University of Utah
UNDERGRADUATE RESEARCH JOURNAL

A Structural Analysis of the Lost City MutY Metagenome

Payton H. Utzman, Mary C. Fairbanks and Martin P. Horvath PhD
School of Biological Sciences

Abstract

MutY is an enzyme heavily conserved through all kingdoms of life with the function of repairing oxidative damage to DNA that has the potential to result in cancerous mutations. By studying MutY from the Lost City (LC), a hydrothermal site similar to conditions of the beginning of life, the evolutionary history of this crucial enzyme can be analyzed. The goal of this project was to analyze the LC MutY metagenome collected by Dr. Brazelton in 2018 with a new structural approach compared to the sequence-driven approach previously attempted. Using Phyre2, structures were predicted for all sequences in the MutY metagenome and then compared to the complete crystal structure of MutY from *Geobacillus stearothermophilus* (Gs) in virtual modeling software. The structural analysis proved to be successful as 163 predicted MutY structures were evaluated as *Complete MutY Structures*, meaning these retained each of the MutY-defining chemical motifs. We virtually docked oxidized guanine (OG), the natural ligand of MutY, to these predicted structures and found binding affinities similar to that of docking OG to Gs MutY. This further encourages that MutY is prevalent in the Lost City environment. This technique of predicting structures and performing virtual docking experiments may be useful for biomedical applications, including drug discovery.

Introduction

MutY is an adenine glycosylase that is found in many organisms. Guanine is prone to oxidative damage, resulting in 8-oxo-7,8-dihydroguanine (OG). During replication of DNA, the syn conformer of OG may hydrogen bond with adenine (OG:A). MutY initiates the guanine oxidative (GO) repair pathway by first recognizing the oxidative damage and hydrolyzing the bond between the adenine and the deoxyribose sugar. Succeeding MutY interaction on the OG:A lesion, a series of DNA repair enzymes work to restore a proper G:C base pair before DNA replication (**Figure 1**). In the case that MutY does not initiate repair on an OG:A lesion, future DNA will contain the permanent deleterious T:A mutation. Accumulation of these T:A mutations puts the organism at an increased risk for cancer and neurological diseases.

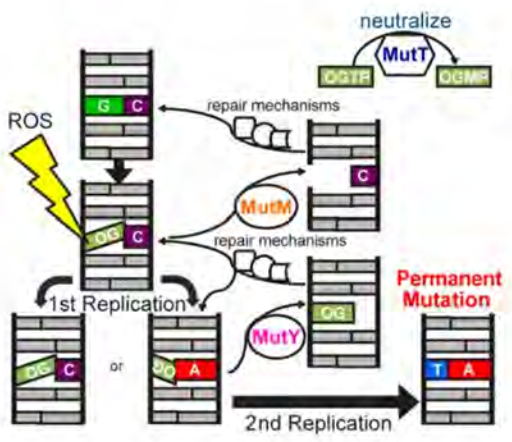


Figure 1. GO Repair Pathway. *Once encountered with a reactive oxygen species (ROS), guanine has the potential to mispair with adenine. MutY, the last line of defense before a permanent mutation, recognizes this error and removes the adenine. Following MutY, a series of future enzymes continue to repair the oxidative damage. Image created by Kyle Knutsen.*

MutY has several chemical characteristics that are highly conserved and define the enzyme. These are the basis for interpreting whether a sequence is likely a functional MutY enzyme or not. From crystal structures of MutY in complex with DNA (Woods et al., 2016), we see in the N-terminal Glu-43 poised for acid-base catalysis, Asp-144 providing electrostatic stabilization of the transition state, and both Asp-144 and Tyr-126 contributing to a hydrogen-bond network that positions the DNA and Glu-43. Peyton Russelburg has determined MutY recognizes OG primarily by a beta loop composed of Phe-307, Ser-308, and His-309, known as the FSH Loop, in the C terminus. Together these six amino acid residues make up the *chemical motifs* of MutY.

Determining the structures of the enzyme and comparing these between different organisms will give insight into what mechanics of the enzyme are conserved and therefore crucial. The Lost City MutY Project aims to take this idea one step further by determining structures of MutY from lineages that diverged billions of years ago.

The Lost City is a hydrothermal field that is an extreme environment in the Northern Atlantic Ocean. The Lost City is of interest to many researchers because it represents a possible origin of life (University of Washington, 2003). Dr. William Brazelton of the University of Utah studies metagenomics of the Lost City. In 2008 Dr. Brazelton completed an expedition to the site to retrieve organic samples (Amador, Bandfield, Brazelton, Kelley, 2017). After the expedition, DNA was extracted from these organic samples and sequenced. All sequences were then BLAST-ed (Basic Local Alignment Tool). Sequences that showed a high level of resemblance to known MutY sequences (known as *MutY Hits*) were generously given to the Horvath Lab, establishing the Lost City MutY project.

In September of 2018, Dr. Brazelton completed a second expedition to the Lost City to collect samples of the organisms present (University of Utah Biology, 2018). Once again, DNA purification was performed, and sequences were BLAST-ed. Sequences that had a high amino acid resemblance to known MutY sequences were sent to Dr. Horvath. In 2019, these 70 MutY hits were analyzed for their amino acid composition, in search of the sequences most similar to known MutY sequences. Eventually, four sequences with percent identity over 75% to known MutY enzymes were chosen as the 2nd Generation of Lost City MutY.

