Boosting Enzyme Activity in Biomass Conversion by Accelerating the Catalytic and Non-Catalytic Processes of Cellobiohydrolases

Han Liu, Yu Ding, Scott Mazurkewich, Wenwen Pei, Xiuxin Wei, Johan Larsbrink, Christophe Chipot, Zhangyong Hong, Wensheng Cai, Zhiyou Zong

Boosting Enzyme Activity in Biomass Conversion by Accelerating the Catalytic and Non-Catalytic Processes of Cellobiohydrolases

Han Liu[1,2,#], Yu Ding[3,#], Scott Mazurkewich[4,#], Wenwen Pei[1,2], Xiuxin Wei[3], Johan Larsbrink[4], Christophe Chipot[5,6,7,*], Zhangyong Hong[3,*], Wensheng Cai,[3,*] and Zhiyou Zong[1,2,*]

[1] *National Engineering Research Center of Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China.*

[2] *National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China.*

[3] *State Key Laboratory of Medicinal Chemical Biology, Tianjin Key Laboratory of Biosensing and Molecular Recognition, Research Center for Analytical Sciences, College of Chemistry, Nankai University, Tianjin 300071, China*

[4] *Wallenberg Wood Science Center, Division of Industrial Biotechnology, Department of Life Sciences, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden.*

[5] *Laboratoire International Associé CNRS and University of Illinois at Urbana−Champaign, LPCT, UMR 7019 Université de Lorraine CNRS, Vandœuvre-lès-Nancy F-54500, France*

[6] *Department of Physics, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

[7] *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637, United States*

***Corresponding E-mail:** *chipot@illinois.edu, hongzy@nankai.edu.cn, wscai@nankai.edu.cn, zongzhy@tib.cas.cn*

ABSTRACT: Cellobiohydrolases (CBHs) from the glycoside hydrolase family 7 are the most significant cellulose-degrading enzymes, the performance of which determines the cost-effective utilization of the renewable lignocellulosic resource. Most engineering strategies for improving CBHs hydrolysis are currently focused on the non-catalytic process, such as increasing the flexibility of eight substrate-enclosing loops (SELs) to accelerate the enzyme-substrate dissociation, which does not take the catalysis into account, or even deteriorates it. Here, in the model *Trichoderma reesei* CBHI, we identified a key SEL that affects the non-catalytic enzyme-substrate dissociation by examining enzyme-enzyme/substrate interactions. Furthermore, through analyzing the hydrogen-bonding network for the catalytic region, we detected a crucial residue D262. Root-mean-square-fluctuation analysis indicates that its replacement by valine (D262V) markedly improves the stability of the catalytic triad. Through QM/MM simulations we determined that this mutation diminished the free-energy barrier against catalysis by 2.2 kcal/mol, and increased k_{cat} by 53.1%, as determined in kinetic experiments. Additionally, the substitution caused a significant enhancement of the SEL flexibility, resulting in a lowered non-catalytic enzyme-substrate dissociation barrier by 2.4 kcal/mol, and a 49.8% cellobiose yield increment. This work unlocks a brand-new engineering direction for the industrially important CBHs, contributing to more efficient depolymerization of renewable lignocellulose.

KEYWORDS: lignocellulose, cellobiohydrolases, bioconversion, enzyme engineering, catalytic and noncatalytic processes, computational simulation, structure-function relationship

INTRODUCTION

Lignocellulose is an abundant renewable biomass resource and mainly refers to wood, energy crops, and agricultural crop residues. In second-generation biorefinery applications, it is the main feedstock for production of eco-friendly, value-added products, such as biomaterials, bioenergy, and a variety of platform chemicals. Improved use of lignocellulose could significantly reduce greenhouse gas emissions, whilst playing a major role in the transition to sustainable societies.[1] Lignin, hemicellulose, and cellulose are the major components of lignocellulose,[2, 3] wherein cellulose can be converted by cellulase cocktails into simple sugar molecules under mild conditions, offering huge potential for biofuel generation.[4] However, although the degradation of cellulose is becoming increasingly more cost-effective, the enzyme-hydrolysis step still accounts for a quarter of the cost of lignocellulose biofuel production.[5, 6] Improvement of enzyme performance is, therefore, critical to enhance the competitiveness of lignocellulosic biofuels, and bolster clean industries.

Cellulase cocktails consist at the core of endo-β-1,4-glucanases (EGs, EC 3.2.1.4), cellobiohydrolases (CBHs, EC 3.2.1.91), β-glucosidases (BGs, EC 3.2.1.21), and nowadays also oxidative lytic polysaccharide monooxygenases (LPMOs, EC 1.14.99.53-56), which altogether degrade cellulose step by step into glucose. The glycoside hydrolase family 7 (GH7) CBHI secreted by *Trichoderma reesei* (*Tr*Cel7A), which catalyzes cellulose depolymerization into cellobiose, is commonly utilized as a key model CBHs in biomass degradation processes.[7] The entire process of *Tr*Cel7A hydrolyzing insoluble microcrystalline cellulose, comprising the catalytic process (glycosylation/deglycosylation reaction,[8] processive movement of polysaccharide,[7] and product expulsion[9]), and the non-catalytic process, i.e. the enzyme-substrate dissociation,[10] has been well defined by Knott et al., Vermaas et al., and Zong et al. (Figure 1). The catalytic process in *Tr*Cel7A governs the catalytic rate constant (k_{cat}), while the non-catalytic enzyme-substrate dissociation, modulated by the flexibility of eight substrate-enclosing loops (SELs),[5] defines the overall product yield.

