–87. doi: 10.1016/j.str.2004.04.022 PMID: 15242594.

45. Venditto I, Najmudin S, Luis AS, Ferreira LM, Sakka K, Knox JP, et al. Family 46 Carbohydrate-Binding Modules contribute to the enzymatic hydrolysis of xyloglucan and beta-1,3–1,4-glucans through distinct mechanisms. The Journal of biological chemistry. 2015; 290(17):10572–86.doi: 10.1074/jbc.M115.637827 PMID: 25713075.

46. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, et al. Protein identification and analysis tools in the ExPASy server. Methods in molecular biology (Clifton, NJ). 1999; 112:531–52. PMID: 10027275.

47. Delgado-Garcia M, Valdivia-Urdiales B, Aguilar-Gonzalez CN, Contreras-Esquivel JC, Rodriguez-Herrera R. Halophilic hydrolases as a new tool for the biotechnological industries. Journal of the science of food and agriculture. 2012; 92(13):2575–80. Epub 2012/08/29. doi: 10.1002/jsfa.5860 PMID: 22926924.

48. Garcia-Fraga B, da Silva AF, Lopez-Seijas J, Sieiro C. Functional expression and characterization of a chitinase from the marine archaeon *Halobacterium salinarum* CECT 395 in *Escherichia coli*. Applied microbiology and biotechnology. 2014; 98(5):2133–43. Epub 2013/07/31. doi: 10.1007/s00253-013-5124-2 PMID: 23893326.

49. Zheng B, Yang W, Zhao X, Wang Y, Lou Z, Rao Z, et al. Crystal structure of hyperthermophilic endo-beta-1,4-glucanase: implications for catalytic mechanism and thermostability. The Journal of biological chemistry. 2012; 287(11):8336–46. doi: 10.1074/jbc.M111.266346 PMID: 22128157; PubMed Central PMCID: PMC3318711.

50. Oyama T, Schmitz GE, Dodd D, Han Y, Burnett A, Nagasawa N, et al. Mutational and structural analyses of *Caldanaerobius polysaccharolyticus* Man5B reveal novel active site residues for family 5 glycoside hydrolases. PloS one. 2013; 8(11):e80448. Epub 2013/11/28. doi: 10.1371/journal.pone.0080448 PMID: 24278284; PubMed Central PMCID: PMC3835425.

51. Sakon J, Irwin D, Wilson DB, Karplus PA. Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca*. Nature structural biology. 1997; 4(10):810–8. PMID: 9334746.

52. Beckham GT, Bomble YJ, Matthews JF, Taylor CB, Resch MG, Yarbrough JM, et al. The O-glycosylated linker from the *Trichoderma reesei* Family 7 cellulase is a flexible, disordered protein. Biophysical journal. 2010; 99(11):3773–81. Epub 2010/11/30. doi: 10.1016/j.bpj.2010.10.032 PMID: 21112302; PubMed Central PMCID: PMC2998629.

53. Zhong L, Matthews JF, Hansen PI, Crowley MF, Cleary JM, Walker RC, et al. Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose Ibeta: the enzyme-substrate complex. Carbohydrate research. 2009; 344(15):1984–92. doi: 10.1016/j.carres.2009.07.005 PMID: 19699474.

54. Vester-Christensen MB, Abou Hachem M, Svensson B, Henriksen A. Crystal structure of an essential enzyme in seed starch degradation: barley limit dextrinase in complex with cyclodextrins. Journal of molecular biology. 2010; 403(5):739–50. Epub 2010/09/25. doi: 10.1016/j.jmb.2010.09.031 PMID: 20863834.

55. Lawson CL, van Montfort R, Strokopytov B, Rozeboom HJ, Kalk KH, de Vries GE, et al. Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. Journal of molecular biology. 1994; 236(2):590–600. doi: 10.1006/jmbi.1994.1168 PMID: 8107143.

56. Wang C-Y, Hsieh Y-R, Ng C-C, Chan H, Lin H-T, Tzeng W-S, et al. Purification and characterization of a novel halostable cellulase from *Salinivibrio* sp. strain NTU-05. Enzyme and Microbial Technology. 2009; 44(6–7):373–9.

57. Hirasawa K, Uchimura K, Kashiwa M, Grant WD, Ito S, Kobayashi T, et al. Salt-activated endoglucanase of a strain of alkaliphilic *Bacillus agaradhaerens*. Antonie van Leeuwenhoek. 2006; 89(2):211–9. Epub 2006/05/20. doi: 10.1007/s10482-005-9023-0 PMID: 16710633.