

RESEARCH

Open Access



# Dynamics of loops surrounding the active site architecture in GH5\_2 subfamily *TfCel5A* for cellulose degradation

Xiuyun Wu<sup>1</sup>, Sha Zhao<sup>1</sup>, Zhennan Tian<sup>1</sup>, Chao Han<sup>3</sup>, Xukai Jiang<sup>2</sup> and Lushan Wang<sup>1\*</sup>

## Abstract

**Background** Lignocellulose is the most abundant natural biomass resource for the production of biofuels and other chemicals. The efficient degradation of cellulose by cellulases is a critical step for the lignocellulose bioconversion. Understanding the structure-catalysis relationship is vital for rational design of more stable and highly active enzymes. Glycoside hydrolase (GH) family 5 is the largest and most functionally diverse group of cellulases, with a conserved TIM barrel structure. The important roles of the various loop regions of GH5 enzymes in catalysis, however, remain poorly understood.

**Results** In the present study, we investigated the relationship between the loops surrounding active site architecture and its catalytic efficiency, taking *TfCel5A*, an enzyme from GH5\_2 subfamily of *Thermobifida fusca*, as an example. Large-scale computational simulations and site-directed mutagenesis experiments revealed that three loops (loop 8, 3, and 7) around active cleft played diverse roles in substrate binding, intermediate formation, and product release, respectively. The highly flexible and charged residue triad of loop 8 was responsible for capturing the ligand into the active cleft. Severe fluctuation of loop 3 led to the distortion of sugar conformation at the – 1 subsite. The wobble of loop 7 might facilitate product release, and the enzyme activity of the mutant Y361W in loop 7 was increased by approximately 40%.

**Conclusion** This study unraveled the vital roles of loops in active site architecture and provided new insights into the catalytic mechanism of the GH5\_2 cellulases.

**Keywords** Cellulase, GH5\_2 subfamily, Active site architecture, Loop dynamics, Enzyme catalysis

## Introduction

Cellulose, an insoluble linear polysaccharide composed of D-glucose units linked by  $\beta$ -1,4-glycosidic bonds, is the most abundant natural biopolymer on earth and ranges in length from 2000 to 25,000 glucose residues [1]. A crucial factor that prevents the conversion of cellulose to other molecules is the recalcitrant nature of its crystalline region [2]. The complete breakdown of cellulose requires the synergistic hydrolytic action of three main enzymes:  $\beta$ -1,4-endoglucanase (EC 3.2.1.4), cellobiohydrolase (exoglucanase; EC 3.2.1.91 and EC 3.2.1.176), and  $\beta$ -glucosidase (EC 3.2.1.21) [3]. The crystalline surface of cellulose can also be disrupted by the oxidative action of

\*Correspondence:

Lushan Wang

lswang@sdu.edu.cn

<sup>1</sup> State Key Laboratory of Microbial Technology, Institute of Microbial Technology, Shandong University, Qingdao 266237, China

<sup>2</sup> National Glycoengineering Research Center, Shandong University, Qingdao 266237, China

<sup>3</sup> Shandong Key Laboratory of Agricultural Microbiology, Shandong Agricultural University, Tai'an 271018, China



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

lytic polysaccharide monooxygenases [4]. Among these enzymes,  $\beta$ -1,4-endoglucanase plays a key role in cellulose degradation. According to the Carbohydrate-Active enZymes database (CAZy; <http://www.cazy.org/>),  $\beta$ -1,4-endoglucanases are classified into glycoside hydrolase (GH) families 5, 6, 7, 8, 9, 10, 12, 26, 44, 45, 48, 51, 124, and 148, which randomly cleave internal glycosidic bonds in amorphous cellulose [5]. Several cellulases have been used as fundamental ingredients in industrial enzyme cocktails to degrade cellulose; however, the intrinsic instability and low catalytic activity of cellulases are the major hurdles in their widespread application in biorefinery industries [6, 7]. Therefore, it is essential to explore the catalytic mechanism of cellulases for efficient cellulose conversion [8, 9].

The GH5 family is one of the largest GH families and the first cellulase family to be described, and it is historically known as “cellulase family A” [10]. The GH5 family has a long evolutionary history, and its members are widely distributed in the kingdoms of bacteria, archaea, and eukaryotes. In recent years, with the accumulation of sequence data, a more detailed subfamily classification has emerged in the GH5 family [11]. In particular, a new GH5 subfamily classification system has been proposed based on the criterion of sequence identity >75%, and in this new system, 51 subfamilies cover more than 80% of the sequences. Currently, there are 56 subfamilies in the GH5 family according to their phylogenetic relationships.  $\beta$ -1,4-endoglucanases are the key members of the GH5 family and are distributed in nine major subfamilies (GH5\_1, 5\_2, 5\_4, 5\_5, 5\_22, 5\_25, 5\_26, 5\_37, and 5\_39 subfamilies). Some subfamilies of enzymes were multifunctional, indicating that mutations of only a few amino acids could lead to functional divergence [11]; this feature was also confirmed by the modification of the GH5\_4 subfamily [12]. Therefore, the structural basis of substrate recognition has become a promising area of research as scientists are seeking to further understand the functional diversity of the GH5 family [13, 14].

GH5 enzymes possess a highly conserved overall structure, namely a canonical  $(\beta/\alpha)_8$  TIM barrel fold with eight loops surrounding the catalytic cleft [15]. This structural fold provides abundant freedom in variation for the loops, which leads to the catalytic diversity of the proteins, including substrate specificity, hydrolytic activity, thermostability, and optimal pH [16]. Recently, several research studies have focused on investigating the connecting loop regions of enzymes to improve the performance of enzyme molecules. The structural arrangements of loops in the GH5\_4 subfamily was found to contribute to the breadth of substrate selectivity and hydrolytic reactions [17]. Zheng et al. enhanced the catalytic activity of *GiCel5* by mutating

the asparagine on loop 6 to alanine or glycine, and the molecular dynamics (MD) simulation suggested that the increased flexibility of loop 6 improved the interactions between the enzyme and the substrate [5]. The loops have also been observed to affect the catalytic properties of enzymes in other GH families. For instance, Yang et al. revealed that the longer loop 3 of GH12 endoglucanases might strengthen the hydrogen network of the active site architecture and thus increase the catalytic efficiency of the enzymes [18]. Designing the flexible region of enzymes could counteract the stability-activity trade-off and enhance the thermostability and catalytic activity [19, 20].

Loop conformational dynamics, particularly “lid loop,” play an important role in the process of substrate selection and recognition as well as in mediating substrate entry into the active clefts of the enzyme [21, 22]. Previous studies have shown that the difference between the state of the enzyme and substrate before and after binding lies mainly in the conformational difference of the loop structure in the active center region [23]. To elucidate how loops in the active site architecture regulate the catalytic activity of enzymes, the structure and dynamics of the loops should be studied. Given the limitations of using experiments alone to probe this aspect, MD simulations provide a visual model to show possible connections between the structure and dynamics of loops by exploring various conformational ensembles. We previously performed MD simulations under continuous systems to investigate changes in the protein dynamics of *TfCel5A* during substrate binding; the findings indicated that conformational imbalance caused by substrate binding was likely to be an important trigger of attenuation observed in cellulose hydrolysis [24]. In another study, we identified a salt-sensitive loop in GH5\_2 enzymes using MD simulations, which facilitated the rational design of novel halotolerant enzymes [25]. Moreover, the flexibility of the tunnel-forming loops could be significantly affected by pH, thereby resulting in changes in processivity and substrate complexation [26].

