J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ARTICLE

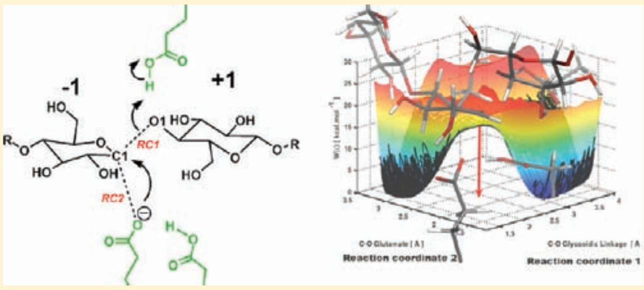
pubs.acs.org/JACS

Molecular Details from Computational Reaction Dynamics for the Cellobiohydrolase I Glycosylation Reaction

Christopher B. Barnett, Karl A. Wilkinson, and Kevin J. Naidoo*

Scientific Computing Research Unit and Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

S Supporting Information



ABSTRACT: Glycosylation of cellobiose hydrolase I (CBHI), is a key step in the processing and degradation of cellulose. Here the pathways and barriers of the reaction are explored using the free energy from adaptive reaction coordinate forces (FEARCF) reaction dynamics method coupled with SCC-DFTB/MM. In many respects CBHI follows the expected general GH7 family mechanism that involves the Glu-X-Asp-X-X-Glu motif. However, critical electronic and conformational details, previously not known, were discovered through our computations. The central feature that ensures the success of the glycosylation reaction are the Glu212 nucleophile's hydrogen bond to the hydroxyl on C2, of the glucose in the −1 position of the cellulosic strand. This Glu212 function restricts the C2 hydroxyl in such a way as to favor the formation of the ^4E ring pucker of the −1 position glucose. A frontier molecular orbital analysis of the structures along the reaction surface proves the existence of an oxocarbenium ion, which has both transition state and intermediate character. The transition state structure is able to descend down the glycosylation pathway through the critical combination of Asp214 (HOMO), ring oxygen (LUMO), and Glu212 (HOMO), anomeric carbon (LUMO) interactions. Using the fully converged FEARCF SCC-DFTB/MM reaction surface, we find a barrier of 17.48 kcal/mol separating bound cellulose chain from the glycosylated CBHI. Taking recrossing into account gives $k_{cat} = 0.415\ s^{-1}$ for cellobiohydrolase glycosylation.

INTRODUCTION

Cellulose, the β-1,4-linked straight chain homopolymer of glucose, plays a key structural role in plants as a component of lignocellulose (along with hemicellulose and lignin) and is the most abundant organic compound in the biosphere. This abundance has made it a key target for a source to produce biofuels such as ethanol. While the commercial use of cellulosic ethanol will stave off environmental and energy catastrophes there are several biochemical and chemical obstacles which make this obvious solution an immediate impossibility. The decline in reaction rate for uninterrupted cellulose digestion by enzymes is an obstacle that is not understood.[1] Notwithstanding the suggestion that the highly crystalline nature of cellulose limits its degradation to smaller dimer and monomer units,[2−4] the molecular details of cellulose hydrolysis remain unknown.

Cellulose is degraded by catalytic hydrolysis via the cellulase class of enzymes, particularly the cellulose β-1,4 cellobiosides (EC 3.2.1.91) of which cellobiohydrolases (CBH) act on crystalline cellulose. Specifically the glycosidic linkages of the polymeric cellulose are cleaved via glycosylation and deglycosylation. CBHIs cleave from the reducing end of a cellulose chain to yield β-cellobiose, and CBHIIs act from the opposite end of the chain, yielding α-cellobiose. In both cases, the CBH cleaves the β-1,4 glycosidic linkage between the main cellulose chain and two terminal glucopyranose residues (designated +1, +2). Both classes of CBHs may either invert or retain the stereochemistry at the anomeric carbon.

The enzymes secreted by the high cellulase-producing filamentous fungus *Trichoderma reesei* (*T. reesei*) are known to be very effective against native, crystalline cellulose. *T. reesei* produces both CBHI and CBHII. CBHI is able to act on the inaccessible, crystalline regions of cellulose,[5] and it cleaves off cellobiose from the chain ends in a processive manner.[6−8] These enzymes are believed to act synergistically to solubilize cellulose, as has been shown for native cotton cellulose.[9]

The CBH acts processively along the cellulosic strand (Figure 1), cleaving off cellobiose units that are then ejected from the catalytic domain. CBHs extract a single chain of cellulose with the aid of the cellulose binding domain (CBD).[6,10,11] The CBD is bound to the cellulase catalytic region by a flexible linker[6] and has conserved tyrosine residues that bind it to the surface of crystalline cellulose.[12] It has been hypothesized that the CBD processes along the surface until it overlaps the reducing end of a broken cellulosic strand. The broken cellulosic strand induces conformational changes[12] in the CBD such that it gains traction on the broken cellulose surface and may direct the cellulose chain toward the catalytic binding channel. The cellulosic strand is directed past the catalytic domain where the glycosylation and deglycosylation reactions take place.

The catalytic domain is a specific part of the lengthy binding channel that is surrounded by the β-jelly-roll scaffolding of

Received: July 28, 2011
Published: October 18, 2011

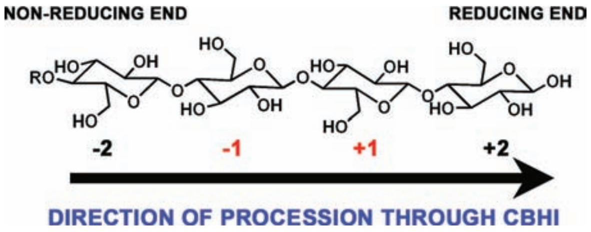


Figure 1. Labeling of the residues of the cellulosic strand. Residues after the glycosidic linkage to be cleaved are labeled positively (+), while those before it are labeled negatively (−). R indicates the remainder of the cellulosic strand.[8]

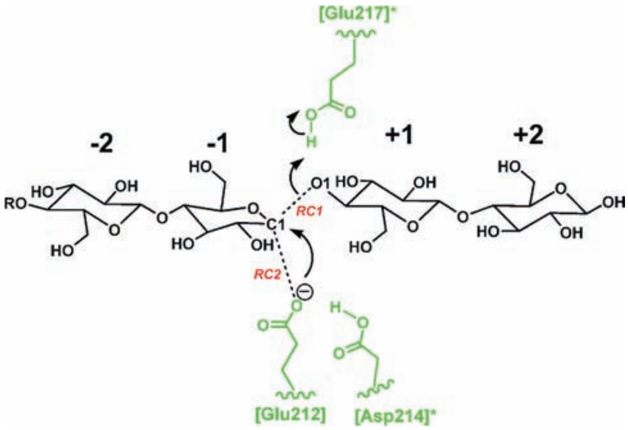


Figure 2. CBH1 skeleton active site showing reaction coordinates labeled RC1 and RC2. R at the left-hand cellulosic oligosaccharide tail represents the remaining four glucose units. Protonated amino acids are starred (*).