We analyzed the Lost City metagenome Dr. Brazelton collected in 2018 with a newly developed, structural approach. It was our hypothesis that some functional MutY enzymes from this metagenome were overlooked due to a different amino acid composition because amino acid composition is far less conserved than protein structure. In June, Dr. Brazelton shared 649 amino acid sequences from the Lost City, of which 90% were not included in the first analysis. Instead of relying on amino acid composition relative to known sequences, structures were predicted from each sequence using the protein homology program Phyre2. This produced 649 PDB files describing the positions of atoms in the predicted MutY structure. Each structure was then superimposed with the complete MutY structure from *Geobacillus stearothermophilus* (Gs) MutY (PDB 6U7T) in the virtual modeling program UCSF Chimera. Through Chimera, a script was written to easily compare chemical motifs and structural geometry. The structural analysis clearly separated candidates complete with MutY chemical motifs from those missing the residues. This newly developed structural analysis proved to be a superior alternative as 163 sequences were found to have a very similar structure to the functional Gs MutY enzyme, encouraging that these are functional enzymes themselves. Furthermore, virtual docking

experiments demonstrate these predicted structures are capable of binding with the MutY ligand OG.

Methods

To accurately determine whether a structure could be a functional MutY enzyme, it was first necessary to define the criteria for what a functional MutY enzyme looks like. Currently, 6U7T is the only complete crystal structure of MutY. 6U7T is Gs MutY in complex with DNA. Using the interactive visualization program UCSF Chimera, the PDB file 6U7T was displayed and annotated to easily see where the necessary chemical motifs are positioned. Glu-43, Tyr-126, Asp-144, Phe-307, Ser-308, His-309, were highlighted by creating surfaces within Chimera (**Figure 2**). This gave insight into how the residues interact with the substrate, information necessary for examining a predicted structure.

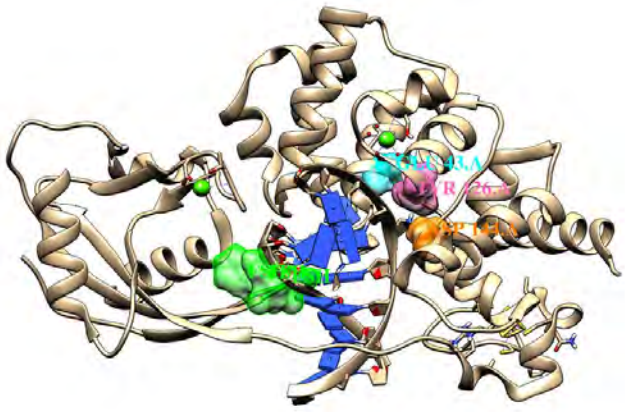


Figure 2. 6U7T. *This image was created in UCSF Chimera with the intent to visualize where the chemical motifs are positioned. Surfaces were colored for each motif: Glu-43 shown in cyan, Tyr-126 in hot pink, Asp-144 in orange, and the FSH Loop is shown in green. These motifs served as the basis for determining whether a predicted structure may be a functional MutY enzyme.*

With the criteria established, the next step was to create structures for each amino acid sequence from the Lost City. Protein Homology/analogY Recognition Engine V 2.0 (Phyre2) was used to create a predicted structure for each of the amino acid sequences. Using the "Batch Job" function, approximately 80 amino acid sequences at a time were uploaded to the server. Phyre2 created a Hidden Markov Model (HMM) for each sequence and aligned it with the HMM database of known proteins, including MutY structures, searching for the closest match to predict the structure. Only the structure with the highest confidence for each sequence was downloaded. The output resulted in a set of PDB files that could be viewed in Chimera to analyze the predicted MutY structures.

Each predicted sequence had to be compared with 6U7T individually, a task that involved superimposing and highlighting the chemical motifs. To streamline this job, a script was created to run within Chimera to efficiently determine whether the predicted structure contained all necessary motifs. The script was written in Sublime Text, a code editing program, and saved within the same directory as the predicted-PDB files downloaded from Phyre2. The script opened a specified predicted-PDB file, opened 6U7T, matched the former structure to the latter, colored the N and C terminus accordingly, and highlighted residues within 1.5 Angstroms of each chemical motif (**Figure 3**).

```
open 6u7t
#open each specified PDB file
open c_000000031207_6.pdb
#color structures so that N terminus of p-PDB is plum, C terminus is cornflower blue
mm #0 #1
color green,r #1
color plum,r #0:0-225.A
color cornflower blue,r #0:235-360.A
background solid dimgrey
~display #0
#Now with the structures mm and colored correctly, it is time to mark motifs.
sel #0:43,126,144,146.A@cb
namesel chemmotif6u7t
select chemmotif6u7t zr<1.5 & ~#0
namesel chemmotifworkset
display chemmotifworkset
rlabel chemmotifworkset
color red chemmotifworkset
sel #0:307,308,309.A@cb
namesel fshmotif6u7t
select fshmotif6u7t zr<1.5 & ~#0
namesel fshmotifworkset
display fshmotifworkset
rlabel fshmotifworkset
```

Figure 3. Script for Comparing MutY Structures. *This script was read in Chimera to quickly superimpose a predicted structure with 6U7T and mark the residues located in the positions of chemical motifs. The predicted structure was colored in a standout green, making it easy to determine whether the structure was complete. The script was adjusted for each individual predicted-PDB file.*

The first step was setting Chimera's working directory to that of the script and predicted-PDB files using the cd command. The script was saved as script.txt, so the next command was "read script". This created a visualization as shown in **Figure 4.** The structure was then binned into one of four groups: *Incomplete MutY Structure*, *Complete MutY-N Structure*, *Complete MutY-C Structure*, or *Complete MutY Structure*. The next predicted-PDB file was then copied into the script, saved, and the process was repeated.

