JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Glycoside Hydrolase Processivity Is Directly Related to Oligosaccharide Binding Free Energy

Christina M. Payne,*[,†,¶] Wei Jiang,*[,‡,¶] Michael R. Shirts,[§] Michael E. Himmel,[∥] Michael F. Crowley,[∥] and Gregg T. Beckham*[,⊥]

[†]Department of Chemical and Materials Engineering and Center for Computational Sciences, University of Kentucky, Lexington, Kentucky 40506, United States

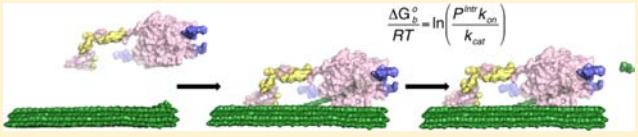
[‡]Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, Illinois 60439, United States

[§]Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22904, United States

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

Supporting Information

ABSTRACT: Many glycoside hydrolase (GH) enzymes act via a processive mechanism whereby an individual carbohydrate polymer chain is decrystallized and hydrolyzed along the chain without substrate dissociation. Despite considerable structural and biochemical studies, a molecular-level theory of processivity that relates directly to structural features of GH enzymes does not exist. Here, we hypothesize that the degree of processivity is directly linked to the ability of an enzyme to decrystallize a polymer chain from a crystal, quantified by the binding free energy of the enzyme to the cello-oligosaccharide. We develop a simple mathematical relationship formalizing this hypothesis to quantitatively relate the binding free energy to experimentally measurable kinetic parameters. We then calculate the absolute ligand binding free energy of cellulose chains to the biologically and industrially important GH Family 7 processive cellulases with free energy perturbation/replica-exchange molecular dynamics. Taken with previous observations, our results suggest that degree of processivity is directly correlated to the binding free energy of cello-oligosaccharide ligands to GH7s. The observed binding free energies also suggest candidate polymer morphologies susceptible to enzyme action when compared to the work required to decrystallize cellulose chains. We posit that the ligand binding free energy is a key parameter in comparing the activity and function of GHs and may offer a molecular-level basis toward a general theory of carbohydrate processivity in GHs and other enzymes able to process linear carbohydrate polymers, such as cellulose and chitin synthases.



INTRODUCTION

Polysaccharides exhibit a wide range of cellular functions, including energy storage, structure, metabolism, and molecular recognition across significant structural diversity.[1−11] As such, many classes of enzymes have evolved to modify carbohydrate polymers including glycosyltransferases, glycoside hydrolases (GHs), lyases, lytic polysaccharide monooxygenases, phosphorylases, and esterases, among others,[12−19] most of which are present in all kingdoms of life. Several carbohydrate polysaccharides, including cellulose, hemicellulose, and chitin, form the cell walls of plants, algae, and fungi, as well as the exoskeletons of arthropods and thus represent the largest source of organic carbon on Earth. To utilize these carbohydrate polymers for food, numerous families of GHs have evolved to decrystallize carbohydrate polymer chains from insoluble polysaccharide substrates and hydrolyze glycosidic bonds via processive mechanisms wherein multiple catalytic events occur before dissociation from the substrate. Microorganisms typically employ synergistic cocktails of GHs and other accessory enzymes for this purpose.[20] A primary driver for GH cocktail synergism is the presence of both processive enzymes and nonprocessive enzymes, which serve complementary functions; for cellulose, these enzymes are termed cellobiohydrolases and endoglucanases, respectively.

In the case of processive GH enzymes that act on cellulose and chitin, processivity is generally defined as the ability to decrystallize single polymer chains from carbohydrate polymer crystals and hydrolyze multiple glycosidic bonds before the dissociation of the substrate from the catalytic domain of an enzyme. Nonprocessive enzymes are thought to hydrolyze linkages randomly in more accessible, amorphous regions of polymer crystals. Multiple types of processivity measurements attempting to quantify processive ability have been developed to date.[21−25,58,59] Recent work has demonstrated that the dissociation rate of processive enzymes may be the rate-limiting step in hydrolysis of recalcitrant polysaccharide crystals.[26,58,59] For multiple systems, processive enzymes have been shown to provide the majority of the hydrolytic potential in conversion of recalcitrant polysaccharides. Because of this large hydrolytic

Received: July 19, 2013
Published: November 26, 2013

   dx.doi.org/10.1021/ja407287f | J. Am. Chem. Soc. 2013, 135, 18831−18839

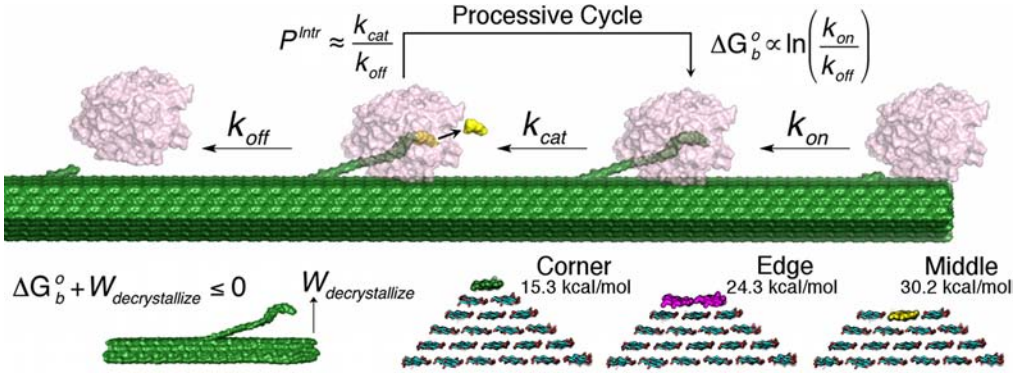


Figure 1. Illustration of the relationship of binding free energy, ΔG_b°, to intrinsic processivity of glycoside hydrolases, P^{Intr}, and work to decrystallize cellulose, $W_{decrystallize}$. The binding free energy is defined as the free energy change between a polysaccharide chain of n monosaccharide units (where n is the chain length required to saturate the GH binding sites) alone and the enzyme−substrate complex in the catalytically active complex. In this study, we approximate the binding free energy of the enzyme binding to the entire polysaccharide as the binding of a cello-oligomer to a GH active site. We define how this quantity is directly related to intrinsic processivity and controls the morphology dependence of cellulose attack by enzymes. At the top of the diagram, the entire processive glycoside hydrolase mechanism, including association, catalysis, and dissociation, is shown with corresponding rate coefficients, k_{on}, k_{cat}, and k_{off}, respectively.[58,59] At the bottom, we illustrate the hypothesized relationship of binding free energy to work required of the enzyme to decrystallize a given cellulose polymorph and morphology. As an example, we illustrate the corner, edge, and middle chain morphologies of cellulose Iβ. The corner chain, in green, has no adjacent hydrogen bonds. Edge chains, in magenta, have one set of adjacent hydrogen bonds to the neighboring chain. Middle chains, in yellow, are hydrogen-bonded to chains on either side. The work required to decrystallize cellononaose from each morphology from our previous study[60] is shown below each chain morphology label. We note that some Family 7 cellobiohydrolases exhibit a Family 1 carbohydrate-binding module and an *O*-glycosylated linker, which is not shown here as these domains are not considered in these studies.

potential, processive enzymes are a primary target of protein engineering strategies aimed at enhancing biomass conversion to sugars for subsequent conversion to fuels.[27−29] Understanding the extent and importance of processivity in GHs is important for design of optimally synergistic enzyme cocktails, to understand how insoluble polysaccharides are degraded mechanistically, and to develop structure−activity relationships for GH enzymes.