CBHI. The binding channel is 40 Å in length,[7] has capacity for nine glycosyl residues,[8] and features a bend that induces conformational strain on the glucose unit in that location.[13] The Glu-X-Asp-X-X-Glu motif is considered key in this family of enzymes, and mutagenesis[14] and crystallographic studies[7] have identified Glu212, Asp214, and Glu217 as critical for catalytic activity. These residues occur in the latter part of the binding channel, within the catalytic domain, and are responsible for glycosylation/deglycosylation. Key residues and interactions required for action on short oligosaccharide strands have been identified from X-ray crystallographic structures.[8] Of particular importance is the interaction of Glu217 with the glycosidic linkage that is to be cleaved.

Overall, the glycosylation step is a substitution reaction in which the oxygen of a glutamate residue (Glu212) substitutes the glycosidic linkage at the anomeric carbon of the −1 position (Figure 2). The cleavage of the glycosidic linkage is assisted by protonation of the glycosidic oxygen by glutamic acid (Glu217). The first step of the proposed mechanism for CBHI, the glycosylation, is a general acid-mediated mechanism.

The important covalent interactions in the glycosylation reaction are the glycosidic linkage protonation, breakage of the glycosidic linkage, and the attack of the nucleophilic glutamate 212 on the anomeric position. We therefore investigate the reaction dynamics of the glycosylation by calculating the reaction surface as a function of two reaction coordinates. We monitor the C1—O1 bond (RC1) and the C1—O (Glu212) bond (RC2). The dynamic evolution of the reaction mechanism is calculated using the free energies from adaptive reaction coordinate forces (FEARCF) approach.[15,16] This is a versatile free energy approach that provides extensive sampling of reaction surfaces, which leads to a glycosylation mechanism that is not prejudiced or guided by constraining or restraining potentials. The transition state structures and barrier heights can be determined from the reaction trajectories and reaction surface. We employed a quantum mechanical/molecular mechanical (QM/MM) methodology where the important reaction facilitating amino acids of the active site and the reacting portion of the sugar are treated quantum mechanically. We are able to gauge the extent to which glycosidases influence glucose ring pucker to lower the reaction barrier. This leads to an understanding of the relationship between cellulose substrate puckering and the mechanism of action of glycosidases.

■ METHODS

Reaction Dynamics and Reaction Surfaces using FEARCF.

The FEARCF is a flat histogram method based on probability distributions. The reaction coordinate space is a discretized *n*-dimensional grid where the sampling frequency for a bin site is recorded for each simulation. This method was used to calculate the two-dimensional reaction surface, $W(\xi)$ for the glycosylation step catalyzed by CBHI. In this method each bond that is formed or broken during a reaction can be allocated to a specific reaction coordinate or (as chemically appropriate) be merged with other bonds as a combination-type reaction coordinate. Several reaction coordinates are possible, making the elucidation of complex reactions possible. The reaction coordinates (ξ_1 = RC1, ξ_2 = RC2, ξ_3 = RC3, ...) are used to produce a multidimensional reaction coordinate hypersurface from which the reaction mechanism may be derived. In the case of the glycosylation reaction only two coordinates (ξ_1, ξ_2) were required and these were gridded into discrete bins. Protonation of the glycosidic linkage occurs naturally, without the assistance of a biasing force; this bond was therefore excluded from the reaction coordinate definitions. The bins were populated from the sampling of the reaction coordinate space which was gained from the collective simulations to that point. This led to the reaction coordinate probability density $P(\xi)$ from which the current reaction surface guess can be calculated

$$W(\xi) = -k_BT \ln P(\xi) \tag{1}$$

To correctly assign a weighting contribution from each simulation the individual histograms from multiple simulations were combined using the weighted histogram analysis method (WHAM).[17−19] This is an extension of the histogram procedure developed by Ferrenberg and Swendsen[20,21] where probability distribution histograms for all simulations overlap optimally to give an accurate, unbiased probability distribution p_k. A self-consistent solution to the WHAM equations,

$$P_k = \frac{\sum_i n_{i,k}}{\sum_i N_i f_i\, e^{-\beta U_i(\xi_k)}} \tag{2}$$

$$N_i = \sum_k n_{i,k}$$

$$f_i = \frac{1}{\sum_k e^{-\beta U_i \xi_i} p_k}$$

is sought, where f_i are the free energy weighting factors that depend on the solution p_k, the total number of configurations is N_i, and each simulation, i, has n_i stored configurations per kth bin of the discretized histogram. Local gradients, calculated from the summed and weighting-corrected sampling of reaction coordinate space, were then used to determine biasing reaction coordinate forces

$$F(\xi_i) = \frac{\partial W(\xi)}{\partial \xi_i} \tag{3}$$

The forces calculated from reaction coordinate histograms for all i simulations were used to drive simulation i+1 toward the under-sampled regions of the reaction space. This was done iteratively until the reaction surface reached convergence (i.e., long trajectories underwent multiple barrier crossings and recrossings between reactant and product wells).

Self-Consistent Charge Density Functional Tight Binding (QM/MM). Hybrid quantum mechanical/molecular mechanical (QM/MM) simulations were performed for the cellulase with cellooctaose substrate using a self-consistent charge density functional tight binding (SCC-DFTB) level of theory as implemented in CHARMM[22] for the QM region. The MM region was treated with the carbohydrate solution force field (CSFF).

SCC-DFTB delivers with speed comparable to that of traditional semiempirical (SE) methods such as the MNDO, AM1, and PM3 schemes. It has the added benefit of producing the best carbohydrate ring puckering behavior of any of the popularly used SE methods.[23] A key difference is that SCC-DFTB is derived from density functional theory (DFT), whereas SE methods are derived from Hartree−Fock (HF) theory. In density functional tight binding (DFTB) theory the total energy is written in terms of a reference electron density (ρ_0), which is equal to a superposition of neutral atomic densities and is then given as

$$\begin{aligned} E[\rho_0(\vec{r})] = {} & \sum_i^{\text{occ}} \rangle\psi_i(\vec{r})|h_i^{\text{KS}}[\rho_0(\vec{r})]|\psi_i(\vec{r})\langle - \frac{1}{2}\int \frac{\rho_0(\vec{r})\rho_0(\vec{r}\,')}{|\vec{r}-\vec{r}\,'|} d\vec{r}\ d\vec{r} \\ & + \mathrm{E}_{\text{xc}}[\rho_0(\vec{r})] + \int \upsilon_{\text{xc}}\rho_0(\vec{r})\ d\vec{r} + \frac{1}{2}\sum_{\alpha\beta}\frac{Z_\alpha Z_\beta}{R_{\alpha\beta}} \end{aligned} \tag{4}$$