In the present study, we integrated structural bioinformatics, MD simulations, and site-directed mutagenesis to investigate the role of loops in the active site architecture of the GH5\_2 cellulase *TfCel5A*, an enzyme from *Thermobifida fusca*. The experiments conducted at atomic and molecular scales revealed that loops 3, 4, 7, and 8 played different roles in enzymatic catalysis, such as initial substrate binding, intermediate formation, and product release. Interestingly, the functions of these loops were highly dependent on their unique structural dynamics following interaction with the cellulose substrate. These findings provided mechanistic insights into the catalytic activity of GH5\_2 cellulases and a new

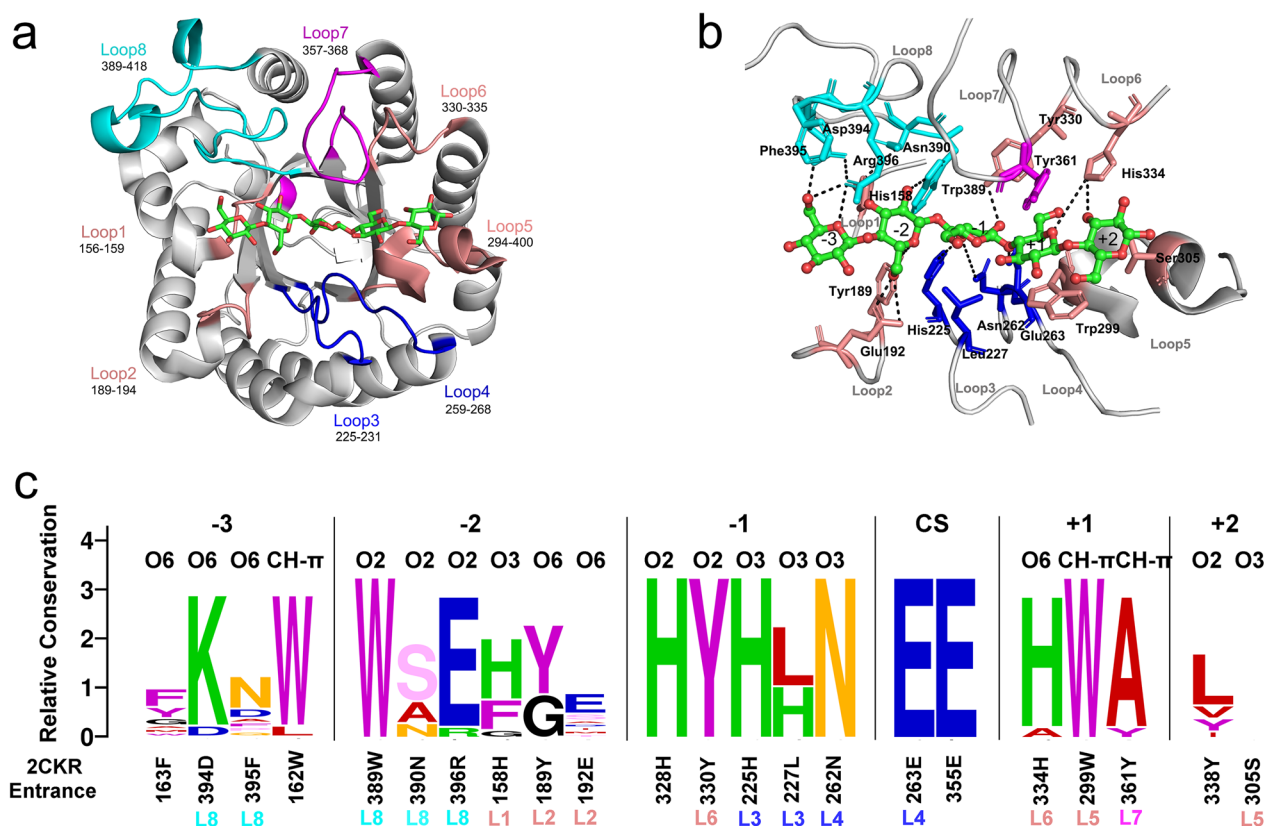
mechanism-based strategy to rationally design cellulases for biomass conversion.

## Results

### Loop features of the active site architecture of GH5\_2 cellulases

Based on the CAZy database, six subfamilies of the GH5 family were analyzed for evolution. Among these subfamilies, the GH5\_2 subfamily had the largest number of characterized enzymes and crystalline structures (Additional file 1: Fig. S1). To show the loop features of the active site architecture of GH5\_2 cellulases, a representative cellulase *TfCel5A* (PDB: 2CKR) from *T. fusca* was used as the reference [24, 25]. The crystalline structure of *TfCel5A* was solved with a cellopentose chain in its active cleft. As shown in Fig. 1a, the active site architecture comprised eight loops, named loop 1 to loop 8 according to the number of residues. The entire active site was divided into five subsites, spanning from -3

subsite (non-reducing end) to +2 subsite (reducing end) according to the cleaving site of the cellulose chain. The interaction network between 17 amino acid residues in the loops and cellopentose was analyzed (Fig. 1b). The catalytic residues were 263E and 355E, with 263E located on loop 4. Bortoli-German et al. clarified that Glu133 and Glu220 in *cel5A* from *Bacillus agaradhaerens* served as proton donors and nucleophiles, respectively, to form glycosylase intermediates [27]. The five amino acids (394D, 395F, 389W, 390N, and 396R) of loop 8 formed hydrogen bond interactions with the sugar rings at the -3 and -2 positions. In addition, hydrogen bond interactions with amino acids at the -2 subsite also included 158H at loop 1 and 189Y and 192E at loop 2. The amino acids 225H at loop 3, 227L at loop 6, 330Y at loop 4, and 262N at loop 4 formed interactions with the -1 sugar ring. At the +1 and +2 subsites, the amino acids 299W and 305S at loop 5, 334H at loop 6, and 361Y at loop 7 formed interactions with the ligands. Moreover, the



**Fig. 1** Loop analysis of the active site architecture of GH5\_2 cellulases. **a** *TfCel5A*-cellopentose complex. The loops of the active site architecture are shown in colors. Loops 1, 2, 5, and 6 are shown in salmon; loops 3 and 4 are shown in blue; and loops 7 and 8 are shown in magenta and cyan, respectively. Cellopentose in the active cleft is shown in green. **b** Hydrogen bond network between the residues in loops 1–8 and the ligand. The subsites of the active cleft are labeled from -3 to +2. **c** Sequence profile of the active site architecture of GH5\_2 enzymes. Based on the cut-off of 5 Å, the amino acid residues surrounding cellopentose in the *TfCel5A*-cellopentose complex were selected to generate the reference using PyMOL software. The size of the letter indicates the degree of conservation of the residues. The sugar atoms interacting with the enzyme are labeled above the residue

hydrogen-bonding networks at  $-3$  to  $-1$  subsites were roughly the same as that of cel5A from *Bacillus agaradhaerens* analyzed by atomic resolution [28]. The difference is that the polar amino acids that interact with the  $-3$  and  $-2$  sugar rings are 267K, 269E in cel5A, while 394D and 396R in *TfCel5A*.

To validate the generality of the significance, we used a structural bioinformatics method to analyze the evolution of the active site residues in the GH5\_2 subfamily (Fig. 1c). The amino acid residues at the  $-1$  and  $-2$  subsites were more conserved in the family than those at the  $-3$ ,  $+1$ , and  $+2$  subsites. Among these residues, 225H at loop 3 (L3), 262N and 263E at loop 4 (L4), 299W at loop 5 (L5), 330Y at loop 6 (L6), and 389W at loop 8 (L8) were absolutely conserved in the GH5\_2 subfamily. Moreover, 394D and 396R at loop 8, 334H at loop 6, and 361Y at loop 7 (L7) were relatively conserved. In cel5A, conserved 98H (225H in *TfCel5A*), 192H (328H in *TfCel5A*) and 198H (334H in *TfCel5A*) were proved to be essential for the catalytic activity of the enzyme, but do not affect its *in vivo* stability, while non-conserved Arg residues played an important role in protein stabilization [27, 29]. In addition, our previous results suggested that alanine substitution of conserved residues of xylanases from GH10 and GH11 family lead to a significant decrease in enzymatic activity [30, 31].

