

Catalytic mechanism and origin of high activity of cellulase *TmCel12A* at high temperature: a quantum mechanical/molecular mechanical study

Peng Lian · Hao-Bo Guo · Jeremy C. Smith ·
Dong-Qing Wei · Hong Guo

Received: 5 March 2013 / Accepted: 25 July 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Understanding the factors that determine the catalytic efficiency of cellulases is of considerable importance in cellulosic ethanol production, especially at high temperature. The cellulase 12A from the hyperthermophile *Thermotoga maritima* (*TmCel12A*) is a possible candidate for accelerating the rate of hydrolysis via temperature elevation up to as high as 95 °C. However, the details of the catalytic mechanism and origin of the activity of *TmCel12A* at high temperature have not been well studied. Here, the enzyme-catalyzed reaction is explored using free energy simulations (potential of mean force) with umbrella sampling and quantum mechanical/molecular mechanical (SCC-DFTB/MM) potential at both relatively low (37 °C) and high (85 °C) temperatures. The free energy barriers for glycosylation and deglycosylation are calculated to be

22.5 ± 0.4 and 24.5 ± 0.7 kcal · mol⁻¹ at 85 °C, respectively. The barrier for deglycosylation is found to decrease with increasing temperature or as a result of the Y61 → G mutation, consistent with experimental observations. The transition state for glycosylation and deglycosylation obtained from the simulations is in an oxocarbenium state with the -1 glucose ring having an E₃ envelop (or ⁴H₃ half-chair) conformation. A unique characteristic of the *TmCel12A* structure seems to be the existence of a stable moiety that may play a role in “holding” cellulose at the binding site with the correct orientation for the reaction even at 85 °C. This stable moiety (comprising hydrogen-bonded E116, E134, E227 and an active-site water molecule) may be one of the important factors for the relatively high activity of *TmCel12A* at high temperature.

P. Lian · D.-Q. Wei (✉)
The State Key Laboratory of Microbial Metabolism,
College of Life Sciences and Biotechnology,
Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: dqwei@sjtu.edu.cn

H.-B. Guo · J. C. Smith · H. Guo
Department of Biochemistry and Cellular and Molecular
Biology, University of Tennessee,
Knoxville, TN 37996, USA

H.-B. Guo · J. C. Smith · H. Guo (✉)
Oak Ridge National Laboratory, UT/ORNL Center
for Molecular Biophysics, 1 Bethel Valley Road,
Oak Ridge, TN 37831-6309, USA
e-mail: hguo1@utk.edu

Keywords Cellulase · Catalytic mechanism ·
Quantum mechanical/molecular mechanical
(QM/MM) · Molecular dynamics (MD)

Introduction

Cellulose, a straight chain of homogenous polymer of glucose units linked together by β-1,4-glycosidic bond, is one of the most abundant carbohydrates on earth and is a potentially important source of renewable bioproducts and biofuels such as ethanol

(Himmel et al. 2007; Demain et al. 2005; Himmel et al. 1999). Cellulases (Zhang et al. 2006) from several microbial species are able to hydrolyze cellulose chains into fragments, and even monomeric glucoses, which can serve as precursors for producing renewable biofuel energy. Developing more efficient cellulases is of considerable interest for industrial biofuel production from cellulose. However, most, if not all, cellulases currently known lack the efficiency due to thermal deactivation, end product inhibition, nonspecific binding to lignin, and some other factors (Zhang et al. 2006; Yang and Wyman 2004).

One possible way to speed up a reaction is to elevate the reaction temperature. However, most enzymes in living organisms are mesophilic or psychrophilic, with optimal functional temperatures around 25–50 and 5–25 °C, respectively (Vieille and Zeikus 2001). Various approaches have been adopted to improve the thermostability of cellulases (Xue et al. 2012; Javed et al. 2011; Viikari et al. 2007; Kang et al. 2007; Haki and Rakshit 2003; Volfova et al. 1985; Vieille and Zeikus 2001), including rational design (Srikrishnan et al. 2012; Yi et al. 2011), directed evolution (Liang et al. 2011a; Nakazawa et al. 2009), genetic engineering (Argyros et al. 2011; Shaw et al. 2008), in vitro recombination (Lee et al. 2010; Murashima et al. 2002) and theoretical techniques (Anbar et al. 2012; Liang et al. 2011b). In order to apply the different approaches, understanding the activities of the enzymes at high temperature is of importance. Interestingly, a small group of enzymes from thermophilic microorganisms is able to withstand temperatures as high as 95 °C (Bronnenmeier et al. 1995; Liebl et al. 1996). Cellulase 12A from the bacterial thermophile *Thermotoga maritima* (*TmCel12A*) is one member of this group and belongs to the family 12 of glycoside hydrolases (GH12)(Bronnenmeier et al. 1995; Liebl et al. 1996; Nelson et al. 1999). The X-ray crystallographic structure of over-expressed *TmCel12A* from the gene MSB8 (ATCC 43589) has been recently solved by Cheng et al. (Liu et al. 2011), and the structural and biochemical studies have provided valuable insight into the function of the enzyme. For instance, it has been demonstrated that the activity of *TmCel12A* can be further enhanced by certain mutation(s). Indeed, the k_{cat} value was found to increase from 791 to 1,155 S⁻¹ when Y61 of *TmCel12A* was mutated to glycine (Cheng et al. 2012). Nevertheless, to the best of our knowledge the catalytic mechanism and activities of

TmCel12A and its mutant at high temperature have not been studied in detail.

In the present study, quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) and free energy simulations were performed on *TmCel12A* at both relatively low (37 °C) and high (85 °C) temperatures to investigate the catalytic mechanism and to understand the activity of the enzyme at high temperature. The relatively high activities at the high temperature and as a result of the Y61G mutation were reproduced computationally. The simulations indicate that a structural moiety comprising E116, E134, E227 and an active-site water molecule (stabilized by hydrogen-bond networks) appears to play a role in maintaining the reactive configuration for cellulose at the high temperature (85 °C). A bioinformatics assessment was further performed to support this hypothesis by exploring the sequence conservation of both hyperthermophilic and non-thermophilic cellulases.

Methods

The CHARMM program (Brooks et al. 1983, 2009) was used for this study. A semi-empirical density-functional approach, self-consistent charge density-functional tight-binding method (SCC-DFTB) (Elstner et al. 1998) implemented in the CHARMM program (Brooks et al. 1983, 2009), was used for the QM atoms in the QM/MM MD and free energy (potential of mean force, PMF) simulations. SCC-DFTB has been widely tested and applied to both organic molecules (Elstner et al. 2003; Sattelmeyer et al. 2006) and biological molecules including enzymes (Banerjee et al. 2005; Guo et al. 2001; Riccardi et al. 2008; Li and Cui 2003; Bondar et al. 2004; Guo et al. 2005a, b; Xu et al. 2006). Although downfalls were found from different versions of DFTB in accurately simulating the structure of water cluster (Maupin et al. 2010; Choi et al. 2013) and carbohydrates (Islam and Roy 2012), the parameters in mio-1-1 set were reported to reproduce the free energy surface obtained with *ab initio* method for the furanose and pyranose ring systems (Barnett and Naidoo 2010; Barnett et al. 2010; Sattelle and Almond 2010). Moreover, a comparison of SCC-DFTB calculation with B3LYP/6-31G* on a similar system has been made previously by Liu et al. (Liu et al. 2010). Only

small differences in bond lengths (less than 0.1 Å) and angles (less than 5°) were observed. Therefore, it is plausible to use SCC-DFTB in the QM/MM simulations of the cellulase *TmCel12A*. However, it should be pointed out that care must be exercised when the SCC-DFTB method is used for other systems.

