TOWARD A STRUCTURAL BASIS FOR THE PH-DEPENDENT ACTIVITY OF BETA-GLUCOSIDASES: COMPUTATIONAL EVIDENCE OF CONFORMATIONAL AND SUBSTRATE-BINDING DYNAMICS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science Environmental Toxicology

by

David Foster Flannelly

August 2015

© 2015 David Foster Flannelly

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 (BGs) enzymes, which convert cellulobiose into two glucose molecules. The activity of these enzymes has been shown to vary with solution pH but the influence of pH-induced changes on the enzyme conformation required for catalytic action on the substrate is not well understood. We applied molecular dynamics simulations to investigate both conformational and substrate binding dynamics in two well-characterized BGs of bacterial (*Clostridium cellulovarans*) and fungal (*Trichoderma reesi*) 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 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 new insights provided by our findings contribute to a structural basis for the pH-dependent substrate binding specificity in BGs.

BIOGRAPHICAL SKETCH

Born and raised in York County, PA. Attended the College of Wooster in Wooster, OH and majored in chemistry and minored in environmental science. After college graduation I taught high school chemistry, AP chemistry, physics, and environmental science at Lee High School in Lee County, AR. Then I entered the field of Environmental Toxicology at Cornell in Dr. Ludmilla Aristilde's lab studying the molecular dynamics of carbon cycling enzymes.

ACKNOWLEDGMENTS

I want to acknowledge the National Science Foundation (NSF) for two fellowships that I was on during my time at Cornell, the NSF IGERT (Food Systems and Poverty Reduction) as well as the NSF GRFP. I would be remiss if I didn't acknowledge the technical and non-technical support I received from my former and present lab mates (Samantha Sasnow, Baoquan Xie, Hua Wei, Amy Pochodylo, Ed Park, Mat Kukurugya, and Johnny Huynh). They are all amazing people, fiercely smart and emotionally warm. I would also like to acknowledge my partner Chantal Koechli for all of her support and encouragement over the years. Last but not least I want to acknowledge my advisor Dr. Ludmilla Aristilde for her instruction and guidance during my time at Cornell University.

TABLE OF CONTENTS

Toward a Structural Basis for the pH-Dependent Activity of Beta-Glucosidases: Computational Evidence of Conformational and Substrate-Binding Dynamics

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.[1] This cleavage is a rate limiting step in the complete degradation of cellulose to glucose.[1,2] 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[3–5] and in engineered biofuel production from lignocellulosic wastes.[1,6–8] Both in natural environments and in engineered bioreactors, BGs are exposed to various aqueous pH values, ranging from acidic to alkaline conditions.[9–12] Because enzymes are comprised of ionizable amino acid residues, pH impacts their conformation, structural stability, and catalytic activity.[13]

A large diversity of pH-activity profiles exists for BGs of different origins, and even within the same species.[14] For instance, optimal activities of BGs from the anaerobic bacterium *Clostridium cellulovorans*[2], the fungi *Asperigillus aculeaus*[15] and *Trichoderma reesi*[2], and the insect *Chilo suprressalis* [16] 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[2], one from the bacterium *C. cellulovorans* and one from the fungus *T. reesi*, which have well defined crystal structures. Both of these enzymes have been shown experimentally to exhibit optimal catalytic activity at pH 6, with decreased activity at higher and lower pH values.[2] 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[2,17,18] (Figure 1; Figure 1). A transition state facilitates the splitting of the C1-C1' link while retaining a covalent bond with the nucleophilic Glu[17] (Figure 1). The reaction is completed by the release of the cleaved part of the substrate following the de-glycosylation step (Figure 1). The initiation of the catalytic action is mediated by a H-bonding interaction between the substrate O1 and the acidic/basic Glu.[2,18] 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.

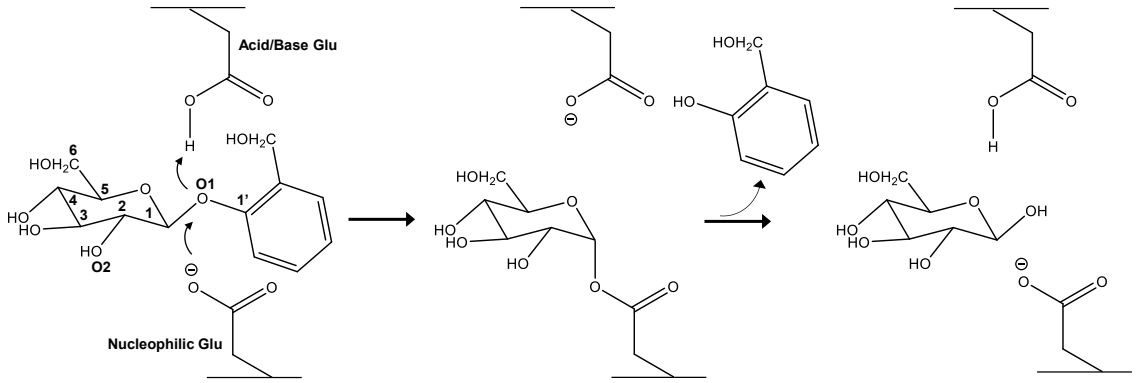

Figure 1. Schematic representation of BG catalytic action on the substrate (salicin).

The nucleophilic Glu residue plays an integral role in stabilizing the intermediate state in the hydrolysis reaction[17] but its role in substrate binding is unresolved. In structures of BGs co-crystallized with substrates in the active site,[19,20] the nucleophilic Glu has adopted an orientation, which 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).[21] 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 (Figure 2): an Asn ($Asn165_{bacterial}$/$Asn164_{fungal}$) residue that immediately precedes the acidic/basic catalytic glutamate residue, a Gln residue ($Gln20_{bacterial}$/$Gln16_{fungal}$) and a Glu residue ($Glu406_{bacterial}$/$Glu424_{fungal}$) which form two H-bonds with the substrate, a Tyr residue ($Tyr296_{bacterial}$/$Tyr298_{fungal}$) which participates in the removal of the bound intermediate, a His residue ($His121_{bacterial}$/$His119_{fungal}$) which mediates the initial binding of the substrate, and a Trp residue ($Trp407_{bacterial}$ /$Trp425_{fungal}$) (Figure 1).[18] These non-catalytic residues have been evaluated for their individual energetic contributions to the catalytic hydrolysis.[18] However, similar to the catalytic Glu residues, the involvement of the conserved non-catalytic residues in the substrate binding under unfavorable pH conditions has not yet been investigated.

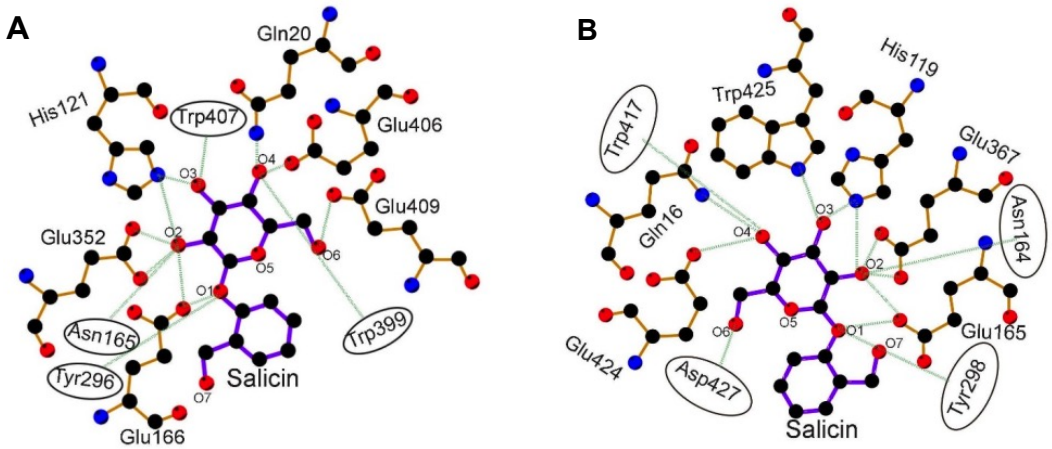

Figure 2. Schematic diagrams of a BG active site from (A) the bacterium *Clostridium cellulovorans* and (B) from the fungus *Trichoderma reesi*. Possible H-bonds that can form between the amino acids and the substrate (salicin) are shown in green. These H-bonds were monitored for interactions throughout the course of simulation.

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 may facilitate unfavorable interactions with solvated waters.

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.[22–24] 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. reesi*[24] 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. reesi,*[25] mutations of carboxyl-carboxyl pairs into amide-carboxyl pairs led to increased activity of the enzyme at higher pH.[25] In this study[25], it was proposed 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.[26–28] 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.[18,29] 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.[30–34] For instance, a MD study[33] 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.[33] With respect to BGs, one MD study investigated the binding dynamics of several known inhibitors whereby the inhibitors assumed different protonation states.[35] 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 determine the structural factors responsible for the pH-dependent substrate binding dynamics in the two aforementioned family 1 BGs from *C. cellulovorans* and from *T. reesi*. We aim (1) to 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) to determine the consequence of these changes on the favorable substrate binding required for catalysis, and (3) to capture the participation of solvated waters in interfering with substrate binding. To meet these objectives, we developed a modeling methodology, which 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 intermolecular interactions with catalytic residues as well as non-catalytic residues.[18] In addition, the explicit solvation approach allowed us to monitor the dynamic role of water molecules both in solvating the substrate and the active site. The molecular perspective illustrated by our MD-optimized structures provides mechanistic insights towards elucidating the link between experimentally-determined pH-dependent activity and pH-induced structural changes in BGs.

Computational Methods

Modeling Platform

Molecular simulations were performed using the all atom force field CHARMM 27 (Chemistry at HARvard Molecular Mechanics),[36] as interfaced through Accelrys's Discovery Studio software package[37] and the open source molecular dynamic software GROMACS{Citation}. This force field was previously validated[36,38] to simulate structures of peptides and proteins.

Validation Simulations of a termite BG

As a pre-requisite to performing the simulations of the bacterial (C. cellulovorans; PDB ID 3AHX[2]) and fungal (T. reesi; PDB ID 3AHY[2]) BGs, we developed a methodology for docking and simulating the substrate in the active site (Figure 3). Because no structural data are available for the aforementioned BGs with a substrate in the active site, we validated our methodology using the structure of another family 1 BG of termite origin (Neotermes koshunsis; PDB ID 3VIL[19]), which was crystallized with a substrate surrogate (salicin) in its active site.