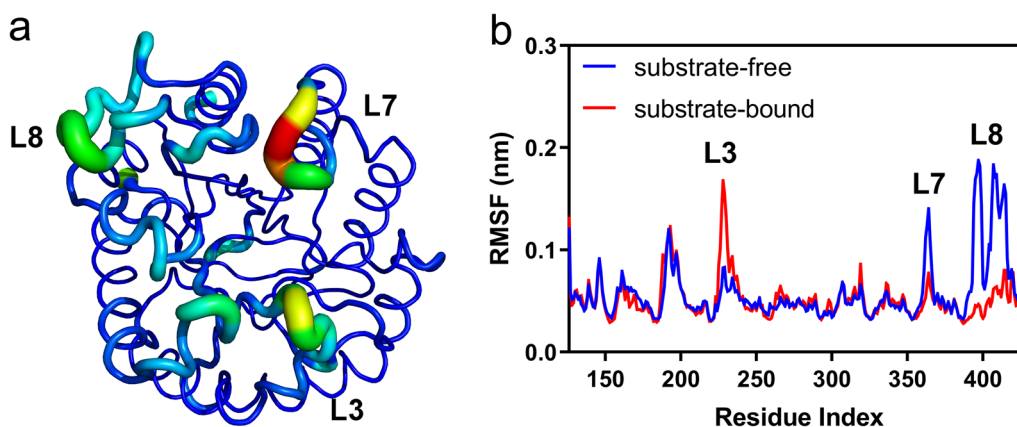
#### Analysis of flexibility of loops in *TfCel5A*

To further investigate the structural flexibility of *TfCel5A*, particularly the dynamics of key loops of the active site architecture, we performed b-factor analysis and RMSF calculation. The b-factor values reflect the “diffusion” of the electron density of the atoms in the crystal, which actually reflects the conformational state

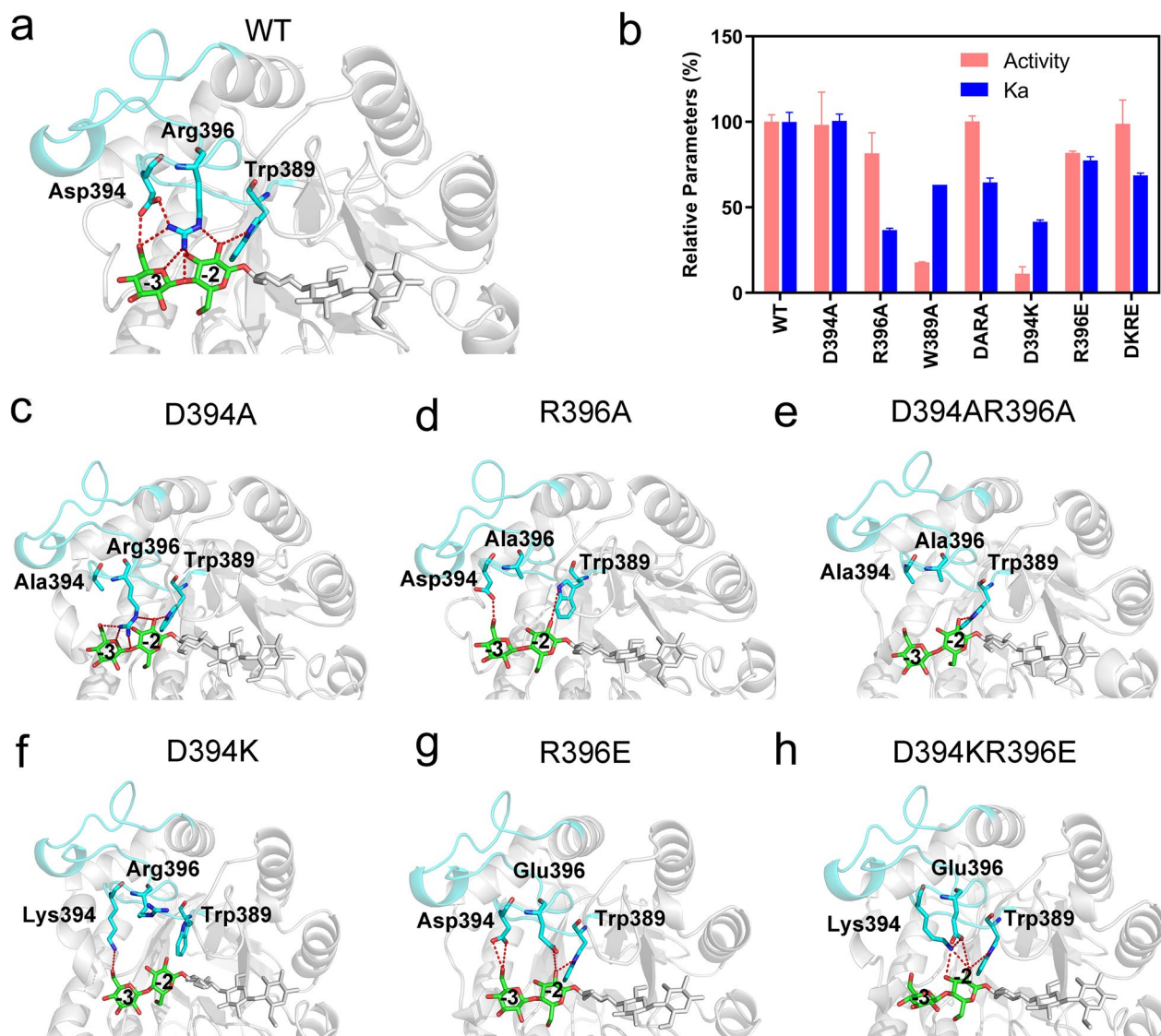
of the protein molecule in the crystal [32]. The b-factors of *TfCel5A* (PDB: 2CKR) overall structure were shown in Fig. 2a, and represented by the thickness and color of the displayed tubes. The b-factor values of loops 3, 7, and 8 were high, indicating that these three loops have high flexibility. We also calculated the root mean square fluctuation (RMSF) changes for the two simulation systems: substrate-free state and substrate-bound state. Although the RMSF values of most protein regions remained almost unchanged, the RMSF values of several regions increased and decreased significantly. In particular, three regions (loops 3, 7, and 8) always showed the largest RMSF change between the substrate-free and substrate-bound states (Fig. 2b). Moreover, the corresponding loops in several other GH5\_2 enzymes exhibited similar structural dynamics (Additional file 1: Fig. S2). This indicated that the unique dynamics of these loops might be a key factor in the catalysis of GH5 enzymes. Importantly, our large-scale MD simulations of 21 enzymes from 6 different GH5 subfamilies revealed that loop 7 and 8 always showed high flexibility (Additional file 1: Fig. S3).

#### Loop 8 mediated the initial substrate binding

Interestingly, according to local structural analysis, loop 8 contains an acidic amino acid (394D in *TfCel5A*) and a basic amino acid (396R in *TfCel5A*) that appeared adjacent to the fully conserved 389W (Fig. 3a and Additional file 1: Fig. S4). The residue triad in loop 8 was located at the non-reducing end of the active site and formed extensive interactions with the sugars at the  $-3$  and  $-2$  subsites. 394D formed a hydrogen bond with 396R and a hydrogen bond with the O6'-hydroxyl group of glucose at the  $-3$  subsite; 396R formed four hydrogen bonds with glucose at the  $-3$  and  $-2$  subsites; and



**Fig. 2** Structural flexibility of *TfCel5A*. **a** B-factor analysis of the entire structure of *TfCel5A*. Visualizing relative disorder or uncertainty in atomic positions is done by coloring by temperature value. Atoms with low temperature values are colored blue, while atoms with high temperature values are colored red. **b** Comparison of the RMSF values between the substrate-free (blue line) and substrate-bound states (red line). Locations of loops 3, 7, and 8 are labeled



**Fig. 3** Interaction between residues in loop 8 and the substrate. **a** The hydrogen bonding network between wild-type *Tf*Cel5A and the substrate at the – 3 and – 2 subsites. **b** Measurements of the catalytic activity relative to WT *Tf*Cel5A and substrate binding affinity of the loop 8 mutants. DARA and KDRE represent double mutants D394A\_R396A and D394K\_R396E, respectively. **c–h** The hydrogen bonding interactions of loop 8 and the substrate revealed by the MD simulations of the mutants

