Advantages of a Distant Cellulase Catalytic Base

Tucker Burgin¹, Jerry Ståhlberg², and Heather B. Mayes¹

From the ¹Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109, USA, and ²Department of Molecular Sciences, Swedish University of Agricultural Sciences, SE-75007, Uppsala, Sweden

Running title: Advantages of a Distant Cellulase Catalytic Base

To whom correspondence should be addressed: Heather B. Mayes, hbmayes@umich.edu

Keywords: bioenergy, bioengineering, biotechnology, cellobiohydrolase, cellulase, enzyme mechanism, glycoside hydrolase, molecular dynamics, protein engineering, Trichoderma reesei

ABSTRACT

The inverting glycoside hydrolase Trichoderma reesei (Hypocrea jecorina) Cel6A is a promising candidate for protein engineering for more economical production of biofuels. Until recently, its catalytic mechanism had been uncertain: the best candidate residue to serve as a catalytic base, D175, is further from the glycosidic cleavage site than in other glycoside hydrolase enzymes. Recent unbiased transition path sampling simulations revealed the hydrolytic mechanism for this more distant base, employing a water wire; however, it is not clear why the enzyme employs a more distant catalytic base, a highly-conserved feature among homologs across different kingdoms. In this work, we describe molecular dynamics simulations designed to uncover how a base with a longer side chain, as in a D175E mutant, affects procession and active site alignment in the Michaelis complex. We show that the hydrogen bond network is tuned to the shorter aspartate side chain, and that a longer glutamate side chain inhibits procession as well as being less likely to adopt a catalytically productive conformation. Furthermore, we draw comparisons between the active site in TrCel6A and another inverting, processive cellulase to deduce the contribution of the wire water to the overall enzyme function, revealing that the more distant catalytic base enhances product release. Our results can inform efforts in the study and design of enzymes by demonstrating how counterintuitive sacrifices in chemical reactivity can have worthwhile benefits for other steps in the catalytic cycle.

In order to tap into the deep reservoir of renewable energy represented by fuels derived from plant matter, an economical means of converting lignocellulose is required (1). Decomposition of the primary component, cellulose, is catalyzed by glycoside hydrolase (GH) enzymes, which are found ubiquitously in nature (2); therefore, improved catalytic efficiency of cellulose decomposition enzymes would help biomass to compete with non-renewable carbon sources. This motivates molecular-level studies into GH enzymatic mechanisms, as such understanding has previously proven invaluable in efforts to engineer variants with increased activities (3–5).

A particularly important GH enzyme is Trichoderma reesei Cel6A (TrCel6A), which plays a key synergistic role in industrial enzyme cocktails for cellulose digestion. This enzyme is a cellobiohydrolase of GH family 6, which cleave β-1,4 glycosidic bonds processively along cellulose chains, from the non-reducing towards the reducing end, to release the glucose dimer cellobiose as the main product (6). This processive mode of action is believed to be key to their efficiency on highly crystalline cellulose. Glycoside hydrolase family 6 (GH6) enzymes exhibit a range of activity on a continuum between cellobiohydrolase processive activity and endoglucanase activity, characterized by cleavage of internal bonds (7–10). Endoglucanases
Advantages of a Distant Cellulase Catalytic Base

Generally exhibit higher catalytic rate constants, only show appreciable activity on less ordered regions of cellulose, and produce a broader range of products (5). A crystal structure of TrCel6A was first solved in 1990 (11), but only recently has its molecular-level mechanism begun to be established. GH6 enzymes function via an inverting mechanism wherein the stereochemistry at the carbon of the cleaved $\beta$-1,4 bond is changed from equatorial to axial. The classical inverting mechanism (Figure 1, left) by which such reactions typically take place requires a catalytic acid-base pair (a proton donor and acceptor), the latter of which activates a nucleophilic water molecule during hydrolysis (12). The identity of the catalytic acid in TrCel6A was identified based on the first crystal structure as D221 without controversy (5, 11), while the identity of the catalytic base was more ambiguous. Initial hypotheses of the catalytic base identity included D263 or D401 (5, 11). Mutational studies of D263 soon excluded this residue as the catalytic base (13); later D401 was also excluded on the basis of activity studies of site mutants of TrCel6A and homologous enzymes (5, 9, 14, 15). Koivula et al. proposed that the two water molecules that consistently appear in the active site of crystal structures would act as a water wire to shuttle a proton to TrCel6A D175, thus allowing this residue to serve as the catalytic base (15). Recent computational studies provide convincing evidence of this hypothesis (16) and other studies have provided evidence of Grotthuss mechanisms in other GH enzymes (17).

