The hydrolysis mechanism of a GH45 cellulase and its potential relation to lytic transglycosylase and expansin function

Received for publication, October 8, 2019, and in revised form, February 12, 2020 Published, Papers in Press, February 13, 2020, DOI 10.1074/jbc.RA119.011406

© Vivek S. Bharadwaj†, Brandon C. Knot‡, Jerry Ståhlberg§, Gregg T. Beckham¶, and Michael F. Crowley‡‡

From the †Biosciences Center and ‡National Bioenergy Center, National Renewable Energy Laboratory, Golden, Colorado 80401 and the §Department of Molecular Sciences, Swedish University of Agricultural Sciences, P. O. Box 7015, 750 07 Uppsala, Sweden

Edited by Gerald W. Hart

Family 45 glycoside hydrolases (GH45) are endoglucanases that are integral to cellulosytic secretomes, and their ability to break down cellulose has been successfully exploited in textile and detergent industries. In addition to their industrial relevance, understanding the molecular mechanism of GH45-catalyzed hydrolysis is of fundamental importance because of their structural similarity to cell wall-modifying enzymes such as bacterial lytic transglycosylases (LTs) and expansins present in bacteria, plants, and fungi. Our understanding of the catalytic itinerary of GH45s has been incomplete because a crystal structure with substrate spanning the –1 to +1 subsites is currently lacking. Here we constructed and validated a putative Michaelis complex in silico and used it to elucidate the hydrolytic mechanism in a GH45, Cel45A from the fungus Humicola insolens, via unbiased simulation approaches. These molecular simulations revealed that the solvent-exposed active-site architecture results in lack of coordination for the hydroxymethyl group of the substrate at the –1 subsite. This lack of coordination imparted mobility to the hydroxymethyl group and enabled a crucial hydrogen bond with the catalytic acid during and after the reaction. This suggests the possibility of a nonhydrolytic reaction mechanism when the catalytic base aspartic acid is missing, as is the case in some LTs (murine transglycosylase A) and expansins. We calculated reaction free energies and demonstrate the thermodynamic feasibility of the hydrolytic and nonhydrolytic reaction mechanisms. Our results provide molecular insights into the hydrolysis mechanism in HiCel45A, with possible implications for elucidating the elusive catalytic mechanism in LTs and expansins.

This work was supported by Department of Energy, Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies Office via Grant DE-AC36-08GO28308. The authors declare that they have no conflicts of interest with the contents of this article. This article contains Figs. S1–S6, Tables S1–S3, Sections S1–S4, and Videos S1 and S2.

1 To whom correspondence may be addressed: E-mail: gregg.beckham@nrel.gov.
2 To whom correspondence may be addressed: E-mail: michael.crowley@nrel.gov.
3 The abbreviations used are: GH, glycosyl hydrolase; LTs, lytic transglycosylase; RC, reaction coordinate; TPS, transition path sampling; LM, likelihood maximization; MC, Michaelis complex; QM, quantum mechanics; MM, molecular mechanics; MD, molecular dynamics; TS, transition state; AS, aimless shooting; OP, order parameter; AMBER, assisted model building with energy refinement; CHARMM, Chemistry at Harvard Macromolecular Mechanics.

Cellulases are glycoside hydrolase (GH)3 enzymes that hydrolyze glycosidic bonds in cellulose and have been the mainstay of enzyme technology platforms for the production of biofuels from cellulosic feedstocks (1). Among the various GH families that have been characterized structurally and biochemically (2), family 45 GHs (GH45) have garnered substantial interest for their industrial utility in textile applications, such as denim polishing, and as an integral component of detergents (3, 4).

GH45s have been classified into three subfamilies (A, B, and C) based on phylogenetics (1). Davies et al. (5, 6) characterized the native and oligosaccharide-bound (cellulbiose and cellotriose) GH45 structures from Humicola insolens (HiCel45A), belonging to subfamily A. Biochemical experiments on HiCel45A revealed that HiCel45A hydrolyzes glycosidic bonds in cellulose-oligomers via a one-step inverting mechanism. Point mutations to three candidate aspartic acid residues (D121A, D10A, and D114A) at the putative active site implicated Asp-121 (proton donor) and Asp-10 (catalytic base) as indispensable catalytic residues for hydrolytic activity in HiCel45A (5) (Fig. 1A). CocrySTALLization of HiCel45A with cellotriose yielded a structure with two cellotriose molecules occupying subsites –4 to –2 and +1 to +3. However, no sugar occupancy was observed at the –1 subsite in this structure (Fig. 1A) nor in any published structure of a member of subfamily A (5). Given that the enzyme hydrolyzes the glycosidic bond between the –1 and +1 binding sites, the absence of the substrate at the –1 subsite leaves a gap in our understanding of the GH45 catalytic itinerary. Furthermore, it has been demonstrated that cellulases facilitate hydrolysis by modifying the conformation of the sugar at this site (7–9). Beyond the static information provided by crystal structures, a more detailed understanding of the catalytic mechanism of HiCel45A remains unknown.