Many structural and biochemical studies[11,21,23,24,30−42] and several computational studies[43−48] have been conducted to understand the molecular-level basis for processivity. The processive ability of a GH has most often been attributed to structural features of the active site. It is thought that the presence of a tunnel or deep cleft is indicative of processivity and that lack thereof indicates that an enzyme is likely nonprocessive.[30,32,33,39,49−54] However, biochemical studies have demonstrated that processivity can be changed dramatically by point mutations or removal of single active site loops indicating this seemingly straightforward classification is not universal.[22,40,55]

We recently proposed three dynamical hallmarks of processivity. These were (1) the degree of ligand solvation, (2) the magnitude of average atomic fluctuation of the ligand as a function of binding site, and (3) the magnitude of overall fluctuations of key catalytic site residues. These proposed dynamical hallmarks of GH processivity were examined on four Family 18 chitinases, the two processive chitinases ChiA and ChiB from the bacterium *Serratia marcescens* and two nonprocessive chitinases, *S. marcescens* ChiC2 and the endochitinase from *Lactococcus lactis*. It was shown that these three dynamical properties were qualitatively related to the experimentally measured processivity of *S. marcescens* chitinases.[22,25,27,40,42,54] These metrics also affect the ligand binding free energy, ΔG_b°; however, a quantitative relationship has yet to be elucidated.

Here, we propose a more quantitative, generalizable description of processivity. We hypothesize that carbohydrate processivity is directly related to the ligand binding free energy of a polymer chain to a GH active site tunnel, illustrated in Figure 1.[56,57,60] In this study, we develop a simple, quantitative relationship between ligand binding free energy and intrinsic processivity. Direct measurement of the binding free energy is difficult, but we can estimate the binding free energy shown in Figure 1 of a GH to the polymer crystal by computing the binding free energy of a cellononaose ligand in solution to the GH also in solution, as described below. This hypothesis can also be extended to understand where a given enzyme productively binds to a crystalline substrate, as proposed earlier in studies of cellulose and chitin decrystallization.[60−62] Therein, we hypothesized that cellulases and chitinases act to depolymerize polymer chains from crystalline surfaces, and so the sum of the free energy of ligand binding and the work for chain decrystallization must be negative for thermodynamic favorability.

To test the hypothesis that binding free energy is directly related to processivity, we use free energy perturbation with replica-exchange molecular dynamics (FEP/λ-REMD)[63] to compute the absolute ligand binding free energy of a cello-oligomer ligand comprising nine glucose units to five GH Family 7 cellobiohydrolases shown in Table 1. To estimate the relative contribution of the celloheptamer 'substrate'-side binding (in the −7 to −1 binding sites, described below) to the overall binding free energy, as would be the case after hydrolysis and cellobiose product expulsion, we also calculate the free energy of binding a celloheptamer to *Hypocrea jecorina* (also known as *Trichoderma reesei*) Cel7A. Family 7 GHs have a common characteristic β-jelly roll fold with two primarily antiparallel β-sheets, which pack to form a curved β-sandwich. The cellobiohydrolases from this family exhibit long loops along the edge of the β-sandwich forming a nearly 50-Å long ligand-binding tunnel as illustrated in Figure 2. The mobility of

Table 1. Cellobiohydrolase Enzymes Examined in This Study with FEP/λ-REMD Simulations To Calculate the Binding Free Energy

enzyme[a]	PDB ID[b]
Heterobasidion irregulare Cel7A (*Hir*Cel7A)	2XSP[70]
Hypocrea jecorina Cel7A (*Hje*Cel7A)	8CEL[31,33]
Melanocarpus albomyces Cel7B (*Mal*Cel7B)	2RFW[68]
Phanerochaete chrysosporium Cel7D (*Pch*Cel7D)	1GPI, 1Z3W[c,35,67]
Trichoderma harzianum Cel7A (*Tha*Cel7A)	2YOK[69]

[a]We note that the letter after "Cel7" does not denote function, and that all 5 of these enzymes have either been confirmed or designated cellobiohydrolases. [b]PDB: Protein Data Bank. [c]We used the rotamer of Arg240 from 1Z3W because it is in the correct position to hydrogen bond to the ligand.

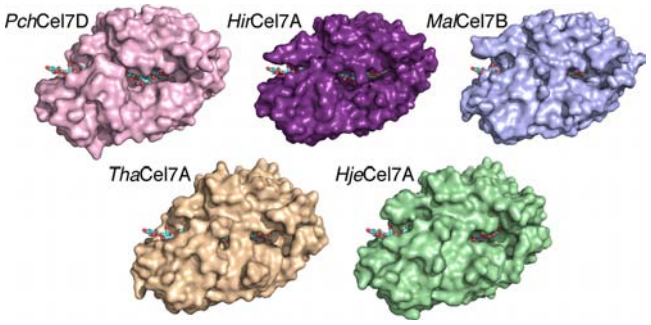


Figure 2. Fungal Family 7 GHs hypothesized to act processively on cellulose crystalline microfibrils. Abbreviations are defined in Table 1. Each cellulase exhibits the same characteristic fold along with attached loop domains forming the tunnel-shaped active site. The active sites (having at least 9 or 10 glucose-binding subsites) are nearly 50 Å long and encompass the cello-oligomer ligand, in cyan stick, to varying degrees. Of the five enzymes, *Pch*Cel7D has the most open of the active site tunnels, whereas *Hje*Cel7A has the most enclosed active site tunnel.

these loops has long been attributed to actuation of processive motion in GH7s, though they likely also allow endo-initiation by virtue of an open−close mechanism.[64] We have chosen to focus on Family 7 GHs as there is a wealth of structural and biochemical data available for this GH family. Furthermore, GH7s represent important protein engineering targets, and for a select number of enzymes studied here, processivity has either been directly measured[38,58,59,65,66] or suggested from crystal structures.[31,33,35,67−70]

METHODS

For each of the five enzymes, the system preparation and simulation methodology is described in detail in the Supporting Information. Briefly, the five enzymes listed in Table 1 were constructed using CHARMM[71] closely following previous work.[48,70] The bound ligand in each case was the cellononaose ligand from the 8CEL structure bound in the −7 to +2 binding subsites (Figure S1).[33] The *Hje*Cel7A celloheptaose binding case also used the ligand from the 8CEL structure with the glucopyranose moieties in the +1/+2 binding subsites removed. Protonation states of the catalytic residues were set to reflect the catalytically active conformation (Figure S1). Each enzyme−ligand complex was minimized, solvated with water and sodium ions, and minimized again. The solvated systems were heated to 300 K and density equilibrated. The equilibrated systems were simulated in NAMD[72] for 250 ns at 300 K. All simulations use the CHARMM27 force field with CMAP correction[73,74] for the proteins, the CHARMM35 carbohydrate force field[75−77] for the ligands, and the TIP3P force field[78,79] for water. From the molecular dynamics (MD) simulations, we calculated various properties likely related to processivity.[47] Additional details are provided in the Supporting Information.

