JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

The Mechanism of Cellulose Hydrolysis by a Two-Step, Retaining Cellobiohydrolase Elucidated by Structural and Transition Path Sampling Studies

Brandon C. Knott,[†] Majid Haddad Momeni,[‡] Michael F. Crowley,[§] Lloyd F. Mackenzie,[∥,#] Andreas W. Götz,[⊥] Mats Sandgren,[‡] Stephen G. Withers,[∥] Jerry Ståhlberg,*[,‡] and Gregg T. Beckham*[,†]

[†]National Bioenergy Center and [§]Biosciences Center, National Renewable Energy Laboratory, Golden, Colorado 80401, United States

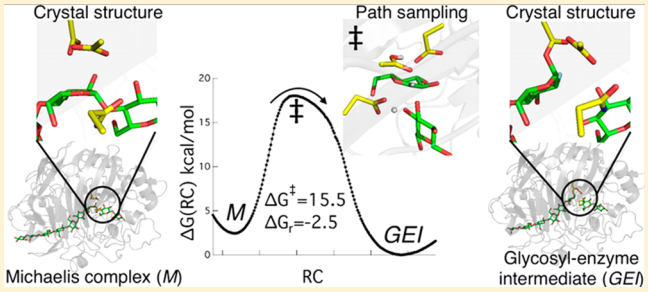
[‡]Department of Molecular Biology, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden

[∥]Department of Chemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1

[⊥]San Diego Supercomputer Center, University of California San Diego, La Jolla, California 92093, United States

Supporting Information

ABSTRACT: Glycoside hydrolases (GHs) cleave glycosidic linkages in carbohydrates, typically via inverting or retaining mechanisms, the latter of which proceeds via a two-step mechanism that includes formation of a glycosyl-enzyme intermediate. We present two new structures of the catalytic domain of *Hypocrea jecorina* GH Family 7 cellobiohydrolase Cel7A, namely a Michaelis complex with a full cellononaose ligand and a glycosyl-enzyme intermediate, that reveal details of the 'static' reaction coordinate. We also employ transition path sampling to determine the 'dynamic' reaction coordinate for the catalytic cycle. The glycosylation reaction coordinate contains components of forming and breaking bonds and a conformational change in the nucleophile. Deglycosylation proceeds via a product-assisted mechanism wherein the glycosylation product, cellobiose, positions a water molecule for nucleophilic attack on the anomeric carbon of the glycosyl-enzyme intermediate. In concert with previous structures, the present results reveal the complete hydrolytic reaction coordinate for this naturally and industrially important enzyme family.



INTRODUCTION

Carbohydrates are the most diverse set of biomolecules in Nature. As such, many classes of enzymes have evolved to assemble, modify, and deconstruct carbohydrates for a vast array of biological functions.[1] Glycoside hydrolase (GH) enzymes are responsible for hydrolytic cleavage of glycosidic linkages. GHs play vital roles in the global carbon cycle via turnover of polysaccharides, such as cellulose and chitin, in immune responses via trimming of peptidoglycans, and in protein folding, stability, and function via modulation of protein glycosylation. Koshland originally postulated that GHs employ either a 'retaining' or 'inverting' hydrolytic mechanism, where the name refers to the fate of the anomeric carbon stereochemistry.[2] Retaining GHs employ a two-step, double-displacement mechanism with an acid/base residue and a nucleophilic residue.[3−5] The first step, commonly referred to as glycosylation, proceeds via proton transfer from the acid/base residue to the glycosidic oxygen, coupled to attack at the anomeric carbon of the carbohydrate in the −1 binding site by the nucleophile to form a glycosyl-enzyme intermediate (GEI). Glycosylation is generally promoted via enzymatic distortion of the −1 glycosyl residue from the ground-state chair conformation.[6] The second step, typically termed deglycosylation, proceeds via an attacking water molecule at the anomeric carbon, which breaks the GEI bond and transfers a proton to the acid/base, thus restoring both the acid/base residue and the nucleophile to complete the catalytic cycle.

Many GH families utilize retaining mechanisms.[1] Of particular biological and industrial significance, GH Family 7 (GH7) cellulases employ retaining mechanisms to hydrolyze β-1,4 glycosidic bonds in cellulose, which are incredibly recalcitrant to uncatalyzed hydrolysis with significantly higher stability than DNA and peptide bonds.[7] GH7 cellulases are categorized as either cellobiohydrolases or endoglucanases, the former of which depolymerize single chains from crystalline cellulose and processively hydrolyze cellobiose units from the reducing ends, whereas the latter randomly cleave glycosidic linkages in more disordered regions of cellulose.[8,9] Structurally, cellobiohydrolases typically exhibit more closed tunnels, whereas endoglucanases have more open active site clefts.[10,11] GH7 cellobiohydrolases are often the major components of

Received: October 7, 2013
Published: December 9, 2013

   dx.doi.org/10.1021/ja410291u | J. Am. Chem. Soc. 2014, 136, 321−329

biomass-degrading enzyme cocktails in many fungi,[12] which are responsible for the majority of plant material degradation on Earth. As central as these enzymes are to fungal biomass degradation, they are also the cornerstone of modern industrial enzyme cocktails for biofuels processes.[9] As such, GH7 cellobiohydrolases are the target of intense structural, mechanistic, and engineering studies.[8,9,13,14] The first GH7 structure, from the ascomycete fungus *Hypocrea jecorina* (*Trichoderma reesei*), revealed a catalytic triad of highly conserved residues in the active site in an EXDXXE motif.[8] In *H. jecorina* Cel7A (*Hje*Cel7A), Glu212 has been identified as the nucleophile, and the acid/base is Glu217.[8,15] Asp214 hydrogen bonds to the nucleophile and strongly promotes catalysis,[15] but its mechanistic role has not been elucidated in detail. Subsequent crystallographic studies of Family 7 CBHs from *H. jecorina* and other organisms revealed binding characteristics of both the substrate[13,16] and product.[11,13,15,16]

Elucidating GH catalytic mechanisms is essential to understanding their activity. In the context of reaction rate theory, the term 'reaction coordinate' (RC) refers to a single variable that quantifies progress along a reaction pathway;[17] thus, knowing the RC for a given process amounts to knowing the detailed molecular mechanism. Crystallographic studies are of paramount importance in elucidating enzymatic RCs by capturing 'snapshots' of stable intermediates along the catalytic itinerary.[5,13,18−21] However, connecting the many geometrical changes between static configurations to the collective variables (CVs) that actually determine the fate of the chemical reaction (i.e., the RC) requires atomic-level dynamical information provided by computational modeling. Furthermore, computational modeling is essential for connecting structural information to free energy barriers and rates. For example, mixed quantum mechanics/molecular mechanics (QM/MM) simulations coupled with free energy methods have yielded potentials of mean force (PMF, or the free energy profile along a given coordinate), characterization of transition states, and full ring puckering itineraries for the hydrolytic action of retaining GHs.[22−25]

