Understanding the Pathogenesis of Acid-Resistant *E. coli:* Computational Modeling of pH-Dependent Conformational Changes in GadB

Hannah S. Kenagy

Half Hollow Hills High School East, Dix Hills, NY

Section I:

I come from an agricultural family: one side of my family owns a 200-acre farm in Oregon and the other owns a plant nursery adjacent to my backyard in New York. Between living right next to the nursery and spending two or three weeks a year on the Oregon farm, I have been exposed to agricultural and horticultural issues my entire life. When report after report of acid-resistant *E. coli* outbreaks hit the news over the past few years, I became quite interested in the issue as a result of my agricultural background (and my "foodie" interests). I was inspired to read more about this problem and the many other problems of today's increasingly industrialized food system.

In the summer of 2010, I had the opportunity to work in the computational structural biology lab of Dr. Carlos Simmerling at Stony Brook University. That first summer, I did not do any original research. First, I learned how to use UNIX, and then I completed various tutorials. Through the tutorials, I learned how to use many components of the Amber Molecular Dynamics Simulation Program, as well as how to use a number of molecular visualization programs. In addition to completing tutorials, I also did some data analysis for a graduate student to gain additional experience.

Prior to this work, I had no experience with computer programming. It was quite a challenge, therefore, to learn to use UNIX and the simulation programs. In addition to just learning how to write input codes for the simulations and manipulate the output data, I also used my time during the first summer to learn some of the physics and math behind the workings of the simulation programs. The math and physics I had learned in high school up to that point felt quite elementary as I tried to understand the linear algebra and quantum physics used in calculations.

After working through many tutorials, during my second summer in the lab my mentor allowed me to take on a project of my own choosing and design. It was quite a daunting task to pick out a system that I wanted to study, but eventually I picked a protein system that was right in line with my agricultural interests. I chose to use the computational techniques I had learned to study a protein involved in the acid-resistance of *E. coli*, one of the biggest threats to the safety of our food system today.

The work was quite frustrating, but, in the end, very rewarding. The world of computational structural biology is extraordinarily interdisciplinary which is, in my opinion, what makes it so exciting. There is, as indicated by the name, much biology involved, particularly the structural biology of proteins and nucleic acids. And since the molecular dynamics simulations I was working with involved the examination of biological molecules on the molecular level, there was, of course, much chemistry involved as well. And, as I mentioned before, computer science, mathematics, and physics also play key roles in running simulations. As I worked on my project, I found it incredibly exciting not only to see the practical applications of all the scientific theory I had learned, but also to see all of these scientific and mathematical disciplines so integrally play key roles in my project. But the most exciting part was that the findings I came up with were a result of a project of my own design and execution.

To any high school student who is thinking about undertaking a research project, go for it! The best advice I can give to all of you aspiring high school researchers is to be patient and persevere. When I finally had all of my simulation input files prepared for my project, I went to run them on the computer cluster in the lab I was working in. I ran a few benchmarks, and realized that the system I was working with was so large that my simulations would take nearly ten years to run on this cluster. I approached my mentor with this problem, and he suggested I run the simulations on a supercomputer with which the lab had an account. However, after running the benchmarks on this computer, I discovered that my simulations would take about ten months to run. At this point, with the deadline for the Intel Science Talent Search only weeks away, I didn't know what to do. It seemed at times that perhaps all the work I had put in thus far was for naught. Once again, I spoke with my mentor. Luckily, he was able to get me sufficient computer time on a brand new GPU computer, and my simulations were completed in under ten days. Without patience and perseverance, I probably would not have been able to complete my project.

It can certainly feel incredibly overwhelming and extraordinarily frustrating at times (some of the most frustrating experiences of my life were a result of this project!), but it is extremely gratifying in the end. In addition to the personal satisfaction I felt of having successfully persevered through a project of my own design to its completion, by the end I was also amazed at how much I had learned. Looking back, I think I learned more over the course of the two summers that I worked on my project than I did during all of my classes during high school. Not only was it satisfying to have learned all of that, but it was immensely rewarding to have done it almost entirely independently.

Section II:

Relevance of Project

The news media over the last few years has been scattered with stories of microbe resistance, whether it be MRSA, other antibiotic resistant microbes, or acid-resistant *E. coli*. Because of the threat such resistant microbes pose, they have become a substantial health concern. As stated by the CDC, "Over its 15-year history, *E. coli* O157:H7 has evolved as a major problem" [1]. The source of toxic strains of *E. coli*, such as O157:H7, has been pinpointed to concentrated animal feeding operations (CAFOs) that raise the beef that our country consumes [2].