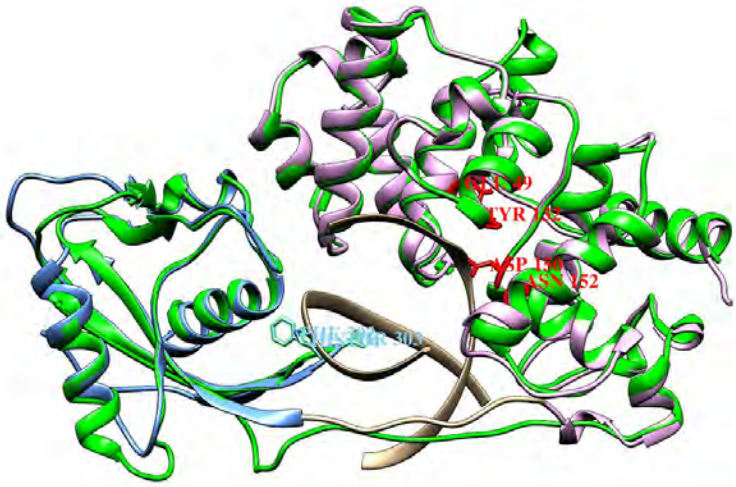


Figure 4. Predicted Structure Superimposed with 6U7T. *Pictured is an image generated by the script read in Chimera. Quickly observed is the complete coverage of the predicted structure shown in green. Residues within 1.5 Å of the chemical motifs of 6U7T were highlighted and labeled. This structure would be binned into the "Complete MutY Structure" directory since it contained all six chemical motifs of a MutY enzyme. This process was carried out for the 649 predicted structures.*

Interested specifically in the Complete MutY Structures, a phylogenetic tree was created to visualize the relatedness between the Lost City sequences and known MutY sequences such as *Halolamina sediminis* and *Homo sapiens* (Human) MutY. This was done by creating a txt. file that contained the amino acid sequences of the Lost City sequences that had been declared Complete MutY Structures along with *Halolamina sediminis* MutY and Human MutY sequences. This file was uploaded to PROMALS3d, where the PDB files for *Escherichia coli* (1WEI), Gs (6U7T), and Human MutY (3N5N) were added to include known MutY structures as

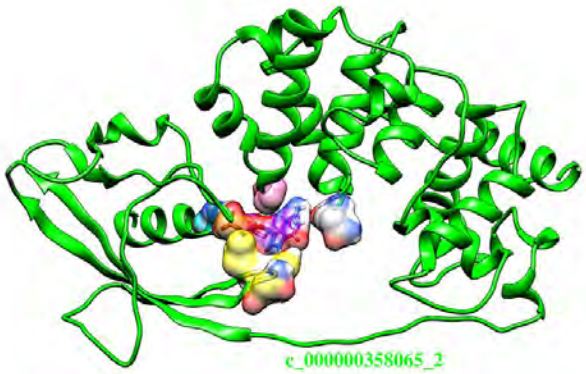
well. The output was saved in CLUSTAL format, where it was uploaded to the ClustalOmega server. Several settings were fine-tuned to prevent Clustal Omega from realigning the sequences: Dealign Input Sequences (no), Order (input). The phylogenetic tree data was then copied and pasted into the Interactive Tree of Life server (iTOL) to create the tree.

Lastly, docking experiments were performed using AutoDock Vina (Vina) and AutoDockTools (ADT). The oxidized guanine (OG) ligand was created by opening the PDB file 6U7T in Chimera, selecting the OG molecule, and then saving only the selected atoms as a new PDB file. The ligand was then prepared in ADT by Ligand<open<choose "adenine.pdb"; Ligand<Torsion Tree<Detect Root; Ligand<Torsion Tree<Set Number of Torsions=0; and then Ligand<Output<Save as PDBQT. Each receptor was prepared by File<Read Molecule; Edit<Hydrogens<Add<Polar Only; Edit<Atoms<Assign AD4 type; Edit<Charges<Compute Gastieger; and then saving as a PDBQT file. ADT was used to create the Grid Box, the area defining the search space, by Grid< Grid box, and adjusting the size and placement of the box.

With the Grid Box, receptor, and ligand prepared, a script was written to run in Vina. To start Vina, Command Prompt was opened, and the directory was set to that which contained vina.exe, the ligand, and the receptor. Once in the directory, the Vina command complete with parameter settings was copied and pasted in the Command Prompt, and Vina was run (**Figure 5**).