A major limitation of free energy methods such as metadynamics or umbrella sampling is that they require user-specified RCs (e.g., bond lengths) along which to sample and compute free energies. Thus, the RC is an *input* to the simulation, and all subsequent computational results depend upon the fitness of one's choice. Although RC selection may often seem trivial or intuitive, even seemingly simple processes such as ion pair dissociation[26] and alanine dipeptide isomerization[27] in water proceed along nonintuitive RCs. Thus, instead of sampling along assumed coordinates, one should first determine the RC with an unbiased approach such as transition path sampling (TPS).[28] TPS harvests an ensemble of reactive trajectories without the need to specify the RC *a priori* (a 'reactive trajectory' is a sequence of atomic configurations that connect a reactant configuration with a product configuration). From this ensemble of unbiased trajectories, one can systematically identify an optimal RC by employing methods such as likelihood maximization (LM).[29] LM selects the best model for the RC from a list of candidate CVs (functions of a configuration that compress many atomic details into a physically significant quantity) by fitting the model to the committor probability (p_B), which is the probability that a given atomic configuration will evolve in time to the product basin B (as opposed to reactant basin A).[29] The accuracy of a RC model can be quantified via the p_B histogram test, in which a histogram is constructed of p_B estimates from an ensemble of putative transition states; a good RC will give a histogram sharply peaked near 0.5.[29]

The catalytic cycle of GH7 cellulases includes several steps. After cellulose chain acquisition in the active site tunnel to form the Michaelis complex, the glycosylation step produces the GEI. Crystal structures of GH7 cellulases reveal two distinct binding modes for the product, which suggest that the cellobiose product pivots slightly outward from an 'unprimed GEI' to a 'primed GEI' binding mode, allowing a water molecule to approach the anomeric carbon of the GEI.[11] This is followed by the second catalytic step, deglycosylation. In the current study, we elucidate the full hydrolytic mechanism of the *Hje*Cel7A cellobiohydrolase via crystallography and TPS. We present new structures of the catalytic domain of *Hje*Cel7A that offer the first experimentally determined structural picture of the Michaelis complex for a GH7 cellobiohydrolase and the first GEI for any GH7 enzyme. To understand the molecular-level steps in the entire hydrolytic cycle, we employ QM/MM TPS to determine optimal RCs for both chemical steps, enabling accurate free energy and reaction rate calculations. Taken with previous structural reports, the present experimental and computational results elucidate the full 'static' and 'dynamic' RC of this vital class of enzymes.

METHODS

Preparation of Crystals and X-ray Structure Determination. Insoluble cellooligosaccharides (provided by Prof. Jürgen Puls) were prepared from water-soluble cellulose acetate with degree of substitution (DS) = 0.7 by enzymatic hydrolysis using a *Humicola insolens* endoglucanase (NovoNordisk SP613), aqueous size exclusion chromatography (SEC), and saponification.[30] Fraction C10, used herein, contained Glc_{10} as the major component and successively smaller amounts of shorter and longer cellooligosaccharides (Figure S1).

Protein preparation and crystallization of the catalytic domain of *Hje*Cel7A E217Q acid/base mutant has been described previously.[13,15] To obtain the Michaelis complex structure, C10 powder (~1 mg/mL) was added to the protein solution and suspended by sonication for ~1 min followed by overnight incubation with occasional sonication prior to crystallization setup. The structure of the GEI was obtained by first incubating 5 μM enzyme for 4 days with 80 μM 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-cellotrioside (DNP-2F-G3)[31] in 10 mM sodium MES buffer pH 6.0. The enzyme was then concentrated and crystallized in the presence of 1 mM fresh DNP-2F-G3 added to the crystallization drops for another 4 days prior to flash freezing of crystals in liquid nitrogen. X-ray diffraction data were collected from single crystals at MAX-lab beamline 711 (Michaelis complex) and ESRF beamline ID14−4 (glycosyl-enzyme intermediate). Further details on the indexing and processing as well as the crystallography data and refinement statistics of structure models can be found in the Supporting Information (SI).

Simulation Setup. The starting structure for all simulations is *Hje*Cel7A bound with cellononaose substrate (PDB code 8CEL).[13] Residues Asp214 and Glu217 are protonated. The protein and substrate are solvated in an equilibrated, cubic box ~80 × 80 × 80 Å^3 of TIP3P water molecules. Overall charge neutrality is achieved by adding Na^+ ions to solution. The total system size is 51 896 atoms. All simulations are performed in the isothermal−isobaric ensemble (NpT) at 300 K and 1.0 bar. The temperature is controlled by the Andersen thermostat. The pressure is controlled via an isotropic barostat. All QM/MM simulations employ a 1 fs time step, 8.0 Å nonbonded cutoff, periodic boundary conditions, and SHAKE-constrained hydrogen bonds (except in the active site).

The QM region for Step 1 consists of the side chains of the nucleophile Glu212, Asp214, acid/base Glu217, the −1 glucosyl unit, and the +1 glucosyl unit; a water molecule is added for Step 2. SCC-

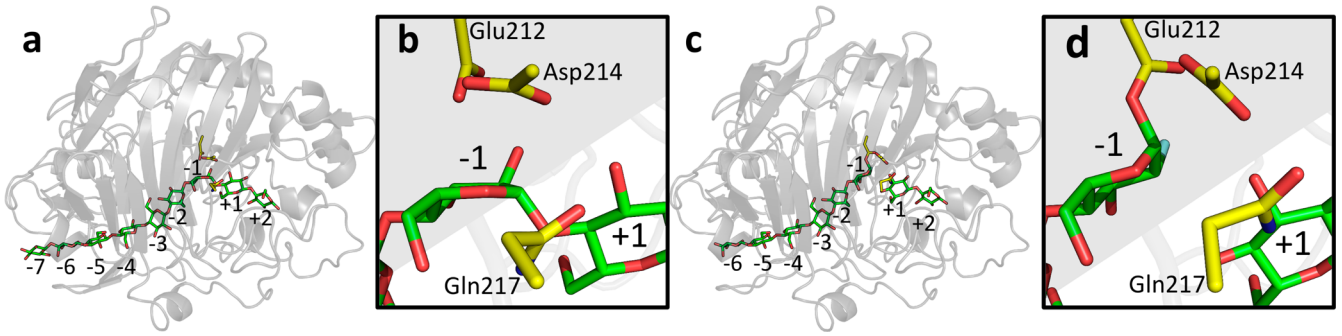


Figure 1. *Hje*Cel7A structures. (a) Overall crystal structure of the Michaelis complex of *Hje*Cel7A. (b) The active site of the Michaelis complex reveals a ^4E conformation of the glucopyranose ring in the −1 site. (c) Overall structure of the glycosyl-enzyme intermediate of *Hje*Cel7A. (d) The active site view of the glycosyl-enzyme intermediate. Comparison with the Michaelis complex in panel b suggests that the nucleophile Glu212 undergoes a dihedral rotation upon formation of the covalent bond with the anomeric carbon. The cellobiose product is in 'unprimed GEI' mode.

