Journal of Structural Biology 191 (2015) 352–364

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Short-time dynamics of pH-dependent conformation and substrate binding in the active site of beta-glucosidases: A computational study

David F. Flannelly [a], Thalia G. Aoki [b], Ludmilla Aristilde [a,b,*]

[a] The Institute for Comparative and Environmental Toxicology, College of Agricultural and Life Sciences, Cornell University, Ithaca, NY 14853, USA
[b] Department of Biological and Environmental Engineering, College of Agricultural and Life Sciences, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form 17 June 2015
Accepted 3 July 2015
Available online 6 July 2015

Keywords:
Cellulose degradation
Enzyme catalysis
pH effect
Molecular dynamics

ABSTRACT

The complete degradation of cellulose to glucose is essential to carbon turnover in terrestrial ecosystems and to engineered biofuel production. A rate-limiting step in this pathway is catalyzed by beta-glucosidase (BG) enzymes, which convert cellulobiose into two glucose molecules. The activity of these enzymes has been shown to vary with solution pH. However, it is not well understood how pH influences the enzyme conformation required for catalytic action on the substrate. A structural understanding of this pH effect is important for predicting shifts in BG activity in bioreactors and environmental matrices, in addition to informing targeted protein engineering. Here we applied molecular dynamics simulations to explore conformational and substrate binding dynamics in two well-characterized BGs of bacterial (*Clostridium cellulovorans*) and fungal (*Trichoderma reesei*) origins as a function of pH. The enzymes were simulated in an explicit solvated environment, with NaCl as electrolytes, at their prominent ionization states obtained at pH 5, 6, 7, and 7.5. Our findings indicated that pH-dependent changes in the ionization states of non-catalytic residues localized outside of the immediate active site led to pH-dependent disruption of the active site conformation. This disruption interferes with favorable H-bonding interactions with catalytic residues required to initiate catalysis on the substrate. We also identified specific non-catalytic residues that are involved in stabilizing the substrate at the optimal pH for enzyme activity. The simulations further revealed the dynamics of water-bridging interactions both outside and inside the substrate binding cleft during structural changes in the enzyme-substrate complex. These findings provide new structural insights into the pH-dependent substrate binding specificity in BGs.

1. Introduction

Beta-glucosidases (BGs) catalyze the cleavage of the 1–4 beta-linkage of cellulobiose (or cellobiose), a glucose dimer, to produce two glucose molecules (Singhania et al., 2013). This cleavage is a rate limiting step in the complete degradation of cellulose to glucose (Singhania et al., 2013; Jeng et al., 2011). A comprehensive understanding of the functional dynamics of BGs under different aqueous conditions is needed to predict the contributions of these enzymes to terrestrial carbon fluxes (Wieder et al., 2013; Knight and Dick, 2004; Mariscal-Sancho et al., 2010) and in engineered biofuel production from lignocellulosic wastes (Singhania et al., 2013; Teugjas and Väljamäe, 2013; Jørgensen et al., 2007; Percival Zhang et al., 2006). Both in natural environments and in engineered bioreactors, BGs are exposed to various aqueous pH values, ranging from acidic to alkaline conditions (Zimmerman et al., 2011; Acosta-Martínez and Tabatabai, 2000; Lauber et al., 2009; Lan et al., 2013). Because enzymes are comprised of ionizable amino acid residues, pH impacts their conformation, structural stability, and catalytic activity (Creighton et al., 1993).

A large diversity of pH–activity profiles exists for BGs of different origins, and even within the same species (Schomburg et al., 2013). For instance, maximum enzymatic activities of BGs from the anaerobic bacterium *Clostridium cellulovorans* (Jeng et al., 2011), the fungi *Aspergillus aculeatus* (Murao et al., 1988) and *Trichoderma reesei* (Jeng et al., 2011), and the insect *Chilo suppressalis* (Zibaee et al., 2009) occur, respectively, at pH values of 6, 3, 6, and 9. A mechanistic basis for the relationship between enzyme structure and pH-induced effects on BG activity is lacking. In the present study, we seek to explore this relationship in two enzymes belonging to the family 1 of BGs (Jeng et al., 2011), one from the bacterium *C. cellulovorans* and one from the fungus *T. reesei*, which have well defined crystal structures. Both of these enzymes have

* Corresponding author at: Department of Biological and Environmental Engineering, College of Agricultural and Life Sciences, Cornell University, 214 Riley-Robb Hall, Ithaca, NY 14853, USA.
E-mail address: ludmilla@cornell.edu (L. Aristilde).

http://dx.doi.org/10.1016/j.jsb.2015.07.002
1047-8477/

been shown experimentally to exhibit optimal catalytic activity at pH 6, with decreased activity at higher and lower pH values (Jeng et al., 2011). The catalytic action on the substrate in family 1 BGs has been well characterized but the structural dynamics underlying the pH-dependent substrate specificity have not been fully examined.

Catalysis in family 1 BGs is mediated by two Glu residues: an acidic/basic Glu residue and a nucleophilic Glu residue (Scheme 1). Protonation of the linking C1-C1′ O atom (O1) of the substrate by the acidic/basic Glu residue ($Glu166_{bacterial}$/$Glu165_{fungal}$) sets the stage for a nucleophilic attack by the other catalytic Glu residue ($Glu352_{bacterial}$/$Glu367_{fungal}$) on the C1 of the substrate (Jeng et al., 2011; Vuong and Wilson, 2010; Badieyan et al., 2012) (Scheme 1; Fig. 1). A transition state facilitates the splitting of the C1–C1′ link while retaining a covalent bond with the nucleophilic Glu (Vuong and Wilson, 2010) (Scheme 1). The reaction is completed by the release of the cleaved part of the substrate following the de-glycosylation step (Scheme 1). The initiation of the catalytic action is mediated by a H-bonding interaction between the substrate O1 and the acidic/basic Glu (Jeng et al., 2011; Badieyan et al., 2012). Thus, the disruption of this H-bonding interaction would impede forward catalytic steps. The effect of pH on this key interaction has not been determined.

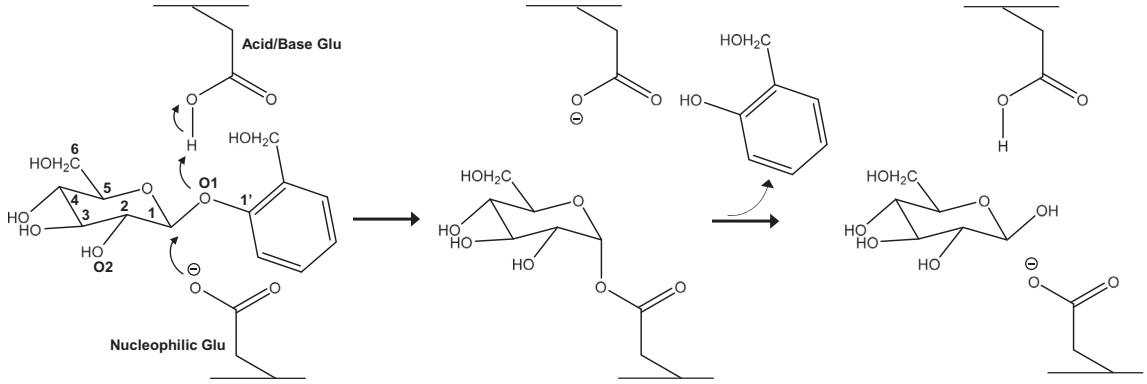
The nucleophilic Glu residue plays an integral role in stabilizing the intermediate state in the hydrolysis reaction (Vuong and Wilson, 2010) but its role in substrate binding is unresolved. In structures of BGs co-crystallized with substrates in the active site (Jeng et al., 2012; Chuenchor et al., 2011), the nucleophilic Glu has adopted an orientation that can facilitate the formation of a H-bond with the H atom connected to the O2 (i.e. H_{O2}) of the substrate. This H-bond could thus aid in orienting the substrate towards a favorable interaction with the catalytic acidic/basic Glu. However, due to the geometry of the nucleophilic Glu$\cdots H_{O2}$ H-bond, this bond would have to be broken prior to the nucleophilic attack by the Glu on the C1 of the substrate in the second catalytic step (Scheme 1) Withers et al., 1992. Therefore, a strong interaction between the nucleophilic Glu and H_{O2} could impede catalysis. The occurrence of this interaction during pH-dependent structural changes in the enzyme remains to be elucidated.

In addition to the catalytic Glu residues, the substrate binding cleft has several non-catalytic amino acid residues, which are highly conserved in family 1 BGs (Fig. 1): an Asn ($Asn165_{bacterial}$/$Asn164_{fungal}$) residue that immediately precedes the acidic/basic catalytic Glu residue, a Gln residue ($Gln20_{bacterial}$/$Gln16_{fungal}$) and a Glu residue ($Glu406_{bacterial}$/$Glu424_{fungal}$) that form two H-bonds with the substrate, a Tyr residue ($Tyr296_{bacterial}$/$Tyr298_{fungal}$) that participates in the removal of the bound intermediate, a His residue ($His121_{bacterial}$/$His119_{fungal}$) that mediates the initial binding of the substrate, and a Trp residue ($Trp407_{bacterial}$/$Trp425_{fungal}$) (Fig. 1) Badieyan et al., 2012. These non-catalytic residues have been evaluated for their individual energetic contributions to the catalytic hydrolysis (Badieyan et al., 2012). However, similar to the catalytic Glu residues, the involvement of the conserved non-catalytic residues in the substrate binding of family 1 BGs under unfavorable pH conditions has not yet been investigated.

It is widely accepted (Badieyan et al., 2012; Chuenchor et al., 2011; White and Rose, 1997; Vocadlo and Davies, 2008; Wang et al., 2011) that water plays an essential role in the BG hydrolysis pathway by stabilizing the active site-bound substrate and intermediate as well as participating in interactions that mediate the release of the final product. It is challenging to capture directly the dynamics of water molecules in enzyme-substrate interactions via experimental means. Consequently, the role of water in the stabilization of the substrate (Chuenchor et al., 2011) and the intermediate (Badieyan et al., 2012) has been inferred from the positioning of O atoms in X-ray crystal structures. For instance, the geometry of a BG co-crystallized with a glucose polymer suggested that amino acid residues outside of the immediate active site can interact with the substrate through water bridges (Chuenchor et al., 2011). A quantum mechanics (QM) analysis has demonstrated that water molecules are especially important in the thermodynamically-favorable binding of the intermediate during BG hydrolysis (Badieyan et al., 2012). Studies on the dynamic role of water interactions in mediating pH-dependent substrate binding are lacking.

We hypothesize that pH-dependent substrate specificity in BGs may result from three main phenomena: (1) changes in ionization states in amino acid residues in the substrate binding pocket alter substrate-binding interactions, (2) pH-induced changes in the ionization states of non-active site residues lead to conformational changes in the active site, and/or (3) changes in enzyme conformation disrupt stabilization by water interactions.