As the cost of corn has decreased significantly over the past few decades due to US government subsidies, many beef producers have begun to feed corn and other cheap grain to their cattle, rather than providing the animals with pastures for grazing. Although this increases their growth rate, feeding grain to cattle results in the occurrence of fermentation in the colon of the bovine intestinal tract, producing acetic, propionic, and butyric acids [3,4]. As such acids accumulate, the pH of the colon decreases, allowing acid-resistant *E. coli* to develop. Such strains are able to survive the acidic conditions of the human stomach, and those such as O157:H7 may then produce toxic chemicals that can cause hemorrhaging or kidney failure. In fact, grain-fed cattle have been shown to have 10^6 times more acid-resistant *E. coli* than cattle whose diet consists primarily of hay [5]. Thus, feeding grain to cattle poses the dangerous health issue of food contamination with acid-resistant *E. coli*, which has the potential to be lethal.

The CAFOs also present another problem: their crowdedness and lack of cleanliness. This allows the toxic *E. coli* present in one cow to spread to nearly every other cow in the

feedlot. By the time the cows reach the slaughterhouse, their hides are caked with manure, both their own and that of other cows present in the feedlot. With slaughterhouses slaughtering 400 cows every hour, inevitably some of the *E. coli* present either in the cows themselves or in the manure they are caked in ends up in the meat. Thus, new acid-resistant toxic strains of *E. coli* end up in the food system, and are easily transported to other farms though factory farm runoff [3].

In order to combat this problem with the existing grain-based bovine diet, more about the mechanism behind the acid-resistance of *E. coli* must be learned. Currently there are three known methods for the acid-resistance of *E. coli*, the most potent of which is the Gad system. In fact, the Gad system has the potential to allow for the survival of *E. coli* for more than two hours in a strongly acidic environment with a pH similar to that of the human stomach [6,7].

The Gad system itself includes three proteins: GadA, GadB, and GadC [8]. All three are required for survival in an acidic environment [9,10,11,12]. It is thought that the Gad system controls the internal pH of the cell by changing glutamate (an acidic substrate) into a neutral compound (GABA) by incorporating hydrogen ions [8]. The GadC protein is thought to then export GABA into the extracellular region, thereby removing protons (hydrogen ions) from the cell and increasing the internal pH of the cell [8]. This process is initiated by a conformational change in the GadB protein that is caused by a decrease in the internal pH of the cell. Currently there is limited knowledge about this conformational change. The more that is known, however, the easier it will be to develop a small molecule that could stabilize the neutral-pH conformation of GadB. By stabilizing the neutral-pH conformation, the pH-driven conformational change would be inhibited, and, as a result, the effectiveness of the Gad system of acid-resistance would be blocked.

Current Structural Knowledge

Recently, the crystal structure of GadB in both acidic (form A in pH 4.6, accessible through PDB code 1pmm) and neutral (form C in pH 7.6, accessible through PDB code 1pmo) environments has been determined, and some comparisons of the two structures has been done.

GadB is considered a homohexamer, as it is made up of six identical chains. It was found that the overall structure of the hexamer does not change, but that significant changes do occur at the N- and C-termini and in the region containing residues 300-313 where a beta-hairpin (a type of

protein secondary structure) is present. In form C (pH 7.6), residues 3-15 do not form any regular secondary structure, but those residues in form A (pH 4.6) are part of alpha helices (a type of protein secondary structure). The beta-hairpin at residues 300-313 in form C is situated such that it narrows the active site funnel. However, in form A, the beta-hairpin is shifted towards the center of the active site funnel. It has also been noted that Lys3-Lys30 form different conformations in forms A and C. Lastly, it has been determined that Asp2, Asp8, Glu12, and Asp15 are all necessary for the conformational change that occurs in the N-terminal region, likely due to changes in protonation states at the varying pH values [8].

In addition to the identification of the aforementioned structural differences, it has also been noted that this pH-driven conformational change involves the uptake of 4-6 protons. Some have proposed that Asp2, Asp8, Glu12, and Asp15 are all likely candidates for this protonation (addition of a proton) [8].

Previous research has also indicated that as the intracellular pH lowers, GadB moves toward the membrane of the cell, which is consistent with the Gad system's method of proton expulsion as described previously. Interestingly, however, this movement towards the membrane does not occur when the N-terminus is deleted from the GadB structure [8].

