Complete Structural Determination and Analysis of tRNA (guanine-N^1)-methyltransferase Enzyme Expressed in *Anaplasma phagocytophilum* with Insights to Inhibition

Damien Edele[1], Christopher Jannotta[1], Danielle Levanti[2]
Eastport South Manor Jr/Sr High School[1], Northport High School[2]

Project Category: Cellular and Molecular Biology

Introduction:

Anaplasma phagocytophilum is a gram-negative bacterium that causes Human Granulocytic Anaplasmosis (HGA), the second most common tick-borne disease in the northeastern United States [1]. It is commonly transmitted by the black-legged tick (*Ixodes scapularis*) along with *Borrelia burgdorferi*, the causative agent of Lyme disease. HGA has the potential to inflict severe and even lethal symptoms to those who are infected because of its pathological mechanisms including obligate intracellular parasitism and tropisms in neutrophils. Since neutrophils produce a wide array of substances and exotoxins to destroy bacteria or produce NETs (Neutrophil Extracellular Traps), the *Anaplasma phagocytophilum* bacteria infiltrate the neutrophil instead, where the bactericidal toxins are produced [2]. This strategy has been employed by bacteria for the vast majority on their evolutionary timescale, and thus humans evolved to counter this strategy by having neutrophils undergo rapid apoptosis once reaching a mature stage, even if bacterial infection is present. However, *A. phagocytophilum* has countered this adaptation by modifying neutrophil genomes/expression to suppress apoptosis-inducing genes such as the Tumor Necrosis Factor (TNF) superfamily and upregulate others such as the CXCR4, which promotes neutrophil mobility[3]. Not only are the pathological mechanisms of HGA advanced and evolving to counter the mammalian immune system, but anthropogenic occurrences such as climate change, population increase, and tick activity, also directly correlate with the increase in HGA cases. On Long Island, it has been documented that deer tick populations have exponentially increased over the past few years, and tick-borne disease has reached epidemic proportions[4].

With an understanding of the pathology of Human Granulocytic Anaplasmosis, the process *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[2]. 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 obtained from the Highly Automated Macromolecular Crystallography Beamline (AMX) at the National Synchrotron Light Source II (NSLS-II) to provide data needed to create a model of a TrmD enzyme majorly involved in the bacterial protein synthesis cycle of 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[5]. 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. This can be hypothesized because the modification of RNA is prevalent in more than one percent of the bacterial genome and TrmD's are conserved across various species; this suggests that they play a critical part in cell vitality. It is essential for life as a determinant to maintain protein synthesis reading-frame. Mutation of the genes coding for this methyltransferase prevents the stabilization of the tRNA structure causing increased +1 frameshift errors, mutating TrmD to result in growth defects.

The discovery of the TrmD class of proteins has stimulated novel and relentless research by many research groups because of the understanding that most 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 determine a completely refined structure of TrmD and provide an analysis on what the model can tell us about the bacterial protein synthesis of tRNA-(guanine-N1)-methyltransferase and possible methods of inhibition. These findings are 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. As TrmD is a leading antimicrobial drug target, the development of TrmD inhibitors proves to be a promising future drug candidate

because of its presence specifically in prokaryotic organisms and its deep-rooted distinction from the human and archaeal counterpart Trm5 (which has the dinucleotide fold 27–30.)

Materials & Methods:

- **Variables and Controls:**

In this study, the control was the reference protein structure used to help provide a basis for the elucidation process of the x-ray diffraction data previously collected from the Synchrotron Light Source. The independent variable was the energy and angles at which the protein crystal was hit during the x-ray diffraction process, which was manipulated to be set to 92 picometers in order to scan the crystal with x-rays but not burn it/destroy bonds and devaluing the data collected. The dependant variable was the outputted diffraction patterns, intensities, electron density map, and eventually the novel protein structure. Extraneous variables were also present in the experiment but steps were taken to reduce their impact during the data analysis phase. Some examples are auger electrons deforming protein structure and thus losing intensity output for that region, and compton scattering which also resulted in loss of outputted intensities.

Materials Consist of Programs Used:
(in Chronological Order)

- HKL-2000
- SCALA
- Truncate
- Pointless
- SFTools
- SFCheck
- Phaser MR
- Molrep
- Refmac5
- COOT
- WaterTidy
- wwPDB Diagnostic "OneDep System"[6]

Methods/Procedure:

1. The initial phase of data analysis was done using an automated program to collect data on the Highly Automated Macromolecular Crystallography (AMX) beamline at the National Synchrotron Light Source II (NSLS-II). "HKL-2000"[7] was utilized to collect information regarding the diffraction pattern data obtained from crystallized TrmD protein exposed to high intensity X-rays. 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"[8]. Some 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 that needed to be calculated, 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 began computerizing the diffraction pattern images and started 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 and x-ray reflection showing on the diffraction image. This was 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 the data begun. 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"[9] accomplished this and filtered 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"[10] was used 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 estimated 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 also had to be collected from the diffraction data. "Pointless"[11] 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 fit into one of the major crystal point group classes and thus the space group and other important information about the crystal were found this way.

5. After the space group and average amplitudes have been calculated using the previous programs, "SFtools"[12] was used to convert all this raw data into a structured form that can be readable for mathematical uses and 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 it was integral for us to ensure the integrity of the files at work. "SFCheck" [13] is a program dedicated to providing insight to the files previously output by SFTools. The program gave us the ability to view very important information about the refinement process so far. Through this, we were able to see an early rendition of our R-Value, the correlation, Luzzati and Wilson plots (check B-factors of the dataset), as well as much more analytic information that allowed 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) were 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"[14] 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 was used in this data analysis is a liganded TrmD from *Anaplasma sp*. In the wwPDB as ID: 4IG6 [15]). The process of creating the electron density map from the reference and collected data was 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 which made 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.

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

*Image is courtesy of the Stony Brook University Physics Department. Image Source: [16]

8. After the electron density map was compiled, a skeleton of the protein structure was made so that there is a correlation with it in order to start focusing on the unique details of this specific TrmD. "MOLREP"[17] is an automated refinement/replacement program that took a similar protein model as a reference (PDB ID: 4IG6 [15]) 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) was also inputted. 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 gave us a starting point for our model by making the first part/its skeleton for further refinement.

9. Once the preliminary model was built, both automatic refinement program "Refmac5"[18] and manual refinement program "Crystallographic Object-Oriented Toolkit (COOT)"[18] were used to detail the model and constrain 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 fit the electron density. Each of the 219 amino acids was carefully reviewed and some stripped of R-groups to complement the electron density map. Since the electron density map is the experimental data, translational modifications were 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 was 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 18.1% was gotten after refinement and was deemed successful. However, RSRZ outliers and clashscore were lower than in higher resolution protein models because of the absence of information about the smaller details of the electron density.

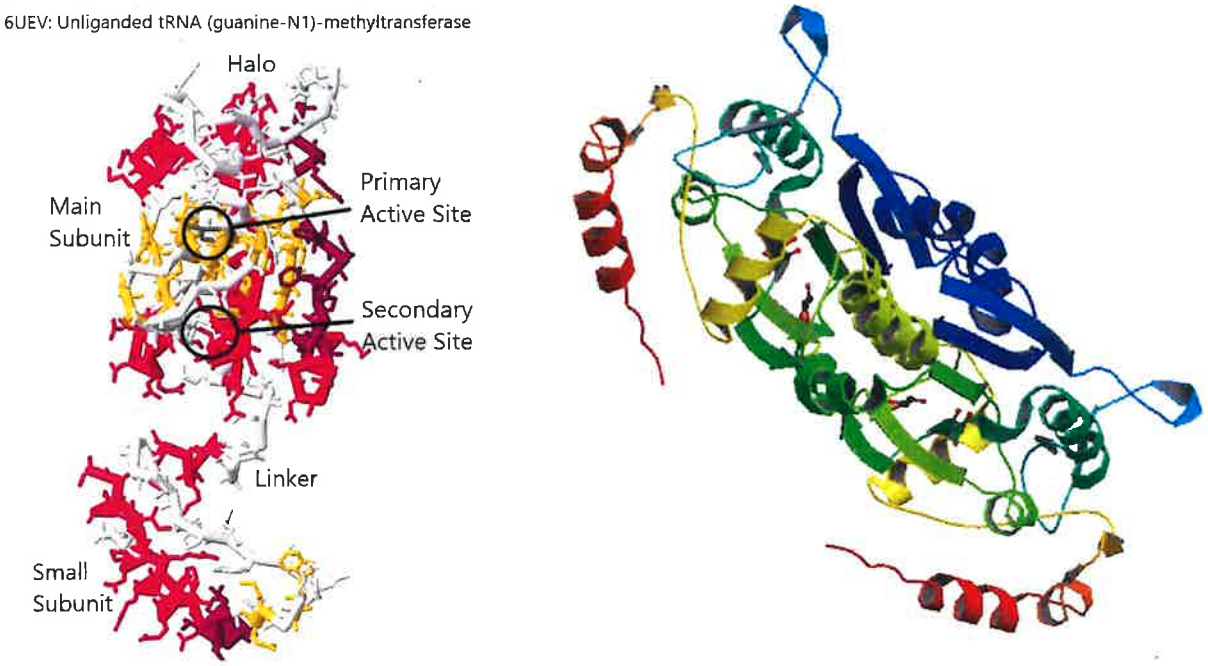
10. After the protein model complete, a final run of "WaterTidy"[18] was run in order to help sort out the water molecules bound by hydrogen bonding into more favorable locations and give them better organization within the protein model's file. Afterward, the protein model (now in mmcif format) and the experimental electron density data file (in mtz format) were sent in for review by the wwPDB deposition team. A diagnostic test was given on the protein's properties in reference to accepted scientific data and the experimental electron density. No major problems were encountered, the model was accepted and published in the RCSB PDB research journal.