An interesting aspect of the mechanism revealed is that the catalytic step wherein D175 accepts the excess proton requires an additional water molecule compared to the canonical mechanism, positioned between the basic carboxylate group at the end of D175 and the nucleophilic (or “attacking”) water, as shown in Figure 1 at right. This “wire” or “bridge” water molecule, observed in crystal structures of the enzyme (5, 18), momentarily forms a hydronium ($H_3O^+$) ion during hydrolysis before offloading its excess proton to D175 (16). The mechanism therefore requires the stabilization of two water molecules in the active site as opposed to only the one participating in the overall chemistry, raising the question as to why the active site of TrCel6A includes the additional water. Previous studies indicate that there is an additional energetic barrier on the order of 5–6 kcal/mol associated with each additional water wire in a Grotthuss mechanism (19). If the additional water molecule is only needed to conduct a proton over the required distance, a mutant base with a longer side chain could obviate the need for the water wire. This suggests investigation of a TrCel6A D175E mutant, since glutamate (E) is identical in structure to aspartate (D) but for an additional CH$_2$ group in its side chain that could project its carboxylate group further into the active site. While no studies appear in the literature on a TrCel6A D175E mutant, mutants of the homologous aspartates have been constructed for the GH6 enzymes Thermobifida fusca Cel6A (TfCel6A, at that time known as Thermomonospora fusca endocellulase E2) (20) and Cellulomonas fimii Cel6A ( CfCel6A, known as CenA at the time) (21). Note that these studies were published in 1999 and 1995, respectively, before Koivula et al.’s seminal 2002 paper informing our current understanding of which residues perform the function of the catalytic base in this family of enzymes. Wolfgang and Wilson’s 1999 work described the TfCel6A catalytic base as D265, but the homologous residue to TrCel6A D175 is D79 (5, 20). Compared to the wild-type enzyme, the D79E mutant showed 0.4%, 0.6%, and 38% activity on carboxymethyl cellulose, phosphoric acid-swollen cellulose, and Whatman number 1 filter paper, respectively. Similarly, Damude et al.’s 1995 paper stated that the CfCel6A catalytic base was D392, which was later found inconsistent with new data. Instead, their CfCel6A D216E mutant is homologous to the TrCel6A D175E mutant. The CfCel6A D216E activity was tested on carboxymethyl cellulose and phosphoric acid-swollen cellulose, showing just under 0.1% of wild-type activity on both substrates. Due to the close sequence and structural similarity between all these Cel6A enzymes, it is highly likely that at TrCel6A D175E mutant would show decreased activity, likely in the range of two to three orders of magnitude lower, rather than an increased rate due to obviating the need for the water wire.

In order to bridge the gap between atomistic enzymatic detail and human chemical intuition, we constructed molecular models of both wild-type and D175E TrCel6A and investigated the influence of the
Advantages of a Distant Cellulase Catalytic Base

mutation on two non-reactive steps in the catalytic cycle: (1) procession of the substrate into the active site; and (2) the transition to the reaction-competent active site conformation following procession. We found in both cases that the longer E175 residue in the mutant was a hindrance to the catalytic cycle, highlighting the importance and intricacy of the remarkable network of hydrogen bonding interactions that stabilizes the active site of the wild-type enzyme. Finally, by means of comparison of the active site to that of a related processive cellulase that functions via the classical inverting mechanism (Tf/Cel9A), we offer an explanation for the TrCel6A mechanism in terms of reduced association between the product and enzyme. Based on the results of these studies, we propose a benefit in cellulas to activation of the nucleophilic water via a wire water by taking into account aspects of the enzymatic cycle outside the reaction itself, broadening the context for rationalizing enzymatic features in carbohydrate-active enzymes.

RESULTS

D175E Mutant: Procession of the substrate into the active site — As a non-reducing-end cellobiohydrolase, each enzymatic cycle in TrCel6A requires the leading ring of the substrate chain advancing from the +1 site to the −2 site (5), described as moving from the “pre-slide” to “slide” positions as shown in Figure 2 (at top). One hypothesis to explain the lower activity of the TrCel6A D175E homologue is that the bulkier side chain could protrude into the active site groove and hinder procession. To study this, we performed umbrella sampling simulations of procession using a collective variable (CV) that tracks the relative positions of the enzyme and substrate (simulation details available in the supporting material). The resulting potential of mean force (PMF) plots for both enzyme types are shown in Figure 2 (bottom), with the zero point on the free energy axis set at around 2 Å, where we did not expect the identity of the residue at position 175 to have a large effect. We note that the CV used for sampling did not capture all key features that change during the transition from the pre-slide to slide positions. Specifically, as discussed in our previous study of TrCel6A wild-type procession, another key order parameter is the puckering of the second-to-leading glycosyl ring as it enters the −1 binding site, near 9 Å on the x-axis of the PMF in 2 (16). Additionally, we found that the serine loop moves from a more-open to less-open position during procession, appearing coincident to the procession at approximately 6 Å on the x-axis of the PMF. The multidimensional CV required to properly sample all these (and potentially more) key feature changing during procession would be computationally prohibitive, and thus the PMF shown for the simplified CV (RMSD only) is most appropriately analyzed in terms of comparative qualitative, not quantitative, differences between the WT and D175E mutant (further discussion of this point is included in the supporting text). Specifically, we note that the PMF for the wild-type enzyme has a pronounced energy well that stabilizes the productive structure just after 10 Å, providing a driving force for spontaneous procession. In contrast, the D175E mutant retains a fairly flat energy profile. This effect can be at least in part attributed to steric clashes with several residues near the catalytic base, as shown in Figure 3, suggesting that one advantage of the shorter catalytic base is to enhance procession by widening the gap between the wall of the active site tunnel and the substrate.