389W formed a hydrogen bond with glucose at the – 2 subsite. Considering the functional importance of the loops in enzymes with a TIM barrel fold [5, 33, 34], we hypothesized that the residue triad in loop 8 was likely to be a unique motif essentially required for the catalysis of *Tf*Cel5A. Therefore, we constructed a series of mutants with mutations at positions 394, 396, and 389 and measured the substrate binding affinity and catalytic activity of the mutants (Fig. 3b). Alanine substitution of 389W sharply decreased the catalytic activity of *Tf*Cel5A by 82.2%, which reflected the critical role of 389W in loop 8 of *Tf*Cel5A. Single mutants D394A and R394A and the

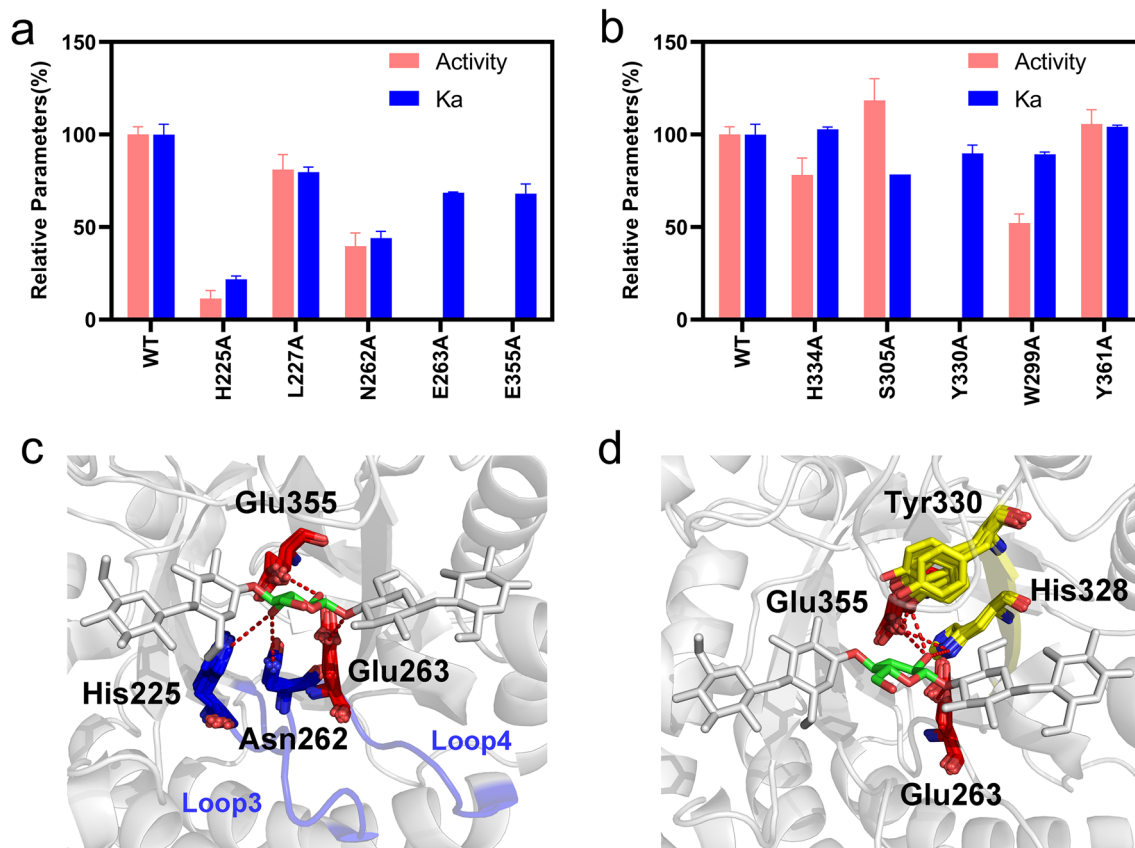
double mutant D394A\_R394A showed catalytic activities similar to that of the wild type (WT). However, the catalytic activity and substrate binding affinity of D394K significantly decreased by 82.3% and 59.4%, respectively. In contrast, the catalytic activity and substrate binding affinity of R396E decreased by approximately 20%. Interestingly, when we swapped the electrochemical properties at positions 394 and 396, the double mutant D394K\_R396E showed similar catalytic activity to the WT. Collectively, these results indicated that the unique residue motif in loop 8 played a vital role in substrate binding and catalytic activity of *Tf*Cel5A. The enzymatic

activities and binding force changes of the other alanine mutations that interact with the sugars at positions  $-3$  and  $-2$  are shown in Additional file 1: Fig. S5. Following alanine mutation, the enzyme activities remained above 60% of the activity of the wild-type enzyme, while the enzyme activities of 395F and 192E remained unchanged.

To further investigate the functional mechanism of the residue triad in loop 8, we performed a series of MD simulations for each mutant. The simulations with D394A showed that 386R and 389W formed extensive hydrogen bonding interactions with the substrate (Fig. 3c) and this finding explained the experimentally measured non-significantly different activity and binding affinity between WT and D394A. In the simulations with R396A, 394D and 389W formed two hydrogen bonds with the sugars (Fig. 3d), resulting in a decrease in its substrate-binding affinity. In the simulations with D394A\_R396A, only 389W formed a hydrogen bond with the sugar at the  $-2$  subsite (Fig. 3e), while the activity of D394A\_R396A surprisingly remained similar to that of the WT. In the simulations with D394K, the electrostatic repulsion

between 394 K and 396R changed the orientation of the sidechains of both 396R and 389W (Fig. 3f); this perturbed the hydrogen bonding network at the  $-2$  and  $-3$  subsites and significantly decreased the catalytic activity of D394K. In the simulations with R396E and D394K\_R396E, loop 8 maintained an extensive hydrogen bonding network with the substrate (Fig. 3g and h), which might account for their similar activities to the WT. The unique structural dynamics enabled the residue triad (i.e., 389W, 394D, and 396R in *TfCel5A*) to search for the substrate in the solution and finally captured the cellulose chain into the active cleft of GH5\_2 cellulases.

**Loops 3 and 4 promoted the formation of the intermediate** To discover the functional determinants of the  $-1$  and  $+1$  subsites in the active site architecture, we constructed mutants with alanine-scanning substitutions and measured the substrate binding affinity and catalytic activity of the mutants (Fig. 4a and b). Compared to the wild-type *TfCel5A*, the alanine substitutions at almost all positions decreased the substrate binding affinity and



**Fig. 4** Functional analysis of key residues in loops 3, 4, and 6. **a** Measurements of the catalytic activity relative to WT *TfCel5A* and substrate binding affinity of loop 3 and 4 mutants. **b** Measurements of the catalytic activity relative to WT *TfCel5A* and substrate binding affinity of loop 5, 6, and 7 mutants. **c** Hydrogen bonding interaction between the active site residues and the O3'-hydroxyl group of the sugar at the  $-1$  subsite. **d** Hydrogen bonding interaction between the active site residues and the O2'-hydroxyl group of the sugar at the  $-1$  subsite

catalytic activity of the mutants, which confirmed the significance of the active site residues in catalysis. Notably, the alanine substitutions at the  $-1$  subsite dramatically decreased the catalytic activities by 88.5% (H225A), 60.5% (N262A), and 100% (Y330A), in addition to the complete inactivation of the catalytic residues by the mutation.