11. After being reviewed and published, the protein model was used to observe the properties of the protein confidently and draw conclusions and find potential areas of further research using "COOT" and "Jmol"[19] to model the protein in 3D space. Since our investigation on this diffraction data has found the data to accurately observe a halo region and linker region electron density, our model provided 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 were found by comparing liganded TrmDs to the model being proposed in this paper. The differences outline the major factors in keeping/changing dimerization and conformation in certain ways to allow for the function to be carried out. Variable regions found from the differences between the unliganded and liganded TrmD 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 acid conformations are changed by the binding of a ligand and the effects they have on the linker region and halo region of the protein, analysis and conclusions, as well as future studies was drawn from this project in hopes of finding a novel method of treating bacterial disease. A research paper publication is currently in the process of being written.

Results/ Scientific Analysis:

Through the use of 3D imaging software (COOT/Jmol) the completed model could be visualized on a computer, allowing for observations to be drawn from the enzyme's structural properties. The enzyme's catalytic domain was one of the first areas of analysis due to the importance it has for overall protein function. Specifically, the TrmD's catalytic domain was believed to have certain sidechain interactions in the presence of a substrate that would generate the temperature factor and conformational differences seen in the enzyme between liganded and unliganded states. To compare what side-chains play major roles in the enzyme's catalytic activity, an inspection was done on the conformation of the catalytic domain when a ligand structurally analogous to the intended substrate was bound and when the protein was in an unliganded state (excluding interactions from water molecules and cryoprotectant). The protein structure experimentally determined in this study (PDB ID: 6UEV) served as the unliganded model while the reference model (PDB ID: 4IG6) served as a liganded form to compare 6UEV with. The reference structure (4IG6) was bound to S-adenosylhomocysteine, a product in the enzyme's catalysis reaction. This molecule was structurally similar to the methyl-donor that the TrmD uses as a reactant and thus simulated most of the intramolecular effects on the enzyme during catalytic activity.

- Image of monomeric and dimeric units of unliganded tRNA (guanine-N^1)-methyltransferase 6UEV[20]:

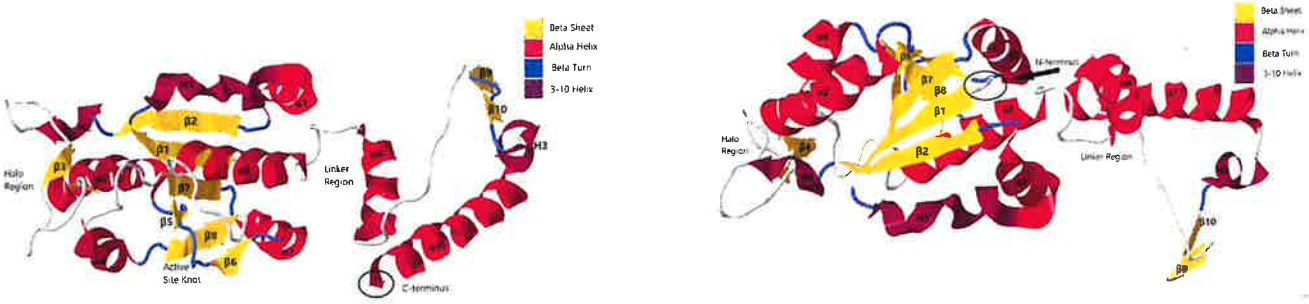


Active Site Interactions:

1. The Knot Region/Adenine Binding Loop:

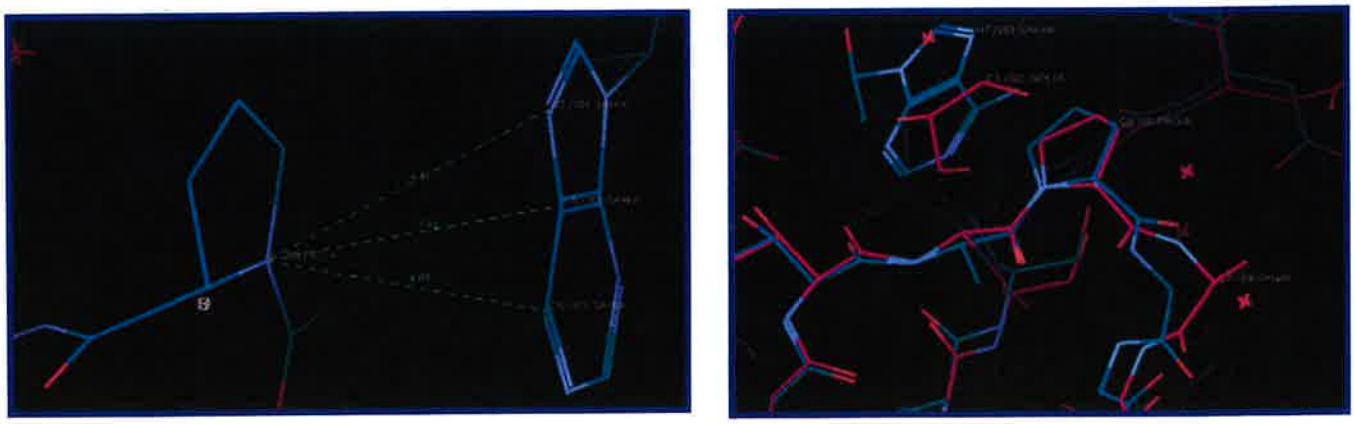
Observing the differences in conformation of the catalytic domain's side amino acid chains due to the presence of a substrate provided insight to which amino acids play a major role in affecting overall protein structure. Most of the observed conformational changes occurred in "knot" structure in the active site. This region of the active site has an adenine binding loop surrounding the active site with a β – sheet going through a hole made by another β – sheet wrapping around it. The knot region of the active site is visualized in Figure-(1), which labels the secondary structure of the protein.

Figure-(1): The TrmD enzyme with its secondary structure mapped out. The "Active Site Knot" outlines where one β – sheet ($\beta 8$) runs through the gap made by β – sheets $\beta 5$ and $\beta 6$. $\beta 8$ was observed to have strong electrostatic interactions with the ligand in structure 4IG6, but failed to interact and thus had conformational changes when looking at 6UEV (the unliganded structure).



The changes that resulted from the presence of a substrate in the active site were observed to come from multiple amino acid side chains, but three major contributors were found on the knot region of the protein. Proline-88 was seen to change conformation in the presence of a substrate because of π – *cation* interactions between the aromatic character of the substrate (S-adenosylhomocysteine) and the partial positive charge on the nitrogen on proline-88. Figure-(2) shows the interaction proline-88 has with the substrate in liganded model 4IG6 and compares it with the proline's conformation when the protein is devoid of a substrate-like ligand (model 6UEV).

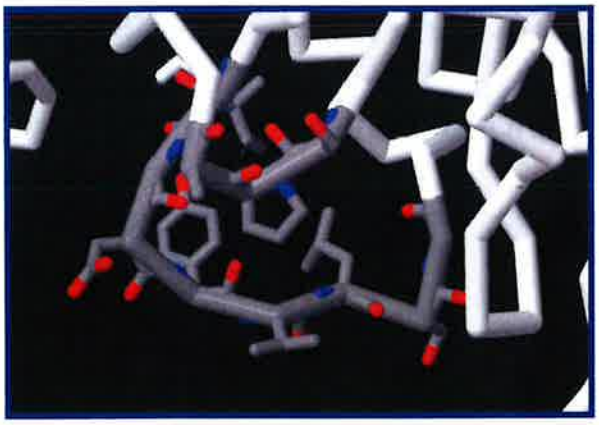
Figure-(2): The leftmost image shows the interaction between proline-88 and the 'purine' component of S-adenosylhomocysteine in liganded model 4IG6[15]. The cationic character of the nitrogen in proline-88 electrostatically interacts with the aromatic character of the purine ring, forming a π – *cation* bond.
The rightmost image compares 4IG6's knot region with 6UEV's. Although proline-88 is only slightly affected, the loss of the π – *cation* bond destabilizes the knot region, increasing the B-factor in the region.
Liganded TrmD = Cyan color Unliganded TrmD = Purple color



The $\pi - cation$ bond is shown to have a molecular distance of about 4-angstroms. This is consistent with the generally accepted distance for significant effects, and the coulombic force applied by the bond is able to keep tension on the nearby knot to keep (β8) sturdy. The substrate molecule is benefited by the bond because the proline-88 acts as a docking site for the purine component, thus allowing for the molecule is decrease its rotational kinetic energy because of the favored parallel conformation of the purine and proline. However, the purine is not only interacting with the proline and thus the coulombic force on the substrate does not increase its translational kinetic energy. Other parts of the active site interact with the other end of the S-adenosylhomocysteine to cancel out the forces. The rightmost image in Figure-(2) agrees with stability hypothesis because in the unliganded model (purple), only a weak ion-dipole force may form between the glycerol molecule and the proline. This glycerol is not well bound to the active site and thus cannot apply a stabilizing force to the proline to give β – sheet eight more tension and stability. The consequence of this is that the B-factor of the region increases and the conformations of nearby side chains are liable to change. This effect is most heavily seen to change the B-factor and conformation of Lysine-89, which shows large positional deviation between the liganded and unliganded models.

The Adenine binding loop itself plays an integral part in docking the correct substrate into the active site of the enzyme by interacting with its bicyclic 'purine' or 'adenine' ring. This site is necessary for making sure substrate molecules can enter the active site easily and interact with side chains to transfer methyl groups from the methyl donor molecule (S-adenosylmethionine) to guanine-37 (the 3' end) of a bacterial tRNA anticodon docked in the p-site of a 70S ribosome[21]. The side chains on the adenine binding loop and above it creates a pocket for the active site to form, favor substrates such as S-adenosylmethionine (SAM) to enter it, and create a pathway for a methyl group to be distributed to an unmethylated SAH on another TrmD enzyme also. Based on the unliganded structure of the enzyme, observing the rotameric positions each sidechain is in could provide insights into how SAM is attracted to the active site and how methyl group transfer is possible between two TrmD's in a dimer complex. Figure-(3) visualizes the binding-pocket formed by the adenine binding loop and the residues that makeup its amino acid sequence.