DFTB governs the QM interactions, and hydrogen link atoms are utilized at the boundary between the QM and MM regions.

Path Sampling. For both Steps 1 and 2, the initial reactive trajectory required for aimless shooting (AS) is obtained from the equilibrated structure by harmonically restraining key bonds to their assumed near-transition-state lengths. From this initial reactive trajectory, we harvest 21 000 trajectories in 10 runs and discard the first 100 from each run when processing with LM. For all AS simulations, Δt = 16 fs. Every trajectory consists of a forward and a backward segment, each 800 fs in length.

Free Energy Calculations. The PMFs for Steps 1 and 2 are calculated via equilibrium path sampling (EPS).[32] For Steps 1 and 2, EPS harvests 10 000 trajectories, each 50 fs in length, with 10 time-slices per trajectory. For Step 1, the distributions are computed in 25 overlapping windows; 20 windows are used for Step 2.

RESULTS

Structural Studies. A Michaelis complex structure of the *Hje*Cel7A acid/base mutant E217Q[15] (Figure 1a) bound with a fully intact cellononaose chain spanning sites −7 to +2 (1.45 Å resolution and R_{work}/R_{free} = 0.171/0.196; PDB code 4C4C) was obtained by cocrystallization of the catalytic domain of *Hje*Cel7A (the linker and carbohydrate-binding module were proteolytically cleaved prior to crystallization) with an insoluble cellooligosaccharide mixture (major component: Glc_{10}; distribution shown Figure S1). The glucose residue in site −1 adopts a ^4E envelope conformation with its anomeric carbon 3.4 Å from the nearest carboxylate oxygen of the catalytic nucleophile Glu212 and its C2 hydroxyl at hydrogen bonding distance from the other carboxylate oxygen of Glu212. The mutated acid/base residue Gln217 displays dual conformations; a major form of the side chain is hydrogen bonded to the −1/+1 glycosidic oxygen (Figure 1b), and an alternate form is rotated toward the +1 glucosyl residue (Figures S2a,c and S3a).

The second *Hje*Cel7A crystal structure is the GEI of the E217Q mutant enzyme with a cellohexaoside covalently linked to the catalytic nucleophile (1.32 Å resolution and R_{work}/R_{free} = 0.168/0.187; PDB code 4C4D), obtained by incubation of the protein with the mechanism-based suicide inhibitor 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-cellotrioside, DNP-2F-G3 (Figure 1c). The structure features cellohexaose in sites −6 to −1 and cellobiose in +1/+2, as there is clear electron density for glucopyranoside units in −6 to +2 (weak density in site −7 suggests that a glucose partially occupies this site also). The only visible conformation for the glucose in the −1 site is a $^4C_1$ chair that is α-linked to the nucleophile Glu212 (Figure 1d). The −1 glucosyl residue is presumably a 2-fluoro-glucoside moiety, although fluorine cannot be distinguished from a hydroxyl at this position. The density of the −1 unit is weaker than for the other sugars, thus it seems that the crystal contains two populations of structures, a major form with a covalently linked 2-fluoro-glucoside moiety in site −1, and a minor form wherein the −1 site is empty. This is further supported by the fact that all three catalytic residues (Glu212, Asp214, Gln217) exhibit dual conformations (Figures S2b,d and S3b). The cellobiose product has tilted slightly outward compared to the Michaelis complex but is still in the 'unprimed GEI' mode. Crystallographic statistics are provided in Table S1, and electron density is shown in Figure S2.

Elucidating the RC for Glycosylation (Step 1). To fully elucidate the 'dynamic' RC for *Hje*Cel7A, we utilize the AS[29] version of TPS to perform QM/MM molecular dynamics (MD) simulations of cellononaose hydrolysis by *Hje*Cel7A (see SI for full description). We harvest 21 000 trajectories for Step 1 (a representative trajectory is shown in Movie S1), 27% of which are reactive (i.e., 27% of the generated trajectories connect the reactant and product basins). Snapshots from a representative trajectory for the glycosylation step are shown in Figure 2a−c for the reactant, transition state, and product, respectively, and Figure 2d contains a schematic representation of the overall reaction. The reactant configuration (Figure 2a and left panel of Figure 2d) represents the Michaelis complex, in which the glycosidic bond is intact and Glu217 is protonated. The transition state (Figure 2b and middle panel of Figure 2d) shows the ring distortion as the anomeric carbon migrates toward the nucleophile, and the proton is transferred from Glu217 to the glycosidic oxygen. The product (Figure 2c and right panel of Figure 2d) shows the GEI in which the glycosidic bond has been broken and the proton transfer is complete. The newly cleaved cellobiose product sits in 'unprimed GEI' mode.

LM analyzes the AS trajectories and systematically determines the most accurate RC from candidate CVs.[29] We screen 149 candidate CVs as well as combinations of 2 and 3 of these (for a total of 551 449 possible RCs). The complete list of the candidate CVs for Steps 1 and 2 is available in Table S2. The candidate CVs are categorized as forming/breaking bonds, hydrogen bonds near the active site, −1 pyranose ring puckering, position and orientation of the catalytic water, orientation of the catalytic residues, protein loop interactions, etc.

LM analysis of the Step 1 AS data indicates that the best three-parameter RC involves: (1) the difference between the lengths of the forming and breaking bonds involving the