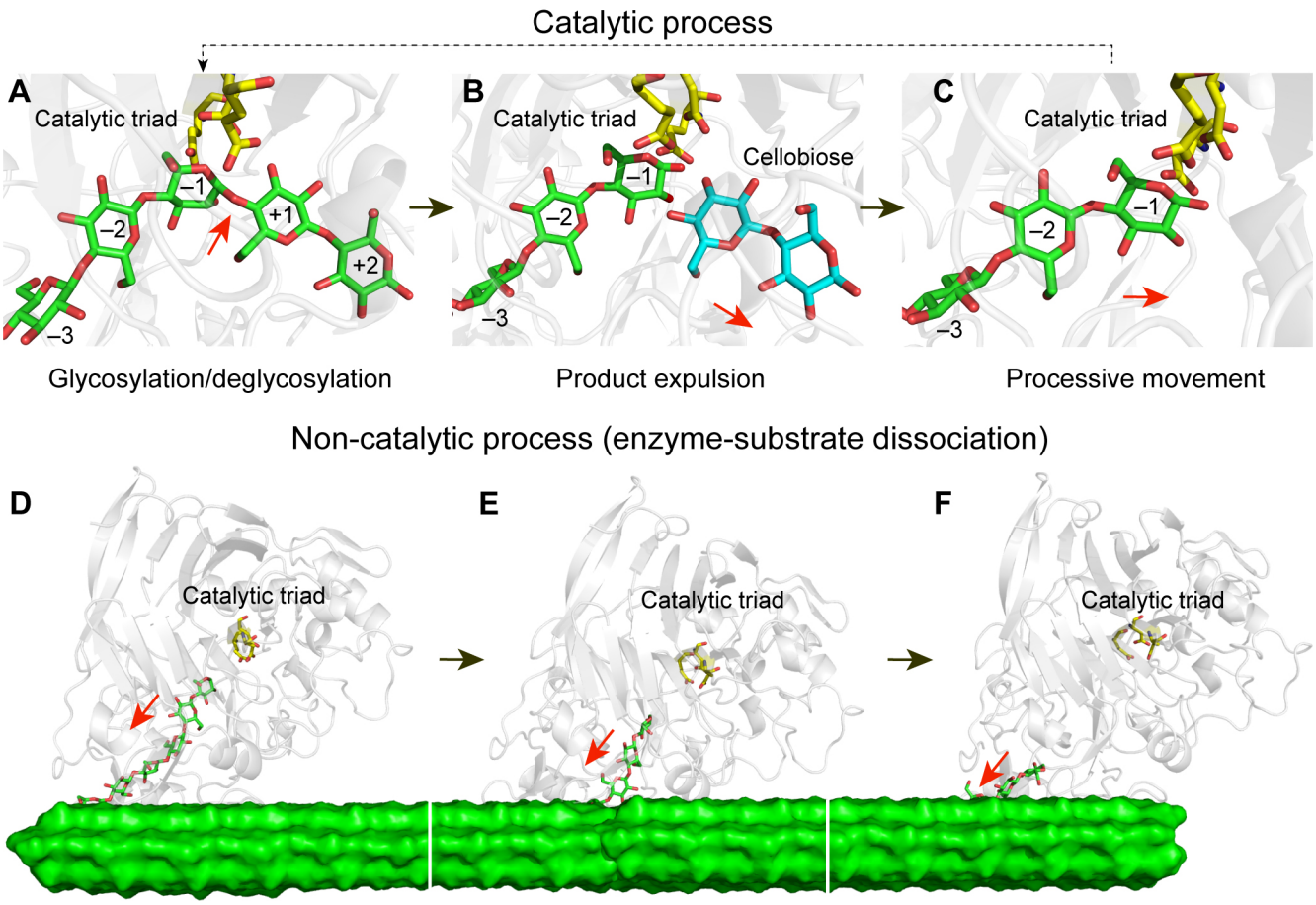

Figure 1. Schematic diagram for the entire hydrolytic process of CBHI. (A-C) Catalytic process. (D-F) Non-catalytic process, i.e., the dissociation of the enzyme from microcrystalline cellulose. The red arrows denote the (A) glycoside bond cleavage site and the directions for (B) cellobiose expulsion, (C) polysaccharide processive movement, and (D-F) dissociation. The catalytic residues, polysaccharide chain, and cellobiose are shown in stick model; their C atoms are colored in yellow, green, and cyan. For clarity, all hydrogen atoms are hidden. The microcrystalline cellulose in (D-F) is depicted as a green surface. Dissociation refers to the enzyme sliding off the polysaccharide chain stage by stage.

Most engineering strategies for improving the properties of CBHs have focused on accelerating the non-catalytic enzyme-substrate dissociation (Figure 1D-F) by improving the flexibility of SELs. In the work of Taylor II et al., *Penicillium funiculosum* CBHI (*Pf*Cel7A) was found to perform better than *Tr*Cel7A.[5] A double mutant of *Tr*Cel7A was, therefore, constructed in reference to the structure of *Pf*Cel7A, leading to an improved flexibility of SELs in *Tr*Cel7A. However, the k_{cat} value in the variant does not reach the level of the reference *Pf*Cel7A, and is even slightly lower than that of the wild-type (WT) *Tr*Cel7A. Furthermore, the enzyme-substrate affinity in the variant was enhanced significantly. In

most instances, high enzyme-substrate affinity could detrimentally affect k_{cat},[11] thereby impeding the catalytic process. The current strategy of accelerating the non-catalytic process towards better CBHs performance, thus, typically yields unsatisfactory results, and warrants serious reconsideration. However, how to develop strategies that take both the catalytic and non-catalytic processes into account represents a daunting challenge, especially from the perspective of single-residue substitutions.

For the sake of achieving the overarching goal of promoting the catalytic and non-catalytic processes in CBHs hydrolysis, we sought clues considering the following two aspects. First, in a study conducted by Igarashi et al., *T. reesei* CBHII (*Tr*Cel6A) was found to enhance the mobility of *Tr*Cel7A on cellulose,[12] which might result from the increased flexibility of the SELs in *Tr*Cel7A. In addition, in our previous work, we observed that a hydrophobin can enhance the product yield through potential protein association, wherein the SELs are considered as the crucial regions of interaction.[13] These findings led us to investigate potential *Tr*Cel7A-*Tr*Cel6A interaction, and explore which SELs in *Tr*Cel7A are affected by the interaction, and whether there are key SELs responsible for the non-catalytic enzyme-substrate dissociation. Second, the stability of the active site, which may be modulated by the hydrogen-bonding interactions between the residues in the catalytic region, is hypothesized here to be a major factor in the catalytic process. For instance, increased stability of the catalytic triad may enhance catalysis by decreasing the free-energy barrier against the catalytic reaction. Screening residue substitutions to strengthen the hydrogen-bonding interactions between the residues in the catalytic region represents, therefore, a potential approach to improve catalysis. Ultimately, finding mutations that can simultaneously regulate the stability of the catalytic triad and the flexibility of the key SELs affecting the non-catalytic enzyme-substrate dissociation is of paramount importance.

In the present work, we identified a key SEL that is the most relevant to dissociation, and designed substitutions of the residues in the catalytic region to improve stability of the active site. Large-scale molecular dynamics (MD) simulations, free-energy calculations, and state-of-the-art QM/MM simulations were employed to elucidate the molecular mechanism responsible for the accelerated catalytic and non-catalytic processes. Furthermore, we performed enzyme kinetics and hydrolytic experiments to

detect the actual performance of the variants on soluble and insoluble substrates. Overall, the present work, combining synergistically theory and experiment, deepens our understanding of the relationship between the enzyme structural features and the hydrolytic function, and is envisioned to help diminish enzyme cost in biomass degradation.

RESULTS AND DISCUSSION

Searching for Key SELs Resulting from *Tr*Cel7A-*Tr*Cel6A Interaction. To improve our understanding of how the interaction of *Tr*Cel7A with *Tr*Cel6A is able to enhance the mobility of *Tr*Cel7A on cellulose, we first evaluated their interaction by protein docking. As depicted in Figure 2A, the optimal docking pose of *Tr*Cel7A-*Tr*Cel6A indicates that *Tr*Cel6A associates with the SELs of *Tr*Cel7A. Close examination of the docking scores for the *Tr*Cel7A-*Tr*Cel6A assembly (Table S1) suggests that the van der Waals energy is the primary interactional force. Using to the binding free-energy estimator 2 (BFEE2), we followed the well-accepted geometrical route[14, 15] to determine rigorously the standard binding free energy ($\Delta G°$) for the *Tr*Cel7A-*Tr*Cel6A assembly. As controls, corresponding values of *Tr*Cel7A interacting with EG, BG, as well as with itself were also examined. $\Delta G°$ of *Tr*Cel7A-*Tr*Cel6A, *Tr*Cel7A-EG, *Tr*Cel7A-BG, and *Tr*Cel7A-*Tr*Cel7A assemblies are equal to –11.3, –3.9, –10.0, and –7.4 kcal/mol (Table S2). These results indicate that the *Tr*Cel7A-*Tr*Cel6A assembly possesses the strongest binding affinity among all enzymatic combinations considered here.

To explore the dynamics of the *Tr*Cel7A-*Tr*Cel6A interaction, we performed microsecond-long MD simulations for the enzyme-enzyme assembly, starting from the optimal docking structure. Based on the MD trajectories, the root-mean-square fluctuation (RMSF) for these eight SELs were computed (Figure S1), which implied that the flexibility of five SELs, namely A1, A3, B1, B2, and B3, improved significantly. The opening of the SELs (Figure 2B) is important for CBHI degrading microcrystalline cellulose.[13, 16] We, therefore, measured the cross-sectional area (CSA) formed by the distances between centers of mass (COMs) of SELs A3, B2, and B3 which revealed, as illustrated in Figure 2C, that the CSA of *Tr*Cel7A expanded considerably from *Tr*Cel6A association. As shown in Figure 2D, after 1-μs MD

simulations, the *Tr*Cel7A-*Tr*Cel6A complex started to dissociate, suggesting that the interaction of *Tr*Cel7A with *Tr*Cel6A is dynamic, and will not impair the accessibility of *Tr*Cel7A to microcrystalline cellulose. Additionally, we found in the experiment that the dynamic *Tr*Cel7A-*Tr*Cel6A interaction is weak and difficult to quantify accurately. Put together, our theoretical results support the view of an interaction between *Tr*Cel7A and *Tr*Cel6A in *T. reesei* cellulases. A3, B2, and B3, which determine the opening of SELs and exhibit the most significant increase in flexibility, are the most markedly affected SELs by *Tr*Cel7A-*Tr*Cel6A interaction.