The CHARMM32 (Mackerell et al. 2004) all-atom force field was employed for the MM atoms. A modified TIP3P (Jorgensen 1981) water model was employed for the solvent. The initial coordinates were obtained from the crystal structure of cellulase 12A from *T. maritima* (*TmCel12A* (Liu et al. 2011), PDB ID: 3AMM, resolution: 1.98 Å). The substrate (cello-tetraose), residues E231 (the general acid/base), E134 and E116 were treated by QM, and the rest of the system by MM. The mio-1-1 (Elstner et al. 1998; Niehaus et al. 2001; Yang et al. 2008; Gaus et al. 2011) set of parameters was chosen for the SCC-DFTB method.

Stochastic boundary conditions (Brooks et al. 1985) were adopted in this study. The system was separated into a reaction zone and a reservoir region and the reaction zone was further divided into a reaction region and a buffer region. The oxygen atom of the glycosidic linkage between the -1 and +1 glucose (Fig. 1) was chosen as the reference center for partitioning the system. The reaction region was a sphere with radius $R = 20$ Å, and the buffer region with $20 \leq R \leq 22$ Å. Atoms in the reaction region were simulated using Newtonian molecular dynamics, whereas atoms in the buffer region were described by Langevin dynamics. The friction constants of the Langevin dynamics for the protein and water atoms were 250 and 62 ps⁻¹, respectively. An integration step of 1 fs was used in all simulations. A minimization using steepest descent method for the MM atoms, followed by the adapted basis Newton–Raphson (ABNR) method for the whole system was performed until the gradient of the system converged to less than 10⁻⁶ kcal · mol⁻¹. The reaction coordinate driving method was employed to find the minimum energy path (MEP).

The catalysis involves both a nucleophilic attack (NA) and a proton transfer (PT) in glycosylation and deglycosylation (Fig. 1). Therefore, a set of proper reaction coordinates (RC) is required to represent both events during the two steps of the catalysis. A rational choice is to use two separate RCs, with RC1 for the PT

and RC2 for the NA, respectively. This choice has been applied in other enzyme systems (Guo et al. 2006). However, a relatively expensive two-dimensional (2D) simulation is required for this approach. An alternative, one-dimensional (1D) approach is to use the linear combination of both RCs, if the NA and PT processes are correlated with each other such that to represent the 2D results (Saharay et al. 2010). In order to test whether the 1D approach is able to reproduce the 2D results, both approaches were applied on the *TmCel12A* system at its optimal temperature (85 °C) (Bronnenmeier et al. 1995; Liebl et al. 1996).

Starting with the structures collected from the MEP calculations, umbrella sampling (Torrie and Valleau 1974) together with the weighted histogram analysis method (WHAM) (Kumar et al. 1992) were applied to determine the free energy (potential of mean force, PMF) along the reaction path. The free energy profiles were calculated at both 37 and 85 °C. In the 1D PMFs at both 37 and 85 °C, 64 windows with a width of 0.1 Å were employed for glycosylation (60 windows for deglycosylation). For the low temperature (37 °C) simulations, the system at each window was heated to 37 °C in 50 ps and equilibrated for 500 ps until the RMSD value of the backbone atoms of enzyme reached a plateau, and the data were collected from a 200 ps production run after the equilibration. In the high temperature (85 °C) simulations, 50 ps heating, 200 ps equilibration and 100 ps production runs were performed starting from the equilibrated conformations obtained at 37 °C. In the umbrella sampling, harmonic potentials with a force constant of 80 kcal · mol⁻¹ · Å⁻² were used. In each case, five independent runs were performed using different initial randomized Gaussian distribution of velocities. The standard deviations of the free energies obtained from the five simulations were subsequently presented as the statistical errors. For the 2D simulations at 85 °C of the wild type (WT) enzyme, two harmonic restraints on the RCs for both NA and PT were applied for a total of 192 windows for glycosylation, and 180 windows for deglycosylation, respectively.

In order to further determine the factors affecting the activity of *TmCel12A*, an additional 500 ps QM/MM MD simulations with all restraints relaxed were performed for the reactant state of the WT, E116G and E116D mutants at both the low (37 °C) and high (85 °C) temperatures, respectively.

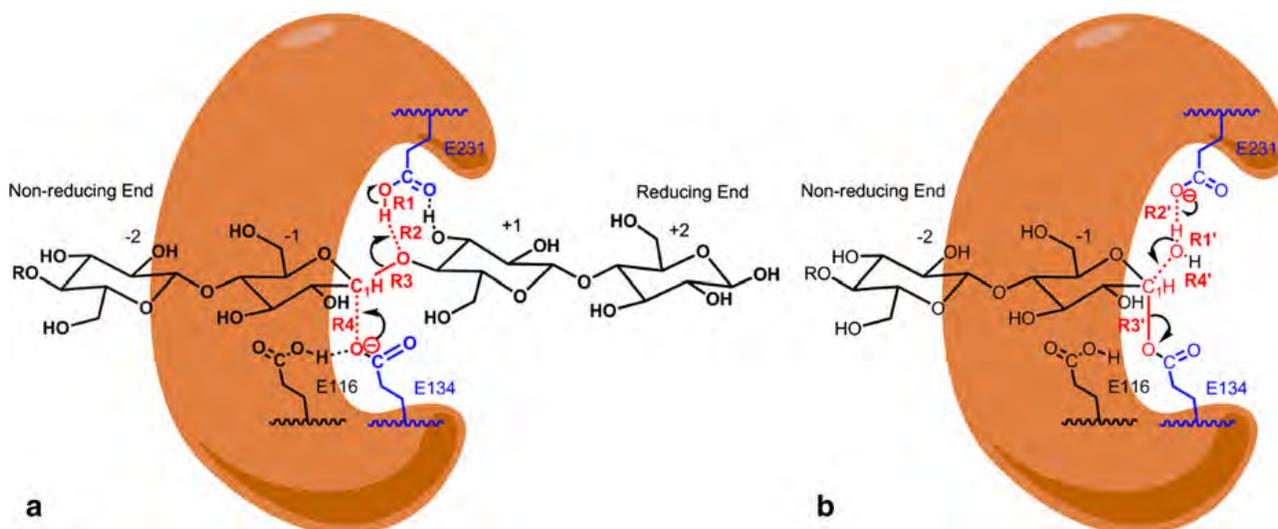


Fig. 1 Reaction coordinates (RC) of **a** glycosylation, the 2D (RC1/RC2) and 1D (RC_{glycol}) RCs are $RC1 = R1 - R2$, $RC2 = R3 - R4$ and $RC_{\text{glycol}} = RC1 + RC2$; and **b** deglycosylation. The 2D ($RC1'/RC2'$) and 1D (RC_{deglycol}) RCs are $RC1' = R1' - R2'$, $RC2' = R3' - R4'$ and $RC_{\text{deglycol}} =$

$RC1' + RC2'$. Atoms used in the reaction coordinate driving are shown in *red* and the catalytic residues are shown in *blue*. Here $R1 = r(O_{E231}-H_{E231})$, $R2 = r(H_{E231}-O_{\text{linkage}})$, $R3 = r(O_{\text{linkage}}-C1)$, $R4 = r(C1-O_{E134})$, $R1' = r(H_{WT}-O_{WT})$, $R2' = r(O_{231}-H_{WT})$, $R3' = r(C1-O_{E134})$ and $R4' = r(O_{WT}-C1)$. (Color figure online)