dx.doi.org/10.1021/ja410291u | J. Am. Chem. Soc. 2014, 136, 321−329

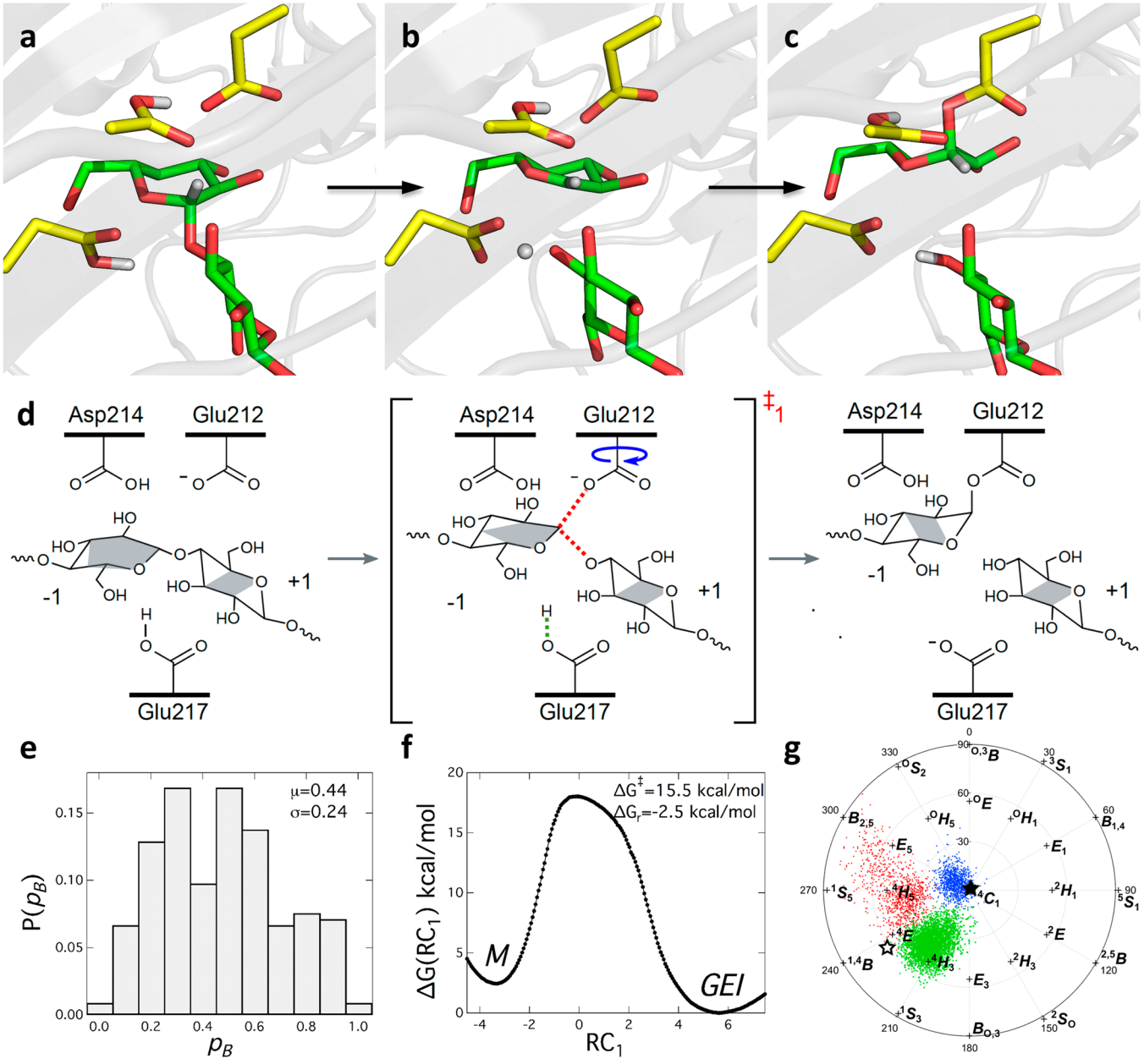


Figure 2. Glycosylation step results. (a) Snapshot of the reactant conformation from a representative AS trajectory (with substrate in green and catalytic residues in yellow) for the glycosylation step. The proton resides on the acid residue, Glu217, and the −1 glucopyranose ring is in the $^4H_5$ conformation. (b) A representative snapshot of the transition state. The −1 glucopyranose ring now adopts a $^4H_3$ conformation. (c) The product of the glycosylation reaction shows the glycosyl-enzyme intermediate with the −1 sugar in the $^4C_1$ conformation. (d) Schematic view of the overall glycosylation reaction with the collective variables identified by LM colored at the transition state. The best three-component RC identified by LM includes the forming/breaking bonds involving the anomeric carbon, the breaking bond between Glu217 and its proton, and the orientation of the nucleophile Glu212. (e) The p_B histogram for Step 1 demonstrates that the three-component RC is a valid model for the glycosylation step RC. (f) Reaction free energy and barrier for Step 1. *M* represents the Michaelis complex, and GEI represents the glycosyl-enzyme intermediate. (g) The Stoddart diagram overlaid with the distributions of the CP parameters for the glycosylation step catalytic itinerary (red dots represent reactant configurations, blue dots represent products, and green dots represent the transition-state ensemble). The open star represents the Michaelis complex of *Hje*Cel7A from the crystal structure, and the solid star indicates the glycosyl-enzyme intermediate crystal structure.

anomeric carbon; (2) the breaking bond distance between the Glu217 oxygen atom and its proton; and (3) the orientation of the nucleophile, Glu212, expressed as the angle between the anomeric carbon and the two carboxylate oxygen atoms of Glu212. In the enzyme−substrate complex, Glu212 hydrogen bonds with Asp214. As glycosylation proceeds, this hydrogen bond is broken, and the Glu212 oxygen formerly involved in this bond attacks the anomeric carbon. The inclusion of the nucleophile orientation in the RC demonstrates the importance of its conformational change to the fate of the reaction.

The p_B histogram for the Step 1 three-parameter RC is shown in Figure 2e. The p_B histograms of the best one-, two-, and three-parameter RCs are shown in Figure S5 (and coefficients in Table S3). Each histogram is constructed from estimating p_B at an ensemble of putative transition states as determined by AS/LM. The RC significantly improves as the

dx.doi.org/10.1021/ja410291u | J. Am. Chem. Soc. 2014, 136, 321−329

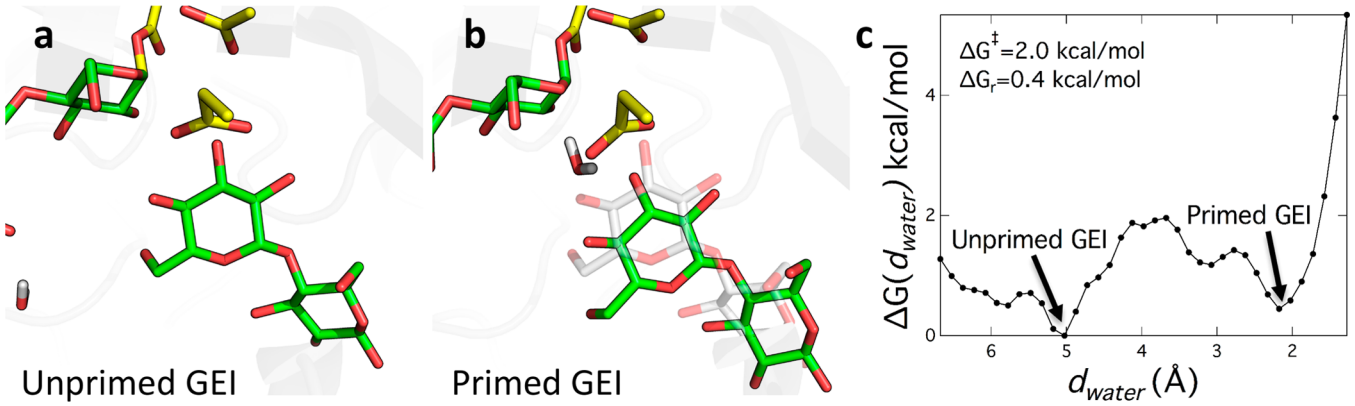


Figure 3. Cellobiose product transition from the unprimed to primed GEI and the approach of the nucleophilic water. (a) Snapshot of the *Hje*Cel7A GEI obtained from AS (with substrate in green and catalytic residues in yellow). A water molecule in the vicinity of the active site that will eventually become the nucleophilic water is also shown (lower left). (b) Snapshot of the 'primed GEI' from an umbrella sampling simulation that exerts a force to 'pull' a water molecule into the active site. The cellobiose product translates downward toward the binding tunnel exit in concert with this water movement. The primed GEI cellobiose product is shown in green, and the unprimed GEI cellobiose product is shown in transparent gray for reference. (c) The free energy as a function of the distance from the nucleophilic water to the active site suggests only a small barrier between these two states.

number of parameters increases, evidenced by the narrowing of the distribution of p_B estimates. For the best three-parameter RC, the mean of the distribution is 0.44, and the standard deviation is 0.24.