GH45s bear structural similarity to other cell wall-modifying enzymes found in plants and bacteria: expansins and lytic transglycosylases (LTs) (1, 10, 11). Expansins facilitate growth in plant cell walls (10, 12), but hydrolytic products of their action have not been detected (13). LTs nonhydrolytically cleave glycosidic linkages in bacterial cell wall peptidoglycans (14), forming an anhydro glycan product (15–17). Expansins as well as one family of LTs (called family 2 in the classification scheme of Blackburn and Clarke (18), classified as GH102 in the...
Carbohydrate-active Enzymes database and represented by *Escherichia coli* MltA (11) share a similar double-ψ β-barrel fold with GH45s and are structurally similar at the active site. Although catalytic activity in most cellulases (including GH45s) is mediated by a proton donor and a base (usually an aspartate or glutamate dyad (19)), the catalytic domain in MltA and the putative active site in expansins is characterized by only a single aspartate. As illustrated in Fig. 1B, MltA and expansins share the conserved HFD domain (20) containing the aspartic acid that serves as the proton donor in *H. insolens* GH45A but display no analog to the base aspartic acid. GH45 subfamily C also features a similar overall fold and conservation of the HFD domain (21) as well as the conspicuous absence of the base aspartic acid at the active site (22). Given the catalytic machinery of *H. insolens* GH45A for conventional hydrolysis, its well-characterized biochemistry, and its active-site similarity (conserved HFD domain) with GH45 subfamily C, we posit that mechanistic details of GH45A hydrolysis will lead us to insights into plausible modes of catalysis where the base aspartate is mutated, which could also be widely applicable to other cell wall–modifying proteins with a similar active-site architecture.

Although crystal structures provide static pictures of long-lived states, molecular simulations provide the requisite spatiotemporal resolution necessary for elucidating atomistic and dynamic details of reaction mechanisms. Though many simulation techniques applied to chemical reactions require a priori specification of the reaction coordinate (RC), which quantifies dynamic progress from reactant to product, transition path sampling (TPS) produces an unbiased ensemble of trajectories (23), which can be analyzed using likelihood maximization (LM) so that the optimal RC is an output of the simulations. TPS has been instrumental in gaining valuable insight into hydrolytic mechanisms in cellulases (24, 25). Knott et al. (24) elucidated the two-step retaining mechanism in *Trichoderma reesei* Cel7A. More recently, Mayes et al. (25) applied TPS to reveal the single-step inverting mechanism in Cel6A from the same fungus. These studies enabled key insights into the RC, free energy barriers for the reactions, sugar puckering itineraries, and rate-limiting steps in these important cellulohydrolases.

In this study, we employ TPS to elucidate the hydrolysis mechanism in the inverting endoglucanase *HiCel45A*. We begin by constructing and validating a Michaelis complex (MC) in silico. Subsequent quantum mechanics/molecular mechanics (QM/MM) TPS simulations result in determination of the RC for hydrolysis by WT *HiCel45A*. In addition, insights gained from the WT hydrolysis mechanism form the basis for an LT-like nonhydrolytic mechanism for the *HiCel45A D10N* mutant, where the catalytic base aspartic acid is mutated to asparagine. Computed free energy barriers demonstrate the feasibility of the nonhydrolytic mechanism which could serve as a critical segue to understanding mechanistic action in LTs and expansins.

**Results**

**In silico construction of the MC**

The MC in *HiCel45A* has not yet been structurally characterized. To construct a model MC for *HiCel45A*, we began with the crystal structure featuring two cellotriose molecules occupying subsites −4 to −2 and +1 to +3 (PDB code 4ENG) and connected these via the sugar occupying the −1 site from another inverting GH, *Thermobifida fusca* Cel6A (TfCel6A, PDB code 4AVO) (26). The sugar conformation at the −1 site in the substrate binding groove is critical for catalysis, as GHs generally stabilize the ring pucker states away from the chair
conformations that are stable in solution (8, 24, 25). Inverting cellulases utilize a single-step nucleophilic substitution involving an oxocarbenium ion at the anomeric carbon (25, 27, 28). The important criteria facilitating the hydrolysis mechanism include the puckered state of the substrate ring (generally $^{2}S_{O}$ for inverting cellulases (28)) at the $-1$ subsite, the presence of a water molecule between the anomeric carbon and the base (Asp-10), and the orientation of the acid (Asp-121) proton toward the glycosidic oxygen at the $-1$ site. Unbiased molecular dynamics (MD) simulations were analyzed for simultaneous satisfaction of all three criteria, starting from the constructed HiCel45A MC, and each frame was categorized as catalytically competent (all above criteria met) or not competent for catalysis (any one or more criteria not met). Of the five independent MD runs, one simulation demonstrated catalytically competent conformations 54.5% of the time. There, the catalytically competent conformations were stable over tens to hundreds of picoseconds, timescales beyond what would be required for a single hydrolytic event (Fig. S3). This result provides sufficient confidence in the stability of the putative MC for more detailed studies of the hydrolysis reaction mechanism.

Further analysis of the equilibrium MD simulations reveals that the hydroxymethyl group of the substrate ring at the $-1$ position is not coordinated by specific direct interactions with the protein, although a hydrogen bond with an active site water is observed ($\sim$14% occupancy over 100 ns simulation). This differs from $T_{f}$Cel6A, where the oxygen atom on the hydroxymethyl group of the substrate ring bound at the $-1$ subsite in GH6 coordinates closely with an active site tyrosine to maintain its puckered conformation in the MC (29). This has implications in the reaction mechanism in the absence of the catalytic base, as described in subsequent sections. The important enzyme–substrate interactions observed from MD simulations involve Asp-178, Asp-10, Asn-179, Glu-48, and Trp-18 (Fig. 2 and Fig. S4).

**Hydrolysis mechanism in HiCel45A**

The catalytically competent configurations from the simulations described above enabled investigation of the HiCel45A hydrolysis reaction mechanism using QM/MM MD simulations. Fig. 3A depicts snapshots of the reactant, transition state, and product configurations of the system, accompanied by chemical structures depicting the bonds being broken and formed. Note the strained conformation of the substrate ring bound at the $-1$ subsite in the reactant and transition state conformations and the development of a strong interaction between the hydrogen atom on O6 and the deprotonating oxygen on Asp-121. Movie S1 depicts one of these trajectories.

The ensemble of reactive and unreactive trajectories was systematically analyzed to determine the optimal RC. LM analysis determined the optimal RC to include two differences in distances: the difference between the length of the forming and breaking bonds involving the acidic proton and the difference between the length of the forming and breaking bonds involving the anomeric carbon (Fig. 3A and Section S2.2). This RC was validated using the $p_{B}$ histogram test (30) (Fig. S2).