D175E Mutant: Reaction-competent active site conformation — During hydrolysis, the catalytic acid loses a proton while the base obtains an excess proton. Before the next reaction, these residues likely exchange a proton when the acid (D221) bends away from the reaction site towards the base (residue 175), as shown for the wild-type enzyme in Figure 4. After re-protonating, the acid must rotate along the dihedral angle indicated in the figure in order to position its proton toward the glycosidic oxygen and realign itself for the next catalytic event (5, 15).

The energy landscape in Figure 4 shows the barriers for the wild type and D175E mutant for this transition, with the acid and base residues further apart at the larger dihedral angles. The PMFs are similar in shape, and the difference in barrier height of no greater than 1 kcal/mol is not expected to significantly affect the overall kinetics of the enzyme, as this reconfiguration has a barrier less than half of that for hydrolysis of the glycosidic bond. However, as shown in Figure 5, the wild-type and mutant conformations corresponding to the right-hand side energy wells in Figure 4 have significantly different
Advantages of a Distant Cellulase Catalytic Base

hydrogen bonding networks. In the wild type, the hydrogen bond between the acid and base residues is broken, and the acid instead hydrogen-bonds with the glycosidic oxygen, in a reactive conformation for hydrolysis (16). In the mutant, the base often remains hydrogen-bonded to the acid, rather than with the active-site “bridge” water, preventing formation of the hydrogen-bonding network that aligns the active site waters for hydrolysis. When unbiased simulations were run with the acid initially at a dihedral angle of 175°, the productive hydrogen bond between the D221 proton and the glycosidic oxygen was never observed over the 1-ns trajectory in the D175E mutant, compared to roughly 5% of the frames in the wild type.

In a separate simulation, the acid-base hydrogen bond in the conformation shown in Figure 5B (between D221 and E175) remained stable during 972 ps of simulation, at no point bonding instead with the attacking water. Efforts to obtain such a conformation indicated that it does not occupy a local energy minimum. As shown in Figure 6, the energy barrier associated with this active conformation was quantified using umbrella sampling with restraints applied to the dihedral angle connecting the β- and γ-carbons. This parameterization approximates the motion that the residue undergoes during unbiased simulation initiated near the high-energy region around 40°. As shown, the free energy trough in this region is shallow and readily degenerates to the inactive state at −50°, and an even lower-energy state is available at −170° (leftmost energy well in Figure 6). A total free energy activation barrier of roughly 7.3 kcal/mol separates the active state from the low-energy state at −170°, posing significant hindrance to hydrolysis.

The simulations used to construct the PMF in Figure 6 began from a structure with D221 initially in a high-dihedral angle conformation from Figure 4, which in the wild type aligns the D221 carboxylic proton with the glycosidic oxygen. However, in the mutant these atoms were not consistently aligned, indicating that 7.3 kcal/mol is an underestimation of the barrier to reactive alignment. Additionally, even when the attacking water is aligned by E175, its lone pair is less-favorably oriented for nucleophilic attack as compared to the wild type. Considering these factors and assuming that the lack of a water wire reduces the hydrolysis barrier by between 4.8 and 5.8 kcal/mol (19), the net effect of the mutation would be to increase the barrier by a conservative (low) estimate of between 1.5 and 2.5 kcal/mol. We find that although E175 may adopt a conformation where it could accept a proton directly from the nucleophilic water (and thus may act as catalytic base in a classical inverting mechanism), activity in this mutant is lower than in the wild type because (a) the “active” conformation of E175 is disfavored, (b) the lone pair of the attacking water is less-favorably oriented for nucleophilic attack at the anomeric carbon, and (c) the catalytic acid D221 is less able to dissociate from E175 (as compared to D175) to enable protonation of the glycosidic bond.

The finding that a glutamate at position 175 disrupts the hydrogen bonding network that catalyzes hydrolysis indicates that active site of TrCel6A is tuned for the shorter, highly-conserved aspartate (5). Several other GH6 enzymes have been crystallized, including TfCel6B (PDB ID: 4B4F (22)), Chaetomium thermophilum Cel6A (PDB ID: 4A05 (23)), Humicola insolens Cel6A (PDB ID: 1BVW (24)), and TfCel6A (PDB ID: 2BOD (25)). The distances between the glycosidic oxygens and catalytic bases, as well as the presence of two active site waters in structures apparently primed for hydrolysis, indicate that members of this family generally perform hydrolysis via a Grotthuss mechanism to a more distant catalytic base. Our simulations clearly indicate why a longer catalytic base decreases activity; however, the question remains as to why the active site is not tuned for a longer side chain, or why the aspartate is not positioned closer to the cleavage site.

Active Site Homology — To better understand why the TrCel6A active site is tuned to require a wire water, we compared its active site to that of another inverting glycoside hydrolase, TfCel9A, a processive endocellulase that is believed to employ the classical mechanism based on crystallography and site-directed mutagenesis studies (26, 27). We created a TfCel9A model based on a product-state crystal structure (PDB ID: 4TF4 (28)) to compare with a product-state structure produced for TrCel6A in the course of our previous work (16), as shown in Figure 7. Panels A and B highlight differences in the hydrogen-bonding network in the product states.
In both cases, hydrogen bonding interactions that helped stabilize the nucleophilic water in the appropriate position for catalysis pre-reaction become hydrogen bonds to the product monomer in the $-1$ position post-reaction. Specifically, the product hydrogen bonds with S181 and the “bridge” water (in turn hydrogen-bound to D175) in \( \text{Tf} \text{Cel6A} \), and to D55 as well as directly to D58 in \( \text{Tf} \text{Cel9A} \). \( \text{Tf} \text{Cel6A} \) D175 and \( \text{Tf} \text{Cel9A} \) D58 are the base residues while \( \text{Tf} \text{Cel6A} \) S181 and \( \text{Tf} \text{Cel9A} \) D55 serve to stabilize the nucleophilic water in the reactant state.

