PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Strategies to reduce end-product inhibition in family 48 glycoside hydrolases

Mo Chen,[1] Lintao Bu,[2] Markus Alahuhta,[2] Roman Brunecky,[2] Qi Xu,[2] Vladimir V. Lunin,[2] John W. Brady,[1] Michael F. Crowley,[2] Michael E. Himmel,[2] and Yannick J. Bomble[2]*

[1] Department of Food Science, Cornell University, Ithaca, New York

[2] National Renewable Energy Laboratory, Golden, Colorado

ABSTRACT

Family 48 cellobiohydrolases are some of the most abundant glycoside hydrolases in nature. They are able to degrade cellulosic biomass and therefore serve as good enzyme candidates for biofuel production. Family 48 cellulases hydrolyze cellulose chains via a processive mechanism, and produce end products composed primarily of cellobiose as well as other cellooligomers (dp $\leq$ 4). The challenge of utilizing cellulases in biofuel production lies in their extremely slow turnover rate. A factor contributing to the low enzyme activity is suggested to be product binding to enzyme and the resulting performance inhibition. In this study, we quantitatively evaluated the product inhibitory effect of four family 48 glycoside hydrolases using molecular dynamics simulations and product expulsion free-energy calculations. We also suggested a series of single mutants of the four family 48 glycoside hydrolases with theoretically reduced level of product inhibition. The theoretical calculations provide a guide for future experimental studies designed to produce mutant cellulases with enhanced activity.

Proteins 2016; 84:295–304.

Key words: glycoside hydrolases; product inhibition; biofuels; cellulose; molecular dynamics.

INTRODUCTION

Lignocellulosic biomass exists abundantly in nature, and is considered the primary resource for producing biofuel. In principle, conversion of lignocellulose into valuable products involves three steps: a pretreatment that separates and removes the lignin and hemicellulose components, an enzymatic treatment that hydrolyzes cellulose into soluble cellodextrins and glucose; and finally, fermentation that converts these sugars to desired products. The biomass conversion process, nevertheless, still remains somewhat inefficient due to biomass recalcitrance and low enzyme specific activity. The prevalence of lignin molecules in biomass, interacting with the cellulose, strongly hinders cellulose degradation.[1] The insolubility of crystalline cellulose in aqueous and most organic solvents further limits its accessibility to the hydrolytic enzymes. Additionally, the catalytic activities of cellulases, which hydrolyze cellulose to soluble oligomers (DP $\leq$ 6), are particularly lower compared to enzymes that work on soluble sugars. It has been reported that high loadings of enzymes ($\sim$25 kg/ton of cellulose) are required to release most of the sugars from biomass at rates compatible with high-throughput processes, and the requirement for such unusually high protein loadings appears to be the single largest cost in the production of cellulosic biofuels, not including the cost of biomass itself.[2] Increasing the catalytic activity of cellulases can potentially reduce the cost of the process and improve the overall biomass conversion rate. To achieve these goals, rational design based on an understanding of the structure–function relationships of various cellulases and their interactions with cellulose remains a useful approach.

Glycoside hydrolase family 48 (GH48) are a major group of processive exocellulases that catalyze cellulose hydrolysis from the reducing end and produce mostly cellobiose molecules. More than twenty GH48s have been isolated from various microorganisms,[3] but among them, only a few have a corresponding X-ray crystal structure. These structures share common features, including an $(\alpha/\alpha)_6$ barrel structure, and a tryptophan-

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Y. J. Bomble; E-mail: Yannick.bomble@nrel.gov

Received 30 July 2015; Revised 2 October 2015; Accepted 6 November 2015
Published online 1 February 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24965

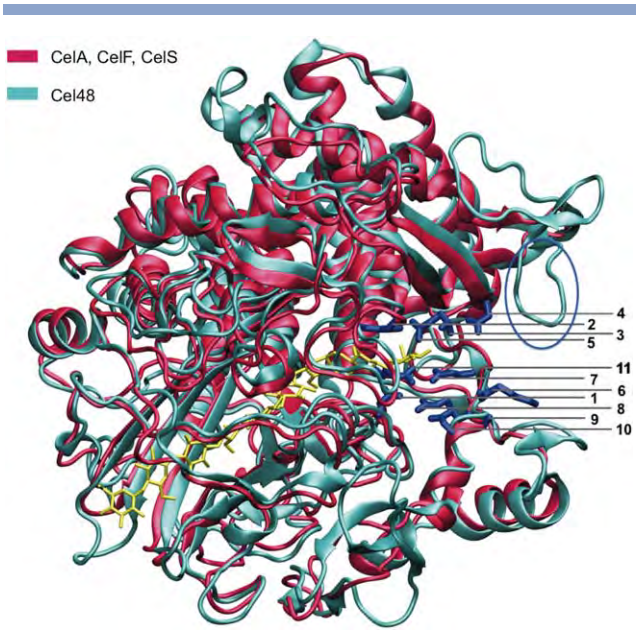

Figure 1

Superimposition of the crystal structures of family 48 glycoside hydrolases. Note that CelA (4EL8), CelS (1L2A), and CelF (2QNO) share very similar structures. One representative structure is shown in red. Cel48 (5BV9) is shown in cyan, and it exhibits several longer loops compared to the other three structures. In particular, one extra loop in Cel48 is located at the tunnel exit and is highlighted by a blue circle. The cellooligomer in the active site tunnel is shown using yellow sticks. The conserved or partially conserved amino acid residues (see Table I) that form strong interactions with the cellobiose product at the tunnel exit are shown in blue sticks with CelA taken as an example where the labels refer to the Group ID in Table I.