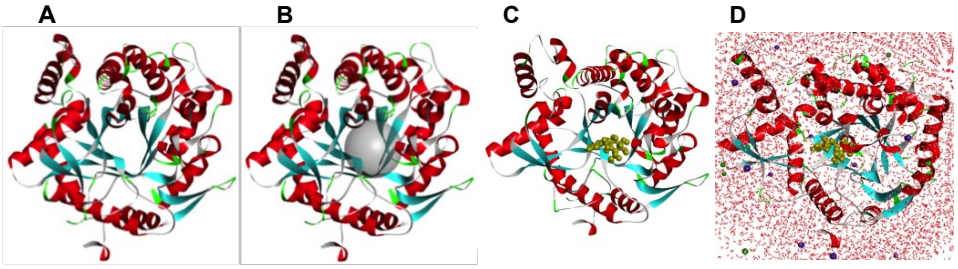

Figure 3. Modeling workflow for obtaining preliminary structures of a BG in a hydrated environment: (A) downloaded X-ray structure of enzyme was prepared at a specific pH, (B) preliminary determination of interaction region of substrate (shown by gray sphere), (C) docking of substrate (shown in yellow), and (D) hydration of periodic 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

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)[19] used for the protein's crystallization by using an algorithm[39] as implemented in Discovery Studio.[39] This algorithm assesses the local environment of each amino acid residue with respect to other residues in the protein structure and the degree of exposure of the amino acid residue to the solvent in order to compute the prominent ionization states of each amino acid residue. In addition, the residue $Glu402_{termite}$[19] was manually deprotonated to reflect the known catalytic ionization state of the nucleophilic base.[17] The substrate

(salicin) was docked into the active site of the enzyme using the CHARMm docking software CDOCKER[40]—the interaction sphere (radius of 7Å) was defined to include the residues, which were shown to interact with the substrate in the crystal structure;[19] the additional docking parameters were the following: pose cluster radius = 0.1, random conformations = 10, dynamic steps = 1,000, and simulated annealing heating to 700 K and then cooling to 300K over 2,000 steps. The lowest-energy configurations were compared to the atomic coordinates of the original X-ray structure[19] whereby successful dockings were evaluated by the conventional standard of Root Means Square Deviation (RMSD) of less than 2.0 Å.[41,42] The docked ligand had an RMSD of 0.382, which met the criteria for a successful docking (Table 1).

Table 1: Validation results comparing interaction distances observed in the crystal structure 3VIL between the active site and the substrate against the mean distance found in a 1 ns MD simulation.

Interaction Pair	Distance in Crystal Structure (Å)	Average Distance in 1 ns MD simulation (Å)
Ser193···O2	2.7	2.56 ± 0.43
Glu402 ···H34	1.77	1.93 ± 0.23
His148···O6	1.91	1.88 ± 0.10
Asn192···O4	1.92	2.02 ± 0.24
Gln45···H35	1.90	2.05 ± 0.25
Glu451···H36	1.78	2.26 ± 0.7

The 3VIL protein was charge-balanced with Na^+ and Cl^- ions following by solvation within a periodic system of explicit water molecules and 0.15 M NaCl. The solvated system was subjected to two EM cycles (1,000 max steps each), heated simulation (300K for 4 ps at 1 fs time step), and MD equilibration (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. Six H-bonding interactions between the substrate and the enzyme, which were deduced from the crystal structure of the active site[2], were monitored throughout the simulations. The MD-predicted values for these H-bond distances were within one standard deviation of the experimental values[19] (Table 1), thus demonstrating the ability of our modeling methodology to

replicate appropriate configurations of the enzyme to facilitate favorable active site-substrate interactions in an aqueous condition very similar to our systems of interest.

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 pH 5, 7, and 7.5 (Figure 2). After the enzymes were retrieved from the PDB database, they were prepared at each pH according to the aforementioned algorithm.[39] We note that the focus of our study was to simulate the enzyme at its most probable ionization state under each pH condition. Therefore, 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—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 [30,31,34,43–45].

In order to model the known starting catalytic ionization states,[2] 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.[6] 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 in the termite BG enzyme and those in 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) was oriented in both the bacterial and fungal active sites was

informed by the binding orientation obtained with the crystal structures of the termite BG[19]—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.[2] All systems were then solvated in a 0.1 M NaCl solution with explicit water molecules in a dodecahedron periodic cell. In order to prevent protein interactions with a periodic image of itself, over 1 nm of water molecules are between the faces of the cell and the protein. These solvated systems were subjected to EM followed by a two-step equilibration. The first equilibration step holds temperature constant (NVT) to equilibrate the system at 300K. During this first equilibration step, the position of the ligand and the protein are restrained and the water molecules are afforded the opportunity to interact and reach a more favorable conformational arrangement with the protein (500 ps). The position restraints on the protein/ligand are lifted for the second equilibration. A pressure coupling thermostat (Parrinello-Rahman barostat) is used in this 2 ns equilibration step to stabilize the pressure of the simulation.

Initially the fungal and the bacterial systems were simulated for 2 ns on local Dell Precision T7610 machines. These results were summarized and submitted to a journal. In order to respond to a request for simulations on the order of 100ns to μs, we obtained access to the High Performance Computer (HPC) resource Stampede (Cite) through the National Science Foundation's network of supercomputers. Subsequent md simulations were run for 200 ns on the Stampede supercomputer using the following parameters. Simulation time step was 2 fs. Short range coulombic and Van der Waals interaction cutoffs were both placed at 1.4 nm. The simulation was run under NPT conditions using the Parinello-Rahman pressure coupling barostat as in the second step of equilibration.

Analysis

The configurations used for analysis were retrieved from the MD production step. For the termite validation study we monitored the dynamics of H-bonding interactions between the substrate and water molecules using Discovery Studio's experimentally-validated[46,47] 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 90-180 degrees). Hydrogen bond analysis of the MD simulations used the default settings (distance < 3.5 Å , angle > 150 degrees) found in GROMACS's g_hbond program. Interaction energies (IEs) were calculated as the sum of the coulombic and Van der Waals forces between two atoms or groupings of atoms.

Results and Discussion

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 2). 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. One Glu became negatively charged and one His became neutral charge from pH 7 to pH 7.5 (Table 2). 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. Three His residues went from being positively charged to neutral from pH 7 to pH 7.5 (Table 2). 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 2). Except for $Glu406_{bacterial}$, all the amino acid residues that underwent a change in their ionization states were located outside of the substrate-binding cleft.

Table 2. 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

pH-Induced Conformational Changes in BGs

To get an idea of the broad conformational dynamics of the bacterial and fungal BGs at different pH values an RMSD (root mean squared deviation) analysis on the protein αC backbone (Figures 3-8) was performed alongside root mean squared fluctuation (RMSF) analysis on specific side chain residues (Figures 9-10). Both measure the structural deviation from the protein's starting conformation (rotation and translation of protein during simulation is accounted for and not included in RMSD/RMSF results). The RMSD of the αC provides a running measure of structural deviation and is often used to assess structural equilibration. RMSF is the average Root Mean Square Fluctuation taken over the entire course of the simulation and provides one summary number.

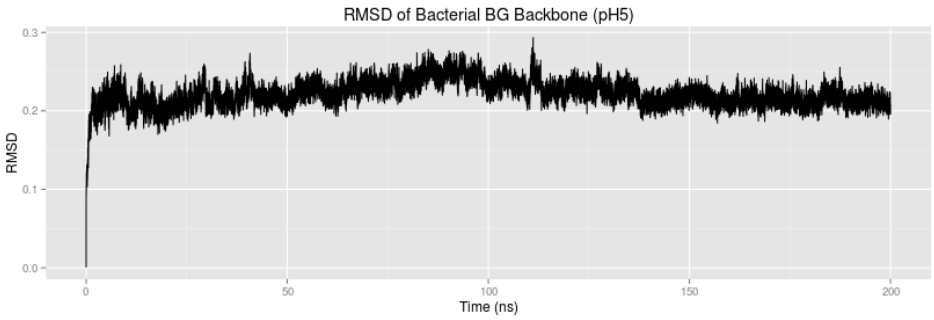

Figure 3. RMSD of protein α-C backbone for Bacterial BG simulation at pH 5.

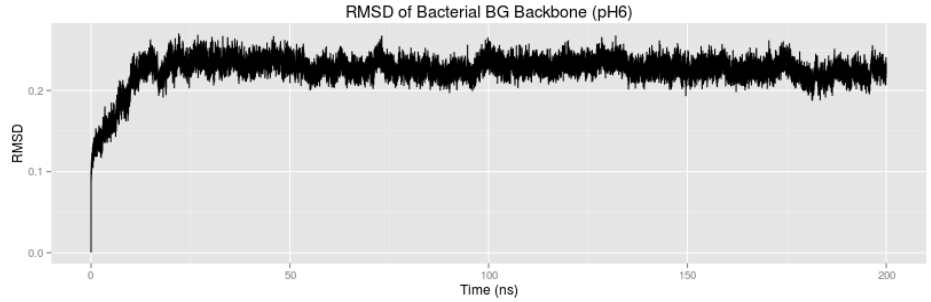

Figure 4. RMSD of protein α-C backbone for Bacterial BG simulation at pH 6.

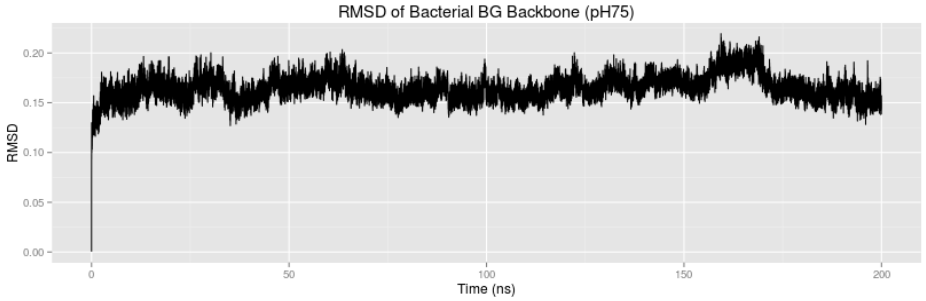

Figure 5. RMSD of protein α-C backbone for Bacterial BG simulation at pH 7.5.

