Volume 1 | Number 1 | Jan 2013 | Pages 1–100

PCCP

Physical Chemistry Chemical Physics
www.rsc.org/pccp

ROYAL SOCIETY OF CHEMISTRY

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

The Impact of Active Site Protonation on Substrate Ring Conformation in *Melanocarpus albomyces* Cellobiohydrolase Cel7B

Timothy C. Schutt [‡]*, Vivek S. Bharadwaj* [‡]*, David M. Granum, C. Mark Maupin**

‡These authors contributed equally

*To whom correspondence should be addressed

Chemical and Biological Engineering Department, Colorado School of Mines, 1500 Illinois Street, Golden, Colorado, 80401, USA.

ABSTRACT The ability to utilize biomass as a feedstock for liquid fuel and value-added chemicals is dependent on the efficient and economic utilization of lignin, hemicellulose, and cellulose. In current bioreactors, cellulases are used to convert crystalline and amorphous cellulose to smaller oligomers and eventually glucose by means of cellulase enzymes. A critical component of the enzyme catalyzed hydrolysis reaction is the degree to which the enzyme can facilitate substrate ring deformation from the chair to a more catalytically active conformation (*e.g.* skewed boat) at the -1 subsite. Presented here is an evaluation of the protonation state for critical active site residues (*i.e.* Glu212, Asp214, Glu217, and His228) in *Melanocarpus Albomyces* (*Ma*) Cellobiohydrolase Cel7B on the substrate's orientation and ring conformation. It is found that the protonation state of the active site can disrupt the intra-enzyme hydrogen bonding network and enhance the sampling of various ring puckering conformations for the substrate ring at the +1 and -1 subsites. In particular it is observed that the protonation state of Asp214 dictates the accessibility of the glycosidic bond to the catalytic acid/base Glu217 by influencing the φ/ψ dihedral angles and the puckering of the ring structure. The protonation-orientation-conformation analysis has revealed an active site that primarily utilizes two highly coupled protonation schemes; one protonation scheme to orient the substrate and generate catalytically favorable substrate geometries and ring puckering conformations and another protonation scheme to hydrolyze the glycosidic bond. In addition to identifying how enzymes utilize protonation state to manipulate substrate geometry, this study identifies possible directions for improving catalytic activity through protein engineering.

1. Introduction

Cellulose is the most abundant polymer on earth and thus unlocking its potential is key to a renewable and sustainable feedstock for the production of liquid fuels and value-added chemicals.[1] However, breaking down cellulose to cellobiose and glucose is hindered by its inherent recalcitrant nature.[2] The recalcitrance of lignocellulosic material primarily arises from its crystalline structure, strong inter- and intra-molecular hydrogen bonding (H-bonding), and the surrounding protective hemicellulose and lignin components.[3] Degradation of cellulose typically relies on solvents and/or high temperatures to overcome H-bonding between sheets as well as inorganic catalysts[4], supercritical water[5], pyrolysis, or gasification[6] to break the strong β-glycosidic linkages. Enzymatic methods for the deconstruction of cellulosic biomass operate at lower temperatures than chemical methods due to the limitations imposed by the use of proteins.[1,7] Cellulases obtained from the fungal species *Trichoderma reesei* are considered to be some of the best enzymes for cellulosic deconstruction, both from an enzyme production and enzyme activity standpoint.[8] However, a more detailed understanding of protein-protein and protein-substrate interactions is needed to improve the catalytic activity and operating constraints of cellulase enzymes *via* protein engineering. An enhanced understanding of the complex enzymes is crucial to unlocking the potential of lignocellulosic material as a renewable and sustainable feedstock for our future fuel and chemical requirements. For example, directed mutations on cellulases have resulted in improved thermostability, thereby enabling enzymatic hydrolysis of cellulose at higher temperatures, which has significant advantages in an industrial setting since mass transport and kinetics are enhanced with increasing temperature. Advancements in understanding how protonation state affects substrate conformation will be of interest not only to the cellulose community, but also for renewable energy, enzymology, and carbohydrate science.

While experimental efforts have been used to optimize the activity, cloning, and expression of cellulases[9], computational modeling and simulation studies have proven to be insightful in revealing, at a molecular level, why particular cellulase enzyme systems achieve higher catalytic activity than others[10] in addition to revealing insights into the substrate conformational preferences in glycosidases.[11] Analyses of the electrostatic

energies[12], threading[13], and procession[14] of enzyme-substrate systems have led to increased understanding of the kinetic factors of enzymatic digestion. Fully atomistic simulation studies have been used to investigate the flexibility and processivity[15-17], active site pK_a values[15], hydrogen bonding and substrate binding in the catalytic groove[18,19], product inhibition characteristics, and thermostability[20,21] for several cellulase systems and other related enzymes[22]. *Ma* Cel7B cellobiohydrolases are known to catalyze hydrolysis by a double displacement net retaining mechanism and therefore the ring conformation at the -1 subsite is instrumental in facilitating the hydrolysis reaction as observed by QM/MM calculations.[23-26] Experimental X-ray data has revealed that the sugar in the -1 subsite adopts a $^{2,5}B$ (boat) and/or a $^{2}S_0$ (skewed boat) conformation.[27,28] The distorted ring at the -1 subsite is hypothesized to facilitate the simultaneous proton transfer from Glu217 to the glycosidic bond oxygen and the nucleophilic attack on the sugar ring at the -1 subsite by Glu212.[29] However, there is no consensus in literature whether the distortion of the ring conformation is due to steric or electrostatic interactions. The molecular dynamics studies presented here reveal how protonation states of the active site residues, and the resulting shifts in the electrostatic environment in addition to the surrounding H-bonding environment, impacts the substrate orientation and ring puckering conformation. In addition, two dominant protonation states are identified where the first protonation state orients the substrate and induces a distorted ring conformation and the second protonation state, which is catalytically primed with a proton on Glu217, initiates the hydrolysis reaction. The detailed knowledge pertaining to the role of specific residues on substrate conformation is needed to help inform directed mutagenesis for developing improved enzymes with greater stability and overall activity.

2. Methods

2.1 Docking of Substrate to *Ma* Cel7B

The systems evaluated in this study were created from the crystal structure of the cellobiohydrolase I from Melanocarpus albomyces (*Ma* Cel7B; PDB ID: 2RFW).[30] To obtain the substrate-enzyme complex the approach outlined in our previous work was implemented.[15,17] In brief, a 10-glucose oligomer was used as

the flexible ligand and docked to the rigid *Ma* Cel7B crystal structure using the AutoDock 4.2 software program.[31] The docked conformatons with the lowest energies were then compared to the aligned structure of *Ma* Cel7B and the crystal structure for *Trichoderma reesei* (*Tr*) Cel7A (PDB ID: 7CEL)[32], which was co-crystalized with a 9-glucose oligomer in the active site. *Tr* Cel7A provides a check for the resulting docked structures due to the fact that both enzymes have extensive structural similarity resulting in good alignment. The docked substrate configuration with the best agreement to the substrate docked to *Tr* Cel7A was then selected for further simulations.

2.2 Constant pH MD Simulations

Constant pH MD (CpHMD) simulations of the substrate bound *Ma* Cel7B were previously conducted to evaluate the active site residue pK_a values and enzyme dynamics.[15,17] Those simulations are further evaluated here with respect to the substrate ring conformation in the active site for pH values of 3, 6, and 8. Provided here is only a brief description of the CpHMD simulations since the methodologies[15,17,33] and theoretical background[34,35] have been reported previously. The CpHMD simulation methodology enables the coupling of protein dynamic motions with the ionization state of protonatable amino acids at a given environmental pH all in the framework of a coupled molecular dynamics and Monte Carlo algorithm. The Metropolis accept/reject criterion is based on the protonation transition free energy, which is given by

$$\Delta G = k_B T \left(pH - pK_{a,ref} \right) \ln 10 + \Delta G_e - \Delta G_{e,ref} \qquad (1)$$

where k_B is the Boltzmann constant, T is the temperature in Kelvin, pH is the specific solvent pH, $pK_{a,ref}$ is the pK_a of the reference coumpound, ΔG_e is the electrostatic component of the protonation transition energy, and $\Delta G_{e,ref}$ is the electrostatic component of the protonation transition energy for the reference compound.