This expression is valid when compared to the Kohn−Sham (KS) total energy functional. The DFTB total energy functional (eq 4) is the sum over all occupied orbitals obtained from the diagonalization of the electronic Hamiltonian (first term), plus the correction for double counting of electron−electron interactions, expanded up to the second order (second, third, and fourth terms) and finally core repulsion (last term). To speed up the process of energy minimizations, the above equation is done only once so that only the shape of the KS orbital and not the changing density is considered. This *nonself-consistent* optimization is the essence of DFTB. A simplification of the first term was made to further increase the computational speed. It was assumed that the solution of the first term resulted in the KS atomic energy of the atomic orbital basis functions (ε) from the diagonal matrix elements and the kinetic and effective potential energy that depends on only two atoms bonded to each other. The off diagonal matrix elements are therefore dependent on the interatomic separation and can be fitted to analytic functions or interpolated from a previously populated lookup table. The pairwise interaction simplification has been extended to the rest of the terms in eq 4 where they collectively can be identified as the effective repulsive energy E_{rep}. This can be computed by fitting to analytic functions or interpolated from tabulations over various distances.

A major weakness in the DFTB approach is the assumption that individual unperturbed atomic charge densities can be summed to represent the charge density of the molecule accurately. This assumption neglects polarization effects between atoms that have very different electronegativities. Elstner et al. corrected this by deriving the SCC-DFTB scheme through a second-order expansion of the DFT total energy functional with respect to the charge density fluctuations $\delta\rho(\mathbf{r})$ around the reference density $\rho(r)$.[24] The density fluctuations are defined as a sum of all the atomic contributions

$$\delta\rho(\vec{r}) = \sum_\alpha \Delta q_\alpha \tag{5}$$

where q is the atomic contribution to the charge. The contribution to the total energy from the density fluctuations about the reference density is then the second-order term

$$\begin{aligned} & \frac{1}{2}\int\int \left[\frac{1}{|\vec{r}-\vec{r}|} + \left.\frac{\delta E_{\text{xc}}}{\delta\rho(\vec{r})\delta\rho(\vec{r})}\right|_{\rho_0} \delta\rho(\vec{r})\delta\rho(\vec{r})\right] \mathrm{d}\vec{r}\ \mathrm{d}\vec{r} \\ & = \sum_{\alpha\beta}^{\text{atoms}} \Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \end{aligned} \tag{6}$$

where γ is the effective inverse distance between neighboring atoms and is related to the atomic hardness. The SCC-DFTB format of the total DFT energy is then written as

$$E[\rho_0(\vec{r})] = \sum_i^{\text{occ}} \rangle\psi_i(\vec{r})|h_i^{\text{KS}}[\rho_0(\vec{r})]|\psi_i(\vec{r})\langle + E_{\text{rep}} + \sum_{\alpha\beta}\Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} \tag{7}$$

The SCC-DFTB implementation is based on generalized gradient approximation (GGA) functionals. Therefore, the method exhibits problems originating from these functionals such as the inaccurate modeling of dispersion interactions.[25−27] The long-range dispersion interactions (C_6/R^6 and higher components) are poorly accounted for[25,26] but can be empirically corrected[28] as has been shown for DFT calculations.[29] This empirical function applied with a damping function, $f(R_{\alpha\beta})$ that rapidly approaches zero, combined with empirical coefficients $C_{\alpha\beta}$ can significantly improve the SCC-DFTB energy.

$$\begin{aligned} E[\rho_0(\vec{r})] = {} & \sum_i^{\text{occ}} \rangle\psi_i(\vec{r})|h_i^{\text{KS}}[\rho_0(\vec{r})]|\psi_i(\vec{r})\langle + E_{\text{rep}} \\ & + \sum_{\alpha\beta}\Delta q_\alpha \Delta q_\beta \gamma_{\alpha\beta} - \sum_{\alpha\beta} f(R_{\alpha\beta})\frac{C_6^{\alpha\beta}}{R^6} \end{aligned} \tag{8}$$

It has been shown that this correction in SCC-DFTB leads to reliable nucleic acid base-stacking interactions[28] and polypeptide (protein) structures.[30] Finally, when biomolecules composed of amino acids and nucleic acids are simulated, a correct representation of hydrogen bonds is critical to the accuracy of the trajectory. SCC-DFTB underestimates each hydrogen bond by as much as 2−3 kcal/mol Here, we included corrections for hydrogen bonding using a damped term.[31]

Simulation Details. The 7CEL.pdb structure as reported by Divne et al.[8] was protonated following a pK_a analysis.[32−34] Aspartates and glutamates greater than 18 Å away from the active site were protonated to achieve an overall neutrally charged MM part of the system. Three active site amino acids, glutamic acid 217 (Glu217), aspartic acid 214 (Asp214), and glutamate 212 (Glu212), were previously found to be significant for transition state stabilization, and were therefore treated quantum mechanically. For glycolysis to occur the glycosidic linkage must be correctly oriented toward Glu217.[8] The cellooctaose was placed inside the catalytic domain (Figure 2, with correct orientation confirmed, Figure 3), and glucoses at the −1 and +1 position of the substrate were treated quantum mechanically. The quantum region, with an overall charge of −1, was modeled using SCC-DFTB/d. The glucose units on either side of the breaking glycosidic bond were treated quantum mechanically (SCC-DFTB/d), and GHO[35] atoms were used to bridge the parts of the sugar that were treated quantum mechanically and those that were treated classically using the CSFF forcefield.[36]