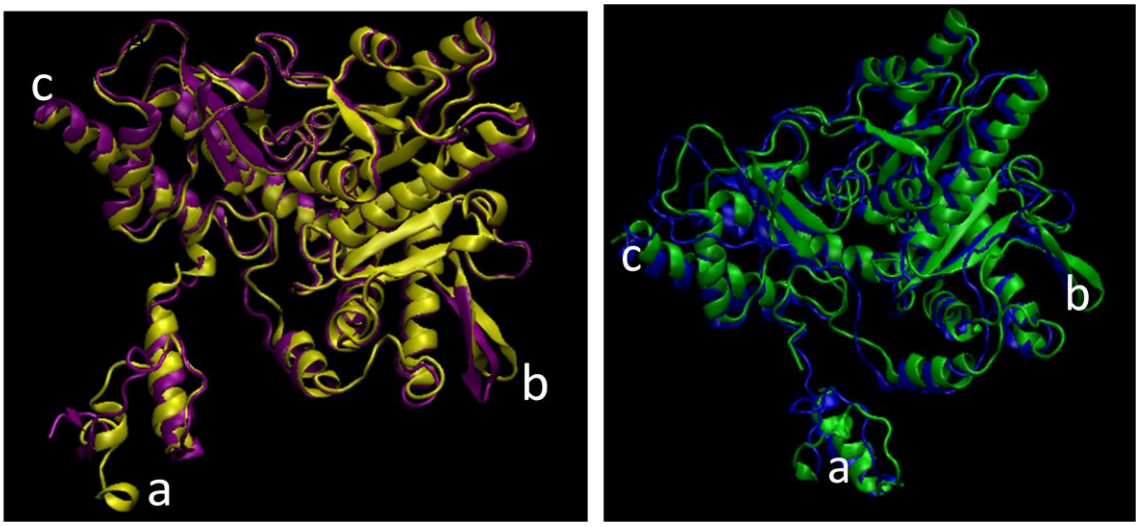
Purpose

The GadB protein undergoes a conformational change when going between acidic and neutral environments, and crystal structures of the protein have been determined for the structure in both pH environments [8]. Specifically, differences in the two structures have been observed at the N-terminus, C-terminus, and active site of the enzyme. It has also been found that this conformational change involves the uptake of 4-6 protons and some residues have been proposed as candidates for this protonation [8]. However, currently it is unknown exactly which residues are protonated, and what the mechanism for the protonation is during this conformational change. Through the use of molecular dynamics simulations, this project explored the effect of altered protonation states of titratable residues (residues which have two protonation states: protonated and deprotonated) in GadB on the conformational changes that accompany a shift in pH. Increased knowledge of such effects could help society develop ways to block the effectiveness of the Gad system, and thereby eliminate the acid-resistant nature of certain *E. coli* strains.

Methods and Results

First, the pKa (pH value at which the functional group is half protonated and half deprotonated) of each ionizable residue in GadB was calculated at a pH of 4.6 (using the PDB structure 1pmm) and a pH of 7.6 (using the PDB structure 1pmo) using a web-based software called H++ [13,14]. These pKa values were used to calculate the protonation states of each residue at the given pH. Using the protonation state predictions from H++, four PDB files were created to serve as initial structures for equilibration and then molecular dynamics simulations: one with the low-pH starting structure and the predicted low-pH protonation states (low), one with the low-pH starting structure and the predicted neutral-pH protonation states (low-to-neutral), one with the neutral-pH starting structure and the predicted low-pH protonation states (neutral-to-low), and one with the neutral-pH starting structure and the predicted neutral-pH protonation states (neutral).

With the structures prepared, Amber (a package of molecular simulation programs) was used to run molecular dynamics simulations of each structure [15]. As in all crystal structures, in the crystal structures of GadB hydrogen atoms were not visible. As such, it is impossible to determine the role of altered protonation states on the dynamics of the protein. Below are two pictures of chain A of the GadB protein. The first shows the overlaid crystal structures of the low-pH (yellow) and neutral-pH (purple) crystal structures and the second shows the overlaid average structures of the low-to-neutral-pH (green) and neutral-to-low-pH (blue) simulations. On each, "a" labels the N-terminus, "b" labels the beta-hairpin region at residues 300-313, and "c" labels the C-terminus.