Enabled by an accurate RC, free energy and reaction rate calculations can be employed in a kinetically meaningful manner. The PMF presented in Figure 2f is computed along the optimal three-parameter RC with the reactant basin at a RC value near −3.5 and the GEI basin at an RC value of roughly 5.5 (in nondimensional units; see the SI for details). The free energy barrier height is 15.5 kcal/mol. The free energy difference between reactants and products is −2.5 kcal/mol downhill. The reaction rate constant is formulated by transition-state theory (TST) as $k = (k_BT/h)(\kappa)(\exp(-\Delta G^{\ddagger})/k_BT)$, where k_B is Boltzmann's constant, T is the absolute temperature, and h is Planck's constant. The transmission coefficient, κ,[33] represents the barrier recrossing term and is calculated by firing trajectories from the barrier (Figure S6). The transmission coefficient for Step 1 is 0.40, and as such, the predicted reaction rate constant is 10.8 s^{-1}.

The puckering of the −1 pyranose ring has long been regarded as an important element of enzymatic catalysis.[18,19,34] Enzymes stabilize distorted nonchair conformations at the −1 glucopyranoside ring that allow the catalytic nucleophile access to the anomeric carbon reaction center. Thus, nonchair −1 glycosides in crystal structures are often interpreted as pre-transition-state conformations indicative of the conformational itinerary for the reaction. To quantify this puckering itinerary, we present the distributions of the Cremer−Pople (CP)[35] parameters for the stable states and the transition-state ensembles for the whole catalytic cycle. The reaction proceeds from a half-chair ($^4H_5$) in the Michaelis complex to $^4H_3$ at the transition state, before relaxing to a stable chair conformation ($^4C_1$) in the product state (Figure 2g). The −1 sugar stays in this conformation until Step 2 commences.

Priming for Deglycosylation. After Step 1, there is not sufficient space for a water molecule to attack the anomeric carbon of the −1 glucopyranose ring to complete the catalytic cycle. However, various crystal structures of GH7 cellulases reveal two binding modes for the cellobiose product, which may correspond to an 'unprimed GEI'[11,13,14,16,36] and a 'primed GEI' mode[11,15] (Figure S7). The GEI structure and the Step 1 product conformation obtained in AS agree quite well with previously solved crystal structures in the unprimed GEI mode. In primed GEI mode, there is sufficient space for a water molecule to reside between the enzyme-bound substrate and the product. Thus, it is hypothesized that between Steps 1 and 2 of the catalytic cycle, the cellobiose product translates toward the tunnel outlet, creating space for a water molecule to move into the active site near Glu217 and serve as the Step 2 nucleophile.

To connect the Step 1 product to the Step 2 reactant, the nucleophilic water is pulled into the active site via a harmonic restraint on the distance between the water oxygen and the center of mass of the −1 glucopyranose anomeric carbon and the Glu217 C_δ (Figure 3a,b, Movie S2). As the water is pulled into the active site, the cellobiose product translates away from the active site toward the binding tunnel exit. Via umbrella sampling,[37] the PMF between the unprimed and primed GEI is calculated along this water distance coordinate (Figure 3c), revealing a 2.0 kcal/mol barrier, and essentially equal stability in the unprimed and primed GEI modes. This water molecule is added to the QM region for Step 2.

Elucidating the RC for Deglycosylation (Step 2). The computational procedure for Step 2 closely follows that of Step 1. AS harvests 21 000 trajectories for Step 2 (representative trajectory shown in Movie S3), with a 32% acceptance rate. Figure 4a−c shows snapshots of the reactant, transition state, and product, respectively, from a representative AS trajectory. LM reveals that the deglycosylation step proceeds via a product-assisted mechanism. The optimal three-parameter RC consists of: (1) the difference between the lengths of the forming and breaking bonds involving the anomeric carbon; (2) the difference between the lengths of the forming and breaking bonds involving the transferred proton; and (3) the sine of the dihedral angle formed by the cellobiose C3 hydroxyl and the nucleophilic water oxygen (dihedral C3−O3−HO3−O_{water}). Figure 4d shows an overall scheme of the deglycosylation step with the RC components highlighted in the middle panel. The histogram test for the best three-component RC is shown in Figure 4e (mean of 0.51 and