In GH enzymes, ring puckering at the active site is an important feature of the catalytic itinerary. Although the 1.5-ps trajectories are long enough to capture bond-breaking phenomena, observations of puckering transitions in the reactant and product basins require longer simulation times. Hence, starting from the ends of a successful reactive trajectory, we also conducted 500-ps QM/MM simulations to evaluate this behavior for the substrate at the $-1$ subsite pre- and post-reaction (Fig. 3B). The substrate in the reactant state starts near the $^{2}S_{O}$ region ($\theta_{cremer-pople}$ (31) $\sim$ 90°) plot and then immediately shifts to the $^{4}C_{1}$ ($\theta_{cremer-pople}$ $\sim$ 10°) chair state, with the hydroxymethyl and 2, 3, and 4 hydroxyls in equatorial positions. In the product state, the substrate ring at the $-1$ subsite again starts in the vicinity of the $^{2}S_{O}$ region and explores various puckers states ($\theta_{cremer-pople}$ $\sim$ 90°) before transitioning to the less common $^{4}C_{4}$ ($\theta_{cremer-pople}$ $\sim$ 180°) state with axial substituents.

The elucidation of an optimal RC sets the stage for free energy calculations along the inverting hydrolysis mechanism.
During free energy calculations, the sugar pucker in the reactant windows (RC $<-1.3$) was restrained to $5S_O$, as this is in analogy with the pucker state observed in GH6, which exhibits similar catalytic machinery (25, 26). Likewise, product windows (RC $>0.7$) were restrained to $1C_4$, as we observed this to be the stable conformation in the product basin. The resulting free energy profile revealed a slightly endergonic $3.8 \pm 1.9$ kcal/mol reaction with a barrier of $23.55 \pm 0.3$ kcal/mol.
The proposed catalytic mechanism in the D10N mutant revealed no hydrolytic activity (for the reaction in the absence of the WT catalytic base Asp-10). This suggests the possibility of Asp-121 acting as both acid and base of the hydroxymethyl proton with the catalytic acid Asp-121 in equivalent position of Asp-10 in only GH45 subfamilies A and B exhibit the catalytic base at the −1 subsite. The catalyst activity for MltA and expansins the conserved catalytic acid, but of these, the analysis for the 500-ps unrestrained QM/MM MD of the product also reveals the presence of a hydrogen bond between the hydroxymethyl group, enables close coordination of the C6 hydroxyl hydrogen with the Asp-121 oxygen as indicated by the average distances ~2 Å in the product state. Hydrogen bond analysis for the 500-ps unrestrained QM/MM MD of the product also reveals the presence of a hydrogen bond between the hydroxymethyl proton and Asp-121 oxygen in 58% of the frames.

Proposed catalytic mechanism in the D10N mutant

The active sites of GH45 subfamilies A, B, and C share with MltA and expansins the conserved catalytic acid, but of these, only GH45 subfamilies A and B exhibit the catalytic base at the equivalent position of Asp-10 in HiCel45A. The coordination of the hydroxymethyl proton with the catalytic acid Asp-121 in our simulations of the hydrolysis mechanism in WT HiCel45A suggests the possibility of Asp-121 acting as both acid and base for the reaction in the absence of the WT catalytic base Asp-10. Although experimental characterization of the HiCel45A D10N mutant revealed no hydrolytic activity (5), this does not preclude other catalytic activity in the HiCel45A D10N mutant, similar to MltA or expansins. The role of aspartic/glutamic acid as the sole catalytic residue mediating glycosidic bond lysis is the established mechanism in LTs (1, 11, 15–17, 32, 33). Fig. 5A illustrates the mechanism proposed in the HiCel45A D10N mutant, which involves initiation of glycosylation by proton transfer from the acid (Asp-121) to the glycosidic oxygen, followed by transfer of the hydroxymethyl hydroxyl proton to Asp-121 acting as a base. This generates a nucleophile O6 and an oxocarbenium ion at the anomeric carbon, which react to form a 1,6-anhydro product (Fig. 5).

We sought to evaluate the relative feasibility of this proposed reaction mechanism via an RC that is the difference in the distances between the anomeric carbon to the glycosidic oxygen and the −1 hydroxymethyl oxygen to the anomeric carbon (the breaking and forming C-O bonds, respectively). This is the most significant component of the WT RC, consistent with past studies (24, 25). The free energy simulations estimate that the proposed reaction mechanism is exergonic by 14.8 ± 0.1 kcal/mol with a free energy barrier of 30.56 ± 0.13 kcal/mol. This free energy barrier is higher than the hydrolysis mechanism by ~7 kcal/mol.

Configurations near the barrier in Fig. 5C were used to initiate further unbiased trajectories for conceptualizing the sequence of events that constitute the reaction. Analysis of 16 of these reactive trajectories reveals that the reaction is predicted to consist of three main stages: glycolysis initiated by proton transfer from Asp-121; transfer of the hydroxymethyl proton to the Asp-121, which now acts as a base; and nucleophilic attack of the hydroxymethyl group onto the oxocarbenium ion at the anomeric carbon (Fig. S6). Movie S2 depicts one of these reactive trajectories.

Discussion

A detailed molecular understanding of the inverting hydrolysis mechanism in HiCel45A has to date been hindered by the lack of an experimentally characterized MC. Here we constructed the MC in silico and elucidated the WT hydrolysis mechanism (including a unique puckering itinerary) via TPS simulations. The insights facilitated exploration of an LT-like mechanism of glycosidic bond cleavage in the absence of the base aspartic acid (i.e. the D10N HiCel45A mutant). The predicted plausibility of this mechanism has potential implications for our understanding of the mode of action in GH45 subfamily C, LTs, and expansins.