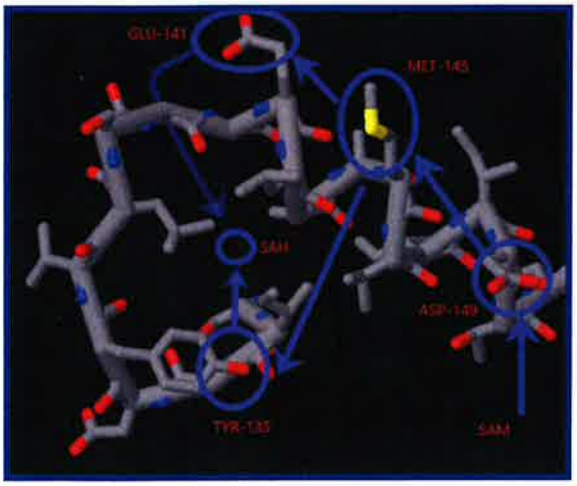
Figure-(3): The leftmost image shows the outer catalytic domain of the TrmD from a top view. The side chains at the upper part of the region tend to have more charged species (proline, aspartic acid, carbonyl groups and secondary amines) while the bottom of the binding loop tends to hold hydrophobic, nonpolar side chains such as the tyrosine ring, leucine, and isoleucine. The rightmost image shows 6UEV in a dimer form. The electron density from the TrmD subfamily of proteins points to the individual enzymes forming dimers to carry out their function[12]. In 6UEV, it was observed that the dimer's substrate pockets were ~15.28 angstroms away as shown in the rightmost image.

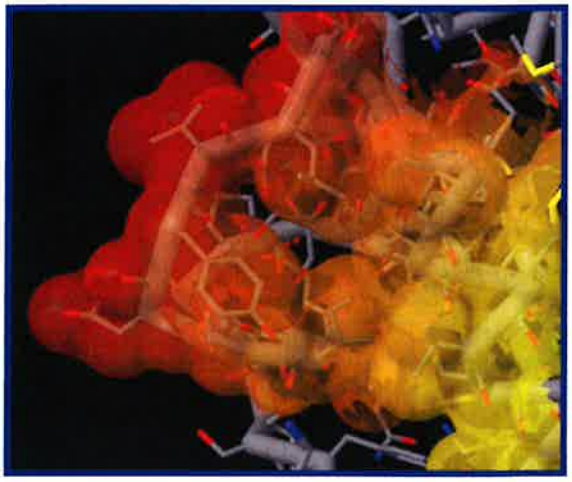


The knot and adenine binding loop play a role in docking SAM and transferring methyl groups to other TrmD's (dimer) that may have an active p-site 3' tRNA attached so that it may continue methylation. Thus one TrmD in the dimer may act as a storage while the other (actively bound to a 3' anticodon of tRNA) receives methyl groups to keep methylating each anticodon. However if this mechanism of methyl group transfer is not observed in future

research, it is possible for each monomer of the dimer switches back and forth between methylating each anticodon as it enters the p-site. However, through studying 6UEV, a plausible mechanism for the first case can be found. The actual path in which the methyl group travels is unknown. However observing the alpha helices around the axis of symmetry provided insight into the transfer mechanism (shown in Figure-(3)). Since the enzymes are arranged antiparallel to each other in dimerized form, an electronegative pathway could be taken by the positive methyl group after leaving the sulfur it is attached to initially in SAM. The movement mechanism would allow the methyl group to be attracted to a negatively charged species on the alpha helices and give it momentum to move along the path, and insulating/hydrophobic groups would have to also alternate with the negatively charged species on the alpha helices to cushion the methyl group. Using the determined structure of 6UEV to map out the side chains in the alpha helices and adenine binding loops, Figure-(4) was constructed to show the possible route taken by the methyl group.

Figure-(4): The leftmost image shows the possible routes that a methyl group could take when crossing the dimer-interface at the two alpha helices. The methyl group would start at the first adenine binding loop where it is removed from SAM. Afterwards it travels through a hole in the hydrophobic field and attracts to aspartic acid-149. From there it may perform an S_e1 electrophilic attack on methionine-145's sulfur, causing a methyl group to be expelled and either travel around the second adenine binding loop through glutamate-141 or it may attract to tyrosine-135 and travel through a hole in the hydrophobic field like it did to get out of the first adenine loop. The rightmost image shows the hydrophobic field hole.

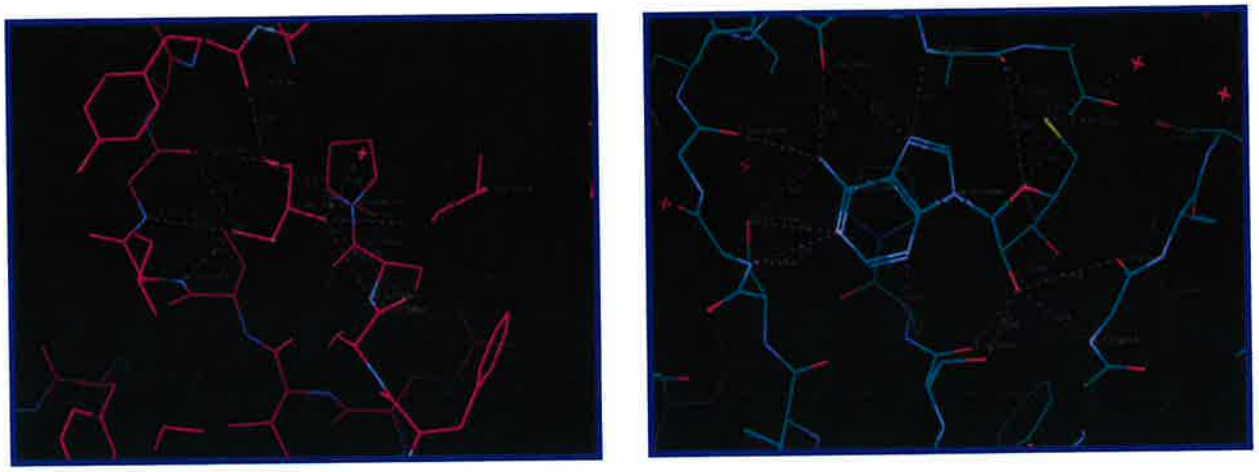




2. The Catalytic Domain Interior/Secondary Active Site:

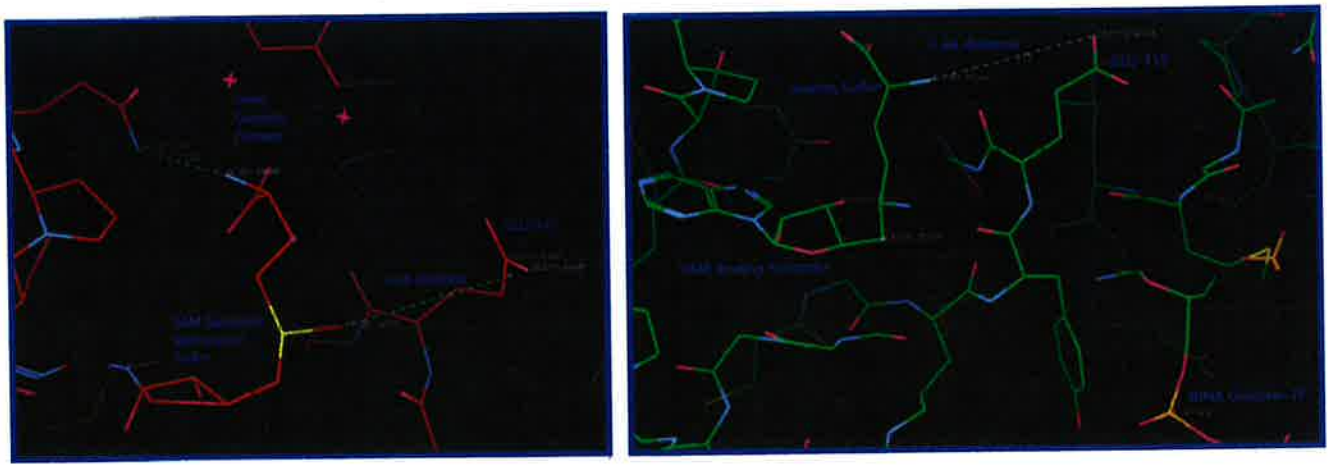
The interior region of the catalytic domain is important in transferring methyl groups from SAM in the active site to the tRNA molecule bound to the other end of the enzyme. However the mechanism behind this transfer is not well described and a sufficient electron path through the protein has not been mapped out. 6UEV is unliganded and thus the pathway for this methyl group transfer mechanism may be observed since the conformation of the active side chains would be in different conformations than in a bound-state enzyme where the presence of a ligand would change the conformation of the contributing amino acids. To differentiate which side chains/amino acids play a major role in demethylation of S-adenosylmethionine (SAM), the active site of 6UEV was compared to the active site of another tRNA (guanine-N^1)-methyltransferase,but bound with SAH (PDB ID: 4IG6[15]). Figure-(5) outlines the major interactions found by comparing 6UEV and 4IG6.

Figure-(5a): The leftmost image shows the conformation of the amino acid R-groups found in the deeper parts of the catalytic domain when the TrmD enzyme is found in an unliganded state. The rightmost image outlines the conformation when the enzyme is bound by a natural product, S-adenosylhomocysteine (SAH). Clear differences can be seen stemming from charged and partially charged side chains and cascading to other parts of the enzyme. The SAH ligand clearly has more interactions with the active site than the glycerol in the unliganded site.