dx.doi.org/10.1021/ja410291u | J. Am. Chem. Soc. 2014, 136, 321−329

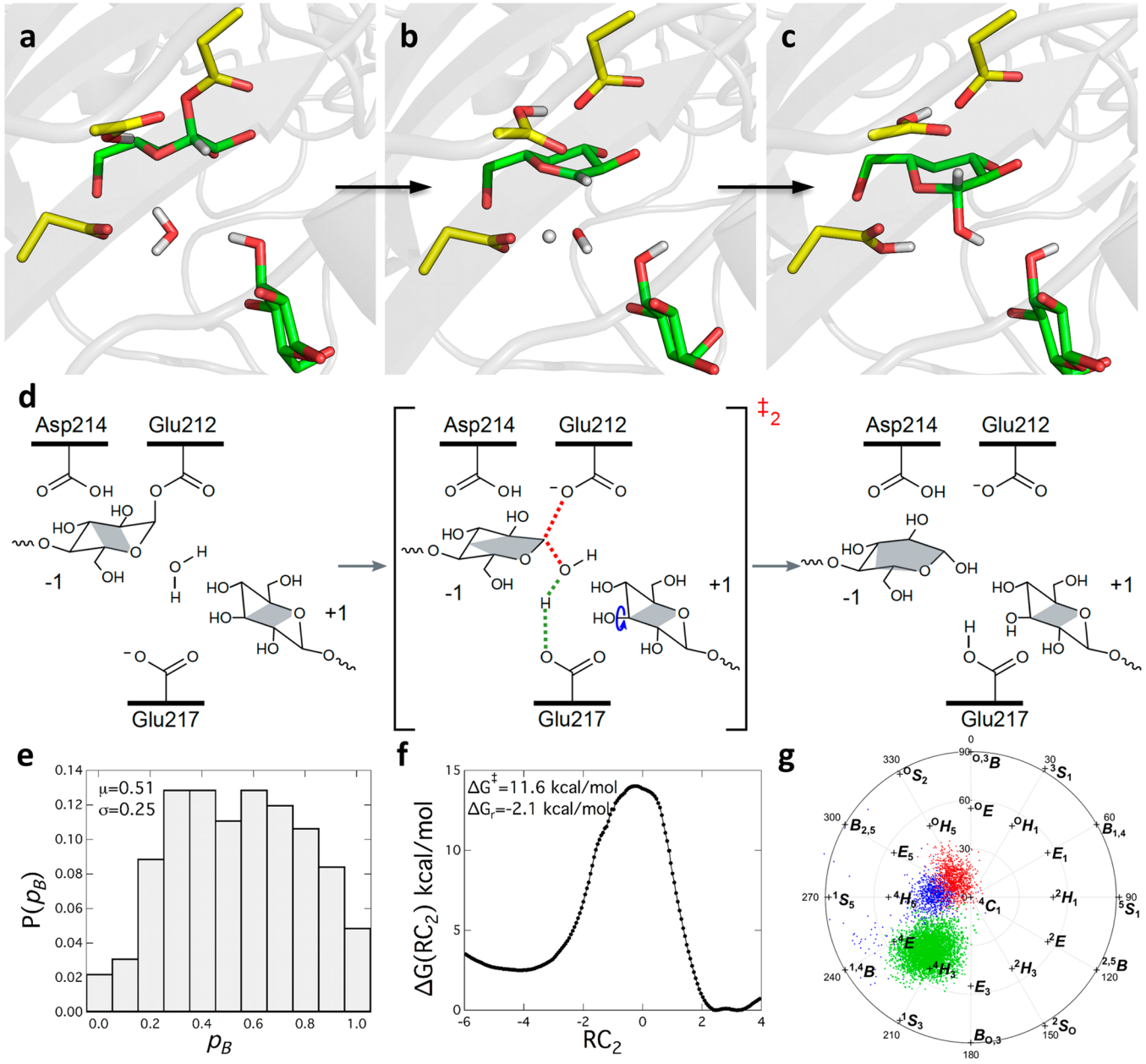


Figure 4. Deglycosylation step results. (a) Snapshot of the reactant conformation from a representative AS trajectory (with substrate in green and catalytic residues in yellow) for the deglycosylation step. The covalent glycosyl−enzyme bond is intact, and the cellobiose product is in primed GEI mode. The −1 glycosyl residue is in the stable $^4C_1$ conformation. (b) A representative snapshot of the transition state. Note the distorted $^4H_3$ conformation of the −1 sugar, as the nucleophilic water molecule is ripped apart. (c) A snapshot of the product in which the glycosyl-enzyme bond has been broken, and the catalytic residues have been regenerated. (d) Schematic view of the overall deglycosylation reaction with the collective variables identified by LM colored at the transition state. The best three-component RC identified by LM includes the forming/breaking bonds involving the anomeric carbon, the forming/breaking bonds involving the transferring proton, and the orientation of the C3 hydroxyl of the +1 sugar. (e) The p_B histogram for Step 2 demonstrates that the three-component RC is a valid model for the deglycosylation step RC. (f) Reaction free energy and barrier for Step 2. (g) The Stoddart diagram overlaid with the distributions of the CP parameters for the deglycosylation step catalytic itinerary (red dots represent reactant configurations, blue dots represent products, and green dots represent the transition-state ensemble).

standard deviation of 0.25), and the complete set of one-, two-, and three-component histograms is shown in Figure S5.

The Step 2 free energy profile shown in Figure 4f reveals an 11.6 kcal/mol barrier for deglycosylation. Similar to Step 1, the products are stabilized by −2.1 kcal/mol relative to reactants, suggesting an overall reaction free energy of −4.2 kcal/mol. Also similar to Step 1, the −1 ring pucker distribution for the Step 2 transition-state ensemble is centered near the $^4H_3$ half-chair, while the overall progression is essentially the reversal of Step 1. The transmission coefficient for Step 2 is 0.24, and TST gives a reaction rate constant of 5.29×10^3 s^{-1}.

■ DISCUSSION AND CONCLUSIONS

Capturing a GH structure with its natural substrate productively bound is exceptionally rare.[18,21,38−40] To our knowledge, the Michaelis complex presented here is the first retaining cellulase with its natural substrate spanning the active site in productive binding mode. The only *Hje*Cel7A crystal

dx.doi.org/10.1021/ja410291u | J. Am. Chem. Soc. 2014, 136, 321−329

structure to date with an occupied −1 binding site is an E212Q variant of the enzyme bound with two cellotetraose molecules occupying sites −7 to −4 and −2 to +2 (PDB code 5CEL).[13] However, two significant twists of the cellulose chain occur across binding sites −4/−3 and −3/−2 that prime the cellulose chain for catalytic hydrolysis.[13] The vacancy at the −3 site results in an unstrained configuration in sites −2 to +2, i.e., the oligomer is nonproductively bound. This same study presented an additional structure in which only the −1 site was vacant (PDB code 7CEL). From this structure, an unbroken cellononaose chain that spanned the −7 to +2 sites was constructed by modeling a near-boat conformation at the −1 site.[13] The cello-oligomer of this theoretical model (PDB code 8CEL) overlaps well with the Michaelis complex presented here, and the binding of the −1 glucosyl by six protein−carbohydrate hydrogen bonds is essentially as predicted.[13] Additionally, the structure of the GEI is the first for a member of GH7 and offers a snapshot along the catalytic itinerary directly after the glycosidic bond has been cleaved and before cellobiose has transitioned from unprimed to primed GEI mode.

Comparison of the two new structures shows that except for the large changes of the −1 sugar ring and a small shift in site +1, the other substrate residues overlap quite closely. The protein structures are practically identical except for a 30° rotation of the carboxyl group of the catalytic nucleophile Glu212. This feature of the 'static' RC corroborates the 'dynamic' RC for Step 1 as revealed by AS/LM, in which the key parameter is the orientation of Glu212 (in addition to the forming/breaking bonds).

The Step 1 simulation results for the −1 site ring puckering closely match the new structures (Figures 1b,d and 2g). In the Step 1 reactant basin, QM/MM simulations give a distribution of ring puckers that includes ^4E and $^4H_5$ configurations. The ^4E envelope configuration of the corresponding enzyme−substrate crystal structure falls within the puckering range of the QM/MM simulations. Postglycosylation, both QM/MM simulations and the crystal structure indicate an unstrained $^4C_1$ chair conformation at the −1 site. Crystal structures of retaining β-glycosidases may suggest that the ring puckering progression is $^{1,4}B/^1S_3 \rightarrow {}^4H_3 \rightarrow {}^4C_1$ for this class of enzymes.[41] Our QM/MM results follow this progression with the exception of the $^4E/^4H_5$ Michaelis complex, which is closer to what has been seen in crystal structures of β-mannosidases.[41] The $^4H_5$ conformation is also quite close to a local minimum on the Stoddart diagram.[42]

The existence of a small barrier between the unprimed and primed GEI (Figure 3c) is consistent with the binary product binding modes in GH7 crystal structures.[11] The slight shift of cellobiose between these two modes may be an essential (and previously unexamined) element of the GH retaining mechanism, as it allows the nucleophilic water access to the −1 anomeric carbon for deglycosylation, without requiring prior product removal. These distinct binding modes and the strength of their interactions with the nearby Cel7A residues are also important to the effect of inhibitors[11] and a key step in the processive cycle of GH7 CBHs. Representations of the retaining mechanism[2] often portray product removal as preceding the second chemical step, but our findings suggest that, if present, the cellobiose product may be involved in the deglycosylation reaction.