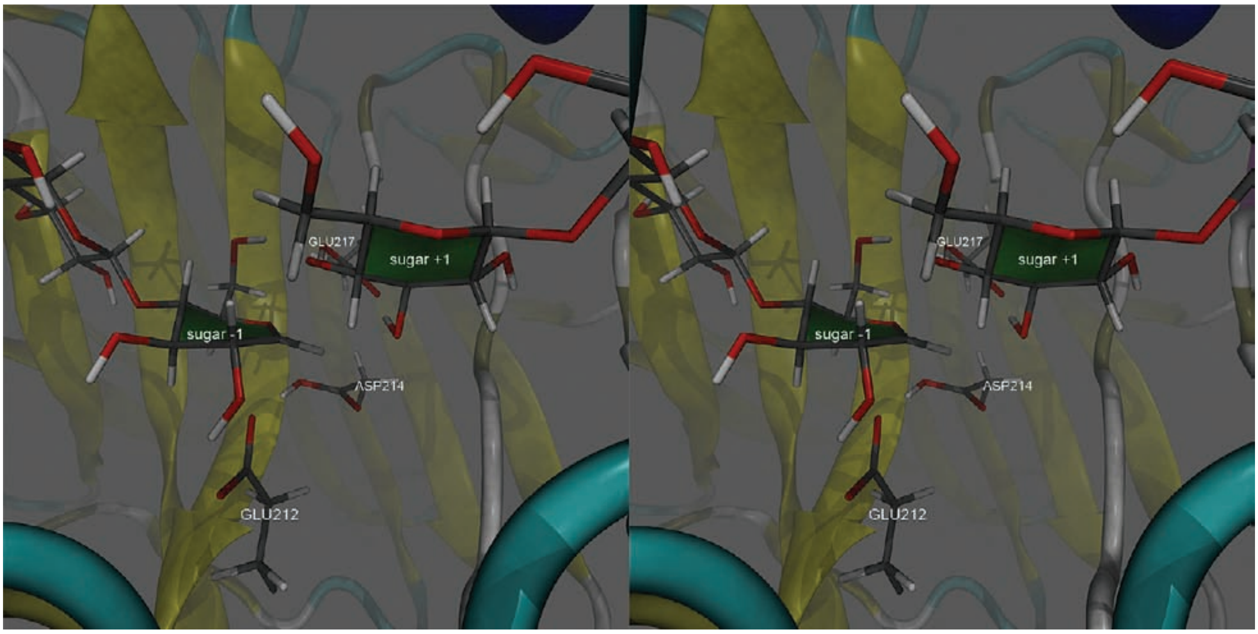


Figure 3. Labeled stereogram of the cellulose active site, with the commonly seen $^4H_3$ transition state structure.

After initial ABNR minimization for 500 steps the protein was solvated in a sphere of TIP3P water, with radius 38 Å, centered on the glucose at the −1 site. The addition of a spherical boundary potential maintains the integrity of the solvent sphere. The system was then heated for 30 ps and equilibrated for 100 ps, implementing leapfrog Langevin dynamics at 298.15 K with group-based cutoffs of 10, 12, and 14 Å force shifting for electrostatics and van der Waals potentials. The nonbonded interactions were updated using CHARMM's builtin heuristic algorithm. The SHAKE algorithm was applied to MM atoms. Each iteration of the FEARCF reaction simulation comprised eight individual QM/MM protein FEARCF calculations of 200 ps in length. Convergence of the free energy surface was reached after 10 iterations of FEARCF QM/MM dynamics.

TS Structure Selection from FEARCF Surface. We selected structures located at the barrier separating the reactants from the products for further examination. A characterization of TS structures was carried out to identify coordinates that lie in the TS region (i.e., between $(2.0 \leq \mathrm{RC1} \leq 2.4)$ and $(2.1 \leq \mathrm{RC2} \leq 2.5)$) and that have a single imaginary frequency. At first structures that that lie about the dividing surface (i.e., as shown by dotted line drawn on reaction surface in Figure 4 c) are considered to be TS structures. Next an analyses of the bond frequencies was done, leading to a more rigorous identification of TS structures. The effect of the environment was taken into consideration when performing the frequency calculations. The frequencies of the QM region (−1, +1, Glu212, Asp214, Glu217 as in Figure 2) were calculated using a B3LYP/6-31+G** level of theory while representing the rest of the protein with MM charges.

Transmission Coefficient Calculation. Only a few of these structures yielded a single imaginary frequency, and these were used in bluemoon-type simulations[37] to estimate the transmission coefficient. A total of 160 QM/MM simulations using structures that revealed TS character were carried out. Each simulation was 5 ps in length and initialized with randomly assigned velocities. Crossing and recrossing can occur in each simulation, and the transmission coefficient was calculated using this information such that $\kappa_{\tau,n} = \langle \xi \rangle_{\tau,n}$. This is the 'ensemble average' of crossing characteristics of the n simulations, where $\xi = 1/\alpha$. When the number of forward crossings is one greater than the number of backward crossings then $\alpha = 1$. It is zero, otherwise.

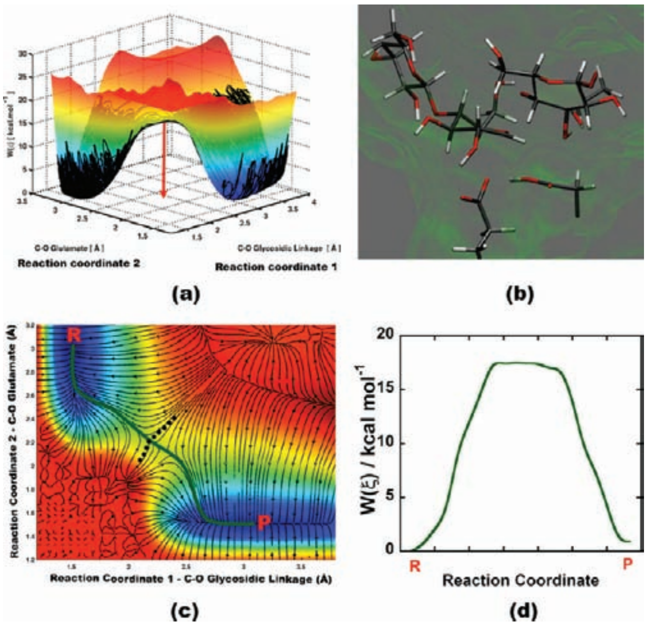


Figure 4. (a) Two-dimensional (2D) free energy surface for the chosen reaction coordinates for glycosylation in CBHI. (b) The most commonly seen transition state $^4H_3$. (c) Plot of the 2D free energy surface showing the vector gradient paths. The minimum energy path is highlighted in green. A stippled line for illustrative purposes indicates a TST dividing surface. (d) The minimum free energy path is extracted as a 1D graph. R represents the reactant and P the product.

Electronic Structure Analysis. Structures along the reaction path minima, including reactant and product minima as well as the transition state, were extracted from coordinates collected during the dynamics simulation. These structures were further analyzed using quantum mechanical calculations. Specifically, the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) were identified at these points. Coordinates and calculation details are included in the Supporting Information.