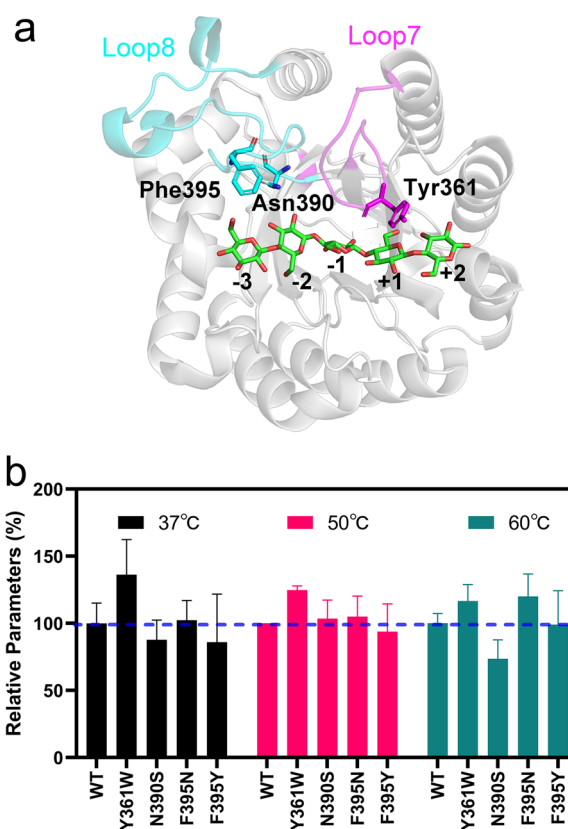
A twisted-boat intermediate conformation of the sugar at the  $-1$  subsite is required for the catalysis of GH5 enzymes [35]. In the *TfCel5A*-substrate complex, the sidechains of the two absolutely conserved residues 225H and 262N formed multiple hydrogen bonds with the O3'-hydroxyl group of the sugar at the  $-1$  subsite (Fig. 4c). 225H and 262N were located at loops 3 and 4, respectively. Interestingly, the MD simulations revealed that loop 3 became more flexible following substrate binding, whereas loop 4 remained rigid in the absence or presence of the substrate (Fig. 2b). Because of the difference in the flexibility of loops 3 and 4, 225H moved frequently, whereas 262N remained immobile. Therefore, the stable hydrogen bonding interaction formed between 262N and the sugar at the  $-1$  subsite might function as an "anchor" of the sugar, whereas the severe fluctuation of 225H led to the distortion of sugar conformation at the  $-1$  subsite. The experimental results of H225A and N262A mutants and the conservation of both residues in the GH5 family further confirmed their key roles in the catalysis of GH5 cellulases.

### Stabilization of the catalytic residues

Following the formation of the intermediate, two catalytic residues (e.g., 263E and 355E in *TfCel5A*) initiated the hydrolysis of the glycosidic bond between the sugars at the  $-1$  and  $+1$  subsites, which required the precise coordination of the catalytic residues. We found that 328H and 330Y on  $\beta$ -sheet 6 of *TfCel5A* were close to the catalytic residues and highly conserved in the GH5 family (Fig. 1c). In the MD simulations, 328H and 330Y formed two hydrogen bonds with the O2'-hydroxyl group of the sugar at the  $-1$  subsite and formed two hydrogen bonds with the sidechain of the nucleophile residue 355E (Fig. 4d). Importantly, 328H and 330Y showed low flexibility during the simulations, which stabilized the carboxyl group of 355E and ensured the precise orientation of the catalytic residues. This explained the significant decrease in the catalytic activity of the Y330A mutant. Additionally, the phenolic hydroxyl group of 330Y helps regulate the deprotonated state of the carboxyl group of 355E with the support of 185R [36]. Taken together, these results indicated that 330Y in loop 6 assisted the precise positioning and protonation state of the nucleophile residues during the catalysis of *TfCel5A*.

### Loop 7 regulated product release

After the hydrolysis of the glycosidic bond between the sugars at the  $-1$  and  $+1$  subsites, the enzyme needs to release the cleaved oligosaccharide product from the active site and initiate the next catalytic reaction. Interestingly, we found that loop 7 of GH5\_2 cellulases was often highly flexible in the substrate-free state but remained stable in the substrate-bound state (Fig. 2b). Notably, Loop 7 interacts with the substrate via a CH- $\pi$  interaction between 361Y and the pyranose. 361Y is not conserved in the GH5 family and the alanine substitution of 361Y exerted a minor effect on the catalytic activity (Figs. 1c and 4b). We then performed a series of mutations targeting the non-conserved amino acids at loops 7 and 8 (Fig. 5a). Unlike the amino acid mutants at loop 8, the activity of Y361W increased by approximately 40% as compared to that of the WT at 37 °C and was approximately 20% higher than that of the WT at 50 °C. At 60 °C, the mutations of Y361W and F395N increased the catalytic activity by approximately 16% and 20%, respectively



**Fig. 5** Comparison of catalytic activities between the WT and its mutants at 37 °C, 50 °C, and 60 °C. **a** The amino acid positions selected for site-directed mutations. Loops 7 and 8 are shown in red and cyan, respectively. The subsites of the active site are labeled from  $-3$  to  $+2$ . **b** The relative catalytic activity of *TfCel5A* mutants at different temperatures

(Fig. 5b). Considering that product release is a key limitation of enzyme catalysis at low temperatures [37], the measurements suggest that the mutation of Y361W benefited the release of the product and thereby significantly enhanced the activity at low temperatures. Interaction energy calculation showed the order of binding force was WT > Y361W > Y361A (Additional file 1: Fig. S6). These results indicated that depending on the swing motion of loop 7, the conformational transition might promote the release of the hydrolyzed product from the reducing end of the active site of *TjCel5A*.

## Discussion

GH5 cellulases are the key components of enzyme cocktails used for hydrolyzing lignocellulosic biomass in modern biorefinery industries. Generally, a classical two-step, retaining mechanism is used to achieve the catalysis [35, 38]. Upon substrate binding to the active site of the enzymes, a glycosyl-enzyme intermediate is rapidly formed through the glycosylation process. Subsequently, the nucleophile residue (e.g., E355 in *TjCel5A*) and the acid/base residue (e.g., E263 in *TjCel5A*) use a water molecule to attack the anomeric carbon of the sugar at the -1 subsite and finally cleave the glycosidic bond [15, 39]. Despite the basic understanding of the catalytic mechanism of cellulases [40], the dynamic behavior of key loops in active site architecture and their functions in substrate binding, catalysis and product release are relatively little studied.

An in-depth analysis of the active site architecture is a useful strategy to clarify the catalytic mechanism [41–43]. The unique spatial arrangement of the amino acid residues in the enzyme active site determines the specificity and efficiency of substrate degradation [44]. Using GH5\_2 cellulases as a model, we systematically investigated their structure-catalysis relationship by integrating multiscale approaches and proposed a theoretical model of the entire catalytic process (Fig. 6).

At the -3 and -2 subsites, depending on the high flexibility of loop 8, a residue triad (389W, 394D, and 396R in *TjCel5A*) functioned as a “grapple-hook” to search for and capture the cellulose chain into the active cleft. Given that the residue triad sits at the entrance of the active site, the interaction of the triad with the substrate might trigger the subsequent binding of the substrate to the enzyme cleft in a zipper-like manner (Fig. 6, Step 1). The experiments of the site-directed mutations at 390N and 395F demonstrated the availability of potential chemical space to optimize the sequence motif of loop 8 for better catalytic activity (Fig. 5b). This provides a promising direction to improve the enzyme activity by modifying the architecture at the -3 and -2 subsites of the active site [37, 45]. When the enzyme-substrate

complex was formed, the flexibility of loop 3 substantially increased, and 225H contributed to the hydrogen bonding interaction with the O3'-hydroxyl group of the sugar at the -1 subsite (Fig. 4c). Through the anchoring effect of 262N at loop 4, the vibration of 225H strongly stimulated the transition of the sugar ring from a chair conformation to a twist-boat conformation (Fig. 6, Step 2), which mediated the formation of a glycosyl-enzyme intermediate [35]. Subsequently, the glycosidic bond between the sugars at the -1 and +1 subsites was cleaved, which requires the precise orientation and protonation state of the catalytic residues [15]. Our results revealed that 328H and 330Y stabilized the side chain of the nucleophile residue (Fig. 6, Step 3) and finally achieved the hydrolysis reaction in cooperation with the acid/base residue [40]. Although the residues at the +1 and +2 subsites were non-conserved in the GH5 family, the motion of loop 7 facilitated product release by generating steric repulsion between the enzyme and the hydrolyzed product (Fig. 6, Step 4). The sequence diversity of loop 7 might reflect the requirement of GH5\_2 cellulases to release the product under different conditions (e.g. under different temperatures and pH) [46]. The significantly increased activity of the mutant Y361W (Fig. 5b) demonstrated that loop 7 was an intriguing target to accelerate product release and thus improve the catalytic activity of GH5\_2 cellulases.

