Anaplasma Phagocytophilum TrmD G-37 Research Plan

Authors: Damien Edele, Christopher Jannotta, Danielle Levanti

Project Category: Cellular and Molecular Biology

Rationale:

- *Anaplasma phagocytophilum* is a pathogenic species of bacteria that causes Human Granulocytic Anaplasmosis (HGA), the second most common tick-borne disease in the northeastern United States, along with many other pathogens that use the same vector (one being *Borrelia burgdorferi*, the causative agent of Lyme disease). Even in first world countries where preventative measures, modern science, and health education are accessible to most, HGA still spreads and has the potential to inflict severe and even lethal symptoms in those who contract it. The pathological mechanism behind HGA includes obligate intracellular parasitism in neutrophils. Since neutrophils produce a wide array of substances/exotoxins to destroy bacteria or produce NETs (Neutrophil Extracellular Traps), the *Anaplasma phagocytophilum* bacteria infiltrate the neutrophil instead, where the toxins are produced but not released. Strategies such as this have been used by bacteria for a considerable time on the evolutionary timescale, and thus humans evolved to counter this by having neutrophils undergo rapid apoptosis once reaching a mature stage, even if bacterial infection is present (Dendritic cells may prolong their lives in certain cases). However, pathogens such as *A. phagocytophilum* have countered this technique by modifying neutrophil genomes to suppress apoptosis-inducing genes such as the tumor necrosis factor (TNF) superfamily and upregulate others such as the CXCR4, which promotes neutrophil mobility. Not only are the pathological mechanisms of HGA advanced and evolving to counter the mammalian immune system, but anthropogenic activities such as climate change increase the population and activity of deer ticks and black-legged ticks, which is directly correlated with the increase in HGA cases seen. On Long Island, it has been documented that deer tick populations have exponentially increased over the past few years, and HGA cases have as well.

- With an understanding of the pathology of Human Granulocytic Anaplasmosis, a rationale behind the research and focuses of the project/experiment can be explained in an efficient and sensible manner. The process that *A. phagocytophilum* uses to induce expressional changes in neutrophil regulating proteins requires the bacterium itself to produce a wide variety of proteins and enzymes to allow these changes to happen. In addition to this, the bacterium needs to produce proteins that can interact with the endosome it is encapsulated in so it does not get digested by the neutrophil. The high-reliance on protein synthesis during the intercellular to intracellular transition period makes ribosomal function imperative to bacterial survival and continuing infection in the host. Thus this experiment focuses on interpreting raw x-ray diffraction data from a National Synchrotron Light Source to provide a model of a TrmD enzyme majorly involved in the bacterial protein synthesis cycle: tRNA (guanine-N1)-methyltransferase. Previous researchers have noted that this enzyme methylates a tRNA anticodon nucleoside, guanine 37, to stabilize the anticodon in such a way that hydrogen bond formation between mRNA and the tRNA anticodon becomes energetically unfavorable and kinetically improbable. If hydrogen bond formation was to be spontaneous, base pairing between tRNA and mRNA would occur and cause some nucleotides in the mRNA to fail the translation procedure, leading to a frameshift mutation that would change the entire sequence of the synthesized protein after that base pair is exempted from translation. Given that HGA relies heavily on protein production, the bacteriostatic properties an inhibitor to the TrmD protein would induce are expected to cause heavy decimation of HGA's mechanism of disease.

- The discovery of the TrmD class of proteins has stimulated novel and relentless research by many research groups because of the understanding that almost all eubacteria are observed to use it while other organisms rely on a different family of proteins for the same processes. However, in many of the published models of these proteins, a linker region, halo region, and unliganded form of the protein have not been observed. These regions are elusive and likely hold valuable details about the protein. Thus since no unliganded form of the *A. phagocytophilum* TrmD that contains a fully continuous model (no gaps or missing sequences in the chain from N to C terminus) has been published yet, the rationale behind this research project is to do so and provide an analysis on what the model can tell us about the bacterial protein synthesis tRNA-(guanine-N1)-methyltransferase and possible methods to inhibition for it. These findings would not only be valuable for the treatment of HGA, but also have the potential to be incorporated into virtually any TrmD carrying bacteria because of their active site conservation and extreme importance to bacterial protein synthesis.