In the current simulations the Monte Carlo algorithm was used every two time steps to sample the protonation states of 17 ionizable residues (*i.e.* D35, Y109, Y145, Y171, D173, D179, K181, D198, E212, D214, E217, H228, Y247, E262, E368, E367, and Y370) lining the active site tunnel. The *Ma* Cel7B enzyme was described by the parm99SB force field[36], the cellulose oligomer was described by the Glycam06 force field[37], and the ionizable residues were described by the constant pH force field. For the CpHMD

5

simulations the solvent was modeled with a Generalized-Born (GB) implicit solvent model with a salt concentration of 0.1 M and a nonbonded cutoff of 30 Å as implemented in the AMBER 12 package.[34,35,38,39] The substrate docked systems were initially minimized for 4000 steps using the steepest descent algorithm followed by 5 ns of equilibration and a production run of 70 ns. For all calculations the temperture was set at 300K, the SHAKE algorithm[40] was used to restrain bonds containing hydrogen atoms, and the global timestep was set to 2 fs within the leapfrog integrator.

Table 1 Energetically feasible protonation schemes of active site residues.

	SysA	SysB	SysC	SysD	SysE
Optimal pH*	> 7	4.9	4.8	< 4	4.3
Probability at pH 6	54.6%	20.3%	0.2%	15.0%	9.9%

Residue	SysA	SysB	SysC	SysD	SysE
GLU 212	-	-	-	-	-
ASP 214	H+	-	H+	-	-
GLU 217	-	H+	-	H+	-
HIS 228	-	-	H+	H+	H+

*The optimal pH was determined by evaluating the probability of each protonation scheme as a function of pH using the modified Henderson-Hasselbalch (Hill's) equation with the pK_a values and Hill coefficients reported in Granum *et al.*[15] The probabilities of each protonation scheme at pH 6 are reported as observed from the CpHMD simulations.

2.3 Classical MD Simulations

The classical, fixed protonation state, Molecular Dynamics (MD) simulations were performed using the AMBER12 package[38] and the aforementioned (section 2.1) substrate bound *Ma* Cel7B system. The enzyme-substrate system was initially solvated with waters described by the TIP3P force field[41], which resulted in a total system with dimension of 71 Å x 71 Å x 71 Å. The fixed protonation state systems were created based on the five most probable protonation states as determined from the CpHMD simulations at pH=6 (Table 1).[15] In all protonation schemes the Glu212 remains unprotonated, which is in line with its role as the nucleophilic residue during hydrolysis.[42] The five systems depicting the different protonation schemes (*i.e.* SysA, SysB, SysC, SysD, and SysE) were subjected to 2000 steps of steepest descent minimization followed by

2000 steps of conjugated gradient minimization. The minimized systems were then equilibrated in the isobaric isothermal (NPT) ensemble for 5 ns at a pressure of 1 atm. The equilibrated systems were then run for 100 ns in the canonical ensemble (NVT), which constituted the production runs. The SHAKE algorithm was used to restrict bonds containing hydrogens and a global timestep of 2 fs was used within the leapfrog integration. All simulations were conducted at a temperature of 300 K and used the Langevin thermostat with a collision frequency of 2 ps^{-1}. In addition to the enzyme-substrate systems, a system containing a fully solvated substrate molecule (*i.e.* 10-glucose oligomer) in TIP3P water was also simulated following the protocol outlined above. Analysis of the entire canonical ensemble trajectories was performed using ptraj[38] and the Visual Molecular Dynamics[43] (VMD) software packages. For the H-bond analysis the bond distance between donor and acceptor was set to < 3 Å and the bond angle to < 20°. The Cremer Pople[44] analysis was used to evaluate the ring puckering and utilized an in-house code.

2.4 Molecular Mechanical Poisson-Boltzmann Solvent Accessibility Calculations

In order to estimate the overall binding energies and residue-substrate interactions, the MD trajectories were also post-processed using the molecular mechanics Poisson-Boltzmann solvent accessibility (MM/PBSA) method as implementation in AMBER12.[45,46] The analysis was conducted using 5,000 frames taken at intervals of 20 ps, and the binding free energies were estimated by using

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand}) \qquad (2)$$

where $G_{complex}$, $G_{receptor}$ and G_{ligand} are the free energies for the 10-glucose oligomer when bound to *Ma* Cel7B, isolated *Ma* Cel7B, and isolated 10-glucose oligomer, respectively. The free energies were estimated according to

$$G = E_{bond} + E_{vdw} + E_{elec} + G_{PB} + G_{SA} - TS_s \qquad (3)$$

where E_{bond} is the energy contribution from bonds angles and dihedrals, E_{vdw} is the van der Waals energy, E_{elec} is the electrostatic energy, G_{PB} and G_{SA} are the polar and non-polar energy contributions, T is the temperature in Kelvin, and S_S is the solute entropy. A physiological salt concentration of 0.15 M and an internal dielectric of 4 was employed for the MM/PBSA calculations.[47] In addition to the substrate binding

energy decomposition analyses was also conducted, which provides insight into specific residue-substrate interactions.[48-50] A total of 29 residues were included in the decomposition analysis, which consisted of the 17 protonatable amino acids from the CpHMD simulations (section 2.2) as well as all other nearby residues potentially involved in enzyme-substrate interactions as identified by Parkkinen *et al.*[30]

3. Results and Discussion

Comparing the different protonation schemes of *Ma* Cel7B indicates interesting trends in the substrate conformation, H-bonding network, accessibility of the glycosidic bond to the acid/base residue Glu217, substrate binding free energy, and protein dynamics (*i.e.* loop fluctuations), which are discussed in detail in the subsequent sections.

3.1 Conformational Analysis of Substrate

The Cremer-Pople analysis[44] describes the different conformations for 6 membered rings using the three polar variables, θ, φ, and Q. The angles θ and φ are generalized puckering angles calculated from the ring co-
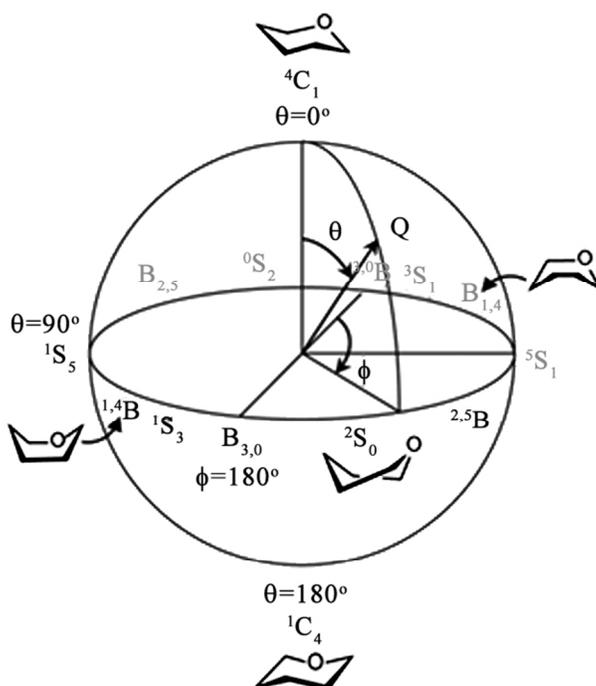
Figure 1 Diagram showing the general mapping of the Cremer Pople ring puckering characterization scheme. Figure adapted from Davies *et al.*[11]

ordinates and the Q value is the amplitude of puckering representing the extent to which a particular conformation is exhibited [44]. For the Cremer-Pople plots the chair conformation is characterized by θ values around 0° and 180° while the boat and skewed boat conformations are present when θ approaches 90°. The skewed boat and boat conformations are alternatingly characterized by φ values every 30° starting with $^{3,0}B$ at $\varphi = 0°$.[44] Figure 1 illustrates the common ring structures and their notations mapped on to the θ, φ coordinate space.[11]