Results and discussion

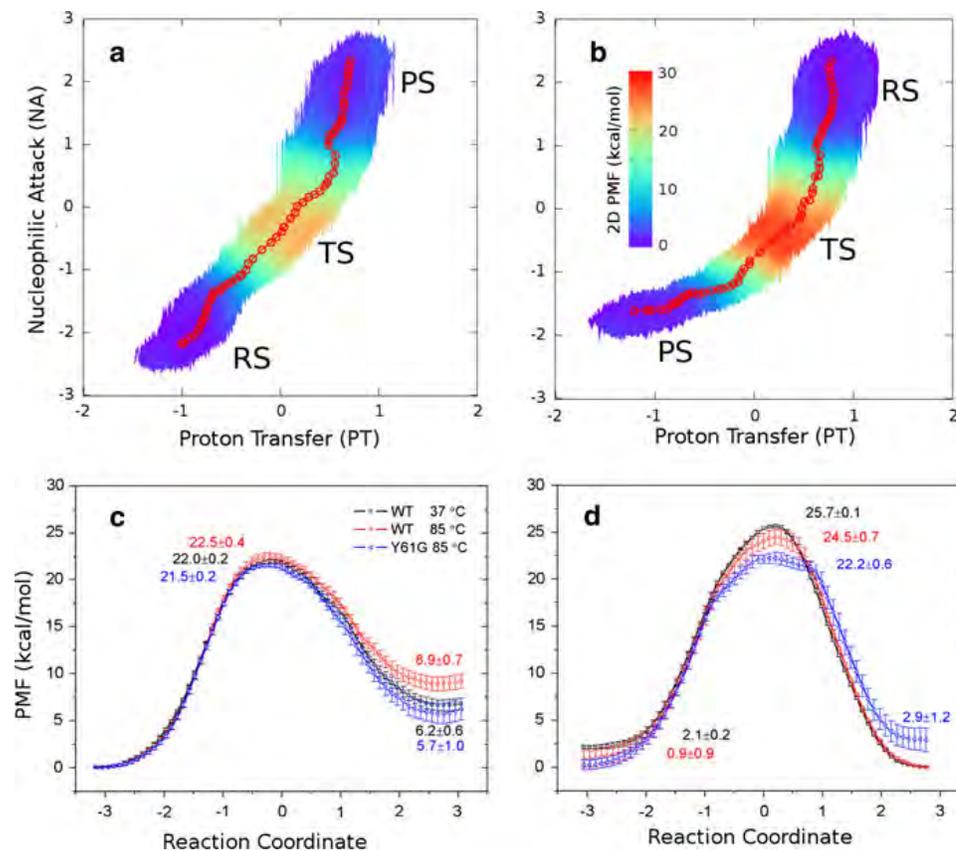
Similar to other members of the family, *TmCel12A* is a retaining glycosidase, i.e., the stereo configuration of the anomeric center is retained after catalysis (Schou et al. 1993). The enzyme hydrolyzes cellulose in two steps: glycosylation and deglycosylation. The catalytic reaction in glycosylation/deglycosylation involves a nucleophilic attack (NA) accompanied by a proton transfer (PT), which can be described by two independent reaction coordinates ($RC1/RC1'$ and $RC2/RC2'$), respectively (Fig. 1). During the glycosylation, the carboxylate of E134 attacks the anomeric carbon atom at the -1 position of the substrate (NA, $RC2$), and at the same time the protonated E231 side chain functions as a general acid to protonate the linkage oxygen atom (PT, $RC1$). This step yields a non-reducing end at the $+1$ position of the cellobiose and a covalent complex between E134 and -1 unit of the substrate. Moreover, the deprotonated E231 could serve as a general base facilitating the deglycosylation step. In deglycosylation, the NA involves the attacking of a water molecule to the anomeric carbon at the -1 position ($RC2'$), which is assisted by the PT from the water molecule to E231 ($RC1'$). This step leads to the cleavage of the C–O bond between cellobiose and E134 and the formation of the C–O bond between the

cellobiose and the hydroxyl group of the water molecule. As a consequence, E231 returns to the protonated state. After deglycosylation the enzyme returns to its initial state such that the catalytic cycle of degradation is complete (Fig. 1).

Potential of mean force profiles of glycosylation and deglycosylation

Two-dimensional (2D) PMF surfaces for both glycosylation and deglycosylation at 85 °C [the optimal growth temperature of *TmCel12A* (Bronnenmeier et al. 1995; Liebl et al. 1996)] were performed using two separate reaction coordinates ($RC2/RC2'$ and $RC1/RC1'$, Fig. 1) to describe the nucleophilic attack (NA) and the proton transfer (PT), respectively. The reaction surfaces from the 2D PMFs are plotted in Fig. 2a, b for glycosylation and deglycosylation, respectively. Concerted NA and PT were observed in glycosylation, i.e., the PT appeared to be synchronous with the NA. However, in deglycosylation (from RS to PS), the NA occurred ahead of the PT, i.e., the PT started mainly after the TS. The free energy barriers obtained with the 2D PMF are 21.4 kcal · mol⁻¹ for glycosylation and 27.1 kcal · mol⁻¹ for deglycosylation.

Fig. 2 2D potential of mean force (PMF) maps and 1D PMF profiles for glycosylation (a, c) and deglycosylation (b, d). 2D PMF (a, b) were obtained at 85 °C for WT, whereas 1D PMF (c, d) were obtained at 37 °C for WT (black squares) and at 85 °C for WT (red circles) or 85 °C for the Y61 → G mutant (blue triangles). The reaction paths (red circle) projected from 1D PMF simulations of WT at 85 °C are shown on the 2D PMF maps. Averaged free energy values together with the standard deviations from five independent simulations for 1D PMF were shown in c and d. (Color figure online)



For comparison, the 1D PMF simulations using a linear combination of the two RCs (Fig. 1) were also carried out for both glycosylation and deglycosylation of WT at 85 °C. It was found that for both steps the reaction paths (Fig. 2c, d) obtained from the five separate 1D PMF simulations agree well with the corresponding result from 2D calculations (Fig. 2a, b) with the free energy barriers deviating by only ~ 1 – 2 kcal \cdot mol $^{-1}$. These observations indicate that the 1D PMFs should be able to address the essential characteristics of the 2D PMFs for the reactions studied here. The transition state (TS) structures of the glycosylation and deglycosylation steps of WT at 85 °C further support the consistency between the 1D and 2D approaches (see below and Fig. 3). Consequently, in the subsequent calculations only the 1D PMF simulations for WT and the mutants at both low (37 °C) and high (85 °C) temperatures were performed to save the computer time.

