JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

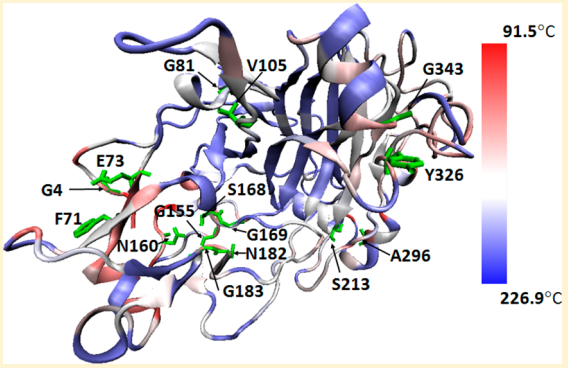
Improving *Trichoderma reesei* Cel7B Thermostability by Targeting the Weak Spots

Shujun Zhang,† Yefei Wang,† Xiangfei Song, Jingbo Hong, Yu Zhang, and Lishan Yao*

Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology Chinese Academy of Sciences, Qingdao, 266061, China

Supporting Information

ABSTRACT: For proteins that denature irreversibly, the denaturation is typically triggered by a partial unfolding, followed by a permanent change (e.g., aggregation). The regions that initiate the partial unfolding are named "weak spots". In this work, a molecular dynamics (MD) simulation and data analysis protocol is developed to identify the weak spots of *Trichoderma reesei* Cel7B, an important endoglucanase in cellulose hydrolysis, through assigning the local melting temperature (T_{mp}) to individual residue pairs. To test the predicted weak spots, a total of eight disulfide bonds were designed in these regions and all enhanced the enzyme thermostability. The increased stability, quantified by ΔT_{50} (which is the T_{50} difference between the mutant and the wild type enzyme), is negatively correlated with the MD-predicted T_{mp}, demonstrating the effectiveness of the protocol and highlighting the importance of the weak spots. Strengthening interactions in these regions proves to be a useful strategy in improving the thermostability of *Tr.* Cel7B.



INTRODUCTION

Cellulose hydrolysis of reducing sugars, which are then fermented to biofuels, is a key process in biomass conversion.[1] Cellulases represent a friendly catalyst for the hydrolysis, and these generally include exoglucanases cleaving cellobiose from cellulose strand ends, endoglucanases cleaving strands randomly, and β-glucosidases converting soluble cellodextrins and cellobiose to glucose. These cellulases usually consist of two domains—a large catalytic domain (CD) and a small carbohydrate-binding module (CBM)—connected by an O-glycosylated linker peptide. The binding of the CBM to cellulose is important for the hydrolysis.[2−4] However, recent work indicates that eliminating CBMs from cellulases results in comparable hydrolysis efficiency if the substrate lignocellulose loading is high.[5−7] Unlike the intact cellulases, which bind irreversibly to lignin after the hydrolysis of lignocellulose, the cellulases without CBMs tend to stay in the supernatant of the reaction mixture so that a much better enzyme recycling can be achieved[5−7] after the finish of the catalysis, which reduces the cost of the hydrolysis process, one of the key obstacles in the commercial application. The CBM and the linker are important for the thermostability of exoglucanase *Tr.* Cel7A[8] as well as endoglucanase *Tr.* Cel7B (discussed below), and removing these two components considerably decreases the melting temperature. To achieve an efficient cellulose hydrolysis at high temperature (e.g., 50 °C, which is the temperature at which the catalysis by commercial cellulases generally operates), the thermostability of the catalytic domains of Cel7A and Cel7B has to be improved.

Rational design plays an important role in protein engineering to, for example, improve protein stability, enzyme substrate selectivity, and catalytic efficiency. Unlike directed evolution,[9] which is another powerful protein engineering method, a three-dimensional (3-D) protein structure is prerequisite for the application of the rational design method. The advancing of rational design not only produces proteins with the desired properties, but also improves our understanding of the protein structural functional relationship. In the rational protein stability engineering process, a total of 19 natural amino acid mutations can be made for each residue, where the mutational effect on the thermostability must be determined. Many computational programs have been developed for this purpose (CC/PBSA,[10] EGAD,[11] FoldX,[12] I mutant 2.0,[13] Rosetta,[14−17] etc.). All these methods use scoring functions to rank the mutations; thus, these are quite efficient, but with only moderate accuracy. Potapov and co-workers[18] performed a benchmark study for the programs mentioned above and showed that the correlation coefficient R_p, between the predicted unfolding free-energy differences $\Delta\Delta G$ (between the wild type and the mutant) and the experimental values, is <0.6. Previously, we showed[19] that combining the FoldX method with free-energy calculations from molecular dynamics simulations is effective in improving the thermostability of a bacterial blue-light photoreceptor YtvA from *Bacillus subtilis*.[20] The free-energy method implicitly assumes that the equilibrium

Received: June 8, 2014

   dx.doi.org/10.1021/ci500339v | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

is formed between the folded state and the unfolded state. For many proteins that unfold irreversibly, the unfolded state is not well-defined. The denaturation generally consists of a partial unfolding, followed by an irreversible process (e.g., aggregation).[21,22] The key then is to find the region that is important for the partial unfolding, namely, the "weak spot",[22−25] and forge stronger interactions in the weak spot. However, identifying the weak spot, which can be done experimentally by making many mutations (e.g., Xxx to Ala) all over the protein and evaluating their effects on thermostability, is very time-consuming.

A funnel-shaped energy landscape is generally assumed for protein folding. The folded state, which is at the bottom of the landscape, corresponds to the most stable conformation. As the protein begins to unfold, more conformations are visited. Protein backbone hydrogen bonds were shown by NMR to respond differently to temperature[26,27] and pressure changes.[28] A NMR residual dipolar coupling study of protein GB1 structure at different pHs[29] suggested that certain regions of the protein start to change conformations, even though the overall protein structure remains folded. These regions with conformational changes are called "melting hot spots", which are reminiscent to weak spots.[29] Molecular dynamics (MD) simulation is a powerful technique in characterizing protein structure and dynamics and has been applied to study protein stability and its behavior at different temperatures.[30−35] Specifically, MD simulations at high temperature have produced results of protein unfolding in good agreement with experimental observations.[36−40] Encouraged by these studies, in this work a multitemperature MD simulation and data analysis protocol is developed to assign local melting temperature T_{mp} to each contacted residue pair. Then, based on the distribution of the local melting temperature, the weak spots are predicted. Disulfide bonds are introduced in the weak spots and all show improved thermostability. Combining multiple mutants further improves the enzyme thermostability. The results demonstrate that the simulation method is effective in identifying weak spots and strengthening interactions in these regions through disulfide bonds is a good method for improving *T. reesei* Cel7B thermostability.