To compute the ligand binding free energy, we used a protocol originally formulated by Deng and Roux[80] and further developed by Jiang et al.[63] wherein free energy perturbation is coupled to Hamiltonian replica-exchange molecular dynamics (FEP/λ-REMD) to improve Boltzmann sampling. Ligand binding free energy calculations of this type are accomplished through two separate free energy calculations, where (1) a bound ligand is decoupled from an enzyme and (2) a solvated ligand is decoupled from solution. The difference between the two values is the ligand binding free energy of the enzyme−ligand complex, illustrated in Figure 3. In this case, the "Enzyme" is a GH7 cellobiohydrolase *in solution* (i.e., not complexed on a crystal) and "Ligand" is a cellononaose or celloheptaose chain.

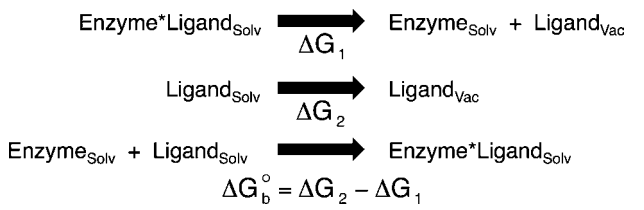


Figure 3. Thermodynamic cycle used to determine ligand binding free energy from FEP/λ-REMD. The subscripts "solv" and "vac" refer to the solvated and vacuum (or decoupled) systems, respectively.

For each free energy simulation, we express the potential energy in terms of four coupling parameters:

$$U(\lambda_{rep}, \lambda_{disp}, \lambda_{elec}, \lambda_{rstr},) = U_0 + U_{rep}\lambda_{rep} + U_{disp}\lambda_{disp} + U_{elec}\lambda_{elec} + U_{rstr}\lambda_{rstr} \quad (1)$$

U_0 is the potential energy of the system with the noninteracting ligand, λ_{rep}, λ_{disp}, λ_{elec}, $\lambda_{rstr} \in [0,1]$ are the thermodynamic coupling parameters, U_{rep} and U_{disp} are the shifted Weeks-Chandler-Anderson (WCA) repulsive and dispersive components of the Lennard-Jones potential,[81] U_{elec} is the electrostatic contribution, and U_{rstr} is the restraining potential. The ligand is decoupled from the enzyme in three steps, via the thermodynamic coupling parameters λ_{rep}, λ_{disp}, and λ_{elec}, controlling the nonbonded interaction of the molecule with its environment. Additionally, the restraints are taken into account via the parameter, λ_{rstr}, to control the translational and orientational restraints. The repulsive contribution, ΔG_{rep}, corresponds to the process

$$U(\lambda_{rep}=0, \lambda_{disp}=0, \lambda_{elec}=0, \lambda_{rstr}=1) \rightarrow U(\lambda_{rep}=1, \lambda_{disp}=0, \lambda_{elec}=0, \lambda_{rstr}=1) \quad (2)$$

The dispersive contribution, ΔG_{disp}, corresponds to process

$$U(\lambda_{rep}=1, \lambda_{disp}=0, \lambda_{elec}=0, \lambda_{rstr}=1) \rightarrow U(\lambda_{rep}=1, \lambda_{disp}=1, \lambda_{elec}=0, \lambda_{rstr}=1) \quad (3)$$

The electrostatic contribution, ΔG_{elec}, corresponds to the process

$$U(\lambda_{rep}=1, \lambda_{disp}=1, \lambda_{elec}=0, \lambda_{rstr}=1) \rightarrow U(\lambda_{rep}=1, \lambda_{disp}=1, \lambda_{elec}=1, \lambda_{rstr}=1) \quad (4)$$

The free energy ΔG_{rstr}, corresponds to the process

$$U(\lambda_{rep}=1, \lambda_{disp}=1, \lambda_{elec}=1, \lambda_{rstr}=1) \rightarrow U(\lambda_{rep}=1, \lambda_{disp}=1, \lambda_{elec}=1, \lambda_{rstr}=0) \quad (5)$$

The insertion of the ligand into the bulk phase is calculated with the same protocol but without any restraints on position or orientation.

dx.doi.org/10.1021/ja407287f | J. Am. Chem. Soc. 2013, 135, 18831−18839

For the FEP/λ-REMD simulations, we used the parallel/parallel replica exchange MD algorithm in NAMD. Each λ-staging FEP window is treated as a replica and the four-stage FEP simulation protocol can be implemented as a single parallel/parallel MPI job where the λ-exchange occurs along the entire alchemical reaction path. The replica-exchange algorithm follows the conventional Metropolis Monte Carlo exchange criterion with λ-swap moves:

$$\exp\left[\frac{1}{k_B T}(U(\lambda_m, \mathbf{r}_m) + U(\lambda_n, \mathbf{r}_n) - U(\lambda_m, \mathbf{r}_n) - U(\lambda_n, \mathbf{r}_m))\right] \geq \text{random}(0, 1) \quad (6)$$

where U denotes the potential energy of the underlying replica, and λ_m denotes the staging parameters. Parameters of the pair of λ values are exchanged (λ-swap) according to the criterion in eq 6. During the MD simulation, the pairs that attempt exchanges alternate every other Monte Carlo period.

Detailed simulation methodology has been provided in the Supporting Information. In short, for each enzyme−ligand complex (started from a 25 ns equilibrated snapshot) and the solvated cello-oligomers, 20 sequential 0.1 ns calculations were performed. The last 10 (totaling 1 ns) were averaged to determine the ligand binding free energy. The simulations use a set of 128 replicas (72 repulsive, 24 dispersive, and 32 electrostatic) with an exchange frequency of one every 100 steps (0.1 ps). The enzyme−ligand complexes included a positional restraint defined by the distance of the center of mass of the ligand to the center of mass of the enzyme. The free energies and statistical uncertainty of the repulsive, dispersive, and electrostatic contributions were determined by using the Multistate Bennett Acceptance Ratio (MBAR) on the energies collected during simulation.[82] To assess progress toward convergence, we monitored the time evolution of the 20 independent FEP calculations. For the last 1 ns of each calculation, we examined each of the 128 replicas, as illustrated in Figures S6−S9.

RESULTS

Here, we demonstrate that ligand binding free energy can be related to the kinetic parameter of processivity through the thermodynamics of chemical equilibrium. However, establishing the relationship first requires discussion of how processivity is quantified, specifically the difference between intrinsic and apparent processivity.

GH processivity is inherently difficult to measure as a result of substrate heterogeneity. From efforts to quantify processivity as a kinetic quantity, two distinct definitions of processivity arise: *apparent processivity*, which includes various environmental limitations such as substrate heterogeneity, and *intrinsic processivity*, the theoretical maximum. Most methods developed to quantify processivity estimate apparent processivity, defined as the average number of consecutive catalytic cycles performed per initiated processive run along the crystalline substrate.[25] Traditionally, apparent processivity in cellulases is approximated in two ways. One method compares the ratio of hydrolytic products, where a high percentage of dimeric products relative to monomers or trimers indicates a higher degree of processivity. This measurement capitalizes on the ability of processive cellulases to regularly produce dimeric soluble products. Alternatively, a second method compares the number of soluble and insoluble reducing ends created as products of hydrolysis. Generally, a higher fraction of soluble reducing ends is indicative of higher relative processive ability. Recently, Kurašin and Väljamäe developed an experimental technique using fluorescence-based detection of insoluble reducing groups to quantify apparent processivity in reducing-end specific cellulases.[58] Each of these methods comes with limitations, as discussed by Horn et al.[25] However, perhaps the most important, general limitation is that measurements of apparent processivity are dependent on the complex, often nonuniform substrate, making comparisons difficult.