The puckering of the substrate rings within the cellulase's catalytic groove has been observed experimentally
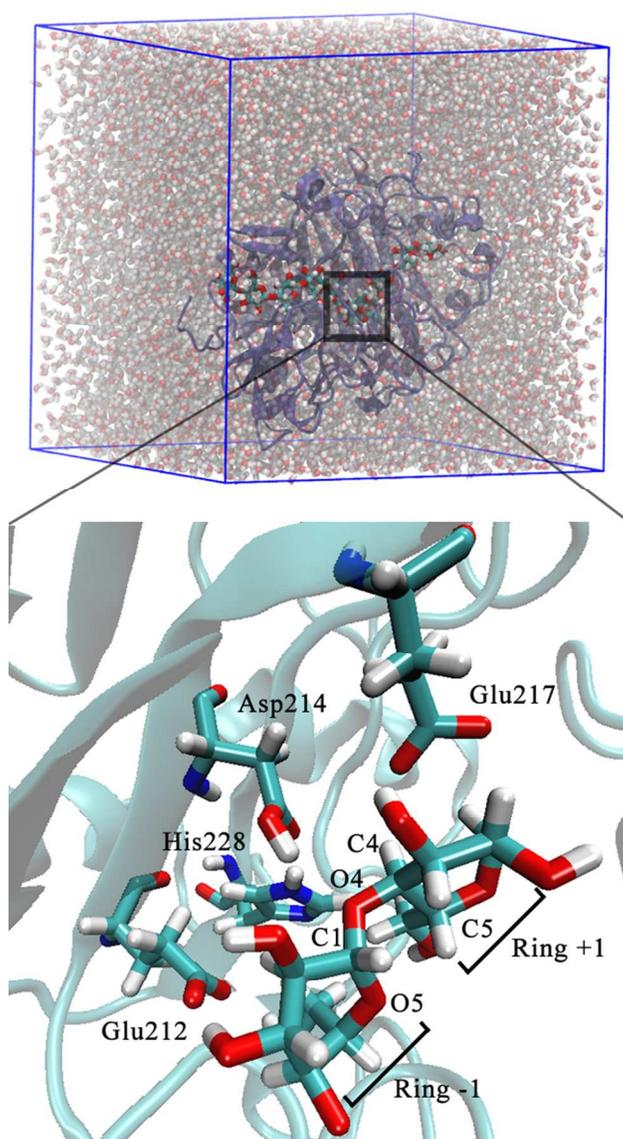
Figure 2 Sugar structure for a cellobiose fragment bound in the active site.

and computationally[27,29], with the glucan ring at the -1 subsite adopting the skewed boat conformation. It has been hypothesized that the skewed conformation of the sugar ring results in the glycosidic bond being more easily accessible to the proton donating acid/base residue Glu217[51], thus priming the glycosidic bond for catalysis. Although much is known about which residues stabilize the substrate in the active site (Figure 2), and facilitate catalysis, the causal factors driving the ring transformation are not yet known. Here we evaluate the impact of the protonation states of important active site residues on the substrate ring conformation. The Cremer Pople analysis on the two substrate rings on either side of the glycosidic bond to be hydrolyzed (*i.e.* +1 and -1 subsites) along with the φ/ψ dihedral plots around the glycosidic bond for the 5 energetically accessible protonation schemes are illustrated in Figure 3. The analysis of a 10-glucan chain in water with no enzyme is also presented as the control system.

From the Cremer Pople plots, it is evident that there are significant changes in the sugar conformation between the different protonation schemes. As expected, the control system almost exclusively exhibited the chair conformation, which is the most energetically favorable conformation for glucose in water.[52] The φ/ψ plot for the control also identifies the primary occupancy around φ=-90° and ψ=-90° with some sampling around φ=-90° and ψ=90°, which is consistent with the energetic analysis of solvated sugars conducted by Bharadwaj *et al.*[53] It is observed that the substrate ring at the +1 subsite exhibits conformations shifted towards φ=180°, which corresponds to $B_{3,O}$, in SysA, SysB, SysD, and SysE. However, SysC does not exhibit a boat conformation for the substrate ring at the +1 subsite, which may be due to the shifted ring conformation distribution at the -1 subsite, the overall active site charge distribution, and/or the more open conformation for loops I and V. These various observations will be explored in more detail in the subsequent sections of this manuscript.

The conformation of the substrate ring at the -1 and +1 subsite are of the most interest since the glycosidic bond to be hydrolyzed resides between these two rings. It has been suggested in literature that the reaction path for Cel7B hydrolysis requires the substrate ring at the -1 subsite to adapt the boat conformation as a precursor to the subsequent hydrolysis of the glycosidic bond.[25] Simulation of the various protonation
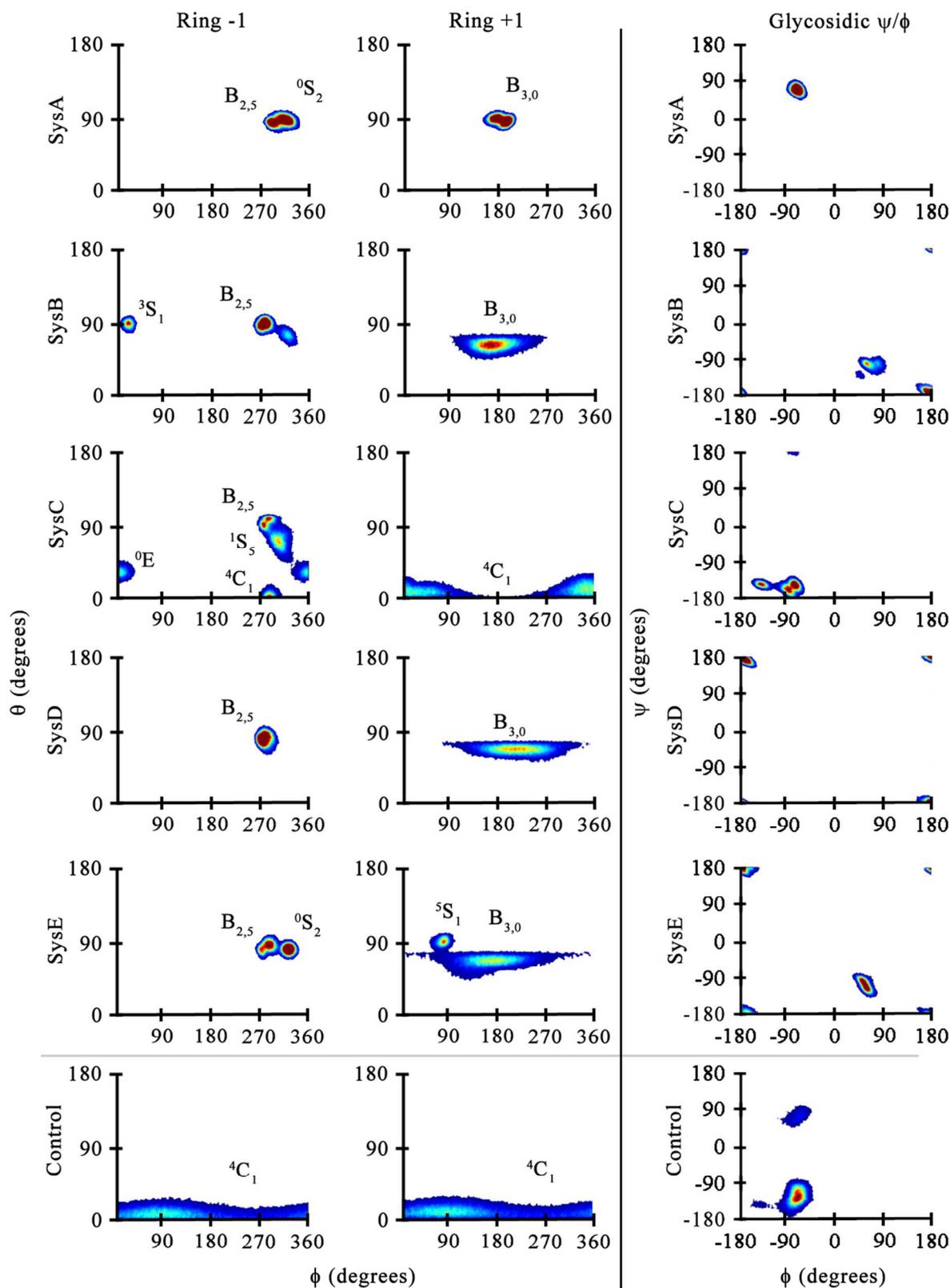
Figure 3 Cremer Pople Analysis of cellulose substrate inside the active site tunnel for each proto-
nation scheme and for the control system (*i.e.* fully solvated substrate).