METHODS AND MATERIALS

MD Simulation. The starting Cel7B model was from the X-ray crystallography structure (WT, Protein Databank (pdb) code: 1EG1[41]). The pK_a values of all ionizable residues were predicted by PROPKA.[42] In addition, those with pK_a smaller than 5.0 were deprotonated while those with pK_a larger than (or equal to) 5.0 were protonated. MD simulations were carried out using Gromacs 4.5,[43,44] with the Charmm27 force field[45] for the protein and TIP3P for water.[46] The protein was solvated by adding 9.0 Å of water in a dodecahedron box and Na^+ ions were used to neutralize the system. Before the production run, 1000 steps of energy minimization, followed by 20-ps MD simulations at constant pressure (1 atm) and temperature (300 K), were performed to equilibrate the system. A harmonic restraint with the force constant of 1000 kJ/nm^2 was used to restrain the backbone atoms of the protein in the equilibration and was removed in the production run. The pressure was regulated using the extended ensemble Parrinello−Rahman approach[47,48] and the temperature was controlled by a modified Berendsen thermostat.[49] The Particle-Mesh-Ewald Method[50,51] was used to evaluate the contribution of the long-range electrostatic interactions. A nonbonded pair list cutoff of 10.0 Å was used and the nonbonded pair list was updated every 5 steps, or 10 fs. All bonds to hydrogen atoms in proteins were constrained by using the LINCS[52] algorithm, whereas bonds and angles of water molecules were constrained by the SETTLE[53] algorithm, allowing a time step of 0.002 ps. The MD simulations were performed at the temperatures starting from 250 K, increased at an interval of 5 K to 500 K, and with three runs (15 ns each, with initial velocities generated randomly) at each temperature. MD snapshots were saved every 50 ps, and only those from the last 10 ns were included in the data analysis.

Cloning, Expression, and Purification of *Tr.* Cel7B. The enzyme and its mutants were expressed in *E. coli* strain origami (DE3). The details about the cloning, expression, and purification of *Tr.* Cel7B were described elsewhere.[54] All the mutations were made by PCR-based site-directed mutagenesis and verified by DNA sequencing. The purity was determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Protein concentration was measured by UV spectroscopy using an extinction coefficient of 62310 M^{-1} cm^{-1} at 280 nm.[55]

Activity on FP and Avicel. A mixture of filter paper (FP, Whatman No. 1) or Avicel PH101 (Sigma−Aldrich) with Cel7B-CD enzyme was incubated in a 2-mL Eppendorf tube (50 mg/mL substrate and 0.14 μM enzyme with a total volume of 400 μL, in a 100 mM NaAc and pH 5.0 buffer). For the specific activity measurements of the WT and the mutants, the mixtures were incubated at 50 °C for 24 h. The reactions were stopped by filtering the mixture through 0.22 μm filter, which effectively removes the reactant. The main products of cellulose hydrolysis by Cel7B are cellobiose and glucose.[56] To completely convert cellobiose to glucose, 25 μL of 2 μM *Thermotoga maritima* β-glucosidase (purified as described in the literature[57]) was added to 75 μL of the product mixture, which was then reacted for 1 h at 65 °C. The concentration of the glucose was measured by PAHBAH[58] assay. To determine the optima temperature of the enzyme, the mixtures were incubated at different temperatures (30, 40, 45, 50, 53, 56, 59, and 62 °C) for 15 min. The reducing sugar released was converted to glucose as described above. The time course of the Avicel and FP hydrolysis was also monitored for the WT and the mutant G4C−F71C/N160C−G183C/S168T at 40, 50, and 53 °C, at the time points of 0.25, 0.5, 1, 2, 4, 6, 9, 12, and 24 h. The activity experiments were done in triplicate, which provided an estimation of the measurement error. The specific activity (U/mg) is defined as the amount of reducing sugar (μmol) released per minute per milligram of enzyme.

T_{50} Measurements. T_{50} is defined as the temperature at which a 3 min of incubation causes a loss of 50% of the activity, relative to a reference sample, which does not undergo incubation. Twenty microliter (20 μL) samples containing 0.25 μM *Tr.* Cel7B (100 mM NaAc, pH 5.0) each were incubated at different temperatures for 3 min. After cooling on ice, 80 μL of CMC (final concentration of 0.8% (w/v), 100 mM NaAc, pH 5.0) were added to the reaction mixture and reacted for 15 min at 50 °C before quenched by DNS.[59] The reducing sugar was measured by the DNS assay.[59] The 3-min incubation at high temperature significantly diminished the enzyme activity at 50 °C, indicating that the unfolding of the enzyme was irreversible. The residual activity versus temperature was plotted and fitted using four-parameter sigmoidal curves.[60]

dx.doi.org/10.1021/ci500339v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

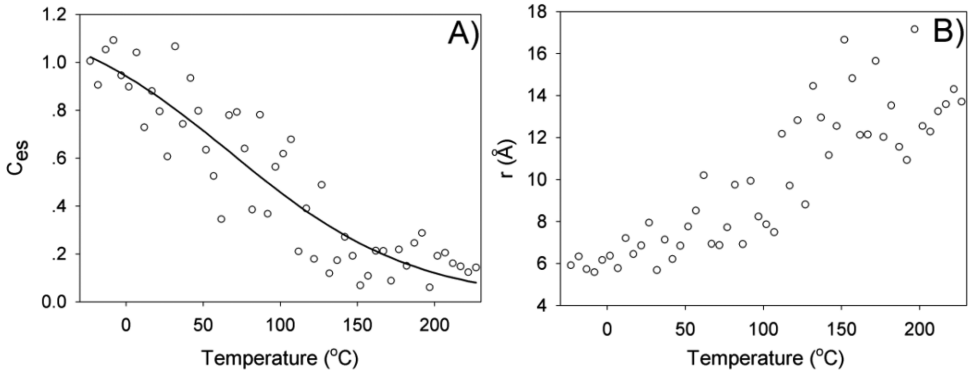


Figure 1. (A) Effective contact c_{es} between P3 and A162 versus the temperature. c_{es} is calculated from the MD simulations at each temperature (see more details in the main text) and fitted using eq 1. A melting temperature of T_{mp} = 91.5 ± 5.7 °C was obtained for the residue pair. (B) Average distance (r) between C_α atoms of P3 and A162 versus the temperature.

