Proceedings of The National Conference
On Undergraduate Research (NCUR) 2017
University of Memphis, TN
April 7-9, 2017

Measuring The Binding Energy Of Glucose To The Glucose/Galactose Binding Protein Computationally

Nicholas R. Rigel
Department of Chemistry
Wittenberg University
200 West Ward Street
Springfield, Ohio 45501- 0720 USA

Faculty Advisor: Dr. Amil G. Anderson

Abstract

The Glucose/Galactose Binding Protein (GGBP) is crucial to bacterial chemotaxis in *E. coli* and other bacteria, binding either to glucose or galactose[1]. GGBP is an α/β protein consisting of two globular Rossman fold domains joined by three peptide segments[2,3]. Since GGBP binds to glucose in *E. coli*, it has the possible use as a bio-indicator in diabetes patients[4]. GGBP normally exists in its closed conformation and opens to bind to glucose[5]. GGBP binds to glucose with a hinge feature that causes the protein to close around the sugar. A hinge angle exists between a residue in each domain and a residue near the binding pocket. An umbrella sampling molecular dynamics method was implemented to not only explore the conformation change when GGBP binds to glucose, but also to measure the Gibb's free energy of binding. An angle change of 23.5° was observed experimentally, while 22.0° was observed during the umbrella sampling[1]. The experimental Gibb's free energy of binding of GGBP to glucose is −9.1 kcal/mol, while a binding energy of −10.2(±0.9) kcal/mol was calculated by umbrella sampling[6].

Keywords: Glucose/Galactose Binding Protein, Free Energy of Binding Calculations, Molecular Dynamics

1. Introduction:

The simulations from the umbrella sampling method utilized Molecular Dynamics (MD) simulations. MD is a classical mechanics method in which the atoms of the system are represented by slightly charged balls and the bonds by springs[7]. MD makes it possible to analyze very large systems at a lesser computational expense than Quantum Mechanics. The "system" consists of the protein, any ligands bound to the protein, and solvation molecules. Most commonly, the system is built in a box, with periodic boundary conditions, with a protein solvated by water molecules. The composition and coordinates for the protein comes from a PDB file that provides the initial positions of all of the protein's atoms. A force field is selected to calculate the interactions between the bonded and non-bonded atoms in the system. GROMOS-96, as implemented in this investigation, is called a united atom model. In GROMOS-96, the carbon and attached hydrogen atoms are represented by a single group centered on the carbon atom[8]. Using the Ergodic Hypothesis, a long simulation of a single protein molecule over a long time period can represent an ensemble of many protein molecules in solution[9]. Strains in the initially built system are eliminated by periodic EM process using MD simulations. Next, a Boltzmann Distribution of velocities representing the simulation temperature is applied to yield initial velocities of atoms[9]. Typically, the simulation required the system at equilibrium and so requires equilibration after this velocity assignment[9]. The system is most commonly equilibrated twice: first with constant volume (NVT), and then with constant pressure (NPT) conditions. The NPT equilibration follows the NVT simulation to ensure the system is fully equilibrated and ready for a Gibbs free energy calculation. Production runs of whichever type of simulation then follow, e.g., pull simulation using umbrella sampling. The results from the production runs are then analyzed, e.g., by weighted histogram analysis (WHAM). The time step for the MD simulations is usually 1 to 2

femtoseconds[9]. The time steps must be small enough to sample the highest frequency motion in the system, bond vibrations in this case[9]. An adequate sampling time must be selected to accurately observe the system. If, for example, a sampling time is too small, interactions in the system will occur more quickly. These quick interactions can cause the system to "blow up," or fail, because it becomes improbable to occur in nature. Position restraints are frequently used throughout the process to prevent the protein from leaving the box, but at an entropy cost due to the position restraints would not normally exist outside the computational system. Therefore, we must account for the amount of position restraints used in each simulation is usually minimized.