## Conclusion

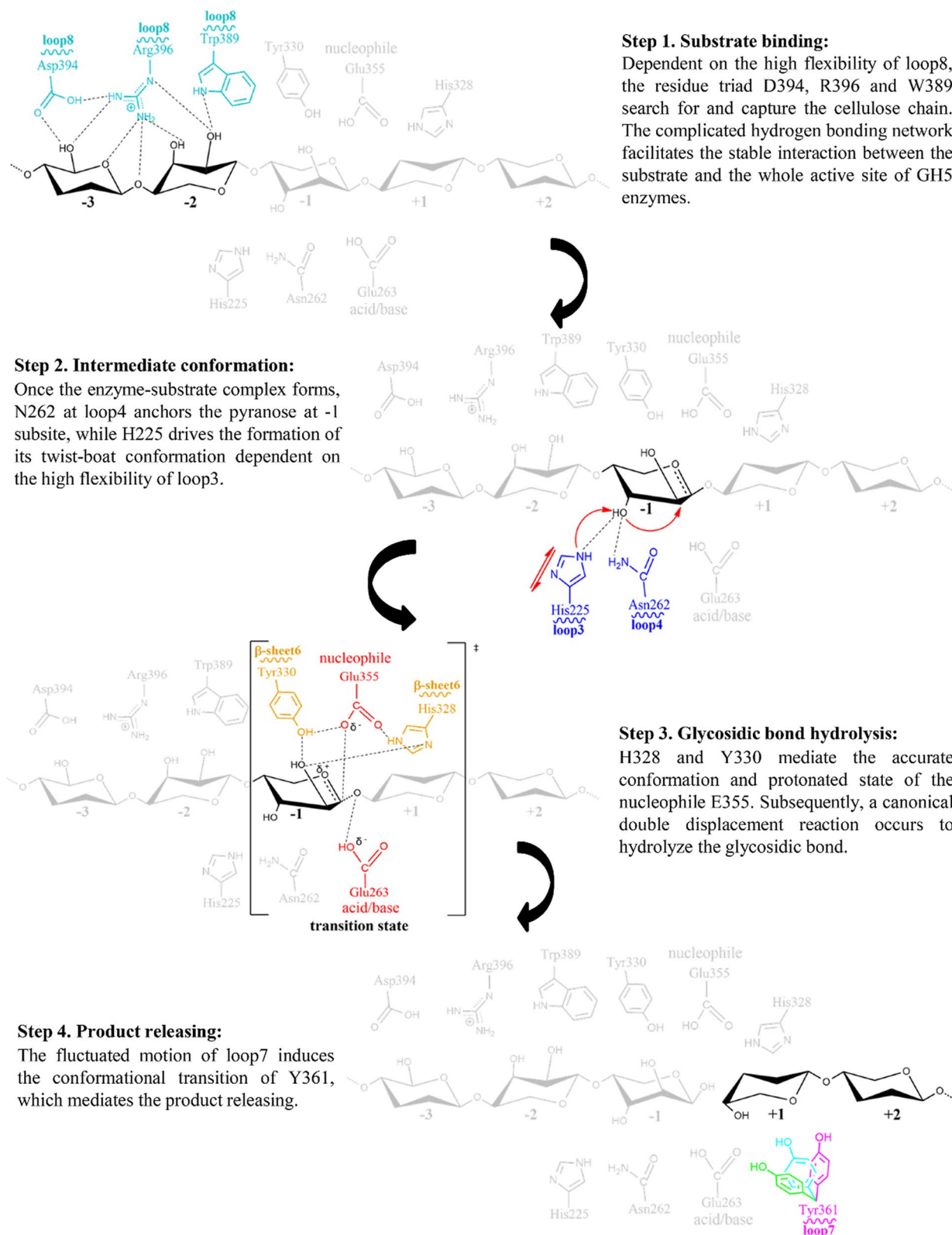
To the best of our knowledge, the present study is the first systematic structure-catalysis relationship study for GH5\_2 cellulases. By integrating experimental and computational approaches, we identified the key determinants of the catalytic activity of GH5\_2 cellulases. Specifically, loops 3, 7, and 8 at the different subsites of the active site showed unique structural dynamic features, which conferred distinct functions, including substrate binding, intermediate formation, stabilization of catalytic residues, and product release. Most functional residues identified in *TjCel5A* were highly conserved in the GH5\_2 family; moreover, our design of *TjCel5A* mutants suggested that the remote subsites provide intriguing chemical space for future engineering of GH5\_2 cellulases. Collectively, our results refreshed the understanding of the catalytic mechanism of GH5\_2 cellulases and guided the rational design of superior biocatalysts for biomass conversion.

## Material and methods

### Materials

*Escherichia coli* T1 (TransGen, Beijing, China) was used for gene cloning and sequencing. *E. coli* BL21 (DE3) (Invitrogen, Carlsbad, CA, USA) was used as a heterologous expression host. The *pEASY-blunt* E1 vector (TransGen)





**Fig. 6** A conceptual model to elucidate the catalysis of GH5\_2 cellulases

was used for constitutive expression in *E. coli*. The *tfcel5A* gene was chemically synthesized (Genewiz, Suzhou, China). The recombinant plasmid *pEASY-blunt E1/tfcel5A*, harboring the endoglucanase-encoding gene *tfcel5A* and a 6×His-tag at the C-terminal, was constructed to express the corresponding *TfCel5A* protein (GenBank accession no. Q01786.2). All chemicals were of analytical grade.

### Mutagenesis of *TfCel5A*

A Fast Mutagenesis System Kit purchased from TransGen was used to perform site-directed mutagenesis in accordance with the manufacturer's instructions. The recombinant plasmid *pEASY-blunt E1/tfcel5A* was used as the PCR template. Mutagenic primers were synthesized by Sangon (Shanghai, China), and their sequences are summarized in Additional file 1: Table S1.

### Enzyme expression and purification

*Escherichia coli* BL21 (DE3) cells were cultured in LB medium supplemented with 50 µg/mL kanamycin at 37 °C until the optical density of the medium reached 0.6–0.8 at 600 nm. Subsequently, for protein induction, isopropyl-β-D-thiogalactopyranoside (IPTG; Solarbio, Beijing, China) was added to the medium at the final concentration of 50 mM. The culture was incubated for 20 h in a shaker (20 °C, 200 rpm). After centrifugation, the precipitate was harvested and resuspended in a lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, pH 8.0). Ni<sup>2+</sup> affinity chromatography (HisTrap™ FF crude; GE Healthcare, Buckinghamshire, UK) was used for protein purification after ultrasonic fragmentation. The eluent was replaced with PC buffer (20 mM sodium phosphate, 10 mM citrate, pH 6.0) and subjected to ultrafiltration through a 3-kDa molecular cutoff membrane (Millipore, Billerica, MA, USA) at 4 °C. SDS-PAGE was performed on a 12% (w/v) polyacrylamide gel, and the protein bands were stained with Coomassie Brilliant Blue R-250 (Sigma–Aldrich, St. Louis, MO, USA). Protein concentration was determined using the classical Bradford method [47].

### Enzymatic property assay

CMC-Na (400–800 centipoises in water at room temperature) was purchased from Sigma–Aldrich as a substrate for enzymatic property assay. The reaction system comprised 150 µL of 1% (w/v) CMC-Na and 15 µg of the purified enzyme in a 300 µL reaction mixture. The reaction mixture was incubated at 60 °C and pH 5.0 for 30 min, and the reaction was terminated by the addition of 300 µL of 3,5-dinitrosalicylic acid (DNS) reagent in a boiling water bath for 10 min. After the sample was cooled-down to ambient temperature, the absorbance

was measured at 540 nm [48]. One unit of enzyme activity (U) was defined as the amount of enzyme (mg) that released 1 µmol of glucose per minute under the optimal assay conditions. The enzyme activities of WT and variants are expressed as mean ± SD (n=3). The relative parameters of WT were set to 100%. The bars indicate the standard errors.