rich active site tunnel providing seven-substrate subsites preceding the hydrolytic cleavage site and two subsites following it at the tunnel exit. In particular, CelA from *Caldicellulosiruptor bescii*,[4] CelS from *Clostridium thermocellum*,[5] and CelF from *Clostridium cellulolyticum*[6–8] have comparable structures; whereas Cel48 from *Bacillus pumilus* exhibits longer loops. Most of these loops are far from the product site except for one, which is a turn composed of residues SDEEGYF starting at position 466 (Fig. 1). However, besides these loops, the structure of Cel48 is extremely close to that of CelA, CelS, and CelF. Regarding catalytic activity, Cel48 and CelF favor mesophilic conditions, CelS is thermophilic, and CelA is extremely thermophilic in that it exhibits optimal activity at 85°C and sustains activity at high temperatures (up to 90°C).[4] The thermophilic cellulases are particularly interesting as they can be added directly to cellulosic biomass immediately after pretreatment, where the temperature remains high, increasing energy efficiency.[9] In addition, CelS and CelF are critical components of cellulosomes,[10] which are self-assembled complexes of cellulases. Cellulosomes have been shown to have high cellulolytic activity on biomass degradation.[11] For example, the GH48 enzyme, CelS, plays an essential role in cellulose digestion and utilization by *C. thermocellum*.[12]

The active site tunnel of GH48 enzymes provides nine substrate subsites serving as a substrate pathway for processive action. These subsites are named as: −7, −6, −5, −4, −3, −2, −1, +1, +2 from the substrate nonreducing end (−7) at the tunnel entrance to the reducing end (+2) at the tunnel exit. It is generally believed that family 48 cellulases can recognize the cellulose chains by their reducing end, and acquire them into the active site tunnel. Subsequently, the cellulose chain processes through the tunnel until it is in position for hydrolytic reaction. Family 48 cellulases follow a concerted, inverting mechanism.[3,13,14] In particular, they use a catalytic acid (glutamic acid) and a catalytic base (aspartic acid) to achieve hydrolysis of glycosidic bonds in cellulose. As a result, a cellobiose product is hydrolyzed and released into the aqueous environment. Next, the cellulose chain processes forward in the tunnel by a cellobiose unit, which continues the catalytic cycle. At some point, the enzymes dissociate from the cellulose substrate, halting the processive cycle. The enzymatic activities of some family 48 cellulases are reported to be extremely low. For example, the activities of *T. fusca* Cel48A acting on swollen cellulose, CM-cellulose, BMCC, and filter article are 0.405, 0.292, 0.191, and 0.068 microMol CB/min /microMol enzyme, respectively.[15] It has been speculated that the small turnover number of GH48 enzymes is due to inefficient acquisition of cellulose by the tunnel entrance, slow processivity of the cellulose substrate in the tunnel, and end-product inhibition. This study focused on understanding the effect of end-product inhibition, in order to provide strategies for improving activity on cellulose.

Several studies have reported that the end-product, cellobiose, strongly inhibits the activity of the family 48 cellulases, such as *C. thermocellum* CelS[16–18] and *T. fusca* Cel48A.[15] Almost complete inhibition of the *C. thermocellum* cellulosome, in which CelS is a major component, was reported at a concentration of 2% (w:v)

Table I

Residues at the Tunnel Exit of the Four Family 48 Glycoside Hydrolases

Group ID	CelA	Cel48	CelF	CelS
1	Glu44	Glu38	Glu44	Glu76
2	Trp412	Trp411	Trp411	Trp439
3	Thr463	Gly500	Thr462	Asp490
4	Asp489	Asp530	Asn490	Ser516
5	Asp493	Asp534	Asp494	Asp520
6	Lys547	Glu591	Gln543	Ala577
7	Glu546	Glu590	Glu542	Glu576
8	Arg548	Arg592	Arg544	Arg578
9	Ala549	Glu593	Gly545	Ala579
10	Asp550	Asp594	Asp546	Asp580
11	Arg613	Arg682	Arg609	Arg643

The residues on the same row are at homologous locations according to protein sequence alignment using the position-specific iterated BLAST method.[41]

cellobiose.[17] Note that this concentration is much lower than the solubility of cellobiose in water (12% w:v). It was postulated that cellobiose could competitively bind to both the tunnel entrance and the tunnel exit. Since it is desirable for cellulases to possess sufficient binding affinity at the tunnel entrance for substrate recognition and acquisition, this study focused on how to reduce the product inhibitory effects at the tunnel exit.

For GH48 enzymes, the mechanism of end-product inhibition at the molecular level has not been fully determined by experimental measurements. As indirect evidence, Zhang and others found that the initial attack of cellulose by *Clostridium phytofermentans* Cel48 generated a perceptible amount of cellotetraose (7% on crystalline cellulose and 4% on amorphous cellulose) and cellotriose (15% on crystalline cellulose and 9% on amorphous cellulose), in addition to the major product cellobiose.[19] It is hypothesized that in the transition state, stabilization of the substrate on the product side is required before initiation of substrate cleavage. Hence, the minor production of cellotriose and cellotetraose at the initial attack of cellulose chains might indicate their weak binding to the tunnel exit. In addition, crystal formation of family 48 cellulases suggests that the tunnel exit possesses strong sugar-binding affinity. For example, the presence of cellobiose is required in forming CelS crystals; and when growing CelS crystals in cellobiose, the solved crystal structure contains a cellobiose only at the tunnel exit, taking up subsites +1 and +2.[5] In the CelF crystal structures, subsites +1, +2, and +3 were identified when using cellotriose and cellotetraose as inhibitors, indicating sufficient sugar-binding potential at the tunnel exit.[7]

Characterization of family 48 crystal structures demonstrates that many residues at the tunnel exit mostly conserved, including a Trp located at subsites +1 and +2; as well as several charged residues, including Arg, Asp, and Glu residing along the tunnel exit. The Trp side-chain stacks onto the two β-glucosyl units at the product side, and might play an important role in stabilizing the substrate in the transition state. It is likely that substituting the Trp at this site can weaken the binding between the product and the tunnel exit. Indeed, for the endocellulase Cel5A from *Acidothermus cellulolyticus*, when mutating the product-binding Tyr 245 into Gly, the catalytic rate was increased by 40% and the inhibitor constant, K_i, was increased to 1480%.[20] However, mutating the product-binding Trp in exocellulases might weaken the processivity of the exocellulases as well, which is not desirable for crystalline cellulose degradation. Furthermore, charged residues can form strong electrostatic interactions with the cellobiose product via hydrogen bonding, hindering its escape into the aqueous environment. More recently, several mutations that were previously predicted computationally at the product site of a family 7 cellulase were shown to increase tolerance to cellobiose.[21]

We used atomic modeling and molecular dynamics simulations of family 48 glycoside hydrolases to better understand end-product inhibition, to evaluate product expulsion energies in the four crystallized family 48 glycoside hydrolases, and to propose rationally designed mutants with reduced end-product inhibition.