Viewed from the reducing end of the substrate, an interesting geometric distinction between the two active sites becomes apparent as depicted in panels C and D. Although \( \text{Tf} \text{Cel9A} \) employs a glutamate residue instead of an aspartate as its catalytic acid, the acid/base pairs in these enzymes are in similar relative positions to one another (with 5.3 Å of separation between the carbon atoms in their carboxylate head groups in the relaxed product state), and the relative positions of the acid and base to the axis of the substrate in each enzyme are the same (making about a 90° angle in each case). However, the whole of the catalytic machinery of \( \text{Tf} \text{Cel6A} \) is rotated roughly 45° about the substrate. Because of the oblong cross-sectional shape of the substrate, the \( \text{Tf} \text{Cel6A} \) base is further from the $-1$ anemic carbon than in \( \text{Tf} \text{Cel9A} \) (6.0 versus 4.1 Å in the product state, respectively), requiring the addition of the wire water molecule to connect the base with the attacking water. While this discussion focuses on the product state, for which a crystal structure of \( \text{Tf} \text{Cel9A} \) is available, the same conclusions should hold in the transition state structures based on the positions of the acid residue $\alpha$-carbons.

One advantageous function of the water wire could be to destabilize the newly-formed product in the $-1$ position after hydrolysis. We quantified this effect by measuring the contribution to the overall free energy of binding from the base residue in each enzyme. To focus on the effect of the mediating water wire alone, not obscured by other differences in the enzymes, we calculated the total change in binding free energy associated with mutating the bases to alanine, as detailed in the supporting text. The base-to-alanine mutations remove all of the polar interactions with the catalytic bases, so the difference between the total binding energy in the mutant and wild type can be interpreted as the contribution only from the bases. As expected, the \( \text{Tf} \text{Cel9A} \) base (which binds directly to the substrate) has a larger contribution to the binding energy ($\sim$7.8 kcal/mol) compared to that of the base in \( \text{Tf} \text{Cel6A} \) ($\sim$2.1 kcal/mol). This result indicates that the GH6 enzyme greatly enhances product release by using the wire water as a buffer between the base and product.

Interestingly, the hydrogen-bonding network in the \( \text{Tf} \text{Cel6A} \) active site stabilizes the $^2S_O$ pucker of the $-1$ ring for the product conformation. In \( \text{Tf} \text{Cel9A} \) this ring is relaxed to the low-energy $^4C_1$ chair, and its C1 hydroxyl group hydrogen bonds directly with the catalytic base D58. This conformation is further stabilized by hydrogen-bonding between the C1 $\alpha$-hydroxyl and the nucleophilic water-stabilizing D55 in \( \text{Tf} \text{Cel9A} \), while the homologous residue in \( \text{Tf} \text{Cel6A} \), S181, stabilizes a puckered product because the latter pulls down the $-1$ ring oxygen. This pucker is known to promote reactivity in \( \text{Tf} \text{Cel6A} \) and occurs spontaneously as the second-to-leading glycosyl ring enters the $-1$ binding site (16). QM calculations of monosaccharides have shown that the energetic cost for puckering an $\alpha$-glucose (as in the product) in the $^2S_O$ orientation is approximately 4 kcal/mol greater than puckering a $\beta$-glucose (as in the reactant) in the $^2S_O$ orientation (29). Holding the product in a puckered state may further promote product release by stabilizing an unfavorable conformation while bound.

In our aforementioned simulation of the \( \text{Tf} \text{Cel6A} \) base mutated to alanine, the $-1$ sugar became free to transition between skew puckers $^2S_O$ and $^1S_3$. Because the product is in the $\alpha$-glucose configuration, this transition stabilizes the bound product by 3 kcal/mol (29). This result points to an additional role of the base (indirectly, through the wire water) in stabilizing the unfavorable $^2S_O$ pucker in the product state and promoting product release.

**DISCUSSION**

The catalytic mechanism of \( \text{Tf} \text{Cel6A} \) involves a counter-intuitive wire water molecule that is not strictly necessary to the overall chemistry of the enzyme’s reaction. Motivated by a desire to rationalize this exception to the classical mechanism for an inverting glycoside hydrolase, we disrupted the active site with a D175E mutation that we hypoth-
Advantages of a Distant Cellulase Catalytic Base