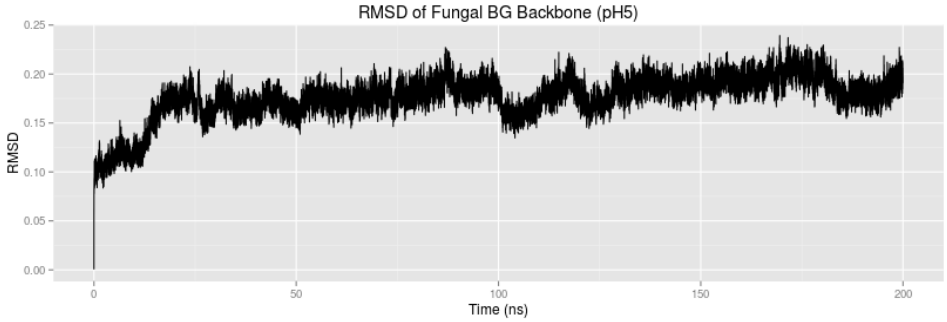

Figure 6. RMSD of protein α-C backbone for Fungal BG simulation at pH 5.

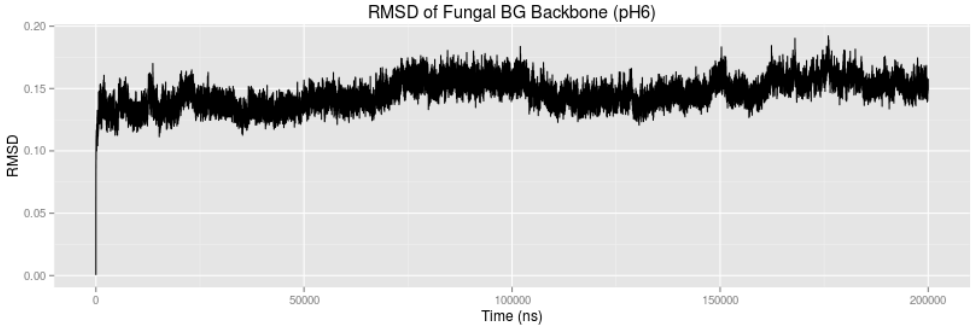

Figure 7. RMSD of protein α-C backbone for Fungal BG simulation at pH 6

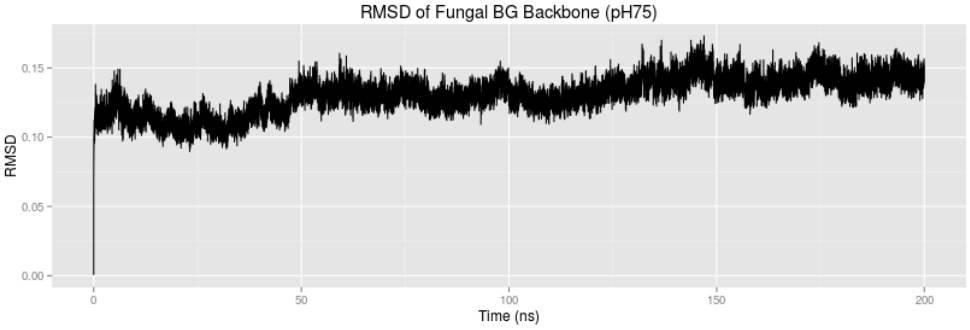

Figure 8. RMSD of protein α-C backbone for Fungal BG simulation at pH 7.5

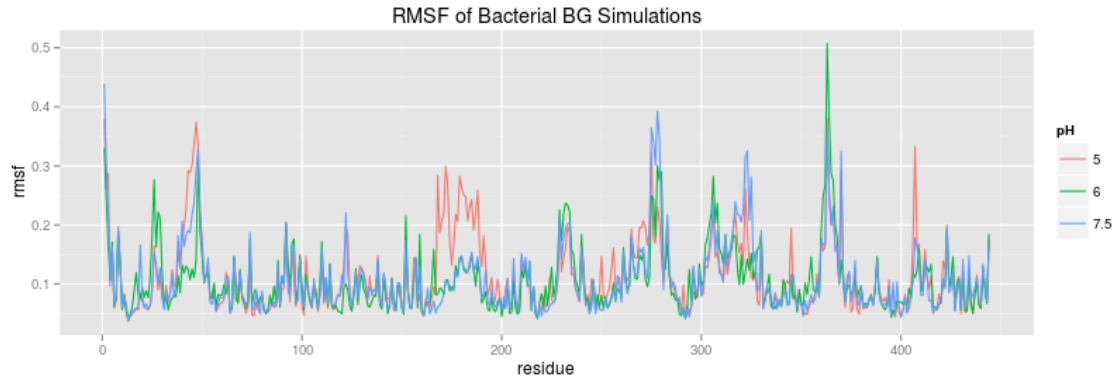

Figure 9. RMSF of protein sidechain backbone for Bacterial BG at pH 5, 6, and 7.5.

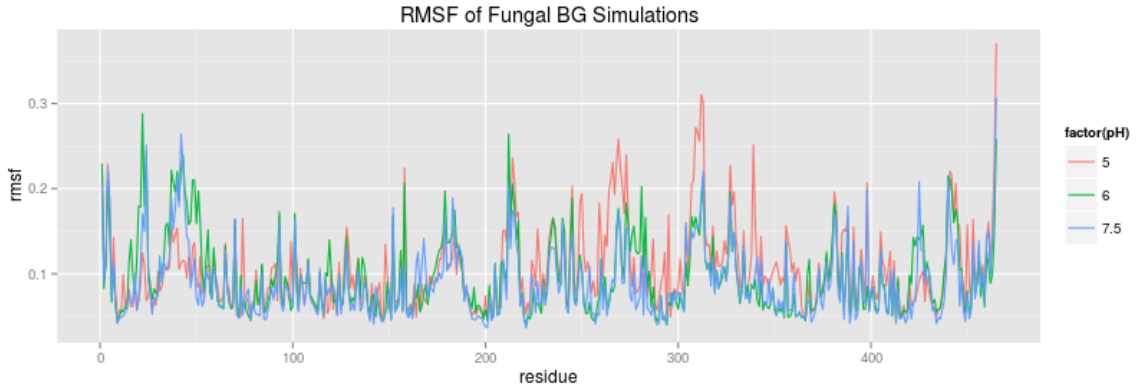

Figure 10. RMSF of protein sidechain backbone for Fungal BG at pH 5, 6, and 7.5.

The bacterial BG simulations at pH 5 and 7.5 display a stabilized RMSD around 0.2 Å and 0.15 Å respectively. This indicates that at the start of the MD simulations the protein structure had equilibrated. The RMSD of the bacterial BG at simulation pH 6 still changes by over 0.15 Å in the first ~ 20 ns of simulation indicating an un-equilibrated structure. After ~ 20 ns of simulation the RMSD of the bacterial BG simulation at pH 6 equilibrates at a RMSD just above 0.2 Å. RMSD values of Fungal BG simulations at pH 6 and 7.5 reveal an equilibrated system at the start of MD with stable RMSD values fluctuating around 0.15 Å. The simulated fungal BG at pH 5 RMSD took ~ 20 ns to reach structural equilibration. RMSF values of amino acid side chain indicate the average structural deviation of specific amino acid side chain residues. Secondary structure has a large impact on the flexibility and consequently structural deviation of amino acid residues with amino acid residues

located on the secondary structure loops being the most flexible (Figure 9, Figure 10). Despite similar RMSF patterns relating to secondary structure, some RMSF values appear to be pH-dependent. For instance when the RMSF values for the active site (Figure 11, Figure 12) are plotted, one can see that some active site side chain structural deviation is pH-dependent while others are more independent of pH. For the bacterial simulations, residues His121, Glu165, Glu166, Tyr296, Glu352, Glu406, and Glu407 have RMSF values that vary with pH, while Gln20 and Trp399 do not. The catalytic glutamate residues, Glu166 and Glu352, exhibit the largest RMSF values at pH6. Fungal RMSF values of the active site (AS) side chains show wide fluctuations at different pH, except for Trp417 and Tyr 298 whose RMSF values remain clustered at different pH conditions. We continue to explore pH-dependent changes in the active site and ramifications for substrate binding in later paragraphs.

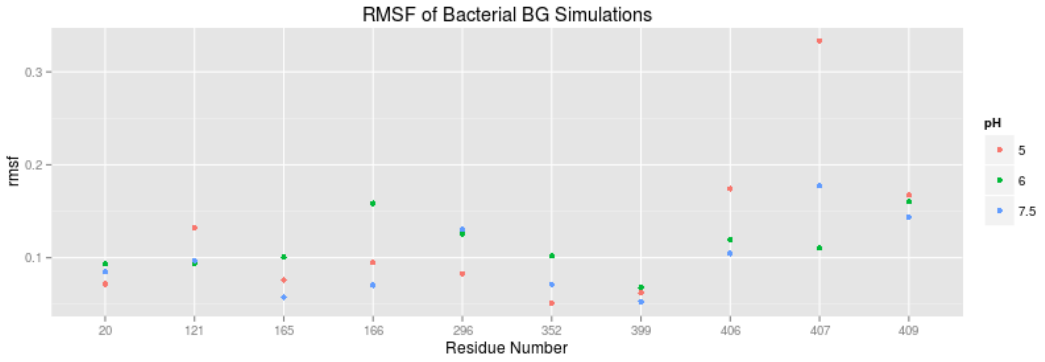

Figure 11. RMSF of protein AS side chain residues (Gln20, His121, Asn165, Glu166, Tyr296, Glu352, Trp399, Glu406, Trp407, Glu409)