METHODS

Structure preparations and molecular dynamics simulations

The atomic models of the glycoside hydrolases were built based on their X-ray crystal structures obtained from the PDB database: Cel48 (5BV9), CelA (4EL8), CelF (2QNO), and CelS (1L2A). The CelF wildtype structure was converted from the crystal structure of its mutant E55Q with the substrate in the active site tunnel following a lower pathway.[8] A celloheptaose and a cellobiose from the crystal structures were placed in the active site tunnel taking the positions from subsite −7 to −2 and from subsite +1 to +2, respectively, representing the state immediately after the hydrolysis. Each of the systems was simulated in an octahedral water box, and the net charge was balanced to be neutral by counter ions.

The CHARMM22 force field[22] with the CMAP correction[23] was used to describe the protein; the CHARMM36 all-atom carbohydrate force field[24] was used for the cellooligomer, and TIP3P served as the water model.[25] The CHARMM program,[26,27] was used to build the molecular systems. The tool CHAMBER[28] was used to convert the coordinate and structural files and the associated force fields in CHARMM format into AMBER format. The PMEMD engine of AMBER[29] was used to carry out the molecular dynamics simulations. For each system, the box size for production run (NVT) and number of water molecules are listed below.

The system preparations prior to production runs included four steps: solvent minimization with 1000 steepest descent steps and 1000 conjugate gradient steps; system minimization with the same strategy; the heating of the solvent was conducted at constant volume from 0 to 300 K for 20 ps; followed by equilibration of the full system in the NPT ensemble at 300 K and 1 atm for 500 ps with a time step of 2 fs. Constant temperature was regulated with a Langevin thermostat, and constant pressure was regulated using the Berendsen weak coupling algorithm. Subsequently, production runs of the systems were collected in the NVT ensemble at 300 K for 20 ns with a step size of 2 fs. The trajectories of the production runs were used to provide a series of starting structures for the steered molecular dynamics (SMD) simulations. The SMD simulation applies forces to the selected atoms to accelerate the movement or

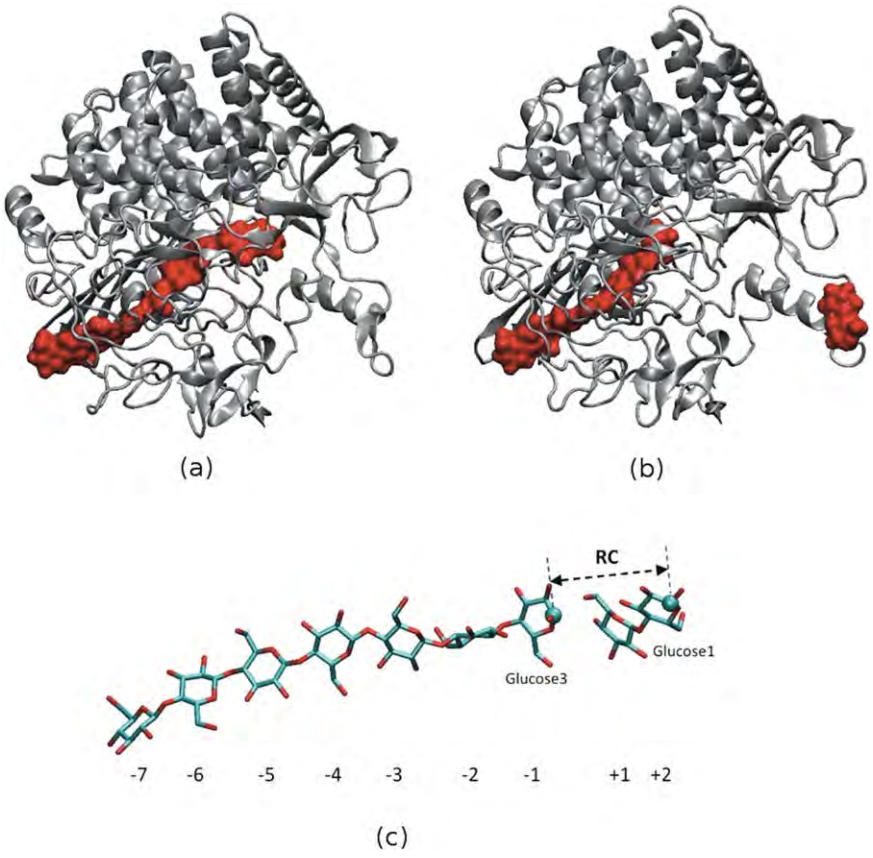

Figure 2

Representative structures of the initial (**a**) and final (**b**) states of the SMD simulations (**c**). The reaction coordinate (RC) was defined to be the distance between the two C1 atoms.

conformational changes of the system at a slow but non-zero rate, therefore permitting the study of hypothetical processes within a reasonable time frame. In this study, 40 streams of SMD simulations were conducted for each system, using 40 different starting structures that were extracted from the production run. The criteria for selecting starting structures was that the distance between the C1 atom of glucose 1 and the C1 atom of glucose 3 be the closest to the most populated distance over the production run. The SMD simulations were performed using the PMEMD engine of AMBER. The reaction coordinate was set to be the distance between the two C1 atoms (Fig. 2). Over the course of each SMD simulation, an external force was exerted on the two C1 atoms coaxial with the line between their centers, to gradually increase the reaction coordinate at a constant slow speed of 1 Å/ns as described by Bu *et al.*,[30] until a total operation distance of 22 Å was reached, simulating the process of product escape into the aqueous environment after the hydrolysis. For constant velocity SMD, instantaneous external force was determined by the change in velocity that is needed to bring the previous instant velocity to the target velocity.[31] During the process of product escape, there was no control over the orientation of the celloheptaose and the cellobiose. The celloheptaose maintained its position in the well-defined tunnel, and thus had no significant changes in neither the Φ, Ψ-torsional angles of the glycosidic linkages nor the Euler angles. For the cellobiose, analysis of the Φ, Ψ-torsional angles over the course of product escape showed that they were within the same range as that of the relaxed cellobiose in aqueous solution. The Euler angles of the cellobiose varied during this process, since cellobiose is not a rigid body, and such variations were considered averaged out through the multiple SMD simulations and contributed to the configurational sampling of the cellobiose molecule, energy, and entropy, during the exit process. The accumulated external work was recorded over the simulations for free energy calculation.