Goals/Expected Outcomes:

- Given the severity of Anaplasmosis as an emerging disease and the lack of information regarding the function of its TrmD proteins' less-observed regions, such as the linker and halo regions, this project's main focus is to publish an accurate 3D visualization of *Anaplasma phagocytophilum's* tRNA (guanine-N1)-methyltransferase that includes the modeling/analysis of the less-observed regions and their interactions with the rest of the enzyme when the enzyme is in not liganded with its natural substrates or similar analogs to stabilize it and consequently change its structure and electron movement network. By making a scientifically sound visualization of the elusive regions of the enzyme, it is expected to uncover the induced changes to the molecule structure when a ligand is bound to it. Based on the differences seen in comparing this model with other models that do contain a substrate in the active site, we hope it provides insights into the significance of these regions for the enzyme's structural properties during the different phases of its activity timeline.

- After the successful modeling and publishing of the protein into the Worldwide Protein Data Bank, we plan to write and publish a research paper discussing the results of the experiment and the future potential that comes out of the findings. Figures that present interesting findings and details about the enzyme will be incorporated into the paper and future research potential regarding what the findings have shown. With this analysis of the protein and its intricacies, it is expected that this research project will make clear how the protein functions and consequently the possible mechanisms of inhibition that can be exploited to make new medicines that target not only Anaplasma species, but a majority of pathogenic bacterial species due to the conserved nature of certain regions of the protein and its integral use as a mutagenic dampener.

Procedure:

-Using X-ray diffraction data produced by using a National Synchrotron Light Source to barrage regions of *Anaplasma phagocytophilum* derived tRNA-(guanine-N1)-methyltransferase protein crystals, the electron density map for a monomer of the enzyme can be calculated, simulated, and used as a reference to create a computerized, visual 3D molecular structure of it. The methods of analyzing the data, including techniques, programs, and other resources, are explained in detail below.

Risk and Safety:

- Understanding the risks and ways to stay safe during any scientific experiment is always a top priority for our research group. Although we have used a Synchrotron Light Source, carried out experimental procedures pertaining to our area of research in government laboratories, and completed extensive online and hands-on training in order to engage in such research and experimentation, this particular project was a multistage, long-term project and initiated in 2018. Thus the procedures regarding expression, crystallization, and x-ray diffraction of the TrmD protein being analyzed have already been completed by previous student researchers and risk assessment/procedures for this portion of experimentation will not be reviewed since we have not technically worked on the initial processes of data collection for this specific protein. Nevertheless, we have gained experience using the National Synchrotron Light Source for other ongoing projects and have followed detailed regulations and procedures to ensure safety and calculate risk from exposure to radiation and hazardous materials. As for the risks and safety regarding our portion of the research and data analysis, sleep-deprivation and eye strain may be associated with the long and specific procedure and manual manipulation of the model needed to engineer an accurate and detailed model of the TrmD enzyme.

Data Analysis and Solutions to Potential Obstacles:

1. The initial phase of data analysis is to be done using an automated program to collect data hot off the National Synchrotron Light Source beamline. "HKL-2000"[1] should be used to undergo this task, collecting important information regarding the diffraction patterns and protein crystal being used to diffract the x-rays off. HKL-2000 automatically corrects physical problems that cause corruption of the data such as "correction for absorption, spindle-axis misalignment, uneven speed of spindle-axis rotation and vibration of the cryogenic loop with the frozen crystal"[2]. Few x-rays do not diffract and instead are absorbed by the crystal, but the ones that do may interact with the nucleus of atoms, cause auger electron emission and thus uncharacteristic x-ray diffraction, and thus can skew the diffraction data. Luckily, by using gaussian integration and the properties of the crystal, HKL-2000 can calculate the necessary correction for this problem. The spindle axis refers to the x,y,z - axes in which x-ray beams are angled at the crystal and the crystal is rotated with respect to. Small misalignments of these axes can skew the Fourier transform result that needs to be calculated later, and thus recognizing and correcting this is carried out by HKL-2000. Lastly, thermo-acoustic oscillations occur when temperature changes cause movement of heat and pressure to make air move back and forth, deteriorating cryogen and adding slight vibrations to the system. HKL-2000 is able to collect data about the system to account for this obstacle.

2. After HKL-2000 corrects the raw data from extraneous variables in the system and refines the x-ray diffraction patterns, integration, and scaling we can begin to computerize the diffraction pattern images and start the process of creating an electron density map from them. HKL-2000 is able to assign an "intensity"- a number that corresponds to the strength an x-ray reflection shows on the diffraction image. This will be done for every diffraction image in a data set (hundreds of diffraction images). The intensities are satisfied by Bragg's Law, which characterizes constructive interference of diffracted x-rays that originated in different planes of the protein crystal. Full constructive interference is observed when the wavelength of the incident x-rays multiplied by any integer is equal to the distance between the two planes of the crystal where these x-rays were reflected. Mathematically, this is characterized by the formula: $2dsin(\Theta) = \lambda(n)$, *where* n = *integer,* d = *distance between planes,* Θ = *reflection angle.* After assigning intensities to all the x-ray diffraction points, scaling may begin. Scaling is the process of taking all the reflections from every diffraction image collected and merging them to create an average of all the data, leaving one value for each point on an x-ray diffraction image. The software program "SCALA"[3] will accomplish this and filters out partial reflection images that would be outliers to the dataset.