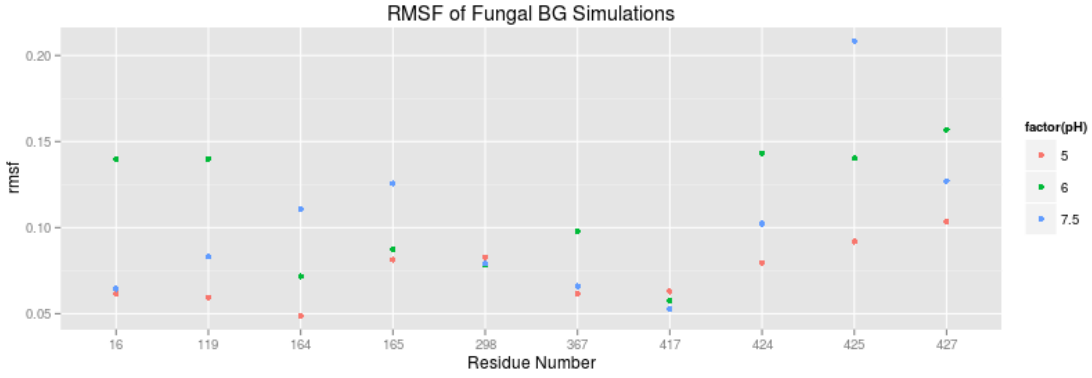

Figure 12. RMSF of protein AS side chain residues (Gln16, His119, Asn164, Glu165, Tyr298, Glu367, Trp417, Glu424, Trp427, Asp427)

pH-Induced Interactions affect Ligand-Active Site and Ligand-Water Dynamics

To get a general idea of the substrate binding dynamics at each enzyme pH condition, summary interaction plots analyzing ligand-substrate interaction energy, ligand-substrate hydrogen bonds, and ligand-water hydrogen bonds were graphed for both the bacterial (Figure 13, 14, 15) and fungal BG (Figure 16, 17, 18) enzyme simulations. Each enzyme pH condition was evaluated relative to the other pH conditions present in the enzyme. The Bacterial BG at pH 5 maintained a high number of AS-ligand interactions (Figure 14), as well as high AS-lig interaction energy (Figure 13) until about 100 ns. After 100 ns, both the number of active site-ligand hydrogen bonds, as well as active site-ligand interaction energy drops down to zero. Simultaneously as ligand-active site interactions decrease, ligand-water interaction energies increase (Figure 15). Visualization of the simulation confirms the ligand becoming solvated and exiting the binding site around 100 ns. Bacterial BG simulations at pH 6 show low ligand-active site hydrogen bond interactions (Figure 14), as well as markedly low interaction energy levels (more than 100 kJ/mol more unfavorable than other pH conditions, Figure 13). The ligand participates in a high degree of hydrogen bond interactions with water (Figure 15) and eventually exits the active site. Bacterial BG simulation at pH 7.5 indicate sustained AS-ligand interaction energy (Figure 13) and active site-ligand hydrogen bonding (Figure 14) throughout the entire simulation. Visualization confirms that while the ligand exited the binding cleft, the ligand remained in the vicinal fringe of the active site, still participating in direct interactions with active site residues. General analysis of the fungal BG simulations at pH 5 indicate tight and strong interactions with the active site initially (Figure 16, Figure 17), but it is soon rapidly solvated (Figure 18) and leaves the active site. The fungal simulation at pH 6 show low active site-ligand hydrogen bond interactions (Figure 17), low interaction energies (Figure 16), and a readiness to solvate (Figure 18). Despite relatively low interactions, the ligand remains near the active site until

around 120 ns of simulations. At pH 7.5 for the fungal simulations the ligand participates in the highest number of interactions (Figure 17) until around 100 ns where the ligand is fully solvated and no further active site ligand interactions are observed. In order to get a finer resolution for which active site residues were responsible for the active site-ligand hydrogen bond interactions and active site- ligand interaction, energies were broken down into active site residue components.

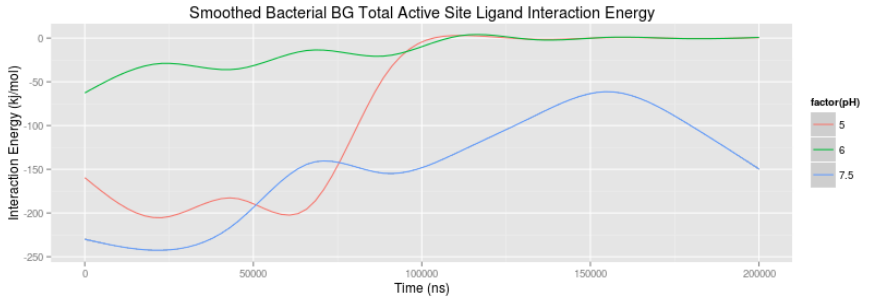

Figure 13. Smoothed total IE for bacterial BG simulations at pH 5, 6, and 7.5 for the 200 ns simulation.

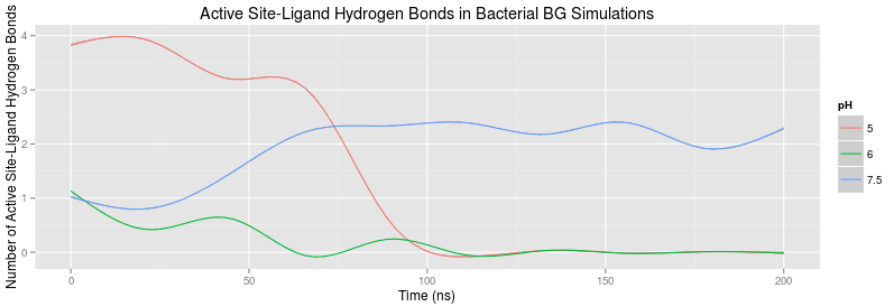

Figure 14. Smoothed number of AS-ligand hydrogen bonds as a function of simulation time of the bacterial BG at pH 5, 6, and 7.5.

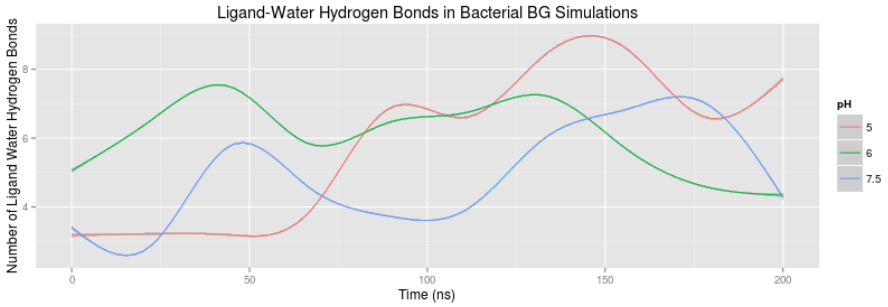

Figure 15. Smoothed number of ligand-water hydrogen bonds as a function of simulation time of the bacterial BG at pH 5, 6, and 7.5.

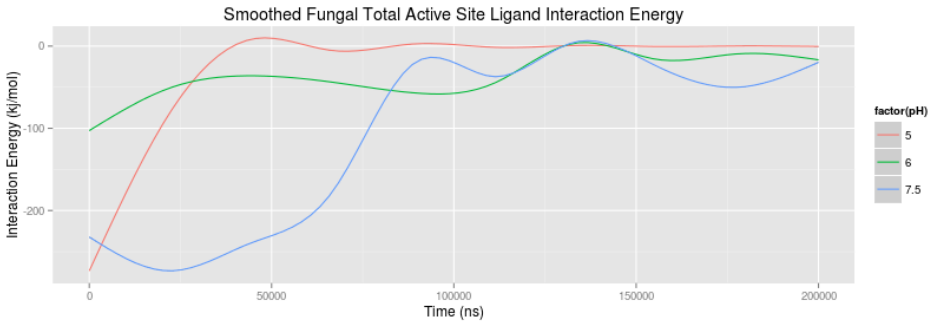

Figure 16. Smoothed total IE for fungal BG simulations at pH 5, 6, and 7.5 for the 200 ns simulation.

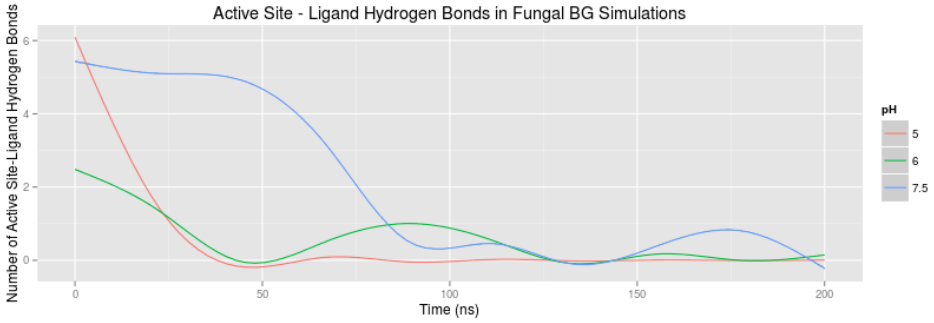

Figure 17. Smoothed number of ligand-water hydrogen bonds as a function of simulation time of the fungal BG at pH 5, 6, and 7.5.

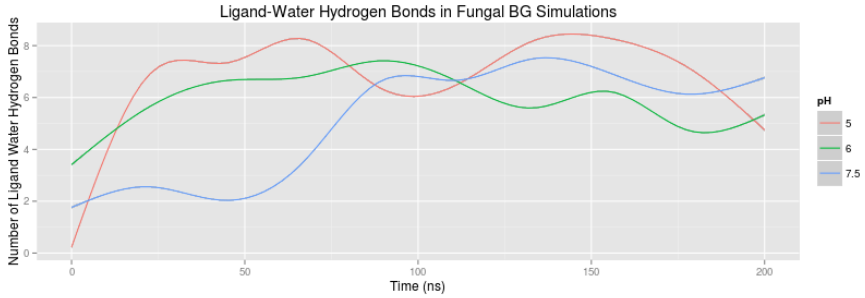

Figure 18. Smoothed number of ligand-water hydrogen bonds as a function of simulation time of the fungal BG at pH 5, 6, and 7.5.

In the bacterial pH 5 BG simulations, the interaction energies hovered around -200 kJ/mol before trailing down to 0 kJ/mol around 100 ns (Figure 19). This interaction energy was broken down into the non-catalytic residue contribution (all active site residues except Glu166 and Glu352), the nucleophillic Glu352 contribution, and the energy resulting from the acidic/basic interaction with the substrate. During the time in which the ligand was interacting with the substrate, about -120 kJ/mol came from non-catalytic residues, -60 kJ/mol from the nucleophillic glutamate residue, and about -20 kJ/mol from the acidic/basic glutamate residue. Multiple catalytic and non-catalytic (NC) residue interactions with the ligand were observed during the bacterial BG simulations at pH 5. The catalytic interaction between acidic/basic Glu166 and O_2 (Figure 20) of the ligand as well as the substrate stabilization interaction of the nucleophillic Glu352 and H_{34} (Figure 21) of the ligand occurred prominently during the first 100 ns. Additional NC interactions involving Glu166/His121/Gln20 and the ligand were strongly observed before the ligands departure from the active site (Figure 22-24).