Free energy calculation using a "fast growth" method

The product expulsion energy, in this case the binding free energy between cellobiose and the glycoside hydrolases, which contains a celloheptaomer bound in the −7

to −1 subsites of the active site tunnel, contributes to the product inhibition. Calculation of protein–ligand binding free energy is intrinsically a very interesting topic. Many methods, for example free energy perturbation, have been successfully implemented in calculating the protein–ligand binding free energy, though such methods for large molecular systems often involve complications at the setup stage.[30]

In this study, we used a nonequilibrium "fast growth" method, also known as Jarzynski's equality, to calculate the product expulsion energy. The product expulsion energy was defined to be the free energy difference between the initial state immediately after the cellobiose was cleaved off the cellulose chain and the final state when the cellobiose was released into the aqueous environment. The "fast growth" method refers to sampling with many irreversible processes, different from the established "slow growth" method, during which the system is driven reversibly from one state to the other.[32]

Jarzynski's equality states that the free energy difference between two states A and B at equilibrium can be estimated using the cumulative work from state A to state B under nonequilibrium conditions.[33,34] Jarzynski's equality and its computational implementation are expressed as:

$$\Delta G_{A\to B} \approx W^{x,N}$$

$$\Delta G_{A\to B} = -\frac{1}{\beta}\ln\langle \exp(-\beta W_{A\to B})\rangle_A$$

$$= -\frac{1}{\beta}\ln\sum_{i=1}^{N}\frac{1}{N}\exp(-\beta W_{i,A\to B})$$

where $W^{x,N}$ is the exponential average of work for N realizations, and W_i is the nonequilibrium accumulative work (for the ith realization) done on the system when going from state A to state B in the forced simulation process. This method generates both statistical and systematic errors.[32] The statistical uncertainty, in terms of standard error, can be calculated using a bootstrap method,[35,36] which is sensitive to the possibility that $W^{x,N}$ might be dominated by one or a few particularly small values of work among all the realizations. In addition, the "fast growth" estimate contains a systematic bias: for finite N, the exponential average of work tends to overestimate $\Delta G_{A\to B}$ by[37]

$$W^{x,N} - \Delta G_{A\to B} \approx \frac{\beta\sigma_w^2}{2N}$$

Normally the statistical errors dominate rather than the systematic errors, and it is suggested that $\sigma_W \geq k_B T$.[32] 100 cycles of bootstrapping analysis were performed to calculate the statistical error every 200 ps.

The "fast growth" method has been shown to converge to the free energy difference of the two states on small molecular systems both computationally[38] and experimentally.[37] It has been used recently to calculate the product expulsion energy of cellobiose in the exocellulase Cel7A from *T. reesei*, and the result (−14.4 kcal/mol) was qualitatively comparable with that calculated by the free energy perturbation method (−11.2 kcal/mol).[30] The product expulsion energies calculated using this method were also qualitatively consistent with the experimental results that product inhibition mostly affects exocellulases (such as Cel7A and Cel6A from *T. reesei*) rather than endocellulases (such as Cel7B from *T. reesei* and Cel6B from *Humicola insolens*).[39]

Table II
Cellobiose Product Expulsion Free Energies of the Four Family 48 Glycoside Hydrolases and their Mutants

Wildtype and mutants	Mean of the product expulsion energy (kcal/mol)	Wildtype and mutants	Mean of the product expulsion energy (kcal/mol)
CelA	7.88 ± 0.21	CelF	11.39 ± 0.29
CelA_R548A	**6.85 ± 0.56**	**CelF_R544A**	8.31 ± 0.76
CelA_D489A	**7.35 ± 0.27**	**CelF_R549A**	7.13 ± 1.07
CelA_D493A	**1.83 ± 0.56**	**CelF_D494A**	8.06 ± 0.79
CelA_D550A	**3.54 ± 0.14**	**CelF_D546A**	9.34 ± 0.46
CelA_E44A	**4.90 ± 0.17**	**CelF_E44A**	4.31 ± 0.56
CelA_E546A	**1.16 ± 0.17**	CelF_E542A	15.99 ± 0.13
CelA_K547A	13.95 ± 1.21		
Cel48	12.60 ± 0.21	CelS	14.82 ± 0.68
Cel48_R592A	**10.90 ± 0.67**	**CelS_R643A**	10.63 ± 0.20
Cel48_R682A	14.60 ± 0.21	**CelS_D490A**	11.51 ± 0.69
Cel48_D530A	14.33 ± 0.28	**CelS_D520A**	8.97 ± 0.19
Cel48_D534A	20.33 ± 0.21	**CelS_E76A**	4.71 ± 0.54
Cel48_D594A	17.12 ± 0.15	CelS_E576A	19.62 ± 0.54
Cel48_E38A	19.19 ± 0.48		
Cel48_E590A	20.14 ± 0.34		
Cel48_E591A	15.12 ± 0.40		
Cel48_E593A	13.21 ± 0.30		

Note: The mutants in bold required reduced level of product expulsion energy compared to the wildtype glycoside hydrolases.

The product inhibition in family 48 cellulases might be related to the composition of amino acid residues at the tunnel exit, particularly the ones that are charged and the ones that form hydrogen bonds with the cellobiose. We designed rational single mutants with the mutation sites along the tunnel exit, aiming at reducing the product expulsion energies so as to reduce the inhibitory level. To identify the mutation sites, we screened the 40 streams of the SMD simulations of each wildtype glycoside hydrolase to detect the tunnel exit residues that have long-term strong interactions with the cellobiose (Table II). In particular, the selected residues needed to be within 4.5 Å of the cellobiose for longer than 2 ns out of the 22 ns simulation. In addition, the interaction energy between each of these residues and the cellobiose was assessed over the course of cellobiose escape for the 40 streams of SMD simulations, in terms of van der Waals interaction energy and electrostatic interaction energy. The mutation sites were determined to be the residues that presented strong VDW interaction (maximum magnitude > 5 kcal/mol) or strong electrostatic

Table III
Product Expulsion Energies for Family 48 Glycoside Hydrolases and Their Rational Mutants