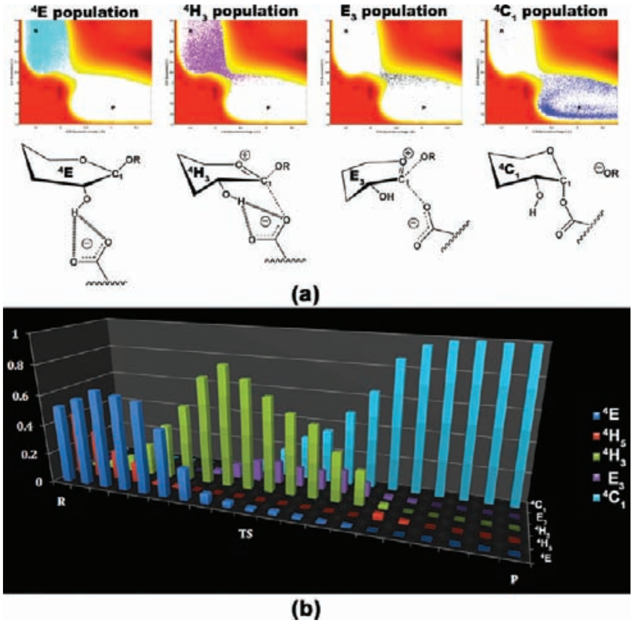


Figure 5. (a) The ^{4}E conformer is prevalent in the reactant region, the $^4H_3$ conformer is prevalent in the reactant and TS region, the E_3 conformer is seen in the TS region. A strained $^4C_1$ conformer is noted on the product side of the TS, which relaxes to $^4C_1$ in the product region. (b) Pucker distributions at chosen points along the minimum free energy path. At each point the pucker populations sum to unity.

RESULTS

Reaction Mechanism from the Free Energy Surface. The glycosylation reaction surface as determined using FEARCF displays two broad wells centered respectively at 1.5 Å and 3.0 Å for the reactant and at 3.05 Å and 1.5 Å for the product (Figure 4a). A single but broad barrier with no intermediate well separates the reactant and product wells. A free energy of 17.48 kcal/mol is required to cross over from the reactant to products, and this TS is represented in Figure 4b. The difference in energy stabilization between reactant and product is 0.89 kcal/mol.

A minimum free energy path can be calculated from the reaction free energy surface (Figure 4c). The gradient vectors are shown on the surface and oriented in the direction of decreasing energy (aimed "downhill"). The dividing surface as understood in TS theory (TST) is derived from these gradient vectors and is shown as a dotted line in Figure 4c. Furthermore, these gradients were used to determine the minimum energy path, which we plot separately (Figure 4d). This 1D free energy curve is reminiscent of a textbook description of a complex reaction that has been reduced to a single conflated reaction coordinate. Nonetheless, the depiction of the glycosylation reaction free energy profile in this way is instructive as it shows the shape of the reaction and the relatively broad transition state that may not be immediately apparent from the 2D surface.

With the converged free energy volume of ring pucker for the reactant calculated previously,[13] it was established that the substrate exhibits ring puckering because CBHI induces preferential puckering of glucose units at the −1 position of the cellulosic strand. In particular, CBHI *destabilizes* the $^4C_1$ conformer and preferentially *stabilizes* the ^{4}E conformer (Figure 5a first frame and Figure 5b) with the aid of Glu212 electrostatic and hydrogen bonding interaction to the C2 hydroxyl. This is significant as similarly strained conformers are observed for TS structures in CBH-catalyzed reactions. For example, Davies et al. have analyzed crystallographic data and concluded that four possible transition-state conformations, $^4H_3$, $^3H_4$, ^{2,5}B, and $B_{2,5}$, are likely in glycosidases.[38] In addition competing puckering conformations of saccharides in the transition state of various glycosidases has been observed for several members of the glycoside hydrolase family.[39] The puckering evidence can be combined with the dominant argument that an oxocarbenium ion is integral to the character of the TS[40] and necessary to the nature of the reaction mechanism. Within this context we examine our simulation structural and electronic data of the −1 sugar ring for evidence of conformational pucker that may facilitate the formation of a transition state with significant oxocarbenium ion character.

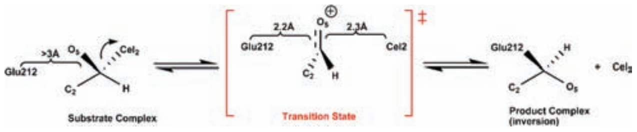


Figure 6. $D_N{}^*A_N$-like mechanism observed, where Cel_2 is cellobiose and atoms adjoining O5 and C2 have been omitted for clarity.[41]

The ring pucker evolves as the reaction progresses, and this is illustrated in Figure 5a. Scatter plots showing the population density of key pucker conformers are shown in the region below 18 kcal/mol and have been overlaid onto a map of the free energy surface. ^{4}E, $^4H_3$, E_3, and $^4C_1$ puckers of the −1 position glucose unit are shown interacting with the Glu212 (Figure 5a). In the reactant region, a strong Glu212−HO2 hydrogen bond predominantly *stabilizes* the ^{4}E conformer. As the reaction proceeds, $^4H_3$ becomes the most preferred conformer. At the TS the Glu212−HO2 interaction weakens as the Glu212 moves along to attack the anomeric carbon. While the $^4H_3$ conformer dominates, some E_3 conformer is present (Figure 5a,b). On the approach to and in the product region the Glu212 forms a covalent bond to the cellulosic strand (now cellohexaose), and the pulling effect (i.e., hydrogen bond) on the C2 carbon hydroxyl disappears. The result is that the −1 sugar ring relaxes into the $^4C_1$ conformation. While the reaction path described here is the most probable, it is possible that the glycosylation reaction may take slightly different minor paths through the TS region. Other possible pathways may include subtly different conformers ($^4H_3$, E_3, and distorted $^4C_1$) that are present in the TS (Figure 5b).

While the RC1 and RC2 distances are almost equal for the TS structure, tracing the minimum free energy reaction path reveals molecular details of the mechanism of glycosylation. When following the path from the edge of the reactant region (1.6 Å, 2.5 Å) to just before the TS (2.0 Å, 2.3 Å) it is clear that the rate of the glycosidic bond breaking is greater than the rate of glutamate carbon formation. This suggests that the leaving group (LG, the oxygen of the glycosidic linkage) dissociates more quickly than the nucleophile (Nu, Glu212) associates to form a bond. An examination of the electron density during the progression of bond formation between the LG and Nu reveals that the electrophilic carbon center in this reaction appears to contain both D_NA_N (Dissociation nucleophile, Association nucleophile) and $D_N{}^*A_N$ mechanistic character (Figure 6).[41] Considering only O5−C1 bond shortening as indicative of the oxocarbenium-like character the $D_N{}^*A_N$ appears to be the resulting mechanism observed.