Figure 5. Running AutoDock Vina. *Docking experiments performed with Vina are initiated through a Command Prompt window (Vina lacks a user interface). First, the working directory had to be set to the directory that contained the vina.exe program, ligand, and receptor accomplished with the cd command. The vina.exe command included information needed by Vina to perform the docking experiment.*

Vina produced a PDBQT file that contained multiple poses of the ligand complexed with the receptor. A log file was also generated containing binding affinities calculated by Vina. The output PDBQT file was then viewed in Chimera, along with the receptor PDBQT file (both were opened as PDB files). In most cases, the output file contained nine ligands individually positioned. It was most useful to only view one at a time with the receptor (**Figure 6**). The receptor and ligand were enhanced visually by creating surfaces.




```
mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -8.3      0.000      0.000
   2        -8.0      1.909      2.135
   3        -7.8     14.102     16.219
   4        -7.8     14.031     16.160
   5        -7.6     10.750     12.529
   6        -7.6      2.718      3.660
   7        -7.6      2.844      4.659
   8        -7.2      4.516      6.120
   9        -7.0      5.829      8.362
Writing output ... done.
```

Figure 6. OG Nucleotide Docking Experiment. *Shown on the left is a complete, predicted structure in green. This structure was prepared as a receptor and docked with MutY's natural ligand, an OG nucleotide. Surfaces were then created to see complementary geometry between the receptor and ligand. Shown on the right is the corresponding log file containing the binding affinities from this docking experiment.*

Results

All 649 sequences were converted into predicted structures through Phyre2. Interestingly, the top hit (structure with the highest confidence) for all 649 structures was a MutY enzyme, mostly commonly 1RRQ from Gs MutY. These predicted structures were compared with the structure of Gs MutY (PDB 6U7T) with the help of a script to make this process more efficient (see methods). Each structure was then cataloged as one of four designations as shown in **Table 1.**

Designation	Number Discovered
Incomplete MutY Structure (IMS)	325
Complete MutY-N Structure (CM-NS)	51
Complete MutY-C Structure (CM-CS)	22
Complete MutY Structure (CMS)	163

Table 1. Designation of Predicted Structures. *Structures were cataloged based on completeness and chemical motifs present.*

Every predicted structure generally had the same structure as 6U7T, the differentiating factor was the completeness of the structures. **Figure 7** shows one predicted structure that was only partially complete but still showed a very similar structure to that of 6U7T. Approximately 325 structures were determined to be *Incomplete MutY Structures* (IMS).

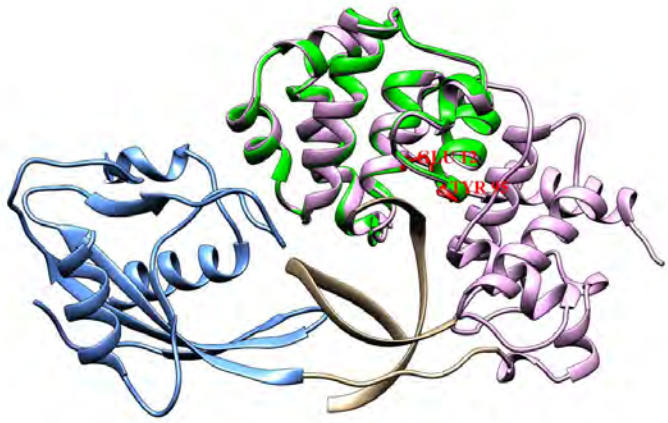


Figure 7. Incomplete MutY Structure. *Shown in green is the predicted structure for sequence c_000000056930_1. Shown in orchid (N-terminus) and purple (C-terminus) is the structure of 6U7T. It is evident that while c_000000056930_1 is very incomplete, the structure still highly resembles 6U7T. Also, worth noting is that this partial structure does contain the motifs covered in its respective portion.*

Many structures were found to contain a complete N or C terminus of MutY while lacking a portion of the other terminus. Designated as *Complete MutY-N Structures* (CM-NS), 51 structures were found to be a complete structure of the N terminus. Additionally, 22 structures were found to be a complete structure of the C terminus of MutY, known as *Complete MutY-C Structures* (CM-CS). These structures maintained the chemical motifs included in their terminus.

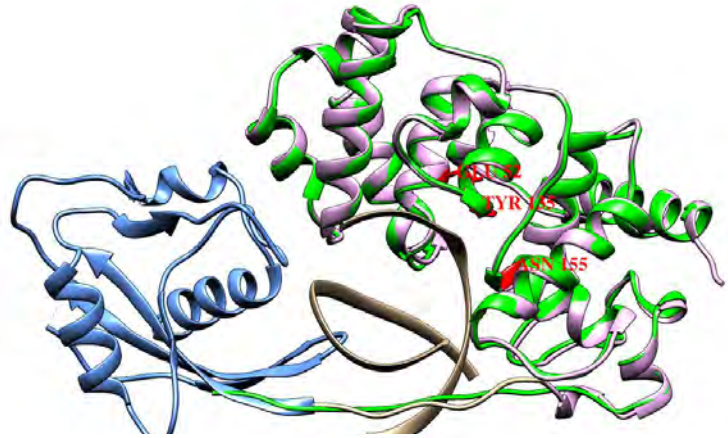


Figure 8. Complete MutY-N Structure. *Shown in green is the predicted structure for c_000000438930_2 superimposed with 6U7T. It is clear that the predicted structure completely covers the N terminus, but lacks a C terminus completely. All three chemical motifs found in the N domain of MutY are conserved. The linker region is also complete. Many other Absolute CM-NS had partial coverage of the C terminus as well.*