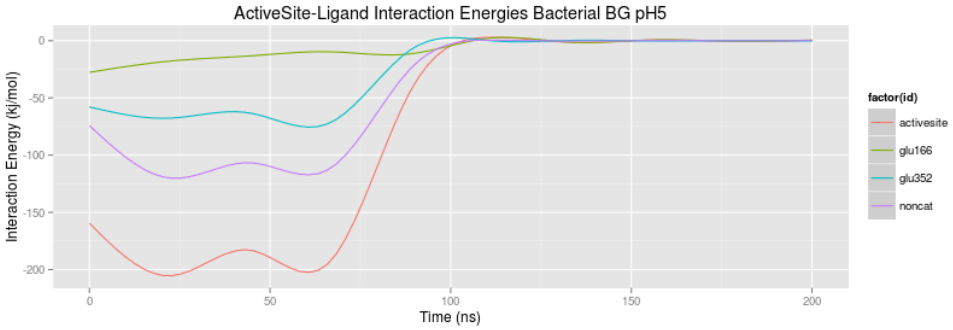

Figure 19. Smoothed IEs of bacterial BG simulations broken down into IE contributions from the acidic/basic Glu166, the nucleophillic Glu352, the remaining eight non-catalytic AS residues, as well as the sum of the entire AS residues.

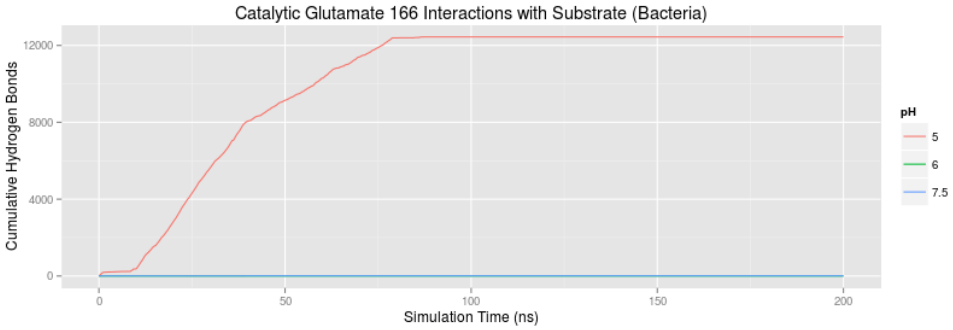

Figure 20. Cumulative total of catalytic Glu166 interactions with the ligand at pH 5, 6, and 7.5)

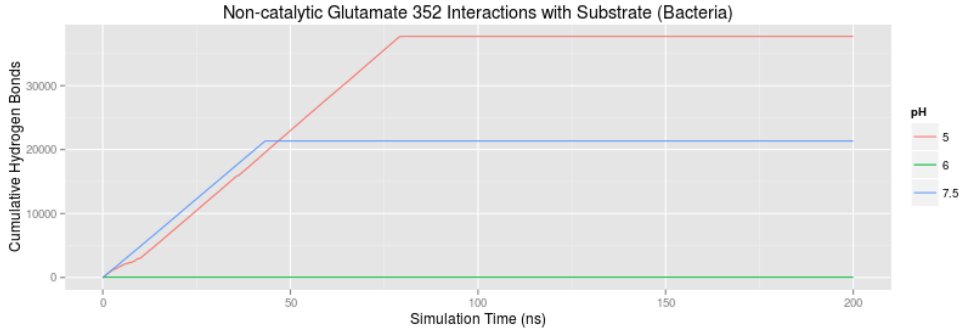

Figure 21. Cumulative total of nucleophillic Glu352 interactions with the ligand at pH 5, 6, and 7.5.

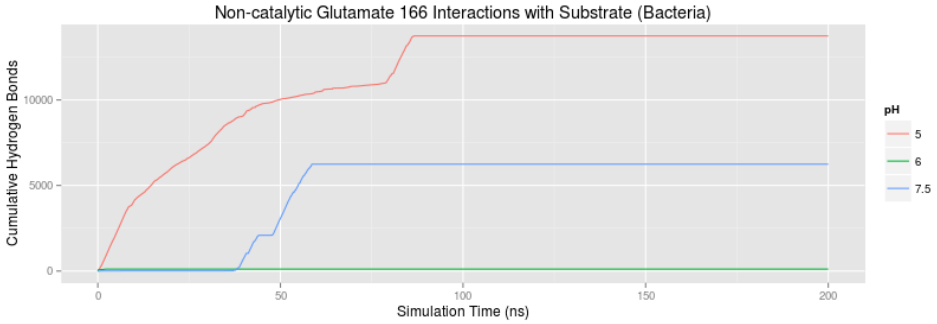

Figure 22. Cumulative total of NC Glu166 interactions with the ligand at pH 5, 6, and 7.5.

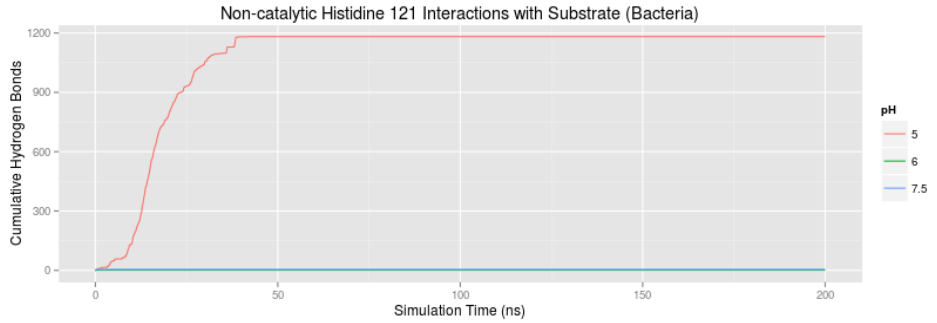

Figure 22. Cumulative total of NC His121 interactions with the ligand at pH 5, 6, and 7.5.

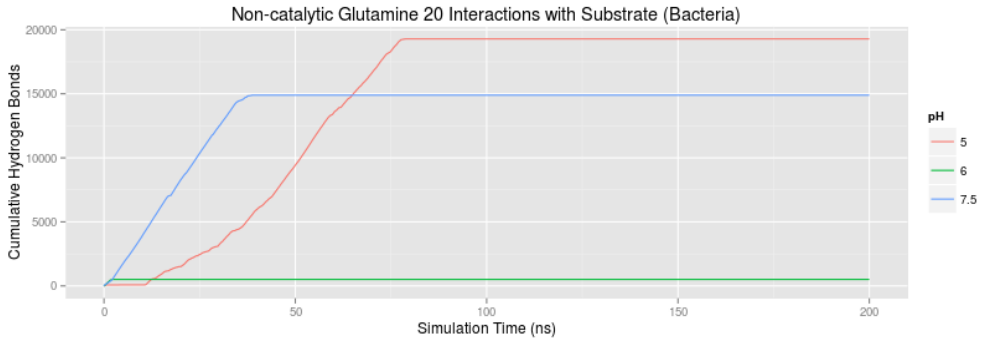

Figure 23. Cumulative total of NC Gln20 interactions with the ligand at pH 5, 6, and 7.5.

The bacterial pH BG simulations start at a much lower -60 kJ/mol and diminish to an IE around 0 kJ/mol by 120 ns. When broken down into energy components, almost the entirety of the IE was contributed from NC residues, with little interaction energy coming from either catalytic glutamate residue (Figure 24). Analysis of the presence of known substrate binding interactions with AS residues confirms the lack of specific AS-ligand binding hydrogen bonds deduced from crystal structures. The ligand still remains in the AS area until around 120 ns, but is weakly bound

energetically by the active site and interacts with the active site in an orientation unfavorable for traditional substrate binding. In fact, the only known AS-lig stabilizing interactions that were observed in the simulations occurred between the ligand and Trp399 for only a brief portion of simulation time (Figure 25).

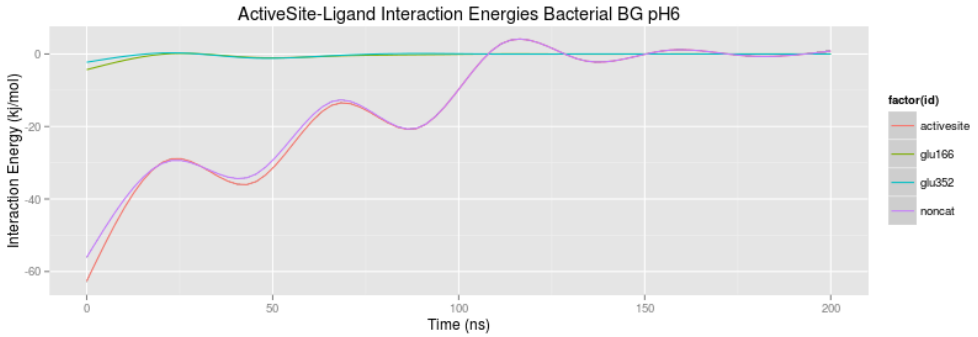

Figure 24. Smoothed IEs of bacterial BG simulations broken down into IE contributions from the acidic/basic Glu166, the nucleophillic Glu352, the remaining eight NC AS residues, as well as the sum of the entire AS residues.

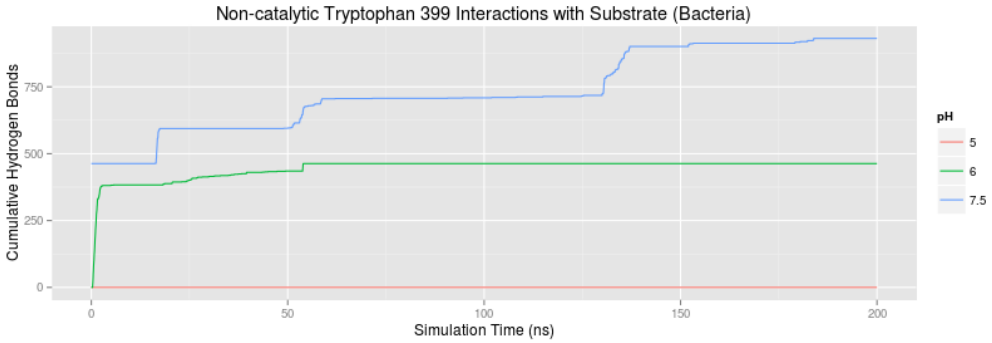

Figure 25. Cumulative total of NC Trp399 interactions with the ligand at pH 5, 6, and 7.5.