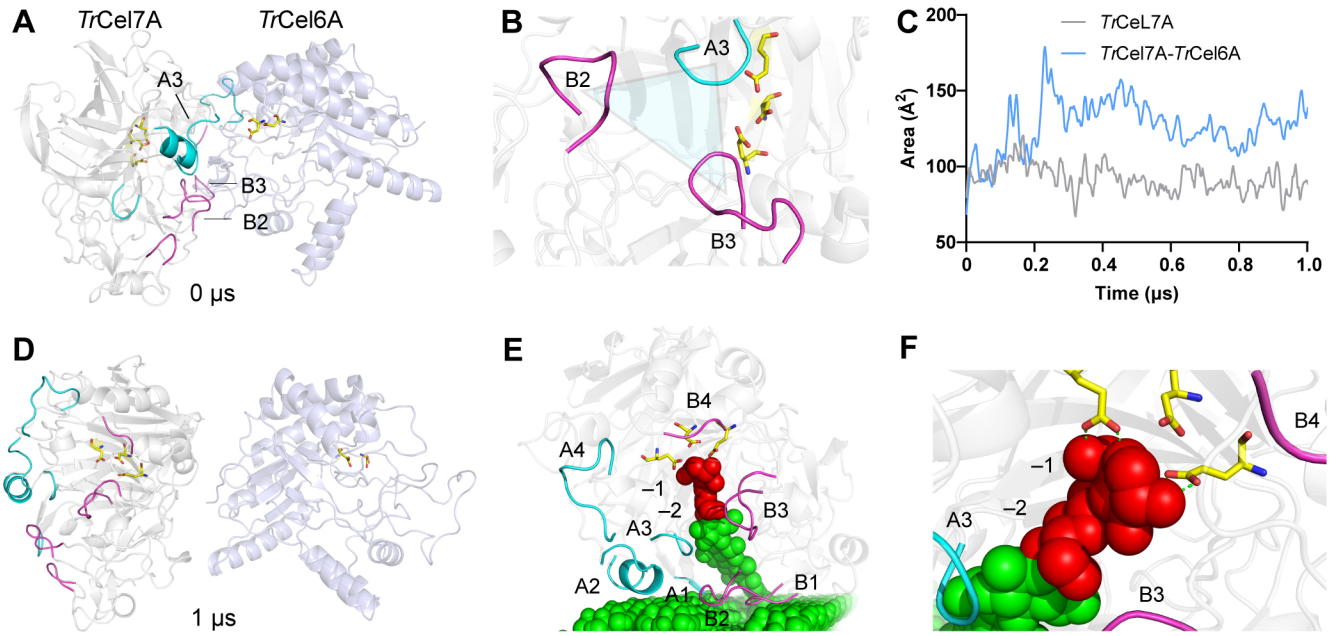

Figure 2. Interaction of *Tr*Cel7A with *Tr*Cel6A. (A) Optimal docking pose of the *Tr*Cel7A-*Tr*Cel6A complex. (B) CSA is characterized by the size of the triangle connecting the COMs of SELs A3, B2, and B3. (C) Time evolution of the CSA in *Tr*Cel7A and *Tr*Cel7A-*Tr*Cel6A assemblies over each 1 μs period. (D) *Tr*Cel7A-*Tr*Cel6A structure after 1-μs MD simulation. (E) Schematic diagram of the enzyme-substrate complex before dissociation. (F) Hydrogen bonds formed between the –1 glucosyl ring and the catalytic triad. The SELs A1-A4 (Q98-K102, T399-S411, D369-N373, and T383-A392) and B1-B4 (Y51-W56, P194-T201, G244-G253, and G339-S342) are colored in cyan and purple, respectively. The SELs A3, B2, and B3 are labeled in (A). The catalytic residues of *Tr*Cel7A (E212, D214, and E217) and *Tr*Cel6A (D175 and D221) are shown in stick model; their C atoms are colored in yellow. In (E and F), the –1 –2 glucosyl rings and the other rings are colored in red and green, respectively. For clarity, all hydrogen atoms are hidden.

The enzyme-substrate model in Figure 2E indicates that SELs A1, A2, B1, and B2 combine with the cellulose surface, while SELs A3 and B3 lie in the solution, implying that the improved flexibility of the

SELs A3 and B3 in response of the *Tr*Cel7A-*Tr*Cel6A interaction can be preserved during hydrolysis. Vermaas et al. have reported that the –1 to –3 transition is the crucial process for the enzyme-substrate dissociation, especially the –1 to –2.[10] As depicted in Figure 2F, the –1 –2 glucosyl rings form strong hydrogen-bonding interactions with the catalytic residues, which hinders dissociation significantly. Compared with SEL A3, B3 is much closer to the –1 –2 glucosyl rings. We speculate that the increased flexibility of SEL B3 may facilitate its interaction with the leading glucosyl rings, and thereby promote the escape of the latter from the active site. Therefore, to simultaneously accelerate the catalytic and non-catalytic processes in enzyme hydrolysis, in the subsequent engineering of the catalytic region, residue substitutions endowed with the potential to modulate the flexibility of SEL B3, are highly desirable.

Design of the Catalytic Region for Enhancing the Stability of the Active Site. *Tr*Cel7A contains three catalytic residues (E212, D214, and E217). To enhance the stability of the catalytic triad, we performed hydrogen-bonding network analysis for the eighteen residues (Figure 3), which lie within 5Å around the active site, as a way to identify potential residue substitution sites that can strengthen the hydrogen-bonding interactions. Four residues, i.e., P258, G260, D262, and W376, were found to lead to much weaker hydrogen-bonding interactions with other residues than the other fourteen ones (Table S3). We, therefore, carried out virtual saturation mutations (except for the inert alanine) on these four sites, and evaluated each variant through a hydrogen-bonding network analysis. All the virtual screening data are provided in Table S4, and the top five are shown in Table 1.

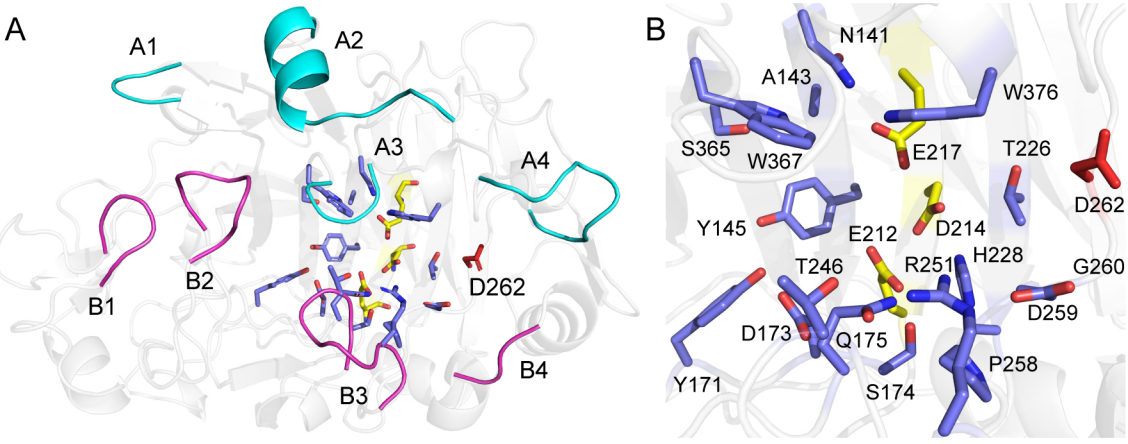

Figure 3. Schematic diagram of the structure of *Tr*Cel7A. (A) Overall view of the eight SELs and the selected eighteen residues in the catalytic region. (B) Close-up perspective of the selected eighteen residues. The catalytic residues are

shown in stick model and their C atoms are colored in yellow. The side chains of residue D262 and the other seventeen adjacent residues around the catalytic triad are shown as red and blue sticks, wherein G260 has no side chain. For clarity, all hydrogen atoms are hidden.

Table 1 Top five variants obtained by hydrogen-bonding network analysis for the catalytic region.

	WT	D262N	D262R	D262I	D262W	D262V
HB No.[a]	50	55	54	54	54	54
HB energy[b]	–19.0	–21.0	–20.6	–20.5	–20.1	–20.5

[a] HB No. denotes the total number of hydrogen bonds formed between the selected eighteen residues in the catalytic region. [b] HB energy denotes the hydrogen-bonding energy (kcal/mol) contributed by the formed hydrogen bonds between the selected eighteen residues in the catalytic region.