RESULTS AND DISCUSSION

A series of MD simulations of *T. reesei* Cel7B CD domain were run at temperatures ranged from −23.2 °C (250 K) to 226.9 °C (500 K) with an increment of 5 °C. Three independent MD runs (15 ns each) were performed at each temperature. The two residues are considered in contact only if they are separated apart by more than four amino acids in the sequence and have a closest distance of <4.0 Å (between two heavy atoms) in the X-ray structure (pdb code: 1EG1). A total of 694 contacts were identified from the X-ray structure.[41] The distance between the C_α atoms of each contacted residue pair was recorded and averaged at each temperature. This distance typically becomes larger at higher temperature due to the unfolding. And the variation of the C_α distances of different contacts can be substantial for large distances so that quantitatively comparing the contacts using the C_α distances is not straightforward. To solve the problem, the distance is converted to an effective contact c_e, by the equation $c_e = \exp(-r/r_0)$, where r is the distance between two C_α atoms and r_0 is a constant (4.0 Å). This equation ensures that the effective contact decreases as the distance increases and the variation of c_e becomes smaller for larger distances. c_e is averaged over the MD snapshots of three runs at each temperature to yield $\langle c_e \rangle$, where $\langle \cdot \rangle$ denotes the ensemble average. The effective contact at different temperatures then is scaled to yield $c_{es}(T_i) = \langle c_e(T_i) \rangle / \langle c_e(T_0) \rangle$, where $\langle c_e(T_0) \rangle$ is the average contact at the lowest temperature of the MD simulation (−23.2 °C). As a result, c_{es} at the lowest MD temperature is 1. In this work, the average $\langle c_e \rangle$ value from the four lowest simulation temperatures is used as the denominator to minimize the numeric noise. c_{es} is plotted against temperature for each contact and fitted to a sigmoid function (see Figure 1, as well as Figure S2 in the Supporting Information):

$$c_{es} = \frac{a}{1 + e^{b(T-c)}} \quad (1)$$

where a, b, and c are three fitting parameters (a corresponds to the c_{es} value at the lowest temperature (e.g., T = −23.2 °C), b reflects the decreasing rate of c_{es}, and c is the temperature at which c_{es} is equal to $a/2$). The temperature with c_{es} = 0.5 is defined as the local melting temperature of the contacted pair (T_{mp}). If a pair has c_{es} > 0.5, even at highest simulation temperature of 226.9 °C, extracting its T_{mp} is not possible, because its value is higher than 226.9 °C. Nevertheless, 226.9 °C is assigned to the pair T_{mp}, corresponding to the lower limit value of this residue pair. The T_{mp} values of different residue pairs are listed in Table S1 in the Supporting Information. To simplify the data, the pairwise T_{mp} is projected to individual residues by assigning the T_{mp} to the two residue components. If a residue forms multiple contact pairs, the lowest T_{mp} from the pairs is assigned to the residue (Figure 2). The mapping of residue specific T_{mp} to the enzyme 3D structure provides a direct view about the weak spot (Figure 3). Several clustered regions, including residues Q2−T7, W40−Y46, V55−E73, G151−S168, R179−L185, S213−G230, C293−L302, and N321−A350, show relatively low T_{mp} (<176.9 °C (450 K)), suggesting that their conformations begin to change at low temperature. In other words, they initiate the Cel7B unfolding. The details about each region will be described below together with the experimental data.

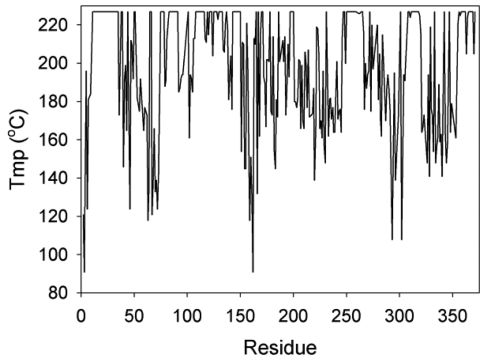


Figure 2. Residue specific melting temperature (T_{mp}). Several regions display relatively low T_{mp} (<176.9 °C (450 K)), including residues Q2−T7, W40−Y46, V55−E73, G151−S168, R179−L185, S213−G230, C293−L302, and N321−A350.

To test the predicted weak spots, disulfide bonds are introduced by using the software *Disulfide by Design* (DbD, version 1.20).[61] Only the disulfide bonds formed in the weak spots are subjected to experimental validations, which include a total of eight cysteine double mutants (see Table 1 and Figure 3). The wild-type Cel7B CD has a T_{50} value of 54.6 °C (Figure 4A), which is considerably lower than the T_m value of the intact *Tr.* Cel7B (67.6 °C),[62] but close to the T_m of *Tr.* Cel7A catalytic domain (51 °C), which is a homologue of Cel7B-CD.[8] The CBM and the linker apparently stabilize *Tr.* Cel7B. Compared to the wild-type Cel7B CD, all the mutants display higher T_{50} values,[60] suggesting that designing a disulfide bond

dx.doi.org/10.1021/ci500339v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

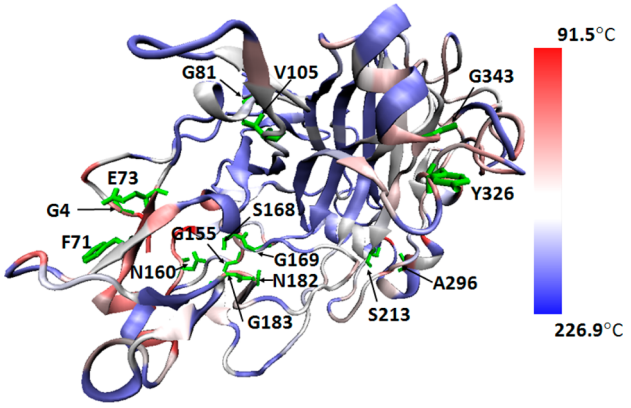


Figure 3. Mapping of T_{mp} to the 3-D X-ray structure of *Tr.* Cel7B (pdb code: 1EG1[41]). The color changes from red through white to blue as the T_{mp} increases. The residues involved in the mutations are labeled and highlighted in green. The figure is made by using the VMD software.[68]

in the regions with low T_{mp} is effective in improving the thermostability of *Tr.* Cel7B. It should be emphasized that the experimental T_{50} value cannot be compared directly with the predicted T_{mp} value, which, for most residues, is higher than 126.9 °C (400 K, Figure 2). The length of each MD simulation is only 15 ns, most likely much shorter than the unfolding time scale of *Tr.* Cel7B. However, the high-temperature simulations permit one to observe the initial unfolding (weak spots, assigned by T_{mp}) so that stronger interactions can be introduced to stabilize these initiation sites. In addition, the experimental T_{50} difference between the wild type and the mutant (ΔT_{50} = T_{50}(mutant) − T_{50}(WT)) shows a negative correlation with the predicted T_{mp} (Figure 4B), suggesting that the predicted weak spots are consistent with the experimental findings. The details about the weak spots are discussed as follows.