The umbrella sampling method is an MD method that can be used to analyze the $\Delta G_{binding}$ of a ligand to a protein. A system containing GGBP (the protein in this study), water, and neutralizing ions is built in a box that is long enough for the glucose to remain in it when it is pulled during the simulations. Next, the energy in the system is minimized and it is then equilibrated. Finally, a series of umbrella pull simulations are performed to pull the glucose out of the binding site to simulate the reverse process of GGBP binding to glucose. The actual umbrella sampling consists of four major steps: 1) Generate a series of configurations along a reaction coordinate, 2) Extract frames from the trajectory in step 1 that correspond to the desired spacing from ligand to protein, 3) Run umbrella pull simulations on each configuration to restrain it within a window, 4) Use the Weighted Histogram Analysis Method (WHAM) to extract the potential of mean force (PMF) curve and calculate the $\Delta G_{binding}$[10]. The "reaction coordinate" is the distance from the binding pocket of the protein. The "configurations" correspond to specific starting and ending distances from the binding pocket. The "frames" extracted from the configurations are used to initiate simulations which produce trajectories that are analyzed by WHAM. The WHAM analysis of the completed pull trajectories produces a series of histograms and a PMF curve[11]. The histograms help determine the adequacy of the sampling by showing the count of samples against the distance of glucose from GGBP (nanometers). There needs to be sufficient overlapping for the PMF calculation to work accurately. The PMF curve shows the energy (kcal/mol) against the corresponding distance of glucose from GGBP (nanometers). The PMF curve should plateau to a maximum line with a small slope. An average and standard deviation are taken of this line to yield a maximum free energy for the pull and an uncertainty, respectively. The minimum free energy of pull value, which occurs early along the reaction coordinate, is subtracted from this averaged maximum to give the free energy of pull. The free energy of pull is the same magnitude, but opposite sign to the free energy of binding. (The uncertainty is propagated from the standard deviation of the maximum plateau line and from the uncertainty of the minimum pull free energy value.)

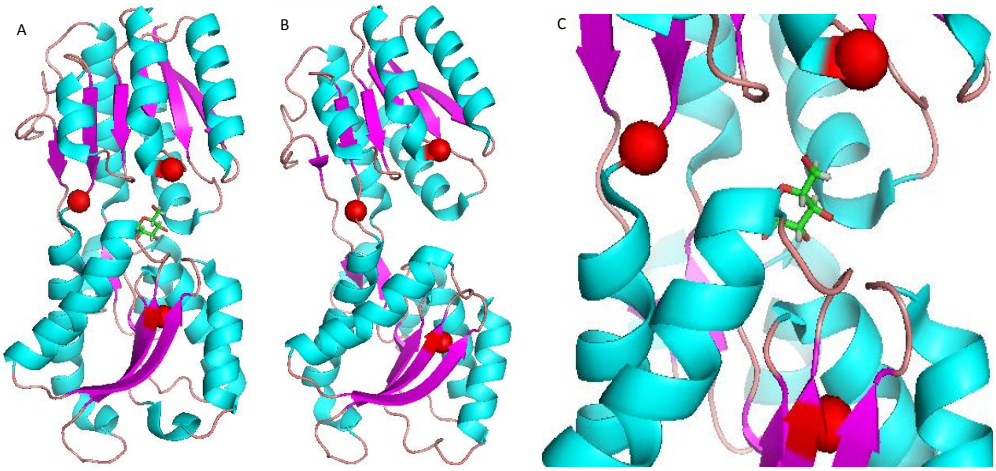

Figure 1. (A) shows GGBP bound to glucose with the C-domain located at the top and the N-domain at the bottom.

The thre hinge angle residues are highlighted as red spheres with Asp-69 at the top and Leu-146 at the bottom. 110-Thr is located near the binding site. This image, rendered by PyMOL from RCSB file 2fvy, shows GGBP at rest before it is built into a system in GROMACS[1]. (B) shows GGBP without glucose in the same orientation as (A). The relative hinge angle between Asp-69, Thr-110, and Leu-146 is large than that of GGBP with bound glucose (A). This is the open conformation that occurs when GGBP binds to glucose[5]. (C) is a zoomed image of the binding site with bound glucose.

Hinge angle analysis was performed to ensure that GGBP follows experimental descriptions when binding to glucose. Since umbrella sampling is the opposite to ligand binding, GGBP should open when glucose is pulled out of the binding site[5]. However, an unbounded, open conformation of GGBP has not been described besides the binding process[5]. A hinge angle analysis shows the behavior of GGBP with and without a bound glucose, which then can be compared to these experimental descriptions to check the reasonableness of the umbrella sampling method. The "hinge angle" of GGBP is more of a relative assessment, although, Borrock has analyzed specific residue segments[1]. The "hinge angle" selected for analysis in this work consisted of three alpha-carbons, from residues Asp-69, Thr-110, and Leu-146[12]. These residues included one residue in the C-domain, one residue in the N-domain, and one residue near the binding site *(Figure 1)*. When comparing the hinge angle of GGBP in the RCSB files with bound glucose (2fvy) and no ligand (2fw0), a difference of 23.5° for the hinge angle exists[1]. For umbrella sampling to be an accurate method of measuring the binding energy of glucose to GGBP, the hinge angle difference for when glucose is bound and when GGBP opens for glucose to be pulled out of the binding site should agree with this 23.5° angle change.