The total number of hydrogen bonds and the relevant hydrogen-bonding energy reflecting the hydrogen-bonding strength, as calculated by Rosetta HbondMetric, were employed to quantify the hydrogen-bonding interactions between the selected eighteen residues in the catalytic region. Table 1 shows that both the total number and the strength of hydrogen bonds in the catalytic region of each variant have improved significantly, compared with the corresponding quantities in the WT enzyme. Additionally, the side chains in these five residue substitutions are bulkier, leading to more pronounced steric hindrances than with the original aspartic acid, which may benefit the enzyme-substrate dissociation. We then constructed for these top five replacements the corresponding Michaelis complex, namely the conformation before catalysis, conducted MD simulations, and measured the RMSF of the catalytic triad in each variant based on the generated trajectories. Our results indicate that the stability of the catalytic triad has been improved with either D262N, D262I, D262W, or D262V, and most notably with the D262V and D262W variants (Figure 4) which were selected as the most promising substitution targets. We considered that if variant D262V or variant D262W could also affect the flexibility of SEL B3, the catalytic and non-catalytic processes in the enzyme hydrolysis could be modulated simultaneously by single residue substitutions.

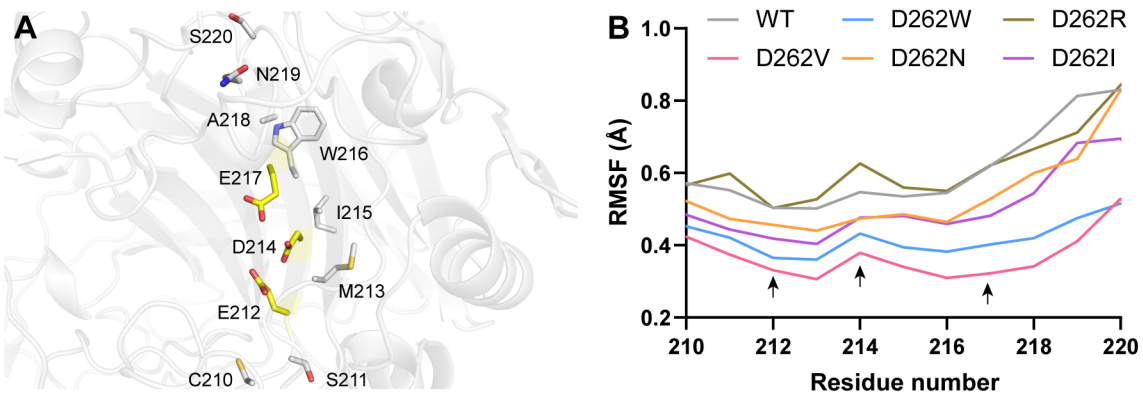

Figure 4. RMSF of the active site in the WT *Tr*Cel7A and the five variants. (A) Schematic diagram of residues from C210 to S220. (B) RMSF of residues from C210 to S220. The arrow points to three catalytic residues, i.e., E212, D214, and E217. In each assembly for investigating the stability of the active site, there is a polysaccharide chain (–7 to +2 glucosyl rings) in the catalytic tunnel; the polysaccharide forms the Michaelis complex with the enzyme, i.e., the conformation before glycosylation.

Enzyme Kinetics. Referring to our previous work, *Talaromyces emersonii* Cel7A (*Te*Cel7A) is not only very similar in sequence (sharing 66% and 78% sequence identity and similarity), structure (C_α distance root mean square deviation of 0.42 Å of *Tr*Cel7A [PDB code 4C4C] compared with the 3PFX model of *Te*Cel7A), and function to *Tr*Cel7A, but it is also easier to express than the latter.[9] Therefore, in order to experimentally probe the effect of the D262V and D262W variants (*Tr*Cel7A numbering) on enzymatic kinetics, the WT *Te*Cel7A and the relevant variants (D262 in *Tr*Cel7A corresponding to D259 in *Te*Cel7A) were synthesized, expressed in *Pichia pastoris*, and purified (Figure S2 and Figure S3). Kinetic experiments for these enzymes were then performed by hydrolysis of the soluble substrate, i.e. *p*-nitrophenol-cellobioside (*p*NPC, to monitor the catalytic rate by release of *p*NP; Figure S4). Table 2 shows that the D259V variant possesses a higher k_{cat} (53.1% increase) than the WT enzyme. Furthermore, we found that the K_m (Michaelis constant) for the D259V variant increased as expected, which may result from the stronger steric hindrance caused by the side chain in valine, bulkier than that of the initial aspartate.

Table 2 Enzyme kinetics for the WT *Te*Cel7A and variants of D259V and D259W*.

	WT	D259V	D259W

k_{cat} (min^{-1})	9.82±0.16	15.03±0.42	3.60±0.03
K_m (mM)	0.25±0.02	0.42±0.05	0.05±0.01

* *p*NPC is used as the substrate; experiments were performed in triplicate and errors represent the standard error of the mean.

QM/MM Simulations for Elucidating the Accelerated Catalytic Process. To interpret the reason of the k_{cat} increase from a mechanistic perspective, we investigated the influence of the valine replacement on the free-energy barrier against glycosylation (the rate-determining step in catalysis) by means of QM/MM simulations, following the reaction coordinate (RC) defined by Knott et al.[8] As illustrated in Figure 5, compared with the WT enzyme, our calculations show that the free-energy barrier against glycosylation, amounting here to 13.4 kcal/mol, is lowered by 2.2 kcal/mol as a result of the substitution with valine, which is the essential reason for the higher catalytic rate. The results of enzymatic kinetics experiments and QM/MM simulations suggest that enhancing hydrogen-bonding interactions of the residues lying in the catalytic region is a feasible strategy to improve the stability of the active site, and, thereby, modulate the catalytic process.

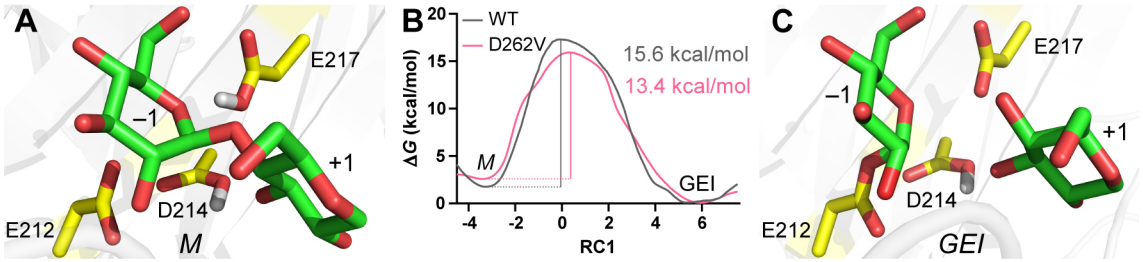

Figure 5. Catalytic reaction. (A) Schematic sketch for the Michaelis complex (*M*) before glycosylation. (B) Potential of mean forces (PMFs) for the glycosylation in the assemblies of WT enzyme and the D262V variant. (C) Schematic sketch for the glycosyl-enzyme intermediate (*GEI*) after glycosylation. The substrate and the catalytic triad are shown in stick model and colored in green and yellow, respectively. D214 and E217 are protonated, wherein E217 is the acid/base residue and E212 is the nucleophile; for clarity, all the other hydrogen atoms are hidden.

Non-Catalytic Enzyme-Substrate Dissociation. The catalysis of CBHI will repeat 10 to 60 times, and then, the enzyme will undergo non-catalytic enzyme-substrate dissociation when encountering an encumbered substrate surface.[7, 12, 17-19] We constructed the so-called preslide models[6, 7, 10] for the WT enzyme and for its D262V variant (Figure 6A), and performed 0.5-μs MD simulations for each

computational assay, from which the interaction energies of the catalytic triad with the crucial –1 –2 glucosyl rings were determined. As depicted in Figure 6B, time evolution of the interaction energy of the catalytic triad with the –1 –2 glucosyl rings in the D262V variant is much lower (3.9-fold decrease; Figure 6C) than that in the WT enzyme, indicating that in the D262V variant, the two leading glucosyl rings can escape from the catalytic triad more easily than in the WT enzyme. As shown in Figure 6D, the time series of the distance between the catalytic triad and the two leading glucosyl rings suggest that in the D262V variant, the latter has separated from the former at the beginning of the simulations (within the first 0.1 μs), while in the WT assembly, the interaction between the two groups was preserved over the entire 0.5-μs MD simulations.