Intrinsic processivity, P^{Intr}, or the average number of catalytic acts before enzyme dissociation regardless of substrate, can be described according to a probability of dissociation, P_D, as shown in eq 7.[25,83] For processive enzymes, the probability of dissocation is low ($\ll 1$) as a result of the protein−substrate interactions allowing consecutive hydrolytic cleavage. The probability of dissociation is related to the dissociation rate coefficient, k_{off}, and the catalytic coefficient, k_{cat}, by eq 8.[25] For processive cellulases, the catalytic rate coefficient is much larger than the dissociation rate coefficient. Thus, intrinsic processivity can be approximated by the ratio of the catalytic rate coefficient over the dissociation rate coefficient, shown in eq 9. Measurement of intrinsic processivity has the advantage of not requiring estimation of the number of initiated processive runs, which is difficult at best and currently impossible for nonreducing end exoinitiations.[25,58]

$$P^{\text{Intr}} = -\frac{1}{\ln(1 - P_D)} \approx \frac{1}{P_D} \quad (7)$$

$$P_D = \frac{k_{off}}{k_{off} + k_{cat}} \quad (8)$$

$$P^{\text{Intr}} \approx \frac{k_{cat}}{k_{off}} \because k_{cat} \gg k_{off} \quad (9)$$

The thermodynamic basis of the relationship of ligand binding free energy to enzyme processivity begins by assuming the enzyme−substrate association process reaches equilibrium, eq 10. The validity of this assumption has been established through kinetic models tested against literature data for cellulase action on insoluble substrates and through identification of the presteady-state regime in this biphasic system.[59,84]

$$\text{Enzyme} + \text{Substrate} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \text{Enzyme*Substrate} \quad (10)$$

At equilibrium, the dissociation constant, K_D, is defined as the concentrations of reactants, the enzyme and substrate independent of one another, over products, the enzyme−substrate complex, and is subsequently related to the association, k_{on}, and dissociation rate coefficient.

$$K_D = \frac{[\text{Enzyme}][\text{Substrate}]}{[\text{Enzyme*Substrate}]} = \frac{k_{off}}{k_{on}} \quad (11)$$

From the Eyring equation, we can relate the association and dissociation rate coefficients to the Gibbs free energy, or ligand binding free energy, ΔG_b°, calculated by FEP/λ-REMD as in eq 12.

$$K_D = \frac{k_{off}}{k_{on}} = \frac{(k_B T/h)\exp(-\Delta G_{off}^{\ddagger}/RT)}{(k_B T/h)\exp(-\Delta G_{on}^{\ddagger}/RT)} = \exp\left(-\left(\frac{\Delta G_{off}^{\ddagger} - \Delta G_{on}^{\ddagger}}{RT}\right)\right) = \exp\left(\frac{-\Delta G_b^0}{RT}\right) \quad (12)$$

Here, k_B is the Boltzmann constant, T is temperature, h is Planck's constant, R is the universal gas constant, and $\Delta G_{off}^{\ddagger}$ and $\Delta G_{on}^{\ddagger}$ are the Gibbs energy of activation of the off and on reaction, respectively. Finally, we reduce the relationship of

ligand binding free energy and association and dissociation rate coefficients, eq 13.

$$\frac{\Delta G_b^\circ}{RT} = \ln\left(\frac{k_{on}}{k_{off}}\right) \quad (13)$$

By combining eq 9 and 13, we relate ligand binding free energy to intrinsic processivity, eq 14.

$$\frac{\Delta G_b^\circ}{RT} = \ln\left(\frac{P^{Intr}k_{on}}{k_{cat}}\right) \quad (14)$$

Equation 14 is thus a general relationship connecting experimentally measurable kinetic variables to the ligand binding free energy, a quantity directly related to the enzyme structure. This approach in relating ligand binding free energy to intrinsic processivity is advantageous because of its relative simplicity, particularly as a first approach to quantitatively relating GH structure to function.

To understand the relationship of binding free energy to structural features of cellulases, we examine five GH7 cellobiohydrolases. Intrinsic processivity has been measured for *Hje*Cel7A and *Pch*Cel7D only. Kurašin and Väljamäe measured values of P^{Intr}, k_{cat}, and k_{off} for both enzymes acting on bacterial microcrystalline cellulose, finding *Hje*Cel7A was more processive on the crystalline substrate than *Pch*Cel7D by approximately a factor of 4.[58] Interestingly, k_{cat} for both enzymes was nearly identical suggesting that the dissociation rate coefficient is the primary limitation of efficient processive action. *Hir*Cel7A, *Mal*Cel7B, and *Tha*Cel7A have been classified as cellobiohydrolases given the long loops enclosing the active site tunnels, which are comparable to loops in GH7 cellobiohydrolases with known processive behavior.[68−70] Previous MD simulations suggest *Hir*Cel7A will have intermediate processive ability relative to *Hje*Cel7A and *Pch*Cel7D.[70]

In this study, we conducted additional MD simulations of *Mal*Cel7B and *Tha*Cel7A to gain insights into their likely processive ability relative to the other three enzymes. Figure 4A shows the root mean square fluctuation (RMSF) of each of the five GH7s from 250 ns MD simulations aligned according to a multiple sequence alignment (Figure S2) to compare appropriate regions with similar structural characteristics. In general, *Pch*Cel7D and *Hir*Cel7A exhibit more fluctuation in the loop regions around the active site tunnel. The RMSF of the ligand as a function of its binding site within the tunnel is a quantitative measure of ligand fluctuation in the MD simulations (Figure 4B), which is also related to the binding free energy.[47] The active sites of GH7 enzymes are numbered from +2 to −7, with catalysis taking place between the +1 and −1 binding sites. The cellobiose product is expelled from the +2/+1 side of the active site. Comparison of the ligand RMSF for the five GH7s reveals that the RMSF of the ligand surrounding the product side and central subsites in *Pch*Cel7D, *Hir*Cel7A, and *Mal*Cel7B is marginally higher than that of *Tha*Cel7A and *Hje*Cel7A; however, near the entrance to the active site tunnel (sites −6 and −7), the RMSF of the ligand is noticeably exaggerated for the putatively less processive enzymes. This behavior is a result of a structural characteristic in *Hir*Cel7A and *Mal*Cel7B, wherein the −7 subsite occasionally sandwiches the glucose monomer between two aromatic residues. As was recently observed for a related GH7 from *Limnoria quadripunctata*,[85] a flexible tyrosine residue over the active site intermittently forms half of the −7 subsite by switching between stable conformations. In *Pch*Cel7D, the −7 subsite in the enzyme is present on a single side only as a result of a loop deletion (Figure S2). Taken together, the dynamic behavior and open architecture of the active sites of putatively less processive enzymes suggests the active site has evolved for more efficient endo-initiation and dissociation rather than processivity, with greater flexibility allowing the enzyme to more effectively bind to amorphous cellulose regions. These results also enable a ranking of likely processive behavior, which can be quantitatively examined with ligand binding free energy calculations.