Group ID	CelA		Cel48		CelF		CelS	
1*	**Glu44**	↘↘	**Glu38**	↗↗	**Glu44**	↘↘	**Glu76**	↘↘
2#	Trp412		Trp411		Trp411		Trp439	
3	Thr463		Gly500		Thr462		**Asp490**	↘
4#	**Asp489**	↘≈	**Asp530**	↗	Asn490		Ser516	
5#	**Asp493**	↘↘	**Asp534**	↗↗	**Asp494**	↘↘	**Asp520**	↘↘
6	**Lys547**	↗↗	**Glu591**	↗	Gln543		Ala577	
7*	**Glu546**	↘↘	**Glu590**	↗↗	**Glu542**	↗↗	**Glu576**	↗↗
8*	**Arg548**	↘	**Arg592**	↘	**Arg544**	↘	Arg578	
9	Ala549		**Glu593**	↗≈	Gly545		Ala579	
10*	**Asp550**	↘↘	**Asp594**	↗↗	**Asp546**	↘	Asp580	
11*	Arg613		**Arg682**	↗	Arg609		**Arg643**	↘

The residues in bold are the ones that possess high interaction energy with the cellobiose product over the course of its escape. ↗ (or ↘) refers to the increase (or decrease) in ΔG compared to the wildtype with the magnitude from 1 to 3 kcal/mol; ≈ means the changes in ΔG is moderate with the magnitude <1 kcal/mol; ↗↗(or ↘↘) refers to much larger level of ΔG changes with the magnitude > 3 kcal/mol. * refers to the residues that form a flat surface on the lower side of the inner tunnel exit. # refers to the residues on the upper side of the tunnel exit (see Fig. 5).

energy (maximum magnitude > 17 kcal/mol) to the cellobiose, and such strong interaction occurred in at least 20 out of the 40 SMD simulations. Interestingly, many of the identified mutation sites were conserved or partially conserved among the four glycoside hydrolases (Table III).

RESULTS AND DISCUSSION

Wildtype family 48 glycoside hydrolases

Whereas the structures and the amino acid residues of the four wildtype family 48 glycoside hydrolases are similar near the product site except for Cel48, the cellobiose product expulsion free energies vary widely. The calculated product expulsion free energies indicate that the product inhibitory level is the highest in CelS, followed by Cel48A and CelF, and the lowest in CelA (Fig. 3). Although the product expulsion energy cannot alone explain the relative activities for these cellulases, it probably makes a large contribution to product inhibition. Compared with the CelA, CelS, and CelF structures, Cel48 exhibits an extra loop structure at the tunnel exit (Fig. 1). This loop was initially speculated to be a cause for the higher product inhibitory level in Cel48, compared to CelA. However, the Cel48 potential of mean force (PMF) profile showed that the free energy change reached a plateau for reaction coordinate values of ~10 Å and beyond, before the cellobiose started to have contact with the additional loop. Therefore, the loop in Cel48 does not seem to affect the level of product inhibition.

The magnitudes of the free energies reported here are in agreement with those reported in earlier studies for similar glycoside hydrolases using steered molecular dynamics and free energy perturbation theory.[30,39] In these SMD simulations, there was no control over the orientation of the celloheptaose and the cellobiose. The celloheptaose maintained its position in the well-defined tunnel, and thus had no significant changes in either the Φ,Ψ-torsional angles of the glycosidic linkages nor the Euler angles. For the cellobiose, analysis of the Φ,Ψ-torsional angles over the course of product escape showed that they were within the same range as that of the relaxed cellobiose in aqueous solution[40] (data not shown). The Euler angles of the cellobiose varied during this process, since cellobiose is not a rigid body, and such variations were considered to be averaged out through the multiple SMD simulations.

It is noteworthy that this study might only represent part of the mechanism for product inhibition. For example, the course of product escape was designed to be such that when moving the cellobiose product out of the tunnel exit, the substrate maintained its position in the substrate side of the tunnel. However, the movement of

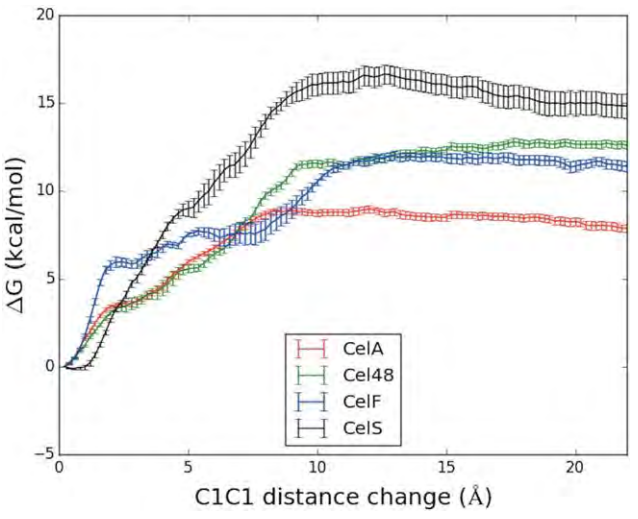

Figure 3
Cellobiose product expulsion free energies of the four wildtype family 48 glycoside hydrolases. Standard error of all data points was below 1 kcal/mol.