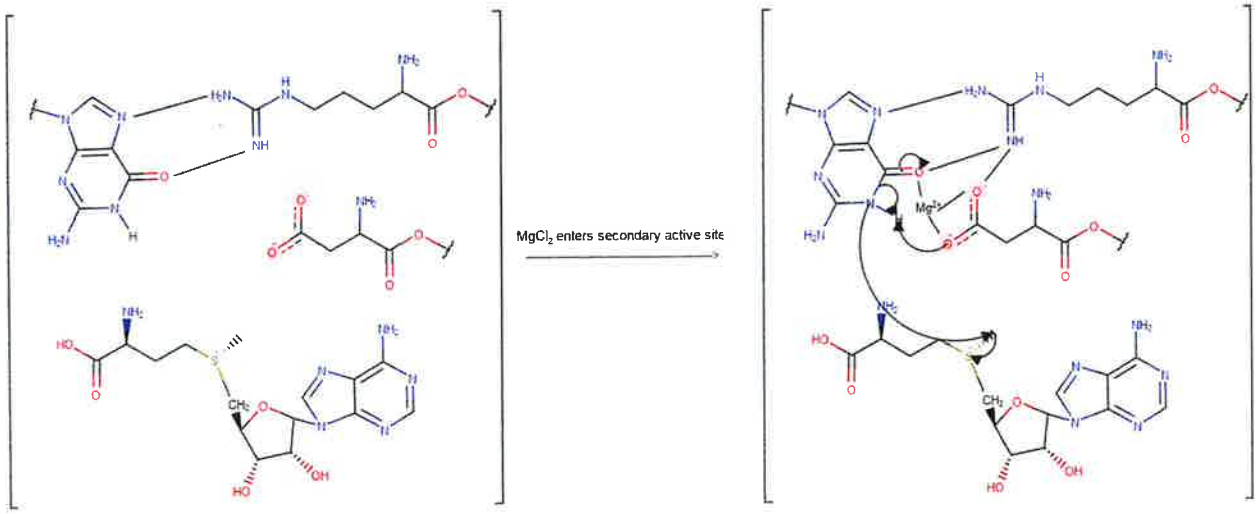
The heteroatoms found in the adenine component of the SAH molecule are vital to docking with both the adenine binding loop and the deeper regions of the catalytic domain. The electrostatic forces are so prevalent in the region that even in the unliganded model, cryoprotectant was able to seep into the active site and form intermolecular forces in the same place that the adenine ring usually would inhabit if liganded. However the adenine ring has significant bonding with the deeper regions of the active site, most notably with glutamic acid 115, which has a negative charge and is inferred to start the enzymatic activity cascade by allowing the methyl group bound to the sulfur on SAM to act as a leaving group and thus allow the positive charge on the sulfur to reduce. With some methyltransferase enzymes that use this mechanism, enzyme 5,10-methylenetetrahydrofolate reductase (MTHFR) will take unbound SAH and remethylate it for use as a methyl donor again[22]. However, in the ribosome where there is constant movement of large tRNAs and a constant need for SAM, it is likely that the TrmD dimer only releases SAH from one of its adenine loops while the other monomer actively methylates tRNA. This could explain why the glutamic acid in structure 4IG6 is positioned away from the sulfur atom in SAH. If it was not, the negative moiety of neutral sulfur and the glutamic acid side chain would block SAH, leaving it unmethylated and stuck in the active site. This may actually happen while a 3' end of a tRNA molecule is bound to the same monomer if the storage-dimer hypothesis is correct. By comparing Figure-(5a) with Figure-(5b), evidence is collected as to if this effect is significant and supported.

Figure-(5b): Using a published structure of a SAM-bound TrmD belonging to *Pseudomonas aeruginosa* (PDB ID: 5WYR[23]), the conformation of glutamic acid-115 can be observed. The leftmost image shows the interaction between SAM and glutamic acid in 5WKY. The rightmost image shows the conformation of the active site when a tRNA molecule is bound to a TrmD dimer with an analog to SAH in the same monomer. A published TrmD structure (PDB ID: 4YVJ[24]) is liganded with both and provides a model to observe the effects of the tRNA.

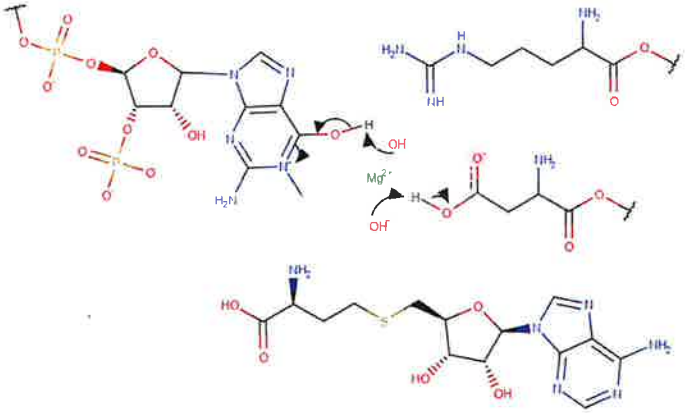


In both images of figure-(5b), the sulfur or 'simulated' sulfur (rightmost figure uses a positive nitrogen species instead) is not close enough to ionically interact and thus draw in glutamic acid-115 to allow the removal of the methyl group and transfer to guanine-37. In the rightmost image of figure-(5b), tRNA can be seen bound near the active site as well, and still had no effect on the movement of glutamic acid. This suggests that the hypothesis is incomplete and the wrong amino acid was being looked at and that another binding force may be necessary to regulate the methylation of tRNA. However Aspartic acid-169 is in stable conformation to collect the methyl group off of the sulfur when tRNA is bound in figure-(5b). By flipping the model using COOT, the protein bound with tRNA and a substrate is able to be seen from an angle that shows a clear path the methyl group takes from the substrate to the guanine-37. However, if this mechanism is correct and the amino acid carrier is correct, then the aspartic acid-169 would have to change conformation and get unusually close to guanine-37's aromatic ring system. Since $\pi - anion$ interactions are extremely rare and need certain conditions to happen, this is most likely not the cause. As stated above, a third ligand or co-factor is most likely at play to allow favorability of the methyl transfer from aspartic acid to N^1 on guanine-37.

Figure-(6): As explained in the two above hypotheses, magnesium cations are likely to be involved in mediating the transfer of the methyl group from S-adenosylmethionine to guanine-37 of a tRNA anticodon. If the storage hypothesis is correct, where one monomer of each dimer is an exchange region for methylating SAH to SAM to supply methyl groups to the other monomer, one Mg^{2+} cation would necessary per dimer. This cation would have to implant itself into the catalytic domain of the enzymatically active monomer as shown below. The unusually long bonds between heteroatoms represent hydrogen bonding and electrostatic interactions. The introduction of cationic magnesium promotes involvement of Aspartic acid-119 via electrostatic interactions, and thus initiating the transfer.



Products of methyl transfer mechanism:



Since the conditions in the bacterial 70S ribosome are slightly basic, hydroxide anions enter the active site to deprotonate the guanine on the tRNA and the aspartate to return the system to a more favorable state and allow the process to restart.

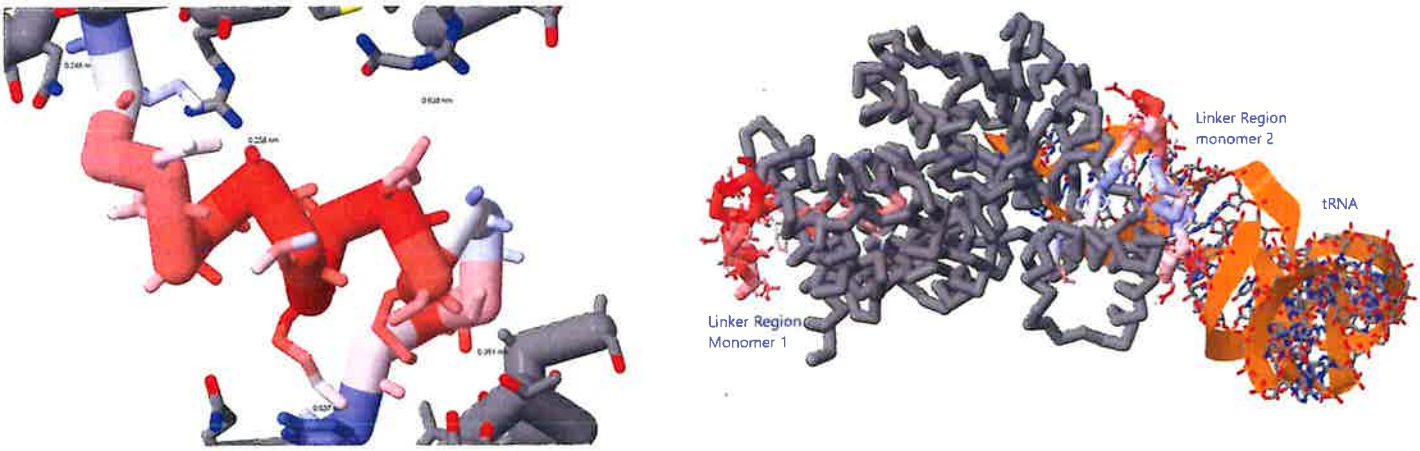
- These images were created by using organic chemistry modelling software program "MarvinSketch"[25].

Interactions Regarding Flexible Outer Regions:

1. TrmD Linker Region:

The linker region of TrmD proteins is rare to see in unliganded models. Even in many liganded models, the linker region still may not fully be present because of its flexibility. Model 6UEV is unliganded with a substrate but still has the presence of a linker region. Figure-(1) shows the linker region as being composed of alpha helix-6 and a variable region. The alpha helix supports the linker region's integrity by giving it a stable secondary structure and forming bonds with the large subunit of the protein where the active site is situated. The variable region merges into the second subunit of the protein, which is significantly smaller but contains the C-terminus and important amino acids for allowing the enzyme to form dimer structures and become functional. Using 6UEV and comparing it with liganded structures that also contain a linker region (4IG6 and 4YVJ) allowed for mapping and analysis of the function of the linker region to be found. The path at which electron networks change the conformation and motility of the linker region was of importance to analyze as it may provide useful evidence as to how ligands affect the linker region and why.