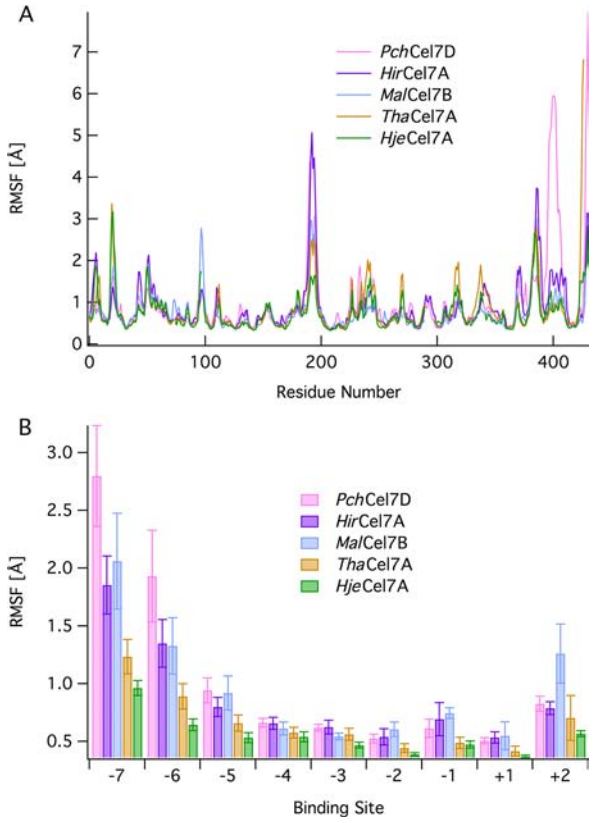


Figure 4. Results from the 250-ns MD simulations of the five GH7s, which have been compared in the text with binding free energies. MD simulation data for *Hje*Cel7A, *Hir*Cel7A, and *Pch*Cel7D are taken from a previous study.[70] (A) RMSF of the protein backbones aligned according to the multiple sequence alignment in Figure S2. (B) RMSF of the active site ligand, cellononaose, as a function of ligand binding subsite calculated using the heavy atoms only in each glucopyranose ring. Error bars were obtained through 2.5 ns block averaging (100 blocks total).

Examination of the structure of the active site topologies and the MD results in Figure 4 leads to the hypothesis that *Hir*Cel7A, *Mal*Cel7B, and *Tha*Cel7A exhibit intermediate processivity compared to *Pch*Cel7D and *Hje*Cel7A. A multiple sequence alignment of the five GH7s (Figure S2) further illustrates this structural variation as shortened loops and deletions in the vicinity of the enzyme active sites. In lieu of processivity measurements for these three GH7s and in light of the MD simulation results in Figure 4, we hypothesize that *Pch*Cel7D, being least processive, is followed in increasing order of processivity by *Hir*Cel7A, *Mal*Cel7B, *Tha*Cel7A, and *Hje*Cel7A, as shown from left to right in Figure 2. We note that we do not include the nonprocessive GH7, *H. jecorina* Cel7B,

dx.doi.org/10.1021/ja407287f | J. Am. Chem. Soc. 2013, 135, 18831−18839

Table 2. Absolute Binding Free Energies of the Five Family 7 GHs and the Solvation Free Energy of the Cello-Oligomer[a]

	ΔG_b°	ΔG_{rep}	ΔG_{disp}	ΔG_{elec}	ΔG_{rstr}
$Glucose_7$		91.06 ± 0.33	−83.68 ± 0.12	−89.33 ± 0.27	
$Glucose_9$		104.13 ± 0.89	−92.19 ± 0.83	−111.47 ± 0.54	
*Pch*Cel7D	−15.19 ± 2.69	167.42 ± 1.41	−152.64 ± 0.86	−129.70 ± 1.46	0.21
*Hir*Cel7A	−20.55 ± 1.50	170.42 ± 1.23	−166.33 ± 0.81	−124.52 ± 0.88	0.35
*Mal*Cel7B	−22.51 ± 1.45	160.95 ± 1.42	−160.75 ± 0.69	−122.38 ± 1.00	0.15
*Tha*Cel7A	−23.89 ± 1.98	165.41 ± 1.46	−158.76 ± 0.59	−130.33 ± 0.49	0.26
*Hje*Cel7A	−27.83 ± 2.10	170.97 ± 1.20	−165.73 ± 0.65	−132.96 ± 0.63	0.36
*Hje*Cel7A Heptamer	−16.72 ± 2.32	128.35 ± 1.87	−123.88 ± 0.31	−103.27 ± 1.24	0.13

[a]Repulsive, dispersive, electrostatic, and restraining contributions sum to the total. All energies are given units of kcal/mol.

which exhibits a cleft instead of a tunnel, in this study as its structural resolution (3.6 Å) is much lower than those in Table 1, and a threaded ligand is relatively less stable during MD simulations and in free energy calculations.[48]

Comparison of the binding free energies calculated by FEP/λ-REMD to the known intrinsic processivities of *Pch*Cel7D and *Hje*Cel7A suggests the processivities of *Hir*Cel7A, *Mal*Cel7B, and *Tha*Cel7A are indeed intermediate. Table 2 lists the binding free energies of the five GHs examined here along with the repulsive, dispersive, electrostatic, and restraint contributions. The free energy of solvation of cellononaose is also given. Of the five GH7s, *Pch*Cel7D has the least favorable binding free energy at −15.2 ± 2.7 kcal/mol. In order of increasing binding free energy favorability, *Hir*Cel7A, *Mal*Cel7B, *Tha*CelA, and *Hje*Cel7A have binding free energies of −20.6 ± 1.5, −22.5 ± 1.5, −23.9 ± 2.0, and −27.8 ± 2.1 kcal/mol, respectively. Table 2 also lists the binding free energy of binding celloheptatose to *Hje*Cel7A, −16.72 ± 2.32 kcal/mol, along with the solvation free energy of celloheptaose.

Error associated with each contribution to the binding free energy (i.e., repulsive, dispersive, and electrostatic) represents 1 standard deviation over the last 1 ns. The error of the binding free energy was obtained by taking the square root of the sum of the squared standard deviations of the relevant cello-oligomer solvation free energy and the free energy to decouple the ligand from the enzyme. Both the binding free energy and the binding free energy as a function of time are given in the Supporting Information, Figures S3 and S4. Whereas the binding free energies qualitatively adhere to the logarithmic relationship described in eq 14, quantitative validation is not possible without additional experimental processivity and rate data. Today, experimental quantities for k_{on}, k_{off}, and k_{cat} have been measured in a self-consistent manner only for *Hje*Cel7A.[59] For *Pch*Cel7A, both k_{off} and k_{cat} have been measured along with the intrinsic processivity discussed above,[58] though k_{on} is unavailable.

DISCUSSION

In this study, we outline a straightforward approach to relate processivity to GH structure through binding free energy calculations. The mathematical relationship described by eq 14 is general to any processive enzyme where the kinetics reach steady state and $k_{cat} \gg k_{off}$. We foresee that this description of GH processivity may eventually find utility as a predictor of processive action; however, extensive experimental measurements are still required to validate the relationship. Unfortunately, quantifying intrinsic processivity and the associated kinetic measurements on insoluble crystalline substrates in consistent manner is extremely difficult; biphasic cellulase kinetics tend to obscure k_{cat}, and the effects of substrate heterogeneity and multiple binding modes contribute uncertainty in determining k_{off}. It is this experimental difficulty that contributes to both the complexity of validation and a general lack of available data by which to do so. All of which underscores the critical need for a theoretical means of describing glycoside hydrolase processivity.