esized would supplant the wire water. We found that the longer E175 residue in the mutant was a hindrance to the catalytic cycle both during process ion and alignment for hydrolysis, highlighting the importance and intricacy of the remarkable network of hydrogen bonding interactions that stabilizes the active site of the wild-type enzyme. During process ion, only the wild-type base promoted hydrogen bonding that kept R174 and N182 out of the tunnel, contributing to an energetically favorable, spontaneous forward motion for the wild-type enzyme, which was not manifest in the mutant. In aligning for hydrolysis, hydrogen bonding in the mutant active site favors positioning the mutant carboxylate group even further from the nucleophilic water than in the wild type, which, compared to it, both makes accepting a proton more difficult and prevents it from helping align the nucleophilic water for attack. Calculations of low-energy conformations of the system in the hydrolysis-ready position with the $-2$ and $-1$ positions occupied by the cellulose chain indicate that the overall reaction barrier in the D175E mutant would be increased by at least between 1.5 and 2.5 kcal/mol. While this is a qualitative result, it corresponds with an activity reduction of approximately two orders of magnitude, and is therefore approximately consistent with measured activity reductions for the homologous mutants TfCel6A D79E (20) and CfCel6A D216E (21). This agreement allows us to address our primary question of why a longer side chain for the nucleophilic residue does not produce a more active enzyme. The effect on alignment for both proton transfer and procession may further explain the range of effects of these homologous mutants on substrates with different degrees of crystallinity.

To further understand the role of the more distant base in TrCel6A, we compared its active site to that of an inverting cellulase, TfCel9A, which does not require a water wire to shuttle a proton to the base during catalysis. While the active site residues align similarly in the two enzymes, the substrate is rotated approximately $45^\circ$ about its axis, creating the additional distance occupied by the “bridge” water in TrCel6A. We propose that the wire water buffers the post-reaction hydrogen bonding between the protonated base and the cellobiose product, easing product release. Consistent with this theory, we calculated an approximately 6 kcal/mol reduction in product binding energy due to differences in catalytic base alignment strategy. We thus suggest that the more distant base benefits the overall TrCel6A mechanism in terms of reduced association between the product and enzyme.

Although not addressed in this study, there are other potential mutations of TrCel6A that could obviate the need for a water wire during hydrolysis. Specifically, the wild-type positions 178 and 180 are occupied by alanines, but mutations to aspartate or glutamate (presumably in conjunction with a mutation of D175 to a non-protonatable residue) might potentially place the side chain of the base close enough to allow direct proton transfer from the nucleophilic water. The alanines at these positions are quite highly conserved in this family of enzymes, although a glutamate is observed at the position analogous to TrCel6A A180 in the H. insolens Cel6B enzyme (5). The large number of possible alternate sequences for these enzymes provides many opportunities to learn from how nature has optimized protein design. It is clear from this study that any such mutants would need to be carefully analyzed for how they could disrupt the highly tuned hydrogen-bonding network in the active site and their effects on multiple parts of the catalytic cycle.

Our results advance cellobiohydrolase enzyme engineering efforts by broadening the focus of the role of active site residues beyond the hydrolysis step. Family GH6 enzymes (and the conservation of their mechanism across different branches of life) expand our understanding of how enzymes can make small sacrifices in reactivity to enhance other aspects of the larger catalytic cycle.

**EXPERIMENTAL PROCEDURES**

*Molecular Dynamics Simulations* — The approach of our investigation using molecular dynamics (MD) simulations is briefly described here, with further detail in the supplemental material. All models were based on crystal structures deposited in the Protein Data Bank (PDB) (30). The initial structure for the procession study was constructed by combining features from two crystal structures. First, the crystal structure PDB ID: 1QK2 (18) (wild-type TrCel6A with a non-hydrolysable cellobetaose)
Advantages of a Distant Cellulase Catalytic Base

was stripped of its non-protein components. Then, the cellohexaose substrate from PDB ID: 4AVO (31) was manually aligned with its leading non-reducing-end ring in the +1 position as in our previous work (16). For the active site conformation studies, we started with the PDB ID: 1QJW (18) crystal structure (TrCel6A Y169F mutant complexed with cellotetraose) reverted to the wild type, and again replaced its substrate with that of PDB ID: 4AVO (31), this time in the active position with the leading ring in the −2 site. The active site water molecules appear in PDB ID: 1QJW and were retained for both the wild-type and D175E mutant models. The crystal structures in PDB format were converted into topology and coordinate files using the CHARMM (32) package and the CHARMM36 forcefield (33–36). The models were solvated in a periodic box using the TIP3 water model (37) and converted into Amber format using the CHAMBER program in ParmEd (38).

All simulations were performed using the Amber14 package (39). Unless otherwise stated, the SHAKE algorithm (40, 41) was used to constrain the lengths of all hydrogen bonds and the cutoff distance for non-bonded interactions was set to at least 8.0 Å. First, structures were minimized without SHAKE over 2500 steps (1250 steepest-descent method followed by 1250 conjugate gradient method). The systems were then heated in a periodic box in the NVT ensemble, using the Andersen thermostat (42) to take the temperature from 100 K to 300 K over 10,000 2-fs time steps followed by 1000 steps of constant-temperature simulation. Velocities were randomized every 1000 steps. MD production simulations were performed under the same conditions as the heating simulations at a constant temperature of 300 K.

Simulations of TrCel6A included positional restraints on the α-carbons of residues S106, A150, D200, N247, A280, I330, and Q437, as in our previous work (16). These atoms have been shown to have a low root mean square fluctuation (43), and restraining them prevents bulk motion of the protein.

Models were visualized using VMD (44). The PMF plots were produced using the umbrella integration method introduced by Kästner and Thiel (45, 46) and implemented by Stroet and Deplazes (47).