2. Methods:

The glucose was pulled off of GGBP using an umbrella sampling method to calculate the free energy of binding of GGBP to glucose. The PDB file (2FVY) for GGBP was downloaded from the RCSB Data Bank[1]. First, glucose, given a ligand label of BGC (Beta GluCose), coordinates had to be extracted from the GGBP PDB file in order to build the ligand into the system in a GROMACS-acceptable manner as a glucose bound to the protein GGBP. This extraction was done by using the Linux grep function, producing a BGC PDB file. This PDB file was then uploaded to the PRODRG server to generate the coordinate and topology files for the ligand[13]. Next, atom lines for BGC and any other unwanted atoms, such as Ca^{2+}, waters, etc., were removed from the GGBP PDB file. This yielded protein-only GGBP molecule files so that the BGC ligand could be pulled out of the binding site. The system needed to be neutral, but GGBP naturally has a -8 charge at neutral pH. Therefore, GROMACS GMX programs were used to build the box for the system, solvate the box with water, and neutralize the system with Na^+ cations and Cl^- anions. No position restraints were generated for BGC because it was a ligand that would be pulled out of the binding site of GGBP by the umbrella sampling simulations. However, an umbrella potential was grown around the BGC molecule with a restrain penalty that will be evaluated in the future. This potential pulls the ligand from the binding site. Positional restraints lower the entropy of the system, causing an increase in the free energy. Studies were previously done by McCoy (2015), Nguyen (2013), and Rigel (2016) that restrained the GGBP molecule in two ways: all atoms fully restrained or only one atom restrained[12,14,15]. However, it was hypothesized that a system with an unrestrained GGBP molecule would produce a free energy of binding with the least amount of free energy penalties due to restraints.

After solvation, the system underwent two processes: energy minimization (EM) and NPT MD equilibration (Berendsen pressure coupling) in order to have an equilibrated starting point for the rest of the investigation. The exact parameters for these processes can be found in Appendix II. To carry out the pull of the glucose off of the GGBP, a series of configurations along a reaction coordinate were generated using Lemkul's method[10]. Over 500 different distances were then extracted from the initial pull data, but only pull configurations in multiples of 10 were simulated. The resources required to simulate all 500 configurations outweigh the need for the data. Each pull window had to be equilibrated again using Parrinello-Rahman pressure coupling (NPT). The Parrinello-Rahman method was used because it distributes the volumes more correctly than the Berendsen method[16]. However, the Parrinello-Rahman method requires the Berendsen method to be used initially in order for the correct average volume to be present in the system[16]. Umbrella sampling simulations were run on each of the selected configurations. Bash scripting was used to simplify the queuing process on the Wittenberg cluster. Finally, the GROMACS WHAM analysis function was used to generate a PMF curve (with error bars calculate by Bootstrapping) and histograms[11]. The PMF curves are used to calculate the free energy change due to the pulling of the glucose ligand away from GGBP (ΔG_{pull}). These curves typically plateau to a maximum with long protein-sugar separation distances. An average of the plateau energy values is calculated for the final energy. Because umbrella sampling runs simulated the reverse of the binding process, the free energy of binding is obtained from the negative of the pull free energy ($\Delta G_{Binding}=-\Delta G_{Pull}$). The histograms generated by WHAM were used to evaluate the spacing between each pull window. This was to ensure that there is adequate overlap of histograms between adjacent windows. If there was not adequate overlap, then the parameters of the pull simulations had to be modified.

A hinge angle analysis was performed with the GANGLE program in GROMACS and PyMOL's get_angle function. The trajectory files from the individual umbrella sampling runs were merged into one long trajectory with the trjcat function in GROMACS. Then, the GANGLE function produced a probability distribution of the hinge angle

throughout the entire pull. The most probable angle is the average angle of GGBP, while the maximum angle is the open conformation of GGBP. Angle analysis was also done on individual pull windows and plotted against distance. This gives the "breathing" angle of GGBP with and without glucose. This "breathing" angle has been described to be about 9° experimentally[5]. An Angle vs. Distance plot is useful when analyzing the interactions in the binding site when glucose first begins to move. PyMOL's get_angle function was used on trajectory files to give more precise measurements than GANGLE when needed and when measuring the exact angle of the GGBP open conformation. The difference between the average hinge angle and open conformation angle gives the change in conformation angle, which was compared to the 23.5° difference from the experimental files (2fvy and 2fw0)[1].