Previous mutagenesis studies have shown that non-catalytic residues of BGs and of other similar classes of enzymes can impact enzyme activity and substrate specificity (Kaper et al., 2002; Huber et al., 2001; Wang et al., 2005). In support of the second hypothesis presented above, a study on the effects of site-directed mutagenesis on the pH–activity profile of a cellulase from *T. reesei* (Wang et al., 2005) reported that mutations in certain amino acid residues on the surface of the enzyme can result in up to 0.6 unit decrease or up to a 1.4 unit increase in the optimal pH for enzyme activity. The mechanisms responsible for these pH shifts for optimal activity were not resolved. In another study on an exocellulase from *T. reesei* (Wohlfahrt et al., 2003), mutations of carboxyl-carboxyl pairs into amide-carboxyl pairs led to increased activity of the



Scheme 1. Schematic representation of the BG catalytic action on the substrate (salicin).

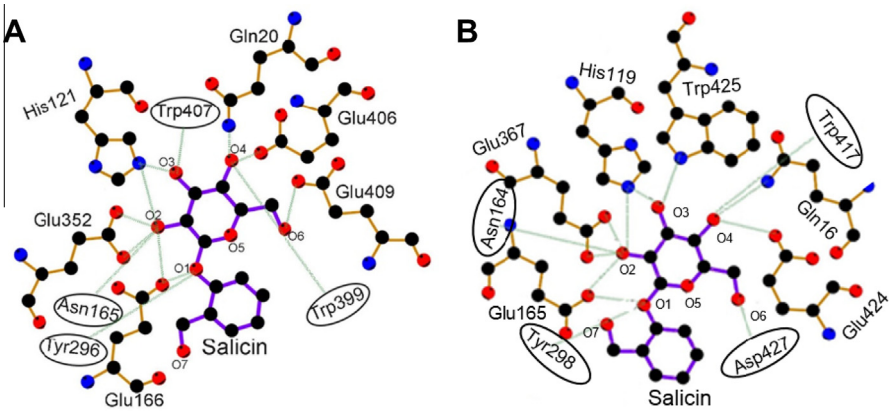


Fig. 1. Schematic diagrams of a BG active site from (A) the bacterium *Clostridium cellulovorans* and (B) from the fungus *Trichoderma reesei*. Possible H-bonds that can form between the amino acids and the substrate (salicin) are shown in green. These H-bonds were monitored for interactions throughout the course of simulation.

enzyme at higher pH (Wohlfahrt et al., 2003). It was proposed (Wohlfahrt et al., 2003) that the deprotonated carboxyl-carboxyl pairs, which would repulse each other, reduced the stability of the enzyme and impede activity at high pH; thus, substituting an amide group for one of the interacting carboxyls would reduce the repulsion, increase stability, and augment activity. The implication of this change in stability on the pH-dependent catalytic binding dynamics was not evaluated.

In silico methods, via both QM and molecular dynamics (MD) simulations, have been applied to analyze reaction mechanisms and reaction pathways in enzyme catalysis (Monard et al., 2003; Senn and Thiel, 2007; Karplus and Kuriyan, 2005). Previous QM molecular modeling of every step along the BG catalytic pathway identified specific amino acids that stabilize the substrate binding at various points during the enzymatic action (Badieyan et al., 2012; Wang et al., 2011). The influence of pH on the catalytic substrate binding in the BG active site, however, has not yet been studied. Performing MD simulations circumvents the limitation of QM studies to capture the conformational dynamics of enzyme structures as a function of changes in pH (Machuqueiro and Baptista, 2007; Bürgi et al., 2002; Langella et al., 2004; Tan et al., 2005; Machuqueiro and Baptista, 2008). For instance, a MD study (Tan et al., 2005) of a proteinase in response to different ionization states of non-catalytic residues in the active site reported pH-dependent binding site perturbations that may be responsible for the experimentally-obtained pH activity profile (Tan et al., 2005). With respect to BGs, one MD study investigated the binding dynamics of several known inhibitors whereby the inhibitors assumed different protonation states (Zhou et al., 2006). To the best of our knowledge, the structural dynamics underlying the pH-dependent BG catalytic action on the substrate have not been characterized.

Building on these previous experimental and modeling studies, the present study adopts an *in silico* approach to explore the structural factors responsible for the pH-dependent substrate binding dynamics in the two aforementioned family 1 BGs from *C. cellulovorans* and from *T. reesei*. We aim to (1) identify the active site and non-active site amino acid residues responsible for changes in the enzyme's ionization states as a function of pH, (2) determine the consequence of these changes on the favorable substrate binding required for catalysis, and (3) explore the participation of solvated waters in mediating substrate binding. To meet these objectives, we applied a methodology that combined molecular docking, energy minimization (EM), and MD algorithms, to conduct molecular simulations of the two enzymes in an explicit solvated environment at different pH values (pHs 5, 6, 7, and 7.5). We explored the dynamics of the docked substrate in the active site and monitored interactions with catalytic residues as well as non-catalytic residues (Badieyan et al., 2012). In addition, the explicit solvation approach allowed us to monitor the dynamic role of water molecules both in solvating the substrate and in bridging interactions between the substrate and the enzyme. Our MD-equilibrated structures thus provide mechanistic insights towards elucidating the link between experimentally-determined pH-dependent activity and pH-induced structural changes in BGs.

2. Computational methods

2.1. Modeling platform

Molecular simulations were performed using the forcefield CHARMm (Chemistry at HARvard Molecular Mechanics) (Brooks et al., 1983), as interfaced in Accelrys's Discovery Studio software package (Software Inc., 2013) or the open source molecular dynamic software GROMACS (GROningen MAchine for Chemical Simulations) (Hess et al., 2008). The CHARMm forcefield was previously validated (Brooks et al., 1983, 2009) for simulating the structures of peptides and proteins. Short-time (up to 2 ns) and long-time (up to 200 ns) were performed, respectively, on local Dell Precision T7610 machines and on the high-performance computer resource Stampede located on EXSEDE (Extreme Science and Engineering Discovery Environment) (Towns et al., 2014).

2.2. Validation simulations of a termite BG

As a pre-requisite to performing the simulations of the bacterial (*C. cellulovorans*; PDB ID 3AHX) (Jeng et al., 2011) and fungal (*T. reesei*; PDB ID 3AHY) (Jeng et al., 2011) BGs, we developed a methodology for docking and simulating the substrate in the active site (Fig. 2). Because no structural data are available for a substrate in the active site of the bacterial and fungal BGs of interest, we validated our methodology using the structure of another family 1 BG of termite origin (*Neotermes koshunensis*; PDB ID 3VIL) (Jeng et al., 2012); this termite BG was crystallized with a substrate surrogate (salicin) in its active site.

For the modeling validation, the ionization states of the residues of the termite enzyme were simulated at the same aqueous conditions (pH 7.25, 0.15 M NaCl) (Jeng et al., 2012) used for the protein's crystallization by using an algorithm (Spassov and Yan, 2008) as implemented in Discovery Studio (Spassov and Yan, 2008). In order to compute the prominent ionization states of each amino acid residue, this algorithm assesses the local environment of each residue with respect to other residues in the protein structure and the degree of exposure of the residue. Before the substrate docking, the residue $Glu402_{termite}$ (Jeng et al., 2012) was manually

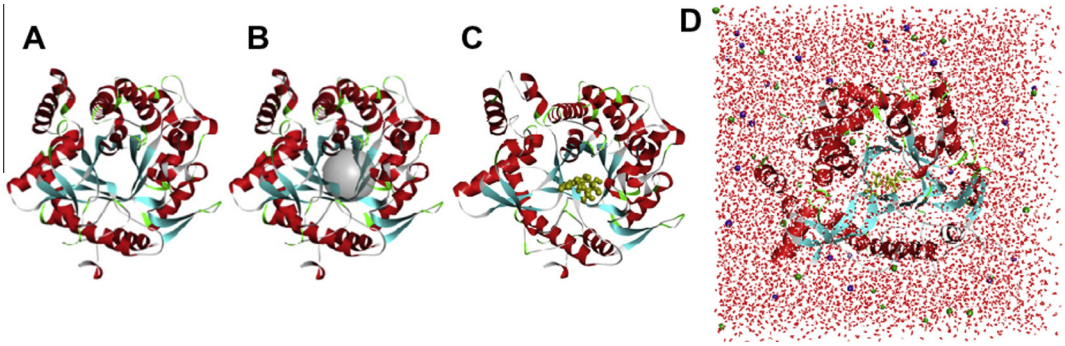


Fig. 2. Modeling workflow for obtaining preliminary structures of a BG in a hydrated environment: (A) downloaded X-ray structure of enzyme was prepared at a specific pH, (B) preliminary determination of interaction region of substrate (shown by gray sphere), (C) docking of substrate (shown in yellow), and (D) view of the hydration of system using explicit solvated water molecules in a NaCl solution (Na and Cl ions are represented as purple and green spheres, respectively). The secondary structure of the enzyme is portrayed in red for alpha helices, blue for beta-sheets, and green for loops.

deprotonated to reflect the known catalytic ionization state of the nucleophilic base (Vuong and Wilson, 2010). The substrate (salicin) was docked into the active site of the enzyme using the CHARMm docking software CDOCKER (Wu et al., 2003). The interaction sphere (radius of 7 Å) was defined to include the residues, which were shown to interact with the substrate in the crystal structure (Jeng et al., 2012). The additional docking parameters were the following: pose cluster radius = 0.1, random conformations = 10, dynamic steps = 1000, and simulated annealing heating to 700 K and then cooling to 300 K over 2000 steps. The lowest-energy configurations were compared to the atomic coordinates of the original X-ray structure (Jeng et al., 2012) whereby successful dockings were evaluated by the conventional standard of Root Means Square Deviation (RMSD) of less than 2.0 Å (Roche et al., 2001; Verdonk et al., 2003). The docked substrate had an RMSD of 0.382, which met the criteria for a successful docking (Supplementary Table S1).

To achieve the background ionic solution (0.15 M) in the solvated 3VIL protein with the docked substrate, the solvation algorithm in Discovery Studio implemented the Na and Cl ions were implemented to both balance charges on the enzyme and randomly in the solvated simulation box. The solvated system was subjected to two EM cycles (1000 maximum steps each), heated simulation (300 K for 4 ps at 1 fs time step), a MD equilibration run (for 10 ps at 2 fs time step) before the final MD run (1 ns at 2 fs time step), from which data were retrieved. The modeling parameters used in the Discovery Studio were the following: NPT (constant pressure and temperature) ensemble with a Nose–Hoover thermostat to set the temperature and a Langevin piston for the pressure simulation; cutoff for non-bonded interactions was 14 Å; bonds involving hydrogen atoms were fixed; the TIP3P water model was used with the CHARMm forcefield.