Prior TPS studies of enzymatic reactions have focused on retaining and inverting cellobiohydrolases and glycosynthases (24, 25, 34), elucidating accurate RCS, free energies, puckering itineraries, and rate-limiting steps. This study extends the application of TPS to the inverting mechanism of an endoglucanase. As illustrated in Fig. 3, the key OPs that constitute the RC (OP#7 and OP#9 in Table S1) in HiCel45A-catalyzed hydrolysis involve differences in distances representing cleaved and formed bonds, akin to reaction mechanisms elucidated in TrCel6A and TrCel7A (24, 25). When taken together with the present findings regarding HiCel45A, these studies overall suggest a general trend where the major components of optimized RCS in GH-catalyzed cellulose deconstruction are relatively simple, involving distances of breaking and forming bonds, especially C-O bonds.

Computational reaction barriers allow comparison with experimentally observed turnover rates. Assuming a transition coefficient of 1, an upper estimate for the rate constant ($k_{cat}$) can be obtained using transition state (TS) theory (Section...
In *HiCel45A*, the experimental turnover $k_{\text{cat}}$ was measured to be 17.7 s$^{-1}$ for the cellohexaose substrate (5). The computed barrier of 23.55 ± 0.3 kcal/mol corresponds to a rate coefficient of 4.01 ± 10$^{-5}$ s$^{-1}$. This may indicate that some TS-stabilizing interactions are absent in our MC as constructed, although these details await experimental determination of the MC structure.

Another difference in the predicted reaction free energy profile is the presence of a much more stable hydrolysis product in the *CtGH8*, *TrCel6A*, and *TrCel7A* mechanisms (24, 25, 36), where the reaction is exergonic, whereas in this study, the reaction is mildly endergonic by 3.8 ± 1.9 kcal/mol. This difference could be attributed to a number of reasons, including, but not limited to, an imperfect MC or the inherent inaccuracies of the semiempirical force fields used here. The observed endergonicity could be related to the unique puckering itinerary observed in *HiCel45A*. In both inverting (GH8 (36) and GH6 (25)) and retaining (GH7 (24, 37) and GH16 (38)) β-glucosidases, the puckering itinerary for the substrate ring bound at the −1 subsite is restricted to the top hemisphere of the Cremer–Pople coordinates (31). This involves the −1 subsite substrate ring transitioning from the $^4C_1$ chair (in solution) to a strained pucker conformation in the MC and transition state followed by return to the $^4C_1$ chair conformation in the product state (24, 25, 36–38). However, in this study, the puckering itinerary in *HiCel45A* involves the reactant binding in the $^4C_1$ chair conformation, the MC puckering to the $^2S_0$ state, and the product transitioning to the $^1C_4$ conformation. Although the $^1C_4$ conformation has not been observed previously in products of cellulose hydrolysis, it has been characterized experimentally and computationally in the GH47 α-mannosidase (39). In *HiCel45A*, the $^1C_4$ state is stabilized by interactions of the hy-
dioxymethyl group with Asp-121, enabling an intramolecular hydrogen bond between the proton on the C3 hydroxyl group and the dioxymethyl oxygen (Fig. 6). Another hydrogen bond is observed between the proton on the C2 hydroxyl group and the glycosidic oxygen connecting to the sugar bound at the −2 subsite.

The ability of HiCel45A to stabilize the 1C4 conformation via interactions of Asp-121 with dioxymethyl groups may be significant, considering the possible similarity of reaction mechanisms in HiCel45A D10N and LTs such as MltA. In general, LTs are essential bacterial cell division, macromolecular transport, and growth by cleaving glycosidic bonds in peptidoglycans without hydrolysis and are characterized by a single catalytic residue (Asp/Glu) at the active site (14). In a recent QM/MM study, Byun et al. (40) described the energetics and puckering itinerary for the LT (MltE) mechanism. It was observed that the lowest-energy pucker conformation in the final 1,6-anhydro product (1,6-anhydroMurNAc) of the LT mechanism is the 1C4 chair. Hence, the ability of the HiCel45A active site to stabilize the 1C4 conformation further bolsters the hypothesis that it may facilitate formation of the 1,6 anhydro product.

The GH45 subfamily A members HiCel45A and Humicola grisea Cel45A hydrolyze xylan (41) (which differs from cellulose by the absence of dioxymethyl exocyclic groups), whereas no xylan activity was observed for a GH45 subfamily C member (PcCel45A) (21). Based on the rapid dissociation of xyloheptaose in MD simulations, Godoy et al. (21) pointed to the importance of the dioxymethyl groups for substrate binding as a key differentiator between cellulose and xylan activity by PcCel45A (21). Our simulations on the HiCel45A D10N mutant indicate that the interaction between the dioxymethyl group of the glucose moiety at the −1 subsite and the catalytic acid is essential for glycolysis in the absence of the catalytic base. These results suggest that, in addition to binding, the absence of dioxymethyl groups on xylan may preclude an important interaction with the catalytic acid, possibly prohibiting xylan hydrolysis at the GH45 subfamily C active site.

Fig. 1B illustrates conservation of the active site domain and the catalytic aspartic acid of MltA and expansins. Although the barrier computed here for 1,6 anhydro product formation in HiCel45A D10N is relatively high, our work still predicts that

The proposed mechanism is potentially feasible, as in LTs such as MltA and MltE as well as expansins, with some differences in stabilizing interactions at the active site. For example, the peptidoglycan substrate of LTs exhibits an acetyl group that stabilizes the TS, whereas HiCel45A has no analogous stabilizing interaction. Consider also that Quay et al. (42) recently characterized structures of a stationary phase survival protein in Bacillus subtilis that has significant similarities with LTs (as well as GH45s and expansins) and observed the presence of 1,6 anhydro GlcNAc (N-acetylgalactosamine) at the active site when cocrystallized with a pentasaccharide (N-acetylgalactosamine). A superposition of this substrate-bound structure with a snapshot from our simulations of the product in the HiCel45A D10N mutant reveals similarities at the active site as well as in the substrate binding modes (Fig. 7).