schemes clearly indicates that the substrate ring at the +1 subsite adopts the $B_{3,0}$ conformation and the sub-

strate ring at the -1 subsite adopts the $B_{2,5}$ conformation (Figure 3), which agrees well with previous QM/MM studies of the reaction pathway.[29] This pair of boat structures places the hydroxymethyl O6 from the substrate ring at the -1 subsite within H-bonding distance of the nucleophilic Glu212, which in turn draws the hydroxymethyl group further away from the glycosidic bond. The φ/ψ of the glycosidic bond is found to be around $\varphi=180°$ and $\psi=180°$, which is energetically unfavorable in an aqueous solvation environment[53], and results in a bent glucan chain that is postulated to make the glycosidic bond easier to break.

For the substrate ring at the -1 subsite, SysC is the only protonation scheme that doesn't exhibit exclusively a skewed boat or boat conformation, and is found to sample the chair and envelope conformations. This observation is of import because it illuminates the role of specific residue's protonation states on ring conformation stability, and points to the role of electrostatics in dictating the adopted ring conformation for the substrate ring at the -1 and +1 subsites. Analysis of the various protonation schemes reveals that placing the right amount of protons in the correct position (*i.e.* protonated Asp214 and His228) the electrostatic environment, and to some extent the H-bonding environment around Asp214, can be manipulated to partially stabilize the chair conformation. Clearly, if the ring conformation was solely dictated by steric interactions the ability to manipulate ring conformations by adjusting protonation states would not be possible. It is interesting to note that SysC, based on geometric analysis, is the least conducive for catalysis amongst the five analyzed protonation schemes, and it was reported to be the highest in free energy (*i.e.* least probable) of the 5 protonation schemes as determined from CpHMD simulations.[15] This energetic observation points to the optimization of the active site residue's pK_a values such that the substrate orientation and ring conformation has been optimally refined for catalysis at a pH of 6.

3.2 MM/PBSA Decomposition Binding Free Energy

Table 2 Interaction energy from MM/PBSA between residue and substrate in kcal/mol and (std. dev.)

Resid	SysA	SysB	SysC	SysD	SysE
Glu212	-19(1)	-16(3)	-15(2)	-15(4)	-18(2)
Asp214	-2(2)	-5(3)	0(1)	-13(3)	-10(2)

Glu217	-15(3)	-4(2)	-8(5)	-4(1)	-9(3)
His228	0(1)	-2(1)	0(2)	-1(2)	0(2)
Asp262	-12(5)	-12(3)	-12(3)	-12(3)	-1(5)

Changes in the individual contributions to substrate binding at the active site, calculated using MM/PBSA indicates that Glu212 is the strongest stabilizing interaction between residues and substrate, and that Glu217 has reduced favorable interactions with the substrate when it is protonated (SysB and SysD in Table 2) as compared to when it is unprotonated. While Glu217 needs to be protonated to initiate the hydrolysis reaction, the partial negative charge of the unprotonated Glu217 is found to be more favorable for binding the substrate, especially in the case of SysA. The dominant role of Glu212 is partially due to the hydroxyl hydrogen off of the sugar methyl group, which H-bonds to Glu212. This H-bond interaction is found to hold the sugar in a twisted orientation such that the glycosidic bond is accessible to the proton donating residue Glu217. This observation would indicate that mutating Glu212 to Asp212 may result in decreased substrate binding due to an increase in the distance between the residue and the substrate. Also, the direct effect of protonating Asp214 on substrate binding can be seen when Asp214 is protonated, as in SysA and SysC, where Asp214 has very little interactions with the substrate. The last residue of the coupled active site, His228, is found to have very minor interactions with the substrate, irrespective of its protonation state. This minor contribution in stabilizing substrate binding is to be expected since His228 is situated on the perimeter of the active site, although His228 does play a critical role, when protonated, as it can donate a proton to the nearby Asp214 that can then donate the proton to Glu217. The importance of the auxiliary residues Asp214 and His228 were also highlighted by Yongchao et al.[54] who found for a Family 9 cellulase that these residues play an important role in a H-bonded network with the catalytic acid/base residue.

An encouraging point of validation for the MM/PBSA analysis is the identification of Asp262 in stabilizing substrate binding. It has been noted[20] that Asp262 interacts with the substrate ring at the +1 subsite, which plays a role in the release of the cellobiose product and is important for the processive motion of the enzyme. The role of Asp262 is linked to the binding of the cellobiose product, and is believed to contribute to the

product inhibition characteristics of the enzyme. This is in agreement with the presented MM/PBSA results that indicate Asp262 as a major contributor to substrate binding.

3.3 Hydrogen Bond Analysis

Table 3 Hydrogen bonding[*] with the -1 glucan ring site.

Hydrogen Bonding to Ring -1					
H-Bond	**SysA**	**SysB**	**SysC**	**SysD**	**SysE**
Asp214 O to Glu217 O	8%	100%	20%	0%	3%
Ring +1 O3 to Glu217 O	64%	38%	54%	69%	46%
Ring -1 O3 to Glu212 O	50%	68%	71%	55%	81%
Ring -1 O6 to Glu212 O	51%	68%	31%	78%	65%
Ring -1 O2 to Glu212 O	0%	0%	21%	0%	0%
Specific Zones of SysC:	$^1S_5$	$B_{2,5}$	^0E	$^4C_1$	
Ring -1 O3 to Glu212 O	72%	73%	66%	73%	
Ring -1 O6 to Glu212 O	9%	62%	37%	0%	
Ring -1 O2 to Glu212 O	42%	0%	0%	58%	

[*]H-bonding is considered to occur only when the angle is less than 20° and the distance is less than 3 Å.

The observed variations in the ring conformations amongst the different protonation schemes could be due to varying enzyme-substrate interactions dictated by the differing charged environments. In order to elucidate how the specific enzyme-substrate interactions change between the various protonation schemes, H-bond analysis was performed. Enzyme-substrate H-bond analysis has identified specific hydrogen bonds that contribute to the stabilization of the substrate and may influence the conformation of the substrate's ring at the -1 subsite. SysC, being the only protonation scheme in which the sugar at the -1 subsite did not completely shift to the boat conformation reveals significant differences in the H-bonding percentage with Glu212 as compared to the other systems evaluated in this study, Table 3. The second Oxygen (O2) of the substrate ring at the -1 subsite was observed to H-bond with Glu212 for around half of the simulation time replacing the typical hydroxymethyl oxygen (O6) interaction with Glu212. To investigate if the changes observed in H-

14

bonding correlate with the change in ring conformation, the trajectory for SysC was parsed into four data sets based on the distinct regions in the Cremer Pople diagram and then the H-bonds were reanalyzed for each sugar conformation. It is found that when the sugar ring is in the chair conformation, Glu212 H-bonds with the O2 atom of substrate's ring at the -1 subsite, and not with the sugar's O6 atom. Surprisingly, the skewed conformation also exhibited a hydrogen bond between Glu212 and O2 of the substrate's ring at the -1 subsite, while the purported catalytically active substrate conformation, $B_{2,5}$, as well as the observed envelope intermediate conformations did not exhibit the O2-Glu212 H-bond. The $B_{2,5}$ and envelope intermediate conformations did possess H-bonding interactions similar to the other four protonation schemes, namely a H-bond between the O3 in the substrate's ring at the -1 subsite and Glu212. It is hypothesized that less time spent in the boat conformation will result in an overall decreased activity as ring distortion is hypothesized to lower the activation energy for the hydrolysis of the glycosidic bond. This result is not too surprising given that SysC is favored by conditions below the optimal pH, perhaps in part indicating why *Ma* Cel7B is deactivated in relatively acidic solutions. In addition, the ability of SysC to stabilize the formation of the chair conformation indicates that protonation states of the active site play a major role in the enzyme's ability to manipulate the ring puckering of the substrate.