Q2−T7. This N-terminal fragment is in contact with multiple residues from the fragments W40−Y46, V55−E73, and G154−S168, and a few contacts display relatively low melting temperatures, such as Q2−G67 (T_{mp} = 120.9 °C, Table S1 in the Supporting Information), P3−A162 (91.5 °C), G4−E73 (142.3 °C), and S6−Y46 (124.7 °C). Two designed disulfide bonds G4C−F71C and G4C−E73C, which link two weak spots Q2−T7 and V55−E73, have T_{50} values that are 2.3 °C higher than the WT (Table 1). It is interesting that the G4C−F71C disulfide bond is present in the Cel7B homologue *Tr.* Cel7A.[63] Based on our calculation, it is expected that this disulfide bond is important for the thermostability of *Tr.* Cel7A as well.

V55−E73. Besides close contacts with Q2−T7, this fragment also forms contact with other residues from G151−S168 and R179−L185. Several contacts show low T_{mp} values, including E63−Y159 (118.3 °C), A64−Y159 (132.0 °C), G67−Y159 (156.6 °C), and I72−A162 (123.9 °C); however, no suitable disulfide bond can be found in this region with fragments G151−S168 or R179−L185 by using the DbD software.[61]

G151−S168. This fragment forms contacts with fragments Q2−T7 and V55−E73, as well as R179−L185. A few contacts have very low T_{mp} values, such as G151−N182 (154.2 °C), G154−G183 (145.3 °C), G155−G183 (145.3 °C), and N160−G183 (151.0 °C). Three disulfide bonds were designed—G155C−G169C, G155C−G182C, and N160C−G183C—and all showed increased T_{50} values (ΔT_{50} = 0.3, 1.9, and 2.2 °C, respectively; see Table 1).

S213−G230. About a dozen contacts are formed between this fragment and fragments C293−L302 and N321−A350. A few of them have low T_{mp}, including S213−A296 (177.0 °C), D220−A296 (139.4 °C), Y229−A335 (157.6 °C), and G230−N334 (148.2 °C). One disulfide bond S231C−A296C was designed, which had a T_{50} value 0.7 °C higher than that of the WT.

N321−A350. This fragment shows contacts with S213−G230 and within itself, among which D321−I348 (164.5 °C), D322−P345 (167.4 °C), Y326−G343 (163.2 °C), N328−S340 (141.5 °C) and S332−S340 (153.9 °C) appear to be relatively unstable. One disulfide bond Y326C−G343C was engineered, which improved the enzyme's T_{50} value by 2.7 °C.

All eight mutants from the weak spots show enhanced thermostability, with T_{50} increases ranging from 0.3 °C to 2.7 °C, suggesting that strengthening interactions in the weak spots through disulfide bond stabilizes Cel7B. We also conducted a

Table 1. T_{50} Values of the Mutants and Their Specific Activities against Filter Paper (FP) and Avicel Cellulose Compared to the WT *Tr.* Cel7B

Tr. Cel7B	T_{50} (°C)[a]	FP[b] (× 10^{-1} U/mg)	Avicel[b] (× 10^{-1} U/mg)
WT	54.6 ± 0.1	0.98 ± 0.04	1.70 ± 0.06
G4C/F71C	56.9 ± 0.1	0.97 ± 0.02	1.22 ± 0.01
G4C/E73C	56.9 ± 0.1	0.72 ± 0.02	1.20 ± 0.02
G81C/V105C	55.9 ± 0.2	0.44 ± 0.04	0.92 ± 0.01
G155C/N182C	56.5 ± 0.1	0.88 ± 0.13	1.34 ± 0.02
G155C/G169C	54.9 ± 0.1	0.34 ± 0.01	0.69 ± 0.01
N160C/G183C	56.8 ± 0.1	1.24 ± 0.05	1.23 ± 0.05
S213C/A296C	55.3 ± 0.1	0.28 ± 0.01	0.93 ± 0.02
Y326C/G343C	57.3 ± 0.2	1.15 ± 0.02	0.91 ± 0.06
G4C−F71C/N160C−G183C	60.4 ± 0.2	0.90 ± 0.12	1.25 ± 0.08
G155C−N182C/N160C−G183C	56.9 ± 0.1	0.85 ± 0.01	1.73 ± 0.05
G4C−F71C/G155C−N182C	60.7 ± 0.2	1.77 ± 0.05	1.90 ± 0.04
G4C−F71C/G155C−N182C/N160C−G183C	60.6 ± 0.1	0.45 ± 0.06	1.20 ± 0.07
G4C−F71C/N160C−G183C/S168T	62.8 ± 0.2	2.44 ± 0.19	2.18 ± 0.10

[a] T_{50} is defined as the temperature at which 50% of the activity is lost. [b] The specific activity was measured at 50 °C for 24 h and more details are described in the main text.

dx.doi.org/10.1021/ci500339v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