The effect of expansins on various plant cell wall polysaccharides has been well characterized, with their activity most prevalent in xylglucan composites and mild extension activity observed in cellulose substrates (43, 44). Despite close structural similarities between expansins and GH45s, the general consensus on the expansins’ mode of action is that they are nonenzymatic (10, 45), based on the lack of observable hydrolytic activity (10, 13). The hydrolytic activity of GHs is quantified by assays that require the presence of a reducing end on a sugar (43). The 1,6 anhydro product does not form a reducing end on the sugar; hence, it is plausible that expansins employ an enzymatic mode of action despite the lack of detectable hydrolytic activity. We also note that an interesting “Newton’s cradle” hydrogen transfer mechanism has been proposed to be involved in catalysis in GH45 subfamily C, which is also lacking the catalytic base of GH45 subfamily B. Combined neutron and X-ray crystallography studies of PcCel45A have revealed amide-imidic acid tautomerization of one strategically positioned asparagine (Asn-92) and a proton transfer chain through the protein from Asn-92 to the catalytic acid Asp-114 (22). However, PcCel45A generates hydrolysis products, i.e.
Reducing sugars. The lack of observed hydrolytic activity in expansins points to the presence of an alternative catalytic strategy for glucosidic bond cleavage without the formation of hydrolytic products, as we have studied here in HiCel45A D10N. Our findings support the hypothesis that the HiCel45A D10N and expansin active sites, as in LTs such as MltA and MltE, might be capable of cleaving glucosidic linkages in the absence of a catalytic base. Furthermore, Godoy et al. (21) reported recently that a peptide pattern recognition analysis revealed that the asparagine residue involved in the mechanism proposed by Nakamura et al. (22) is not universally conserved among all subfamily C GH45s, suggesting the distinct possibility of an alternative mechanism for catalytic action in the absence of the base aspartic acid.

Although this study predicts the feasibility of 1,6 anhydro glucan formation on cellulosic substrates, albeit with a high free energy barrier, other substrates with functional groups on the C2 and C3 hydroxyl groups may promote transition state stabilization. This could rationalize the fact that expansins show greater activity on xylol glucan substrates, which are decorated with acetyl/methyl functional groups, compared with the undecorated cellulose substrate (44). Hence, the insights gained in this study suggests the need for re-evaluation of potential enzymatic activity in expansins and development of an experimental assay to test the presented hypothesis.

**Experimental procedures**

**Preparing the MC**

The MC is not available from reported structures for the inverting cellulase HiCel45A. Among the other inverting GH families that act on glucosides (e.g. GH8 and GH6), the MC is observed to adopt the $^2\text{S}_2^g$ pucker conformation for the substrate at the $-1$ subsite (25, 28, 36). The MC for TjCel6A presents a template for constructing the MC for HiCel45A, as its ligand spans six subsites across the active site, including a skewed sugar ring at $-1$ site (PDB code 4AVO) (26). The glucan chain substrate in the TjCel6A structure spans subsites $-2$ to $+4$, whereas in the HiCel45a structure (PDB code 4ENG (5)), it is in two pieces, across subsites $-4$ to $-2$ and $+1$ to $+3$, missing the key $-1$ sugar. Transferring cellohexaose to HiCel45A based on substrate-based alignment of HiCel45A and TjCel6A at sites $-2$ to $+2$ enabled construction of the MC for HiCel45A.

**Molecular dynamics simulations**

The CHARMM36 forcefield was used for modeling solvated systems of native and cellohexaose-bound HiCel45A, whereas the TIP3P forcefield was used for modeling water (14, 46, 47). The seven disulfide bridges observed in the crystal structure (5) were included, and the protonation states for amino acid side chains were determined using the H++ web server for a pH of 5, the pH at which HiCel45A has been assayed (1, 41, 48). The crystal structure contains a mutation of the catalytic base Asp-10 to asparagine; this was converted back to the native aspartic acid, and the active-site proton donor Asp-121 (catalytic acid) was protonated. All water molecules observed in the crystal structure were retained before solvating the systems with a 12 Å buffer. The net charge on the system was neutralized with addition of sodium ions. The equilibration protocol involved initial restraints on the protein and the substrate, which were gradually released in subsequent simulations (Table S1). During these simulations, dihedral restraints were employed to maintain the pucker state of the substrate ring bound at the $-1$ subsite and distance restraints for maintaining the acid proton orientation toward the glucosidic oxygen; details are described in Section S1. This was followed by five independent unrestrained production runs, each for 100 ns at 300 K. The DOMDEC engine was used with the CHARMM molecular simulation package for these MD simulations with a 2-fs time step and bonds to hydrogen atoms constrained using the SHAKE algorithm (49–51). The simulations were analyzed using the CHARMM analysis tools (49). Catalytically competent configurations (as defined in Fig. S3) were selected from these equilibrium simulations for subsequent QM/MM simulations.