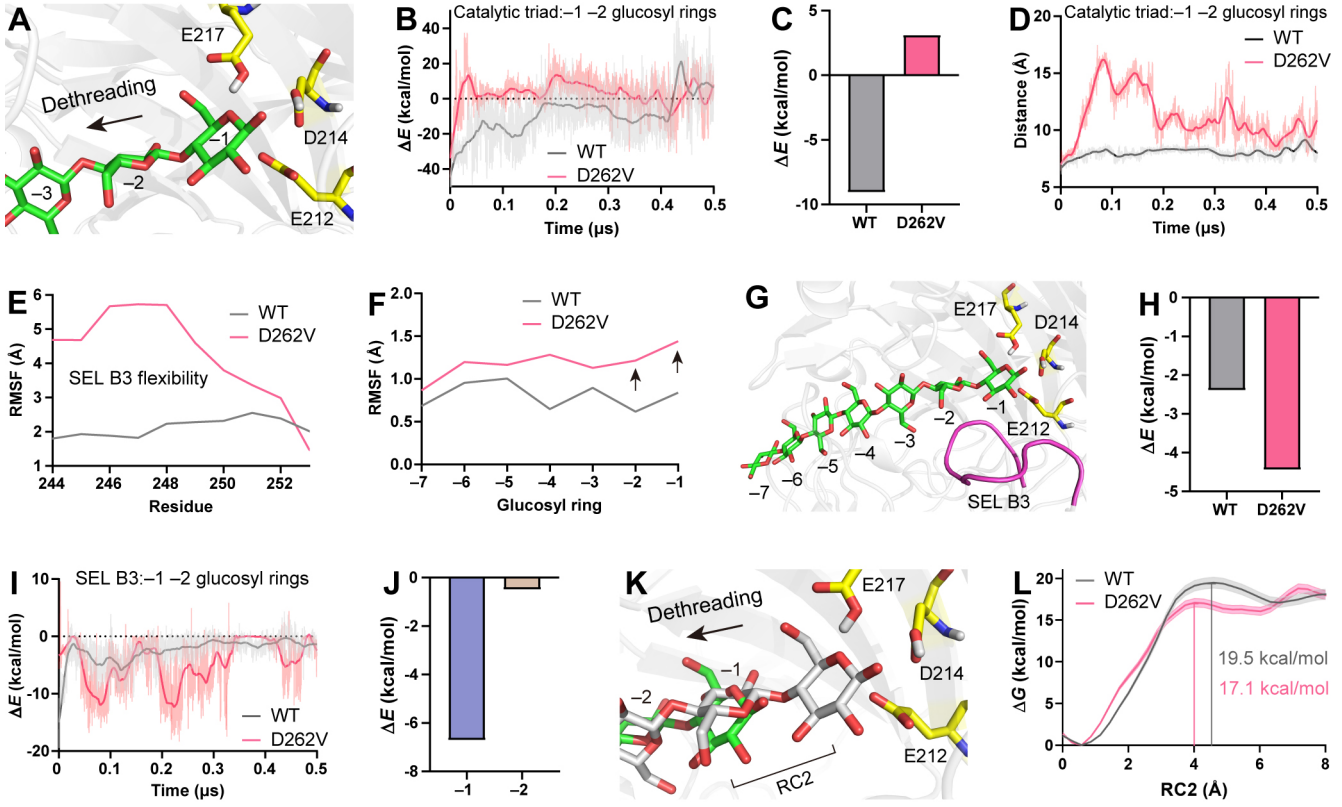

Figure 6. Enzyme-substrate dissociation. (A) The constructed enzyme-substrate complex for investigating dissociation. In each assembly for investigating dissociation, there is a polysaccharide chain (–7 to –1 glucosyl rings) in the catalytic tunnel, while cellobiose has been expulsed from the tunnel, i.e., the preslide state. E217 and D214 were protonated. (B) Time evolution of interaction energies and (C) the average energies of the catalytic triad with the –1 –2 glucosyl rings in the WT and the D262V assemblies over each 0.5 μs period. (D) Time evolution of the distance between the COMs of the catalytic triad and the –1 –2 glucosyl rings in the WT and the D262V assemblies over each 0.5 μs period. A distance greater than 10 Å indicates that the leading glucosyl rings have liberated from the restraint of the catalytic triad. (E) and (F) RMSF of the SEL B3 and the

polysaccharide chain (–7 to –1 glucosyl rings) in the WT and the D262V assemblies obtained from each 0.5-μs MD trajectory. (G) Schematic sketch for the SEL B3 and its adjacent glucosyl rings. (H) Average interaction energies and (I) time evolution of the energies of the SEL B3 with the –1 –2 glucosyl rings in the WT and the D262V assemblies over each 0.5 μs period. (J) Average interaction energies of the SEL B3 with the –1 and –2 glucosyl rings in the D262V assembly over each 0.5 μs period. (K) RC2 denotes the distance of the –1 glucosyl ring movement during the initial dissociation. The C atoms of the polysaccharide chain before and after the initial dissociation are colored in white and green. (L) Free-energy profiles for the escape of the –1 glucosyl ring from the catalytic triad in the WT and the D262V assemblies over each 0.32 μs period. Shaded regions around the free-energy profiles are the standard error of the mean computed from three independent runs. For clarity, all the hydrogen atoms are hidden, except for D214 and E217, which are protonated in the preslide model.

To elucidate the reason for the accelerated escape of the –1 –2 glucosyl rings from the active site in the D262V variant, we calculated the RMSF for the polysaccharide chain (i.e., –7 to –1 glucosyl rings) in the catalytic tunnel, and for the identified key SEL, namely B3. As illustrated in Figure 6E and 6F, flexibility of SEL B3 and of the –1 –2 glucosyl rings are increased significantly by the valine replacement. Furthermore, the –1 –2 glucosyl rings are much closer to SEL B3, compared with the other glucosyl rings (i.e., –7 to –3 glucosyl rings; Figure 6G). We infer that the interaction of SEL B3 with the leading glucosyl rings may facilitate the latter escape from the active site, and, therefore, we calculated their interaction energies. The time evolution of the latter quantities suggests that SEL B3 in the D262V variant interacts in a much stronger way with the two leading glucosyl rings (1.9-fold increase; Figure 6H) than that in the WT enzyme (Figure 6I). In light of these findings, we consider that the valine substitution endows SEL B3 with enhanced flexibility, which leads the loop to interact with the –1 –2 glucosyl rings more frequently and more strongly, thereby accelerating the escape of the leading glucosyl rings from the active site. Additionally, as depicted in Figure 6J, by analyzing the interaction energies of SEL B3 with the –1 and –2 glucosyl rings, we found that the SEL interacts with the –1 glucosyl ring 13.4-fold stronger than with the –2 glucosyl ring (see the time evolution of the interaction energy in Figure S5). This result indicates that SEL B3 modulates the non-catalytic enzyme-substrate dissociation primarily through its interaction with the –1 glucosyl ring. To quantify the influence of the valine replacement on this non-catalytic process, we estimated the free-energy barrier against the escape process of the –1 glucosyl ring from the active site

(see the definition of the RC in Figure 6K). As shown in Figure 6L, the free-energy barrier against dissociation in the D262V variant (viz., 17.1 kcal/mol free-energy barrier), is lowered by 2.4 kcal/mol, compared with that in the WT assay (viz., 19.5 kcal/mol free-energy barrier).