The RC identified by LM as the best of a given set of candidates is validated by the p_B histogram test. The two criteria for a good RC are that its histogram is centered at 0.5 and sharply peaked. To our knowledge, this is the most complex process for which histogram test results have been reported. Past applications have been to simpler processes such as nucleation,[33] small protein conformational changes (including folding[43] and isomerization[27]), polymorph transformation,[44] ion pair dissociation,[26] and proton/hydride transfer.[45] The spread in the p_B distributions presented in Figures 2e and 4e indicates that our RCs describe the catalytic steps well but that other coordinates are likely involved, albeit less directly.

Past simulation work on *Hje*Cel7A has computed free energy barrier heights for glycosylation, though always along an assumed, unverified RC. The barrier heights computed range from 14.1[46] to 32.6 kcal/mol.[25] For example, Barnett et al. computed a barrier height of 17.5 kcal/mol using QM/MM umbrella sampling simulations along the forming and breaking bonds that involve the anomeric carbon.[24] We find the difference between these two bond lengths to be the best one-parameter RC, though the best three-component RC constitutes a significant improvement (Figure S5a), which we use to calculate the reaction rate constant. We compute a Step 1 reaction rate constant of 10.8 s^{-1}. Utilizing high-speed atomic force microscopy, Igarashi et al. found the rate of processive cellobiose hydrolysis by *Hje*Cel7A on a crystalline cellulose surface to be 7.1 ± 3.9 s^{-1}.[47] If it is assumed that the cellulose chain processivity or product release is not rate-limiting, then our computed rate is within the error bars of the experimental measurement.

Past computational studies of retaining GHs have often either ignored Step 2[23−25] or assumed that product release precedes deglycosylation.[22] However, our AS results indicate that the product actually participates in the deglycosylation reaction. Prior to Step 2, the C3 hydroxyl of the +1 glucosyl ring forms a hydrogen bond to the catalytic water. As the nucleophilic water moves toward the −1 anomeric carbon, this hydrogen bond is broken resulting in a rotation of the C3 hydroxyl. This dynamical transition is a crucial reaction coordinate component for deglycosylation. On the basis of the product binding modes found in GH7 crystal structures (and the positions of water molecules therein), Ubhayasekera et al. discussed that product release might not happen until after deglycosylation.[11] The Step 2 RC we report from TPS simulations indicates that this is indeed the case.

To our knowledge, the reaction rate for deglycosylation has not been previously computed, probably because glycosylation is generally considered to be the rate-limiting step on natural substrates in these enzymes, as indicated by enzyme kinetics studies with artificial substrates.[15,48] Our comparison of the reaction rates for the two chemical steps is in agreement with Step 1 being rate-limiting. It has been possible to manipulate the reaction rates through the use of DNP-2-deoxy-2-fluoro-cellooligosaccharides as substrates, thus making deglycosylation rate-limiting and enabling the trapping of a covalent GEI of GH7 endoglucanases.[49] To achieve the same with cellobiohydrolase Cel7A, it was further necessary to use an acid/base-crippled mutant of the enzyme, which is described further in the Supporting Information.

GH7 cellulases are immensely important to both the natural and industrial world. Here we presented novel crystal structures of the most extensively studied GH7 cellobiohydrolase at critical stages of its catalytic itinerary. We also performed QM/MM path sampling simulations for the full hydrolytic itinerary, thus elucidating atomic and dynamical details not readily

available experimentally. The results reveal a product-assisted mechanism for the deglycosylation step as well as the first comparison of the rate constants for both steps of the catalytic cycle, with Step 1 being slower than Step 2 by nearly 3 orders of magnitude. Beyond glycoside hydrolases, we have demonstrated a powerful computational methodology to studying enzymatic reactions capable of elucidating structure−activity relationships. When verified computationally (via the histogram test) and validated experimentally (via reaction rate comparison), this methodology has the unique ability to produce, rather than intuit, accurate reaction mechanisms and kinetics.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Further details on purification, crystallization, and structure determination, including data collection and refinement statistics, electron density maps, and alternate conformations. Simulation parameters and protocols. Complete list of collective variables screened via likelihood maximization. Movies showing representative trajectories for the glycosylation and deglycosylation reactions as well as an illustrative trajectory for cellobiose product 'priming.' This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

Jerry.Stahlberg@slu.se
Gregg.Beckham@nrel.gov

Present Address

[#]Aquinox Pharmaceuticals Inc., Suite 430, 5600 Parkwood Way, Richmond, BC, Canada, V6 V 2M2

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

B.C.K. and G.T.B. thank the NREL Director's Fellowship Program for funding. G.T.B. also acknowledges funding from the Department of Energy BioEnergy Technologies Office and NREL Laboratory Directed Research and Development funds. M.F.C. and A.W.G. acknowledge the US Department of Energy SciDAC program for funding. J.S., M.S., and M.H.M. acknowledge financial support from the European Commission DG-12 (BIO4-CT96-0580), Swedish Foundation for Strategic Research, Swedish Structural Biology Network (SBNet), Swedish Natural Science Research Council (NFR), Swedish Council for Forestry and Agricultural Research (SJFR), Bo Rydin's Foundation for Scientific Research, and the Faculty for Natural Resources and Agriculture at the Swedish University of Agricultural Sciences through the research program 'Micro-DrivE.' S.G.W. and L.M. thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support. Computer time was provided by the Kraken cluster at the National Institute of Computational Sciences through the XSEDE allocation MCB09159. The authors thank Heather B. Mayes for assistance with figures, Prof. Jürgen Puls for providing the insoluble cellooligosaccharides, and Maria Boström, Sabah Mahdi, Dr. Mark Harris, Dr. Christina Divne, and Prof. T. Alwyn Jones for help in association with protein crystallization and structure determination. The authors are also grateful to Prof. John Brady and Dr. Udo Schnupf for sharing preliminary results from an independent study of *Hje*Cel7A.