While I cannot publish the main data of this study, with the molecular dynamics simulations that I ran, I was able to look at how the protein's conformation changed as a result of the protonation states that I changed. I looked at potential energy values for the protein during the course of the simulation and calculated various root mean square deviation (RMSD) values (this found the average distance between the backbone atoms of two protein structures). The equations for these two calculations are shown below. I was able to look at how different parts of the protein deviated from their starting structure during the simulation as a result of the altered protonation states of the titratable residues.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{i=N} \delta_i{}^2}$$

Where δ is the distance between *N* pairs of corresponding atoms (backbone)

$$V(\mathbf{r}) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_o)^2 + \sum_{dihedrals} (V_n/2)(1+\cos[n\phi-\delta] + \sum_{nonbij} (A_{ij}/r_{ij}^{12}) - (B_{ij}/r_{ij}^{6}) + (q_iq_j/r_{ij})$$

Accounts for energies due to bonds, angles, dihedrals, and non-bonded atoms

:su

about the conformational change that GadB undergoes as its pH environment changes. The more we learn about the conformational change, the closer we are to the development of a small molecule that could stabilize GadB's neutral conformation. This would inhibit the activation of the Gad system, thereby blocking its effectiveness and eliminating the acid-resistant nature of certain *E. coli* strains.

References

[1] R.E. Fontaine, S. Arnon, W.T. Martin, et al. "Raw hamburger: an interstate common source of human salmonellosis," *The American Journal of Epidemiology*, vol. 107, pp. 36-45, 1978.

[2] M. Pollan, *The Omnivore's Dilemma*, New York: Penguin, 2006.

[3] R. Kenner, R. Pearce, E. Schlosser, M. Robledo, and W. Pohlad, *Food, Inc.*, Los Angeles: Magnolia Home Entertainment, 2009.

[4] J.B. Russell, F. Diez-Gonzalez, G.N. Jarvis, "Potential effect of cattle diets on the transmission of pathogenic *Escherichia coli* to humans," *Microbes and Infection*, vol. 2, pp. 45-53, 2000.

[5] F. Diez-Gonzalez, T.R. Callaway, M.G. Kizoulis, J.B. Russell, "Grain Feeding and the Dissemination of Acid-Resistant *Escherichia coli* from Cattle," *Science*, vol. 281, pp. 1666-1668, 1998.

[6] J. Lin, I.S. Lee, J. Frey, J.L. Slonczewski, and J.W. Foster, "Comparative analysis of extreme acid survival in *Salmonella typhimurium, Shigella flexneri* and *Escherichia coli*," *The Journal of. Bacteriology*, vol. 177, pp. 4097-4104, 1995.

[7] J. Lin, M.P. Smith, K.C. Chapin, H.S. Baik, G.N. Bennett, and J.W. Foster, "Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*," *Applied and Environmental Microbiology*, vol. 62, pp. 3094-3100, 1996.

[8] G. Capitani, D. De Baise, C. Aurizi, H. Gut, F. Bossa, and M. Grütter, "Crystal structure and functional analysis of *Escherichia coli* glutamate decarboxylase," *The EMBO Journal*, vol. 22, no. 16, pp. 4027-4037, 2003.

[9] D. De Biase, A. Tramonti, F. Bossa, and P. Visca, "The response to stationary-phase stress conditions in *Escherichia coli*: role and regulation of the glutamic acid decarboxylase system," *Molecular Microbiology*, vol. 32, pp. 1198-1211, 1999.

[10] P.D. Cotter, C.G. Gahan, and C. Hill, "A glutamate decarboxylase system protects Listeria monocytogenes in gastric fluid," *Molecular Microbiology*, vol. 40, pp. 465-475, 2001.

[11] B.M. Hersh, F.T. Farooq, D.N. Barstad, D.L. Blankenhorn, and J.L. Slonszewski, "A glutamate-dependent acid resistance gene in *Escherichia coli*." *Journal of Bacteriology*, vol. 178, pp. 3978-3981, 1996.

[12] S.R. Waterman and P.L. Small, "Identification of sigma S-dependent genes associated with the stationary-phase acid-resistance phenotype of *Shigella flexneri*." *Molecular Microbiology*, vol. 21, pp. 925-940, 1996.

[13] J.C. Gordon, J.B. Myers, T. Folta, V. Shoja, L.S. Heath and A. Onufriev, "H++: a server for estimating pKas and adding missing hydrogens to macromolecules," *Nucleic Acids Research*, vol. 33, pp. W368-71, 2005.

[14] Myers J, Grothaus G, Narayanan S, Onufriev A, "A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules, " *Proteins*, vol. 63, pp. 928-938, 2006.

[15] D. A. Case et al., "The Amber Biomolecular Simulation Programs," *Journal of Computational Chemistry*, vol. 26, no. 15, pp. 1668-1688, December 2005.