The PMF profiles for glycosylation exhibit only one obvious peak, with a free-energy barrier of 22.0 ± 0.2 and 22.5 ± 0.4 kcal \cdot mol $^{-1}$ at 37 and 85 °C,

respectively (Fig. 2c). The deglycosylation reaction is believed to be the rate-limiting step for some cellulases utilizing the same two-step mechanism (Liu et al. 2010), and this could be the case for *TmCel12A* as well as indicated the relatively higher barriers for deglycosylation from the simulations (Fig. 2d). Figure 2 shows that at 85 °C the calculated free-energy barrier for deglycosylation is 24.5 ± 0.7 kcal \cdot mol $^{-1}$. At 37 °C, the free-energy barrier of deglycosylation was calculated to be 25.7 ± 0.1 kcal \cdot mol $^{-1}$. The elevation of temperature appears to have a relatively larger effect on the free-energy barrier of deglycosylation than on glycosylation. The relatively lower free-energy barrier at the higher temperature for deglycosylation is consistent with a recent experimental observation (Barnett et al. 2011) that the activity of the enzyme increases with increasing temperature (up to ~ 95 °C). The calculated difference in the free energy barriers of the WT enzyme is still too small to make any conclusion and limited by the accuracy of the computational approaches (also considering the statistical errors). However, it is of interest to note that the

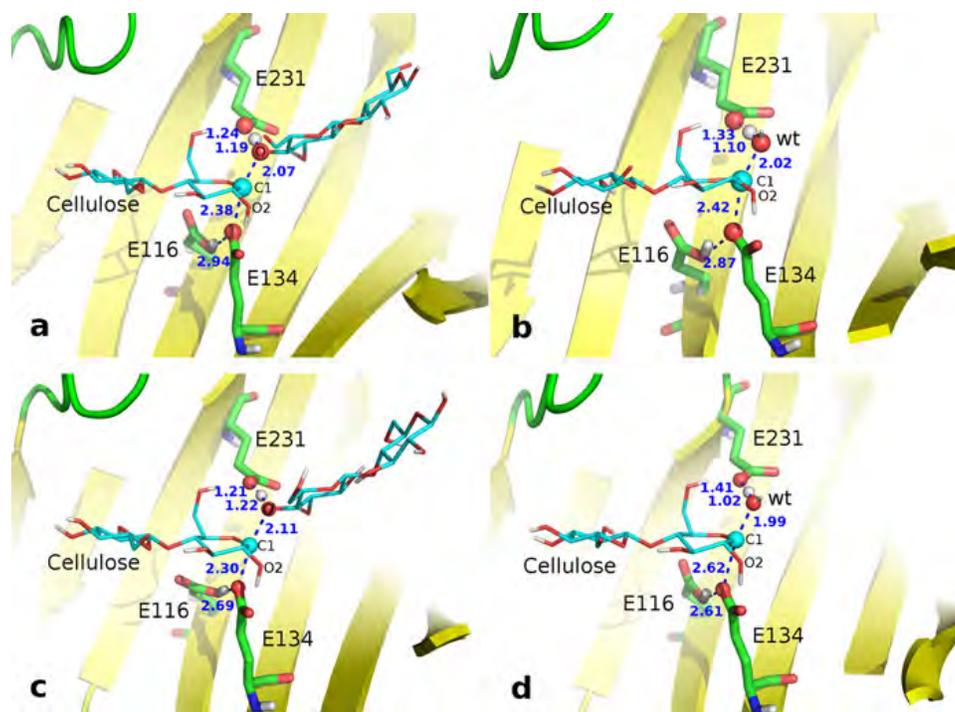


Fig. 3 The transition state structures for glycosylation (**a**, **c**) and deglycosylation (**b**, **d**) obtained from the 2D (**a**, **b**) and 1D (**c**, **d**) free energy simulations. The substrate is shown in thin sticks (cyan) and other active site atoms in sticks (green). The

atoms associated with the reaction coordinates are shown in scaled Van der Waals spheres, and the related bond lengths are labeled. (Color figure online)

calculated free-energy barriers of $22.5 \pm 0.4 \text{ kcal} \cdot \text{mol}^{-1}$ for glycosylation and $24.5 \pm 0.7 \text{ kcal} \cdot \text{mol}^{-1}$ for deglycosylation at the enzyme's optimal temperature ($85 \text{ }^\circ\text{C}$) are reasonably consistent with the experimental measurement of $22.7 \text{ kcal} \cdot \text{mol}^{-1}$ for the degradation of carboxymethylcellulose by *TmCel12A* at $95 \text{ }^\circ\text{C}$ (Bronnenmeier et al. 1995). It should be pointed out, however, that the relative free energy barriers, as opposed to the absolute barriers, are expected to be more reliable and less sensitive to the choice of the QM method and experimental structures due in part to the cancellation of the errors.

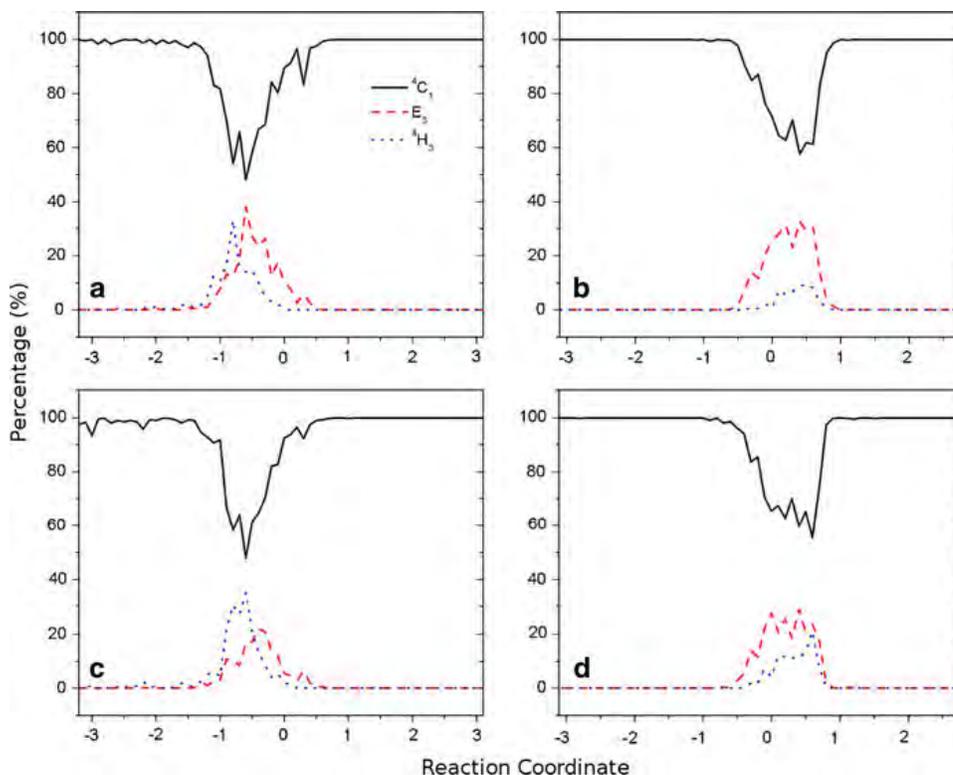
One interesting experimental observation for *TmCel12A* is that the $\text{Y61} \rightarrow \text{G}$ mutation can enhance the activity of the enzyme (Cheng et al. 2012), with k_{cat} increasing from 791 to $1,155 \text{ S}^{-1}$ when Y61 is changed to G. Consistent with this experimental observation, Fig. 2 shows that the free energy barrier for deglycosylation is decreased by $\sim 2 \text{ kcal} \cdot \text{mol}^{-1}$ as a result of the $\text{Y61} \rightarrow \text{G}$ mutation. The statistical errors are negligible compared to the decrease of the free energy barrier of the $\text{Y61} \rightarrow \text{G}$ mutation compared to the WT.

Transition state analysis

The structures near the transition state (TS) are of considerable importance for understanding the possible origin of the TS stabilization. The TS in this work is approximated as the average configuration that possesses the highest free energy on the 1D curve or the saddle point on the 2D map (for simplicity, we will call the TS structures/configurations). The typical TS configurations of glycosylation and deglycosylation for WT at $85 \text{ }^\circ\text{C}$ are presented in Fig. 3; the structures from both the 2D (Fig. 3a, b) and 1D simulations (Fig. 3c, d) are given. Consistent with the free energy profiles/surfaces, the corresponding 1D and 2D TS configurations are in good agreement with each other.