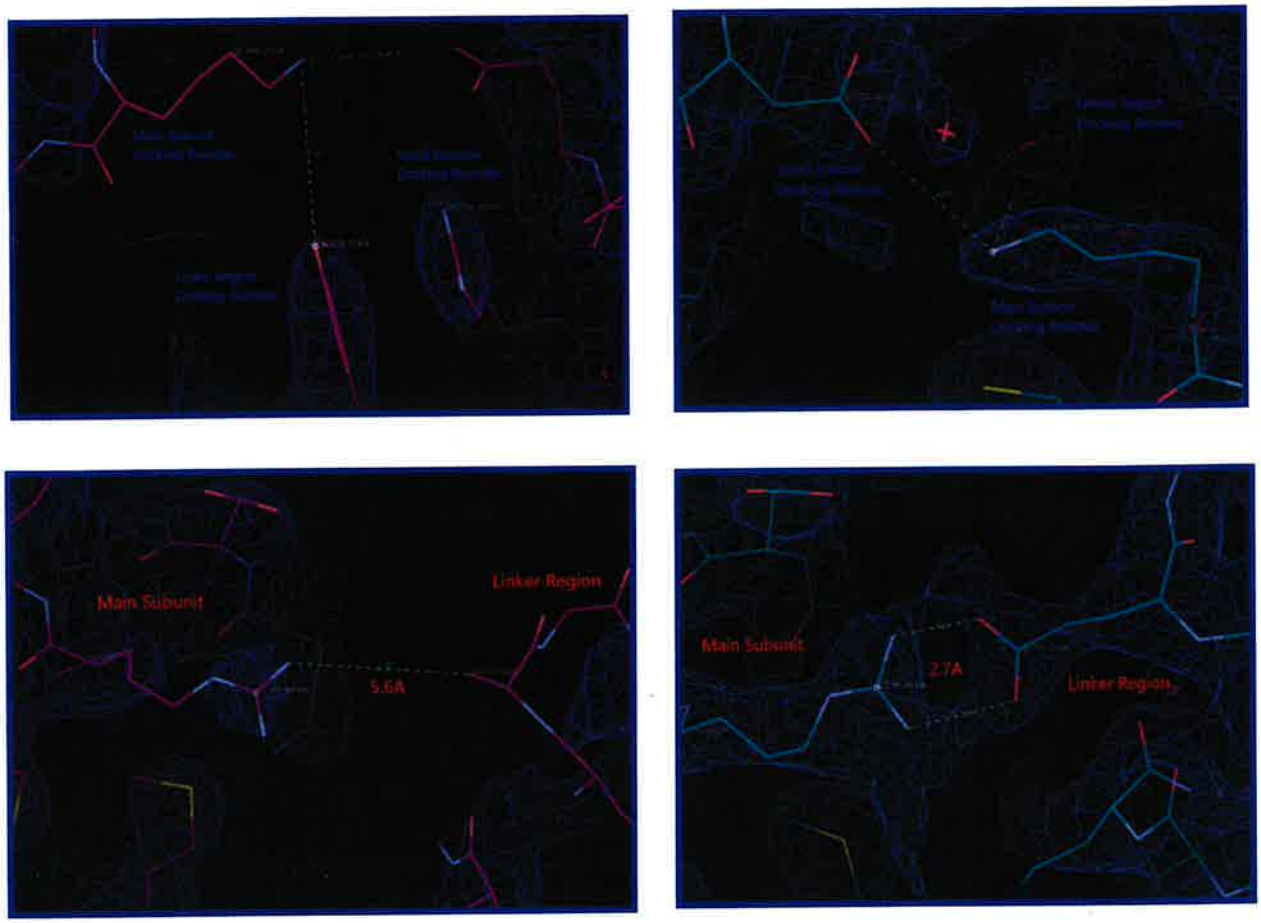
Figure-(7): shows the general properties of the linker region. The leftmost image shows the structure of it gathered from the diffraction data to make model 6UEV. The rightmost image compares the temperature factor (B-factor) of an unliganded linker region to that of a liganded linker region with tRNA present and bound to a second catalytic domain closely connected to the deep catalytic domain where the substrate SAM would enter (PDB ID: 4YVJ). *Red regions indicate high B-factor while blue regions indicate tense and sturdy amino acids.



Since the enzymes are connected in an antiparallel fashion, the linker region of one monomer will actually be opposite to the side its main subunit is on. For example, monomer one in figure-(7) donates the methyl group from its active site to the tRNA bound to the linker region of subunit two. By examining the beta factors of the different linker regions, it is apparent that the binding on substrates helps enforce a strict conformation (low B-factor) on the linker region. If this was not the case, the electron movement pathway would be compromised because of the motility of the linker region, causing the methyl group to fail to transfer to guanine-37. This explains why there is an island of high conservation around reisudes 179-183. These residues are integral to keeping the second subunit tucked in to make room for tRNA and sterically inhibit it from entering the primary active site (figure-13).

Figure-(8): The specific bonding sites on the linker region that allow it to stay stable when substrates are bound to the dimer are characterized by examining unliganded structure 6UEV and structures 4IG6 and 4YVJ. Since 6UEV only has glycerol molecules in its active site, it suggests that the linker region will have a high B-factor as barely any electron networks are being stimulated and no molecules are directly in contact with the linker region. 4IG6 has S-adenosylhomocysteine in its active site, which electrostatically interacts with multiple amino acids in the catalytic domain. Consequently, the electron binding networks can be traced from the active site to the linker region, exposing how a true substrate changes the stability of the linker region. Lastly, the tRNA molecule binds adenine-38 and guanine-37 into the space between the linker region and active site of the dimer. Now multiple areas of the dimer are interacting with another large molecule, forming an extremely stable structure. This is seen in figure-(6).

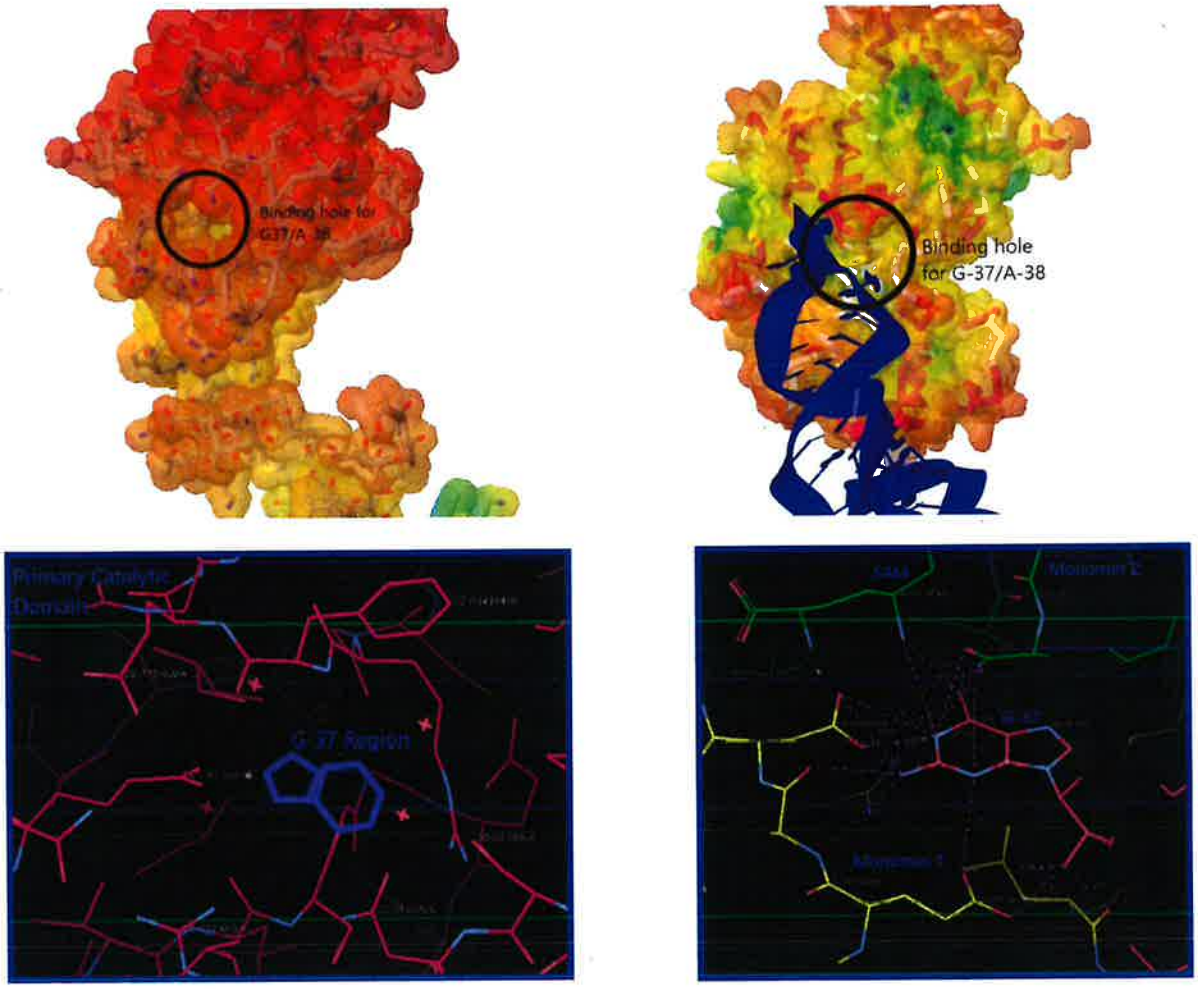
Top/Bottom Leftmost Structures: 6UEV Top/Bottom Rightmost Structure: 4IG6



The reason behind the change in B-factor originates from the active site residues changing conformation via the induced fit model. This cascade leads to lysine-166 to protrude out more from the main subunit, forming an electrostatic bond with tyrosine-176 on the linker region and glutamic acid-203 on the small subunit of the protein. Thus lysine-166 is extremely important in initializing the stabilization of the linker region and changing its conformation when tRNA arrives at the second catalytic domain. In figure-(8), the distance between the three charged species is about 5.0-angstroms on average in 6UEV and about 4.0-angstroms in 4IG6. The maximum distance at which electrostatic interactions can apply a significant force is about 4.0-angstroms[26], showing that substrate binding that moves lysine-166 even one angstrom closer to the other species can have a significant effect. Once tRNA moves into the region, it will be attracted to the linker region of the dimer and move into the space between the two subunits. This movement brings the flexible linker region upwards and locks the tRNA into place as shown in figure-(9). The guanine-37 and adenine-38 of the tRNA then interact with the protein as shown in figure-(6) by moving into the small electrostatic holes shown in figure-(9). The flexible linker at this point is completely stable and it important for keeping the massive tRNA docked on the TrmD until guanine-37 is successfully methylated. As the linker region locks the tRNA into place, handing it over to the halo region to direct further, it also serves a second function, and arguably a more important one. When the linker region moves closer to the main subunit of the protein, it causes the secondary subunit to eclipse the adenine binding loop of the primary catalytic domain. This blocks tRNA molecules from entering the domain but still allows small monomers such as SAM to implant itself into the site.