At pH 7.5, the simulations of the bacterial BG, the AS-ligand IEs and the number of AS-ligand hydrogen bonds undulated up and down the entire length of the simulations with the IEs and number of hydrogen bonds never reaching 0 kJ/mol indicating a not entirely complete departure from the active site (Figure 26). During the first ~ 75 ns of simulation the catalytic glutamates contributed slightly to the IE profile and from ~ 75-200 ns the predominant driving force for IE was the NC residues. When the hydrogen bond interactions are broken down by AS residues an interesting story emerges. There are several residues (Glu166, Glu352, Gln20) which hydrogen bond with the ligand

for a period of time and then stop participating with interactions with the substrate (Figure 20, Figure 21, Figure 23). The ligand, however, made consistent interactions with the NC residue Glu406 throughout the simulations remaining in the exterior of the active site but removed from the binding cleft (Figure 27). At pH 7.5 the residue Glu406 becomes negatively charged, a state that is not present at pH 5, and pH 6. It could be that the additional IE coming from the increased columbic interaction with a charged Glu406 could be enough to make the ligand "catch" just outside of the binding cleft in the exterior of the active site. This identified spot potentially could compete with the binding cleft for ligand binding at pH 7.5.

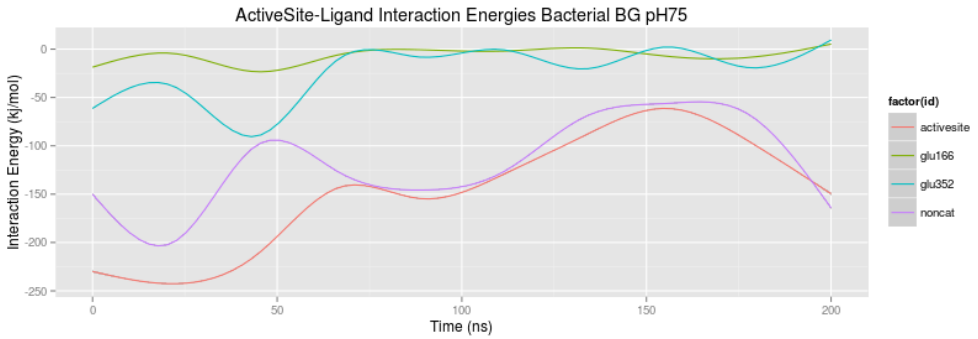

Figure 26. Smoothed IEs of bacterial BG simulations broken down into IE contributions from the acidic/basic Glu166, the nucleophillic Glu352, the remaining eight NC AS residues, as well as the sum of the entire AS residues.

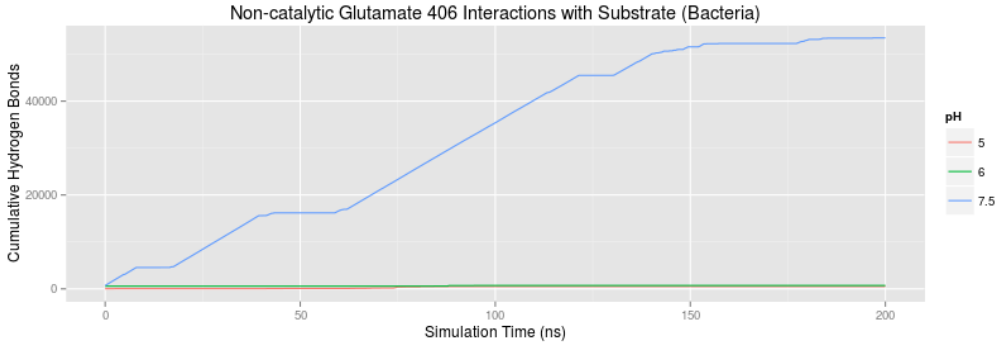

Figure 27. Cumulative total of NC Glu406 interactions with the ligand at pH 5, 6, and 7.5.

Simulations of the fungal BG at pH 5 display a high initial favorable interaction with the ligand that quickly reduces down to 0 kJ/mol IE by ~ 45 ns of simulation (Figure 28). This IE is comprised of predominantly NC and nucleophillic Glu interactions with the ligand. Substrate binding interactions

with the nucleophillic Glu 367 (Figure 29) and the NC AS residues His119, Glu424, Gln16, Asn427, and Trp417 (Figure 30-34) are observed during the first 25 ns of simulation indicating a transition from the binding cleft (first 25 ns) to the exterior of the active site (25-50 ns) followed by a complete departure of the ligand from the active site.

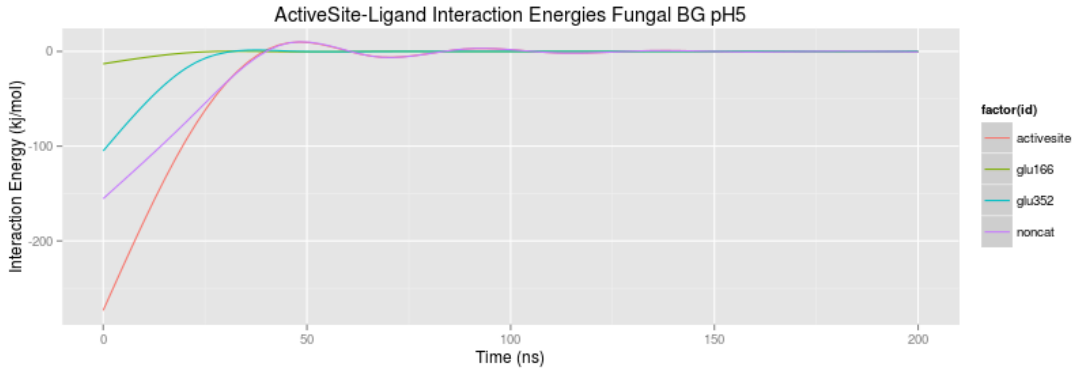

Figure 28. Smoothed IEs of fungal BG simulations broken down into IE contributions from the acidic/basic Glu165, the nucleophillic Glu367, the remaining eight NC AS residues, as well as the sum of the entire AS residues.

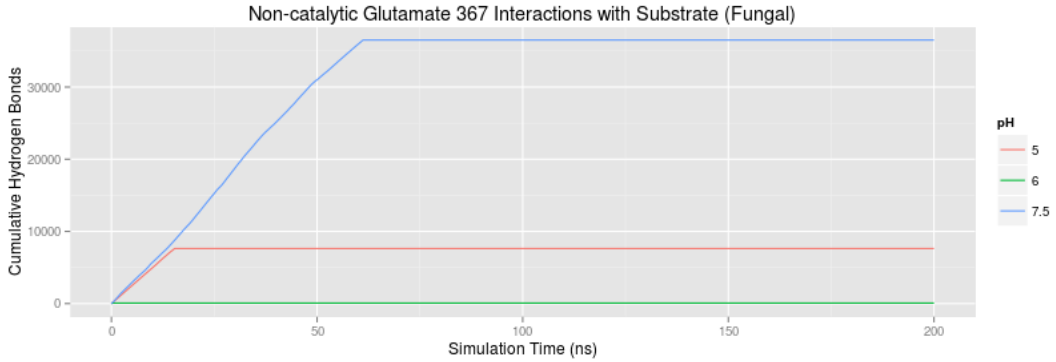

Figure 29. Cumulative total of nucleophillic Glu367 interactions with the ligand at pH 5, 6, and 7.5.

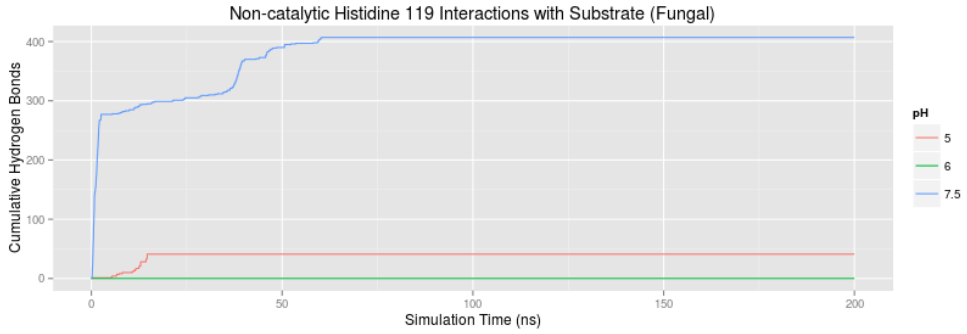

Figure 30. Cumulative total of NC His119 interactions with the ligand at pH 5, 6, and 7.5.

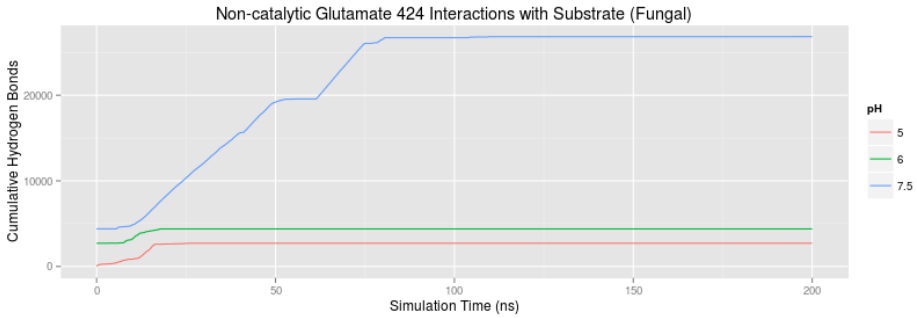

Figure 31. Cumulative total of NC Glu424 interactions with the ligand at pH 5, 6, and 7.5.

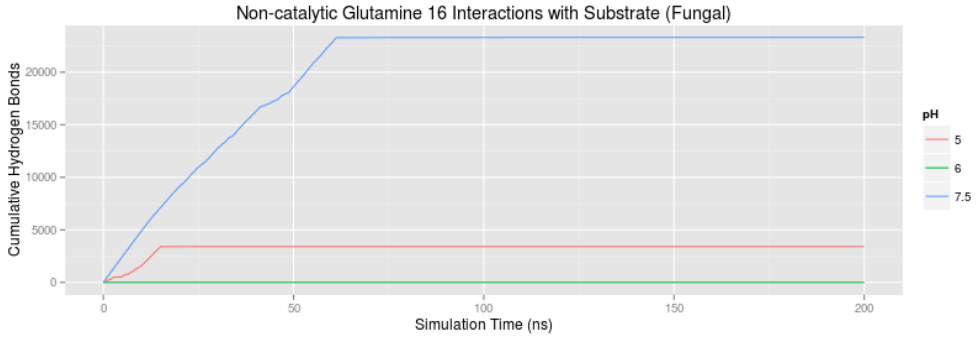

Figure 32. Cumulative total of NC Gln16 interactions with the ligand at pH 5, 6, and 7.5.