The percentage of time spent in the skewed ($80° < \theta < 100°$) conformation for each substrate ring observed from CpHMD simulations, which enabled the dynamic protonation of all ionizable residues in the active site tunnel, is displayed in Figure 4. Most notably, at the pH of 6 the system exhibits the highest percentage of skewed conformations for the substrate's ring at the -1 subsite. In the constant protonation state simulations SysA, SysB, SysD, and SysE exclusively sampled the skewed conformation for the substrate's ring at the -1 subsite whereas with dynamic protonation, the skewed conformation was observed for only 52% of the simulation time. Together these observations further support the hypothesis that the active site protonation strongly impacts substrate ring conformations. The lower probability of skewed boat occurrences in CpHMD, as opposed to the fixed protonation MD, is due to the dynamic nature of the active site protonation states. In the case of the CpHMD, the active site dynamically shifts between many protonation schemes with
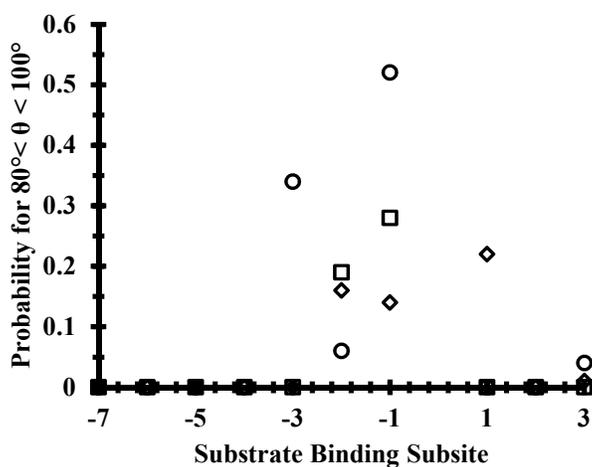
Figure 4 Percentage of time spent in skewed conformation of each ring at different pH's during dynamic protonation simulations (CpHMD). The three pH values depicted are pH=3(□), pH=6(○), and pH=8 (◇).

the five presented in Table 1 being the most commonly sampled schemes. In contrast, the fixed protonation state simulations give only a response of the substrate to an ideal fixed protonation scheme. Therefore, one would expect that the CpHMD's substrate would spend a certain amount of time responding to the changing environment resulting in an overall lower percentage of time spent occupying one particular ring conformation.

Using the pK_a and Hill coefficient values determined from the CpHMD titration simulations[15], the modified Henderson-Hasselbalch (Hill's) equation was used to calculate the trends in probability of a given protonation scheme for various pH values (Figure 5). The values from Hill's equation are meant to illustrate relative trends in the probability of each protonation scheme over a continuous pH range in contrast to the discrete point pH values observed from CpHMD. Inspection of Figure 5 reveals a cross over point between SysA and SysB just below a pH of 6, which is the optimal pH for *Ma* Cel7B.[55] The protonation scheme in SysA, which is the most conducive to orienting the substrate's ring at the -1 subsite into a favorable geometry for the hydrolysis reaction, is found to be in balance with protonation scheme in SysB, which corresponds to the protonated acid/base residue. The protonated acid/base residue is a key prerequisite for catalysis because this residue donates its excess proton to the substrate during the hydrolysis reaction;[29] thus when Glu217

16

is protonated it will be referred to as 'primed' for catalysis. This interplay between SysA and SysB refines

our earlier hypothesis on the coupled active site residues (*i.e.* Glu212, Asp214, Glu217, and His228) by in-

corporating the substrate geometry and ring conformations.[15,17] The current hypothesis now focuses on SysA

and SysB as the dominant protonation schemes, which favorably orients the substrate (SysA) and then trans-

fers a proton from Asp214 to Glu217 to prime the catalytic acid/base residue (SysB). As seen in the H-

bonding analysis between Asp214 and Glu217 presented in Table 3, it is observed that the extra proton

on Glu217 is H-bonded with the Asp214 residue during the entire simulation for SysB. However, in

SysA the Asp214 and Glu217 residues were observed to H-bond only 8% of the simulation time. In both

SysA and SysB an ample number of opportunities to transfer a proton between one another were ob-

served. It is noted that H-bonding between Glu217 and the substrate was seen to decrease in SysB as

compared to that of SysA.

3.4 Interatomic Distance Analysis

Ab initio simulation studies have revealed that the Glu212 residue attacks the carbon at the base of
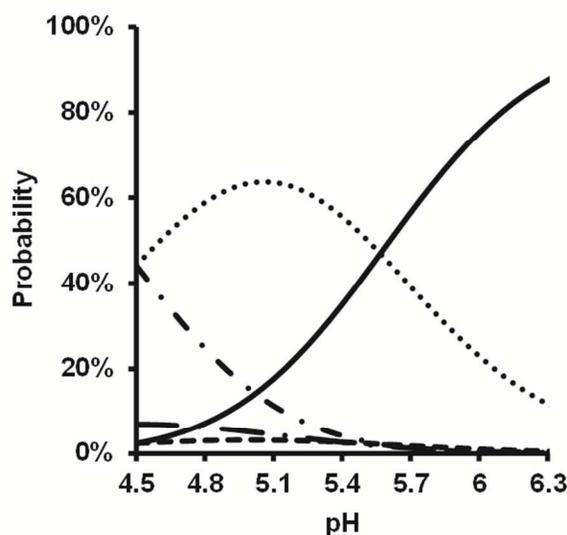
Figure 5 Probability of each protonation scheme as a function of pH using the modified Henderson-Hasselbalch (Hill's) equation. The graphs depicts SysA (——), SysB (······), SysC (— —), SysD (— · —), and SysE (— · · —) over the relevant pH around the optimal enzyme pH.
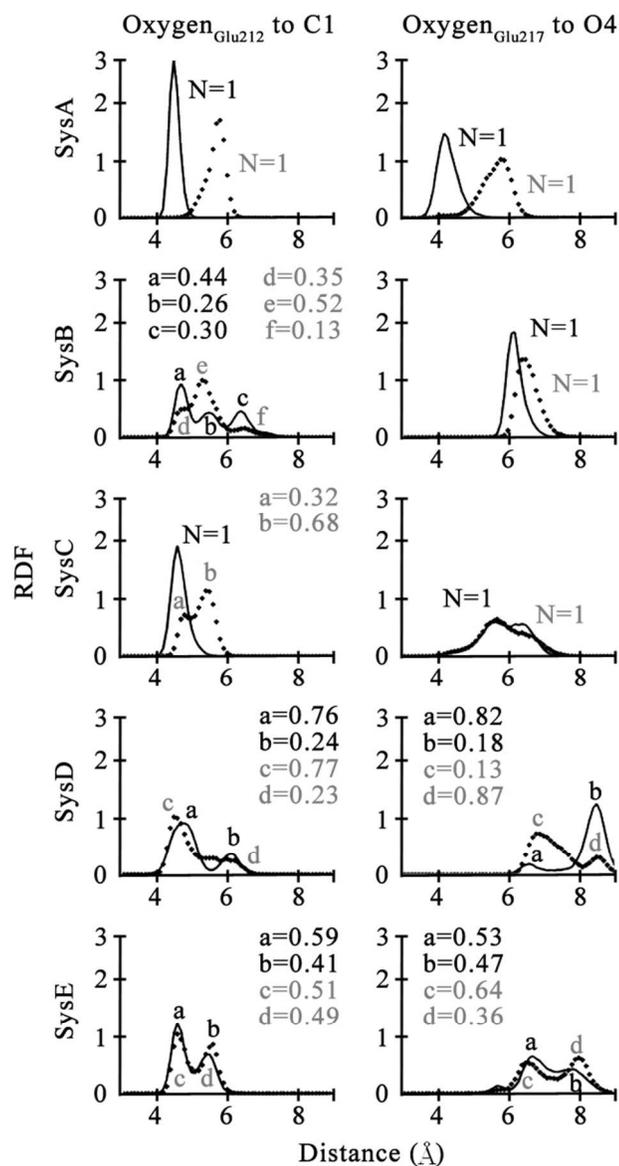
Figure 6 Interatomic distance plots for Glu212 and Glu217 OE1 oxygen (dotted) and OE2 oxygen (solid) to the C1 on the substrate and glycosidic oxygen respectively for each protonation scheme. Peaks are labeled with letters and the corresponding coordination numbers are reported using $N = 4\pi\rho\int_{r_1}^{r_2} r^2 g(r)dr$.