Figure-(9): The leftmost image shows the electrostatic potential of the experimentally determined model, 6UEV. The induced fit effects are not active because no true ligands are bound to the monomer. Since the leftmost image is

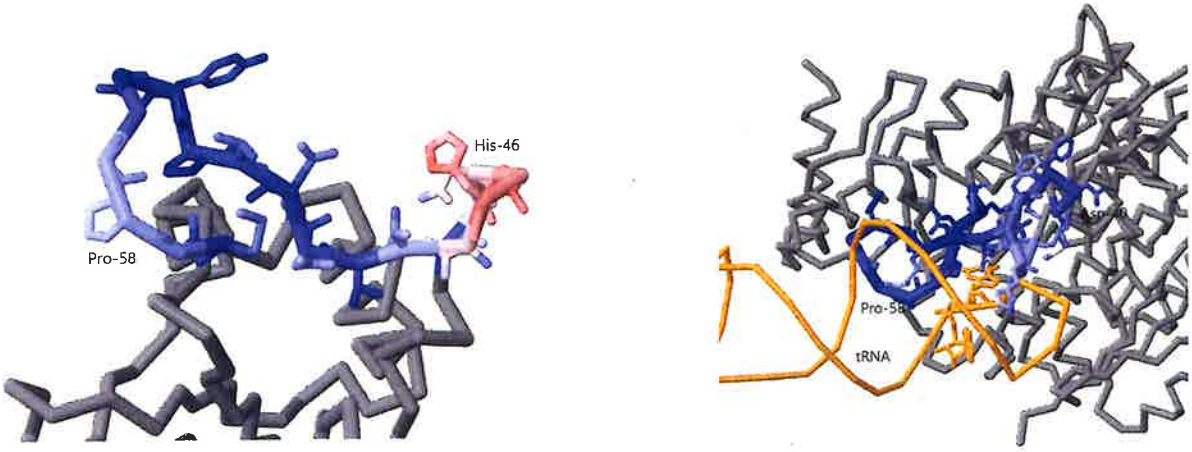
a monomer, the linker region that would bind and lock tRNA into place is not present. However this allows for a clear view of the secondary catalytic domain pocket that the tRNA would fit into. The rightmost image shows tRNA bound to a dimerized TrmD, 4YVJ. The linker region of a monomer locks the tRNA into the pocket of the other monomer, forming a complex that can undergo methylation with an extremely high success rate. The bottom images are models of the secondary catalytic domain with tRNA bound to it in structure 4YVJ (right) compared to in 6UEV.



2. TrmD Halo Region:

The halo region of the protein is also usually absent from TrmD models when in an unliganded state. This region is also believed to be extremely flexible and interact to help form a dimer structure. However research on this region is sparse, especially when characterizing the region in an unliganded model. With model 6UEV, the halo region can be compared with liganded models to elucidate the mechanisms within the protein that allow for it to be flexible or stable in certain states and why it is important for the protein's function.

Figure-(10): Shows the general properties of the halo region of the protein based on B-factor and interactions with substrates. In 6UEV (leftmost image), the halo region is part of a monomer and has no true stabilization. It may interact with the antiparallel protein not shown that forms a dimer complex, but despite this, high motility is still evident. The rightmost image shows the interactions of the halo region when a substrate is bound to the active site of the protein and tRNA is in the vicinity. This was modelled using published TrmD PDB ID: 4YVJ [24]. It complements the data gathered on the linker region, as the halo region seems to be more stable when in a bound-state and most likely contributes to stabilizing the antiparallel monomers in the dimer.



The leftmost image in figure-(10) shows that the halo region is generally more stable than the linker region, as many electrostatically interactive species populate it such as tyrosine, proline, and histidine. However, the histidine is not able to interact with the rest of the protein, thus failing to stabilize that portion of the region. This is because the histidine needs to stay relatively flexible so it can help dock tRNA into the enzyme. While the linker region did provide some support for the tRNA to anchor into the secondary active site, it mainly disallowed tRNA from being able to bind into the main active site where SAM is suppose to bind. Since SAM and guanine are extremely similar in aromatic nature, it would be possible for tRNA to bind to the primary active site and fail to be methylated if the linker region did not change conformation to block the adenine binding loop from having large molecules such as tRNA from entering the vicinity. The cationic R-group and aromatic side chains that litter the halo region explain its functional significance. All these R-groups are able to strongly interact with tRNA through $\pi - stacking$, $\pi - cation$ and hydrogen bonding.

Figure-(11): Comparing the primary active sites of liganded (right) and the novel unliganded (left) model showed that the reason for the decrease in B-factor in liganded forms not only stems from tRNA binding, but also from the increased rigidity put on the trefoil knot and adenine binding loop as discussed in figure-(2). This increased rigidity favors a single conformation of trefoil knot residue arginine-89, which plays a significant role in supporting the halo.

Genetic/Amino Acid Sequencing:

1. Phylogenetic Tree

Using "ClustalOmega"[26], we created a phylogenetic tree comparing many of the deposited and sequenced TrmD proteins. On this tree, we observed three main families of the TrmD. According to the phylogenetic tree produced,

the most similar protein to the one our team deposited was Q2GIL5 (HZ strain *A. phagocytophilum*). This shows that many of the features between what we have produced and recorded in our research are also conserved for the Q2GIL5. This allows us to draw conclusions about the structure of our own protein and provided insight for our refinement techniques. The usage of the phylogenetic tree allowed us to further dictate which proteins show the most conservation to our own.

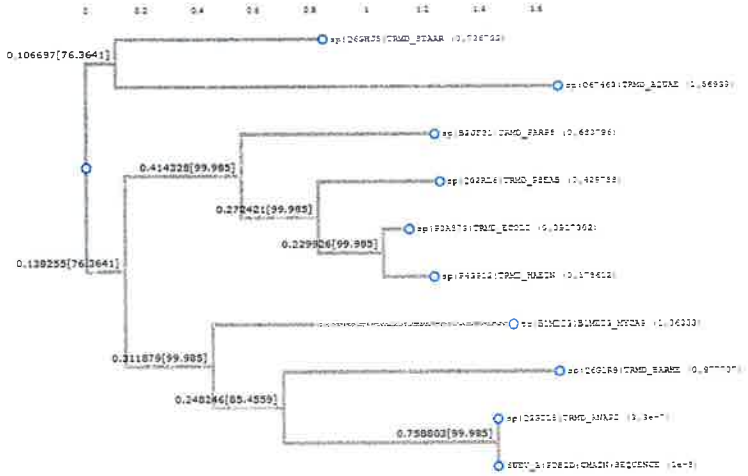


Figure-(12): Phylogenetic Tree. This shows three major families of the TrmD protein (dubbed A, B, C, from top to bottom). 6UEV is shown as the bottommost protein on the tree, in family C. Proteins that are closer on the tree are more conserved and are generally closely related (relative to the other proteins). We can observe in family C that 6UEV is directly next to Q2GIL5, leading us to infer that the two are very highly conserved.

2. Amino Acid Comparisons

In addition to allowing us to make phylogenetic trees, Clustal Omega allowed us to create sequence comparisons. The alignments created by the program show the conservation status among a group of TrmDs. We can see that at certain points there are much higher amounts of conservation, an example being the assumed active site. As all of these proteins serve mainly the same function, it is safe to assume that the active site would be highly conserved something which is seen in the alignments. This program also enabled us to see the conservation of alpha helices and beta sheets between the different proteins, something which proved to be useful in manual protein refinement. This helped us in finding and confirming our belief in where the active sites were- a group of highly conserved regions between all the TrmD proteins in the family.

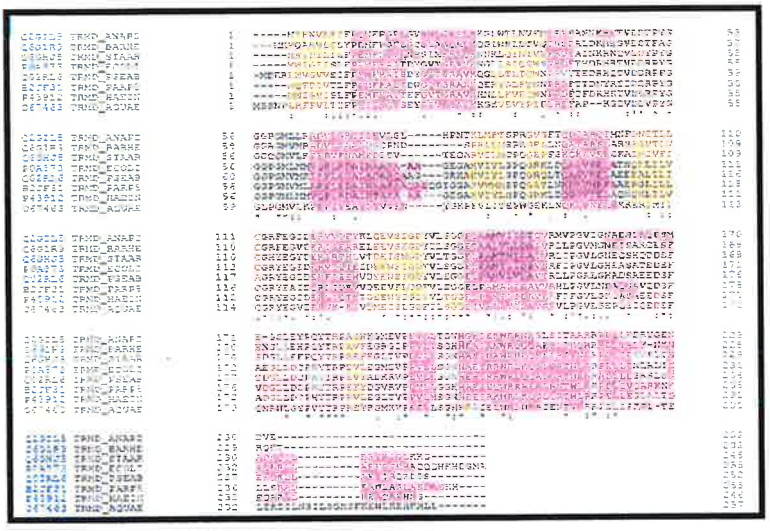


Figure-(13): Protein Alignment. Shows the alignment between similar TrmD proteins. Shown above is each amino acid conservation status with respect to the amino acid in the same spot on the other TrmD proteins. ("*" indicates perfect match between all in alignment, with "," indicating poor conservation and ":" indicating moderate conservation. None of these indicates little or no conservation.) In this, we are also able to see the alpha helices and beta sheets of each protein, and an easy comparison to the others at the same position. Amino acids with a pink highlight are indicative of an alpha helix, and the

yellow highlight indicative of a beta sheet.