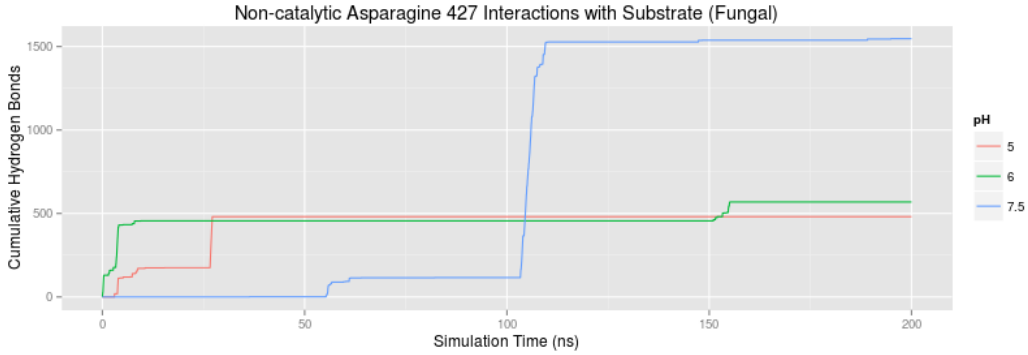

Figure 33. Cumulative total of NC Asn427 interactions with the ligand at pH 5, 6, and 7.5.

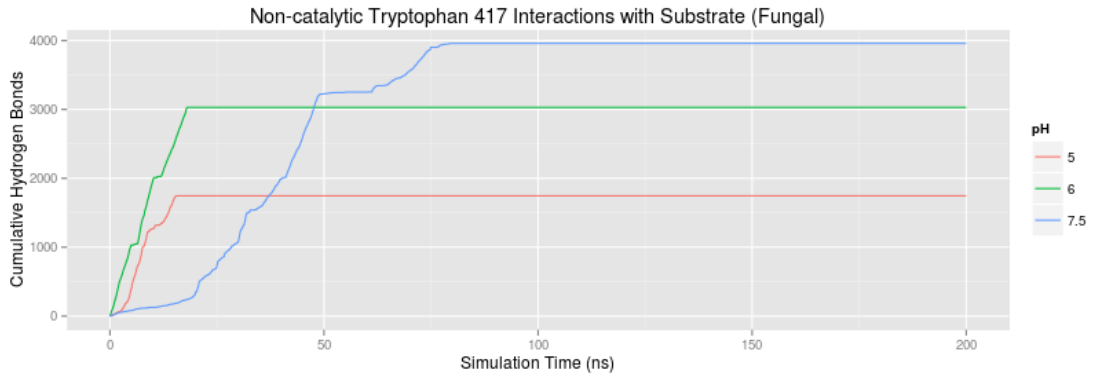

Figure 34. Cumulative total of NC Trp417 interactions with the ligand at pH 5, 6, and 7.5.

The fungal simulations at pH 6 had over 100 kJ/mol less favorable IE than the next closest IE. When broken down into its contributions the IE has a large NC component, but interestingly a large component from interactions with the catalytic glutamate (Glu165) residue (Figure 35). It is the only simulation where, on average, the non-charged acidic/basic glutamate residue contributes more energy to the total IE than the charged nucleophillic glutamate residue (Glu367). This story is corroborated with AS-ligand hydrogen bond information (Figure 36). No ligand-stabilizing hydrogen bond interactions were observed between the nucleophillic glutamate (Figure 29) and the ligand, while strong catalytic and NC interactions were recorded from the catalytic glutamate residue Glu165 (Figure 36). During the course of simulations the substrate-stabilizing hydrogen bonds appear to occur strongly over small time frames. Over the course of the first 25 ns NC AS-ligand interactions involving Glu165, Asn427, Trp417 occur strongly, but briefly (Figure 37, Figure 33, Figure 34). Then, at around 75 ns of simulation, the catalytic Glu165-O_2 interaction occurs intensely for about 25 ns before subsiding (Figure 36). This time frame corresponds to the ligands exit from the binding site to the exterior of the AS. While this interaction is the foundation to the hydrolysis of the ligand, none of the other documented substrate stabilizing interactions is occurring during this same simulation time window possibly indicating unfavorable conditions for catalysis.

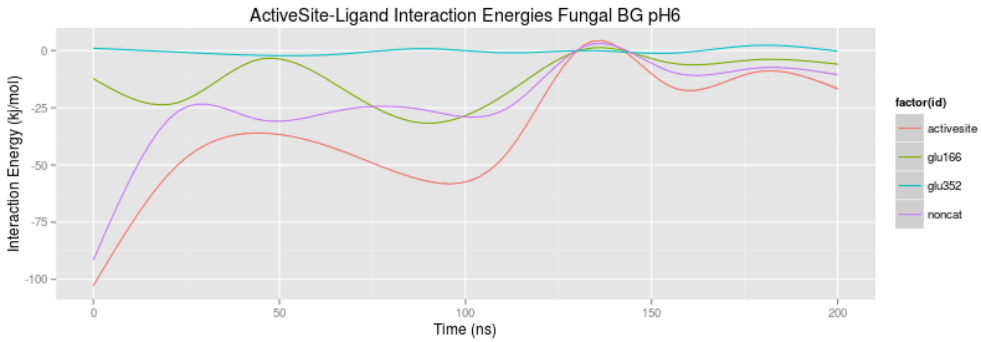

Figure 35. Smoothed IEs of fungal BG at pH 6 simulations broken down into IE contributions from the acidic/basic Glu165, the nucleophillic Glu367, the remaining eight NC AS residues, as well as the sum of the entire AS residues.

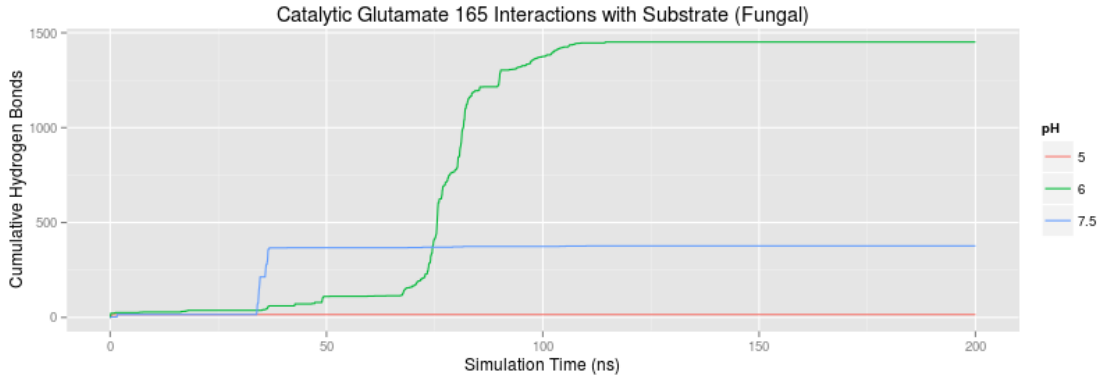

Figure 36. Cumulative total of Catalytic Glu165 interactions with the ligand at pH 5, 6, and 7.5.

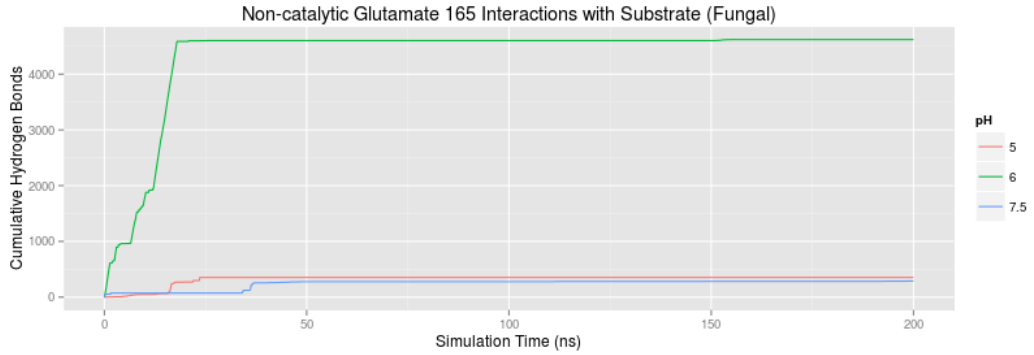

Figure 37. Cumulative total of NC Glu165 interactions with the ligand at pH 5, 6, and 7.5.

The fungal simulation of the BG enzyme at pH 7.5 showed initial strong IE with the ligand and multiple hydrogen bond interactions with the AS . Energy components of the IE show that a large portion of the favorable IE is coming from the IE of the nucleophillic glutamate's interaction with the substrate (accounts for ~ -100kj/mol out of the total ~ -250 kj/mol, Figure 38). IE begins to decline

around 50 ns of simulation and by 100 ns hovers around 0 kJ/mol. Corresponding to the nucleophillic glutamate's large IE contribution, there are consistent substrate-binding hydrogen bonds involving the nucleophillic glutamate (Glu367) for the first 60 ns of the simulation (Figure 29). The catalytic acidic/basic Glu165 interaction (Figure 36) was observed strongly for a brief period of simulation time around 40 ns. NC interactions involving Glu165, Gln16, Glu424, Asn165, and His119 were observed in the first 100 ns of simulation, with most of the interactions occurring before 60 ns of simulation (Figure 37, Figure 32, Figure 31, Figure 39, Figure 30).