Six H-bonding interactions between the substrate and the enzyme, which were deduced from the crystal structure of the active site (Jeng et al., 2011), were monitored throughout the simulations. The MD-predicted values for these H-bond distances were within one standard deviation of the experimental values (Jeng et al., 2012) (Supplementary Table S1). Thus, though this exercise, we validated that our modeling methodology could replicate appropriate configurations of the enzyme to facilitate favorable active site-substrate interactions in an aqueous condition very similar to our systems of interest.

2.3. Molecular simulations of the bacterial and fungal BGs

Following the successful development and validation of our methodology, we employed a similar workflow to simulate the conformational and substrate binding dynamics of the bacterial and fungal BGs enzymes at the optimal pH for catalytic activity of these enzymes, pH 6, as well as at pHs 5, 7, and 7.5 (Fig. 2). After the enzymes were retrieved from the PDB database, they were prepared at each pH according to the aforementioned algorithm (Spassov and Yan, 2008). Once the most prominent protonation states of the amino acid residues were determined at each pH, they were not allowed to fluctuate over the course of the simulations while taking into account changes in intrinsic pK_a values as a function of conformational dynamics during the simulation. We acknowledge that this approach does not account for the simultaneous ensemble of (less prominent) protonation states that can exist in equilibrium at a given pH (Machuqueiro and Baptista, 2007; Bürgi et al., 2002; Machuqueiro and Baptista, 2008; Baptista et al., 2002; Dlugosz and Antosiewicz, 2004; Campos et al., 2010). The focus of our study was to simulate the enzyme at its most probable ionization state under each pH condition.

In order to model the known starting catalytic ionization states (Jeng et al., 2011), the nucleophilic Glu residue important for catalysis ($Glu352_{bacterial}/Glu367_{fungal}$) was deprotonated and the acidic/basic Glu residue ($Glu166_{bacterial}/Glu165_{fungal}$) was protonated (Teugjas and Väljamäe, 2013). Because these two enzymes were not crystallized with a substrate bound to the active site, it was necessary to obtain first the conformation required to accommodate the substrate. We achieved this conformation by overlaying the active site residues of each enzyme (3AHX and 3AHY) onto the termite (3VIL) active site, which was crystallized around a substrate analog (salicin). This superimposition was executed by creating tethered atom pairings between the atoms of the termite BG enzyme and those of each of the BG enzyme being studied. Specifically, α carbons of the 12 active site amino residues as well as the carboxylate moiety of the nucleophilic Glu were paired. Subsequently, the atom pairings were subjected to a geometry optimization step by minimizing the sum of the squared distances between all of the pairs. The initial orientation of the substrate (salicin) in both the bacterial and fungal active sites was informed by the binding orientation obtained with the crystal structures of the termite BG (Jeng et al., 2012); our simulations were performed with salicin, a commonly-used substrate analog of the natural substrate cellobiose, because experimental studies of the relationship between pH and enzymatic activity for both enzymes were conducted with salicin (Jeng et al., 2011). All systems were then solvated in a 0.1 M NaCl solution with explicit water molecules in a periodic environment. These solvated systems were subjected subsequently to a series of EM and MD series, as previously described, consisting of two minimization steps, a heating step, a preliminary MD equilibration step, followed by a 2 ns MD production step used for analysis.

Acknowledging, that longer simulations are customary in molecular dynamics studies, we have conducted longer MD simulations (up to 200 ns) for the systems. Following the 2 ns MD production step detailed above, the systems were subjected to an additional 2 ns MD equilibration step and a 200 ns MD production step. These latter simulations (time step = 2 fs) were run in a NPT ensemble using a pressure-coupled thermostat (Parrinello-Rahman barostat) and a distance of 1.4 nm for coulombic and Van der Waals interaction cutoffs. Throughout the course of the 200-ns MD run in several of the systems, we found that the substrate migrated outside of the substrate binding cleft, interacted with residues on the external periphery of the active site, and became increasingly solvated by water molecules (Supplementary Figs. S1 and S2). We attributed this phenomenon to the destabilization of unfavorable substrate binding interactions with the active site residues when the simulations are too long. Therefore, extended MD simulation time was not an appropriate approach here to examine the substrate binding structures across both favorable and unfavorable pH conditions wherein the enzyme conformations are not always optimal for substrate binding. In fact, it is this pH-dependent unfavorable substrate binding that we are investigating in this study. The analysis of the migration of the substrate away from the active site was beyond the scope of the present study but will be investigated in the future. The shorter simulation period chosen for analysis allowed us to probe specifically pH-dependent substrate binding with both catalytic and non-catalytic residues in the active site substrate remained in the active site for all the investigated pH conditions.

2.4. Analysis

Mapping of the hydrophobic and hydrophilic regions on the enzyme was carried out using the last MD-optimized configuration. Hydrophobic surface mappings were generated using Discovery Studio algorithms that calculate the hydrophobic sections of the enzyme based on the chemical properties of each residue, as well as the chemical properties of the neighboring atoms. We monitored the dynamics of H-bonding interactions between the substrate and water molecules using Discovery Studio's experimentally-validated (Baker and Hubbard, 1984; Bissantz et al., 2010) algorithm for non-bonding interactions to capture any potential H-bonding by setting a maximum H-acceptor distance of 3.1 Å and a range of donor-H-acceptor angles between several angles (all between 90 and 180 degrees). A more selective criteria (distance < 2.8 Å and angles between 120 and 180 degrees) were applied to substrate-active site interactions to screen for stronger interactions. For statistical analysis, the simulation results were divided into 200 ps segments in order to obtain the mean and standard deviation of the presence of an interaction throughout the simulation. Statistical significance was determined using an unpaired two-tailed *t*-test comparison. Co-occurrence correlations between multiple interactions were analyzed using a Pearson's test, followed by a screening for significance using a *p*-value ($p < 0.05$) adjusted for multiple hypotheses and for a correlation strength of an *R* greater than or equal to 0.15. Finally, we performed a RMSD analysis to explore the pH-dependent structural dynamics at the level of amino acid residues in the enzyme. The amino acid residues of interest were those that underwent a change in ionization states and those involved in catalytic, non-catalytic, and water-bridging interactions. For each residue of interest, we monitored the RMSD in distance from residue's starting position for each frame in the final MD run. A Pearson's correlation test of the RMSD fluctuations determined whether the residues were moving away from their original position at statistically-significant similar or distinct times. After the threshold for significance was adjusted to account for the number of correlation pairs tested, all non-significant correlations values were removed.

3. Results

3.1. pH-dependent changes in ionization states

Several amino acid residues in both enzymes underwent changes in their prominent ionization states when the enzymes were subjected to increasing pH conditions (Table 1). In the bacterial enzyme, four amino acids (one Asp and three Glu) became negatively charged and one His went from being positively-charged to being neutral from pH 5 to pH 6, one Glu residue became negatively charged and two His residues became neutral from pH 6 to pH 7, and one Glu became negatively charged and one His became neutral from pH 7 to pH 7.5 (Table 1). In the fungal enzyme, seven amino acids (four Asp and three Glu) became negatively charged from pH 5 to pH 6, one Asp became negatively charged and one His became neutral from pH 6 to pH 7, and three His residues went from being positively charged to neutral from pH 7 to pH 7.5 (Table 1). Therefore, the bacterial enzyme gained five negative charges from pH 5 to pH 6 and two additional ones from pH 6 to pH 7.5; the fungal enzyme gained 7 negative charges from pH 5 to pH 7 and five additional ones from pH 6 to pH 7.5 (Table 1).

Except for $Glu409_{bacterial}$, all the amino acid residues that underwent a change in their ionization states were located outside of the substrate-binding cleft. It is important to note, as stated in Section 1, that pH 6 is the optimal pH for enzyme activity for the two enzymes, with decreased activity recorded at pH values lower and higher than pH 6 (Jeng et al., 2011). Surface mappings of the hydrophobic and hydrophilic regions of each enzyme illustrated the consequence of the pH-dependent changes in the ionization states on the conformation of the substrate-binding cleft (Fig. 3 and Supplementary Fig. S3). Noticeably, the overall hydrophilic nature of the substrate-binding cleft persisted under all four pH conditions, which is consistent with the lack of major changes in the ionization states of active site residues at the different pH conditions. However, this did not preclude structural changes in the substrate-binding pocket, as shown in Fig. 3. These conformational changes around the active site may induce disruption of favorable substrate binding dynamics required for catalysis in the BG

Table 1
List of amino acid residues with pH-dependent ionization states in the bacterial and fungal BG enzymes.

Enzyme	Ionizable residue	pH 5	pH 6	pH 7	pH 7.5
Bacterial	Asp252	0	–	–	–
	Glu22	0	0	–	–
	Glu153	0	–	–	–
	Glu286	0	–	–	–
	Glu370	0	–	–	–
	Glu406	0	0	0	–
	His49	+	0	0	0
	His240	+	+	0	0
	His340	+	+	+	0
	His419	+	+	0	0
Fungal	Asp105	0	–	–	–
	Asp121	0	0	–	–
	Asp142	0	–	–	–
	Asp227	0	–	–	–
	Asp256	0	–	–	–
	Glu18	0	–	–	–
	Glu144	0	–	–	–
	Glu244	0	–	–	–
	His127	+	+	0	0
	His194	+	+	+	0
	His287	+	+	+	0
	His305	+	+	+	0

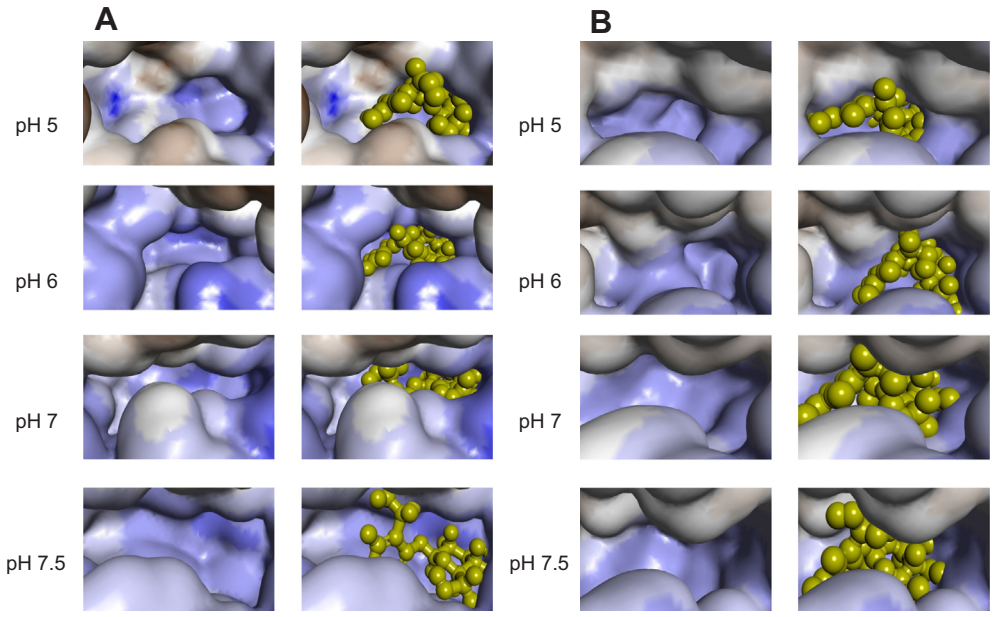


Fig. 3. A close-up snapshot of hydrophobic (brown) and hydrophilic (blue) regions of the substrate-binding pocket in MD-equilibrated configuration of the bacterial (A) and fungal (B) BGs at pH 5, pH 6, pH 7, and pH 7.5 (from top to bottom, respectively) presented without (left) and with (right) the substrate. Surface maps of hydrophobic and hydrophilic regions over the entire surface of the enzymes are shown in Supplementary Fig. S3.

enzymes. We explored these changes in binding dynamics in the following sections.