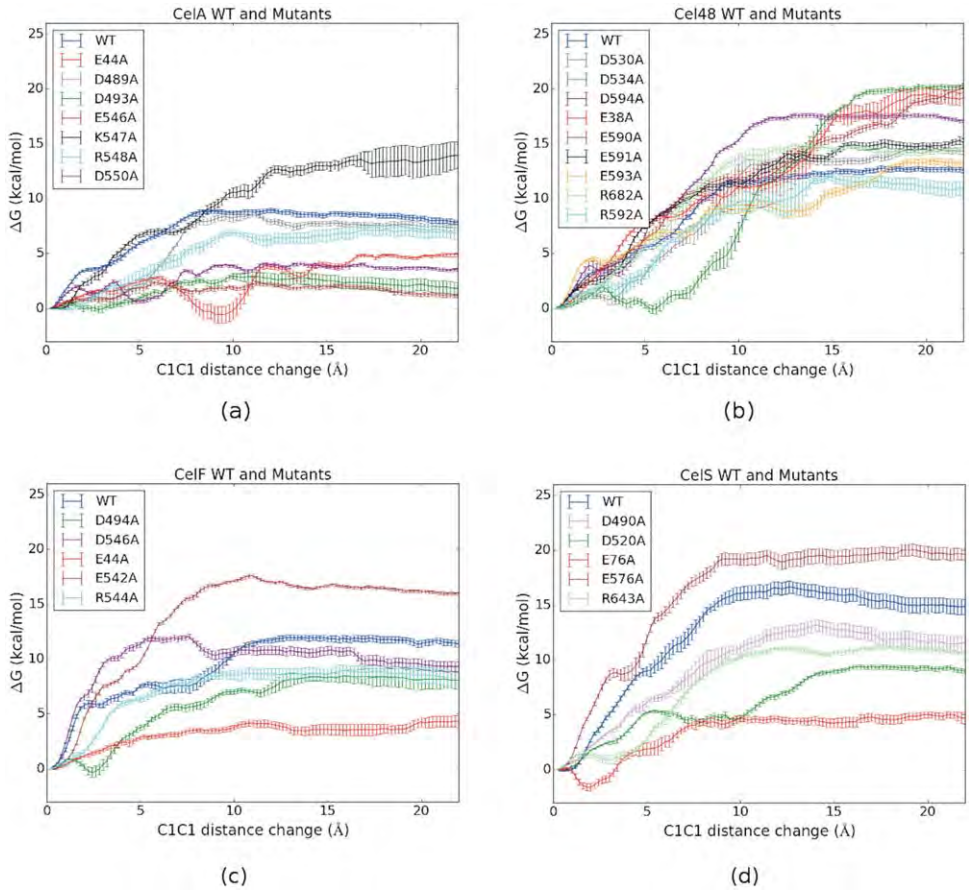

Figure 4

Calculated cellobiose product expulsion energy in the wildtype (WT) and rational mutants of CelA (**a**), Cel48 (**b**), CelF (**c**), and CelS (**d**). Standard error of all data points was below 1 kcal/mol. The homologous mutants of the four glycoside hydrolases (Table I) are plotted in the same color.

the substrate toward the product side and the product escape might happen simultaneously, in which case, the substrate gradually binds to the subsites +1 and +2, reducing the free energy difference between these initial and final states. Additionally, the crystal structures of the glycoside hydrolases might not represent all the conformations over the course of product escape, and previous efforts in guiding the substrate (celloheptamer in this system) to proceed over the active site in simulations were rather unsuccessful due to the narrow path at the active site. Overall, as the goal was to find mutation sites along the tunnel exit of the glycoside hydrolases, and particularly the sites beyond the subsites +1 to reduce binding to the cellobiose product over its escape, it was decided not to involve the process of substrate advance in the simulations.

Rational mutants of family 48 glycoside hydrolases

As described above, the composition of amino acid residues at the tunnel exit affects the product expulsion energy. We rationally designed single mutants at sites along the tunnel exit to reduce the product expulsion free energies. The product expulsion energies of each of the wildtype glycoside hydrolases and their rational mutants are shown in Table II and Figure 4, and are further presented in a more intuitive representation in Table III, where the homologous amino acid residues among the four family 48 glycoside hydrolases are listed in the group. The mutants with reduced product expulsion energies indicated possibly good candidates with reduced levels of product inhibition. Promising mutations were found in all family 48 glycoside hydrolases except for Cel48. The lack of success in reducing the binding free energy in Cel48 seems to come from the extra turn close to the tunnel exit. This turn composed of amino acids SDEEGYF starting at position 466 appears to increase the amino acid packing rendering the tunnel exit less deformable even in the mutants.

Mutating the tunnel exit residues of the glycoside hydrolases can change the topological features of the region and thus affect its affinity to the cellobiose product over the course of product escape. In particular, at

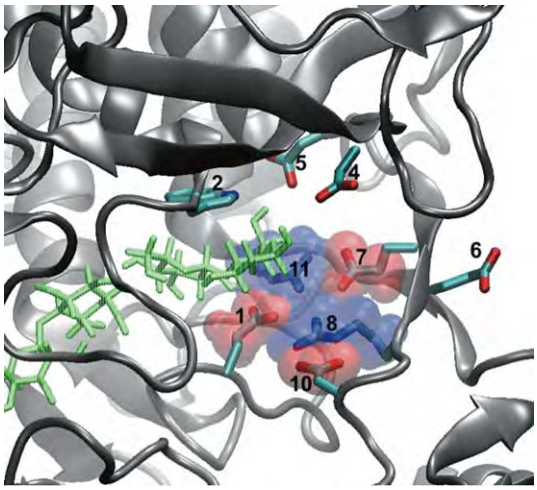

Figure 5

The residues that affect the product escape at the tunnel exit of family 48 glycoside hydrolases. The residues are labeled by their Group IDs. The residue Groups 1, 7, 8, 10, and 11 that form a flat surface at the inner part of the tunnel exit are presented in licorice and transparent vdW spheres. The red vdW spheres refer to acidic amino acid residues, and the blue ones refer to basic residues. The representation corresponds to the crystal structure of Cel48. The Groups 3 and 9 are not shown since Group 3 is further away from the active site and Group 9 is neither conserved nor has a large effect on product escape.

the initial stage of product escape, the cellobiose is stuck in a narrow pocket (Fig. 5). On the bottom side of the tunnel exit, the conserved residues including two Args, two Glus, and one Asp in the Groups 1, 7, 8, 10, and 11, forming a flat vdW surface via multiple salt bridges (Table III). These charged residues can form multiple hydrogen bonds with the cellobiose. On the other side, the conserved residues Trp and Asp in the Groups 2 and 5 form stacking interactions and hydrogen bonds with the cellobiose, respectively. It is likely that the planar bottom surface and the Trp residue on the top side function together in stabilizing the substrate prior to its hydrolysis. Substituting the Group 8 Arg into Ala reduces the product expulsion energy, and this is probably due to Arg associating with three acidic amino acids in the Groups 1, 7, and 10, and plays a key role in forming the flat surface. Therefore, the removal of this Arg would truncate the flat surface region and ease the product release. The interaction energy between each of the mutation sites and the cellobiose over the course of product escape is dominated by electrostatic interactions (Fig. 6 and Supporting Information Fig. S1), rather than vdW interactions that are negligible (data not shown). Analysis of the intermolecular electrostatic interactions indicates that the acidic residues induce attractive interactions whereas the basic residues of the Groups 8 and