the glycosidic bond on the substrate's ring at the -1 subsite (C1), thereby facilitating proton transfer from

Glu217 to the glycosidic oxygen.[56] Interatomic distance analyses were used to evaluate accessibility of the glycosidic bond to the active site acid/base residue and the nucleophile (Figure 6). The protonation scheme in SysA was found to bring the Glu217 oxygens closer to the glycosidic bond than any of the other protonation schemes. This increased accessibility presumably decreases the activation energy for hydrogen transfer from the acid/base residue to the glycosidic bond thus favorably impacting the hydrolysis reaction. Interestingly, SysA also constrains the conformation of the substrate's ring at the +1 subsite (*i.e.* closest ring to Glu217) to the $B_{3,O}$ conformation. It is found that SysA creates a more tightly bound substrate (Table 2) that
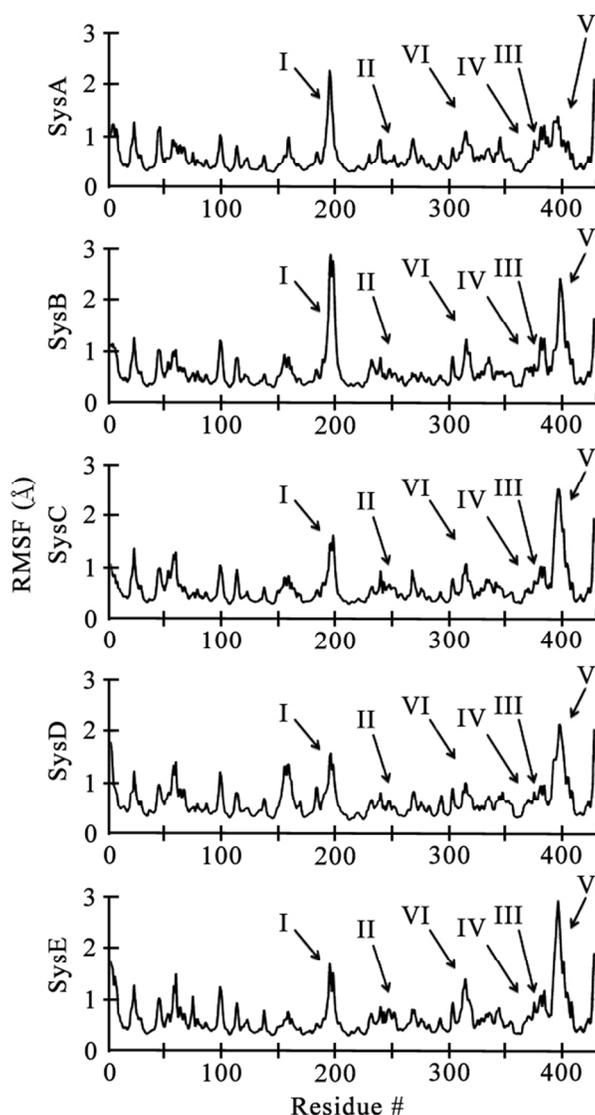
Figure 7   The RMSF of all 5 protonation schemes over all the residues of *Ma* Cel7B.

19

also displays more favorable distance distributions between substrate and both the nucleophile Glu212 and the acid/base Glu217. This is in contrast to the other four protonation schemes, which only loosely constrain the conformation of the substrate's ring at the +1 subsite and sample a wide range of φ values. These observations support the hypothesis that SysA is the most conducive for setting up the substrate for the hydrolysis reaction prior to the hydrogen transfer from the Asp214 to Glu217 and ultimately to the glycosidic bond. In SysA, the interatomic distances also reveal that the C1 carbon atom is not bifurcated by the two oxygens of Glu212. Although either oxygen may interact with C1, both oxygens do not simultaneously interact with C1. The interatomic distances clearly indicate that one oxygen is closer (~5 Å) and points towards the glycosidic bond while the other oxygen is further away (~7 Å). For the protonation schemes where Glu217 is protonated (*i.e.* in SysB and SysD) it is found that the distance between Glu217 and the glycosidic bond reside at a greater distance (>6 Å) indicating that the hydrogen resides between the Glu217 oxygen and the glycosidic bond (*i.e.* H-bonded). The interatomic distances for SysA are found to possess the closest interaction distance between the substrate's glycosidic bond and Glu217, which is either the result or the cause of a tight φ/ψ dihedral distribution for the glycosidic bond (φ=-90°, ψ=90°) that exposes the linkage to Glu217. Based on our previous CpHMD and kinetic modeling studies, it is hypothesized that SysA, which is the lowest energy protonation scheme and therefore the most common protonation scheme with substrate bound at pH=6, represents the initial stage of the hydrolysis reaction. Starting from SysA there would then be a proton transfer from the Asp214 to Glu217 that would prime the enzyme for the hydrolysis reaction. The proton transfer scheme proposed by Granum *et al.*[15] accounts for this behavior as well as dynamic fluctuations of the enzyme.

3.5 Root Mean Square Fluctuations

An analysis of the root mean square fluctuations (RMSF) is an important way to link the impact of the protonation scheme to the overall protein dynamics, which has been hypothesized to correlate to enzymatic activity.[17] The RMSF in Figure 7 shows that loop regions enclosing the active site tunnel *i.e.* residue numbers 189-206 (loop I) and 387-409 (loop V) are significantly impacted by