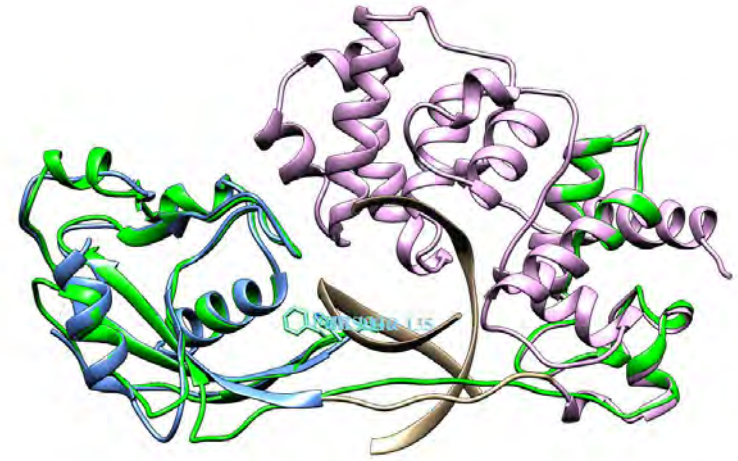


Figure 9. Complete MutY-C Structure. *Shown in green is the predicted structure for c_000000649452_1 superimposed with 6U7T. This predicted structure, while slightly different, completely covers the C terminus of 6U7T (blue ribbon) and possesses the MutY defining FSH Loop for OG recognition. The structure also contains a small portion of the N terminus (orchid)*

Out of the 649 predicted structures, 163 were found to completely cover 6U7T's structure and contain the six motifs. These structures were labeled as *Complete MutY Structures* (CMS). Of the candidates analyzed, ~25% turned out to fall into this CMS category, meaning there is a high probability that functional MutY is prevalent at the Lost City. The predicted structures were remarkably similar to that of Gs MutY despite diverging billions of years ago. Shown in **Figure 10** are a few examples of the Complete Structures that were discovered using this structural approach.

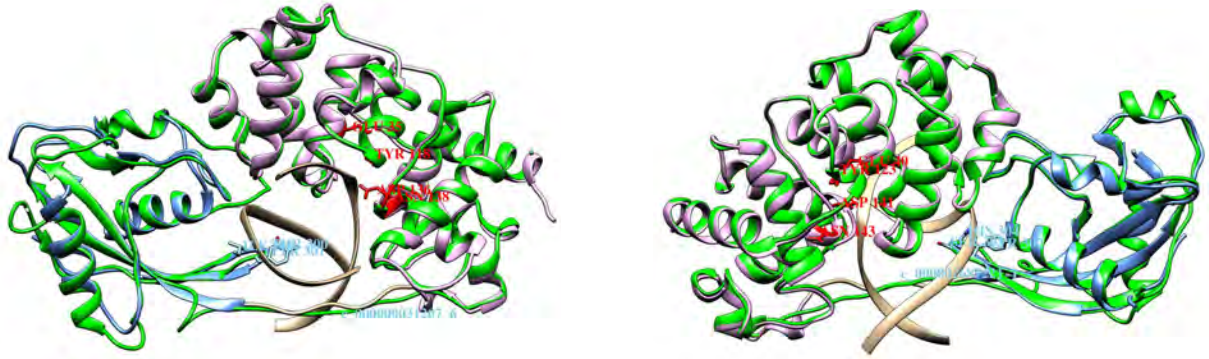


Figure 10. Complete MutY Structures. *Shown on the left is the structure of c_000000031207_6 and on the right is c_000002687221_1. Both structures completely resemble that of 6U7T and contain all chemical motifs. Although there is slight variation in the orientation of alpha helices, this is expected as MutY from the Lost City and Gs MutY have been isolated for billions of years, leading to different evolutions of the enzyme.*

A phylogenetic tree containing the CMS and known MutY sequences and structures was generated to visualize relationships through evolution. Naturally, the longer branch length corresponds with greater sequence dissimilarity, indicating a longer time since these MutYs diverged during evolution. The tree was split into 11 sections, based on relatedness (**Figure 11**).

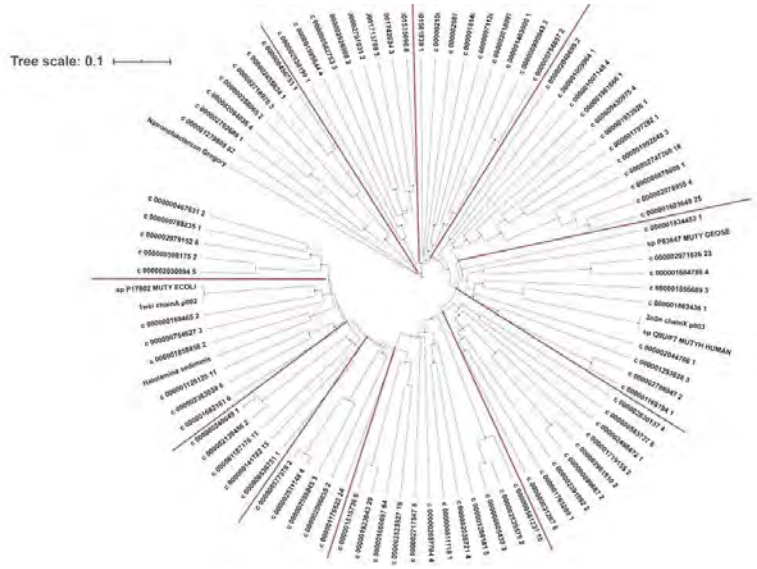


Figure 11. Phylogenetic Tree. *All CMS's amino acid sequences, multiple known MutY sequences, Gs MutY structure, and Human MutY structure were aligned in PROMALS3D before a tree was generated. The tree was split into 11 different sections, loosely based on branch length from one another.*