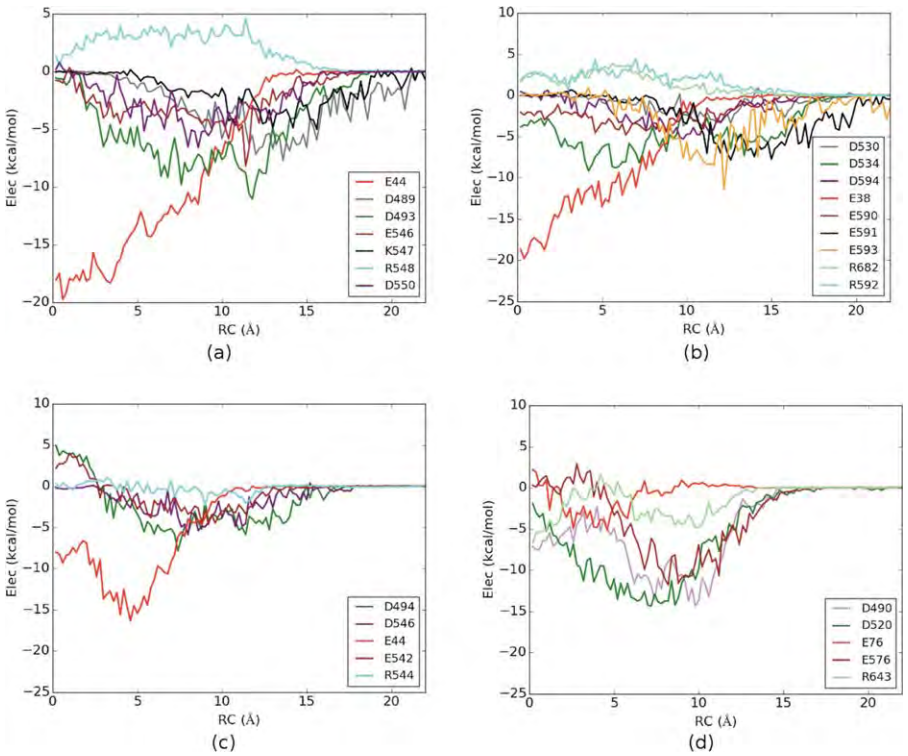

Figure 6

The electrostatic interaction between each mutation site and the cellobiose product in CelA (a), Cel48 (b), CelF (c), and CelS (d). The homologous mutants of the four glycoside hydrolases (Table I) are plotted in the same color.

11 tend to induce repulsive interactions to the cellobiose (Fig. 6). The homologous mutation sites, however, did not always follow the same trend in the changes of the product expulsion energies compared to the wildtype enzymes. For example, mutating the Group 1 Glu or the Group 5 Asp reduces the product expulsion energies in CelA, CelF, and CelS, but increases it in Cel48. Also, mutating the Group 7 Glu causes the increase in the product expulsion energies in Cel48, CelF, and CelS, but not CelA.

At the later stage of product escape, the partially conserved Group 4 Glu and the charged Group 6 residues seem to have impact on the product escape. Interestingly, mutating the Group 6 Lys in CelA severely increases the product expulsion energy, and mutating the Group 6 Glu in Cel48 also increases the expulsion energy. This combined with the result that each of these residues form attractive interactions with the cellobiose product (Fig. 6), indicates that the charged residues might assist in passing along the product more easily than other amino acid species. This suggests that including a charged residue at this location, particularly Lys, might facilitate the product escape. It appears that the optimal mutant could be obtained by some combination of these single point mutations to help the fine tuning of interactions at the product site to further lower the binding free energy of the product.

CONCLUSIONS

End-product inhibition of family 48 processive exocellulases is one possible explanation for their low catalytic activities. The crystal structures of the glycoside hydrolases revealed that a cellobiose unit stays in a pocket at the tunnel exit of the glycoside hydrolases. Also, a previous study on the glucose distribution around the surface of a family 48 glycoside hydrolase, CelF, suggested a higher binding affinity between the cellobiose product and the tunnel exit compared to the protein surface in general. We hypothesize that the end-product binds to the tunnel exit of the glycoside hydrolase, inhibits its functioning. In this study, we evaluated the binding affinity of cellobiose at the tunnel exit of the glycoside hydrolases by use of free energy calculations of the product expulsion. We also designed and evaluated rational single mutants with the mutation sites along the tunnel exit, with the aim of reducing the product expulsion energies. We show that certain single mutants at the conserved or partially conserved residue sites seem to have lower binding affinity to cellobiose compared to the wildtype enzymes. In particular, mutating the residues in the product-binding site groups 1, 5, and 8 into Ala in most of the family 48 glycoside hydrolases may be effective in reducing the product inhibitory effect. Mutating the residues in Group 6 into Lys or other charged amino acids might also ease the product escape. Thus, we have identified theoretically plausible mutants of family 48 glycoside hydrolases. Further experimental studies are needed to verify the effectiveness of these mutants in improving the turnover numbers of the cellulases.

ACKNOWLEDGMENTS

We thank Gregg T. Beckham from the National Renewable Energy Laboratory for useful discussions at the inception of this project. We also thank the BioEnergy Science Center for funding. The BioEnergy Science Center is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