System	Distance (Å)
A	22(4)
B	21(2)
C	30(4)
D	18(3)
E	28(4)
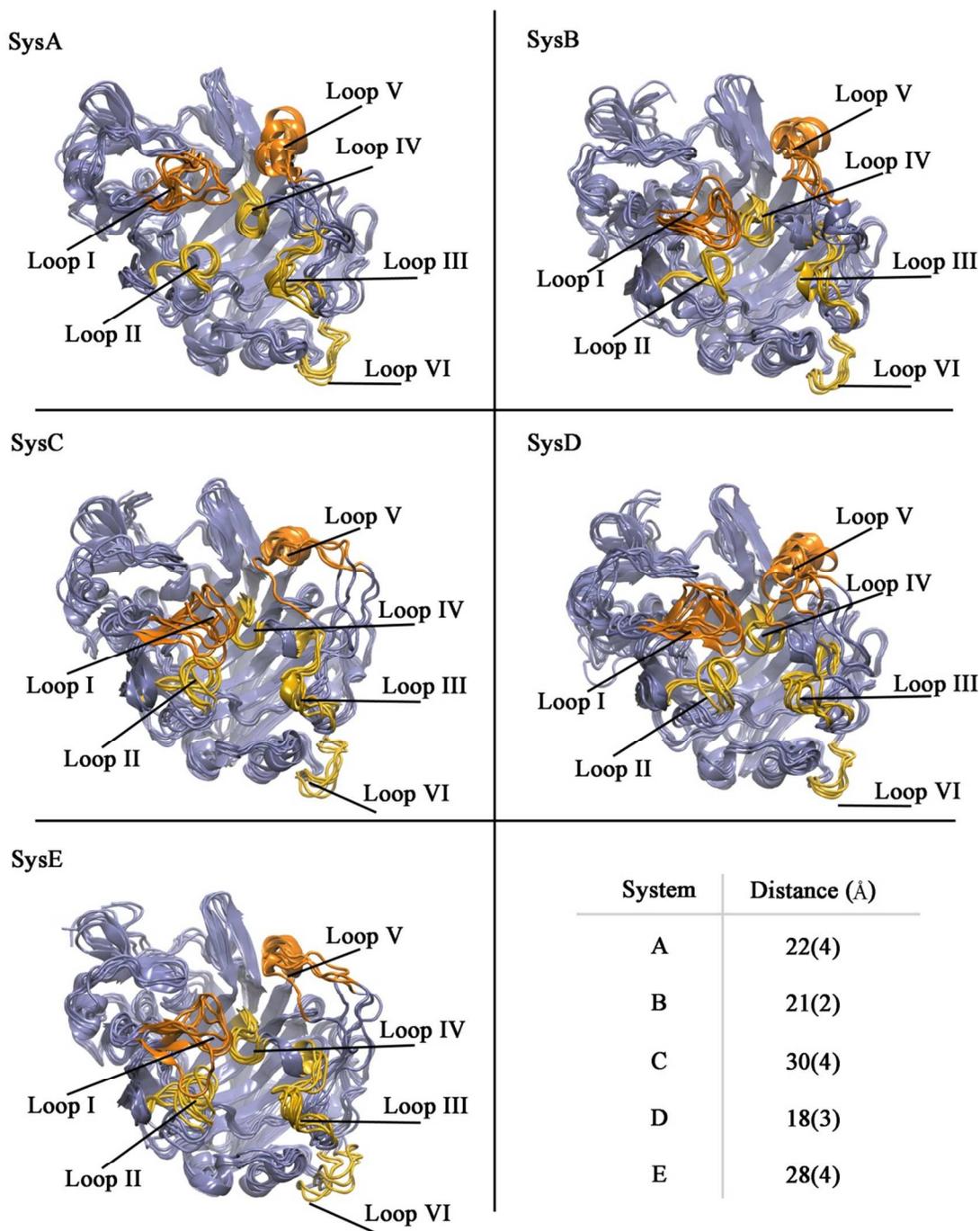
Figure 8 *Ma* cellobiohydrolase with 10 glucose oligomer chain in the active tunnel with distal loop regions displayed in yellow showing the relative changes in conformation throughout the simulation.

the protonation scheme in the active site.[17] Loops I and V are the two loops residing over the top of the active site tunnel and are involved in stabilizing bound substrate through various non-covalent interactions in addition to their position and flexibility (*i.e.* clamping down). The fluctuations for SysA and SysB are found

21

to be the highest in the regions around loop I, whereas SysA has very low fluctuations of the loop V region. Fluctuations of loops I and V are thought to be related to processivity and binding of the substrate in the active site groove.

An interesting trend observed in the simulations is that loop V possesses reduced flexibility for SysA as compared to the other four schemes. Fluctuations in the substrate itself are also smallest in SysA with the glycosidic bond fluctuating by only half an angstrom (data not shown). Based on the observations of the RMSF, the loops enclosing the active site (*i.e.* loop V and loop I) are found to modulate the overall character of the active site from groove to tunnel. The closing of the loops (*i.e.* clamping down) has been linked to the protonation of residues in the interior of the enzyme and an overall stabilization of the bound substrate. From a molecular-level this is observed by a reduced RMSF for the loop regions and a smaller separation distance between the respective loops.[17] As a metric for how closed or open these two loops are, the distance between the alpha carbons of residues 195 and 397 was measured for each simulation frame. The choice of residue 195 and 397 was made because these are the closest two residues residing on either distal loop. Analysis of this distance reveals that SysC and SysE are significantly more open with distances of 29.6 ± 4 Å and 28.4 ± 4 Å, respectively, as compared to an average of 20.7 ± 3 Å for the other three systems. Figure 8 highlights visually, the position of these loops relative to the catalytic tunnel of *Ma* Cel7B occupied by a 10-glucan cellulose oligomer as well as illustrates the conformational change in the shape of loop V visible in SysC and SysE related to the opening of the loops.

The fluctuation of loops I and V have been found to be partially dependent upon the protonation scheme of the active site as determined by the charge coupling analysis from CpHMD simulations of *Ma* Cel7B when substrate was bound in the active site tunnel.[15] From the CpHMD and fixed protonation state MD it is observed that the presence of substrate disrupts the native H-bonding network and causes a dramatic increase in the enzymes overall flexibility, which is most pronounced at pH = 6. The trends observed by Granum *et al.*[15,17] are also observed in the current simulations where relatively high flexibilities of loop I is observed for SysA and SysB. Experimentally, the optimal pH for *Ma* Cel7B is around 6, which corresponds to SysA and

SysB (Figure 5) while SysC and SysD represent a lower (<4.8) pH scenario and exhibit a decreased flexibility in loop I.[55] This reduced flexibility in loop I at lower pH values is due in part to the aforementioned charge coupling stabilization, which is consistent with the Cremer Pople results as well as the MM/PBSA results discussed previously.

4. Conclusions:

This study illustrates how the active site protonation schemes in *Ma* Cel7B cellobiohydrolase impact the sugar ring conformation of the bound substrate as well as φ/ψ dihedrals of the glycosidic linkage, which have significant implications for the catalytic reaction. It is found that the lowest energy protonation scheme of the active site, SysA, favorably orients the substrate so as to expose the glycosidic bond to Glu217 while simultaneously orienting the substrate's C1 in proximity to Glu212. Exposure of the glycosidic bond in SysA was achieved primarily through bending of the sugar chain by means of the φ/ψ dihedral angles around the glycosidic bond. In addition, the ring conformation is found to also aid accessibility of the glycosidic oxygen to the acid/base residue as evidenced in interatomic distance and Cremer Pople analyses. The final pre-catalytic step would then contain the single proton transfer event from Asp214 to Glu217 (SysA to SysB), which would catalytically prime the enzyme. Further evaluation of the system with respect to relatively low pH values has indicated that the H-bonding network around the active site is disrupted by acidic conditions, which may explain the decreased activity of *Ma* Cel7B in solutions of pH less than 6.

To improve the overall activity of *Ma* Cel7B enzymes the mutation of Asp262 to Ala262 is hypothesized to help reduce product inhibition by encouraging release of the cellobiose product after hydrolysis as seen both in the MM/PBSA results and the H-bond analysis, and as previously suggested by Silveria *et al.*[20] The fluctuation and shape of the enclosing loops around the active tunnel were also found to be correlated to substrate binding energy in addition to the effective active site pH. The linking of the active site protonation scheme to the substrate orientation in the active site and ring conformation has expanded our understanding of the active site for glycosidases and how they utilize ionizable residues during catalysis. Utilizing the con-

clusions from Granum *et al.*[15,17] and the insights from this work, it is proposed that mutations removed from the active site, and in the vicinity of His228, would be ideal to shift the optimal pH of the enzyme while still retaining the ability of the active site to favorably manipulate the substrate's ring conformation. The results presented here further our understanding of protein-protein and protein-substrate interactions for glycosidases in particular and enzymes in general in addition to providing crucial information pertaining to biomass processing studies.

AUTHOR INFORMATION

Corresponding Author

*Email: cmmaupin@mines.edu. Phone: 303-273-3197. Fax: 303-273-3730

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ABBREVIATIONS

Ma = *Melanocarpus Albomyces*, MD = Molecular Dynamics, MMPBSA = Molecular Mechanics/Poisson Boltzman Surface Area, CpHMD = Constant pH Molecular Dynamics

Physical Chemistry Chemical Physics Accepted Manuscript

REFERENCES

(1)   Monserrate, E.; Leschine, S. B.; Canale-Parola, E. *International Journal of Systematic and Evolutionary Microbiology* **2001**, *51*, 123.

(2)   Langan, P.; Gnanakaran, S.; Rector, K. D.; Pawley, N.; Fox, D. T.; Cho, D. W.; Hammel, K. E. *Energy & Environmental Science* **2011**, *4*, 3820.

(3)   Lee, D.-S.; Wi, S. G.; Lee, S. J.; Lee, Y.-G.; Kim, Y.-S.; Bae, H.-J. *Bioresource Technology* **2014**, *158*, 239.