3. Results:

This study was a continuation of previous work which had issues with positional restraint penalties and protein movement. A position restraint on a single atom (alpha-carbon of Leu-303) gave a $\Delta G_{binding}$ of -11.5(±1.0) kcal/mol[12]. This is over 2 kcal/mol greater than the experimental value −9.1 kcal/mol[6]. An energy penalty also arose from the protein moving freely in the box. The protein would drift out of one side of the box and be mirrored on the opposite side. The $\Delta G_{binding}$ from a system with these out-of-box movements was –9.9(±0.4) kcal/mol[15]. The potential mean of force curve (PMF) had a maximum plateau with a slope of 0.127 *(Figure 2).* The maximum plateau should come a horizontal line. Therefore, changes needed to be made to avoid restraint and movement penalties. This study used a wider box (10.56nm) than previous unrestrained work (8.56nm) and positioned the protein about 1.5nm lower in the box.

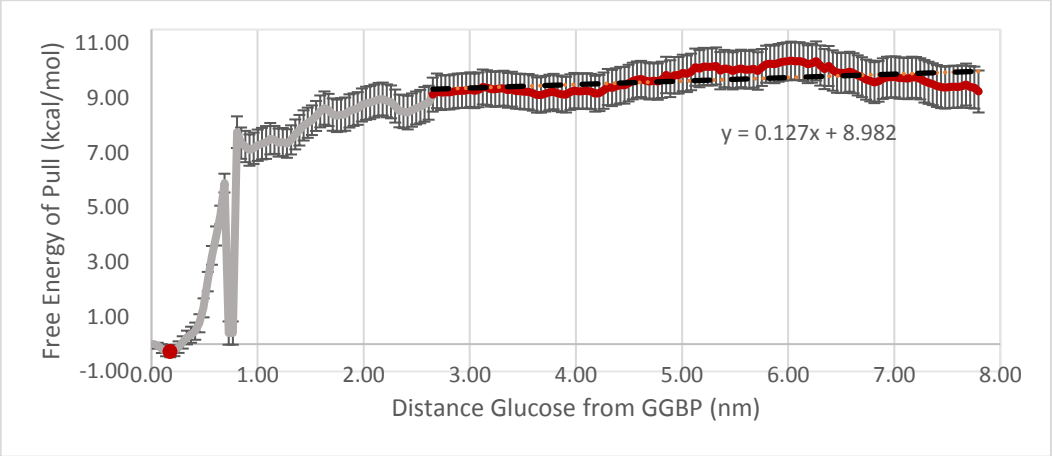

Figure 2. This PMF curve is of wild GGBP umbrella sampling without an applied restraint.

The plateau region (red) is from2.65nm to 7.84nm and the minimum point (red) is at 0.177nm. There is a hole in the plot at 0.75nm from inadequate sampling. The plateau has a larger slope than that of Figure 3, meaning that the system did not maximize as well. The calculated $\Delta G_{binding}$ is −9.9±0.4 kcal/mol. Error bars were generated from bootstrapping in GROMACS.

The elimination of protein restraints and out-of-box movements during umbrella sampling gave a measured $\Delta G_{binding}$ of −10.2(±0.9) kcal/mol. The PMF curve *(Figure 3)* has a slope of 0.0097, meaning the system does come to an energy maximum. The minimum energy point is lower than previous studies and more sampling around 0.3nm needs to be simulated to give a more precise and accurate energy minimum. Histogram analysis of the umbrella sampling *(Figure 4)* of this system showed adequate sampling between pull windows.