Fluorescent spectrometry [49] was used to detect the substrate binding constant (K<sub>a</sub>). Briefly, 100 µL of the purified enzyme and 100 µL of acetate buffer (50 mM, pH 5.0) were added to a colorimeter tube. The excitation wavelength was 295 nm, and the emission wavelength ranged from 300 to 500 nm. Kinetic parameters were measured using an appropriate equivalent amount of the enzyme diluted with varying concentrations of CMC-Na (1–10 mg/mL). The reaction was performed in 50 mM acetate buffer (pH 5.0) at 60 °C for 30 min. Each experiment was performed in triplicate. The relative parameters of WT were set to 100%. The bars indicate the standard errors.

### Analysis of loops in the active site architecture of GH5\_2 subfamily

All protein crystal structures used in the present study were obtained from the Protein Data Bank (PDB) archive (<http://www.rcsb.org>). Homologous sequence alignments were performed using CLUSTAL [50]. Phylogenetic trees were constructed by the neighbor-joining method using MEGA7.0 [51] and optimized by iTOL online software [52]. Sequence profiling of the active site architecture of the GH5\_2 subfamily was performed by the online software WebLogo [53]. All structural diagrams were drawn using PyMOL software (<https://pymol.org/>). The b-factor of the enzyme molecule is recorded in the atomic coordinate file of the PDB structure and can be displayed by PyMOL. Temperature coloring is used to visualize the uncertainty of each atom.

### MD simulations

GROMACS 4.5 was used to perform all-atom simulations for *TfCel5A* and additional 21 GH5 cellulases. Two MD-simulation systems: substrate-free state and substrate-bound state were constructed to investigate the dynamic behavior of the enzyme *TfCel5A*. *TfCel5A* binds cellopentose in a substrate-bound system. The cellopentose was removed from the complex to keep the dissociated state in substrate-free system. The protein was solvated using SPC water models in a cubic box, and 0.1 M sodium chloride was used to neutralize the system. Van der Waals and electrostatic interaction energies between the enzyme and the substrate were calculated using the Amber force field equation [54]. The simulation

details were refined to match unbiased MD simulations [55, 56].

### Abbreviations

GH	Glycoside hydrolase
CAZy	Carbohydrate-active enZymes;
MD	Molecular dynamics
RMSF	Root mean square fluctuation
PDB	Protein Data Bank
IPTG	Isopropyl- $\beta$ -D-thiogalactopyranoside
DNS	3,5-dinitrosalicylic acid

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13068-023-02411-2>.

**Additional file 1: Table S1.** Primers of *TfCel5A* mutants. **Figure S1.** Phylogenetic tree of GH5 main subfamilies. **Figure S2.** Changes in the structural flexibility of GH5\_2 enzymes before and after substrate binding. Root Mean Square Fluctuation (RMSF) values were calculated by molecular dynamics simulations. PDB codes of the used enzymes are labeled. **Figure S3.** Structural flexibility of enzymes from major GH5 subfamilies. The PDB codes of the enzymes are labeled. Visualizing the root mean square fluctuation changes in atomic positions is done by coloring by RMSF value. Atoms with low RMSF values are colored blue, while atoms with high RMSF values are colored red. **Figure S4.** The hydrogen bonding network constructed by the triplet residues at – 3 and – 2 subsites. **a** The wild-type *TfCel5A* (PDB: 2CKR). **b** *Bacillus subtilis* endo-1,4-beta-glucanase (PDB: 3PZT). **c** *Salipaludibacillus agaradhaerens* endoglucanase Cel5A (PDB: 1H11). **d** *Dickeya dadantii* cellulase Cel5 (PDB: 1EGZ). **Figure S5.** Measurements of catalytic activity and substrate binding affinity of the loop 1, 2, and 8 mutants. **Figure S6.** The interaction energy of WT, Y361A and Y361W mutants with ligand, respectively.

### Acknowledgements

We thank the HPC Cloud Platform of Shandong University.

### Author contributions

XW participated in data analysis and drafted the manuscript. SZ and ZT performed the experiments. CH and XJ revised the manuscript. LW conceived and designed the experiments and revised the manuscript. All authors read and approved the final manuscript.

### Funding

This work was supported by National Natural Science Foundation of China (32100022, 31770054), Key Research and Development Project of Shandong Province (2020CXGC010601), Major Project of Natural Science Foundation Shandong Province (ZR2018ZB0420).

### Availability of data and materials

The data generated or analyzed during this study are included in this published article and its additional files. Further datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 25 July 2023 Accepted: 12 October 2023

Published online: 18 October 2023

### References

- Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*. 2007;315(5813):804–7.
- Zhang KD, Li W, Wang YF, Zheng YL, Tan FC, Ma XQ, Yao LS, Bayer EA, Wang LS, Li FL. Processive degradation of crystalline cellulose by a multimodular endoglucanase via a wirewalking mode. *Biomacromol*. 2018;19(5):1686–96.
- Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Res*. 2014;42(Database):D490–495.
- Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VG. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science*. 2010;330(6001):219–22.
- Zheng F, Tu T, Wang X, Wang Y, Ma R, Su X, Xie X, Yao B, Luo H. Enhancing the catalytic activity of a novel GH5 cellulase GtCel5 from *Gloeophyllum trabeum* CBS 900.73 by site-directed mutagenesis on loop 6. *Biotechnol Biofuels*. 2018;11:76.
- Yennamalli RM, Rader AJ, Kenny AJ, Wolt JD, Sen TZ. Endoglucanases: insights into thermostability for biofuel applications. *Biotechnol Biofuels*. 2013. <https://doi.org/10.1186/1754-6834-6-136>.
- Liu G, Qu Y. Integrated engineering of enzymes and microorganisms for improving the efficiency of industrial lignocellulose deconstruction. *Eng Microbiol*. 2021;1:100005.
- Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM. Thermophilic lignocellulose deconstruction. *FEMS Microbiol Rev*. 2014;38(3):393–448.
- Liu G, Qu Y. Engineering of filamentous fungi for efficient conversion of lignocellulose: tools, recent advances and prospects. *Biotechnol Adv*. 2019;37(4):519–29.
- Henrissat B, Claeysens M, Tomme P, Lemesle L, Mornon JP. Cellulase families revealed by hydrophobic cluster analysis. *Gene*. 1989;81(1):83–95.
- Aspeborg H, Coutinho PM, Wang Y, Brumer H 3rd, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol*. 2012;12:186.
- Chen Z, Friedland GD, Pereira JH, Revoco SA, Chan R, Park JI, Thelen MP, Adams PD, Arkin AP, Keasling JD, et al. Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J Biol Chem*. 2012;287(30):25335–43.
- Liu S, Shao S, Li L, Cheng Z, Tian L, Gao P, Wang L. Substrate-binding specificity of chitinase and chitosanase as revealed by active-site architecture analysis. *Carbohydr Res*. 2015;418:50–6.
- Tian L, Liu S, Wang S, Wang L. Ligand-binding specificity and promiscuity of the main lignocellulolytic enzyme families as revealed by active-site architecture analysis. *Sci Rep*. 2016;6:23605.
- Payne CM, Knott BC, Mayes HB, Hansson H, Himmel ME, Sandgren M, Stahlberg J, Beckham GT. Fungal cellulases. *Chem Rev*. 2015;115(3):1308–448.
- Huang PS, Feldmeier K, Parmeggiani F, Velasco DAF, Hocker B, Baker D. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol*. 2016;12(1):29–34.
- Glasgow EM, Kemna EI, Bingman CA, Ing N, Deng K, Bianchetti CM, Takasuka TE, Northen TR, Fox BG. A structural and kinetic survey of GH5\_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis. *J Biol Chem*. 2020;295(51):17752–69.
- Yang H, Shi P, Liu Y, Xia W, Wang X, Cao H, Ma R, Luo H, Bai Y, Yao B. Loop 3 of fungal endoglucanases of glycoside hydrolase family 12 modulates catalytic efficiency. *Appl Environ Microbiol*. 2017. <https://doi.org/10.1128/AEM.03123-16>.
- Luo J, Song C, Cui W, Han L, Zhou Z. Counteraction of stability-activity trade-off of Nattokinase through flexible region shifting. *Food Chem*. 2023;423:136241.
- Yang P, Wang X, Ye J, Rao S, Zhou J, Du G, Liu S. Enhanced thermostability and catalytic activity of *Streptomyces mobaraensis* transglutaminase by rationally engineering its flexible regions. *J Agric Food Chem*. 2023;71(16):6366–75.