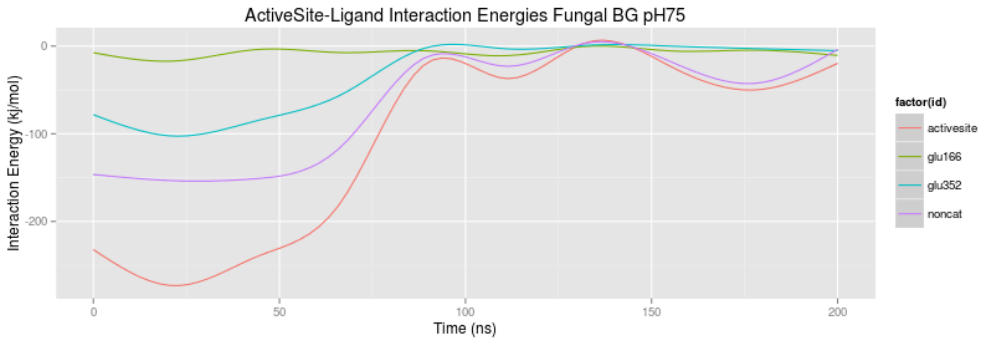

Figure 38. Smoothed IEs of fungal BG at pH 7.5 simulations broken down into IE contributions from the acidic/basic Glu165, the nucleophillic Glu367, the remaining eight NC AS residues, as well as the sum of the entire AS residues.

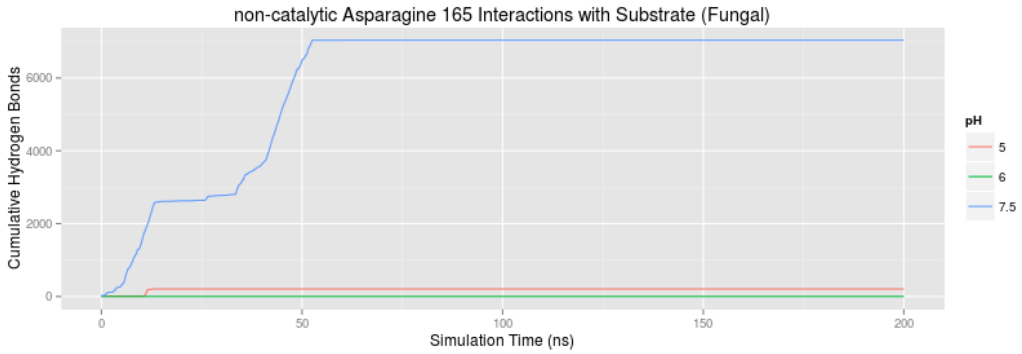

Figure 39. Cumulative total of NC Asn165 interactions with the ligand at pH 5, 6, and 7.5.

Conclusion

The reviewer critique of needing more coverage of simulation time where substrate-binding interactions occur was correct. Extending the simulation run from 2 ns to 200 ns did not increase the amount of time observing AS-lig binding interactions. In retrospect, it is no surprise that the ligand

will explore favorable conformational arrangements that involve solvation and leaving the active site during the course of a 200 ns simulation run. Maybe the product of hydrolysis, a known inhibitor, would have remained tightly bound to the active site, but substrate binding is different. Going forward, I would recommend taking the equilibrated structures after approximately 20 ns of simulation and then proceed with the ligand positioning and subsequent MD steps. Instead of MD simulations of 200 ns I would create 10 identical starting simulations for each pH enzyme condition, randomize the starting velocities of the atoms, and then perform 5-10ns MD runs and analyze the following output. This would give a better statistical representation of the effect of pH on ligand-binding.

There is a silver lining to performing the longer simulation runs. There were a lot of free acquired resources that came with figuring out how to simulate our system more efficiently for longer simulation time periods (GROMACS, Stampede HPC). We now have the resources to fail faster and hopefully fail forwards faster with molecular modeling. Setting up the right modeling system for the right question is essential to getting informative answers and often it is only after the simulations are performed that one knows whether it was the correct or incorrect way to model the system. With the diagnostic information gathered from our 200 ns runs on the super computer we applied for a larger research allocation request from Stampede HPC for extensions to this project. Additionally there are possibly interesting results buried in the 200 ns simulation runs. What are the mechanisms through which the ligand exits the active site? During the parts of the simulation where the ligand left the active site it often interacted with the exterior surface of the protein. Identifying sections of the protein where organic moieties are attracted to could generate some hypotheses on where these extracellular enzymes may bind to organic matter.

REFERENCES

1. Singhania, R. R., Patel, K., Sukumaran, R., Larroche, C. & Pandey, A. Role and significance of beta-glucosidases in the the hydrolysis of cellulose for bioethanol production. *Bioresour. Technol.* **127,** 500–507 (2013).
2. Jeng, W.-Y. *et al.* Structural and functional analysis of three β-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis. *J. Struct. Biol.* **173,** 46–56 (2011).
3. Wieder, W. R., Bonan, G. B. & Allison, S. D. Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Change* **3,** 909–912 (2013).
4. Knight, T. R. & Dick, R. P. Differentiating microbial and stabilized β-glucosidase activity relative to soil quality. *Soil Biol. Biochem.* **36,** 2089–2096 (2004).
5. Mariscal-Sancho, I., Santano, J., Mendiola, M.-A., Peregrina, F. & Espejo, R. Carbon Dioxide Emission Rates and [beta]-Glucosidase Activity in Mediterranean Ultisols Under Different Soil Management. *Soil Sci. Sept. 2010* **175,** 453–460 (2010).
6. Teugjas, H. & Väljamäe, P. Selecting β-glucosidases to support cellulases in cellulose saccharification. *Biotechnol. Biofuels* **6,** 105 (2013).
7. Jørgensen, H., Kristensen, J. B. & Felby, C. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. *Biofuels Bioprod. Biorefining* **1,** 119–134 (2007).

8. Percival Zhang, Y.-H., Himmel, M. E. & Mielenz, J. R. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol. Adv.* **24,** 452–481 (2006).
9. Zimmerman, A. R. & Ahn, M.-Y. in *Soil Enzymology* (eds. Shukla, G. & Varma, A.) 271–292 (Springer Berlin Heidelberg, 2011). at <http://link.springer.com/chapter/10.1007/978-3-642-14225-3_15>
10. Acosta-Martínez, V. & Tabatabai, M. A. Enzyme activities in a limed agricultural soil. *Biol. Fertil. Soils* **31,** 85–91 (2000).
11. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* **75,** 5111–5120 (2009).
12. Lan, T. Q., Lou, H. & Zhu, J. Y. Enzymatic Saccharification of Lignocelluloses Should be Conducted at Elevated pH 5.2–6.2. *BioEnergy Res.* **6,** 476–485 (2013).
13. Creighton, T. E. *Proteins: Structures and Molecular Properties*. (Macmillan, 1993).
14. Schomburg, I. *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* **41,** D764–D772 (2013).
15. Murao, S., Sakamoto, R. & Arai, M. Cellulases of Aspergillus aculeatus. *Methods Enzymol.* **160,** 274–299 (1988).
16. Zibaee, A., Bandani, A. R. & Ramzi, S. 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,** 633–641 (2009).

17. Vuong, T. V. & Wilson, D. B. Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnol. Bioeng.* **107,** 195–205 (2010).
18. Badieyan, S., Bevan, D. R. & Zhang, C. Probing the Active Site Chemistry of β-Glucosidases along the Hydrolysis Reaction Pathway. *Biochemistry (Mosc.)* **51,** 8907–8918 (2012).
19. Jeng, W.-Y. *et al.* 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,** 829–838 (2012).
20. Chuenchor, W. *et al.* The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.* **173,** 169–179 (2011).
21. Withers, S. G., Rupitz, K., Trimbur, D. & Warren, R. A. J. Mechanistic consequences of mutation of the active site nucleophile Glu 358 in Agrobacterium .beta.-glucosidase. *Biochemistry (Mosc.)* **31,** 9979–9985 (1992).
22. Kaper, T. *et al.* Substrate Specificity Engineering of β-Mannosidase and β-Glucosidase from Pyrococcus by Exchange of Unique Active Site Residues†. *Biochemistry (Mosc.)* **41,** 4147–4155 (2002).
23. Huber, R. E., Hlede, I. Y., Roth, N. J., McKenzie, K. C. & Ghumman, K. K. His-391 of β-galactosidase (Escherichia coli) promotes catalyes by strong interactions with the transition state. *Biochem. Cell Biol.* **79,** 183–193 (2001).

24. Wang, T. *et al.* Directed evolution for engineering pH profile of endoglucanase III from Trichoderma reesei. *Biomol. Eng.* **22,** 89–94 (2005).

25. Wohlfahrt, G., Pellikka, T., Boer, H., Teeri, T. T. & Koivula, A. Probing pH-Dependent Functional Elements in Proteins: Modification of Carboxylic Acid Pairs in Trichoderma reesei Cellobiohydrolase Cel6A†. *Biochemistry (Mosc.)* **42,** 10095–10103 (2003).

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

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

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

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

30. Machuqueiro, M. & Baptista, A. M. The pH-Dependent Conformational States of Kyotorphin: A Constant-pH Molecular Dynamics Study. *Biophys. J.* **92,** 1836–1845 (2007).

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

32. Langella, E., Improta, R. & Barone, V. Checking the pH-Induced Conformational Transition of Prion Protein by Molecular Dynamics Simulations: Effect of Protonation of Histidine Residues. *Biophys. J.* **87,** 3623–3632 (2004).
33. Tan, J. *et al.* 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,** 25–40 (2005).
34. Machuqueiro, M. & Baptista, A. M. Acidic range titration of HEWL using a constant-pH molecular dynamics method. *Proteins Struct. Funct. Bioinforma.* **72,** 289–298 (2008).
35. Zhou, J.-M., Zhou, J.-H., Meng, Y. & Chen, M.-B. 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,** 157–165 (2006).
36. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4,** 187–217 (1983).
37. Accelrys Software Inc. *Discovery Studio Modeling Environment*. (Accelrys Software Inc., 2013).
38. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614 (2009).
39. Spassov, V. Z. & Yan, L. A fast and accurate computational approach to protein ionization. *Protein Sci.* **17,** 1955–1970 (2008).

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

41. Roche, O., Kiyama, R. & Brooks, C. L. Ligand−Protein DataBase: Linking Protein−Ligand Complex Structures to Binding Data. *J. Med. Chem.* **44,** 3592–3598 (2001).

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

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

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

45. Campos, S. R. R., Machuqueiro, M. & Baptista, A. M. Constant-pH Molecular Dynamics Simulations Reveal a β-Rich Form of the Human Prion Protein. *J. Phys. Chem. B* **114,** 12692–12700 (2010).

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

47. Bissantz, C., Kuhn, B. & Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **53,** 5061–5084 (2010).