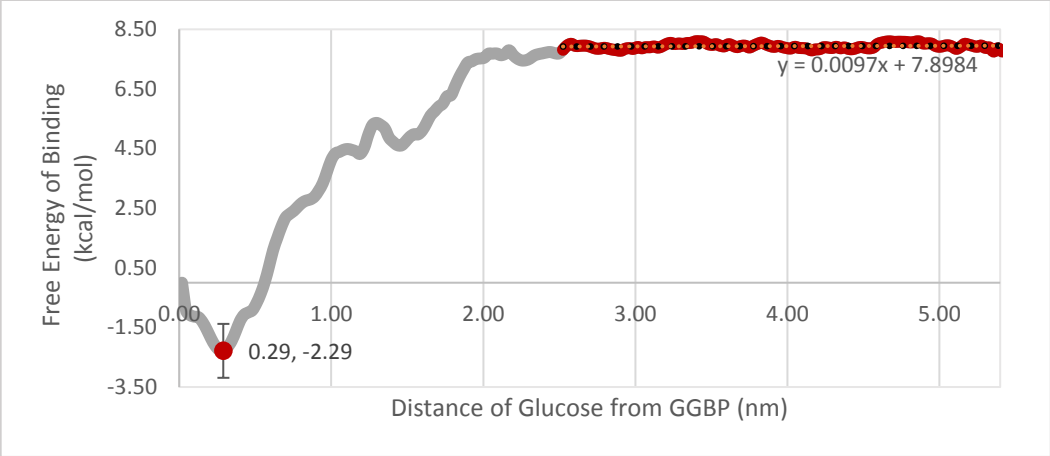

Figure 3. PMF curve for wild GGBP from this study with no applied protein positional restraints, nor any out-of-box movement occurrence.

The minimum energy point occurs at 0.29nm with a value of –2.3 kcal/mol and the average maximum energy is 7.9 kcal/mol. The $\Delta G_{binding}$ for this system is −10.2(±0.9) kcal/mol, where the high uncertainty originates from the uncertainty of the minimum energy point given by GROMACS from bootstrapping during WHAM analysis. This free energy of binding value does not agree with the accepted value of −9.1 kcal/mol[6].

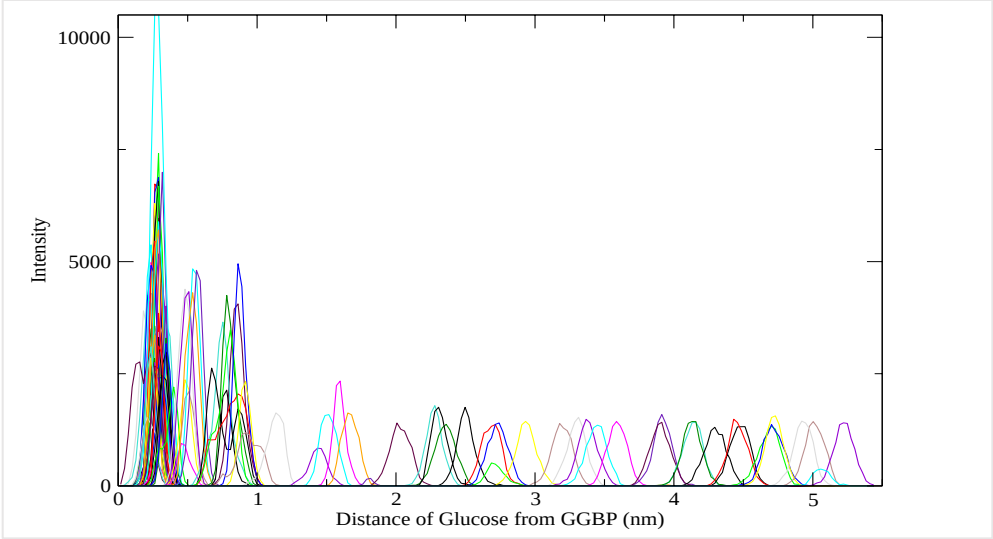

Figure 4. Histogram analysis plot from WHAM function in GROMACS.

The plot shows adequate sampling throughout the pull. The high intensity from 0 to 0.5nm is from intentional oversampling simulated to increase the accuracy of the minimum region on the PMF curve. If multiple pull windows overlap over a small distance, the intensity increases.