■ REFERENCES

(1) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233.
(2) Koshland, D. E. *Biol. Rev.* **1953**, *28*, 416.
(3) Sulzenbacher, G.; Schülein, M.; Davies, G. J. *Biochemistry* **1997**, *36*, 5902.
(4) Mackenzie, L. F.; Sulzenbacher, G.; Divne, C.; Jones, T. A.; Woldike, H. F.; Schülein, M.; Withers, S. G.; Davies, G. J. *Biochem. J.* **1998**, *335*, 409.
(5) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. *Biochemistry* **1998**, *37*, 11707.
(6) Vocadlo, D. J.; Davies, G. J. *Curr. Opin. Chem. Biol.* **2008**, *12*, 539.
(7) Wolfenden, R.; Lu, X. D.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 6814.
(8) Divne, C.; Ståhlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. *Science* **1994**, *265*, 524.
(9) Chundawat, S. P. S.; Beckham, G. T.; Himmel, M. E.; Dale, B. E. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 121.
(10) Kleywegt, G. J.; Zou, J. Y.; Divne, C.; Davies, G. J.; Sinning, I.; Ståhlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1997**, *272*, 383.
(11) Ubhayasekera, W.; Munoz, I. G.; Vasella, A.; Ståhlberg, J.; Mowbray, S. L. *FEBS J.* **2005**, *272*, 1952.
(12) Martinez, D.; Berka, R. M.; Henrissat, B.; Saloheimo, M.; Arvas, M.; Baker, S. E.; Chapman, J.; Chertkov, O.; Coutinho, P. M.; Cullen, D.; Danchin, E. G. J.; Grigoriev, I. V.; Harris, P.; Jackson, M.; Kubicek, C. P.; Han, C. S.; Ho, I.; Larrondo, L. F.; de Leon, A. L.; Magnuson, J. K.; Merino, S.; Misra, M.; Nelson, B.; Putnam, N.; Robbertse, B.; Salamov, A. A.; Schmoll, M.; Terry, A.; Thayer, N.; Westerholm-Parvinen, A.; Schoch, C. L.; Yao, J.; Barbote, R.; Nelson, M. A.; Detter, C.; Bruce, D.; Kuske, C. R.; Xie, G.; Richardson, P.; Rokhsar, D. S.; Lucas, S. M.; Rubin, E. M.; Dunn-Coleman, N.; Ward, M.; Brettin, T. S. *Nat. Biotechnol.* **2008**, *26*, 553.
(13) Divne, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309.
(14) Kern, M.; McGeehan, J. E.; Streeter, S. D.; Martin, R. N. A.; Besser, K.; Elias, L.; Eborall, W.; Malyon, G. P.; Payne, C. M.; Himmel, M. E.; Schnorr, K.; Beckham, G. T.; Cragg, S. M.; Bruce, N. C.; McQueen-Mason, S. J. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10189.
(15) Ståhlberg, J.; Divne, C.; Koivula, A.; Piens, K.; Claeyssens, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1996**, *264*, 337.
(16) Parkkinen, T.; Koivula, A.; Vehmaanpera, J.; Rouvinen, J. *Protein Sci.* **2008**, *17*, 1383.
(17) Hanggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251.
(18) Tews, I.; Perrakis, A.; Oppenheim, A.; Dauter, Z.; Wilson, K. S.; Vorgias, C. E. *Nat. Struct. Biol.* **1996**, *3*, 638.
(19) Sulzenbacher, G.; Driguez, H.; Henrissat, B.; Schülein, M.; Davies, G. J. *Biochemistry* **1996**, *35*, 15280.
(20) Vocadlo, D. J.; Davies, G. J.; Laine, R.; Withers, S. G. *Nature* **2001**, *412*, 835.
(21) Guerin, D. M. A.; Lascombe, M. B.; Costabel, M.; Souchon, H.; Lamzin, V.; Beguin, P.; Alzari, P. M. *J. Mol. Biol.* **2002**, *316*, 1061.
(22) Liu, J. L.; Wang, X. M.; Xu, D. G. *J. Phys. Chem. B* **2010**, *114*, 1462.
(23) Petersen, L.; Ardevol, A.; Rovira, C.; Reilly, P. J. *J. Am. Chem. Soc.* **2010**, *132*, 8291.
(24) Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J. *J. Am. Chem. Soc.* **2011**, *133*, 19474.
(25) Yan, S. H.; Li, T.; Yao, L. S. *J. Phys. Chem. B* **2011**, *115*, 4982.
(26) Ballard, A. J.; Dellago, C. *J. Phys. Chem. B* **2012**, *116*, 13490.
(27) Ma, A.; Dinner, A. R. *J. Phys. Chem. B* **2005**, *109*, 6769.
(28) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291.
(29) Peters, B. *Mol. Simul.* **2010**, *36*, 1265.
(30) Lee, S. J.; Altaner, C.; Puls, J.; Saake, B. *Carbohydr. Polym.* **2003**, *54*, 353.

(31) Mackenzie, L. F.; Wang, Q. P.; Warren, R. A. J.; Withers, S. G. *J. Am. Chem. Soc.* **1998**, *120*, 5583.
(32) Peters, B.; Zimmermann, N. E. R.; Beckham, G. T.; Tester, J. W.; Trout, B. L. *J. Am. Chem. Soc.* **2008**, *130*, 17342.
(33) Pan, A. C.; Chandler, D. *J. Phys. Chem. B* **2004**, *108*, 19681.
(34) Ford, L. O.; Machin, P. A.; Phillips, D. C.; Tjian, R.; Johnson, L. N. *J. Mol. Biol.* **1974**, *88*, 349.
(35) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354.
(36) von Ossowski, I.; Stahlberg, J.; Koivula, A.; Piens, K.; Becker, D.; Boer, H.; Harle, R.; Harris, M.; Divne, C.; Mahdi, S.; Zhao, Y. X.; Driguez, H.; Claeyssens, M.; Sinnott, M. L.; Teeri, T. T. *J. Mol. Biol.* **2003**, *333*, 817.
(37) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(38) van Aalten, D. M. F.; Komander, D.; Synstad, B.; Gaseidnes, S.; Peter, M. G.; Eijsink, V. G. H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8979.
(39) Hovel, K.; Shallom, D.; Niefind, K.; Belakhov, V.; Shoham, G.; Baasov, T.; Shoham, Y.; Schomburg, D. *EMBO J.* **2003**, *22*, 4922.
(40) Hehemann, J. H.; Smyth, L.; Yadav, A.; Vocadlo, D. J.; Boraston, A. B. *J. Biol. Chem.* **2012**, *287*, 13985.
(41) Davies, G. J.; Planas, A.; Rovira, C. *Acc. Chem. Res.* **2012**, *45*, 308.
(42) Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129*, 10686.
(43) Rhee, Y. M.; Pande, V. S. *Chem. Phys.* **2006**, *323*, 66.
(44) Beckham, G. T.; Peters, B.; Starbuck, C.; Variankaval, N.; Trout, B. L. *J. Am. Chem. Soc.* **2007**, *129*, 4714.
(45) Quaytman, S. L.; Schwartz, S. D. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 12253.
(46) Li, J. H.; Du, L. K.; Wang, L. S. *J. Phys. Chem. B* **2010**, *114*, 15261.
(47) Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttila, M.; Ando, T.; Samejima, M. *Science* **2011**, *333*, 1279.
(48) Becker, D.; Braet, C.; Brumer, H.; Claeyssens, M.; Divne, C.; Fagerstrom, B. R.; Harris, M.; Jones, T. A.; Kleywegt, G. J.; Koivula, A.; Mahdi, S.; Piens, K.; Sinnott, M. L.; Ståhlberg, J.; Teeri, T. T.; Underwood, M.; Wohlfahrt, G. *Biochem. J.* **2001**, *356*, 19.
(49) Mackenzie, L. F.; Davies, G. J.; Schülein, M.; Withers, S. G. *Biochemistry* **1997**, *36*, 5893.