For both steps of the reaction, the sugar ring at the -1 position is distorted to a similar, high-energy ${}^4\text{H}_3$ half-chair configuration (see below) that bears the sp^2 hybridized anomeric carbon atom. This configuration has been previously reported for the TS structures of cellulases of the GH5, GH7, GH18 and GH20 families (Davies et al. 2003; Barnett et al. 2011).

Fig. 4 Conformational evolution of the -1 glucose ring during the reaction. **a** glycosylation at $37\text{ }^{\circ}\text{C}$; **b** deglycosylation at $37\text{ }^{\circ}\text{C}$; **c** glycosylation at $85\text{ }^{\circ}\text{C}$; **d** deglycosylation at $85\text{ }^{\circ}\text{C}$. ${}^4\text{C}_1$, E_3 and ${}^4\text{H}_3$ conformers are shown in *black (solid)*, *red (dashed)* and *blue (dotted)*, respectively. (Color figure online)



A hydrogen bond between O(E116) and O(E134) and another one between O(E134) and O2(cellulose) were observed in both TS configurations (Fig. 3). In the TS for glycosylation, the proton is located between E231 and O(cellulose). For deglycosylation, the formation of the new C–O bond between the C1 and attacking water molecule seems to take place earlier than the proton transfer from the water molecule to O(E231) (see above). These are also reflected in the TS structures which show that the PT is concerted with NA in the glycosylation and ahead of PT in the deglycosylation (Fig. 3).

Glucose ring conformations

The hydrolysis reaction catalyzed by the cellulases involves the evolution of the glucose ring conformation at the -1 position. The conformation of the six-membered pyranose ring can be classified into 38 basic conformations: 2 chairs, 6 boats, 6 skew-boats, 12 half-chairs and 12 envelopes (Ionescu et al. 2005; Biarnes et al. 2007). During a reaction, the glucose ring evolves from a stable conformation (reactant or intermediate) via possible higher energy conformations to other stable ones (product or intermediate).

This may also involve the transitions between sp^2 and sp^3 hybridized states of the anomeric carbon at the -1 position.

Conformational changes of the -1 glucose ring during both glycosylation and deglycosylation are presented in Fig. 4. In both steps, the conformational changes of the -1 glucose ring lead to a $\beta \rightarrow \alpha \rightarrow \beta$ conversion at the anomeric carbon atom. It is clear that the reactant, product, and cello-enzyme intermediate all adopt the ${}^4\text{C}_1$ chair conformation, which is one of the two most stable conformations (the other one being the ${}^1\text{C}_4$ chair). In contrast, in a previous crystallographic analysis (Davies et al. 2003), the ${}^4\text{H}_3$ and ${}^3\text{H}_4$ half-chair conformations and ${}^{2,5}\text{B}$ and $\text{B}_{2,5}$ boat conformations were considered as four possible transition-state configurations. However, in a recent QM/MM study (Barnett et al. 2011), both the E_3 envelop and ${}^4\text{H}_3$ half-chair conformations were observed in the transition states of cellobiose hydrolyase I (CBHI) of the GH7 family. This seems to be reasonable based on the results from previous *ab initio* free energy landscape calculations on β -D-glucopyranose (Biarnes et al. 2007; Jeffrey and Yates 1979). These studies showed that the positions of both conformations were close to the 2nd lowest free energy minimum [M2 in Biarnes et al. (2007)], which

was only $2.6 \text{ kcal} \cdot \text{mol}^{-1}$ higher than the global minimum of ${}^4\text{C}_1$ chair conformation.

Here, the calculated transition states of both the glycosylation and deglycosylation steps are in either the E_3 envelop or ${}^4\text{H}_3$ half-chair conformations. E_3 is slightly more stable than ${}^4\text{H}_3$, especially in the deglycosylation step (Fig. 4b).

Simulations of WT and mutants, and origin of high activity at high temperature

A close examination of the structures of *Tm*Cel12A and its mutants reveals certain features that might contribute to the hyperthermophilic nature of the enzyme. In particular, the side chains of E116, E134 and E227, together with the active-site water molecule, form a stable structural moiety that interacts with the cellulose via a series of stable hydrogen bonds at both low (37°C) and high (85°C) temperatures. Notable ones among these hydrogen bond are those between E134 and the glucose ring at the -1 position and between the water molecule and the sugar ring at the -2 position (Fig. 5a, b).

In wild type (WT) *Tm*Cel12A, E116 and the water molecule may play an important role in organizing and

stabilizing the structural moiety mentioned above. QM/MM MD simulations were also performed on the E116G mutant, in which an additional water molecule was found to occupy the position of the E116 carboxylate in the WT enzyme, with the formation of a different hydrogen bond network (Fig. 5c, d). For the E116D mutant, a stable unit comprising D116, E227 and a water molecule (Fig. 5e, f) was observed in the simulations. In both the E116G and E116D mutants, the interaction between E134 and cellulose remains. However, the original hydrogen-bond networks connecting E134 and E227 in the WT (through E116 and the active-site water molecule) are lost. Moreover, for the simulation at 85°C , the hydrogen bond networks in E116G were broken during the simulations, and the substrate moved away from the active site (within the first 50 ps, Fig. 5d). The active-site structure of the E116D mutant seems to be less flexible than that of E116G, but failed to “hold” the substrate in the reactive conformation after a 500 ps QM/MM MD simulation (Fig. 5f).

Table 1 summarizes the root-mean-square fluctuations (RMSFs), the average distances and the lifetime of the interactions between E134, the general acid (E231) and the substrate in simulations at 37 and

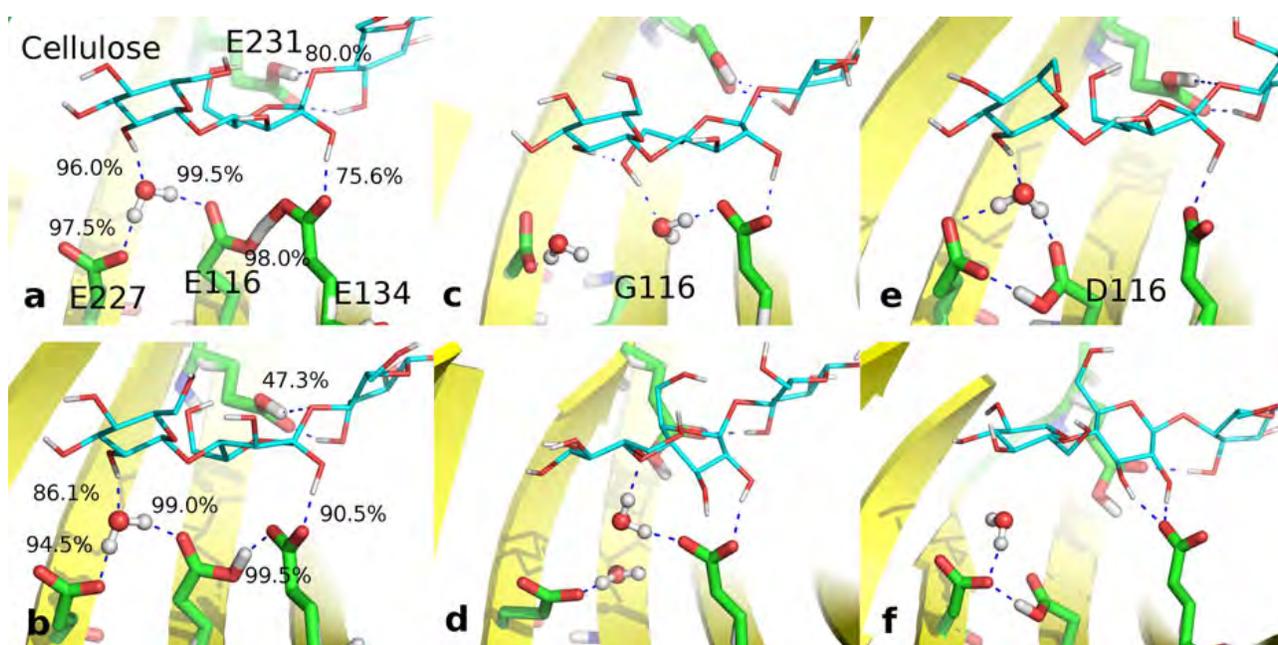


Fig. 5 Thermostability of the structural moiety formed by E116, E134, E227 and the active-site water molecule. **a** WT at 37°C ; **b** WT at 85°C ; **c** E116G at 37°C ; **d** E116G at 85°C ; **e** E116D at 37°C ; **f** E116D at 85°C . Lifetime of the hydrogen