REFERENCES

1. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 2007;315:804–807.
2. DOE., U. S. Breaking the biological barriers to cellulosic ethanol: a joint research agenda. US Dep. Energy, 2006. Report from the December 2005 Workshop, DOE/SC-0095. U.S. Department of Energy Office of Science (2006).
3. Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem J 1996;316:695–696.
4. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MWW, Lunin VV, Himmel ME, Bomble YJ. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 2013;342: 1513–1516.
5. Guimarães BG, Souchon H, Lytle BL, David Wu JH, Alzari PM. The crystal structure and catalytic mechanism of cellobiohydrolase celS, the major enzymatic component of the *Clostridium thermocellum* cellulosome. J Mol Biol 2002;320:587–596.
6. Parsiegla G, Juy M, Reverbel-Leroy C, Tardif C, Belaich JP, Driguez H, Haser R. The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 Å resolution. Embo J 1998;17:5551–5562.
7. Parsiegla G, Reverbel-Leroy C, Tardif C, Belaich JP, Driguez H, Haser R. Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. Biochemistry 2000;39:11238–11246.
8. Parsiegla G, Reverbel C, Tardif C, Driguez H, Haser R. Structures of mutants of cellulase Cel48F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action. J Mol Biol 2008; 375:499–510.
9. Brunecky R, Hobdey SE, Taylor II LE, Tao L, Tucker MP, Himmel ME, Decker SR. High temperature pre-digestion of corn stover biomass for improved product yields. Biotechnol Biofuels 2014;7:170.
10. Reverbel-Leroy C, Pages S, Belaich A, Belaich JP, Tardif C. The processive endocellulase CelF, a major component of the Clostridium cellulolyticum cellulosome: purification and characterization of the recombinant form. J Bacteriol 1997;179:46–52.
11. Bayer Ea, Belaich JP, Shoham Y, Lamed R. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 2004;58:521–554.
12. Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR. Deletion of the Cel48S cellulase from *Clostridium thermocellum*. Proc Natl Acad Sci USA 2010;107:17727–17732.

13. Barr BK, Hsieh YL, Ganem B, Wilson DB. Identification of two functionally different classes of exocellulases. Biochemistry 1996;35: 586–592.
14. Sinnott ML. Catalytic mechanism of enzymic glycosyl transfer. Chem Rev 1990;90:1171–1202.
15. Irwin DC, Zhang S, Wilson DB. Cloning, expression and characterization of a Family 48 exocellulase, Cel48A, from *Thermobifida fusca*. Eur J Biochem 2000;267:4988–4997.
16. Kruus K, Andreacchi A, Wang WK, Wu JHD. Product inhibition of the recombinant CelS, an exoglucanase component of the Clostridium thermocellum cellulosome. Appl Microbiol Biotechnol 1995;44: 399–404.
17. Lamed R, Kenig R, Setter E, Bayer EA. Major characteristics of the cellulolytic system of Clostridium thermocellum coincide with those of the purified cellulosome. Enzyme Microb Technol 1985;7:37–41.
18. Morag E, Halevy I, Bayer EA, Lamed R. Isolation and properties of a major cellobiohydrolase from the cellulosome of Clostridium thermocellum. J Bacteriol 1991;173:4155–4162.
19. Zhang XZ, Zhang Z, Zhu Z, Sathitsuksanoh N, Yang Y, Zhang YHP. The noncellulosomal family 48 cellobiohydrolase from Clostridium phytofermentans ISDg: heterologous expression, characterization, and processivity. Appl Microbiol Biotechnol 2010;86:525–533.
20. Baker JO, McCarley JR, Lovett R, Yu CH, Adney WS, Rignall TR, Vinzant TB, Decker SR, Sakon J,Himmel ME. Catalytically enhanced endocellulase Cel5A from Acidothermus cellulolyticus. Appl Biochem Biotechnol 2005;121124:129–148.
21. Atreya, ME, Strobel, KL, Clark, DS. Alleviating product inhibition in cellulase enzyme Cel7A. Biotechnol Bioeng 2015. doi:10.1002/bit.25809.
22. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102: 3586–3616.
23. MacKerell AD, Feig M, Brooks CL. 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 2004; 25:1400–1415.
24. Guvench O, Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, MacKerell AD. Additive empirical force field for hexopyranose monosaccharides. J Comput Chem 2008;29:2543–2564.
25. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
26. Brooks BR, Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek 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. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.
27. Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
28. Crowley MF, Williamson MJ, Walker RC. CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. Int. J. Quantum Chem. 2009;109:3767–3772.
29. Case DA, Berryman JT, Betz RM, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Giese TJ, Gohlke H, Goetz AW, Homeyer N, Izadi S, Janowski P, Kaus J, Kovalenko A, Lee TS, LeGrand S, Li P, Luchko T, Luo R, Madej B, Merz KM, Monard G, Needham P, Nguyen H, Nguyen HT, Omelyan I, Onufriev A, Roe DR, Roitberg A, Salomon-Ferrer R, Simmerling C, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wu X, York DM, Kollman PA. AMBER 12. University of California San Francisco, Vol. 1, 2012.
30. Bu L, Beckham GT, Shirts MR, Nimlos MR, Adney WS, Himmel ME, Crowley MF. Probing carbohydrate product expulsion from a processive cellulase with multiple absolute binding free energy methods. J Biol Chem 2011;286:18161–18169.
31. Ho BK, Agard DA. An improved strategy for generating forces in steered molecular dynamics: the mechanical unfolding of titin, e2lip3 and ubiquitin. PLoS One 2010;5:e13068.
32. Hendrix DA, Jarzynski C. A 'fast growth' method of computing free energy differences. J Chem Phys 2001;114:5974–5981.
33. Jarzynski C. Nonequilibrium equality for free energy differences. Phys Rev Lett 1997;78:2690.
34. Jarzynski C. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. Phys Rev E 1997;56:5018.
35. Efron B, Efron B. The jackknife, the bootstrap and other resampling plans. SIAM 1982;38. Available at: http://dx.doi.org/10.1137/1.9781611970319.
36. Efron B, Tibshirani RJ. An introduction to the bootstrap. Boca Raton: CRC press, 1994.
37. Liphardt J, Dumont S, Smith SB, Tinoco I, Bustamante C. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. Science 2002;296:1832–1835.
38. Xiong H, Crespo A, Marti M, Estrin D, Roitberg AE. Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. Theor Chem Acc 2006;116:338–346.
39. Bu L, Nimlos MR, Shirts MR, StÍhlberg J, Himmel ME, Crowley MF, Beckham GT. Product binding varies dramatically between processive and nonprocessive cellulase enzymes. J Biol Chem 2012; 287:24807–24813.
40. Frankel AE, Burbage C, Fu T, Tagge E, Chandler J, Willingham MC. Ricin toxin contains at least three galactose-binding sites located in B chain subdomains 1α, 1β, and 2γ. Biochemistry 1996;35:14749–14756.
41. E. Michael Gertz. BLAST scoring parameters. 2005. Available at: ftp://ftp.ncbi.nlm.nih.gov/blast/documents/developer/scoring.pdf.