Our computational investigation of cello-oligomer binding to cellobiohydrolases suggests binding free energy is directly related to intrinsic processivity. Using a robust free energy method, we calculated the free energy of binding cellononaose to five GH7s. Comparing calculated binding free energies to the known intrinsic processivity values for *Hje*Cel7A and *Pch*Cel7D and estimated processivity from MD for *Mal*Cel7B, *Hir*Cel7A, and *Tha*Cel7A, a general trend emerges suggesting more processive enzymes possess greater ability to associate tightly with the substrate. This result aligns closely with the supposition that active site architecture is a primary determinant of processive ability. MD simulations further suggest the active site architecture, namely active site tunnel loops and key residues at the tunnel entrances, contributes to association with the insoluble substrate and gives rise to differences in the extent of GH7 processivity observed in the enzymes examined here.

Calculating the free energy of binding celloheptaose to *Hje*Cel7A provides an estimate of the contribution of substrate (−7 to −1 sites) and the product site (+1/+2) binding to the overall affinity of the enzyme for a cellulose chain. As illustrated in Figure 1, GH7s must bind the entire cellononaose chain as part of the processive mechanism, with the possible exception being upon initial attack. During processive action, the cellobiose product will be hydrolytically cleaved and subsequently released from the active site. The free energy of binding cellononaose and celloheptaose to *Hje*Cel7A was found to be −27.83 ± 2.10 and −16.72 ± 2.32 kcal/mol, respectively. The difference of the two values, −11.11 ± 3.13 kcal/mol, indicates the 'product' sites (+1/+2) bind the ligand more tightly than the 'substrate' sites (−7 to −1), normalized by the number of binding sites, which likely contributes to product inhibition in cellobiohydrolases.[38,86−94] This value for binding uncleaved cellobiose on the product side of the active site, obtained by difference, is in excellent agreement with a previous determination of binding cellobiose to *Hje*Cel7A in the +1/+2 binding sites (−11.2 ± 0.6 kcal/mol).[94] Overall, this implies the entirety of the GH7 binding site plays an active role in decrystallizing the insoluble cellulose substrate, leading in part the second hypothesis illustrated in Figure 1 relating binding free energy to polysaccharide polymorph and morphology.

Processive GHs are responsible for deconstructing polysaccharide crystals of varying polymorphs and morphologies. Various substrate pretreatment strategies generating non-native

cellulose polymorphs have been shown to enhance deconstruction by interrupting the hydrogen bond network.[95,96] However, little is known about the preference of an enzyme for the various morphologies given the inherent experimental difficulty of characterizing substrate heterogeneity at the molecular level. We have previously hypothesized that the sum of the binding free energy of a given enzyme (negative, thermodynamically favorable) and the work required of the enzyme to decrystallize its ligand (positive, thermodynamically unfavorable) must be sufficiently thermodynamically favorable as to not contribute an overall positive free energy to the processive catalytic cycle. In this previous study, we used simulation to quantify the work to decrystallize a variety of chains from different cellulose polymorphs as a first step toward testing our hypothesis.[60] Focusing here on cellulose Iβ, the most naturally abundant of polymorphs, we recount the work required to decrystallize a corner, edge, and middle chain from the surface of a crystalline microfibril (Figure 1). The result of our previous study, as expected, was that a middle chain was the most difficult to decrystallize as a result of the need to break hydrogen bonds with the crystal on either side of the chain, requiring 30.2 kcal/mol per cellononaose. Similarly, the edge chain and corner chain followed in ease of decrystallization at 24.3 kcal/mol per cellononaose and 15.3 kcal/mol per cellononaose, respectively. The work required to decrystallize a cellulose chain from the surface of a microfibril, calculated using umbrella sampling, is the same order of magnitude as the calculated binding free energies. Taking this a step further, comparison of the amount of work to decrystallize a cellulose chain from a particular morphology to the binding free energies suggests these enzymes have morphological preferences; namely, a middle chain is precluded as a point of attack by processive cellobiohydrolases as the processive catalytic cycle would no longer be energetically downhill. Only the corner chain morphology is universally overall downhill, suggesting these GH7 cellobiohydrolases are all capable of processively decrystallizing corner chain morphologies.

Finally, recent experimental results suggest the relationship we derive here to connect binding free energy to processivity (eq 14) may also describe elements of the enzymatic activity differences observed on different cellulose polymorphs. In a biochemical study of cellulase binding and activity on two cellulose polymorphs Iβ and III_I, Gao et al. concluded that reduced binding affinity, measured by percent of enzyme bound, resulted in enhanced hydrolytic activity.[96] The authors developed a kinetic model including a processive decrystallization rate, k_{slide}, to explain seemingly anomalous experimental findings. Gao et al. observed that increasing k_{slide} without increasing binding affinity increased catalytic efficiency. Furthermore, increasing k_{slide} alongside reduced association (k_{on}) resulted in increased hydrolytic capacity and decreased binding affinity, which was achieved during assays of cellulase behavior on the more easily decrystallized cellulose III_I polymorph.[60]

Here, we suggest k_{slide} is equivalent to apparent processivity encompassing substrate heterogeneity, nonproductive binding, and variation in hydrophilicity of the polymorph. The experimentally observed behavior of cellulase activity on cellulose III_I qualitatively corresponds to the relationship in eq 14 for the intrinsic processivity. As an addendum to Gao et al.'s conclusions, we posit that the experimental observation of increased catalytic activity and processivity accompanied by decreased association and binding affinity represents polymorphic preference as apparent processivity approaches intrinsic processivity. In other words, as the need for binding affinity decreases as it does in the in case of cellulose III_I, apparent processivity approaches intrinsic processivity, yet to maintain a constant binding free energy, k_{cat} must also increase, k_{on} must decrease, or some combination of the two. We suggest that the findings of Gao et al. for increased activity despite reduced binding affinity on cellulose III_I is well described by the relationship of binding free energy to processivity.

■ CONCLUSIONS

In this study, we relate the binding free energy, which quantitatively includes topological and dynamical contributions from the enzyme active site, to an experimentally measurable quantity, intrinsic processivity. Using an accelerated sampling free energy methodology, we determined the binding free energy of five GH7 enzymes for which we know or can estimate intrinsic processivity. We find that thermodynamic favorability of the binding free energy increases with intrinsic processivity, in agreement with the derived relationship. MD simulations also provide insight into molecular-level active site dynamics contributing to the calculated binding free energy suggesting less processive enzymes exhibit a greater degree of flexibility associated with more open active site architectures. The topology and dynamics of the two binding subsites that form the entrance of the active site tunnel appear to be a primary determinant of tight binding, and ultimately processive ability. Well-established architecture of these two entrance binding subsites enables the ligand to maintain stability and form hydrogen bonds with the surrounding protein. The experimental difficulty associated with processivity measurements for GHs has led to a dearth of information for direct comparison. It is exactly this experimental difficulty that emphasizes the critical need to develop such a relationship and demonstrates the utility of theory and simulation, as effectively quantifying and understanding processivity in GHs is key to cost-effective, enzymatic biomass conversion. This study presents the framework for a major component of the thermodynamics of processivity and sheds light on structure-based trends and morphological preference of GH7s that cannot currently be addressed directly by experimental means.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional computational protocol of the MD and free energy calculations; cartoon and stick representation of *Hje*Cel7A simulation setup; GH7 multiple sequence alignment; graphical representation of the free energy results; MD simulation results; active site dynamics; FEP/λ-REMD convergence analysis example for *Mal*Cel7B. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

Christy.Payne@uky.edu
wjiang@alcf.anl.gov
Gregg.Beckham@nrel.gov

Author Contributions

¶C.M.P. and W.J. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Department of Energy (DOE) BioEnergy Technologies Office (to C.M.P., M.E.H., M.F.C., and G.T.B.). Computational resources were provided by the Argonne Leadership Computing Facility at Argonne National Laboratory supported by the DOE Office of Science under contract DE-AC02-06CH11357, the National Institute for Computational Sciences Kraken cluster under the National Science Foundation Extreme Science and Engineering Discovery Environment (XSEDE) Grant MCB090159, and the National Renewable Energy Laboratory Computational Sciences Center supported by the DOE Office of Energy Efficiency and Renewable Energy under contract DE-AC36-08GO28308. The *Hje*Cel7A free energy calculations and method development were performed under an INCITE allocation. *Pch*Cel7D, *HirCel7A*, *Mal*Cel7B, and *Tha*Cel7A free energy calculations were performed with the XSEDE allocation.