3.2. pH-induced changes in substrate interactions with catalytic and non-catalytic residues

As discussed in Section 1, the hydrolytic catalysis is initiated by an interaction between the acidic/basic Glu ($Glu166_{bacterial}/Glu165_{fungal}$) and the substrate O1 atom (Jeng et al., 2011; Badieyan et al., 2012). The simulations with the bacterial enzyme revealed that the catalytic $Glu166_{bacterial}\cdots O1$ interaction was the most abundant at pHs 6, 7, and 7.5 but occurred significantly less at pH 5 than at pH 6 (Fig. 4). We also observed that the occurrence frequency of the $H_{O2}\cdots Glu166_{bacterial}$ interaction followed a similar pattern as the catalytic interaction $Glu166_{bacterial}\cdots O1$, peaking around pH 6 and then decreasing at higher and lower pH values (Fig. 4); this trend is analogous to the experimental pH–activity profile (Jeng et al., 2011).

With the fungal enzyme, the catalytic $Glu165_{fungal}\cdots O1$ interaction in the fungal enzyme was the highest at the pH values of 6 and 7; and, the abundance of the interaction at the enzyme's optimal pH of 6 was statistically higher than at pH 5 or pH 7.5 (Fig. 5). The $H_{O2}\cdots Glu165_{fungal}$ interaction occurred at a higher abundance at the catalytically-optimal pH of 6 than at the other pH values. This interaction correlated strongly with the occurrence of the catalytic $Glu165_{fungal}\cdots O1$ interaction (Fig. 5). These findings with both enzymes indicate that the catalytic H-bond between the carboxylic acid moiety of $Glu166_{bacterial}/Glu165_{fungal}$ and the substrate O1 atom was stabilized by a H-bond between the carbonyl group of the Glu residue and the OH group adjacent to O1 of the substrate (Figs. 4 and 5). In sum, the substrate interaction dynamics with the acidic/basic Glu residue in the fungal enzyme that mirrored more the experimental pH activity profile than the corresponding dynamics with the bacterial enzyme (Jeng et al., 2011).

We also monitored pH-dependent changes in the substrate interactions with other residues in the substrate-binding cleft. First, we investigated the interactions with the nucleophilic Glu ($Glu352_{bacterial}/Glu367_{fungal}$), whose interactions during substrate binding are not well understood. In the MD simulations of the bacterial BG, the abundance of the nucleophilic Glu interaction, $H_{O2}\cdots Glu352_{bacterial}$, exhibited the same trend as the catalytic $Glu166_{bacterial}\cdots O1$ interaction: minimal abundance at pH 5, highest abundance at pH 6, and statistically similar abundance at pH 7 and pH 7.5 (Fig. 4). Furthermore, the Pearson's test for co-occurrence correlation revealed that the presence of the $H_{O2}\cdots Glu352_{bacteiral}$ interaction (or the presence of a pair of interactions that included the $H_{O2}\cdots Glu352_{bacterial}$) correlated with the occurrence of the catalytic $Glu166_{bacterial}\cdots O1$ interaction at pHs 5, 6, and 7 (SI, Table S2). On the other hand, in the simulations of the fungal enzyme, there was no correlation between the pH-dependent trend of the frequency abundances of the nucleophilic Glu interaction ($H_{O2}\cdots Glu367_{fungal}$) and the acidic/basic Glu interaction ($Glu165_{fungal}\cdots O1$) (Fig. 5). The decoupling of these two interactions was especially evident at pH 7 and 7.5 where the $H_{O2}\cdots Glu367_{fungal}$ was highly abundant whereas there was little to no occurrence of the $Glu165_{fungal}\cdots O1$ interaction at these pH values. Correlation analysis confirmed this lack of significant co-occurrence of these interactions (Supplementary Table S2).

In addition to the two catalytic Glu residues, we examined the pH-dependent dynamics of substrate interactions with non-catalytic active site residues (Figs. 4 and 5; Supplementary Figs. S4 and S5). As listed in the Introduction, the relevant non-catalytic residues ($Gln20_{bacterial}/Gln16_{fungal}$, $Glu406_{bacterial}/Glu424_{fungal}$, $Tyr296_{bacterial}/Tyr298_{fungal}$, $His121_{bacterial}/His119_{fungal}$, and $Trp407_{bacterial}/Trp425_{fungal}$) are highly conserved in the substrate binding cleft of family 1 BGs (Badieyan et al., 2012) (Figs. 4 and 5). In addition to these residues, our MD simulations captured the participation of substrate interactions with two others: a Trp ($Tp399_{bacterial}/Trp417_{fungal}$) and a Glu/Asp ($Glu409_{bacterial}/Asp427_{fungal}$) (Figs. 4 and 5; Supplementary Figs. S4 and S5). We specifically examined the pH dependence of these non-catalytic interactions and their co-occurrence or lack thereof with the key catalytic interaction involving the acidic/basic Glu ($Glu166_{bacterial}\cdots O1$ and $Glu165_{fungal}\cdots O1$).

In the bacterial enzyme, the $Trp399_{bacterial}\cdots O4$ interaction occurred at higher abundance at pH 5 than at pHs 6, 7, and 7.5; this pattern was inversely related to the trend of the catalytic $Glu166_{bacterial}\cdots O1$ interaction (Fig. 4). The frequency of another interaction, $H_{O4}\cdots Glu406_{bacterial}$, was the lowest at the optimal pH 6 (Fig. 5), thus also occurring at the inverse trend of the

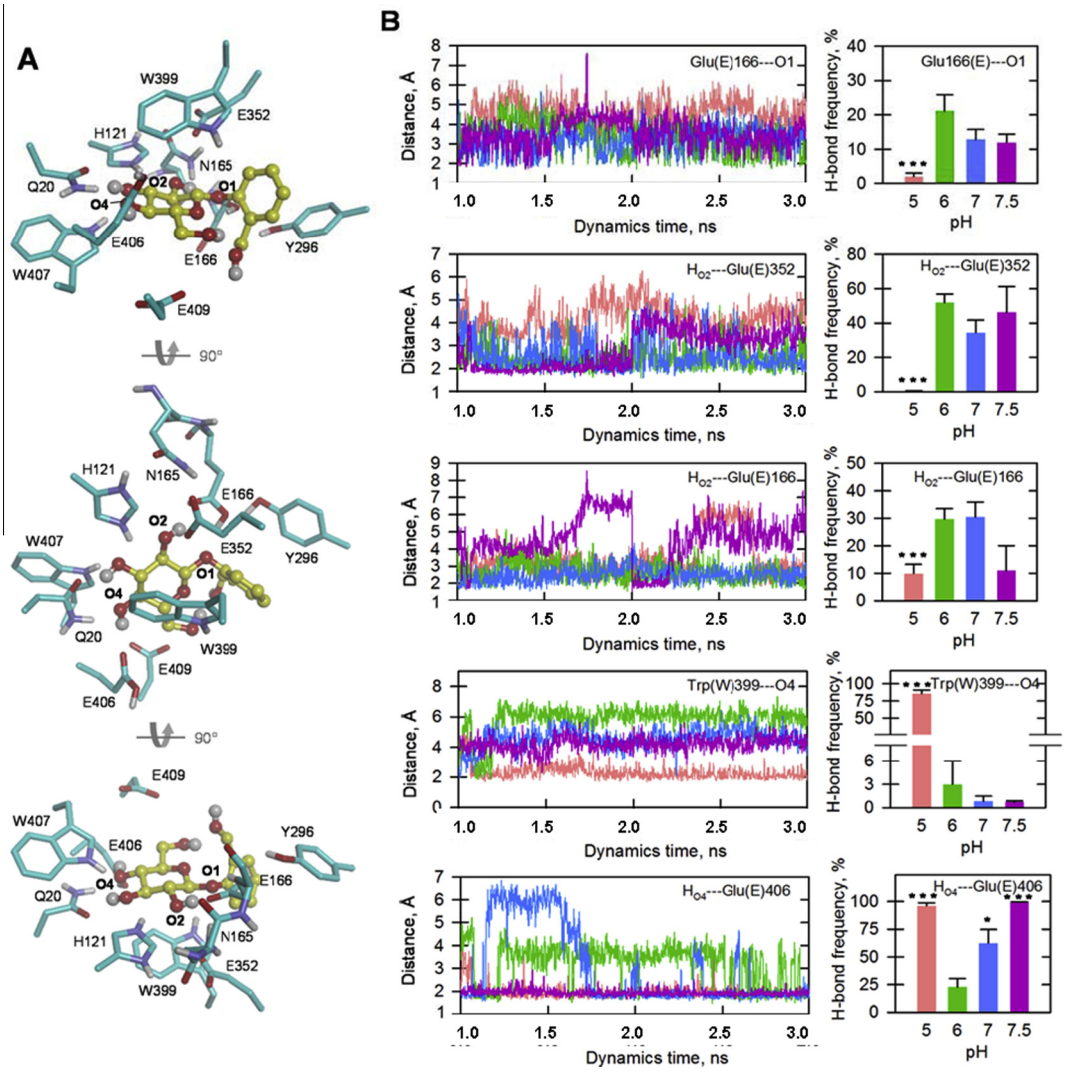


Fig. 4. (A) Close-up views of the substrate in the active site of a BG from the bacterium *C. cellulovorans* and (B) distances (left) and frequency of H-bond abundances (right) between active site residues and the substrate in the MD-simulated configurations at pH 5 (pink), pH 6 (green), pH 7 (blue), and pH 7.5 (purple). Statistical analysis: p-value < 0.05 = *; p-value < 0.001 = ***.

interaction with catalytic residue. With the fungal enzyme, the MD simulations reflected a similar abundance pattern for the corresponding non-catalytic interaction (Fig. 6). The Trp417$_{fungal}$···O4 interaction, which was absent in the simulations at pHs 6 and 7 (the two pH conditions that had the highest occurrence of the catalytic Glu165$_{fungal}$···O1 interaction) was also inversely correlated with the catalytic interaction (Fig. 5). In addition, we found that the Gln16$_{fungal}$···O4 interaction occurred statistically less at pH 6 than at the other pH conditions (Fig. 5). The Asn164$_{fungal}$···O2 interaction displayed a similar frequency pattern to that of the nucleophilic Glu interaction (H_{O2}···Glu367$_{fungal}$), with high abundances at pH 7 and 7.5. An inverse pattern was observed for the occurrence of the His119$_{fungal}$···O2 interaction, which was much lower at pH 7 and 7.5 than at pH 5 and pH 6 (Fig. 5).