Acknowledgments: This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (48), which is supported by National Science Foundation grant number ACI-1548562. The authors thank Dr. Gregg Beckham, National Renewable Energy Laboratory, for helpful discussions.

Conflict of interest: The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions: T. B. performed the simulations, completed the free-energy calculations, drafted the manuscript, and created the figures. H. B. M. designed and coordinated the study. All authors analyzed the data. H. B. M. and J. S. critically revised the manuscript. All authors reviewed the results and approved the final version of the manuscript.
REFERENCES

1. Biddy, M. J., Davis, R., Humbird, D., Tao, L., Dowe, N., Guarnieri, M. T., Linger, J. G., Karp, E. M., Salvachúa, D., Vardon, D. R., and Beckham, G. T. (2016) The techno-economic basis for coproduct manufacturing to enable hydrocarbon fuel production from lignocellulosic biomass. ACS Sustain. Chem. Eng. 4, 3196–3211, 10.1021/acssuschemeng.6b00243

2. Cragg, S. M., Beckham, G. T., Bruce, N. C., Bugg, T. D. H., Distel, D. L., Dupree, P., Etxabe, A. G., Goodell, B. S., Jellison, J., McGeehan, J. E., McQueen-Mason, S. J., Schnorr, K., Walton, P. H., Watts, J. E. M., and Zimmer, M. (2015) Lignocellulose degradation mechanisms across the Tree of Life. Curr. Opin. Chem. Biol. 29, 108–119, 10.1016/j.cbpa.2015.10.018

3. García-Guevara, F., Avelar, M., Ayala, M., and Segovia, L. (2015) Computational tools applied to enzyme design – a review. Biocatalysis 1, 109–117, 10.1515/boca-2015-0009

4. Renata, H., Wang, Z. J., and Arnold, F. H. (2015) Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. Angew. Chem., Int. Ed. 54, 3351–3367, 10.1002/anie.201409470

5. Payne, C. M., Knott, B. C., Mayes, H. B., Hansson, H., Himmel, M. E., Sandgren, M., Ståhlberg, J., and Beckham, G. T. (2015) Fungal cellulases. Chem. Rev. 115, 1308–1448, 10.1021/cr500351c

6. Boisset, C., Fraschini, C., Schülein, M., Henrissat, B., and Chanzy, H. (2000) Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from Humicola insolens and its mode of synergy with cellobiohydrolase Cel7A. Appl. Environ. Microbiol. 66, 1444–1452, 10.1128/AEM.66.4.1444-1452.2000

7. Ståhlberg, J., Johansson, G., and Pettersson, G. (1993) Trichoderma reesei has no true exo-cellulase: All intact and truncated cellulases produce new reducing end groups on cellulose. Biochim. Biophys. Acta 1157, 107–113, 10.1016/0304-4165(93)90085-M

8. Mertz, B., Kuczzenski, R. S., Larsen, R. T., Hill, A. D., and Reilly, P. J. (2005) Phylogenetic analysis of family 6 glycoside hydrolases. Biopolymers 79, 197–206, 10.1002/bip.20347

9. Beckham, G. T., Ståhlberg, J., Knott, B. C., Himmel, M. E., Crowley, M. F., Sandgren, M., Soolt, M., and Payne, C. M. (2014) Towards a molecular-level theory of carbohydrate processivity in glycoside hydrolases. Curr. Opin. Biotechnol. 27, 96–106, 10.1016/j.copbio.2013.12.002

10. Nakamura, A., Tasaki, T., Ichiwata, D., Yamamoto, M., Okuni, Y., Visootsat, A., Maximilien, M., Noji, H., Uchiyama, T., Samejima, M., Igarashi, K., and Iino, R. (2016) Single-molecule imaging analysis of binding, processive movement, and dissociation of cellobiohydrolase Trichoderma reesei Cel6A and its domains on crystalline cellulose. J. Biol. Chem. 291, 22404–22413, 10.1074/jbc.M116.752048

11. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C., and Jones, T. A. (1990) Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. Science 249, 380–386

12. Koshland, Jr., D. E. (1953) Stereocchemistry and the mechanism of enzymatic reactions. Biol. Rev. 28, 416–436

13. Spezio, M., Wilson, D. B., and Karplus, P. A. (1993) Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry 32, 9906–9916, 10.1021/bi00089a006

14. Davies, G. J., Brzozowski, A. M., Dauter, M., Varrot, A., and Schülein, M. (2000) Structure and function of Humicola insolens family 6 cellulases: structure of the endoglucanase, Cel6B, at 1.6 Å resolution. Biochem. J. 348, 201–207

15. Koivula, A., Ruohonen, L., Wohlfahrt, G., Reinikainen, T., Teeri, T. T., Piens, K., Claeyssens, M., Weber, M., Vasella, A., Becker, D., Sinnott, M. L., Yu Zou, J., Kleywegt, G. J., Szardenings, M., Ståhlberg, J., et al. (2002) The active site of cellobiohydrolase Cel6A from Trichoderma reesei: The roles of aspartic acids D221 and D175. J. Am. Chem. Soc. 124, 10015–10024, 10.1021/ja012659q

16. Mayes, H. B., Knott, B. C., Crowley, M. F., Broadbelt, L. J., Ståhlberg, J., and Beckham, G. T. (2016) Who’s on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolase.
Advantages of a Distant Cellulase Catalytic Base