To further determine whether these structures may be functional MutY enzymes, docking experiments with MutY's natural substrate, an OG nucleotide, were carried out using Vina. OG is a damaged form of the normal G nucleotide; MutY functions to prevent mutations by finding this OG nucleotide in DNA. While testing all 163 predicted CMS would be beneficial, it was more feasible to randomly choose one CMS from each section of the phylogenetic tree. Singularly, each structure was prepared in ADT with the grid box generously covering the active site of the enzyme. The OG nucleotide was isolated from the 6U7T PDB file and prepared for the docking experiments. Each docking experiment produced nine different positions where the

ligand was most likely to bind in the form of a PDBQT file (**Figure 12**). In all cases, one of those nine positions was very similar to that of the OG in the crystal structure of Gs MutY with DNA (6U7T).

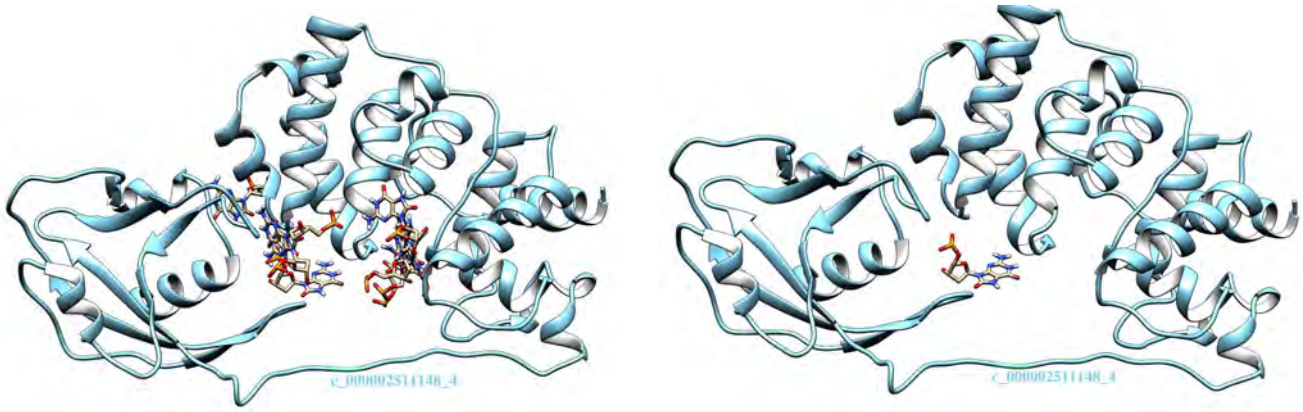


Figure 12. Docking Experiment. *Predicted structure c_000002511148_4 was docked with the OG nucleotide MutY binds to. Vina produced a PDBQT file that contained nine possible positions for the ligand to bind to the receptor. This is shown in the image on the left. In all 11 docking experiments, there were at least one of the nine ligand positions in which is highly resembled the OG position within 6U7T's structure, as shown in the image on the right.*

Vina also generated a log file that contained the calculated binding affinity energy for each ligand position to the receptor. Negative binding affinities signify the ligand is likely to bind. All binding affinities between the OG ligand and the 11 receptors that were tested were negative, many over -7.0 kcal/mol. This means it is highly likely the ligand would naturally bind to these predicted structures from the Lost City Metagenome. **Figure 13** displays the affinities produced from the 11 docking experiments in the form of a histogram. The results of all 11 docking experiments strongly indicate there is an attraction between the predicted structures and the biologically relevant ligand for MutY (OG), implying that these Lost City proteins function similarly to MutY.

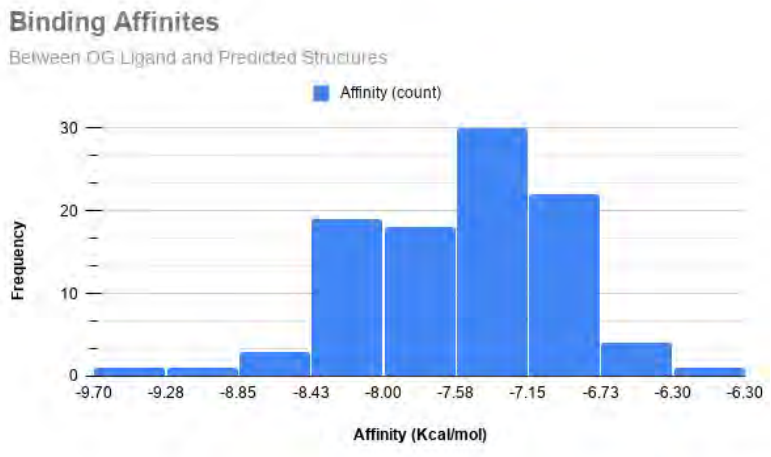


Figure 13. Histogram of Binding Affinities. *Presented are the binding affinities calculated by Vina for all 11 docking experiments (nine outcomes per experiment). All of the affinities were significantly negative and consistent. This means it is likely the OG ligand would bind to any of these predicted structures in the active site.*