bonds in WT are shown in **a**, **b**. Hydrogen bonds with the distance between two heavy atoms less than the threshold of 3.0 \AA are taken into account

Table 1 Root-mean-square fluctuations, average distances and the lifetime of the interactions between catalytic residues and substrate from the QM/MM MD simulations at 37 and 85 °C, respectively

Group	T (°C)	RMSF (Å)				Distance (Å)		Lifetime (%)	
		E116	E134	E227	Substrate	E134-Sub ^a	E231-Sub ^b	E134-Sub	E231-Sub
WT	37	0.05 ± 0.01	1.59 ± 0.02	3.15 ± 0.06	0.54 ± 0.41	3.78 ± 0.21	1.95 ± 0.21	47.8	60.2
	85	0.06 ± 0.01	1.91 ± 0.05	3.58 ± 0.06	0.48 ± 0.31	3.55 ± 0.28	2.47 ± 0.86	80.6	24.9
E116G	37	–	1.30 ± 0.05	3.09 ± 0.06	0.41 ± 0.22	3.22 ± 0.21	2.43 ± 0.36	100	10.9
	85	–	2.61 ± 0.07	6.54 ± 0.16	0.51 ± 0.31	5.19 ± 0.36	4.82 ± 0.76	0.5	0.0
E116D	37	0.05 ± 0.01	1.39 ± 0.03	3.21 ± 0.09	0.47 ± 0.29	3.33 ± 0.28	2.25 ± 0.43	100	8.62
	85	4.07 ± 0.06	1.36 ± 0.03	3.32 ± 0.06	0.49 ± 0.32	3.83 ± 0.29	3.60 ± 1.31	38.5	0.0

^a Distance from OE2 of Glu134 to C1 of the sugar ring at –1 position

^b Distance from the proton of Glu231 to the linkage oxygen atom between the –1 and +1 glucoses

85 °C. Mass-weighted side chain atoms were used to calculate the RMSF of each residue. In order to show the distances between the catalytic residues and the substrate intuitively, the ‘lifetime’ of an interaction here is monitored based on the percentage of its appearance during the whole simulation. The averaged distances between the two catalytic residues, E134 and E231, and the substrate (cellulose) at 37 °C are 3.78 and 1.95 Å, and these distances were used as thresholds for estimating the lifetimes of the interactions between the two glutamates and the substrate. For WT, E116, E134 and E227 are relatively stable at both temperatures. For the E116G mutant, the flexibilities of E134 and E227 are found to be much larger at 85 °C than at 37 °C, and the increased distances between E134/E231 and the substrate in this mutant at 85 °C indicate that the substrate has moved away from the active site. For the E116D mutant, although the fluctuations of E134 and E227 are similar at both 37 and 85 °C (and to those in the WT), the distances between the substrate and the catalytic residues became larger, suggesting that the thermostability of the E116D mutant may decrease, with lower activity relative to WT. In summary, based on these observations and the PMF calculations at different temperatures, it seems that the stable structural moiety comprising E116, E134, E227 and the active site water may be an important factor contributing to the high activity of *TmCel12A* at high temperatures.

Bioinformatics sequence analysis of GH12 domain

To further probe the origin of the hyperthermophilicity, a bioinformatics survey was performed for the

catalytic domain (residues 89–262) of *TmCel12A*. For this work, the Pfam database (Punta et al. 2012), ClustalW2 (Larkin et al. 2007) and WebLogo 3 (Crooks et al. 2004) were used to explore the sequence conservation of this domain. 316 out of 386 sequences from mesophilic or psychrophilic microbes were classified as non-hyperthermophilic domains (Fig. 6a), whereas 29 sequences from microbes with optimal growth temperatures above 75 °C were considered as hyperthermophilic domains (Fig. 6b).

As is shown in Fig. 6, in addition to some residues that are known to form a hydrophobic pocket around the active site, several hydrophilic residues are highly conserved among the sequences of all cellulases, including those corresponding to E116, E134, E227 and E231 in *TmCel12A*. As the key residues for the catalysis, the residues corresponding to E134 and E231 are highly conserved among all glycoside hydrolase sequences. Interestingly, Asp at the D116 position dominates in non-hyperthermophilic microorganisms, while Glu at the E116 position is preferred within the hyperthermophilic species including *Thermotoga*, *Thermococcus* and *Pyrococcus*. Glu at the E227 position is also more conserved in hyperthermophilic microbes. Therefore, this survey seems to be consistent with the QM/MM MD simulations indicating that E116 may play an important role in organizing a stable structural moiety comprising E116, E134, E227 and the water molecule so as to contribute to the hyperthermophilicity of *TmCel12A*. This hypothesis deserves further experimental and theoretical testing in order to facilitate the design of thermostable cellulases with high activity for application in industrial biofuel production.

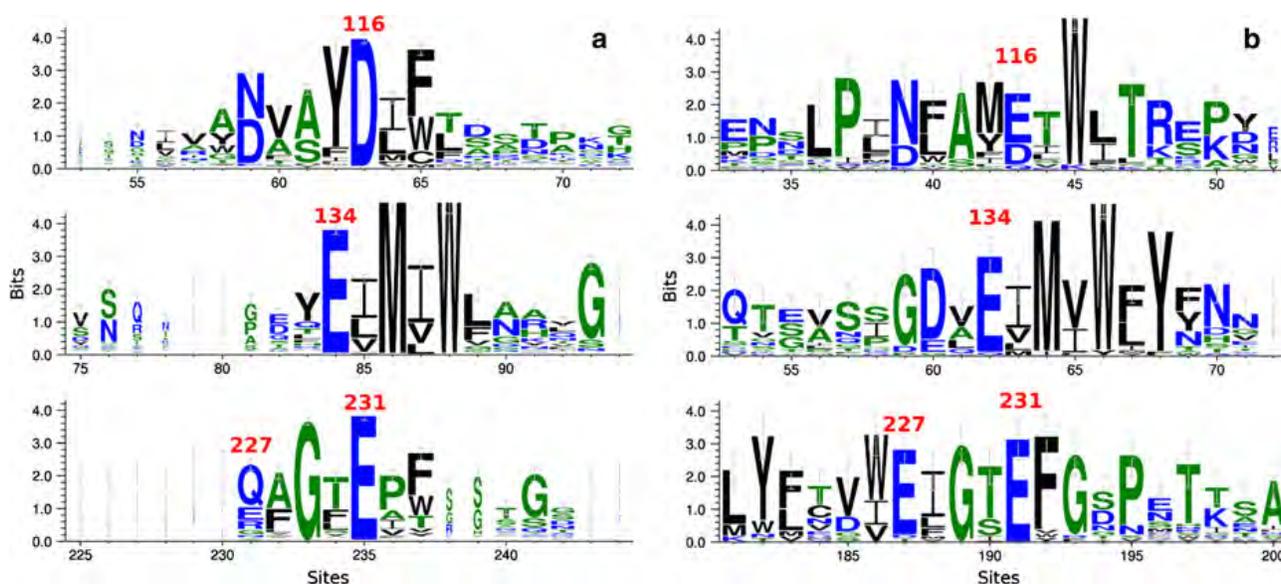


Fig. 6 Sequence conservation of GH12-family domains without (a) and with (b) the hyperthermophilic characteristics. **a** Represents conservation of the hydrolysis domain among

mesophilic and psychrophilic microbes and **b** for that among hyperthermophilic organisms. The residue numbers are based on those of *TmCel12A*