Chem. Sci. 7, 5955–5968, 10.1039/C6SC00571C

17. Gregg, K. J., Suits, M. D. L., Deng, L., Vocadlo, D. J., and Boraston, A. B. (2015) Structural analysis of a family 101 glycoside hydrolase in complex with carbohydrates reveals insights into its mechanism. J. Biol. Chem. 290, 25657–25669, 10.1074/jbc.M115.680470

18. Zou, J., Kleywegt, G. J., Stählberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T. T., and Jones, T. A. (1999) Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in celllobiohydrolase Cel6A from Trichoderma reesei. Structure 7, 1035–1045, 10.1016/S0969-2126(99)80171-3

19. Riccardi, D., König, P., Guo, H., and Cui, Q. (2008) Proton transfer in carbonic anhydrase is controlled by electrostatics rather than the orientation of the acceptor. Biochemistry 47, 2369–2378, 10.1021/bi701950j

20. Wolfgang, D. E. and Wilson, D. B. (1999) Mechanistic studies of active site mutants of Thermomonospora fusca endocellulase E2. Biochemistry 38, 9746–9751, 10.1021/bi990401v

21. Damude, H. G., Withers, S. G., Kilburn, D. G., Miller, Jr., R. C., and Warren, R. A. J. (1995) Site-directed mutation of the putative catalytic residues of endoglucanase CenA from Cellulomonas fimii. Biochemistry 34, 2220–2224, DOI: 10.1021/bi00007a016

22. Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B. R., Larenas, E. A., Stählberg, J., and Hansson, H. (2013) The structure of a bacterial celllobiohydrolase: The catalytic core of the Thermobifida fusca family GH6 celllobiohydrolase Cel6B. J. Mol. Biol. 425, 622–635, 10.1016/j.jmb.2012.11.039

23. Thompson, A. J., Heu, T., Shaghasi, T., Benyamino, R., Jones, A., Friis, E. P., Wilson, K. S., and Davies, G. J. (2012) Structure of the catalytic core module of the Chaetomium thermophilum family GH6 celllobiohydrolase Cel6A. Acta Crystallogr., Sect. D: Biol. Crystallogr. 68, 875–882, 10.1107/S0907444912016496

24. Varrot, A., Hastrup, S., Schülein, M., and Davies, G. J. (1999) Crystal structure of the catalytic core domain of the family 6 celllobiohydrolase II, Cel6A, from Humicola insolens, at 1.92 Å resolution. Biochem. J. 337, 297–304, 10.1042/0264-6021:3370270

25. Larsson, A. M., Bergfors, T., Dultz, E., Irwin, D. C., Roos, A., Driguez, H., Wilson, D. B., and Jones, T. A. (2005) Crystal structure of Thermobifida fusca endoglucanase Cel6A in complex with substrate and inhibitor: The role of tyrosine Y73 in substrate ring distortion. Biochemistry 44, 12915–12922

26. Zhou, W., Irwin, D. C., Escovar-Kousen, J., and Wilson, D. B. (2004) Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes. Biochemistry 43, 9655–9663, 10.1021/bi049394n

27. Li, Y., Irwin, D. C., and Wilson, D. B. (2007) Processivity, substrate binding, and mechanism of cellulose hydrolysis by Thermobifida fusca Cel9A. Appl. Environ. Microbiol. 73, 3165–3172, 10.1128/AEM.02960-06

28. Sakon, J., Irwin, D. C., Escovar-Kousen, J., and Karplus, A. P. (1997) Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. Nat. Struct. Biol. 4, 810–818, 10.1038/nsb1097-810

29. Mayes, H. B., Broadbelt, L. J., and Beckham, G. T. (2014) How sugars pucker: Electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis. J. Am. Chem. Soc. 136, 1008–1022, 10.1021/ja410264d

30. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242, 10.1093/nar/28.1.235

31. Wu, M., Bu, L., Vuong, T. V., Wilson, D. B., Crowley, M. F., Sandgren, M., Stählberg, J., Beckham, G. T., and Hansson, H. (2013) Loop motions important to product expulsion in the Thermobifida fusca glycoside hydrolase family 6 celllobiohydrolase from structural and computational studies. J. Biol. Chem. 288, 33107–33117, 10.1074/jbc.M113.502765

32. Brooks, B. R., Brooks, III, C. L., Mackerell, Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won,
Advantages of a Distant Cellulase Catalytic Base

Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., et al. (2009) CHARMM: The biomolecular simulation program. J. Comput. Chem. 30, 1545–1614, 10.1002/jcc.21287

33. Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, Jr., A. D. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone $\phi$, $\psi$ and sidechain $\chi$ 1 and $\chi$2 dihedral angles. J. Chem. Theory Comput. 8, 3257–3273, 10.1021/ct300400x.Optimization

34. MacKerell, Jr., A. D., Feig, M., and Brooks, III, C. L. (2004) Improved treatment of the protein backbone in empirical force fields. J. Am. Chem. Soc. 126, 698–699, 10.1021/ja036959e

35. Mackerell, Jr., A. D., Bashford, D., Bellott, M., Dunbrack, Jr., R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 102, 3586–3616, 10.1021/jp973084f