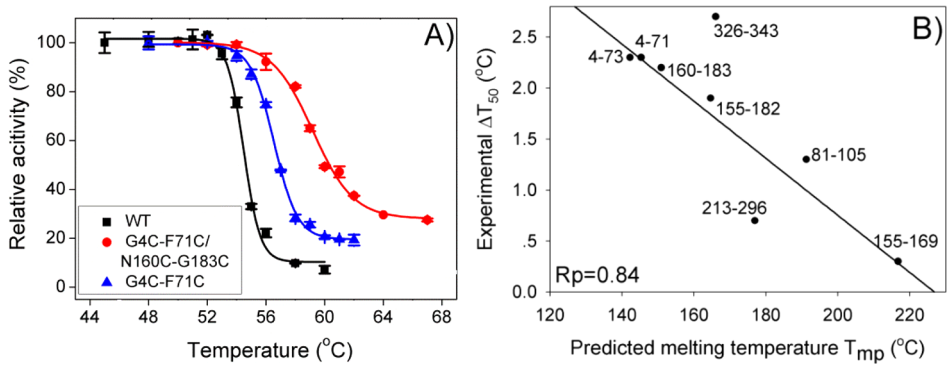


Figure 4. (A) Relative activity at 50 °C after incubation at different temperatures versus the incubation temperature for the wild type (WT), and mutants G4C−F71C and G4C−F71C/N160C−G183C. The T_{50} value was estimated by fitting the data to a sigmoid function.[60] (B) Correlation between the predicted pairwise T_{mp} (see Table S1 in the Supporting Information) and experimental ΔT_{50}, defined as the T_{50} difference between the mutant and wild type *Tr.* Cel7B ($\Delta T_{50} = T_{50}$(mutant) − T_{50}(WT)). The best fitted line is $\Delta T_{50} = 14 - 0.028T_{mp}$. The pairwise T_{mp} was from the fitting of c_{es} to eq 1.

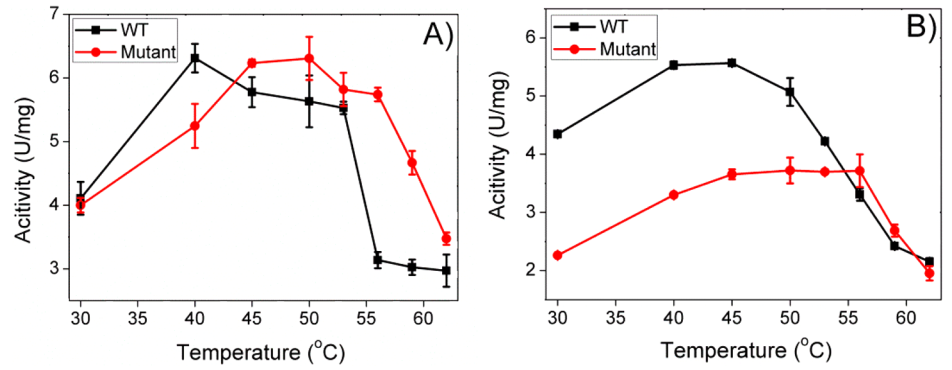


Figure 5. Temperature optimum curves of the WT Cel7B and the mutant G4C−F71C/N160C−G183C/S168T in the hydrolysis of (A) Avicel and (B) FP. The enzyme substrate mixtures were reacted for 15 min.

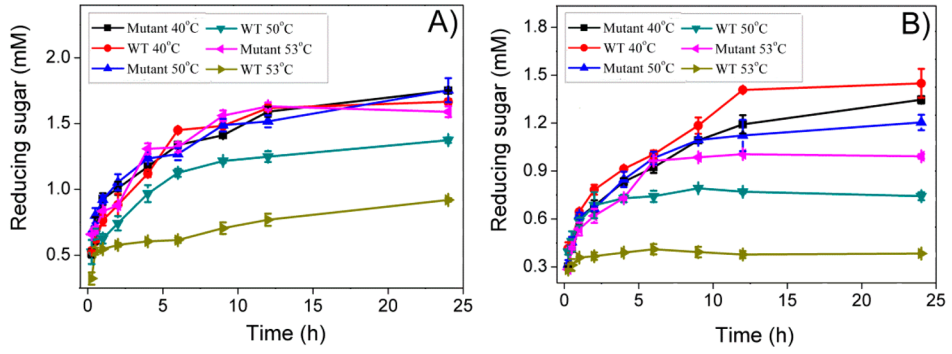


Figure 6. Degradation of (A) Avicel and (B) FP by the WT Cel7B and the mutant G4C−F71C/N160C−G183C/S168T. The time course of the hydrolysis was followed for the WT and the mutant at 40, 50, and 53 °C for 24 h.

control study by designing a disulfide bond mutant T15C−S87C which is not from the weak spot. This mutant has a T_{50} of 55.3 °C, slightly higher than the WT (ΔT_{50} = 0.7 °C). However, this T_{50} value is lower than six of the eight disulfide bond mutants from the weak spots. Thus, the authors argue that designing a disulfide bond from the weak spots is a better way to improve protein stability. The negative correlation between predicted T_{mp} and experimental ΔT_{50} also underlines the importance of the weak spot for the disulfide bond selection (see Figure 4B).

Combination of Mutations. As discussed above, all the disulfide bond mutations improve the enzyme thermostability. The enzyme-catalyzed hydrolysis against filter paper and Avicel cellulose was also measured, with most of the mutants having slightly reduced activities (Table 1). A few mutants with relatively high activities are combined, where G4C−F71C/N160C−G183C and G4C−F71C/G155C−N182C further improve the T_{50} to ∼60 °C (Table 1), indicating that these two disulfide bonds are independent from each other. However, the combination G155C−N182C/N160C−G183C has a T_{50} value of 56.9 °C, which is close to the individual disulfide bond mutants, so that the stabilization effect is not additive, probably due to the proximity of the two disulfide bonds. A single mutation S168T, also from the weak spot, further improves the T_{50} values by 2.4 after combination with G4C−F71C/N160C−G183C.

dx.doi.org/10.1021/ci500339v | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Specific activity of the mutant G4C−F71C/N160C−G183C/S168T against Avicel and FP was determined as a function of temperature (30−63 °C, 15 min of reaction time). When using Avicel as the substrate (Figure 5A), the temperature optimum for the mutant is ∼50 °C (∼10 °C higher than that of the WT), and the maximum activity of the mutant is similar to the WT. In comparison, when using FP as the substrate, the mutant has a lower specific activity than the WT at low temperature, although it still has a temperature optimum ∼10 °C higher than the WT (Figure 5B). The activity of the G4C−F71C/N160C−G183C/S168T mutant against Avicel and FP was also measured in a time series (up to 24 h) at 40, 50, and 53 °C. For the Avicel substrate, the mutant displays a similar activity profile at three temperatures, whereas the WT shows a decreased activity at high temperature (Figure 6A). A similar effect is seen for the WT when using FP as the substrate (Figure 6B). The decrease in the activity of the WT is most likely due to its poor thermostability. As for the mutant, a small activity decrease is also observed for the FP substrate as temperature increases, but the amount of reducing sugar released is still ∼2.5-fold of that produced by the WT at 53 °C. Overall, the mutant displays a higher activity than the WT at high temperature, because of its improved thermostability.