Discussion

Reanalyzing the LC MutY Metagenome with a structural approach proved to be successful as 163 sequences were determined to likely be functional MutY enzymes. The predicted structures for these sequences completely matched that of Gs MutY's known structure and contained all chemical motifs that define the enzyme. Most of these sequences were overlooked in the analysis performed in 2019 that was based on amino acid composition. This is likely because many amino acids share similar properties, and after much time, mutations can arise throughout the sequences that are neutral. These resulting divergences would then produce the same functional MutY enzyme with a similar structure but with a dissimilar amino acid sequence. This explanation describes why Lost City sequences that had a percent identity less than 50% to that of Gs MutY maintained the same polypeptide structure. The Lost City is also a very isolated site, being located at the ocean floor, so diversification is expected, compared to sequences from life above water such as Gs MutY, which is found in soil and hot springs.

Docking experiments performed in Vina further add confidence that these sequences are functional MutY enzymes. The experiments between MutY's true ligand, OG, and the predicted structures resulted in binding affinities reasonable for a functional MutY enzyme. Perhaps the most persuasive result is the location Vina predicted the OG nucleotide would bind. For all 11 experiments, Vina docked the OG nucleotide to a very similar position to the OG nucleotide in Gs MutY's complete structure with the full DNA double helix (6U7T). Even without the full structure of the DNA helix, there was a natural affinity for the OG ligand to bind to the correct location in these predicted structures. This was a surprising pattern because the absence of DNA would create a large volume for the ligand to add at other positions, yet the results showed the ligand in the correct position. This strongly encourages that these sequences from the Lost City are MutY enzymes and possess the necessary properties if these are to operate as MutY does to surveil the genome and intercept OG:A lesions.

There were consistent variations between the predicted structures from the Lost City and 6U7T's structure that are of interest. With 51 sequences coding only for the N terminus, around 8% of the candidates investigated, there must be a reason for this consistent incompletion. Perhaps organisms from the hydrothermal vent have lost the FSH Loop due to evolution and employ alternative residues located in the N terminus for recognition of the OG substrate. However, this would contradict Sonia Seghal's findings in 2017 on the 1st Generation of LC MutY. The 1st Generation was notoriously incomplete, lacking a C terminus. Seghal tested for the function of the sequences by use of a mutation suppression assay in 2017. The results showed little or no activity. (Sehgal, 2019). Possibly then, these are not MutY enzymes at all, and instead an unknown enzyme that evolved from an original MutY sequence. This could explain the lack of OG binding to the polypeptide, observed by Sehgal.

Similarly, many CMS showed consistent variations. As shown in **Figure 14**, many predicted structures contained an alpha helix in the C terminus that was slightly different compared to that of 6U7T, with some structures completely lacking it altogether. With the function of this alpha helix unknown, it may be interesting to perform assays with sequences of this structure to determine whether the specificity of the enzyme is affected by this variation. Such information would provide further insight into how this crucial enzyme functions.

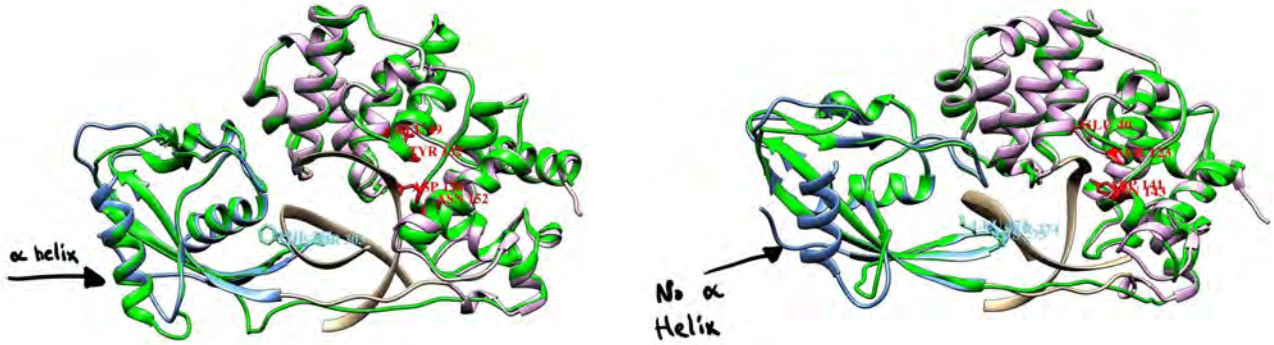


Figure 14. Alpha Helix Variation in the C terminus. *Many predicted structures that were deemed CMS had variations in a specific alpha helix located in the C terminus. Shown in green are the predicted structures, and in purple is 6U7T. Numerous sequences lacked this alpha helix altogether, as shown in the image on the right. The function of this alpha helix is unknown.*

Many predicted structures also had a slightly different curvature in the critical FSH Loop, as shown in **Figure 15**. As earlier stated, the FSH Loop is responsible for the recognition of the OG ligand. The slightly different curvature of this loop may affect the selectivity of the enzyme, something that would need to be tested once back in the lab.