■ REFERENCES

(1) Carpita, N. C. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1996**, *47*, 445−476.
(2) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683−720.
(3) Delmer, D. P. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 245−276.
(4) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370−2376.
(5) Ball, S. G.; Morell, M. K. *Annu. Rev. Plant Biol.* **2003**, *54*, 207−233.
(6) Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredez, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H. *Science* **2004**, *306*, 2206−2211.
(7) Smith, A. M.; Zeeman, S. C.; Smith, S. M. In *Annual Review of Plant Biology*; Annual Reviews: Palo Alto, CA, 2005; Vol. *56*, pp 73−98.
(8) Seeberger, P. H. *Nature* **2005**, *437*, 1239−1239.
(9) Cosgrove, D. J. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 850−861.
(10) Finkelstein, J. *Nature* **2007**, *446*, 999.
(11) Stern, R.; Jedrzejas, M. J. *Chem. Rev.* **2008**, *108*, 5061−5085.
(12) Davies, G. J.; Henrissat, B. *Biochem. Soc. Trans.* **2002**, *30*, 291−297.
(13) Davies, G. J.; Gloster, T. M.; Henrissat, B. *Curr. Opin. Struct. Biol.* **2005**, *15*, 637−645.
(14) Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res.* **2009**, *37*, D233−D238.
(15) Lombard, V.; Bernard, T.; Rancurel, C.; Brumer, H.; Coutinho, P. M.; Henrissat, B. *Biochem. J.* **2010**, *432*, 437−444.
(16) Vaaje-Kolstad, G.; Westereng, B.; Horn, S. J.; Liu, Z. L.; Zhai, H.; Sørlie, M.; Eijsink, V. G. H. *Science* **2010**, *330*, 219−222.
(17) Quinlan, R. J.; Sweeney, M. D.; Lo Leggio, L.; Otten, H.; Poulsen, J. C. N.; Johansen, K. S.; Krogh, K.; Jorgensen, C. I.; Tovborg, M.; Anthonsen, A.; Tryfona, T.; Walter, C. P.; Dupree, P.; Xu, F.; Davies, G. J.; Walton, P. H. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 15079−15084.
(18) Beeson, W. T.; Phillips, C. M.; Cate, J. H. D.; Marletta, M. A. *J. Am. Chem. Soc.* **2012**, *134*, 890−892.
(19) Levasseur, A.; Drula, E.; Lombard, V.; Coutinho, P.; Henrissat, B. *Biotechnol. Biofuels* **2013**, *6*, 41.
(20) Lynd, L. R.; Weimer, P. J.; van Zyl, W. H.; Pretorius, I. S. *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 506−577.
(21) Barr, B. K.; Hsieh, Y. L.; Ganem, B.; Wilson, D. B. *Biochemistry* **1996**, *35*, 586−592.
(22) Horn, S. J.; Sorbotten, A.; Synstad, B.; Sikorski, P.; Sørlie, M.; Varum, K. M.; Eijsink, V. G. H. *FEBS J.* **2006**, *273*, 491−503.
(23) Igarashi, K.; Koivula, A.; Wada, M.; Kimura, S.; Penttila, M.; Samejima, M. *J. Biol. Chem.* **2009**, *284*, 36186−36190.
(24) Igarashi, K.; Uchihashi, T.; Koivula, A.; Wada, M.; Kimura, S.; Okamoto, T.; Penttila, M.; Ando, T.; Samejima, M. *Science* **2011**, *333*, 1279−1282.
(25) Horn, S. J.; Sørlie, M.; Varum, K. M.; Väljamäe, P.; Eijsink, V. G. *Methods Enzymol.* **2012**, *510*, 69−95.
(26) Fox, J. M.; Levine, S. E.; Clark, D. S.; Blanch, H. W. *Biochemistry* **2012**, *51*, 442−452.
(27) Horn, S. J.; Sikorski, P.; Cederkvist, J. B.; Vaaje-Kolstad, G.; Sørlie, M.; Synstad, B.; Vriend, G.; Varum, K. M.; Eijsink, V. G. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 18089−18094.
(28) Zhang, Y. H. Z.; Himmel, M. E.; Mielenz, J. R. *Biotechnol. Adv.* **2006**, *24*, 452−481.
(29) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804−807.
(30) Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K.; Jones, T. A. *Science* **1990**, *249*, 380−386.
(31) Divne, C.; Ståhlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J. K. C.; Teeri, T. T.; Jones, T. A. *Science* **1994**, *265*, 524−528.
(32) 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*, 383−397.
(33) Divne, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309−325.
(34) Zhang, S.; Irwin, D. C.; Wilson, D. B. *Eur. J. Biochem.* **2000**, *267*, 3101−3115.
(35) Muñoz, I. G.; Ubhayasekera, W.; Henriksson, H.; Szabó, I.; Pettersson, G.; Johansson, G.; Mowbray, S. L.; Ståhlberg, J. *J. Mol. Biol.* **2001**, *314*, 1097−1111.
(36) Breyer, W. A.; Matthews, B. W. *Protein Sci.* **2001**, *10*, 1699−1711.
(37) Varrot, A.; Frandsen, T. P.; von Ossowski, I.; Boyer, V.; Cottaz, S.; Driguez, H.; Schulein, M.; Davies, G. J. *Structure* **2003**, *11*, 855−864.
(38) von Ossowski, I.; Ståhlberg, 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*, 817−829.
(39) Proctor, M. R.; Taylor, E. J.; Nurizzo, D.; Turkenburg, J. P.; Lloyd, R. M.; Vardakou, M.; Davies, G. J.; Gilbert, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 2697−2702.
(40) Zakariassen, H.; Aam, B. B.; Horn, S. J.; Varum, K. M.; Sørlie, M.; Eijsink, V. G. H. *J. Biol. Chem.* **2009**, *284*, 10610−10617.
(41) Mertz, B.; Gu, X.; Reilly, P. J. *Biopolymers* **2009**, *91*, 478−495.
(42) Sørlie, M.; Zakariassen, H.; Norberg, A. L.; Eijsink, V. G. H. *Biocatal. Biotransform.* **2012**, *30*, 353−365.
(43) Mulakala, C.; Reilly, P. J. *Proteins: Struct., Funct., Bioinf.* **2005**, *61*, 590−596.
(44) Mulakala, C.; Reilly, P. J. *Proteins: Struct., Funct., Bioinf.* **2005**, *60*, 598−605.
(45) Mertz, B.; Hill, A. D.; Mulakala, C.; Reilly, P. J. *Biopolymers* **2007**, *87*, 249−260.
(46) Payne, C. M.; Bomble, Y.; Taylor, C. B.; McCabe, C.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J. Biol. Chem.* **2011**, *286*, 41028−41035.
(47) Payne, C. M.; Baban, J.; Horn, S. J.; Backe, P. H.; Arvai, A. S.; Dalhus, B.; Bjørås, M.; Eijsink, V. G. H.; Sørlie, M.; Beckham, G. T.; Vaaje-Kolstad, G. *J. Biol. Chem.* **2012**, *287*, 36322−36330.
(48) Taylor, C. B.; Payne, C. M.; Himmel, M. E.; Crowley, M. F.; McCabe, C.; Beckham, G. T. *J. Phys. Chem. B* **2013**, *117*, 4924−4933.
(49) Koivula, A.; Kinnari, T.; Harjunpaa, V.; Ruohonen, L.; Teleman, A.; Drakenberg, T.; Rouvinen, J.; Jones, T. A.; Teeri, T. T. *FEBS Lett.* **1998**, *429*, 341−346.
(50) van Aalten, D. M. F.; Synstad, B.; Brurberg, M. B.; Hough, E.; Riise, B. W.; Eijsink, V. G. H.; Wierenga, R. K. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5842−5847.
(51) 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*, 8979−8984.