The following interactions, which were statistically different at pH 6 when compared with the other pH values, exhibited no clear trend: His121$_{bacterial}$···O2, His121$_{bacterial}$···O3, Gln20$_{bacterial}$···O4, Trp407$_{bacterial}$···O3, H_{O6}···Glu409$_{bacterial}$, and Tyr298$_{bacterial}$···O1 for the bacterial enzyme; and, Trp425$_{fungal}$···O3, His119$_{fungal}$···O3, and H_{O6}···Asp427$_{fungal}$ for the fungal enzyme (Supplementary Figs. S4 and S5). The pH-dependent significance of these latter interactions with respect to the catalytic interactions could not be resolved. A noticeable general trend in both the bacterial and fungal BG simulations at pH 6 was the abundance discrepancy between the catalytic interactions, which occurred during 22% and 17% of the total simulation time, compared to several non-catalytic interactions, which were observed in nearly every simulation frame (Supplementary Figs. S4 and S5). Specifically, His121$_{bacterial}$···O2, His121$_{bacterial}$···O3, and Trp407$_{bacterial}$···O3 as well as H_{O4}···Glu424$_{fungal}$, Gln16$_{fungal}$···O2, and His119$_{fungal}$···O3 were relatively abundant across all pH conditions.

3.3. pH-dependent involvement of solvated waters in substrate binding dynamics

We investigated direct water-substrate interactions as well as water-bridging interactions. A water-bridging interaction was characterized by the occurrence of H-bonds connecting an amino acid residue and the substrate to the same water molecule. Expectedly, the MD simulations revealed the highest amount of direct water-substrate interactions with the solvent-exposed aromatic ring (O7 and H_{O7}) of the substrate (Supplementary Fig. S6). Solvating water interactions also occurred with the atoms of the substrate's glucose ring localized inside the binding cleft as well as with the glycosidic O atom (O1) (Supplementary Fig. S6). Specifically, the solvated bacterial BG revealed H-bonds between

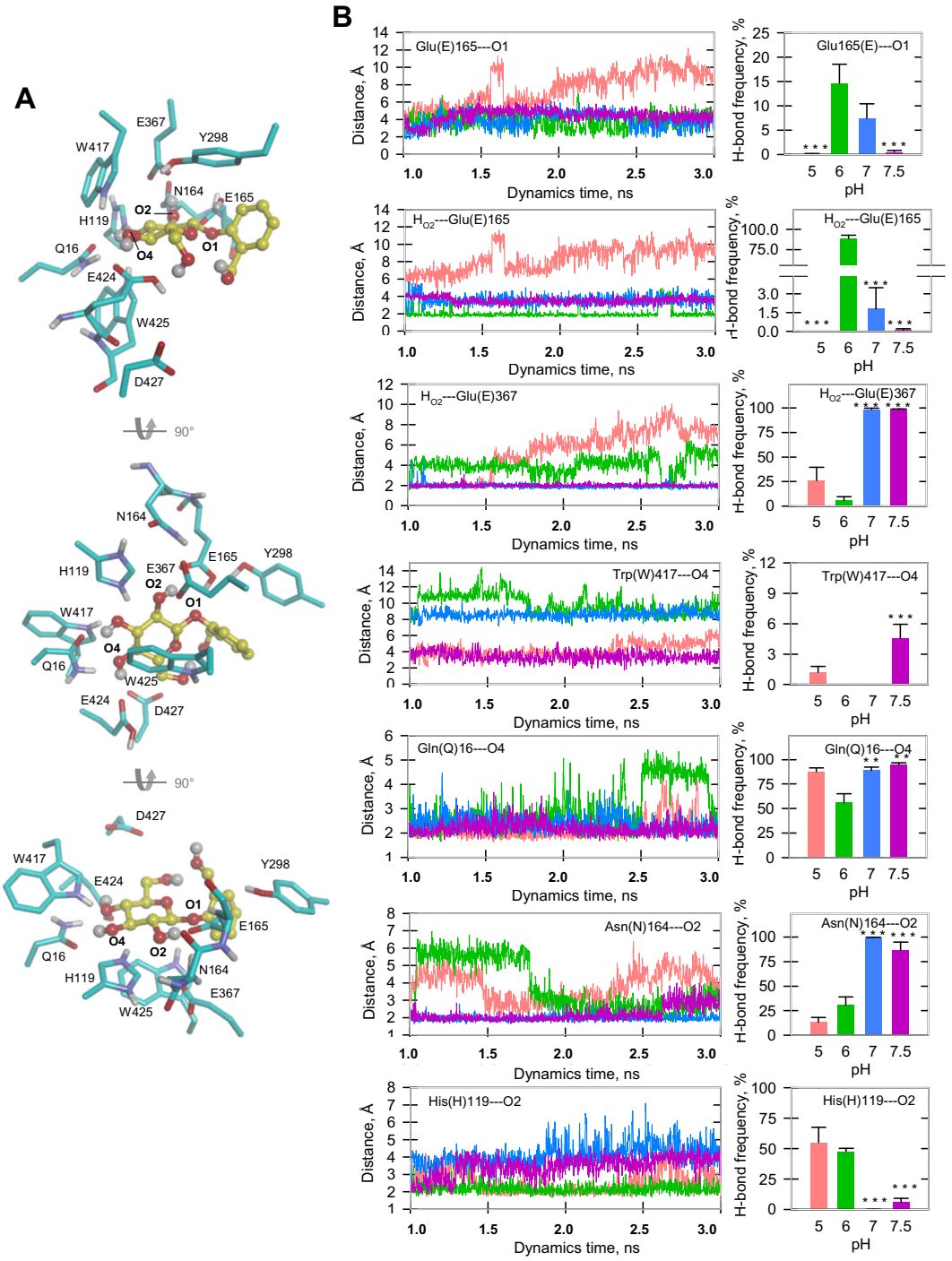


Fig. 5. (A) Close-up views of the substrate in the active site of a BG from the fungus *T. reesei* and (B) distances (left) and frequency of H-bond abundances (right) between active site residues and the substrate in the MD-simulated configurations at pH 5 (pink), pH 6 (green), pH 7 (blue), and pH 7.5 (purple). Statistical analysis: p value < 0.01 = **, p value < 0.001 = ***).

water molecules and the substrate atoms O4, O5, O6, H_{O4}, and H_{O6} and the solvated fungal BG displayed water interactions with the substrate atoms O3, O4, O5, O6, H_{O3}, H_{O4}, and H_{O6}.

With respect to water bridging interactions, we accounted between 634 and 2712 interactions involving 10 to 22 unique amino acid residues (Fig. 7 and Supplementary Fig. S6). Fig. 6 illustrates the differences in the water-bridging profiles at the different pH conditions for each enzyme (Fig. 6 and Supplementary Fig. S6). The total sum of water-bridging interactions was the lowest at the catalytically-optimal pH of 6 for the bacterial enzyme (Supplementary Fig. S7). The total number of water-bridging interactions formed in the fungal enzyme decreased as a function of increasing pH, spanning from 2712 water-bridges at pH 5 to 634 water-bridges at pH 7.5 (Supplementary Fig. S7).

We also observed distinct water-bridging interaction patterns at the different pH conditions. With the bacterial enzyme, the two most abundant water-bridging interactions at pH 6 involved Lys $413_{bacterial}$ and Tyr $296_{bacterial}$. Lys $413_{bacterial}$ is located outside of the immediate active site and only formed water-bridging interactions at pH 6. Tyr$296_{bacterial}$, which is located in the immediate

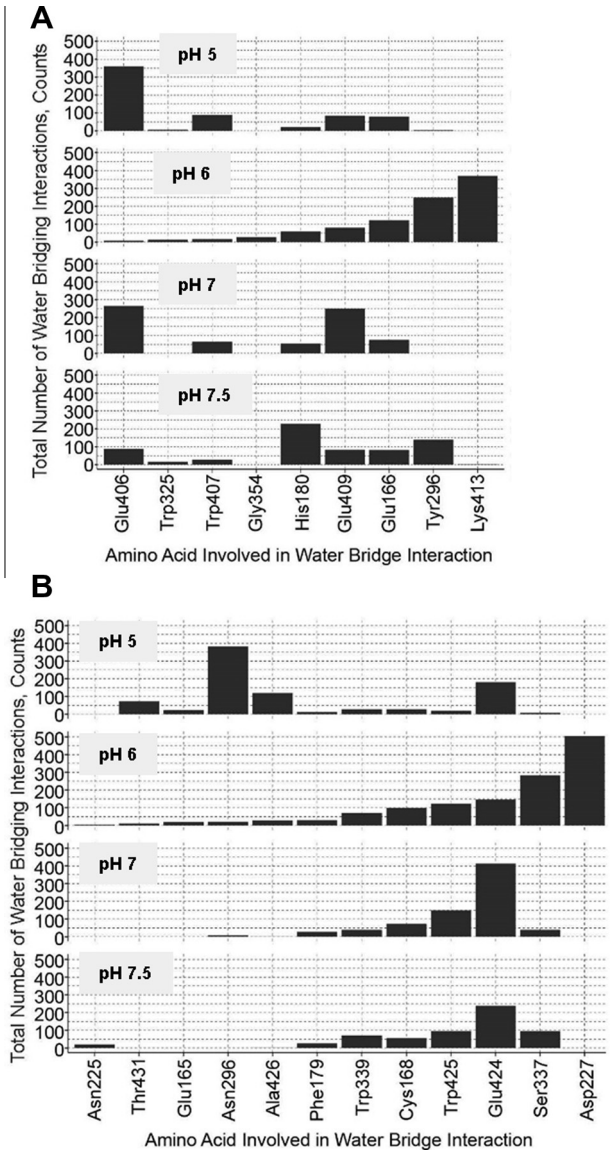


Fig. 6. MD prediction of total amount of water-bridging interactions between the substrate and the bacterial BG (A) and the fungal BG (B) enzymes at the different pH values. Amino acids are listed (from left to right) according to their increasing participation in water-bridging interactions at pH 6.

active site, recorded a high abundance of water-bridging interactions (Fig. 6), but was found to rarely participate in direct H-bonding with the substrate (Supplementary Fig. S6). $Glu406_{bacterial}$, a residue within the immediate active site, was involved in the least abundant water-bridging interaction at pH 6 (Fig. 7). At other pH conditions, however, $Glu406_{bacterial}$ water-bridging interactions occurred in high abundance, a pattern which closely follows the abundances of the direct H-bonding between $Glu406_{bacterial}$ and the substrate (Figs. 5 and 6).