For a protein that unfolds reversibly, its stability is governed by the free-energy difference between the folded and unfolded states. By increasing this free-energy difference, the protein stability can be improved. However, if the protein unfolds irreversibly, meaning that after the partial or full unfolding, some permanent change occurs to the protein, the protein stability is determined by the (partial) unfolding activation energy immediately before the irreversible step.[21,22] For such a system, it is critical to characterize this unfolding step. Pikkemaat et al. used the MD simulation to probe the flexibility of haloalkane dehalogenase and designed a disulfide bond to restrain a flexible region, which increased the protein's stability.[32] A similar approach has been adopted to improve lipase thermostability by targeting the residues with high X-ray B-factors.[64] The root-mean-square fluctuation (RMSF) of *Tr.* Cel7B backbone heavy atoms from MD simulations at 26.9 °C shows that a few high-RMSF regions overlap with those with low T_{mp} values (see Figure S1 in the Supporting Information). However, the direct correlation between the two is poor (Figure 7), suggesting that the local T_{mp} provides new information that cannot be retrieved from the RMSF obtained from MD simulations at a single temperature. The constraint network analysis (CNA)[65−67] developed by Gohlke et al. can also be used to predict the weak spots. The advantage of CNA is its efficiency, which is much higher than the MD method described in this work. It will be interesting to see whether the weak spots predicted by CNA are consistent with the MD method.

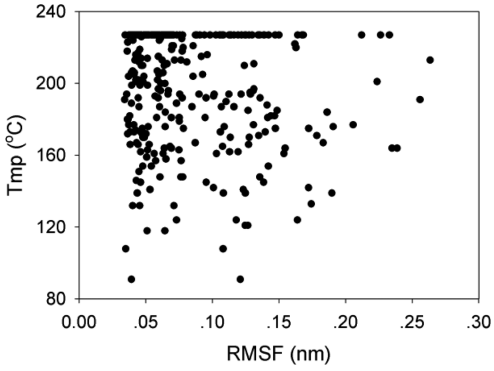


Figure 7. Local melting temperature (T_{mp}) versus the root-mean-square fluctuation (RMSF) of backbone C_α calculated from the 26.9 °C MD simulations. The poor correlation (R_p = 0.14) between the two suggests that the weak spots extracted from T_{mp} cannot be retrieved from the RMSF.

CONCLUSIONS

In summary, a MD simulation and data analysis protocol has been developed to investigate the weak spots of *Tr.* Cel7B, whereas engineering disulfide bonds in these weak spots results in improvement of the enzyme's thermostability, with lower-melting-temperature regions displaying more-pronounced effects. The protocol can be easily employed to study other protein systems to probe the weak spot or improve the thermostability. For the disulfide bond mutants that decrease the enzyme activity, it is likely that the mutation perturbs the enzyme structure or dynamics. Directed evolution with a library built based on the weak spots may minimize the perturbation and thus improve the thermostability without compromising the activity. Further studies are needed to test this idea.

ASSOCIATED CONTENT

Supporting Information

Figures showing the root-mean-square fluctuations (RMSFs) of C_α (Figure S1) and the effective contact c_{es} versus the temperature, relative to various melting temperatures (Figure S2); a table is provided that lists the local melting temperature (T_{mp}) values. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel.: 86 532 80662792. Fax: 86 532 80662778. E-mail: yaols@qibebt.ac.cn. Postal Address: QIBEBT, Room 413, Qingdao 266061, PRC.

Author Contributions

†These two authors contributed equally to the work.

Notes

The authors declare the following competing financial interest(s): A Chinese patent has been filed, using part of the results given in this paper.

ACKNOWLEDGMENTS

This work was supported in part by 100 Talent Project of Chinese Academy of Sciences, National Nature Science Foundation of China (Grant Nos. 21173247 and 31270785 to L.Y.), the Foundation for Outstanding Young Scientist in Shandong Province (Grant No. JQ201104 to L.Y. and Grant No. ZR2011BQ008 to Y.W.), and the "135" Projects Fund of the CAS-QIBEBT Director Innovation Foundation.

REFERENCES

(1) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. *Science* **2007**, *315*, 804−807.

(2) Boraston, A. B.; Bolam, D. N.; Gilbert, H. J.; Davies, G. J. Carbohydrate-Binding Modules: Fine-Tuning Polysaccharide Recognition. *Biochem. J.* **2004**, *382*, 769−781.