36. Beglov, D. and Roux, B. (1994) Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. J. Chem. Theory Comput. 10, 9050–9063, 10.1063/1.466711

37. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935, 10.1063/1.445869

38. Crowley, M. F., Williamson, M. J., and Walker, R. C. (2009) CHAMBER: Comprehensive support for CHARMM force fields within the AMBER software. Int. J. Quantum Chem. 109, 3767–3772

39. Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Gohlke, H., Goetz, A. W., Gusarov, S., Homeyer, N., Janowski, P., Kauz, J., et al. (2014) Amber 14, University of California, San Francisco

40. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. J. Comput. Phys. 23, 327–341, 10.1016/0021-9991(77)90098-5

41. Miyamoto, S. and Kollman, P. A. (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. J. Comput. Chem. 13, 952–962, 10.1002/jcc.540130805

42. Andrea, T. A., Swope, W. C., and Andersen, H. C. (1983) The role of long ranged forces in determining the structure and properties of liquid water. J. Chem. Phys. 79, 4576–4584, 10.1063/1.446373

43. Bu, L., Crowley, M. F., Himmel, M. E., and Beckham, G. T. (2013) Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. J. Biol. Chem. 288, 12175–12186, 10.1074/jbc.M113.462465

44. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. J. Mol. Graph. 14, 33–38, 10.1016/0263-7855(96)00018-5

45. Kästner, J. and Thiel, W. (2005) Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration”. J. Chem. Phys. 123, 144104, 10.1063/1.2052648

46. Kästner, J. and Thiel, W. (2006) Analysis of the statistical error in umbrella sampling simulations by umbrella integration. J. Chem. Phys. 124, 234106, 10.1063/1.2206775

47. Stroet, M. and Deplazes, E. Python implementation of the umbrella integration method for potential of mean force (pmf) calculations, https://doi.org/10.5281/zenodo.164996

48. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., and Wilkens-Diehr, N. (2014) XSEDE: Accelerating scientific discovery. Comput. Sci. Eng. 16, 62–74, http://doi.ieeecomputer-society.org/10.1109/MCSE.2014.80
Advantages of a Distant Cellulase Catalytic Base

FOOTNOTES
This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (48), which is supported by National Science Foundation grant number ACI-1548562. Abbreviations: TfCel6B, Thermobifida fusca Cel6A; TfCel9A, Thermobifida fusca Cel9A; TrCel6A, Trichoderma reesei Cel6A; CV, collective variable; GH, glycoside hydrolase; GH6, GH family 6; MD, molecular dynamics; PDB, Protein Data Bank; PMF, potential of mean force; and RMSD, root mean square deviation.

FIGURES

Figure 1: A comparison of the “classical” inverting cellulase hydrolase mechanism proposed by Koshland (12) (left) to that used by TrCel6A (right). The overall reaction chemistry is the same in each case, but TrCel6A requires the presence of the additional “wire water” in the active site due to the increased distance between the attacking water molecule and the catalytic base. Atoms and bonds belonging initially to the nucleophilic water and the wire water are depicted in cyan and red, respectively.
Figure 2: PMFs for substrate procession. Snapshots of substrate positions from the simulations with the wild-type enzyme are shown at RMSD values of 0.25 and 10.25 Å, respectively. The substrate binding sites −2 to +3 are labeled, along with two catalytic residues in the wild type. The RMSD compares the positions of the leading two glucose rings relative to the “pre-slide” structure (left). The dashed lines indicate the RMSD values of the “pre-slide” and “slide” conformations.
Figure 3: Snapshot of the TrCel6A wild type and D175E mutant mid-procession, corresponding to an RMSD value of 6.5 Å from Figure 2. While the wild-type base hydrogen bonds with nearby R174 and N182 to keep all three residues tucked away from the processing substrate, the added length of E175 disfavors this binding and leaves the residues to clash with the leading ring.
Figure 4: PMFs from umbrella sampling for dihedral rotation around the D221 Cα-Cβ bond (dihedral angle indicated in green in the snapshots shown from the wild-type enzyme simulations), representing the transition from the conformation for catalytic base/acid proton transfer (at around 80°) toward the D221 position for glycosidic cleavage (near 170°).
**Figure 5:** Diagram of the active site of *Tr*Cel6A for the wild type and D175E mutant. An exquisite network of hydrogen bonding interactions stabilizes the wire water between the attacking water (hydrogen bonding with S181 in both cases) and the wild-type base, D175. The longer side chain of the mutant E175 residue compels it out of the active site, chasing after the void left by the acid D221 after the rotation described in Figure 4 and resulting in an inactive conformation.
Figure 6: PMF for rotation of dihedral around the bond between the $\beta$- and $\gamma$-carbons in the mutant E175 residue. Three distinct minima are observed at $-170^\circ$, $-50^\circ$, and $40^\circ$. The minimum highest in free energy corresponds to a potentially hydrolytically active conformation, with E175 hydrogen bonding with the nucleophilic water as shown in the corresponding snapshot.
Figure 7: Comparison of the product-state active sites of TrCel6A (A and C) and TfCel9A (B and D). In A and B, the dashed lines highlight specific hydrogen bonds further discussed in the text. Panels C and D show the same conformation from a different orientation, to highlight the differences in relative positions of the product and acid/base residues.