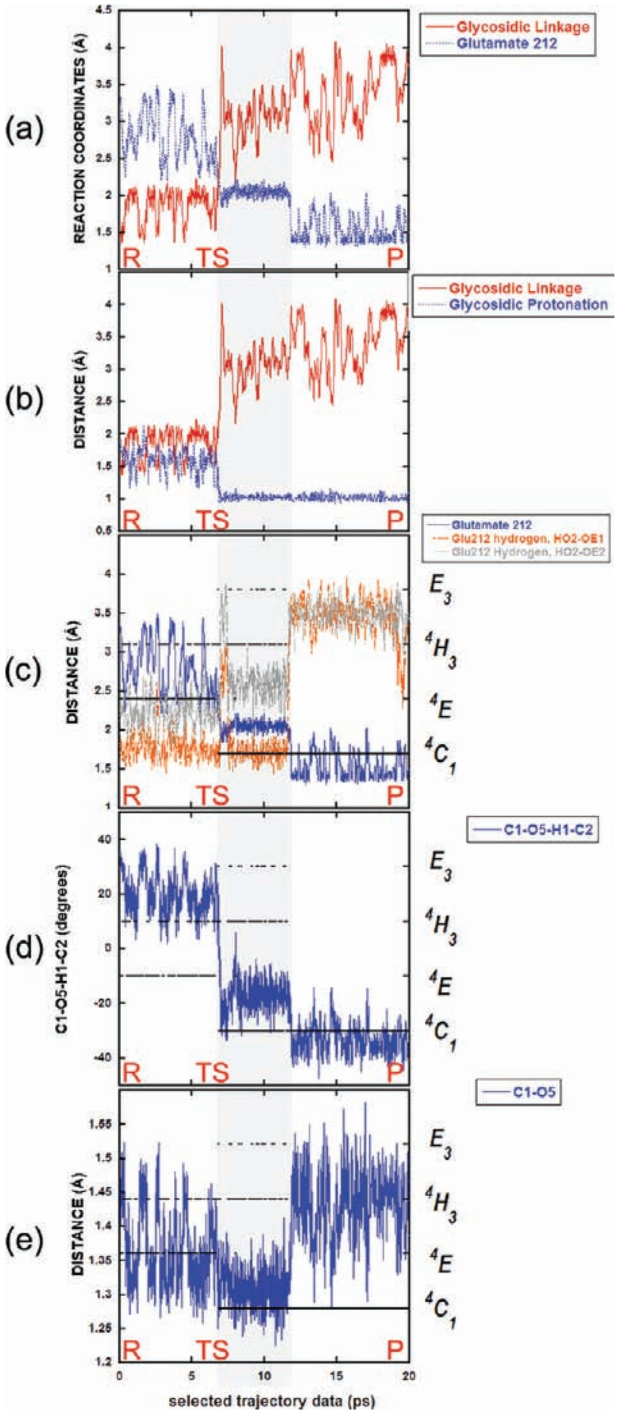


Figure 7. Short FEARCF trajectory showing (a) the reaction coordinates progressing from reactant to product, (b) protonation distance vs glycosidic distance, (c) RC2, the distance between the glutamatic acid oxygens (OE1, OE2) and the HO2 hydrogen and frequency of pucker conformers (d) the C1−O5−H1−C2− torsional angle and frequency of pucker conformers shown during glycosylation, and (e) the C1−O5 bond distance and the puckering.

Monitoring the reaction coordinate distances over part of a FEARCF trajectory (Figure 7a) shows that this transition occurs in about 6 ps which is on the order of the time scale expected.[40] Crossing of the reaction coordinates in this trajectory occurs at (2.2 Å, 2.2 Å) as seen in the 2D reaction free energy surface (Figure 4a). The glycosidic dissociation occurs in conjunction with protonation of the glycosidic oxygen by Glu217 (Figure 7b). The protonation is accelerated as it makes the easy progression from a 'perpetual' hydrogen bond formed between the Glu217 and the glycosidic linkage.

The relationship of ring pucker to molecular events within the reaction mechanism can be evaluated by analyzing the evolution from reactants to products of the −1 ring puckering frequency alongside that of the bonds, angles, and dihedrals during the reaction. The change in the glutamic acid reaction coordinate (RC2), the HO2−glutamatic acid interaction distance, and the frequency of E_3, $^4H_3$, ^4E, and $^4C_1$ pucker conformations are shown for a typical FEARCF trajectory (Figure 7c). While the reactant exists, at least one of the Glu212 nucleophile oxygens forms a strong interaction with the hydroxyl at C2. However, as the product forms, the interaction is lost. This supports the proposed structures shown in Figure 5.

The planarity of atoms C1, O5, H1, and C2 can be assessed by measuring the dihedral angle they make. In Figure 7d the extent of reaction and planarity of the dihedral are compared. In the reactant well, the hydrogen is α (i.e., points downward), in the transition state it is in the plane of the pyranose ring atoms O5−C1−C2, while in the product well it points upward and is β. To travel between these two extremes the structure must go through a planar structure with potential sp^2 character. The planarity of these atoms does not definitively prove that the anomeric carbon is sp^2 hybridized, for example in the trigonal bipyramidal transition state of S_N2 reactions (D_NA_N) these same atoms would appear planar.

However, the shortening of the C1−O5 bond length along with the planarity of atoms C1, O5, H1, and C2 provides geometric support for the existence of an oxocarbenium in the TS. Comparing the C1−O5 bond length and pucker in Figure 7e and reaction coordinates in Figure 7a confirms that the C1−O5 bond shortens at the TS. That is, when the hydrogen is in the plane of C1, C2, and O5; therefore, the C1−O5 bond gains double bond character; the anomeric carbon is sp^2 hybridized. The time series and pucker frequencies shown in Figure 7 provide evidence for the existence of an oxocarbenium ion as has been postulated previously.[40] The oxocarbenium character is taken to be an integral part of the reaction mechanism (Figure 6).[41]

Electronic Structure Analysis. An investigation into the electronic nature of this reaction is done using a HOMO/LUMO analysis of structures (Figure 8a−e) extracted from the reaction pathway identified in Figure 4d. These structures were taken from the trajectory presented in Figure 7 and coincide with regions of interest identified from the 1D and 2D free energy surfaces shown in c and d of Figure 4. Specifically, the reactant well (Figure 8a), a point associated with the cleavage of the C1−Ogly (Figure 8b), the transition state (Figure 8c), a point associated with the formation of the C1−Oglu bond (Figure 8d) and the product well (Figure 8e).