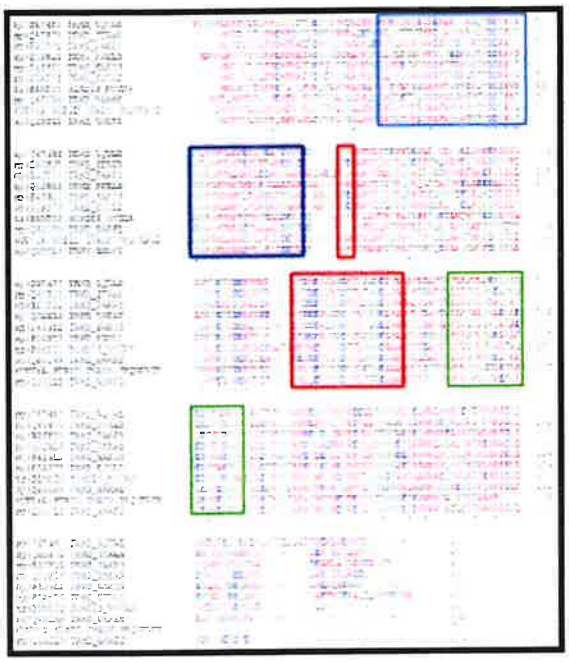


Figure-(14): Protein Alignment (color-coded by amino acid). Similar to figure-(13), the above is a comparison between major TrmD proteins, including the one we researched. Each amino acid residue is color coded as to provide an easier comparison between proteins. Labeled above are the following areas of the protein: Active Sites: 88-190, 130-150 bp (Red Outlines). Linker Region: 160-180 bp (Green Outline). Halo Region: 40-60 bp (Blue Outline). As shown, the active sites are highly conserved between all of the TrmD proteins. There are many very similar, or even the same in most cases, of the same amino acids in the same spots respective to their other TrmD counterparts in the outlined active site area. This leads us to the conclusion that all of the proteins are very similar in their structure and function.

*Images figure-(13/14) were created using software program "ClustalOmega"[26].

Discussion/Conclusions:

1. Catalytic Domains/Active Sites:

We have concluded that there are two active sites within each TrmD enzyme. The primary active site, where the methyl donor binds, and is composed of a trefoil knot and adenine binding loop. The other identified active site is situated above the linker region and is deemed the secondary catalytic domain. Using the program Jmol, we analyzed the electron density above the linker region for unliganded and fully liganded TrmDs, where we found a highly negative and electrically prevalent region with a pocket for tRNA guanine-37 to fit into. Upon further research and investigation, we came to the conclusion that there was more functionality of this region than was initially thought. The conclusions made about the analysis of each region using 6UEV and comparisons follow:

- Primary Catalytic Domain:

The trefoil knot and adenine binding loop of primary catalytic domain were heavily studied because of previous research pointing to the significance in that area. The results show that there are multiple amino acids that contribute to docking the purine component of S-adenosylmethionine to the region. Proline-88 plays arguably the most important role in the outer primary domain because of its ability to dock the purine component via a strong $\pi - cation$ interaction. In return, the substrate puts tension on the trefoil knot by applying an equal and opposite coulombic force on proline-88, causing an attraction. This applied tension increases the structural integrity of the adenine binding loop and trefoil knot (originating at β8). Comparisons between liganded TrmDs and the novel structure (6UEV) showed that without this interaction, both the adenine binding loop and the halo region of the monomer lose their rigidity. Figure-(11) proves the rotameric shift arginine-89 has when a ligand is present is significant in keeping the halo region together, as it creates two glutamic acid bridges that will not form when the primary active site is unliganded and thus the adenine binding loop flexible. However, this flexibility is most likely beneficial as it keeps the primary active site open for SAM molecules to transfer into it through the diminishment of permanent electrostatic potential barriers seen in figure-(4).

The deeper end of the primary active site has a mechanism to help remove methyl groups from sulfur on SAM molecules. However, the initiation of the mechanism requires bound tRNA to the secondary catalytic domain and

the presence of a divalent metal such as the magnesium cation. The deeper end consists of multiple negatively charged R-groups such as glutamic acid 115 and 141 in reference to 6UEV along with other electrostatic side chains to help mediate the transfer of a positively charged methyl group. Aspartic acid has been referenced as a possible amino acid to mediate the movement of a methyl group from SAM to G-37, but the conservation of aspartic acid is not apparent when viewing the amino acid sequences. Instead it seems not only is aspartic acid capable of mediating this transfer, but other negatively charged and fairly long amino acids may take its place such as glutamic acids or arganines. Thus while the outer trefoil knot and adenine binding loop are built to primarily interact with the purine component of SAM and similar substrates (SAH), the interior pocket of the primary active site acts as a bridge between the primary catalytic domain of one monomer and the linker region of the other. In monomeric function, the binding status of the primary catalytic domain (liganded strongly vs poorly) will determine the subsequent flexibility and b-factors of the linker and halo. If the trefoil knot and adenine loop are interacting with a ligand, the halo region will become rigid and locked. If the inner pocket and ends of the adenine loop are bound well, the linker region will become better stabilized and start to recede cause the small subunit to block the primary active site from tRNA.

- Secondary Catalytic Domain:

The secondary catalytic domain only forms when a dimer of the protein is present. One unit alone cannot form all the necessary components of the secondary active site. This is because the active site relies on the primary catalytic domain and halo region from one monomer but also needs the aromatic side chains and linker region of another monomer as well. The high density of aromatic or cationic R-groups near the end of the first subunit and beginning of the linker region now are understood. They are necessary for forcing tRNA to correctly bind guanine-37 to the secondary active site pocket. The movement of the tRNA itself into the secondary active site causes immense losses in rotational and translational kinetic energy for the dimer complex, stabilizing the entire system and allowing it to modify the RNA accordingly. The acquired stability also affects the active site and halo region of the dimer. The methods at which the tRNA leaves the secondary active site and when exactly it gets methylated are still under debate, but it most likely occurs when the tRNA is at the a- or p-site of the 70S ribosome. One possible method in which the tRNA becomes unbound from the dimer is by allowing the dimer to disassociate after methylation has been completed. If the dimer dissociates, the secondary active site and the factors keeping the tRNA bound fall apart and move, leaving the tRNA to float out of what is left of the secondary active site on one of the monomers. Since the tRNA interacts with the halo region of the monomer whose primary catalytic domain is being used, it is most likely the tRNA would stay attached to the small subunit of that monomer until the monomer floats away, and the flexible linker region twisting and turning until the tRNA anticodon falls out and binds with mRNA codons or gets transferred to the p-site.

- The Dimer Storage Hypothesis:

Figure-(4) showed that by analyzing the dimer interface at the outer primary catalytic domains, only a relatively small distance existed between the two. The alternating hydrophobic and hydrophilic amino acids creating a path for a positive charge to travel and the small holes in the electrostatic potential fields under the adenine binding loops suggest that there is a mechanism in which methyl groups from one monomer could potentially travel to the primary catalytic domain of the monomer, and from there cross the domain bridge and enter the secondary catalytic domain, where it would meet up with its original TrmD monomer again and could be used for G-37 methylation. No definitive proof exists for this pathway, but the structural analysis of 6UEV suggests this is a plausible mechanism. It also provides an explanation to other observed pharmacodynamic effects on the TrmD enzyme. For example, when considering the amount of each reagent the dimer needs to function, it was found that two SAM molecules are taken up, but only one magnesium ion has to be used for each dimer. "Interestingly, the binding stoichiometry of Mg2+ to TrmD, as measured by the enzyme activity as a function of the metal ion, shows that it is one metal ion required for the activity of one dimer"[21]. This suggests that methyl transfer only occurs in one secondary catalytic domain as it is also known that a magnesium cation is necessary to allow the modification of guanine-37. Thus given the structural evidence gathered from model unliganded 6UEV and the work on previous researchers, it is a valid question to consider if this could be a possible mechanism within the dimer.

2. Halo Region/Small Subunit:

Each TrmD monomer displays a flexible halo region above the primary catalytic domain, capping the protein's top portion. This region extends twenty amino acids long and contains aromatic and electronegative amino acids in high density. The C-terminus of the small subunit also has moiety of the halo region and is believed to have an analogous function. Figure-(10) visualizes the B-factor of the halo and exposes the reactive residues, most notably histidine and prolines. Based on the high efficacy polar and aromatic residues have in docking and inducing changes in purine bases, as seen in both active sites of TrmD, the halo region too is capable of attracting purine rings of a compound. When TrmD is dimerized and enzymatically active it is observed that the halo region of the catalytically active enzyme (one that directly methylates G-37) creates part of the secondary active site, enlarging its area, attracting functional groups into the vicinity, and stabilizing guanine-37 and adenine-38 on the tRNA molecule by directing it into the active site pocket via a plethora of electrostatic interactions. The most flexible region of the halo region (colored red in figure-10) is able to stretch and and help dock/stabilize different regions of the tRNA molecule. The tRNA-bound dimer in figure-(10) suggests that the stretching of the halo region results in electrostatic interactions forming with the D-loop of the tRNA. This stabilization reduces the movement of tRNA and the rotational kinetic energy of both structures, consequently allowing the methyltransferase activity to happen uninterrupted and ensuring the right molecule is methylated. Since tRNA is only found to be modified in one enzyme per dimer[26], it is unknown what the function of the halo region is for the antiparallel monomer that does not undergo catalysis.

The small subunit of the non-catalytic enzyme in the dimer complex serves multiple functions both before and after a tRNA molecule has entered the secondary catalytic domain. As described earlier, the C-terminus at the small subunit is chemically similar to the halo region, as it contains histidine and other polar functional groups. This moiety allows the small subunit to act as a 'tRNA net' and move around until it comes into proximity with a tRNA anticodon in the ribosome. At this point, the small subunit electrostatically binds to a part of the tRNA, railing in the macromolecule until it gets near enough to the flexible halo region to begin micromovements into the active site pocket. As the tRNA progresses into the secondary active site, alpha helix α7 changes its position and migrates upwards away from the secondary catalytic domain and instead towards the primary catalytic domain and to the rear frontal region of the protein. In unliganded forms of the protein, the small subunit is already unbound and can migrate without any obstructions, thus creating the 'crab-like' alpha helix arms of the dimer. In models that have bound tRNA and other ligands, this movement still occurs but becomes harder to notice because the electrostatic interactions the ligands induce upon the linker region and small subunit lower the B-factor of the area. This keeps α7 closer to the main subunit of the enzymatically active protein.