**QM/MM setup**

The CHAMBER utility was used to convert the CHARMM files to the AMBER format for QM/MM simulations (52). All reactions were modeled using the QM/MM MD suite in AMBER (53). The QM region was modeled using the self-consistent charge density functional tight binding method (54, 55), which has been demonstrated previously to be well-suited for studying hydrolysis reactions in GHs (24, 25). Although self-consistent charge density functional tight binding is a semipirical approach, it enables access to many thousands of reactive and unreactive trajectories required for judicious analysis of a transition state ensemble from path sampling. The method has also been demonstrated to reliably reproduce energies and geometries for reactions involving C-O and O-H bond formation and breakage encountered in this study (56). The QM/MM boundary was set across C-C bonds, and hydrogen atoms were used for link atoms as in the default AMBER QM/MM protocol. The QM/MM region consisted of the 89 atoms that include the side chains of the acid and base (Asp-121 and Asp-10), the hydrolytic water, the substrate rings at the $-1$ and $+1$ subsites, and Thr-6, Tyr-8, and Asp-114, which have close substrate interactions. All QM/MM simulations were conducted with a 1-fs timestep, a 10 Å nonbonded cutoff, periodic boundary conditions, and in the isothermal–isobaric ensemble at 300 K and 1.0 bar. Particle mesh Ewald was used for calculation of long-range electrostatics (57). Temperature control was maintained using an Andersen thermostat.

**TPS and free energy calculations**

The aimless shooting (AS) (30) variety of TPS (58), as described in detail in Section S2, starts with guessing a transition state configuration and initiating trajectories with randomized velocities in the “forward” direction and the opposite of these velocities in the “backward” direction. A Monte Carlo procedure is then carried out, with the acceptance criterion for a trajectory being that it connects the reactant and product basins (as defined in Fig. S1). Initial guesses for the transition states were obtained by restraining key bonds (namely, anomeric C to glycosidic oxygen and water oxygen to anomeric carbon) to putative transition state values from catalytically competent structures observed from the MD simulations of...
the constructed MC. This produced 32 near-TS structures, each of which seeded an independent AS run. For each of the 32 AS runs, 1,000 trajectories were obtained from each, the first 100 of which were discarded for equilibration purposes. In this way, 28,800 production AS trajectories were obtained.

The RC is constructed from candidate order parameters (OPs), which measure the geometric changes that characterize the transition of the system from reactant to product. A total of 99 OPs (Table S2) were analyzed in this transition state ensemble using LM (59, 60) to predict the best RC. LM (59, 61) predicts the best RC from these OPs and their combinations by fitting the model to \( p_B \), defined as the fraction of trajectories that commit to the product basin from a given configuration (59). The best predicted RC was validated using the committor probability (30) \( (p_B) \) test (Section S2.3, Table S3, and Fig. S2).

The best predicted RC was used to set up umbrella sampling windows for the calculation of the free energy profile for the reaction. Thirty-one windows were seeded from a reactive AS trajectory spanning the RC values between -3.5 to +2.5, describing the transition from reactant to product. In AMBER, in-house modifications were made to the sander utility source code to enable restraining the windows to custom-specified RCs that can be combinations of distance and angle values and differences between them. The windows were restrained to the desired RC values using this modified sander routine (rxncoor) with a force constant of 15 kcal/(mol-Å²) and run for 100 ps each. The weighted histogram analysis method was used to construct the free energy profile (62).

**Mutant simulations**

The D10N mutant system was created by choosing a QM/MM snapshot of the WT system in the reactant state and mutating Asp-10 to asparagine. The RC considered for this calculation involved the difference in the distance between the anomic carbon to the glycosidic oxygen and the -1 hydroxy-methyl oxygen to the anomeric carbon (the breaking and forming C-O bonds, respectively). In analogy with past studies (24, 25), the values of the RC for the reactant and product state (Fig. S5) were known to be -2.25 and 1.75, respectively. Restrained RC simulations using rxncoor were used to prepare seeds for 81 windows spanning these RC values, starting from the reactant state. Following a short equilibration of 2 ps in each window, production runs of 50 ps with a force constant of 100 kcal/(mol-Å²) were conducted, and the weighted histogram analysis method was used to construct the free energy profile (62). A more detailed description of the setup for mutant simulations is provided in Section S3.

**Author contributions**—V. S. B. data curation; V. S. B. formal analysis; V. S. B., B. C. K., J. S., and M. F. C. investigation; V. S. B. visualization; V. S. B., B. C. K., and M. F. C. methodology; V. S. B., B. C. K., and M. F. C. writing-original draft; V. S. B., B. C. K., J. S., G. T. B., and M. F. C. writing-review and editing; B. C. K., J. S., G. T. B., and M. F. C. supervision; G. T. B. conceptualization; M. F. C. funding acquisition; M. F. C. validation.

**Acknowledgments**—This work was authored in part by the Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the Department of Energy under Contract DE-AC36-08GO28308. We are also grateful for access to and use of National Renewable Energy Laboratory Computational Sciences resources (Peregrine and Eagle), supported by the Department of Energy Office of Energy Efficiency and Renewable Energy under Contract DE-AC36-08GO28308.

**References**

1. 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 CrossRef Medline

2. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-enzyme Active site database (CAZy): an expert resource for glycoscience. *Nucleic Acids Res.* **37**, D233–D238 Medline

3. Schüle, M., Kauppinnen, M. S., Lange, L., Lassen, S. F., Andersen, L. N., Klysner, S., and Nielsen, J. B. (1998) in *Enzyme Applications in Fiber Processing* (Eriksson, K.-E. L. and Cavaco-Paulo, A. eds) pp. 66–74, American Chemical Society, San Francisco, CA

4. Rasmussen, G., Mikkelsen, J. M., Schüle, M., Patkar, S. A., Hagen, F., Hjort, C. M., and Hastrup, S. (February 15, 1995) A cellulase preparation comprising an endoglucanase enzyme. EPO Patent 53137281

5. Davies, G. J., Tolley, S. P., Henrissat, B., Hjort, C., and Schüle, M. (1995) Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 Å resolution. *Biochemistry* **34**, 16210–16220 CrossRef Medline

6. Davies, G. J., Dodson, G., Moore, M. H., Tolley, S. P., Dauter, Z., Wilson, K. S., Rasmussen, G., and Schüle, M. (1996) Structure determination and refinement of the *Humicola insolens* endoglucanase V at 1.5 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 7–17 CrossRef Medline