(52) Aronson, N. N.; Halloran, B. A.; Alexyev, M. F.; Amable, L.; Madura, J. D.; Pasupulati, L.; Worth, C.; Van Roey, P. *Biochem. J.* **2003**, *376*, 87−95.
(53) Watanabe, T.; Ariga, Y.; Sato, U.; Toratani, T.; Hashimoto, M.; Nikaidou, N.; Kezuka, Y.; Nonaka, T.; Sugiyama, J. *Biochem. J.* **2003**, *376*, 237−244.
(54) Baban, J.; Fjeld, S.; Sakuda, S.; Eijsink, V. G. H.; Sørlie, M. *J. Phys. Chem. B* **2010**, *114*, 6144−6149.
(55) Davies, G. J.; Brzozowski, A. M.; Dauter, M.; Varrot, A.; Schulein, M. *Biochem. J.* **2000**, *348*, 201−207.
(56) Beckham, G. T.; Bomble, Y. J.; Bayer, E. A.; Himmel, M. E.; Crowley, M. F. *Curr. Opin. Biotechnol.* **2011**, *22*, 231−238.
(57) Chundawat, S. P. S.; Beckham, G. T.; Himmel, M. E.; Dale, B. E. *Annu. Rev. Chem. Biomol. Eng.* **2011**, *2*, 121−145.
(58) Kurašin, M.; Väljamäe, P. *J. Biol. Chem.* **2011**, *286*, 169−177.
(59) Cruys-Bagger, N.; Elmerdahl, J.; Praestgaard, E.; Tatsumi, H.; Spodsberg, N.; Borch, K.; Westh, P. *J. Biol. Chem.* **2012**, *287*, 18451−18458.
(60) Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 4118−4127.
(61) Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J. Phys. Chem. Lett.* **2011**, *2*, 1546−1550.
(62) Beckham, G. T.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 4516−4522.
(63) Jiang, W.; Hodoscek, M.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 2583−2588.
(64) Ståhlberg, J.; Johansson, G.; Pettersson, G. *Biochim. Biophys. Acta* **1993**, *1157*, 107−113.
(65) Irwin, D. C.; Spezio, M.; Walker, L. P.; Wilson, D. B. *Biotechnol. Bioeng.* **1993**, *42*, 1002−1013.
(66) Medve, J.; Karlsson, J.; Lee, D.; Tjerneld, F. *Biotechnol. Bioeng.* **1998**, *59*, 621−634.
(67) Ubhayasekera, W.; Muñoz, I. G.; Vasella, A.; Ståhlberg, J.; Mowbray, S. L. *FEBS J.* **2005**, *272*, 1952−1964.
(68) Parkkinen, T.; Koivula, A.; Vehmaanpera, J.; Rouvinen, J. *Protein Sci.* **2008**, *17*, 1383−1394.
(69) Textor, L. C.; Colussi, F.; Silveira, R. L.; Serpa, V.; de Mello, B. L.; Muniz, J. R. C.; Squina, F. M.; Pereira, N.; Skaf, M. S.; Polikarpov, I. *FEBS J.* **2013**, *1*, 56−69.
(70) Momeni, M. H.; Payne, C. M.; Hansson, H.; Mikkelsen, N. E.; Svedberg, J.; Engström, Å.; Sandgren, M.; Beckham, G. T.; Ståhlberg, J. *J. Biol. Chem.* **2013**, *288*, 5861−5872.
(71) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(72) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(73) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(74) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(75) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543−2564.
(76) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009**, *5*, 2353−2370.
(77) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. *J. Chem. Theory Comput.* **2011**, *7*, 3162−3180.
(78) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926−935.
(79) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198−2202.
(80) Deng, Y. Q.; Roux, B. *J. Chem. Theory Comput.* **2006**, *2*, 1255−1273.
(81) Deng, Y. Q.; Roux, B. *J. Phys. Chem. B* **2004**, *108*, 16567−16576.
(82) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, 129.
(83) Lucius, A. L.; Maluf, N. K.; Fischer, C. J.; Lohman, T. M. *Biophys. J.* **2003**, *85*, 2224−2239.
(84) Cruys-Bagger, N.; Elmerdahl, J.; Praestgaard, E.; Borch, K.; Westh, P. *FEBS J.* **2013**, *280*, 3952−3961.
(85) 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*, 10189−10194.
(86) Hsu, T. A.; Gong, C. S.; Tsao, G. T. *Biotechnol. Bioeng.* **1980**, *22*, 2305−2320.
(87) Wald, S.; Wilke, C. R.; Blanch, H. W. *Biotechnol. Bioeng.* **1984**, *26*, 221−230.
(88) Philippidis, G. P.; Smith, T. K.; Wyman, C. E. *Biotechnol. Bioeng.* **1993**, *41*, 846−853.
(89) Nidetzky, B.; Zachariae, W.; Gercken, G.; Hayn, M.; Steiner, W. *Enzyme Microb. Technol.* **1994**, *16*, 43−52.
(90) Oh, K. K.; Kim, S. W.; Jeong, Y. S.; Hong, S. I. *Appl. Biochem. Biotechnol.* **2000**, *89*, 15−30.
(91) Bezerra, R. M. F.; Dias, A. A. *Appl. Biochem. Biotechnol.* **2004**, *112*, 173−184.
(92) Gruno, M.; Väljamäe, P.; Pettersson, G.; Johansson, G. *Biotechnol. Bioeng.* **2004**, *86*, 503−511.
(93) Bezerra, R. M. F.; Dias, A. A. *Appl. Biochem. Biotechnol.* **2005**, *126*, 49−59.
(94) Bu, L. T.; Beckham, G. T.; Shirts, M. R.; Nimlos, M. R.; Adney, W. S.; Himmel, M. E.; Crowley, M. F. *J. Biol. Chem.* **2011**, *286*, 18161−18169.
(95) Chundawat, S. P. S.; Bellesia, G.; Uppugundla, N.; Sousa, L. D.; Gao, D. H.; Cheh, A. M.; Agarwal, U. P.; Bianchetti, C. M.; Phillips, G. N.; Langan, P.; Balan, V.; Gnanakaran, S.; Dale, B. E. *J. Am. Chem. Soc.* **2011**, *133*, 11163−11174.
(96) Gao, D.; Chundawat, S. P. S.; Sethi, A.; Balan, V.; Gnanakaran, S.; Dale, B. E. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10922−10927.