Hydrolysis Experiment. Given that the catalysis as well as the non-catalytic enzyme-substrate dissociation is primarily governed by the catalytic domain (CD),[17, 20] we, therefore, expressed and purified the CD of *Te*Cel7A to perform the hydrolysis experiment of microcrystalline cellulose, and test our *in-silico*-based hypotheses. As depicted in Figure 7, in the D259V system, increases of 75.0%, 32.4%, and 49.8% in cellobiose yield were observed at 24 h, 48 h, and 72 h respectively, compared with the WT group. Overall, our theoretical and experimental results suggest that the catalytic and non-catalytic processes in the enzyme have been simultaneously promoted by a single valine replacement. Additionally, the sequence-alignment results of CBHI from twelve diverse species show that every CBHI possesses a conserved aspartate residue at position 262 (*Tr*Cel7A numbering).[21] We extended this sequence analysis to embrace all recognized GH7 members and observed aspartate in about 70% of the members at this position, while glycine and all the other amino acids collectively account for 25% and less than 5% of the members at this position, respectively (Figure S6). These results underscore the significance of the proposed valine substitution at position 262 (*Tr*Cel7A numbering) for the evolution of the CBHI family.

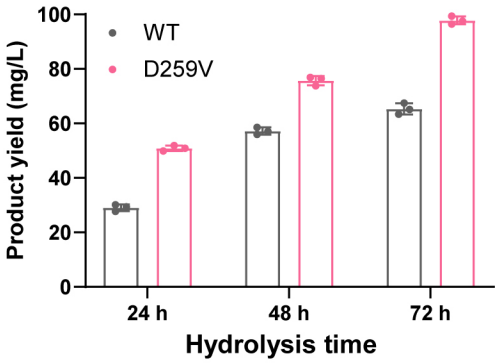

Figure 7. Hydrolysis of microcrystalline cellulose. Cellobiose released from microcrystalline cellulose at different time points (24 h, 48 h, and 72 h) with the hydrolysis of the WT *Te*Cel7A and the D259V variant (corresponding to the D262V in *Tr*Cel7A). Experiments were performed in triplicate and error bars represent the standard error of the mean.

CONCLUSIONS

In the lignocellulosic bioeconomy, cellulose-degrading enzymes, particularly CBHs, are of utmost significance.[6] In the last decade, the rate-determining step of the model CBHs, i.e., *Tr*Cel7A, hydrolyzing insoluble substrates has been demonstrated unambigously to be the non-catalytic enzyme-substrate dissociation.[10] However, in industry, hydrolysis of the lignocellulosic biomass by cellulase cocktails is a complex process modulated by multiple factors. Modifying the hydrolysis conditions, by adding, for instance, accessory enzymes (e.g., endoglucanases, non-reducing end specific CBHs, and LPMOs), or changing the substrate types and actual load will lead to an alteration of the rate-determining step from the non-catalytic enzyme-substrate dissociation to the catalysis for the hydrolysis in CBHs. Developing engineering strategies that modulate both the catalytic and non-catalytic processes is, therefore, of paramount importance to elevate their hydrolytic properties. In the present work, turning to large-scale computational modeling supported by experimental investigations, we are making a first attempt to improve the enzymatic catalysis as well as the non-catalytic process by a single-residue substitution, which is in stark contrast with previous strategies that only focus on one of the processes, thereby driving a brand-new direction for the enzyme engineering. In general, the present work is envisioned to contribute to advancing the properties of industrially important enzymes that hydrolyze complex natural substrates, and promote the economical and effective use of lignocellulosic biomass.

METHODS

Enzyme-Enzyme Docking. To investigate the interactions between enzymes, we performed protein-protein docking by HADDOCK Web Server (https://bianca.science.uu.nl/haddock2.4/)[22, 23] to obtain the optimal docking poses. The crystal structures of EG (PDB code 1KS5), *Tr*Cel7A (PDB code 4C4C, Q217 was mutated back to Glu) and *Tr*Cel6A (PDB code 1HGW, A175 was mutated back to Asp) were downloaded from the protein data bank (PDB). BG, secreted by *Penicillium piceum*, was modeled by homology, using SWISS-MODEL (swissmodel.expasy.org),[24-28] wherein PDB codes 3ZYZ and 4IIB, which have the highest homology with the target BG, were employed as the templates. To obtain as many combination patterns as possible, we selected all residues located on the enzyme surface as docking sites.

50,000 rigid docking simulations were carried out using the HADDOCK software version 2.4 for each assembly. The top 1,000 complex structures were then selected for semi-flexible refinement, followed by a final refinement by means of short MD simulations in water. The docking procedures were carried out following the HADDOCK protocols. The optimal docking pose of each enzyme-enzyme assembly was automatically provided by the software.

Simulation Design. To investigate the enzyme-enzyme interactions after docking, MD simulations were carried out for the individual *Tr*Cel7A and the *Tr*Cel7A-*Tr*Cel6A complexes. There is no substrate in each assembly. Then all the assemblies were solvated in an equilibrated box of water separately; the overall charge neutrality was achieved by adding Na^+ ions to the solution. Each molecular system before production simulations underwent two-step pretreatments, i.e., (i) 5000 steps minimization and 100 ps MD simulation with protein restrained, and (ii) 5 ns of water equilibration without restrained. A total of 2-μs MD simulations were carried out for the individual *Tr*Cel7A and the *Tr*Cel7A-*Tr*Cel6A complexes. Furthermore, to explore the effect of the designed five variants (D262N, D262R, D262I, D262W, and D262V) on the stability of the catalytic triad *in silico*, the Michaelis complex for each variant (PDB code 4C4C, Q217 was mutated back to Glu), which contains a polysaccharide chain (–7 to +2 glucosyl rings) in the catalytic tunnel ready for glycosylation, was constructed. After two-step pretreatments, a total of 6-μs MD simulations were conducted for the five variants and the WT *Tr*Cel7A assemblies.

In order to investigate the effect of the valine mutation on the non-catalytic enzyme-substrate dissociation, a cellobiose-free model was constructed by removing the cellobiose molecule from the initial structure of *Tr*Cel7A (PDB code 4C4D, Q217 was mutated back to Glu). After two-step pretreatments, a total of 1-μs MD simulations were performed for the D262V variant and the WT *Tr*Cel7A assemblies. All the brute-force MD simulations reported in the present work were conducted using the scalable program NAMD 3.0[29] with the CHARMM36 force field,[30-32] and the TIP3P water model.[33] PyMOL molecular graphics system (version 2.4 Schrödinger, LLC) and VMD 1.9.4 program[34] were employed to visualize and analyze the obtained MD trajectories. Each molecular system before production simulations underwent a standard operation procedure by referring to the work of Knott et al.[8] A total of 10 molecular

assemblies were simulated over an aggregate time of 9 μs, details of which are provided in Supplementary Table 5.

$\Delta G°$ Calculations. Each molecular assembly was loaded in the BFEE2 software to automatically generate all simulation input files. All calculations were carried out using the NAMD 3.0 software and followed by the published protocol.[35-37] Details of the molecular assemblies are shown in Table S2. The free-energy contributions provided in Table S2 were calculated by means of one-dimensional PMF calculations using the Gaussian-accelerated well-tempered meta-extended adaptive biasing force (GaWTM-eABF) algorithm.[38] The corresponding error was estimated by dividing the one-dimensional PMF calculation into two sub-runs of equal lengths, and by computing the standard deviation over the different PMFs.

Hydrogen Bonding Network Analysis. Eighteen residues located within 5 Å around the catalytic triad of *Tr*Cel7A (PDB code 4C4C, Q217 was mutated back to Glu) were selected. All the non-protein molecules were removed from the initial structure. Then the protein was relaxed by the Rosetta FastRelax Mover[39, 40] and scored using the REF2015 Rosetta energy function.[41] The number of the formed hydrogen bonds as well as the relevant energy were calculated using the Rosetta HbondMetric (implemented in RosettaScripts).[42] Four residues, which formed the least hydrogen bonds in the selected residues, were chosen for virtual saturation mutations. The seventy-two virtual variants were further evaluated by the above procedure *in silico*.

QM/MM Simulations. PMFs of glycosylation were determined for the WT *Tr*Cel7A (PDB code 4C4C, Q217 was mutated back to Glu) and the D262V variant in aqueous solution, by QM/MM simulations. The QM region of the WT assembly, which has an overall charge of –1, includes E212, D214, E217, as well as the –1 and +1 glucosyl units. The QM area of D262V variant, which has an overall charge of –1, contains the QM atoms of the WT and a D262V residue. The free energy profiles characterizing glycosylation were calculated following the approach reported by Knott et al.[8] The total calculative time amounted to 5 ns. Details of the molecular assemblies are depicted in Table S6.