With respect to the fungal BG enzyme, the most abundant water-bridging interactions at pH 6 involved the residues $Asp227_{fungal}$ and $Ser337_{fungal}$, both of which are outside of the immediate active site (Fig. 6). These interactions occurred either uniquely ($Asp227_{fungal}$) or in higher abundance ($Ser337_{fungal}$) at pH 6, when compared with the simulations at the other pH conditions (Fig. 6). In addition, the lowest amount of water-bridging interactions at pH 6 occurred with $Asn225_{fungal}$, which is localized outside of the immediate active site; this low abundance was found in all pH conditions (Fig. 6).

3.4. pH-dependent structural dynamics in the enzyme

The pH-dependent frequency patterns of both the direct substrate binding and water-bridging interactions suggested conformational changes of the enzymes' structures under different pH conditions (Figs. 4–6). Correlated structural dynamics were monitored by examining the geometrical fluctuations (i.e. RMSD) of the residues involved in changes in ionization states, catalytic interactions, non-catalytic interactions, and water-bridging interactions (Figs. 7 and 8). Positive correlations between pairs of amino acid residues indicated that these residues are increasing or decreasing their displacement from their original location at similar times throughout the simulation. Negative correlations, conversely, indicated pairs of amino acid residues whose RMSD values moved in opposing directions at the same time. Our discussion is focused on significant RMSD correlations with either the acidic/basic Glu ($Glu166_{bacterial}$/$Glu165_{fungal}$) or the nucleophilic Glu ($Glu352_{bacterial}$/$Glu367_{fungal}$), which occurred uniquely at the optimal pH (pH 6) for enzymatic activity.

The MD simulations of the bacterial BG indicated fewer RMSD correlations at pH 6 than at any other simulated pH condition (Fig. 7). The RMSD values of the residues $His180_{bacterial}$ (a residue involved in water bridging) and $Glu153_{bacterial}$ (an ionizable residue which becomes negative at pH 6), correlated positively with the acidic/basic Glu ($Glu165_{bacterial}$) of the bacterial BG (Fig. 7). No uniquely significant positive RMSD correlations were observed with the nucleophilic Glu residue ($Glu352_{bacterial}$) at pH 6 (Fig. 7). Furthermore, no negative RMSD correlations were observed with both catalytic Glu residues at pH 6 (Fig. 7). By contrast to the correlation patterns of the structural dynamics in the bacterial enzyme, the corresponding correlation patterns for the fungal enzyme showed a large number of statistically-significant correlations at pH 6 (Fig. 8). Specifically, positive correlations between the RMSD fluctuations of the acidic/basic $Glu165_{fungal}$ and those of $Trp425_{fungal}$ and $Tyr298_{fungal}$ (both involved in non-catalytic substrate binding) were only observed at pH 6 (Fig. 8). In addition, the RMSD fluctuations of the nucleophilic $Glu367_{fungal}$ were positively correlated with those of $Trp425_{fungal}$ (non-catalytic) and $Asp256_{fungal}$ (ionizable residue), and were negatively correlated with those of $His127_{fungal}$ (ionizable residue) and $Thr431_{fungal}$ (water-bridging residue) (Fig. 8).

The correlation results (Figs. 7 and 8) thus illustrated that pH-induced conformational changes, which are evidently unique to the enzyme structure, can impact the correlated movement of proximal and distal amino acid residues (Figs. 7 and 8). Despite the fact that the amino acid residues in the substrate cleft are highly conserved, the fluctuations of neighboring amino acid residues in the active site did not always correlate with each other (Figs. 7 and 8). Furthermore, structural fluctuations with residues in the substrate-binding cleft were significantly correlated with fluctuations of several residues (ionizable residues and water-bridging interaction residues) localized outside of the immediate active site. And, many of these correlations were found to be pH- and enzyme-dependent (Figs. 7 and 8).

4. Discussion

4.1. Changes in ionization states and disruption of substrate binding interactions

In Section 1, we put forth three hypotheses regarding how changes in pH may impact substrate binding dynamics. In strong support for the second hypothesis, our findings revealed that the specific residues, which were sensitive to changes in ionization states, were primarily localized on the surface of enzyme, and

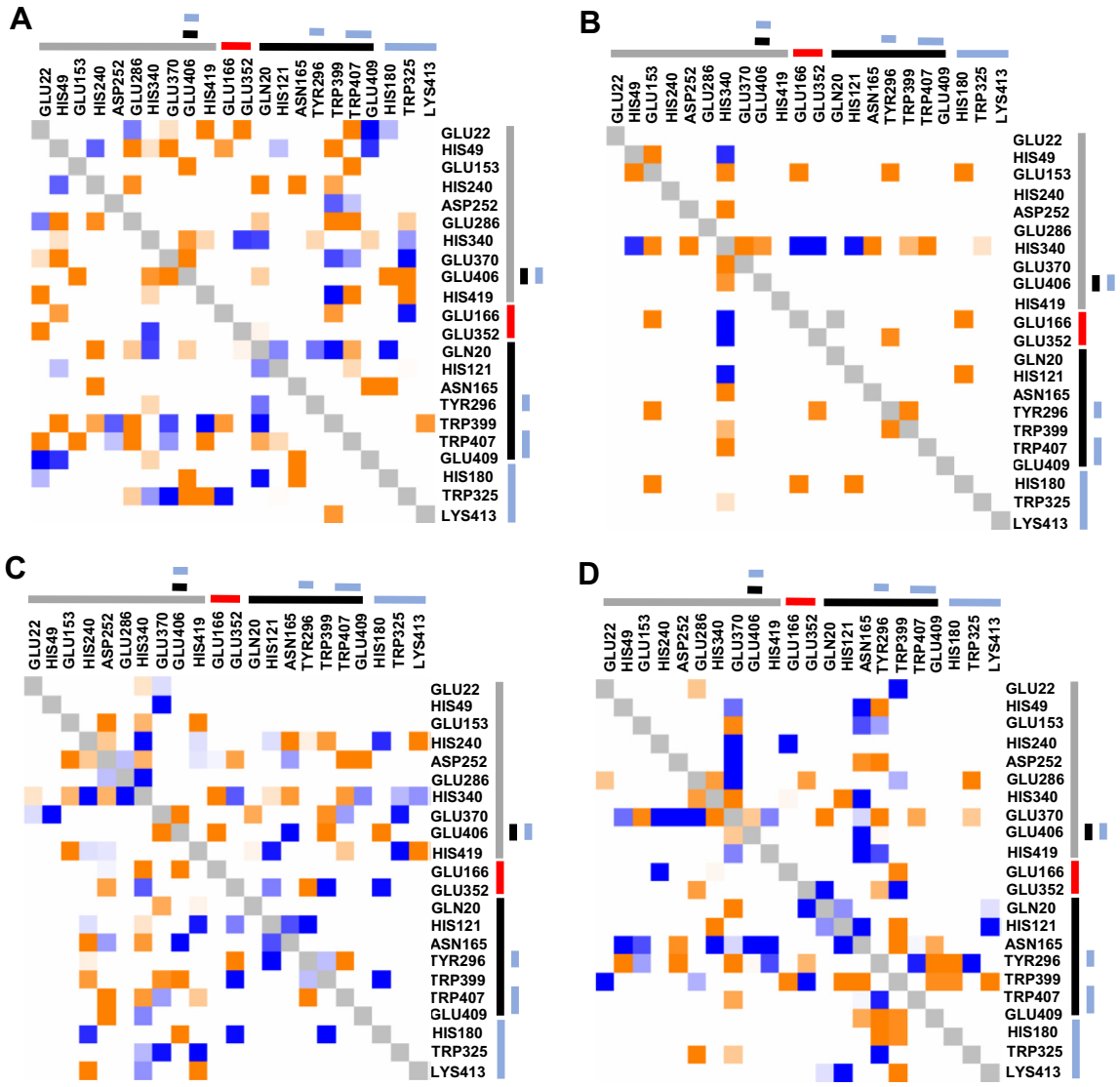


Fig. 7. Correlation matrices of amino acid-resolved structural dynamics of the bacterial BG enzyme at pH 5 (A), pH 6 (B), pH 7 (C), and pH 7.5 (D). The RMSD of each amino acid of interest was monitored over the course of the MD simulation. Positive, negative, non-significant, and self-pairing correlations are shown, respectively, in orange, blue, white, and gray squares. The bars alongside the residue labels designate residues that change ionization states (gray), catalytic residues (red), non-catalytic residue involved in direct substrate binding (black), residues involved in water-bridging interactions with the substrate (light blue).

away from the active site; the only exception was a Glu localized in the active site of the bacterial BG. Mutations of amino acid residues on the surface of enzymes have been shown previously to result in a shift in the optimal pH for enzyme activity of an endocellulase (Wang et al., 2005) and an exocellulase (Wohlfahrt et al., 2003) from *T. reesei* but the mechanisms were not fully elucidated. The experiments with the exocellulase indicated that neighboring carboxyl-carboxyl pair repulsion destabilized the enzyme (Wohlfahrt et al., 2003). In our simulations, the prominent ionizable residues were not located immediately next to each other. However, it was clear that the increasing negative charges arising from these residues (one Asp, five Glu, and four His residues in the bacterial enzyme; five Asp, three Glu, and four His residues in the fungal enzyme) as a function of increasing pH resulted in conformational changes in the enzyme (Table 1; Fig. 3). This led to alterations in the substrate binding interactions as evidenced by the pH-dependent frequency patterns of the H-bonding interactions (Figs. 4 and 5). Our work thus provides for the identification of residues, which could be targeted in protein engineering towards improving enzymatic activity at unfavorable pH conditions.