Conclusions

Quantum mechanical/molecular mechanical free energy calculations were performed to understand the mechanistic details of the catalysis of the cellulase *TmCel12A*. The simulations provide some clues to the possible origin of the high activity at high temperature of *TmCel12A*. At 85 °C, the calculated free energy barrier of deglycosylation is higher than that of glycosylation. A nucleophilic attack (NA) accompanied by a proton transfer (PT) was found in both glycosylation and deglycosylation steps of the *TmCel12A* hydrolysis. In the glycosylation step, the NA and the PT processes occurred synchronously based on the simulations, whereas in deglycosylation the NA occurred ahead of the PT. In both steps, an inversion of the glucose ring at the -1 position was observed. Both E_3 and 4H_3 were found to be possible conformations at the transition states. According to both the QM/MM MD simulations and bioinformatics analysis, a stable hydrogen-bonded structural moiety containing the side chains of E116, E134, E227 and an active-site water molecule may be one of the factors contributing to the hyperthermophilic nature of *TmCel12A*, although other factors may be involved as well and still need to be identified.

Acknowledgments This work is supported in part by the National Science Foundation Award (Grant No. 0817940 to H.G.) and in part by grants from the National High-Tech R&D Program (863 Program Contract No. 2012AA020307 to D.Q.W.), the National Basic Research Program of China (973 Program) (Contract No. 2012CB721000 to D.Q.W.), and the Key Project of Shanghai Science and Technology Commission (Contract No. 11JC1406400 to D.Q.W.). P.L. is supported by a fellowship from Shanghai Jiao Tong University. JCS acknowledges support from the Bioenergy Science Center, funded by Biological and Environmental Research in the Office of Science of the U.S. Department of Energy.

References

- Anbar M, Gul O, Lamed R, Sezerman UO, Bayer EA (2012) Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. *Appl Environ Microb* 78(9):3458–3464
- Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC (2011) High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl Environ Microb* 77(23):8288–8294
- Banerjee A, Yang W, Karplus M, Verdine GL (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature* 434(7033):612–618
- Barnett CB, Naidoo KJ (2010) Ring puckering: a metric for evaluating the accuracy of AM1, PM3, PM3CARB-1, and SCC-DFTB carbohydrate QM/MM simulations. *J Phys Chem B* 114(51):17142–17154. doi:10.1021/jp107620h

- Barnett CB, Wilkinson KA, Naidoo KJ (2010) Pyranose ring transition state is derived from cellobiohydrolase I induced conformational stability and glycosidic bond polarization. *J Am Chem Soc* 132(37):12800–12803. doi:10.1021/ja103766w
- Barnett CB, Wilkinson KA, Naidoo KJ (2011) Molecular details from computational reaction dynamics for the cellobiohydrolase I glycosylation reaction. *J Am Chem Soc* 133(48):19474–19482
- Biarnes X, Ardevol A, Planas A, Rovira C, Laio A, Parrinello M (2007) The conformational free energy landscape of beta-D-glucopyranose. Implications for substrate preactivation in beta-glucoside hydrolases. *J Am Chem Soc* 129(35):10686–10693. doi:10.1021/ja068411o
- Bondar AN, Elstner M, Suhai S, Smith JC, Fischer S (2004) Mechanism of primary proton transfer in bacteriorhodopsin. *Structure* 12(7):1281–1288
- Bronnenmeier K, Kern A, Liebl W, Staudenbauer WL (1995) Purification of *Thermotoga maritima* enzymes for the degradation of cellulosic materials. *Appl Environ Microb* 61(4):1399–1407
- Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4(2):187–217
- Brooks CL, Brunger A, Karplus M (1985) Active-site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers* 24(5):843–865
- Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caffisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodosek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30(10):1545–1614
- Cheng YS, Ko TP, Huang JW, Wu TH, Lin CY, Luo WH, Li Q, Ma YH, Huang CH, Wang AHJ, Liu JR, Guo RT (2012) Enhanced activity of *Thermotoga maritima* cellulase 12A by mutating a unique surface loop. *Appl Microbiol Biotechnol* 95(3):661–669
- Choi TH, Liang R, Maupin CM, Voth GA (2013) Application of the SCC-DFTB method to hydroxide water clusters and aqueous hydroxide solutions. *J Phys Chem B* 117(17):5165–5179. doi:10.1021/jp400953a
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14(6):1188–1190
- Davies GJ, Ducros VMA, Varrot A, Zechel DL (2003) Mapping the conformational itinerary of beta-glycosidases by X-ray crystallography. *Biochem Soc Trans* 31:523–527
- Demain AL, Newcomb M, Wu JHD (2005) Cellulase, clostridia, and ethanol. *Microbiol Mol Biol Rev* 69(1):124–154
- Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* 58(11):7260–7268
- Elstner M, Frauenheim T, Suhai S (2003) An approximate DFT method for QM/MM simulations of biological structures and processes. *J Mol Struct THEOCHEM* 632:29–41
- Gaus M, Cui QA, Elstner M (2011) DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput* 7(4):931–948
- Guo H, Cui Q, Lipscomb WN, Karplus M (2001) Substrate conformational transitions in the active site of chorismate mutase: their role in the catalytic mechanism. *Proc Natl Acad Sci USA* 98(16):9032–9037
- Guo HB, Rao N, Xu Q, Guo H (2005a) Origin of tight binding of a near-perfect transition-state analogue by cytidine deaminase: implications for enzyme catalysis. *J Am Chem Soc* 127(9):3191–3197
- Guo HB, Wlodawer A, Guo H (2005b) A general acid–base mechanism for the stabilization of a tetrahedral adduct in a serine-carboxyl peptidase: a computational study. *J Am Chem Soc* 127(45):15662–15663. doi:10.1021/ja0520565
- Guo HB, Wlodawer A, Nakayama T, Xu Q, Guo H (2006) Catalytic role of proton transfers in the formation of a tetrahedral adduct in a serine carboxyl peptidase. *Biochemistry* 45(30):9129–9137
- Haki GD, Rakshit SK (2003) Developments in industrially important thermostable enzymes: a review. *Bioresour Technol* 89(1):17–34
- Himmel ME, Ruth MF, Wyman CE (1999) Cellulase for commodity products from cellulosic biomass. *Curr Opin Biotechnol* 10(4):358–364
- Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315(5813):804–807
- Ionescu AR, Berces A, Zgierski MZ, Whitfield DM, Nukada T (2005) Conformational pathways of saturated six-membered rings. A static and dynamical density functional study. *J Phys Chem A* 109(36):8096–8105
- Islam SM, Roy P-N (2012) Performance of the SCC-DFTB model for description of five-membered ring carbohydrate conformations: comparison to force fields, high-level electronic structure methods, and experiment. *J Chem Theory Comput* 8(7):2412–2423
- Javed MM, Ikram-ul-Haq, Mariyam I (2011) Multistep Mutagenesis for the over-expression of cellulase in *Humicola insolens*. *Pak J Bot* 43(1):669–677
- Jeffrey GA, Yates JH (1979) Stereographic representation of the Cremer-Pople ring-puckering parameters for pyranoid rings. *Carbohydr Res* 74:319–322
- Jorgensen WL (1981) Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers: application to liquid water. *J Am Chem Soc* 103(2):335–340
- Kang HJ, Uegaki K, Fukada H, Ishikawa K (2007) Improvement of the enzymatic activity of the hyperthermophilic cellulase from *Pyrococcus horikoshii*. *Extremophiles* 11(2):251–256
- Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* 13(8):1011–1021
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947–2948