Hinge angle analysis of umbrella sampling of this system gave a change in conformation angle of 22.0°, relative to the angle segment Asp-69, Thr-110, and Leu-146. The probability distribution from the GANGLE function in GROMACS shows an average angle of 69.5° and a maximum angle of 91.5° *(Figure 5)*. After further assessment with PyMOL, the exact open conformation angle was measured as 91.1°. This open conformation occurred around 0.88nm into the pull, shown in the Angle vs. Distance plot *(Figure 6)*. A minimum angle of 57.5° occurred at 0.38nm, after the minimum energy point (0.29nm). This may be where glucose begins to significantly move in the binding site, causing the binding pocket to collapse in on itself until water molecules can enter and displace the glucose. However,

trajectories of the binding site are difficult to visualize, and the work has not been completed at this time. A breathing angle of 11° was measured from umbrella sampling by measuring the angle range of the plateau region where there is no bound glucose. The hinge angle analysis showed that GGBP behaves similarly computationally as it has been described experimentally. The angle difference between its two conformations, as observed in the two crystallographic structures *(Figure 1)*, is 23.5° and the breathing angle has been described as 9° experimentally[1,5].

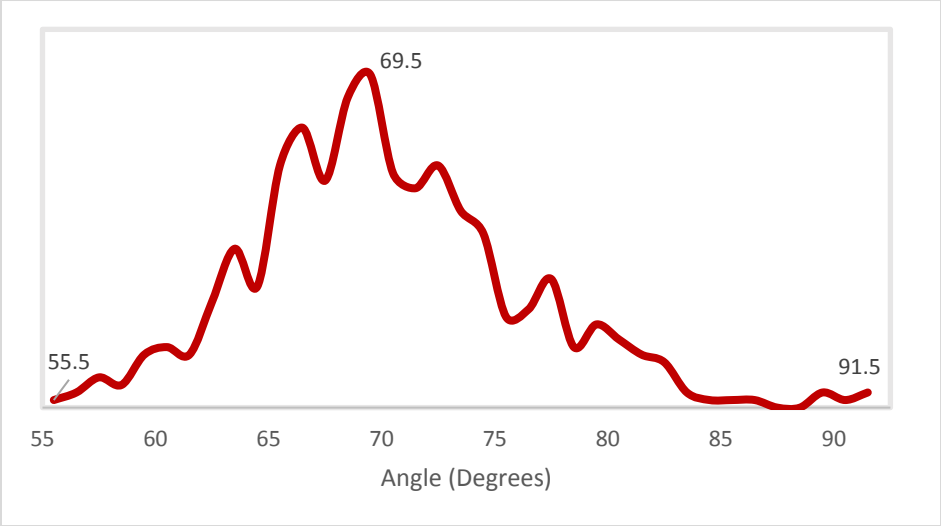

Figure 5. Probability distribution of angles throughout umbrella sampling pull.

The maximum angle from this GANGLE GROMACS analysis is 91.5°, while the average angle is 69°. The minimum angle, 55.5°, occurs around 0.4nm.

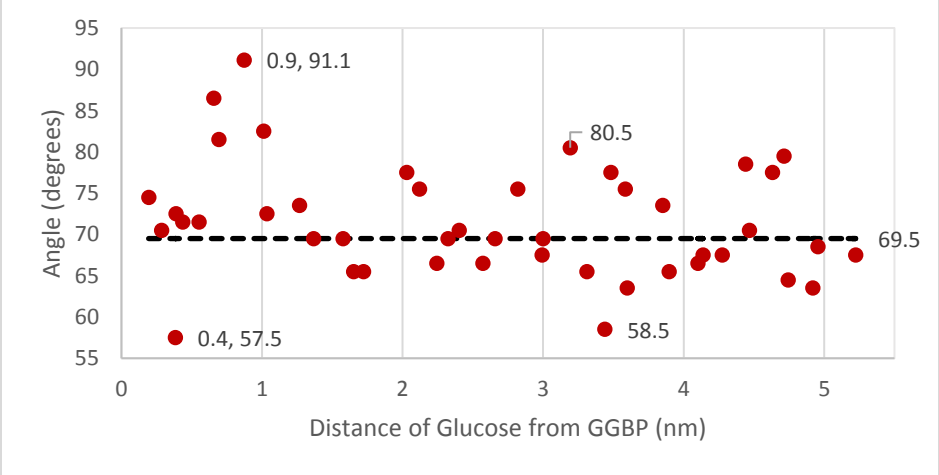

Figure 6. Angle vs. Distance plot for umbrella sampling pull.

he average angle, 69.5°, is highlighted by a dashed line. A breathing angle of 11° is observed from both xpansion and reduction of the binding pocket after glucose is pulled out. A maximum angle of 91.1° occurs at .9nm and a minimum angle occurs at 0.4nm, which needs to be further investigated.