In the bacterial BG enzyme, low abundances of the H-bonding interaction between the substrate and $Glu166_{bacterial}$ (the acidic/basic Glu) at pH 5 was in agreement with the pH–activity profile (Jeng et al., 2011) (Fig. 4). At the pH values higher than pH 6, the dynamics of this interaction did not correlate with the trend in activity reduction observed in experiments (Jeng et al., 2011) (Fig. 4). Correlation analysis indicated that the $H_{O2}\cdots Glu166_{bacterial}$ interaction may help align Glu166 in a favorable position for the catalytic interaction, $Glu166_{bacterial}\cdots O1$, to occur at pH 6. Furthermore, the interaction of $Glu166_{bacterial}$ with O1 of the substrate correlated positively with the interaction of $Glu352_{bacterial}$ (the nucleophilic Glu) with H_{O2} of the substrate (Fig. 4). This is consistent with the nucleophilic Glu aiding in the positioning of the substrate for the catalytic action, the protonation of O1, by the acidic/basic Glu residue. The correlation analysis further implies the participation of both $Trp399_{bacterial}\cdots O4$ and $H_{O4}\cdots Glu406_{bacterial}$ in the destabilization of the

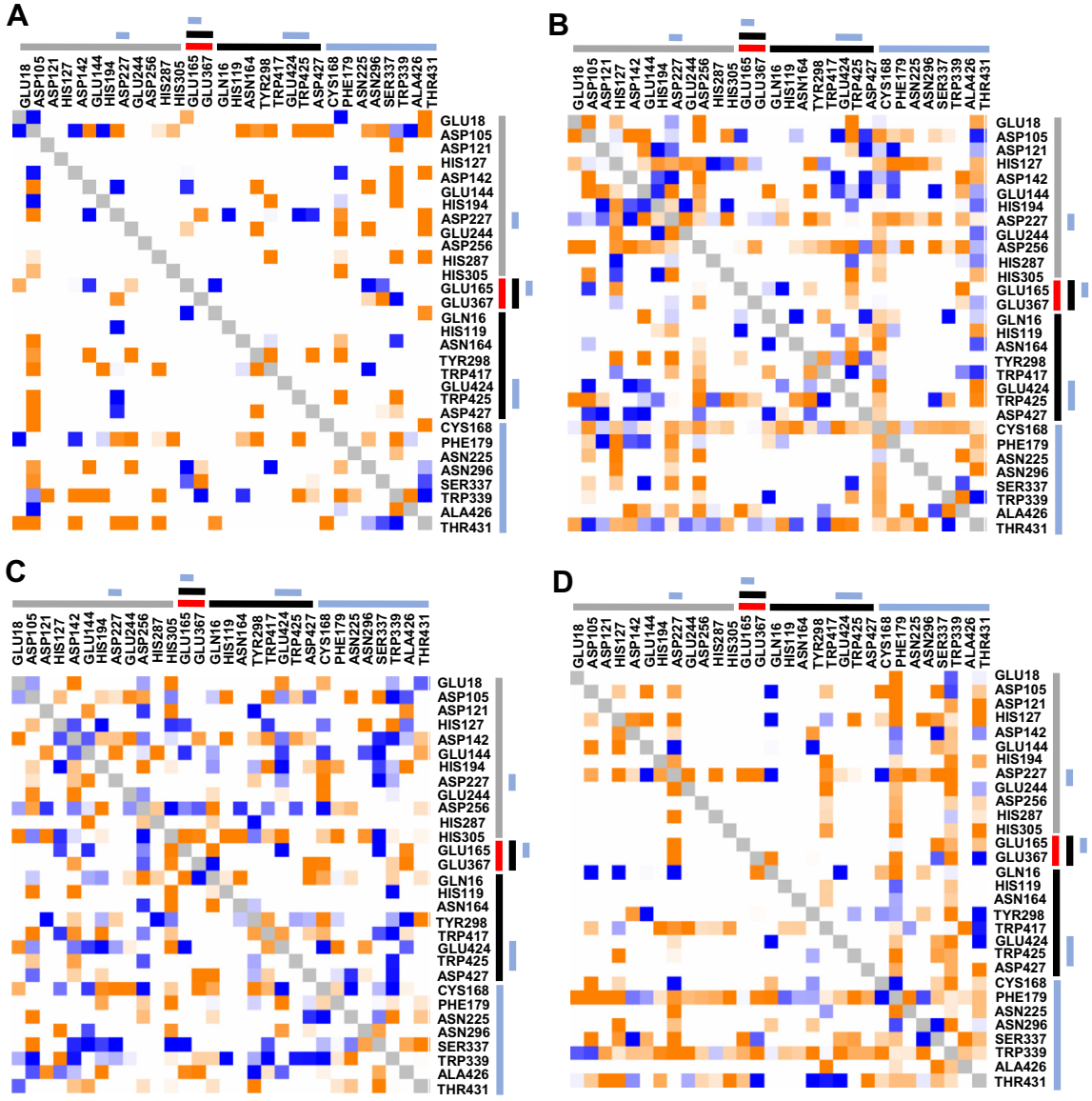


Fig. 8. Correlation matrices of amino acid-resolved structural dynamics of the fungal BG enzyme at pH 5 (A), pH 6 (B), pH 7 (C), and pH 7.5 (D). The RMSD of each amino acid of interest was monitored over the course of the MD simulation. Positive, negative, non-significant, and self-pairing correlations are shown, respectively, in orange, blue, white, and gray squares. The bars alongside the residue labels designate residues that change ionization states (gray), catalytic residues (red), non-catalytic residue involved in direct substrate binding (black), residues involved in water-bridging interactions with the substrate (light blue).

catalytically-optimal substrate orientation in the binding site (Fig. 7).

In contrast, the MD simulations of the fungal BG indicated the highest abundance of H-bond between the nucleophilic Glu ($Glu367_{fungal}$) and the substrate at pH 7 and pH 7.5 whereas there was a low abundance of H-bond between the acidic/basic Glu ($Glu165_{fungal}$) and the substrate O1 atom (Fig. 5). Thus, a conformational arrangement in which strong interactions of the substrate with the nucleophilic Glu367 did not lead to a favorable orientation of the substrate for the H-bonding interaction with the acidic/basic Glu and the O1 of the substrate. Because it was determined that the fungal BG exhibited decreasing activity at pH greater than 6 (Jeng et al., 2011), these simulation results implied that the H-bond interaction involving $Glu367_{fungal}$ may be a barrier for catalysis in the fungal enzyme. Our correlation analysis indicated the supporting participation of $H_{O2}\cdots Glu165_{fungal}$ for the occurrence of the acidic/basic $Glu166_{fungal}\cdots O1$ interaction (Figs. 6 and 8). Additionally, our dynamics results are consistent with the role of the $Gln16_{fungal}\cdots O4$, $Trp417_{fungal}\cdots O4$, and $Asn164_{fungal}\cdots O2$ interactions in impeding the favorable catalytic interactions (Fig. 8).

The relatively low abundance of the catalytic acidic/basic Glu interaction with O1, in comparison to several non-catalytic interactions suggests a predominant involvement of non-catalytic interactions with the highly conserved amino acid residues in the initial substrate binding. In sum, the MD predictions of the pH-dependent interactions between the substrate and the non-catalytic amino acid residues in the substrate binding cleft provided further insights on the consequence pH on favorable substrate binding.

4.2. Water molecules in the stabilization of substrate binding

Water molecules have been shown to be essential in stabilizing the transition state and in the removal of the retained portion of the cleaved substrate in the BG hydrolysis pathway (Badieyan et al., 2012; White and Rose, 1997; Vocadlo and Davies, 2008; Wang et al., 2011). Less is known, however, about the role of

waters in substrate binding of BGs (Chuenchor et al., 2011). In structures of BGs co-crystallized with a bound substrate, several water molecules were localized in the active site as well as within interaction distance of the glycosidic O atom (Badieyan et al., 2012). A QM/MM modeling reported greater energy contribution of water interactions during stabilization of the intermediate in the active site than during the substrate binding (Badieyan et al., 2012). This QM/MM computation was based on the static position of the water molecules in BG's active site (Badieyan et al., 2012). By modeling explicitly the solvated waters in our MD simulations, we were able to monitor the dynamic interactions of water molecules with the substrate.

In agreement with our third hypothesis, our simulations revealed a pH-dependent network of water-bridging interactions between the substrate and amino acid residues inside and outside of the immediate active site (Fig. 7). We note that Lys $413_{bacterial}$, $Asp227_{fungal}$, $Ser337_{fungal}$, all of which are located outside of the active site, were able to participate in substrate binding through water-bridging connections (Fig. 6). In addition to the solvation of the solvent-exposed aromatic ring of the substrate, there were water-bridging interactions with several atoms of the substrate ring localized within the pocket of the binding cleft (Supplementary Fig. S5). This is apparently the first time that these interactions have been reported for BG substrate binding.

These findings corroborate a greater and more dynamic role of water molecules in substrate stabilization than previously thought. For instance, $Tyr296_{bacterial}$, found in the immediate active site of the bacterial BG, infrequently interacted directly with the substrate at pH 6, but participated in a large amount of water-bridging interactions. This is consistent with a conformational arrangement of the active site wherein $Tyr296_{bacterial}$ does not interact directly with the substrate but still adopts an orientation that allows for multiple water-bridging interactions (Figs. 4 and 6). On the other hand, $Glu406_{bacterial}$, a residue in the immediate active site of the bacterial BG, participated in markedly low abundances in both direct and water-bridging interactions with the substrate at pH 6 whereas were abundant at other pH values, thus presenting a conformation of the enzyme in which either direct or water-bridge interactions between $Glu406_{bacterial}$ and the substrate were not critical. Connections between water-bridging interaction abundances and direct residue-substrate interactions were less evident in the fungal BG simulations. Water bridging profiles thus expand our view of which amino acids are involved in substrate stabilization.

5. Implications and caveats

We conducted MD simulations to gain insights on the effects of pH on the substrate binding dynamics, the initial step of catalysis, in two family 1 BG enzymes from *T. reesei* and *C. cellulovorans*. Experimental determinations of structures for one BG enzyme are obtained typically at one aqueous pH condition that yields the best crystals (Jeng et al., 2011, 2012). Therefore, it is challenging to resolve the structural dynamics underlying the pH-dependent BG catalytic action on the substrate, albeit it is well known that enzymatic activity of BGs and related enzymes is influenced significantly by the aqueous pH conditions (Jeng et al., 2011; Zibaee et al., 2009; Yan and Wu, 2013). Furthermore, the direct experimental characterization of the role of interacting water molecules in BG substrate binding is limiting (Badieyan et al., 2012), necessitating complementary insights from molecular modeling. Our computational study presents the first molecular perspective on the consequence of pH-dependent conformational changes on substrate-binding interactions in BGs.

Changes in ionizable residues on the enzyme surface induced disruptive changes in the favorable conformation for substrate binding interactions with the catalytic residues. Our findings further demonstrated that the role of the catalytic nucleophilic Glu residue in either facilitating or disrupting the catalytic substrate binding by the acidic/basic Glu was dependent on the enzyme structure and the pH condition. In addition, we identified substrate binding interactions with specific non-catalytic residues which may be unfavorable for the forwarding steps in catalytic pathway. The explicit solvation identified several water-bridging interactions, which were important at the optimal pH for enzyme activity, were disrupted at other pH values due to structural re-arrangements in the enzyme conformation. Furthermore, correlation analysis of amino acids' movements indicated that the synchronicity (or lack thereof) of pH-dependent geometrical fluctuations in a network of amino acids, which were localized both outside and inside the substrate binding cleft, was responsible for the substrate binding specificity.

The relevance of our findings should be considered in relation to two important factors. First, we conducted our analysis using equilibrated structures obtained during short-time dynamics conducted for several nanoseconds. We have also performed longer simulations (at 200 ns), which are typical in MD simulation studies. However, our findings revealed that long-time simulations are not appropriate to probe the substrate binding, especially under unfavorable pH conditions. In such conditions, we found that the substrate migrated outside the substrate binding cleft and interacted with residues on the periphery of the active site. Our findings thus indicated that, in order to increase the sampling set of substrate interactions specifically in the active site, a more appropriate approach is to equilibrate a range of docked substrate structures for short-time dynamics time. Future explorations of the substrate affinity to residues outside of the active site will provide insights into the influence of pH on the tunneling pathway of the substrate towards the active site. Second, an evaluation of the pH-dependent structural dynamics on the entire catalytic reaction pathway is requisite to a comprehensive understanding of the structural basis for the pH-dependent activity of BGs. The present study provides an account of the pH-dependent substrate binding interactions in response to pH-induced conformational dynamics, thus laying the foundation for future investigations on the influence of pH on the subsequent steps in the catalytic pathway.