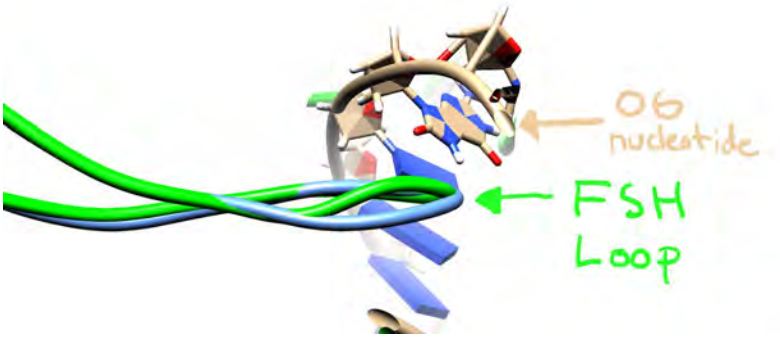


Figure 15. FSH Loop Curvature. *Shown in green is a predicted structure; shown in purple is 6U7T. The curvature of the FSH Loop of the predicted structure differs slightly from that of 6U7T. This subtle difference may affect the selectivity of the enzyme.*

It is remarkable how similar the Lost City predicted structures are to Gs MutY. Despite being remotely isolated from one another for millions, perhaps billions, of years, the structures are incredibly homogenous. Conservation of the chemical motifs in the Lost City structures further confirms the importance of these residues for MutY enzymatic activity. The evolution of MutY seems to be minimal, only seen in the amino acid composition and not polypeptide structure when looking at the predicted structures. MutY's substrate, a DNA helix containing an OG:A pair, may be the reason for this high conservation of structure. DNA's structure is large with respect to that of the MutY enzyme and inserted within the active site. The structure of MutY is cast around the DNA. Because DNA's structure is highly maintained throughout all life on Earth, major structural changes of MutY due to evolution may not be possible because of DNA's position within the enzyme. DNA can then be considered an evolutionary constraint of MutY; any structure that diverges too far would fail to bind with DNA and that lineage would lead to extinction.

It is important to note while the predicted structures of the sequences have been generated and analyzed, they are only that, predicted. Consequently, the predicted structures are expected to show high levels of compatibility to known structures as Phyre2 bases its prediction off of those very known structures. Therefore, while the predicted structures show promise, further research

will need to be conducted. To confirm whether these Complete MutY Structures from the Lost City MutY Metagenome are functional, a few sequences will be chosen and ordered to perform a mutation suppression assay in vivo soon in the lab. Sequences from different sections of the phylogenetic tree will be chosen; these sequences will be a part of the *3rd Generation of Lost City MutY*. MutYs from the Lost City that display high enzymatic activity on the OG substrate will be purified and crystalized to fully determine the structure.

Performing crystallography and determining a new structure of MutY would be extremely valuable as currently, Gs MutY is the only complete structure of MutY. The structural analysis developed can be used to find MutY's with higher diversities and lead to a broader understanding of the crucial enzyme. Furthermore, by comparing a Lost City structure to that of Gs or Human MutY in the future, further knowledge of necessary residues can be found. A more complete understanding of the mechanisms MutY employs to recognize the substrate and perform its function would be beneficial for enzyme engineering or drug creation to prevent cancer and save lives.

Acknowledgments
Special thanks to Dr. William Brazelton for sharing the metagenomic data used in this study. Thank you, Peyton Russelberg, for training on UCSF Chimera to generate such detailed visuals. We are especially grateful to conduct research in an environment with team members that are passionate, hard-working, and caring.

Office of
UNDERGRADUATE RESEARCH
THE UNIVERSITY OF UTAH

This work was supported by funding from the Undergraduate Research Opportunities Program at the University of Utah awarded to Payton Utzman and Mary Fairbanks.

References

Elena S. Amador, Joshua L. Bandfield, William J. Brazelton, and Deborah Kelley.Astrobiology.Nov 2017.1138-1160.http://doi.org/10.1089/ast.2016.1606

O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, Journal of Computational Chemistry 31 (2010) 455-461 DOI 10.1002/jcc.21334

Russelburg, Demir, Knutsen, Sehgal, Cao, Horvath, David, & O'Shea Murry. (submitted, under review). Structural basis for finding OG lesions and avoiding undamaged G by the DNA glycosylase MutY. University of Utah. doi: 10.2210/pdb6Q0C/pdb

Sehgal, S. L. (2019, May 5). University of Utah. Retrieved from https://docs.google.com/document/u/0/d/1ZqSV2kB6X4F-SNtLZ4Lr3RSn5IZk6ABGv37AVZmp8/mobilebasic

The Phyre2 web portal for protein modeling, prediction and analysis Kelley LA et al. Nature Protocols 10, 845-858 (2015)

UCSF Chimera--a visualization system for exploratory research and analysis. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. J Comput Chem. 2004 Oct;25(13):1605-12.

University of Utah. (2018, September). Return to the Lost City 2018. Retrieved October 28, 2019, from https://lostcity.biology.utah.edu/research/.

University of Washington, U. (2003). About The Lost City Hydrothermal field. Retrieved from http://www.lostcity.washington.edu/story/About

Woods, R. D., O'Shea, V. L., Chu, A., Cao, S., Richards, J. L., Horvath, M. P., & David, S. S. (2016). Structure and stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to retaining glycosidases. Nucleic Acids Research, 44(2), 801-810. https://doi.org/10.1093/nar/gkv1469\