3. Linker Region:

The main function of the linker region is to grab and anchor the tRNA anticodon for methylation. When not in the monomer form, we notice that the linker region is nearest to the halo region/secondary active site of the other protein. The linker region "grabs" the tRNA and disallows it from moving away, preventing the tRNA from getting to the primary active site where the methyl donor (S-adenosylmethionine - SAM) would bind to. The small subunit connected to the linker region and the adenine binding loop of the primary catalytic domain then only allow SAM to molecules get in, and initiating the methyl transfer mechanism. The linker region is now intertwined and electrostatically bound to various nucleotides of the tRNA and retreats closer to the enzymatically active protein so the strain on the linker is relieved and the flexible halo region can adopt the 3' end of the tRNA to guide into the secondary active site's ligand pocket. The linker region becomes more electrostatically bound with the opposing monomer as ligands such as S-adenosylmethionine when they enter the primary active site as it changes the conformation in the interior of the site and leads to the protrusion of docking residues such as arginine-166 (with respect to 6UEV). This is not only seen with antiparallel monomers but also when the protein is in a monomeric state as well. It is clear that linker region and small subunit of the inactive enzyme in the dimer complex plays a crucial role in tRNA docking and structural support for the secondary active site, but the linker region of the active enzyme has no conclusive function based on this analysis. Since both enzymes still do get bound with agonists, the linker region still does become stabilized on the other side of the dimer, but whether this linker region acts as a net for the next tRNA anticodon or structurally supports the dimer in the ribosome is unable to be determined without future study. If the former scenario is true, this may suggest that TrmDs switch back and forth between which enzyme in the dimer complex will be active with methylation of tRNA.

4. Insights to Inhibition:

One of the major reasons the TrmD protein subfamily has gained attention in the scientific community over the past years is because of the realization of the potential the enzyme has in becoming a target for the next generation of antibiotics. Not only was the rationale for this experiment to help expand on the intricate details of the bacterial protein synthesis pathway, but also to provide more information on the protein's structure and how it ties in with function, thus exposing critical mechanisms within the TrmD subfamily that could be targeted by future drugs. Not only is the TrmD enzyme used by most eubacteria, but the function it plays forces it to have heavily conserved catalytic domains (aa 88-90 / 130-150), halo regions (aa 40-60), and linker regions (aa 160-180). Figure-(14) clearly shows this conservation across multiple species of both pathogenic and non-pathogenic bacteria in different families. Thus studying a TrmD from any bacterium may provide useful insights that can be applied across multiple other families of bacteria. Future research can be aimed at testing the effect different potential inhibitors have on the enzymatic activity (Km) of the dimer. Ligands to consider for inhibitors should be aimed to obstruct the primary active site, disallow the formation of the dimer by hindering dimer interfaces, or obstruct the primary-secondary active site methyl transfer mechanism by interacting with the Mg^{2+}/G-37/Asp activated complex.

- Inhibition via Invasion of the Primary Catalytic Domain:

The most heavily researched inhibitors for Trm-family enzymes are competitive inhibitors of the primary active site, where the binding of essential methyl donor molecules such as S-adenosylmethionine (SAM) can bind and transfer methyl groups to unmethylated tRNA of an anticodon in the ribosome of an organism. Since the active site has a protruding adenine binding loop built to attract in SAM, creating antagonistic molecules is largely based off mimicking the structure of SAM but not allowing any methyl groups to be transferred off the potential inhibitor. Since docking SAM is largely dependent on adenine loop and trefoil knot interactions with a purine in the substrate, inhibitor molecules should contain regions that are or resemble a purine in terms of aromaticity and hydrogen bonding. In models such as 4YVJ[24], a purine base is attached to a pentose sugar to resemble SAM, but instead of having a small sulfur-containing side chain entering the inner domain, the small chain uses primary amines and a carboxyl group to bury itself deep into the inter-primary/secondary domain space via strong hydrogen bonding. This acts as an almost irreversible inhibitor to the TrmD and deprives the system from free cationic methyl groups. Thus primary active site inhibitors should contain the following three regions:

1. An exposed nitrogenous aromatic ring at one end of the ligand (purines).
2. A pentose sugar with an oxygen heteroatom and similar functional group moiety to ribose.
3. A protruding 5-6 carbon chain consisting of electronegative functional groups (amines/carboxylic acids).

- Inhibition via Obstruction of Monomer-Monomer Interfaces:

The significance of having the TrmD structures assembled in an antiparallel dimer complex has been discussed and explained through findings in the results combined with previous research on the topic. It is clear that without the protein being in a dimer conformation, the catalysis reaction cannot take place as the secondary active site depends on the linker region of the inactive monomer interlocking with the inter-primary/secondary domain space of the catalyzing monomer to form a viable secondary active site that can favor the transfer of cationic methyl groups onto guanine-37 and the docking of large tRNA molecules. By obstructing the monomeric TrmD proteins from joining together, this dimeric complex cannot form, leaving two monomers with incomplete secondary active sites and little stability to dock tRNA molecules successfully. No comprehensive analyses on molecules that would disrupt the dimer itself have been done and thus are another interesting possibility in future research. Without molecular screening and enzymatic assays to test what molecules inhibit the formation of a dimer best, only speculation based on the given structural data can be done. Since the interfaces of the dimers occur between alpha helices, an inhibitor may need to mimic the structure of an alpha helix by being long in length but short in width. The 'zipper' interaction is usually seen between interface regions such as that in figure-(3), which allows a zigzagging of hydrogen bonds or disulfide linkages. However TrmD do not heavily rely on cysteine links and thus focus should be directed on inhibiting the electron networks formed between alpha helices at dimer interfaces via ligand screening/catalysis.

- Inhibition via Obstruction of Secondary Catalytic Domain:

Like the monomer-monomer interface obstruction pathways, nothing can be used to predict possible inhibitors to the secondary catalytic active site methyl transfer mechanism until ligand screening is done and confirmed with enzymatic activity assays and determination of the now inhibited structure. When testing to see the effects of a primary active site inhibitor, the determined protein structure would show an inhibitor in that active site.In the case of a monomer-monomer interface inhibitor, the dimerization of the protein would be obstructed and thus the expected electron density would show that ligand either bound between a partially connected dimer or complete obstruction where no dimer could form to any degree, leaving the proteins as monomers. In the case with a secondary catalytic inhibitor, the region in which it would be found would be in the inter-domain region where it would interact with electronegative amino acids and potentially with Mg^{2+} cations to change their positions, causing the area to become electrically unfavorable for a positive methyl group to transfer through. Little research has been done in this kind of inhibitor, but there is potential as unlike the monomer-monomer interface, the inter-domain region is highly conserved and requires a specific conformation and the presence of a metallic cofactor. Thus future ligand screening for components that could fit into and obstruct the region are plausible when considering ways to inhibit the function of the TrmD.

Bibliography:

1. 27, J. (2018, June 22). Anaplasmosis: Tick-Borne Disease Is Gearing Up and On the Rise. Retrieved from https://universityhealthnews.com/daily/pain/anaplasmosis-gearing-season/

2. Garyu, J. W. A., Choi, K.-seong, Grab, D. J., & Dumler, J. S. (2005, February). Defective phagocytosis in Anaplasma phagocytophilum-infected neutrophils. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC547103/

3. Grab, D. J., Nyarko, E., Barat, N. C., Nikolskaia, O. V., & Dumler, J. S. (2007, November). Anaplasma phagocytophilum-Borrelia burgdorferi coinfection enhances chemokine, cytokine, and matrix metalloprotease expression by human brain microvascular endothelial cells. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2168173/

4. Fighting the Tick-Borne Disease Epidemic. (n.d.). Retrieved from https://renaissance.stonybrookmedicine.edu/medicine/infectious_diseases/tick-borne-diseases-epidemic-symposium

5. Tucker, A. (n.d.). Retrieved from https://collab.its.virginia.edu/access/content/group/f85bed6c-45d2-4b18-b868-6a2353586804/2/Ch25_Tucker_A_Haemophilus_Influenzae_TrmD-_-/Ch25_Tucker_A_Haemophilus_Influenzae_TrmD_tRNA_guanine_N_1_methyltransferase.html

6. wwPDB.org. (n.d.). wwPDB OneDep System. Retrieved September 20, 2019, from http://www.wwpdb.org/deposition/system-information

7. 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).

8. 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

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

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

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

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

13. SFCHECK (CCP4: Supported Program). (n.d.). Retrieved January 24, 2020, from http://www.ccp4.ac.uk/html/sfcheck.html

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

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

16. 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

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

18. 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

19. Introduction to Jmol. (n.d.). Retrieved August 2, 2019, from http://cbm.msoe.edu/includes/modules/jmolTutorial/jmolGettingStarted.html

20. Jannotta, C., Edele, D., Levanti, D., Carson, M., Prucha, G., Caesar, J., … Bolen, R. (2019). Structure of unliganded tRNA (guanine-N1)-methyltransferase found in Anaplasma phagocytophilum. doi: 10.2210/pdb6uev/pdb

21. Hou, Y. M., Matsubara, R., Takase, R., Masuda, I., & Sulkowska, J. I. (2017). TrmD: A Methyl Transferase for tRNA Methylation With m1G37. *The Enzymes*, *41*, 89–115. doi:10.1016/bs.enz.2017.03.003

22. Froese, D. S., Kopec, J., Rembeza, E., Bezerra, G. A., Oberholzer, A. E., Suormala, T., … Yue, W. W. (2018). Structural basis for the regulation of human 5,10-methylenetetrahydrofolate reductase by phosphorylation and S-adenosylmethionine inhibition. *Nature communications*, *9*(1), 2261. doi:10.1038/s41467-018-04735-2

23. Jaroensuk, J., Liew, C., Atichartpongkul, S., Chionh, Y., Wong, Y., Zhong, W., … Fuangthong, M. (2017). Crystal structure and catalytic mechanism of the essential m1G37 tRNA methyltransferase TrmD from Pseudomonas aeruginosa. doi: 10.2210/pdb5wyr/pdb

24. Yoshida, K., Ito, T., & Yokoyama, S. (2015). Crystal Structure of H. influenzae TrmD in complex with sinefungin and tRNA variant (G36C). doi: 10.2210/pdb4yvj/pdb

25. "Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, Marvin 20.1.0, 2019, ChemAxon (http://www.chemaxon.com)"

26. PMC, E. (n.d.). The EMBL-EBI search and sequence analysis tools APIs in 2019. Retrieved from http://europepmc.org/article/MED/30976793

27. Hydrogen bonds (H bonds). (n.d.). Retrieved November 20, 2019, from http://biomodel.uah.es/en/water/hbonds.htm