3. Subsequently, after making one unified data set with the x-ray intensities, a program called "Truncate"[4] is necessary for taking the intensities and finding the amplitude of the x-ray diffraction patterns instead. Traditionally, converting is relatively easy and requires square rooting all the positive reflection intensities. However this fails to account for negative values and alternatively, Truncate is able to estimate the best values for amplitude by using the distribution of intensities with respect to each resolution shell. This makes all negative reflections get 'inflated' and become positive and thus included in the amplitude calculations to help better the accuracy of the data.

4. In addition to the intensities of the diffracted x-rays needed to gather positional data for calculating the electron density function of the protein, the space group and other statistics about the protein will also be needed from the diffraction data. "Pointless"[5] is a software program that finds the Patterson/Laue group of the protein crystal using the unmerged diffraction pattern files. This is done by scanning the diffraction patterns of the protein crystal and use its centering, near-zero intensities, and matching symmetry between reflections given the cell dimension restraints. The data will fit into one of the major crystal point group classes and thus the space group and other important information about the crystal is found this way.

5. After the space group and average amplitudes have been calculated using the previous programs, "SFtools"[6] will be used to convert all this raw data into a structured form that can be readable for mathematical uses/calculating the electron density of the protein. The outputted form of the data is called the structure factors and can come in the form of a preliminary MTZ file. However, this file is not yet transformed/integrated to code for the electron density map of the data and instead is capable of being read by molecular replacement software to help finish the later process of forming the electron density map.

6. At this point in the process is it integral to ensure the integrity of the files at work. "SFCheck" [12] is a program dedicated to providing insight to the files previously output by SFTools. The program will give us the ability to view very important information about the refinement process so far. Through this, we are able to see an early rendition of our R-Value, the correlation, Luzzati and Wilson plots, as well as much more analytic information that will allow us to gather further insight into the refinement process so far.

7. After the structure factors (amplitudes of the x-rays and their distribution based on the crystal space group/diffraction patterns) have been calculated, the electron density could be theoretically calculated. However, the data gathered alone only gives the absolute value of the 3-dimensional points at which the diffraction patterns were made because the scanners that pick up the reflected x-rays cannot determine the phases of them. This allows for a variety of potential solutions when using the positions for calculating an electron density map. "Phaser MR"[7] is a molecular replacement program that overcomes the phasing problem by using an input electron density maps of a similar protein as a reference when calculating our enzyme's electron density. The reference structure and electron density that will be used in this data analysis is a liganded TrmD from *Anaplasma sp.* In the wwPDB as ID: 4IG6 [8]). The process of creating the electron density map from the reference and collected data will be done by using inverse Fourier synthesis, which uses the phases of the similar protein electron density map (Fourier transform of the reference's electron density can retrieve the +/- signs) and combines it with the structure factors found using the previous programs to make it possible to calculate and computerize/visualize an electron density map for our enzyme. While Fourier transforms are used to breakdown a signal/function into a series of sine waves, inverse Fourier synthesis uses a sum of components (structure factors) to make an electron density function.

The function to the right represents the electron density p(x) in terms of the structure factors F(h,k,l) being integrated via inverse Fourier synthesis, all with respect to reciprocal volume. Source: [9]

$$\rho(\mathbf{X}) = \int F(h,k,l)e^{2\pi i \mathbf{X}\cdot(h,k,l)}\, dV$$

8. After the electron density map is compiled, a skeleton of the protein structure is to be constructed so that there is a correlation with it in order to start focusing on the unique details of this specific TrmD. "MOLREP"[10] is an automated refinement/replacement program that took a similar protein model as a reference (PDB ID: 4IG6 [8]) and aligns it to fit the experimental electron density map. The amino acid sequence of the protein (which was gathered from previous experimentation with expressing the protein from *E. coli*) will also be input. Together MOLREP automatically changes the conformation, molecular makeup, and other details of the reference protein to make it align much better with the experimental electron density map. This gives us a starting point for our model by making the first part/its skeleton for further refinement.