7. Davies, G. J., Planas, A., and Rovira, C. (2012) Conformational analyses of the reaction coordinate of glycosidases. *Acc. Chem. Res.* **45**, 308–316 CrossRef Medline

8. 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 CrossRef Medline

9. Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase catalysis. *Curr. Opin. Chem. Biol.* **12**, 539–555 CrossRef Medline

10. Cosgrove, D. J. (2000) Loosening of plant cell walls by expansins. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 565–586 CrossRef Medline

11. van Straaten, K. E., Dijkstra, B. W., Vollmer, W., Thunnissen, A.-M. (2005) Crystal structure of MltA from *Escherichia coli* reveals a unique lytic transglycosylase fold. *J. Mol. Biol.* **352**, 1068–1080 CrossRef Medline

12. McQueen-Mason, S., Durachko, D. M., and Cosgrove, D. J. (1992) Two endogenous proteins that induce cell wall extension in plants. *Plant Cell* **4**, 1425–1433 CrossRef Medline

13. McQueen-Mason, S. J., and Cosgrove, D. J. (1995) Expansin mode of action on cell walls (analysis of wall hydrolysis, stress relaxation, and binding). *Plant Physiol.* **107**, 87–100 CrossRef Medline

14. Scheurerwater, E., Reid, C. W., and Clarke, A. J. (2008) Lytic transglycosylases: bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **40**, 586–591 CrossRef Medline

15. van Straaten, K. E., Barends, T. R. M., Dijkstra, B. W., Thunnissen, A.-M. (2007) Structure of *Escherichia coli* Lytic transglycosylase MltA with bound chitoheptose: implications for peptidoglycan binding and cleavage. *J. Biol. Chem.* **282**, 21197–21205 CrossRef Medline

16. Lee, M., Domínguez-Gil, T., Hesek, D., Mahasen, K. V., Lastochkin, E., Hermoso, J. A., and Mobashery, S. (2016) Turnover of bacterial cell wall by StbB3, a multidomain lytic transglycosylase of *Pseudomonas aeruginosa*. *ACS Chem. Biol.* **11**, 1525–1531 CrossRef Medline

17. Fibriansah, G., Gliubich, F. I., Thunnissen, A.-M. (2012) On the mechanism of peptidoglycan binding and cleavage by the endo-specific lytic transglycosylase fold. *Humicola insolens* GH45 catalytic mechanism
Humicola insolens GH45 catalytic mechanism

transglycosylase MtE from Escherichia coli. Biochemistry 51, 9164–9177 CrossRef Medline

18. Blackburn, N. T., and Clarke, A. J. (2001) Identification of four families of peptidoglycan lytic transglycosylases. J. Mol. Evol. 52, 78–84 CrossRef Medline

19. Davies, G., and Henrisat, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure 3, 853–859 CrossRef Medline

20. Sampredo, J., and Cosgrove, D. J. (2005) The expansin superfamily. Genome Biol. 6, 242 CrossRef Medline

21. Godoy, A. S., Pereira, C. S., Ramia, M. P., Silveira, R. L., Camilo, C. M., Kadowski, M. A., Lange, L., Busk, P. K., Nascimento, A. S., Skaf, M. S., and Polikarpov, I. (2018) Structure, computational and biochemical analysis of PcCel45A endoglucanase from Phanerochaete chrysosporium and catalytic mechanisms of GH45 subfamily C members. Sci. Rep. 8, 3678 CrossRef Medline

22. Nakamura, A., Ishida, T., Kusaka, K., Yamada, T., Fushinobu, S., Tanaka, I., Kaneko, S., Ohta, K., Tanaka, H., Inaka, K., Higuchi, Y., Niimura, N., Samejima, M., and Igarashi, K. (2015) “Newton’s cradle” proton relay with amide-imide acid tautomerization in inverting cellulase visualized by neutron crystallography. Sci. Adv. 1, e1500263 CrossRef Medline

23. Peters, B. (2016) Reaction coordinates and mechanistic hypothesis tests. Annu. Rev. Phys. Chem. 67, 669–690 CrossRef Medline

24. Knott, B. C., Haddad Momeni, M., Crowley, M. F., Mackenzie, L. F., Götz, A. W., Sandgren, M., Withers, S. G., Ståhlberg, J., and Beckham, G. T. (2014) The mechanism of cellulose hydrolysis by a two-step, retaining cellobiohydrolase elucidated by structural and transition path sampling studies. J. Am. Chem. Soc. 136, 321–329 CrossRef Medline

25. 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 hydrolases. Chem. Sci. 7, 5955–5968 CrossRef Medline

26. Wu, M., Bu, L., Vuong, T. V., Wilson, D. B., Crowley, M. F., Sandgren, M., Ståhlberg, J., Beckham, G. T., and Hansson, H. (2013) Loop motions important to product expulsion in the Thermofibrobacillus fusca glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. J. Biol. Chem. 288, 33107–33117 CrossRef Medline

27. Koshland, D. E. (1953) Stereochemistry and the mechanism of enzymatic reactions. Biol. Rev. 28, 416–436 CrossRef Medline

28. Ardévol, A., and Rovira, C. (2015) Reaction mechanisms in carbohydrate-active enzymes: glycoside hydrolases and glycosyltransferases: insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. J. Am. Chem. Soc. 137, 7528–7547 CrossRef Medline

29. Larsson, A. M., Bergfors, T., Dultz, E., Irwin, D. C., Roos, A., Driguez, H., Larsson, A. M., Bergfors, T., Dultz, E., Irwin, D. C., Roos, A., Driguez, H., Ido, Y., and Roos, A. (2009) Mechanism of cellulose hydrolysis by inverting GH8 endoglucanases: a QM/MM molecular dynamics study. J. Phys. Chem. B 113, 7331–7339 CrossRef Medline