Free Energy Calculations. PMFs describing the –1 glucosyl rings dissociation in the assemblies of

WT *Tr*Cel7A and D262V variant were determined using the umbrella sampling algorithm.[43] The dissociative distance of 8 Å was divided into 32 overlapping windows. Each overlapping window ran for 10 ns with constraint force constant of 5 kcal/mol·Å^2. The calculated PMFs were obtained using the weighted histogram analysis method (WHAM).[44] The total calculative time amounted to 1.92 μs. Detail of the molecular assemblies are provided in Table S7.

Construction of *Te*Cel7A and Variants of D259V and D259W. The CD of *Te*Cel7A (see the protein sequence in Figure S2) was employed to experimentally examine the mutation effect on the substrates. The WT *Te*Cel7A and the variants of D259V and D259W (corresponding to D262V and D262W in *Tr*Cel7A) were generated by total gene synthesis and connected to yeast expression vector pPICZαA (Invitrogen, Carlsbad, CA). The codons were optimized; His-tag was added to the C-terminus for purification. All the synthetic experiments were conducted at the Beijing Genomics Institute.

Enzyme Expression. Genes encoding the WT *Te*Cel7A and the variants of D259V and D259W were linearized by Sac I (Thermo Fisher Scientific, Runcorn, Cheshire, UK) and transformed into the protein production host organism *P. pastoris* strain X-33 (Invitrogen, Carlsbad, CA) by electroporation.[45] Transformants were spread on YPDS plates and incubated for three days at 30 °C. The single colonies of the WT *Te*Cel7A and the variants from the plates were inoculated into buffered complex medium containing glycerol (BMGY) at 30 °C until the cell turbidity OD reached to 2-6. The cells were obtained by centrifugation at 4,000 g for 15 min and further resuspended into a buffered minimal medium containing methanol (BMM). The cultures were then induced by supplementing 0.5 % methanol every 12 h and grew for three days at 30 °C and 240 rpm.

Enzyme Purification. Cultures were centrifuged at 4,000 g for 30 min; the supernatants for the WT *Te*Cel7A and the variants of D259V and D259W were adjusted to neutral and filtered by 0.45 μm polypropylene membranes (Seahorse Bioscience, North Bellerica, MA). The targeted proteins were purified by a cOmplete His-Tag Purification Resin (Roche, Mannheim, Germany); the purity was analyzed by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE; Figure S3).

Enzyme Kinetics. *p*NPC (Sigma, St. Louis, MO) was employed as the soluble substrate to measure

enzymatic kinetics, in concentrations ranging between 0.05 and 9.0 mM. The assay was done as described in our previous work.[9] Kinetic parameters (i.e., k_{cat} and K_m) were determined by nonlinear regression using GraphPad Prism 5 software (GraphPad Software Inc., San Diego, CA). All experiments were performed in triplicate and errors are reported as standard error of the mean.

Degradation of Microcrystalline Cellulose and Activity Assay. Each reaction volume consists of 10 mg/mL Avicel PH-101 (Sigma, St. Louis, MO) and 1.33 μM purified enzyme in 50 mM citrate buffer, pH 5.0. All assays were incubated for 24 h, 48 h, and 72 h at 60 °C with constant rotational mixing and finally boiled for 5 min at 95 °C to stop the reactions. All experiments were performed in triplicate. Then the samples were sedimented by centrifugation at 8,000g for 10 min and the supernatant filtered through 0.25 μm polypropylene membrane (Seahorse Bioscience, North Bellerica, MA). Ultra-Performance Liquid Chromatography (UPLC) system (Waters, Milford, Massachusetts, USA) with a refractive index (RI) detector equipped with a 250 × 4.6 mm (length × inner diameter) Supersil NH2-S 5 μm column (Elite, Dalian, China) was employed to quantify the cellobiose concentrations. Compounds were eluted at a flow rate of 1.0 mL by a mobile phase of ultrapure water with 65% acetonitrile. Quantification was performed by external calibration with a set of cellobiose solutions in a range of 10-500 mg/L.

ASSOCIATED CONTENT

Supporting Information

Additional Tables and Figures related to the Results and Discussion section (PDF).

Author Contributions

Z.Z., W.C., C.C., and Z.H. conceived the study. Z.Z. supervised the study. Z.Z., H.L., Y.D., X.W, and W.P. did the simulations and experiments. Z.Z., C.C, S.M., J.L., and L.L. drafted the manuscript and revised by all the authors.

Competing Interests

The authors declare no competing interests.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (22005157), National Key R&D Program of China (2023YFC3403600), and Tianjin Synthetic Biotechnology Innovation Capacity Improvement Project (TSBICIP-CXRC-077). Z.Z. gratefully acknowledges the financial support from China Scholarship Council (201706205014). Furthermore, the FP7 WeNMR (261572), H2020 West-Life (675858) and the EOSC-hub (777536) European e-Infrastructure projects are acknowledged for the use of their web portals, which make use of the EGI infrastructure with the dedicated support of CESNET-MetaCloud, INFN-PADOVA, NCG-INGRID-PT, TW-NCHC, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, Taiwan and the US Open Science Grid.

REFERENCES

(1) Liao, Y.; Koelewijn, S.-F.; Van den Bossche, G.; Van Aelst, J.; Van den Bosch, S.; Renders, T.; Navare, K.; Nicolaï, T.; Van Aelst, K.; Maesen, M.; Matsushima, H.; Thevelein, J. M.; Van Acker, K.; Lagrain, B.; Verboekend, D.; Sels, B. F., A Sustainable Wood Biorefinery for Low–Carbon Footprint Chemicals Production. *Science (New York, N.Y.)* **2020,** *367* (6484), 1385-1390.

(2) Chen, F.; Dixon, R. A., Lignin Modification Improves Fermentable Sugar Yields for Biofuel Production. *Nature biotechnology* **2007,** *25* (7), 759-61.

(3) 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 (New York, N.Y.)* **2007,** *315* (5813), 804-7.

(4) Bornscheuer, U.; Buchholz, K.; Seibel, J., Enzymatic Degradation of (Ligno)cellulose. *Angew. Chem. Int. Ed. Engl.* **2014,** *53* (41), 10876-93.

(5) Taylor, L. E., 2nd; Knott, B. C.; Baker, J. O.; Alahuhta, P. M.; Hobdey, S. E.; Linger, J. G.; Lunin, V. V.; Amore, A.; Subramanian, V.; Podkaminer, K.; Xu, Q.; VanderWall, T. A.; Schuster, L. A.; Chaudhari, Y. B.; Adney, W. S.; Crowley, M. F.; Himmel, M. E.; Decker, S. R.; Beckham, G. T., Engineering Enhanced Cellobiohydrolase Activity. *Nat. Commun.* **2018,** *9* (1), 1186.

(6) Payne, C. M.; Knott, B. C.; Mayes, H. B.; Hansson, H.; Himmel, M. E.; Sandgren, M.; Ståhlberg, J.; Beckham, G. T., Fungal Cellulases. *Chem. Rev.* **2015,** *115* (3), 1308-448.

(7) Knott, B. C.; Crowley, M. F.; Himmel, M. E.; Ståhlberg, J.; Beckham, G. T., Carbohydrate-Protein Interactions that Drive Processive Polysaccharide Translocation in Enzymes Revealed from a Computational Study of Cellobiohydrolase Processivity. *J. Am. Chem. Soc.* **2014,** *136* (24), 8810-9.

(8) Knott, B. C.; Haddad Momeni, M.; Crowley, M. F.; Mackenzie, L. F.; Götz, A. W.; Sandgren, M.; Withers, S. G.; Ståhlberg, J.; Beckham, G. T., The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies. *J. Am. Chem. Soc.* **2014,** *136* (1), 321-9.