(3) Suurnakki, A.; Tenkanen, M.; Siika-Aho, M.; Niku-Paavola, M. L.; Viikari, L.; Buchert, J. *Trichoderma Reesei* Cellulases and Their Core Domains in the Hydrolysis and Modification of Chemical Pulp. *Cellulose* **2000**, *7*, 189−209.
(4) Linder, M.; Teeri, T. T. The Roles and Function of Cellulose-Binding Domains. *J. Biotechnol.* **1997**, *57*, 15−28.
(5) Le Costaouec, T.; Pakarinen, A.; Varnai, A.; Puranen, T.; Viikari, L. The Role of Carbohydrate Binding Module (CBM) at High Substrate Consistency: Comparison of *Trichoderma Reesei* and *Thermoascus Aurantiacus* Cel7A (CBHI) and Cel5A (EGII). *Bioresour. Technol.* **2013**, *143*, 196−203.
(6) Pakarinen, A.; Haven, M. O.; Djajadi, D. T.; Varnai, A.; Puranen, T.; Viikari, L. Cellulases Without Carbohydrate-Binding Modules in High Consistency Ethanol Production Process. *Biotechnol. Biofuels* **2014**, 7.
(7) Varnai, A.; Siika-aho, M.; Viikari, L. Carbohydrate-Binding Modules (CBMs) Revisited: Reduced Amount of Water Counterbalances the Need for CBMs. *Biotechnol. Biofuels* **2013**, 6.
(8) Hall, M.; Rubin, J.; Behrens, S. H.; Bommarius, A. S. The Cellulose-Binding Domain of Cellobiohydrolase Cel7A from *Trichoderma Reesei* is Also a Thermostabilizing Domain. *J. Biotechnol.* **2011**, *155*, 370−376.
(9) Giver, L.; Gershenson, A.; Freskgard, P. O.; Arnold, F. H. Directed Evolution of a Thermostable Esterase. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12809−12813.
(10) Benedix, A.; Becker, C. M.; de Groot, B. L.; Caflisch, A.; Bockmann, R. A. Predicting Free Energy Changes Using Structural Ensembles. *Nat. Methods* **2009**, *6*, 3−4.
(11) Pokala, N.; Handel, T. M. Energy Functions for Protein Design: Adjustment with Protein−Protein Complex Affinities, Models for the Unfolded State, and Negative Design of Solubility and Specificity. *J. Mol. Biol.* **2005**, *347*, 203−227.
(12) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369−387.
(13) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant 2.0: Predicting Stability Changes Upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306−W310.
(14) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 830−838.
(15) Dantas, G.; Corrent, C.; Reichow, S. L.; Havranek, J. J.; Eletr, Z. M.; Isern, N. G.; Kuhlman, B.; Varani, G.; Merritt, E. A.; Baker, D. High-Resolution Structural and Thermodynamic Analysis of Extreme Stabilization of Human Procarboxypeptidase by Computational Protein Design. *J. Mol. Biol.* **2007**, *366*, 1209−1221.
(16) Leaver-Fay, A.; Tyka, M.; Lewis, S. M.; Lange, O. F.; Thompson, J.; Jacak, R.; Kaufman, K.; Renfrew, P. D.; Smith, C. A.; Sheffler, W.; Davis, I. W.; Cooper, S.; Treuille, A.; Mandell, D. J.; Richter, F.; Ban, Y.-E. A.; Fleishman, S. J.; Corn, J. E.; Kim, D. E.; Lyskov, S.; Berrondo, M.; Mentzer, S.; Popovic, Z.; Havranek, J. J.; Karanicolas, J.; Das, R.; Meiler, J.; Kortemme, T.; Gray, J. J.; Kuhlman, B.; Baker, D.; Bradley, P. ROSETTA3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. *Method. Enzymol.* **2011**, *487*, 545−574.
(17) Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally Designed Libraries for Rapid Enzyme Stabilization. *Protein Eng. Des. Sel.* **2014**, *27*, 49−58.
(18) Potapov, V.; Cohen, M.; Schreiber, G. Assessing Computational Methods for Predicting Protein Stability Upon Mutation: Good on Average but Not in the Details. *Protein Eng., Des. Sel.* **2009**, *22*, 553−560.
(19) Song, X. F.; Wang, Y. F.; Shu, Z. Y.; Hong, J. B.; Li, T.; Yao, L. S. Engineering a More Thermostable Blue Light Photo Receptor *Bacillus Subtilis* YtvA LOV Domain by a Computer Aided Rational Design Method. *PLoS Comput. Biol.* **2013**, 9.
(20) Losi, A.; Polverini, E.; Quest, B.; Gartner, W. First Evidence for Phototropin-Related Blue-Light Receptors in Prokaryotes. *Biophys. J.* **2002**, *82*, 2627−2634.
(21) Eijsink, V. G. H.; Gaseidnes, S.; Borchert, T. V.; van den Burg, B. Directed Evolution of Enzyme Stability. *Biomol. Eng.* **2005**, *22*, 21−30.
(22) Eijsink, V. G. H.; Bjork, A.; Gaseidnes, S.; Sirevag, R.; Synstad, B.; van den Burg, B.; Vriend, G. Rational Engineering of Enzyme Stability. *J. Biotechnol.* **2004**, *113*, 105−120.
(23) Eijsink, V. G. H.; Veltman, O. R.; Aukema, W.; Vriend, G.; Venema, G. Structural Determinants of the Stability of Thermolysin-Like Proteinases. *Nat. Struct. Biol.* **1995**, *2*, 374−379.
(24) Mansfeld, J.; Vriend, G.; Dijkstra, B. W.; Veltman, O. R.; VandenBurg, B.; Venema, G.; UlbrichHofmann, R.; Eijsink, V. G. H. Extreme Stabilization of a Thermolysin-Like Protease by an Engineered Disulfide Bond. *J. Biol. Chem.* **1997**, *272*, 11152−11156.
(25) Vriend, G.; Berendsen, H. J. C.; van den Burg, B.; Venema, G.; Eijsink, V. G. H. Early Steps in the Unfolding of Thermolysin-Like Proteases. *J. Biol. Chem.* **1998**, *273*, 35074−35077.
(26) Cordier, F.; Grzesiek, S. Temperature-Dependence Properties as Studied by of Protein Hydrogen Bond High-Resolution NMR. *J. Mol. Biol.* **2002**, *317*, 739−752.
(27) Hong, J. B.; Jing, Q. Q.; Yao, L. S. The Protein Amide H−1(N) Chemical Shift Temperature Coefficient Reflects Thermal Expansion of the N−H Center Dot Center Dot Center Dot O═C Hydrogen Bond. *J. Biomol. NMR* **2013**, *55*, 71−78.
(28) Nisius, L.; Grzesiek, S. Key Stabilizing Elements of Protein Structure Identified through Pressure and Temperature Perturbation of Its Hydrogen Bond Network. *Nat.Chem.* **2012**, *4*, 711−717.
(29) Ding, K. Y.; Louis, J. M.; Gronenborn, A. M. Insights into Conformation and Dynamics of Protein GB1 during Folding and Unfolding by NMR. *J. Mol. Biol.* **2004**, *335*, 1299−1307.
(30) Wang, T.; Duan, Y. Probing the Stability-Limiting Regions of an Antibody Single-Chain Variable Fragment: A Molecular Dynamics Simulation Study. *Protein Eng. Des. Sel.* **2011**, *24*, 649−657.
(31) Meharenna, Y. T.; Poulos, T. L. Using Molecular Dynamics To Probe the Structural Basis for Enhanced Stability in Thermal Stable Cytochromes P450. *Biochemistry* **2010**, *49*, 6680−6686.
(32) Pikkemaat, M. G.; Linssen, A. B. M.; Berendsen, H. J. C.; Janssen, D. B. Molecular Dynamics Simulations as a Tool for Improving Protein Stability. *Protein Eng.* **2002**, *15*, 185−192.
(33) Merkley, E. D.; Parson, W. W.; Daggett, V. Temperature Dependence of the Flexibility of Thermophilic and Mesophilic Flavoenzymes of the Nitroreductase Fold. *Protein Eng. Des. Sel.* **2010**, *23*, 327−336.
(34) Huang, X.; Zhou, H.-X. Similarity and Difference in the Unfolding of Thermophilic and Mesophilic Cold Shock Proteins Studied by Molecular Dynamics. *Biophys. J.* **2006**, *91*, 2451−2463.
(35) Motono, C.; Gromiha, M. M.; Kumar, S. Thermodynamic and Kinetic Determinants of *Thermotoga Maritima* Cold Shock Protein Stability: A Structural and Dynamic Analysis. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 655−669.
(36) Alonso, D. O. V.; Daggett, V. Staphylococcal protein A: Unfolding Pathways, Unfolded states, and Differences between the B and E Domains. *Eur. J. Med. Chem.* **2000**, *97*, 133−138.
(37) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. Protein Folding and Unfolding in Microseconds to Nanoseconds by Experiment and Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518−13522.
(38) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W. N.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. Unifying Features in Protein-Folding Mechanisms. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13286−13291.
(39) Ferguson, N.; Day, R.; Johnson, C. M.; Allen, M. D.; Daggett, V.; Fersht, A. R. Simulation and Experiment at High Temperatures: Ultrafast Folding of a Thermophilic Protein by Nucleation−Condensation. *J. Mol. Biol.* **2005**, *347*, 855−870.