The effect of the protonation of the glycosidic oxygen can be seen from the growth of the HOMOs in the region of the glycosidic oxygen that is oriented toward the incoming proton (blue density shown in the top centers of the first frames in a and b of Figure 8). The structure that results from the subsequent cleavage of the C1−O glycosidic bond to form the oxocarbenium TS is shown in Figure 8c. The C1 carbon shows an increase in LUMO density (central frame Figure 8c) indicative of a carbocation, while the O5 ring oxygen shows an increase in HOMO

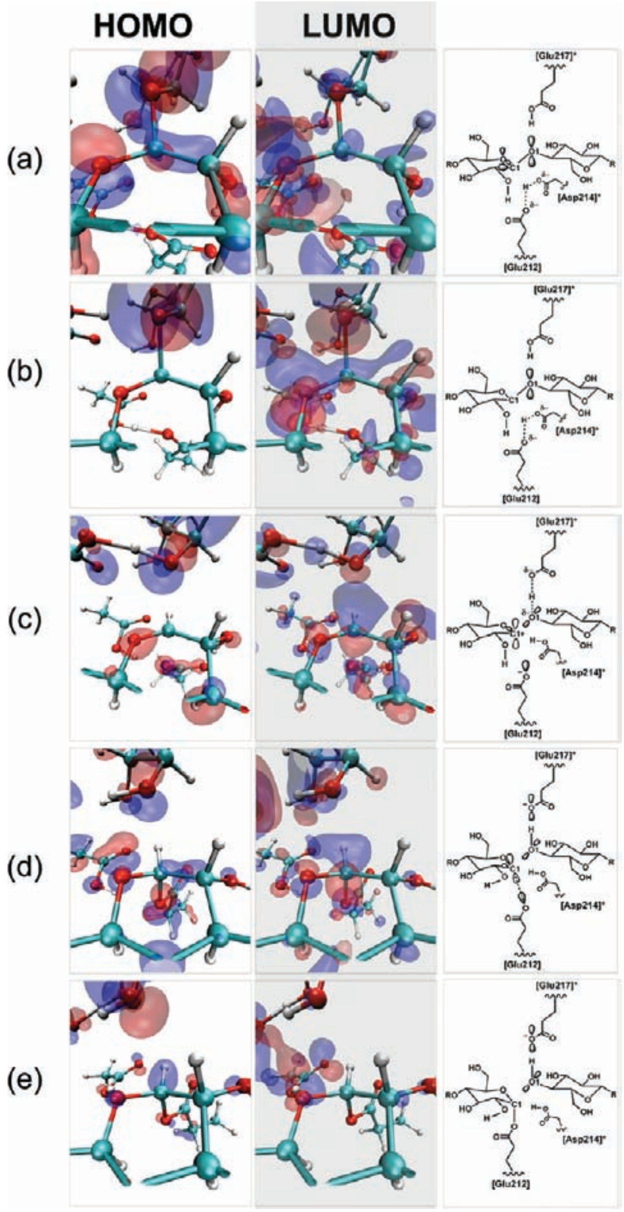


Figure 8. HOMO (left) and LUMO (center) of structures along the reaction coordinate forming the C1−Glu212 bond. Here the reactant is shown in (a), the reactant along with leaving group in (b), the oxocarbenium TS in (c), the product formation in (d), and the product in (e).

density (first frame Figure 8c). The C1−O5 bond therefore increases in strength as there is greater interaction in these frontier molecular orbitals in the TS. A vibrational analysis (B3LYP/6-31+G**) of the TS coordinates yields a single negative frequency with strong intensity at −89.547 cm^{-1} (cf. Supporting Information).

Next the Asp214 HOMO transfers electron density to the O5 LUMO, weakening the C1−O5 bond (to the left of the first two frames in Figure 8d). The presence of this interaction is supported by the shift in the position of the Asp214 residue hydrogen. In Figure 8a−c the O5−Asp214 hydrogen distance ranges from 2.3 to 2.6 Å, and a hydrogen bond between Asp214 and Glu212 is observed; however, in d and e of Figure 8 this reduces significantly to 1.8 Å, and the hydrogen bond is absent.

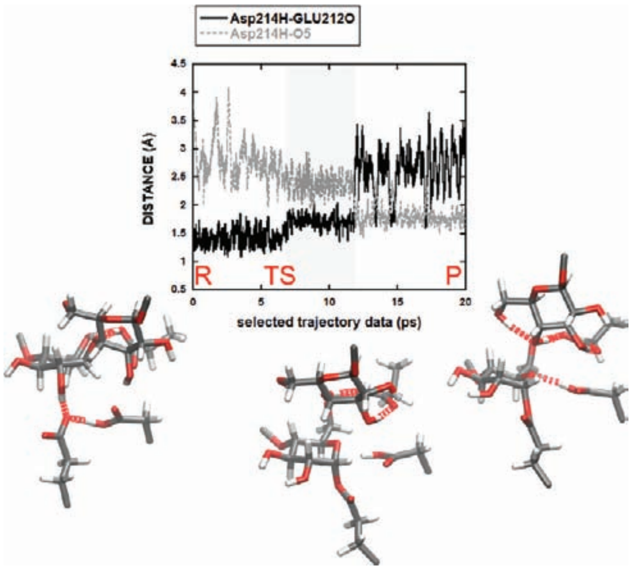


Figure 9. Plot from a short FEARCF trajectory which gives an indication of the Asp214 hydrogen donation ability by illustrating the distance between the Asp214 hydrogen and both the oxygens of Glu212 and the O5 ring oxygen (−1 position). Images indicating the hydrogen bonds as dashed tubes (calculated with the standard cutoffs in vmd) are included for clarity.

At the same time, the Glu212 no longer interacts with the C2 hydroxyl but instead the Glu212 HOMO transfers electron density to the C1 LUMO (center in frames 1 and 2 of Figure 8d). The conditions for bond formation between Glu212 and the anomeric carbon are favorable, which leads to the glycosylation of cellulose (Figure 8e). Asp214 plays a stabilizing role during the reaction, as it is a hydrogen bond donor. During the reactant phase Asp214 hydrogen bonds to Glu212 which breaks in the TS following which a hydrogen bond is made with the O5 ring oxygen (Figure 9).

Rate of Reaction. The computed activation free energy we find here can be compared to experimental values derived from k_{cat} using Eyring transition state theory (TST).[42] Transition state theory provides a connection between k_{cat} and the free energy of activation,

$$k_{\mathrm{cat}} = \kappa \frac{k_{\mathrm{B}}T}{h} \exp\left(-\Delta A''/RT\right) \qquad (9)$$

where the rate constant k depends on a transmission coefficient κ, Boltzmann's constant k_{B}, Planck's constant h, the universal gas constant R, the temperature T in Kelvin, and the change in the Helmholtz free energy $\Delta A''$ (kJ/mol).