21. Kamerlin SC, Rucker R, Boresch S. A molecular dynamics study of WPD-loop flexibility in PTP1B. *Biochem Biophys Res Commun.* 2007;356(4):1011–6.
22. Li Z, Meng S, Nie K, Schwaneberg U, Davari MD, Xu H, Ji Y, Liu L. Flexibility regulation of loops surrounding the tunnel entrance in cytochrome P450 enhanced substrate access substantially. *ACS Catal.* 2022;12:12800–8.
23. Wu M, Bu L, Vuong TV, Wilson DB, Crowley MF, Sandgren M, Ståhlberg J, Beckham GT, Hansson H. Loop motions important to product expulsion in the *Thermobifida fusca* glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. *J Biol Chem.* 2013;288(46):33107–17.
24. Jiang X, Wang Y, Xu L, Chen G, Wang L. Substrate binding interferes with active site conformational dynamics in endoglucanase Cel5A from *Thermobifida fusca*. *Biochem Biophys Res Commun.* 2017;491(1):236–40.
25. Song Y, Wu X, Zhao Y, Jiang X, Wang L. Comparative molecular dynamics simulations identify a salt-sensitive loop responsible for the halotolerant activity of GH5 cellulases. *J Biomol Struct Dyn.* 2021. <https://doi.org/10.1080/07391102.2021.1930167>.
26. Bu L, Crowley MF, Himmel ME, Beckham GT. Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. *J Biol Chem.* 2013;288(17):12175–86.
27. Bartoli-German I, Haiech J, Chippaux M, Barras F. Informational suppression to investigate structural functional and evolutionary aspects of the cellulase EGZ. *J Mol Biol.* 1995;246(1):82–94.
28. Varrot A, Davies GJ. Direct experimental observation of the hydrogen-bonding network of a glycosidase along its reaction coordinate revealed by atomic resolution analyses of endoglucanase Cel5A. *Acta Crystallogr D Biol Crystallogr.* 2003;59(3):447–52.
29. Chapon V, Czjzek M, El Hassouni M, Py B, Juy M, Barras F. Type II protein secretion in gram-negative pathogenic bacteria: the study of the structure/secretion relationships of the cellulase cel5 (formerly EGZ) from *Erwinia chrysanthemi*. *J Mol Biol.* 2001;310(5):1055–66.
30. Wu X, Shi Z, Tian W, Liu M, Huang S, Liu X, Yin H, Wang L. A thermostable and CBM2-linked GH10 xylanase from *Thermobifida fusca* for paper bleaching. *Front Bioeng Biotechnol.* 2022;10:939550.
31. Wu X, Zhang S, Zhang Q, Zhao Y, Chen G, Guo W, Wang L. The contribution of specific subsites to catalytic activities in active site architecture of a GH11 xylanase. *Appl Microbiol Biotechnol.* 2020;104(20):8735–45.
32. Sun Z, Liu Q, Qu G, Feng Y, Retz MT. Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. *Chem Rev.* 2019;119(3):1626–65.
33. Kumagai Y, Yamashita K, Tagami T, Uraji M, Wan K, Okuyama M, Yao M, Kimura A, Hatanaka T. The loop structure of *Actinomyces* glycoside hydrolase family 5 mannanases governs substrate recognition. *FEBS J.* 2015;282(20):4001–14.
34. Liang PH, Lin WL, Hsieh HY, Lin TY, Chen CH, Tewary SK, Lee HL, Yuan SF, Yang B, Yao JY, et al. A flexible loop for mannan recognition and activity enhancement in a bifunctional glycoside hydrolase family 5. *Biochim Biophys Acta Gen Subj.* 2018;1862(3):513–21.
35. Davies GJ, Mackenzie L, Varrot A, Dauter M, Brzozowski AM, Schulein M, Withers SG. Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry.* 1998;37(34):11707–13.
36. Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich JP, Davies GJ, Haser R. Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure.* 1995;3(9):939–49.
37. Wu X, Tian Z, Jiang X, Zhang Q, Wang L. Enhancement in catalytic activity of *Aspergillus niger* XynB by selective site-directed mutagenesis of active site amino acids. *Appl Microbiol Biotechnol.* 2018;102(1):249–60.
38. Vocado DJ, Davies GJ. Mechanistic insights into glycosidase chemistry. *Curr Opin Chem Biol.* 2008;12(5):539–55.
39. Zheng B, Yang W, Zhao X, Wang Y, Lou Z, Rao Z, Feng Y. Crystal structure of hyperthermophilic endo- $\beta$ -1,4-glucanase: implications for catalytic mechanism and thermostability. *J Biol Chem.* 2012;287(11):8336–46.
40. Ardevol A, Rovira C. Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases: insights from ab Initio quantum mechanics/molecular mechanics dynamic simulations. *J Am Chem Soc.* 2015;137(24):7528–47.
41. Fenwick MK, Mehta AP, Zhang Y, Abdelwahed SH, Begley TP, Ealick SE. Non-canonical active site architecture of the radical SAM thiamin pyrimidine synthase. *Nat Commun.* 2015;6:6480.
42. Keen BA, Jozwiakowski SK, Bailey LJ, Bianchi J, Doherty AJ. Molecular dissection of the domain architecture and catalytic activities of human PrimPol. *Nucleic Acids Res.* 2014;42(9):5830–45.
43. Rutten L, Mannie JP, Stead CM, Raetz CR, Reynolds CM, Bonvin AM, Tommassen JP, Egmond MR, Trent MS, Gros P. Active-site architecture and catalytic mechanism of the lipid A deacylase LpxR of *Salmonella typhimurium*. *Proc Natl Acad Sci USA.* 2009;106(6):1960–4.
44. Chakladar S, Wang Y, Clark T, Cheng L, Ko S, Vocado DJ, Bennet AJ. A mechanism-based inactivator of glycoside hydrolases involving formation of a transient non-classical carbocation. *Nat Commun.* 2014;5:5590.
45. Zhang X, Wang S, Wu X, Liu S, Li D, Xu H, Gao P, Chen G, Wang L. Subsite-specific contributions of different aromatic residues in the active site architecture of glycoside hydrolase family 12. *Sci Rep.* 2015. <https://doi.org/10.1038/srep18357>.
46. Liu MQ, Li JY, Rehman AU, Xu X, Gu ZJ, Wu RC. Laboratory evolution of GH11 endoxylanase through DNA shuffling: effects of distal residue substitution on catalytic activity and active site architecture. *Front Bioeng Biotechnol.* 2019;7:350.
47. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 1976;72:248–54.
48. Miller GL. Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem.* 1959;31(3):426–8.
49. Bacia K, Hausteiner E, Schwille P. Fluorescence correlation spectroscopy: principles and applications. *Cold Spring Harb Protoc.* 2014;2014(7):709–25.
50. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. Clustal W and clustal X version 2.0. *Bioinformatics.* 2007;23(21):2947–8.
51. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870–4.
52. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44(Web Server):W242–5.
53. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* 2004;14(6):1188–90.
54. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc.* 1996;118(9):2309–2309.
55. Jiang XK, Chen GJ, Wang LS. Structural and dynamic evolution of the amphipathic N-terminus diversifies enzyme thermostability in the glycoside hydrolase family 12. *Phys Chem Chem Phys.* 2016;18(31):21340–50.
56. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts MR, Smith JC, Kasson PM, van der Spoel D, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics.* 2013;29(7):845–54.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