4. Conclusion:

Umbrella sampling without any positional restraints applied to GGBP gave a free energy of binding of −10.2(±0.9) kcal/mol. Although this is not in complete agreement with the accepted value of −9.1 kcal/mol, it is still closer than any previous measurement from this group[6]. A hinge angle analysis was performed to assess how the protein conformation changes during pull simulations. A closed angle of 68.5° was measured from the RCSB high-resolution file generated from experimental work[1]. This same file was used in umbrella sampling and had an average angle of 69.5° during sampling. An angle of 92.0° was measured from an open conformation file (without glucose) from the same experimental work[1]. A maximum angle of 91.1° was measured from umbrella sampling when glucose was pull completely out of the binding pocket. This shows that the protein changed conformation and supports that an umbrella sampling pull simulation is the opposite to protein-ligand binding. Finally, GGBP returned to the closed conformation and a breathing angle of 11° was observed after glucose was pulled completely out of GGBP. GGBP has been described experimentally to exist only in closed conformation (unless binding to a sugar) and to have a breathing angle of 9° with no bound ligand[5].

5. Citations:

1. Borrock, M. J., Kiessling, L. L., and Forest, K. T. "Conformational changes of glucose/galactose binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures." *Protein Science: A Publication of the Protein Society, 16.6 (2007): 1032-1041.*
2. Vyas, N. K., Vyas, M. N., and Quiocho, F. A. "The 3 Å resolution structure of a D-galactose-binding protein for trasnport and chemotaxis in *E. coli*." *Biochemistry 80: 1792-1796. (1983)*; "A novel calcium-binding site in the galactose-binding protein of bacterial trasnport and chemotaxis." *Nature 327: 635-638. (1987)*
3. Mowbray, S. L., Smith, R. D., and Cole, L. B. "Structure of the periplasmic glucose/galactose receptor of *Salmonella typhimurium*." Receptor 1: 41-53. (1990)
4. Khan, F. Saxl, T. E., and Pickup, J. C. "Fluorescence intensity-and lifetime-based glucose sensing using an engineered high-K d mutant of glucose/galactose-binding protein." *Analytical biochemistry, 399(1), 39-43. (2010)*
5. Flocco, M. M., and Mowbray, S. L. "The 1.9 Å X-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*." *J. Biol. Chem. 269: 8931-8936. (1994)*
6. Zukin, R. S., P. G. Strange, L. R. Heavy, and D. E. Koshland Jr. "Properties of the galactose binding protein of *Salmonella typhimurium* and *Escherichia coli*." *16, 3, 381, 386, Biochemistry. (1977)*
7. Allen, Michael P. "Introduction to Molecular Dynamics Simulation." Computational Soft Matter: From Synthetic Polymers to Proteins, Lecture Notes. *(2004)*
8. Van Gunsteren, W. F. and H. J. C. Berendsen. *Groningen Molecular Simulation (GROMOS) Library Manual*, BIOMOS b.v., Groningen. *(1987)*
9. Anderson, Amil. Personal Communication and Guidance. *(2017)*
10. Lemkul, Justin A. "GROMACS Tutorials" *GROMACS Tutorials.* Department of Pharmaceutical Science University of Maryland, Baltimore. *(2015)*
11. Hub, J. S., B. L. de Groot and D. van der Spoel g_wham - A free weighted histogram analysis implementation including robust error and autocorrelation estimates *J. Chem. Theory Comput. 6 (2010). pp. 3713-3720*
12. McCoy, Christopher L. "Measuring the Free Energy Change of the Glucose/Galactose Binding Protein using Molecular Dynamics and Umbrella Sampling with Limited Position Restraints." Wittenberg University. Summer Research Report. *(2015)*
13. Schüettelkopf, A. W., and Van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D: Biological Crystallography, 60*(8), 1355-1363. *(2004)*
14. Nguyen, C., Anderson, A. "Free Energy Change for Glucose binding to the Glucose/Galactose Binding Protein Using Molecular Dynamics Simulations." Wittenberg University. Summer Research Report. *(2013)*
15. Rigel, Nicholas R. and Anderson, A. G. "Measuring the Free Energy of Binding of Wild GGBP and Mutated Trp-213 GGBP by Umbrella Sampling." Wittenberg University. Research Report. *(2016)*
16. Shirts, Michael R. Department of Chemical Engineering, UVA, Charlottesville, VA. *J. Chem. Theory Comput., 9*(2), 909-926. *(2013)*