- Lee CY, Yu KO, Kim SW, Han SO (2010) Enhancement of the thermostability and activity of mesophilic *Clostridium cellulovorans* EngD by in vitro DNA recombination with *Clostridium thermocellum* CelE. *J Biosci Bioeng* 109(4): 331–336
- Li GH, Cui Q (2003) What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. *J Am Chem Soc* 125(49):15028–15038
- Liang CN, Fioroni M, Rodriguez-Ropero F, Xue YF, Schwaneberg U, Ma YH (2011a) Directed evolution of a thermophilic endoglucanase (Cel5A) into highly active Cel5A variants with an expanded temperature profile. *J Biotechnol* 154(1):46–53
- Liang CN, Xue YF, Fioroni M, Rodriguez-Ropero F, Zhou C, Schwaneberg U, Ma YH (2011b) Cloning and characterization of a thermostable and halo-tolerant endoglucanase from *Thermoanaerobacter tengcongensis* MB4. *Appl Microbiol Biotechnol* 89(2):315–326
- Liebl W, Ruile P, Bronnenmeier K, Riedel K, Lottspeich F, Greif I (1996) Analysis of a *Thermotoga maritima* DNA fragment encoding two similar thermostable cellulases, CelA and CelB, and characterization of the recombinant enzymes. *Microbiology* 142:2533–2542
- Liu JL, Wang XM, Xu DG (2010) QM/MM study on the catalytic mechanism of cellulose hydrolysis catalyzed by cellulase Cel5A from *Acidothermus cellulolyticus*. *J Phys Chem B* 114(3):1462–1470. doi:10.1021/jp909177e
- Liu JR, Cheng YS, Ko TP, Wu TH, Ma YH, Huang CH, Lai HL, Wang AHJ, Guo RT (2011) Crystal structure and substrate-binding mode of cellulase 12A from *Thermotoga maritima*. *Proteins* 79(4):1193–1204
- Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415
- Maupin CM, Aradi B, Voth GA (2010) The self-consistent charge density functional tight binding method applied to liquid water and the hydrated excess proton: benchmark simulations. *J Phys Chem B* 114(20):6922–6931. doi:10.1021/jp1010555
- Murashima K, Kosugi A, Doi RH (2002) Thermostabilization of cellulosomal endoglucanase EngB from *Clostridium cellulovorans* by in vitro DNA recombination with non-cellulosomal endoglucanase EngD. *Mol Microbiol* 45(3): 617–626
- Nakazawa H, Okada K, Onodera T, Ogasawara W, Okada H, Morikawa Y (2009) Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. *Appl Microbiol Biotechnol* 83(4):649–657
- Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson LD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL, Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399(6734):323–329
- Niehaus TA, Elstner M, Frauenheim T, Suhai S (2001) Application of an approximate density-functional method to sulfur containing compounds. *J Mol Struct (Thoechem)* 541(1):185–194
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Bournsall C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40(D1):D290–D301
- Riccardi D, Konig P, Guo H, Cui Q (2008) Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor. *Biochemistry* 47(8): 2369–2378
- Saharay M, Guo H, Smith JC (2010) Catalytic mechanism of cellulose degradation by a cellobiohydrolase, CelS. *PLoS ONE* 5(10):e12947. doi:10.1371/journal.pone.0012947
- Sattelle BM, Almond A (2010) Less is more when simulating unsulfated glycosaminoglycan 3D-structure: comparison of GLYCAM06/TIP3P, PM3-CARB1/TIP3P, and SCC-DFTB-D/TIP3P predictions with experiment. *J Comput Chem* 31(16):2932–2947. doi:10.1002/jcc.21589
- Sattelmeyer KW, Tirado-Rives J, Jorgensen WL (2006) Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules. *J Phys Chem A* 110(50):13551–13559
- Schou C, Rasmussen G, Kaltoft MB, Henrissat B, Schulein M (1993) Stereochemistry, specificity and kinetics of the hydrolysis of reduced cellodextrins by 9 cellulases. *Eur J Biochem* 217(3):947–953
- Shaw AJ, Podkaminer KK, Desai SG, Bardsley JS, Rogers SR, Thorne PG, Hogsett DA, Lynd LR (2008) Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. *Proc Natl Acad Sci USA* 105(37):13769–13774
- Srikrishnan S, Randall A, Baldi P, Da Silva NA (2012) Rationally selected single-site mutants of the *Thermoascus aurantiacus* endoglucanase increase hydrolytic activity on cellulosic substrates. *Biotechnol Bioeng* 109(6):1595–1599
- Torrie GM, Valleau JP (1974) Monte-Carlo free-energy estimates using non-Boltzmann sampling: application to subcritical Lennard-Jones fluid. *Chem Phys Lett* 28(4): 578–581
- Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65(1):1–43
- Viikari L, Alapuranen M, Puranen T, Vehmaanpera J, Siika-Aho M (2007) Thermostable enzymes in lignocellulose hydrolysis. *Adv Biochem Eng Biotechnol* 108:121–145
- Volfova O, Suchardova O, Panos J, Krumphanzl V (1985) Ethanol formation from cellulose by thermophilic bacteria. *Appl Microbiol Biotechnol* 22(4):246–248
- Xu Q, Guo H, Wlodawer A, Guo H (2006) The importance of dynamics in substrate-assisted catalysis and specificity. *J Am Chem Soc* 128(18):5994–5995
- Xue DS, Chen HY, Ren YR, Yao SJ (2012) Enhancing the activity and thermostability of thermostable beta-glucosidase from a marine *Aspergillus niger* at high salinity. *Process Biochem* 47(4):606–611
- Yang B, Wyman CE (2004) Effect of xylan and lignin removal by batch and flowthrough pretreatment on the enzymatic digestibility of corn stover cellulose. *Biotechnol Bioeng* 86(1):88–95
- Yang Y, Yu H, York D, Elstner M, Cui Q (2008) Description of phosphate hydrolysis reactions with the self-consistent-

- charge density-functional-tight-binding (SCC-DFTB) theory. 1. Parameterization. *J Chem Theory Comput* 4(12): 2067–2084. doi:[10.1021/ct800330d](https://doi.org/10.1021/ct800330d)
- Yi ZL, Pei XQ, Wu ZL (2011) Introduction of glycine and proline residues onto protein surface increases the thermostability of endoglucanase CelA from *Clostridium thermocellum*. *Bioresour Technol* 102(3):3636–3638
- Zhang YHP, Himmel ME, Mielenz JR (2006) Outlook for cellulase improvement: screening and selection strategies. *Biotechnol Adv* 24(5):452–481