Acknowledgments

D.F.F. acknowledges an Integrative Graduate Education and Research Traineeship (IGERT) research fellowship and a Graduate Research Fellowship, both from the U.S. National Science Foundation. We acknowledge technical assistance and support from the Cornell Advanced Computing facility. We thank Hua Wei, Ed Park, Amy Pochodylo, and Samantha Sasnow of the Aristilde Research Group (Department of Biological and Environmental Engineering, Cornell University) for technical support, and Chantal Koechli (Department of Microbiology, Cornell University) for aid in graphics generation. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. This research was funded by a start-up package from Cornell University.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2015.07.002.

References

Acosta-Martínez, V., Tabatabai, M.A., 2000. Enzyme activities in a limed agricultural soil. Biol. Fertil. Soils 31 (1), 85–91.

Badieyan, S., Bevan, D.R., Zhang, C., 2012. Probing the active site chemistry of β-glucosidases along the hydrolysis reaction pathway. Biochemistry (Mosc.) 51 (44), 8907–8918.

Baker, E.N., Hubbard, R.E., 1984. Hydrogen bonding in globular proteins. Prog. Biophys. Mol. Biol. 44 (2), 97–179.

Baptista, A.M., Teixeira, V.H., Soares, C.M., 2002. Constant-pH molecular dynamics using stochastic titration. J. Chem. Phys. 117 (9), 4184–4200.

Bissantz, C., Kuhn, B., Stahl, M., 2010. A medicinal chemist's guide to molecular interactions. J. Med. Chem. 53 (14), 5061–5084.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4 (2), 187–217.

Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al., 2009. CHARMM: the biomolecular simulation program. J. Comput. Chem. 30 (10), 1545–1614.

Bürgi, R., Kollman, P.A., van Gunsteren, W.F., 2002. Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. Proteins Struct. Funct. Bioinf. 47 (4), 469–480.

Campos, S.R.R., Machuqueiro, M., Baptista, A.M., 2010. Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. J. Phys. Chem. B 114 (39), 12692–12700.

Chuenchor, W., Pengthaisong, S., Robinson, R.C., Yuvaniyama, J., Svasti, J., Cairns, J.R.K., 2011. The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. J. Struct. Biol. 173 (1), 169–179.

Creighton, T.E., 1993. Proteins: Structures and Molecular Properties. Macmillan.

Dlugosz, M., Antosiewicz, J.M., 2004. Constant-pH molecular dynamics simulations: a test case of succinic acid. Chem. Phys. 302 (1–3), 161–170.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4 (3), 435–447.

Huber, R.E., Hlede, I.Y., Roth, N.J., McKenzie, K.C., Ghumman, K.K., 2001. His-391 of β-galactosidase (*Escherichia coli*) promotes catalyses by strong interactions with the transition state. Biochem. Cell Biol. 79 (2), 183–193.

Jeng, W.-Y., Wang, N.-C., Lin, M.-H., Lin, C.-T., Liaw, Y.-C., Chang, W.-J., Liu, C.-I., Liang, P.-H., Wang, A.H.-J., 2011. Structural and functional analysis of three β-glucosidases from bacterium *Clostridium cellulovorans*, fungus *Trichoderma reesei* and termite Neotermes koshunensis. J. Struct. Biol. 173 (1), 46–56.

Jeng, W.-Y., Wang, N.-C., Lin, C.-T., Chang, W.-J., Liu, C.-I., Wang, A.H.-J., 2012. High-resolution structures of *Neotermes koshunensis* β-glucosidase mutants provide insights into the catalytic mechanism and the synthesis of glucoconjugates. Acta Crystallogr. D Biol. Crystallogr. 68 (7), 829–838.

Jørgensen, H., Kristensen, J.B., Felby, C., 2007. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels Bioprod. Biorefining 1 (2), 119–134.

Kaper, T., van Heusden, H.H., van Loo, B., Vasella, A., van der Oost, J., de Vos, W.M., 2002. Substrate specificity engineering of β-mannosidase and β-glucosidase from pyrococcus by exchange of unique active site residues. Biochemistry (Mosc.) 41 (12), 4147–4155.

Karplus, M., Kuriyan, J., 2005. Molecular dynamics and protein function. Proc. Natl. Acad. Sci. U.S.A. 102 (19), 6679–6685.

Knight, T.R., Dick, R.P., 2004. Differentiating microbial and stabilized β-glucosidase activity relative to soil quality. Soil Biol. Biochem. 36 (12), 2089–2096.

Lan, T.Q., Lou, H., Zhu, J.Y., 2013. Enzymatic saccharification of lignocelluloses should be conducted at elevated pH 5.2–6.2. BioEnergy Res. 6 (2), 476–485.

Langella, E., Improta, R., Barone, V., 2004. Checking the pH-induced conformational transition of prion protein by molecular dynamics simulations: effect of protonation of histidine residues. Biophys. J. 87 (6), 3623–3632.

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil ph as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 75 (15), 5111–5120.

Machuqueiro, M., Baptista, A.M., 2007. The pH-dependent conformational states of kyotorphin: a constant-pH Molecular dynamics study. Biophys. J. 92 (6), 1836–1845.

Machuqueiro, M., Baptista, A.M., 2008. Acidic range titration of HEWL using a constant-pH molecular dynamics method. Proteins Struct. Funct. Bioinf. 72 (1), 289–298.

Mariscal-Sancho, I., Santano, J., Mendiola, M.-A., Peregrina, F., Espejo, R., 2010. Carbon dioxide emission rates and [beta]-glucosidase activity in mediterranean ultisols under different soil management. Soil Sci. 175 (9), 453–460.

Monard, G., Prat-Resina, X., González-Lafont, A., Lluch, J.M., 2003. Determination of enzymatic reaction pathways using QM/MM methods. Int. J. Quantum Chem. 93 (3), 229–244.

Murao, S., Sakamoto, R., Arai, M., 1988. Cellulases of *Aspergillus aculeatus*. Methods Enzymol. 160, 274–299.

Percival Zhang, Y.-H., Himmel, M.E., Mielenz, J.R., 2006. Outlook for cellulase improvement: screening and selection strategies. Biotechnol. Adv. 24 (5), 452–481.

Roche, O., Kiyama, R., Brooks, C.L., 2001. Ligand–protein database: linking protein–ligand complex structures to binding data. J. Med. Chem. 44 (22), 3592–3598.

Schomburg, I., Chang, A., Placzek, S., Sohngen, C., Rother, M., Lang, M., Munaretto, C., Ulas, S., Stelzer, M., Grote, A., et al., 2013. BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. Nucleic Acids Res. 41 (D1), D764–D772.

Senn, H.M., Thiel, W., 2007. QM/MM studies of enzymes. Curr. Opin. Chem. Biol. 11 (2), 182–187.

Singhania, R.R., Patel, K., Sukumaran, R., Larroche, C., Pandey, A., 2013. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. Bioresour. Technol. 127, 500–507.

Accelrys Software Inc., 2013. Discovery Studio Modeling Environment. Accelrys Software Inc., San Diego.

Spassov, V.Z., Yan, L., 2008. A fast and accurate computational approach to protein ionization. Protein Sci. 17 (11), 1955–1970.

Tan, J., Verschueren, K.H.G., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B., Mesters, J.R., et al., 2005. PH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses. J. Mol. Biol. 354 (1), 25–40.

Teugjas, H., Väljamäe, P., 2013. Selecting β-glucosidases to support cellulases in cellulose saccharification. Biotechnol. Biofuels 6 (1), 105.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., et al., 2014. XSEDE: accelerating scientific discovery. Comput. Sci. Eng. 16 (5), 62–74.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., 2003. Improved protein–ligand docking using GOLD. Proteins Struct. Funct. Bioinf. 52 (4), 609–623.

Vocadlo, D.J., Davies, G.J., 2008. Mechanistic insights into glycosidase chemistry. Curr. Opin. Chem. Biol. 12 (5), 539–555.

Vuong, T.V., Wilson, D.B., 2010. Glycoside hydrolases: catalytic base/nucleophile diversity. Biotechnol. Bioeng. 107 (2), 195–205.

Wang, T., Liu, X., Yu, Q., Zhang, X., Qu, Y., Gao, P., Wang, T., 2005. Directed evolution for engineering pH profile of endoglucanase III from *Trichoderma reesei*. Biomol. Eng. 22 (1–3), 89–94.

Wang, J., Hou, Q., Dong, L., Liu, Y., Liu, C., 2011. QM/MM studies on the glycosylation mechanism of rice BGlu1 β-glucosidase. J. Mol. Graph. Model. 30, 148–152.

White, A., Rose, D.R., 1997. Mechanism of catalysis by retaining β-glycosyl hydrolases. Curr. Opin. Struct. Biol. 7 (5), 645–651.

Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. Nat. Clim. Change 3 (10), 909–912.

Withers, S.G., Rupitz, K., Trimbur, D., Warren, R.A.J., 1992. Mechanistic consequences of mutation of the active site nucleophile Glu 358 in Agrobacterium.beta.-glucosidase. Biochemistry (Mosc.) 31 (41), 9979–9985.

Wohlfahrt, G., Pellikka, T., Boer, H., Teeri, T.T., Koivula, A., 2003. Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* cellobiohydrolase Cel6A. Biochemistry (Mosc.) 42 (34), 10095–10103.

Wu, G., Robertson, D.H., Brooks, C.L., Vieth, M., 2003. Detailed analysis of grid-based molecular docking: a case study of CDOCKER – A CHARMm-based MD docking algorithm. J. Comput. Chem. 24 (13), 1549–1562.

Yan, S., Wu, G., 2013. Prediction of optimal pH in hydrolytic reaction of beta-glucosidase. Appl. Biochem. Biotechnol. 169 (6), 1884–1894.

Zhou, J.-M., Zhou, J.-H., Meng, Y., Chen, M.-B., 2006. Molecular dynamics simulation of iminosugar inhibitor–glycosidase complex: insight into the binding mechanism of 1-deoxynojirimycin and isofagomine toward β-glucosidase. J. Chem. Theory Comput. 2 (1), 157–165.

Zibaee, A., Bandani, A.R., Ramzi, S., 2009. Enzymatic properties of alpha- and beta-glocusidases extracted from midgut and salivary glands of rice striped stem borer, *Chilo suppressalis* Walker (Lepidoptera: Pyralidae). C. R. Biol. 332 (7), 633–641.

Zimmerman, A.R., Ahn, M.-Y., 2011. Organo-mineral–enzyme interaction and soil enzyme activity. In: Shukla, G., Varma, A. (Eds.), Soil Enzymology, Soil Biology. Springer, Berlin Heidelberg, pp. 271–292.