(9) Zong, Z.; Li, Q.; Hong, Z.; Fu, H.; Cai, W.; Chipot, C.; Jiang, H.; Zhang, D.; Chen, S.; Shao, X., Lysine Mutation of the Claw-Arm-Like Loop Accelerates Catalysis by Cellobiohydrolases. *J. Am. Chem. Soc.* **2019,** *141* (36), 14451-14459.
(10) Vermaas, J. V.; Kont, R.; Beckham, G. T.; Crowley, M. F.; Gudmundsson, M.; Sandgren, M.; Ståhlberg, J.; Väljamäe, P.; Knott, B. C., The Dissociation Mechanism of Processive Cellulases. *Proc. Natl. Acad. Sci. USA* **2019,** *116* (46), 23061-23067.
(11) Kari, J.; Molina, G. A.; Schaller, K. S.; Schiano-di-Cola, C.; Christensen, S. J.; Badino, S. F.; Sørensen, T. H.; Røjel, N. S.; Keller, M. B.; Sørensen, N. R.; Kolaczkowski, B.; Olsen, J. P.; Krogh, K.; Jensen, K.; Cavaleiro, A. M.; Peters, G. H. J.; Spodsberg, N.; Borch, K.; Westh, P., Physical Constraints and Functional Plasticity of Cellulases. *Nat. Commun.* **2021,** *12* (1), 3847.
(12) Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttilä, M.; Ando, T.; Samejima, M., Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface. *Science (New York, N.Y.)* **2011,** *333* (6047), 1279-82.
(13) Zong, Z.; He, R.; Fu, H.; Zhao, T.; Chen, S.; Shao, X.; Zhang, D.; Cai, W., Pretreating Cellulases with Hydrophobins for Improving Bioconversion of Cellulose: An Experimental and Computational Study. *Green Chem.* **2016,** *18* (24), 6666-6674.
(14) Gumbart, J. C.; Roux, B.; Chipot, C., Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013,** *9* (1), 794-802.
(15) Woo, H. J.; Roux, B., Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. USA* **2005,** *102* (19), 6825-6830.
(16) Bu, L.; Crowley, M. F.; Himmel, M. E.; Beckham, G. T., Computational Investigation of the pH Dependence of Loop Flexibility and Catalytic Function in Glycoside Hydrolases. *J. Biol. Chem.* **2013,** *288* (17), 12175-86.
(17) Kurasin, M.; Väljamäe, P., Processivity of Cellobiohydrolases is Limited by the Substrate. *J. Biol. Chem.* **2011,** *286* (1), 169-77.
(18) Sørensen, T. H.; Windahl, M. S.; McBrayer, B.; Kari, J.; Olsen, J. P.; Borch, K.; Westh, P., Loop Variants of the Thermophile Rasamsonia emersonii Cel7A with Improved Activity against Cellulose. *Biotechnol. Bioeng.* **2017,** *114* (1), 53-62.
(19) Bansal, P.; Hall, M.; Realff, M. J.; Lee, J. H.; Bommarius, A. S., Modeling Cellulase Kinetics on Lignocellulosic Substrates. *Biotechnol. Adv.* **2009,** *27* (6), 833-848.
(20) Cruys-Bagger, N.; Tatsumi, H.; Ren, G. R.; Borch, K.; Westh, P., Transient Kinetics and Rate-Limiting Steps for the Processive Cellobiohydrolase Cel7A: Effects of Substrate Structure and Carbohydrate Binding Domain. *Biochemistry* **2013,** *52* (49), 8938-48.
(21) Hobdey, S. E.; Knott, B. C.; Haddad Momeni, M.; Taylor, L. E.; Borisova, A. S.; Podkaminer, K. K.; VanderWall, T. A.; Himmel, M. E.; Decker, S. R.; Beckham, G. T.; Ståhlberg, J.; Cullen, D., Biochemical and Structural Characterizations of Two Dictyostelium Cellobiohydrolases from the Amoebozoa Kingdom Reveal a High Level of Conservation between Distant Phylogenetic Trees of Life. *Appl. Environ. Microb.* **2016,** *82* (11), 3395-3409.
(22) van Zundert, G. C. P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A., The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016,** *428* (4), 720-725.
(23) Honorato, R. V.; Koukos, P. I.; Jiménez-García, B.; Tsaregorodtsev, A.; Verlato, M.; Giachetti, A.; Rosato, A.; Bonvin, A., Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **2021,** *8*, 729513.
(24) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018,** *46*, 296-303.
(25) Bienert, S.; Waterhouse, A.; de Beer, T. A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository-New Features and Functionality. *Nucleic Acids Res.* **2017,** *45*, 313-319.
(26) Guex, N.; Peitsch, M. C.; Schwede, T., Automated Comparative Protein Structure Modeling with SWISS-MODEL and Swiss-PdbViewer: a Historical Perspective. *Electrophoresis* **2009,** *30*, 162-73.
(27) Studer, G.; Rempfer, C.; Waterhouse, A. M.; Gumienny, R.; Haas, J.; Schwede, T., QMEANDisCo-Distance

Constraints Applied on Model Quality Estimation. *Bioinformatics (Oxford, England)* **2020,** *36* (6), 1765-1771.
(28) Bertoni, M.; Kiefer, F.; Biasini, M.; Bordoli, L.; Schwede, T., Modeling Protein Quaternary Structure of Homo- and Hetero-Oligomers Beyond Binary Interactions by Homology. *Sci. Rep.* **2017,** *7* (1), 10480.
(29) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *Journal of computational chemistry* **2005,** *26* (16), 1781-802.
(30) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr., CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *Journal of computational chemistry* **2010,** *31* (4), 671-90.
(31) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-616.
(32) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd, Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004,** *126* (3), 698-9.
(33) 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* (2), 926-935.
(34) Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.
(35) Coderc de Lacam, E. G.; Blazhynska, M.; Chen, H.; Gumbart, J. C.; Chipot, C., When the Dust Has Settled: Calculation of Binding Affinities from First Principles for SARS-CoV-2 Variants with Quantitative Accuracy. *J. Chem. Theory Comput.* **2022,** *18* (10), 5890-5900.
(36) Fu, H.; Chen, H.; Blazhynska, M.; Goulard Coderc de Lacam, E.; Szczepaniak, F.; Pavlova, A.; Shao, X.; Gumbart, J. C.; Dehez, F.; Roux, B.; Cai, W.; Chipot, C., Accurate Determination of Protein:Ligand Standard Binding Free Energies from Molecular Dynamics Simulations. *Nat. Protoc.* **2022,** *17* (4), 1114-1141.
(37) Blazhynska, M.; Goulard Coderc de Lacam, E.; Chen, H.; Roux, B.; Chipot, C., Hazardous Shortcuts in Standard Binding Free Energy Calculations. *J. Phys. Chem. Lett.* **2022,** *13* (27), 6250-6258.
(38) Chen, H.; Fu, H.; Chipot, C.; Shao, X.; Cai, W., Overcoming Free-Energy Barriers with a Seamless Combination of a Biasing Force and a Collective Variable-Independent Boost Potential. *J. Chem. Theory Comput.* **2021,** *17* (7), 3886-3894.
(39) Khatib, F.; Cooper, S.; Tyka, M. D.; Xu, K.; Makedon, I.; Popovic, Z.; Baker, D.; Players, F., Algorithm Discovery by Protein Folding Game Players. *Proc. Natl. Acad. Sci. USA* **2011,** *108* (47), 18949-53.
(40) Maguire, J. B.; Haddox, H. K.; Strickland, D.; Halabiya, S. F.; Coventry, B.; Griffin, J. R.; Pulavarti, S.; Cummins, M.; Thieker, D. F.; Klavins, E.; Szyperski, T.; DiMaio, F.; Baker, D.; Kuhlman, B., Perturbing the Energy Landscape for Improved Packing During Computational Protein Design. *Proteins* **2021,** *89* (4), 436-449.
(41) Park, H.; Bradley, P.; Greisen, P., Jr.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; DiMaio, F., Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016,** *12* (12), 6201-6212.
(42) Fleishman, S. J.; Leaver-Fay, A.; Corn, J. E.; Strauch, E. M.; Khare, S. D.; Koga, N.; Ashworth, J.; Murphy, P.; Richter, F.; Lemmon, G.; Meiler, J.; Baker, D., RosettaScripts: a Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PloS one* **2011,** *6* (6), e20161.
(43) Torrie, G. M.; Valleau, J. P., Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977,** *23* (2), 187-199.
(44) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of computational chemistry* **1992,** *13* (8), 1011-1021.
(45) Becker, D. M.; Guarente, L., High-Efficiency Transformation of Yeast by Electroporation. *Methods Enzymol.* **1991,**

194, 182-7.