The rate forward is based on the change in Helmholtz free energy from reactants to transition state (17.48 kcal/mol; 73.14 kJ/mol). This barrier height should result in a reasonable estimate of the reaction rate since the use of TST for energy barriers above ∼10 kcal/mol is considered reliable.[43]

Assuming the transmission coefficient is unity, the rate of reaction is 0.956 s^{-1}. However, calculating the transmission coefficient gives an improved estimate of the rate, κ. The transmission coefficient is calculated using a set of bluemoon simulations that are initiated from the TS. The TS structures were validated by frequency analysis as described above. The transmission is $\kappa = 0.4346$ which now gives a rate of 0.415 s^{-1}. Rate enhancements for such reactions are about 10^{17}.[38] While this is not the complete hydrolysis reaction, a comparison of this rate with the spontaneous hydrolysis of 1-methoxy-β-glucose at 25 °C (4.7×10^{-15} s^{-1}) is a speed up on the order of 10^{13}.[44,45]

CONCLUSIONS

While the mechanism of glycosylation has been presumed and proposed from crystallographic, mutation, and kinetic isotope effect studies, here a comprehensive dynamic study of the electronic and conformational transitions of the glycosylation reaction, using FEARCF dynamics, clarifies many of the assumptions made. The glycosylation reaction is sensitive to the dynamic puckering of the −1 glucose away from the stable $^4C_1$ conformation found in solution to $^4H_3$ and ^4E ring puckers. This is done by the active site nucleophile Glu212 that pulls the C2 hydroxyl, thus distorting the −1 ring to a ^4E conformer that is uniquely stable within the CBH1 reaction pocket.

This frontier orbital analysis of the reaction presents novel electronic detail of the Glu-X-Asp-X-X-Glu motif's role in the progress of the reaction. While the roles of Glu212 and Glu217 have been previously correctly guessed, here we show that the Asp214 accelerates the reaction by injecting electron density into the ring oxygen, thereby relaxing the double bond character making the carbocation more readily amenable to bond formation with Glu212.

At its pinnacle, this reaction requires 17.48 kcal/mol to overcome the barrier to glycosylation. As Glu212 covalently bonds to the anomeric position, its influence on the C2 hydroxyl ceases, and so the $^4C_1$ conformer is preferred. We calculate the rate of this part of the cellobiohydrolase reaction (taking barrier recrossing into consideration) to be 0.415 s^{-1}.

ASSOCIATED CONTENT

Supporting Information. Details of electronic structure calculations, SCC-DFTB setup, protein setup, rates calculation, selected coordinate files. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

Kevin.Naidoo@uct.ac.za

ACKNOWLEDGMENT

This work is based upon research supported by the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation (to K.J.N.). C.B.B. and K.A.W. thank SARChI for doctoral and postdoctoral fellowship support.

REFERENCES

(1) Yang, B.; Willies, D. M.; Wyman, C. E. *Biotechnol. Bioeng.* **2006**, *94*, 1122–1128.
(2) Hall, M.; Bansal, P.; Lee, J. H.; Realff, M. J.; Bommarius, A. S. *FEBS J.* **2010**, *277*, 1571–1582.
(3) Ding, S. Y.; Himmel, M. E. *J. Agric. Food Chem.* **2006**, *54*, 597–606.
(4) 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.
(5) Vrsanska, M.; Biely, P. *Carbohydr. Res.* **1992**, *227*, 19–27.
(6) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L. T.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *J. Phys. Chem. B* **2010**, *114*, 1447–1453.
(7) Divne, C.; Stahlberg, J.; Reinikainen, T.; Ruohonen, L.; Pettersson, G.; Knowles, J.; Teeri, T.; Jones, T. *Science* **1994**, *265*, 524–528.
(8) Divne, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309–325.
(9) Kleman-Leyer, K. M.; Siika-Aho, M.; Teeri, T. T.; Kirk, T. K. *Appl. Environ. Microb* **1996**, *62*, 2883–2887.
(10) Mulakala, C.; Reilly, P. J. *Proteins* **2005**, *60*, 598–605.
(11) Zhong, L.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L., III; Adney, W. S.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984–1992.
(12) Nimlos, M. R.; Matthews, J. F.; Crowley, M. F.; Walker, R. C.; Chukkapalli, G.; Brady, J. W.; Adney, W. S.; Cleary, J. M.; Zhong, L.; Himmel, M. E. *Protein Eng., Des. Sel.* **2007**, *20*, 179–187.
(13) Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J. *J. Am. Chem. Soc.* **2010**, *132*, 12800–12803.
(14) Ståhlberg, J.; Divne, C.; Koivula, A.; Piens, K.; Claeyssens, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1996**, *264*, 337–349.
(15) Strümpfer, J.; Naidoo, K. J. *J. Comput. Chem.* **2010**, *31*, 308–316.
(16) Barnett, C. B.; Naidoo, K. J. *Mol. Phys.* **2009**, *107*, 1243–1250.
(17) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(18) Kumar, S.; Payne, P. W.; Vasquez, M. *J. Comput. Chem.* **1996**, *17*, 1269–1275.
(19) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339–1350.
(20) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1988**, *61*, 2635–2638.
(21) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
(22) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569–585.
(23) Barnett, C. B.; Naidoo, K. J. *J. Phys. Chem. B* **2010**, *114*, 17142–17154.
(24) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.
(25) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175–180.
(26) Perezjorda, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134–137.
(27) Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316–325.
(28) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149–5155.
(29) Wu, Q.; Yang, W. T. *J. Chem. Phys.* **2002**, *116*, 515–524.
(30) Liu, H. Y.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Hermans, J.; Yang, W. T. *Proteins: Struct., Funct., Genet.* **2001**, *44*, 484–489.
(31) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861–10873.
(32) Nielsen, J. E. *J. Mol. Graphics* **2007**, *25*, 691–699.
(33) Nielsen, J. E.; McCammon, J. A. *Protein Sci.* **2003**, *12*, 1894–901.
(34) Sondergaard, C. R.; McIntosh, L. P.; Pollastri, G.; Nielsen, J. E. *J. Mol. Biol.* **2008**, *376*, 269–87.
(35) Gao, J.; Amata, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A.* **1998**, *102*, 4714–4721.
(36) Kuttel, M. M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.
(37) Ciccotti, G.; Ferrario, M.; Laria, D.; Kapral, R. In *Progress of Computational Physics of Matter: Methods, Software and Applications*; Reatto, L., Manghi, F., Eds.; World Scientific: River Edge, NJ, 1995; pp 150−190.
(38) Davies, G. J.; Ducros, V. M. A.; Varrot, A.; Zechel, D. L. *Biochem. Soc. Trans.* **2003**, *31*, 523–527.
(39) Fushinobu, S.; Mertz, B.; Hill, A. D.; Hidaka, M.; Kitaokac, M.; Reilly, P. J. *Carbohydr. Res.* **2008**, *343*, 1023–1033.

(40) Schramm, V. L. *Curr. Opin. Struct. Biol.* **2005**, *15*, 604–613.

(41) Schramm, V. L.; Shi, W. X. *Curr. Opin. Struct. Biol.* **2001**, *11*, 657–665.

(42) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107–115.

(43) Siegbahn, P. E. M.; Borowski, T. *Acc. Chem. Res.* **2006**, *39*, 729–738.

(44) Wolfenden, R.; Snider, M. J. *Acc. Chem. Res.* **2001**, *34*, 938–945.

(45) Wolfenden, R.; Lu, X. D.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 6814–6815.