(40) Sharpe, T.; Jonsson, A. L.; Rutherford, T. J.; Daggett, V.; Fersht, A. R. The Role of the Turn in Beta-Hairpin Formation during WW Domain Folding. *Protein Sci.* **2007**, *16*, 2233−2239.
(41) Kleywegt, G. J.; Zou, J. Y.; Divne, C.; Davies, G. J.; Sinning, I.; Stahlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. The Crystal Structure of the Catalytic Core Domain of Endoglucanase I from *Trichoderma Reesei* at 3.6 Angstrom Resolution, and a Comparison with Related Enzymes. *J. Mol. Biol.* **1997**, *272*, 383−397.
(42) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p*K*(a) Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.
(43) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(44) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(45) 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: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(47) Nose, S.; Klein, M. L. Constant Pressure Molecular-Dynamics for Molecular-Systems. *Mol. Phys.* **1983**, *50*, 1055−1076.
(48) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals—A New Molecular-Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(49) Berendsen, H. J. C. Transport-Properties Computed by Linear Response through Weak-Coupling to a Bath. *Nato. Adv. Sci. I E-Appl.* **1991**, *205*, 139−155.
(50) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald—An *N*·log(*N*) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(51) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(52) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(53) Miyamoto, S.; Kollman, P. A. Settle—An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(54) Zhang, Y.; Yan, S.; Yao, L. A Mechanistic Study of *Trichoderma Reesei* Cel7B Catalyzed Glycosidic Bond Cleavage. *J. Phys. Chem. B* **2013**, *117*, 8714−8722.
(55) Claeyssens, M.; Vantilbeurgh, H.; Kamerling, J. P.; Berg, J.; Vrsanska, M.; Biely, P. Studies of the Cellulolytic System of the Filamentous Fungus *Trichoderma Reesei* QM-9414 Substrate Specificity and Transfer Acitivity of Endoglucanse-I. *Biochem. J.* **1990**, *270*, 251−256.
(56) Murphy, L.; Cruys-Bagger, N.; Damgaard, H. D.; Baumann, M. J.; Olsen, S. N.; Borch, K.; Lassen, S. F.; Sweeney, M.; Tatsumi, H.; Westh, P. Origin of Initial Burst in Activity for *Trichoderma reesei* endo-Glucanases Hydrolyzing Insoluble Cellulose. *J. Biol. Chem.* **2012**, *287*, 1252−1260.
(57) Xue, Y. M.; Song, X. F.; Yu, J. J. Overexpression of Beta-Glucosidase from *Thermotoga Maritima* for the Production of Highly Purified Aglycone Isoflavones from Soy Flour. *World J. Microbiol. Biotechnol.* **2009**, *25*, 2165−2172.
(58) Lever, M. Carbohydrate Determination with 4-Hydroxybenzoic Acid Hydrazide (Pahbah)—Effect of Bismuth on Reaction. *Anal. Biochem.* **1977**, *81*, 21−27.
(59) Miller, G. L. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **1959**, *31*, 426−428.
(60) Wu, I.; Arnold, F. H. Engineered Thermostable Fungal Cel6A and Cel7A Cellobiohydrolases Hydrolyze Cellulose Efficiently at Elevated Temperatures. *Biotechnol. Bioeng.* **2013**, *110*, 1874−1883.
(61) Dombkowski, A. A. Disulfide by Design: A Computational Method for the Rational Design of Disulfide Bonds in Proteins. *Bioinformatics* **2003**, *19*, 1852−1853.
(62) Lantz, S. E.; Goedegebuur, F.; Hommes, R.; Kaper, T.; Kelemen, B. R.; Mitchinson, C.; Wallace, L.; Stahlberg, J.; Larenas, E. A. *Hypocrea Jecorina* CEL6A Protein Engineering. *Biotechnol. Biofuels* **2010**, *3*, 20.
(63) Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. The 3-Dimensional Crystal Structure of the Catalytic Core of Cellobiohydrolase-I from *Trichoderma-Reesei*. *Science* **1994**, *265*, 524−528.
(64) Reetz, M. T.; D Carballeira, J.; Vogel, A. Iterative Saturation Mutagenesis on the Basis of B factors as a Strategy for Increasing Protein Thermostability. *Angew. Chem.—Int. Ed.* **2006**, *45*, 7745−7751.
(65) Pfleger, C.; Rathi, P. C.; Klein, D. L.; Radestock, S.; Gohlke, H. Constraint Network Analysis (CNA): A Python Software Package for Efficiently Linking Biomacromolecular Structure, Flexibility, (Thermo-)Stability, and Function. *J. Chem. Inf. Model.* **2013**, *53*, 1007−1015.
(66) Radestock, S.; Gohlke, H. Protein Rigidity and Thermophilic Adaptation. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1089−1108.
(67) Rathi, P. C.; Radestock, S.; Gohlke, H. Thermostabilizing Mutations Preferentially Occur at Structural Weak Spots with a High Mutation Ratio. *J. Biotechnol.* **2012**, *159*, 135−144.
(68) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph. Model.* **1996**, *14*, 33−38.