30. Birnir, B., Akinwumi, K. O., and Naidoo, K. J. (2011) Molecular details from computational reaction dynamics for the cellobiohydrolase I glycosylation reaction. J. Am. Chem. Soc. 133, 19474–19482 CrossRef Medline

31. Biarnés, X., Ardèvol, A., Iglesias-Fernández, J., Planas, A., and Rovira, C. (2011) Catalytic itinerary in 1,3,1,4-β-glucanase unraveled by QM/MM metadynamics: charge is not yet fully developed at the oxocarbenium ion-like transition state. J. Am. Chem. Soc. 133, 20301–20309 CrossRef Medline

32. Thompson, A. J., Dabin, J., Iglesias-Fernández, J., Ardèvol, A., Dinev, Z., Williams, S. J., Bande, O., Siriwardena, A., Moreland, C., Hu, T.-C., Smith, D. K., Gilbert, H. J., Rovira, C., and Davies, G. J. (2012) The reaction coordinate of a bacterial GH47 α-mannosidase: a combined quantum mechanical and structural approach. Angew. Chem. Int. Ed. Engl. 51, 10997–11001 CrossRef Medline

33. Byun, B., Mahasenan, V. K., Dik, D. A., Marous, D. R., Speri, E., Kumarasiri, M., Fisher, J. F., Hermoso, J. A., and Mobashery, S. (2018) Mechanism of the Escherichia coli MtE lytic transglycosylase, the cell-wall-penetrating enzyme for type VI secretion system assembly. Sci. Rep. 8, 4110 CrossRef Medline

34. Vlasenko, E., Schülein, M., Cherry, J., and Xu, F. (2010) Substrate specificity of family 5, 6, 7, 9, 12, and 45 endoglucanases. Bioresour. Technol. 101, 2405–2411 CrossRef Medline

35. Quay, D. H., Cole, A. R., Cyar, A., Thalassinos, K., Williams, M. A., Bhakta, S., and Kepp, N. H. (2015) Structure of the stationary phase survival protein YucC from B. subtilis. BMC Struct. Biol. 15, 12 CrossRef Medline

36. McQueen-Mason, S., and Cosgrove, D. J. (1994) Disruption of hydrogen bonding between plant cell wall polysaccharides by proteins that induce wall expansion. Proc. Natl. Acad. Sci. U.S.A. 91, 6574–6578 CrossRef Medline

37. Whitney, S. E., Gilley, M. J., and McQueen-Mason, S. J. (2000) Probing expansin action using cellulose/hemicellulose composites. Plant J. 22, 327–334 CrossRef Medline

38. Yennawar, N. H., Li, L.-C., Dudzinski, D. M., Tabuchi, A., and Cosgrove, D. J. (2006) Crystal structure and activities of EXPBI (Zea m 1), a β-expansin and group-I pollen allergen from maize. Proc. Natl. Acad. Sci. U.S.A. 103, 14664–14671 CrossRef Medline

39. Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., and MacKerell, A. D., Jr. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and sidechain χ1 and χ2 dihedral angles. J. Chem. Theory Comput. 8, 3257–3273 CrossRef Medline

40. 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 CrossRef Medline

41. Anandakrishnan, R., Aguilar, B., and Onufriev, A. V. (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res. 40, W237–W2541 CrossRef Medline

42. Brooks, B. R., Brooks, C. L., 3rd, MacKerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cuí, Q., Dinner, A. R., Feig, M., et al. (2009) CHARMM: The biomolecular simulation program. J. Comput. Chem. 30, 1545–1614 CrossRef Medline

43. Hyninnen, A. P., and Crowley, M. F. (2014) New faster CHARMM molecular dynamics engine. J. Comput. Chem. 35, 406–413 CrossRef Medline

44. Ryckaert, J.-P., Cicotti, 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 CrossRef Medline

45. Crowley Michael, F., Williamson Mark, J., and Walker Ross, C. (2009) CHAMBER: comprehensive support for CHARMM force fields within the AMBER software. Int. J. Quantum Chem. 109, 3767–3772 CrossRef Medline

46. Case, D. A., Ceurdti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., et al. (2017) Amber 16 and AmberTools17. University of California, San Francisco.

47. de M. Seabra, G., Walker, R. C., Elstner, M., Case, D. A., and Roitberg, A. E. (2007) Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the Amber molecular dynamics package. J. Phys. Chem. B 111, 5655–5664 CrossRef Medline
55. Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S., and Seifert, G. (1998) Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. Phys. Rev. B 58, 7260–7268 CrossRef

56. Krüger, T., Elstner, M., Schiffels, P., and Frauenheim, T. (2005) Validation of the density-functional based tight-binding approximation method for the calculation of reaction energies and other data. J. Chem. Phys. 122, 114110 CrossRef Medline

57. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an N\!\!/H18528log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092 CrossRef

58. Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L. (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annu. Rev. Phys. Chem. 53, 291–318 CrossRef Medline

59. Peters, B., and Trout, B. L. (2006) Obtaining reaction coordinates by likelihood maximization. J. Chem. Phys. 125, 054108 CrossRef Medline

60. Peters, B., Beckham, G. T., and Trout, B. L. (2007) Extensions to the likelihood maximization approach for finding reaction coordinates. J. Chem. Phys. 127, 034109 CrossRef Medline

61. Peters, B. (2010) Recent advances in transition path sampling: accurate reaction coordinates, likelihood maximisation and diffusive barrier-crossing dynamics. Mol. Simul. 36, 1265–1281 CrossRef

62. Grossfield, A. WHAM: the weighted histogram analysis method. 2.0.9 Ed.

63. Georgelis, N., Yennawar, N. H., and Cosgrove, D. J. (2012) Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. Proc. Natl. Acad. Sci. U.S.A. 109, 14830–14835 CrossRef Medline