(4)   Qi, X.; Watanabe, M.; Aida, T. M.; Smith Jr, R. L. *Green Chemistry* **2009**, *11*, 1327.

(5)   Sasaki, M.; Fang, Z.; Fukushima, Y.; Adschiri, T.; Arai, K. *Industrial & Engineering Chemistry Research* **2000**, *39*, 2883.

(6)   Chundawat, S. P. S.; Beckham, G. T.; Himmel, M. E.; Dale, B. E. *Annual Review of Chemical and Biomolecular Engineering* **2011**, *2*, 121.

(7)   Carere, C.; Sparling, R.; Cicek, N.; Levin, D. *International Journal of Molecular Sciences* **2008**, *9*, 1342.

(8)   Gusakov, A. V. *Trends in Biotechnology* **2011**, *29*, 419.

(9)   Li, D.-C.; Li, A.-N.; Papageorgiou, A. C. *Enzyme Research* **2011**, *2011*, 9.

(10)   Bhat, K. M.; Gaikwad, J. S.; Maheshwari, R. *Journal of General Microbiology* **1993**, *139*, 2825.

(11)   Davies, G. J.; Planas, A.; Rovira, C. *Accounts of chemical research* **2011**, *45*, 308.

(12)   Zhang, Z.; Zheng, Y.; Petukh, M.; Pegg, A.; Ikeguchi, Y.; Alexov, E. *PLOS Comp. Biol.* **2013**, *9*, e1002924.

(13)   Lin, Y.; Beckham, G. T.; Himmel, M. E.; Chu, J. *J. Phys. Chem. B.* **2013**, *117*.

(14)   Igarashi, K.; Koivula, A.; Wada, M.; Kimura, S.; Penttila, M.; Samejima, M. *J. Biol. Chem.* **2009**, *284*.

(15)   Granum, D. M.; Vyas, S.; Sambasivarao, S. V.; Maupin, C. M. *The Journal of Physical Chemistry B* **2013**, *118*, 434.

(16)   Knott, B. C.; Crowley, M. F.; Himmel, M. E.; Ståhlberg, J.; Beckham, G. T. *Journal of the American Chemical Society* **2014**, *136*, 8810.

(17)   Granum, D. M.; Schutt, T. C.; Maupin, C. M. *The Journal of Physical Chemistry B* **2014**, *118*, 5340.

(18)   Wahlström, R.; Rahikainen, J.; Kruus, K.; Suurnäkki, A. *Biotechnology and Bioengineering* **2014**, *111*, 726.

(19)   Sambasivarao, S. V.; Granum, D. M.; Wang, H.; Maupin, C. M. *AIMS Molecular Science* **2014**, *1*, 59.

(20)   Silveira, R. L.; Skaf, M. S. *The Journal of Physical Chemistry B* **2014**.

(21)   Zhang, S.; Wang, Y.; Song, X.; Hong, J.; Zhang, Y.; Yao, L. *Journal of chemical information and modeling* **2014**, *54*, 2826.

(22)   Zhang, Z.; Teng, S.; Wang, L.; Schwartz, C. E.; Alexov, E. *Hum. Mutat.* **2010**, *31*, 1043.

(23)   Petersen, L.; Ardèvol, A.; Rovira, C.; Reilly, P. J. *The Journal of Physical Chemistry B* **2009**, *113*, 7331.

(24)   Saharay, M.; Guo, H.-B.; Smith, J. C.; Guo, H.; Oak Ridge National Laboratory (ORNL); Vol. 1052.

(25)   Biarnes, X.; Ardevol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *Journal of the American Chemical Society* **2007**, *129*, 10686.

(26)   Jitonnom, J.; Lee, V. S.; Nimmanpipug, P.; Rowlands, H. A.; Mulholland, A. J. *Biochemistry* **2011**, *50*, 4697.

(27)   Guérin, D. M. A.; Lascombe, M.-B.; Costabel, M.; Souchon, H.; Lamzin, V.; Béguin, P.; Alzari, P. M. *Journal of molecular biology* **2002**, *316*, 1061.

(28)   Zou, J.-y.; Kleywegt, G. J.; Ståhlberg, J.; Driguez, H.; Nerinckx, W.; Claeyssens, M.; Koivula, A.; Teeri, T. T.; Jones, T. A. *Structure* **1999**, *7*, 1035.

(29)   Yan, S.; Li, T.; Yao, L. *The Journal of Physical Chemistry B* **2011**, *115*, 4982.

(30)   Parkkinen, T.; Koivula, A.; Vehmaanperä, J.; Rouvinen, J. *Protein Science* **2008**, *17*, 1383.

(31)   Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. *Journal of computational chemistry* **2009**, *30*, 2785.

(32)   Divne, C.; Stahlberg, J.; Teeri, T. T.; Jones, T. A. *Journal of molecular biology* **1998**, *275*, 309.

(33)   McDougal, O. M.; Granum, D. M.; Swartz, M.; Rohleder, C.; Maupin, C. M. *J Phys Chem B* **2013**, *117*, 2653.

(34)   Mongan, J.; Case, D. A.; McCammon, J. A. *J Comput Chem* **2004**, *25*, 2038.

(35)   Mongan, J.; Case, D. A. *Curr Opin Struct Biol* **2005**, *15*, 157.

(36)   Wang, J. M.; Cieplak, P.; Kollman, P. A. *Journal of computational chemistry* **2000**, *21*, 1049.

(37)   Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *Journal of computational chemistry* **2008**, *29*, 622.

(38)   D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman; University of California: San Fransisco, 2012.

(39)   Onufriev, A.; Case, D. A.; Ullmann, G. M. *Biochemistry* **2001**, *40*, 3413.

(40)   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *Journal of Computational Physics* **1977**, *23*, 327.

(41)   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *The Journal of Chemical Physics* **1983**, *79*, 926.

(42)   MacKenzie, L. F.; Sulzenbacher, G.; Divne, C.; Jones, T. A.; Wöldike, H. F.; Schülein, M.; Withers, S. G.; Davies, G. J. *Biochemical Journal* **1998**, *335*, 409.

(43)   Humphrey, W.; Dalke, A.; Schulten, K. *Journal of Molecular Graphics & Modelling* **1996**, *14*, 33.

(44)   Cremer, D.; Pople, J. A. *Journal of the American Chemical Society* **1975**, *97*, 1354.

(45)   Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2012**, *8*, 3314.

(46)   Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S. H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. *Accounts of Chemical Research* **2000**, *33*, 889.

(47)   Hou, T.; Wang, J.; Li, Y.; Wang, W. *Journal of Chemical Information and Modeling* **2010**, *51*, 69.

(48)   Bharadwaj, V. S.; Dean, A. M.; Maupin, C. M. *Journal of the American Chemical Society* **2013**, *135*, 12279.

(49)   Zoete, V.; Meuwly, M.; Karplus, M. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 79.

(50)   Cao, R. Y.; Jin, Y. D.; Xu, D. G. *J. Phys. Chem. B* **2012**, *116*, 6087.

(51)   Voutilainen, S. P.; Boer, H.; Alapuranen, M.; Jänis, J.; Vehmaanperä, J.; Koivula, A. *Applied microbiology and biotechnology* **2009**, *83*, 261.

(52)   Polavarapu, P. L.; Ewig, C. S. *Journal of computational chemistry* **1992**, *13*, 1255.

(53)   Bharadwaj, V.; Schutt, T.; Ashurst, T.; Maupin, C. M. *PCCP* **2015**, *(Accepted)*.

(54)   Li, Y.; Irwin, D. C.; Wilson, D. B. *Applied and environmental microbiology* **2007**, *73*, 3165.

(55)   Haakana, H.; Miettinen-Oinonen, A.; Joutsjoki, V.; Mäntylä, A.; Suominen, P.; Vehmaanperä, J. *Enzyme and Microbial Technology* **2004**, *34*, 159.

(56)   Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J. *Journal of the American Chemical Society* **2011**, *133*, 19474.

TOC