9. After the preliminary model is built, both automatic refinement program "Refmac5"[11] and manual refinement program "Crystallographic Object-Oriented Toolkit (COOT)"[11] are to be used to detail the model and constraint it to the bounds set by the calculated electron density map and known biochemical principles such as Ramachandran outliers, rotamer probability, and determining other variables in the data such as water molecules and cryoprotectant. Refmac5 uses a variety of probability and physics formulas to determine the best angles/conformation of the protein's components. COOT also displays valuable information regarding favorability of conformations/rotamers, B-factor of different areas, and how well regions for the electron density. Each of the 219 amino acids will need to be carefully reviewed and altered if necessary to complement the electron density map. Since the electron density map is the experimental data, translational modifications are to be done to make sure the protein model agrees with the experimental data. Eventually, refinement using hundreds of Refmac5 cycles and countless hours of COOT to make sure the protein is scientifically sound, the structure is then complete and ready for deposition/publication into the Worldwide Protein Data Bank. Since the data is at a 2.7-angstrom resolution, an R-factor of about 18-20% is expected is refinement is successful. However, RSRZ outliers and clashscore are likely to be lower than in higher resolution protein models because of the absence of information about the smaller details of the e^- density.

10. After the protein model is complete, a final run of "WaterTidy"[11] is to be run in order to help sort out the water molecules bound by hydrogen bonding into more favorable locations and give them better organization within the protein model's file. Afterward, the protein model (now in mmcif format) and the experimental electron density data file (in mtz format) will be sent in for review by the wwPDB deposition team. A diagnostic test is given on the protein's properties in reference to accepted scientific data and the experimental electron density. If no major problems are encountered, the model is accepted to be published in the RCSB PDB research journal.

11. After being reviewed and published, the protein model can be used to observe the properties of the protein confidently and draw conclusions and potential areas of further research on it. Since the previous investigation on this diffraction data has found the potential for a halo region and linker region electron density, if those regions can be accurately modeled in the above steps, it can provide possible answers to multiple questions pertaining to how the TrmD enzyme functions. Since it is rare for TrmD enzymes to be seen in an unliganded state and still have electron density for its linker and halo regions, questions about how the electron movement network works between the halo region, linker region, and the catalytic domain of the protein can be found by comparing liganded TrmDs to the model being proposed in this paper. The differences may outline the major factors in keeping/changing dimerization and conformation in certain ways to allow for the function to be carried out. If variable regions can be found from the differences between the two, it is possible to suggest further study of possible molecules that could inhibit the electron network and thus decrease/increase the instance of conformational change in the protein. Based on which amino acids are changed and how by the binding of a ligand and the effects they have on the linker region and halo region of the protein, analysis and conclusions/future studies can be drawn from this project in hopes of finding a novel method of treating bacterial disease.

Bibliography:

1. Z. Otwinowski and W. Minor, " Processing of X-ray Diffraction Data Collected in Oscillation Mode ", Methods in Enzymology, Volume 276: Macromolecular Crystallography, part A, p.307-326, 1997,C.W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York).

2. Minor, W., Cymborowski, M., Otwinowski, Z., & Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. *Acta Crystallographica Section D Biological Crystallography*, *62*(8), 859–866. doi: 10.1107/s0907444906019949

3. SCALA (CCP4: Supported Program). (n.d.). Retrieved June 23, 2019, from http://www.ccp4.ac.uk/html/scala.html#ref7

4. TRUNCATE (CCP4: Supported Program). (n.d.). Retrieved June 23, 2019, from http://www.ccp4.ac.uk/html/truncate.html

5. POINTLESS (CCP4: Supported Program). (n.d.). Retrieved June 23, 2019, from http://www.ccp4.ac.uk/html/pointless.html

6. SFTOOLS (CCP4: Supported Program). (n.d.). Retrieved July 2, 2020, from http://www.ccp4.ac.uk/html/sftools.html

7. Tutorial: solving a structure with Phaser-MR. (n.d.). Retrieved July 2, 2019, from https://www.phenix-online.org/documentation/tutorials/mr.html

8. Crystal structure of a tRNA (guanine-N1)-methyltransferase from Anaplasma phagocytophilum bound to S-adenosylhomocysteine. (2012). *RCSB PDB*. doi: 10.2210/pdb4ig6/pdb

9. AMS :: Feature Column :: X-ray Crystallography and the Fourier Transform. (n.d.). Retrieved July 2, 2019, from http://www.ams.org/publicoutreach/feature-column/fc-2011-10

10. MOLREP (CCP4: Supported Program). (n.d.). Retrieved July 7, 2019, from http://www.ccp4.ac.uk/html/molrep.html

11. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., … Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D Biological Crystallography*, *67*(4), 355–367. doi: 10.1107/s0907444911001314
12. SFCHECK (CCP4: Supported Program). (n.d.). Retrieved January 24, 2020, from http://www.ccp4.ac.uk/html/sfcheck.html

Research Plan Addendum

No Addendums Exist