Robin Jones

RBIF 101 Final Project

Base excision repair is a DNA repair pathway that identifies and removes aberrant DNA bases. It is present in a wide variety of organisms from bacteria to humans, though may operate differently in higher eukaryotes. One enzyme family involved in base excision repair is the glycosylases. These enzymes can recognize a large assortment of DNA base damaging modifications including alkylation, mismatch, oxidization and deamination (Wyatt et al., 1999). This system is crucial for inhibiting the accumulation of genetic damage and avoiding cell death.

Bacterial 3-Methyladenine glycosylase 1 (3-MPG-I) is a glycosylase specifically involved in the recognition and removal of 3-methyladenine (3-meA) from bacterial DNA (Wyatt et al., 1999). Once a 3-meA lesion is identified in bacterial DNA, MPG binds to the 3-meA lesion and cleaves the glycosidic bond between the damaged base (3-meA) and the sugar backbone of the DNA (Wyatt et al., 1999). This results in the removal of the 3-meA base, creating a site that no longer has a base. This is called an abasic site. Once the abasic site is created, several possible repair mechanisms are possible, but all result in enzymatic replacement of the damaged base with an appropriate new base at the abasic site (Wyatt et al., 1999).

The closely related 3-Methyladenine glycosylase II enzyme is also found in bacteria and human and is able to excise 3-meA lesions as well. However, MPG-II is able to recognize and remove multiple other DNA base lesions whereas MPG-I is specific to 3-meA lesions (Metz et al., 2007). This makes MPG-I much more targeted to repairing just 3-meA lesions.

Because glycosylases such as MPG-I repair DNA and prevent genetic instability, they are ultimately able to promote cell survival and appropriate replication of bacterial cells (Metz et al., 2007). This makes them a good target for antibiotic therapy; pharmacologic molecules could be designed to bind the glycosylase active site and indirectly prevent cell survival and replication by inhibiting DNA repair (Mechetin et al., 2020). For example, the methylation of adenine inhibits proper bacterial cell replication, and removal of this adenine modification by 3-MPG-I helps to restore normal replication (Zhu et al., 2012). By targeting this enzyme with a small molecule inhibitor, it could be possible to inhibit bacterial replication and stop progression of further bacterial infection.

There is a dearth of information about the effects of knock-out or knock-down of 3-MPG-I in bacterial cells and bacterial survival. However, two recent studies point to this enzyme as a potential drug target. Nayak et al., 2019 undertook a comprehensive comparative bioinformatic genomics screening analysis of Streptococcus Pneumoniae in an attempt to identify new antibiotic targets. The gene expression patterns were compared across other species and the genes most essential to bacterial survival identified. Multiple new targets were identified and among them, 3-methyladenine glycosylase I was suggested as a potential multi-species drug target since the protein was identified in multiple gram positive and negative bacteria. Grin et al., 2021 have recently developed an uracil-DNA glycosylase inhibitor that specifically targets human cancer cells. This inhibitor is designed to stop uracil glycosylase from repairing DNA damage caused by methotrexate in cancer cells. Though not targeted to the bacterial enzyme, it is still important evidence toward the druggability of glycosylases and the potential for their inhibition to be effective in halting DNA repair. The authors also suggest that uracil glycosylases could also be potential antibiotic targets.

The purpose of this study is to investigate the suitability of 3-Methyladenine glycosylase I for potential pharmacological inhibition. A comprehensive analysis of the protein based on both sequence and structure is undertaken. Both sequence and structural alignments are provided and discussed, as well as an analysis of any protein domains, the protein binding site and overall druggability of the protein. Finally, an analysis of the potential for drug specificity is given.

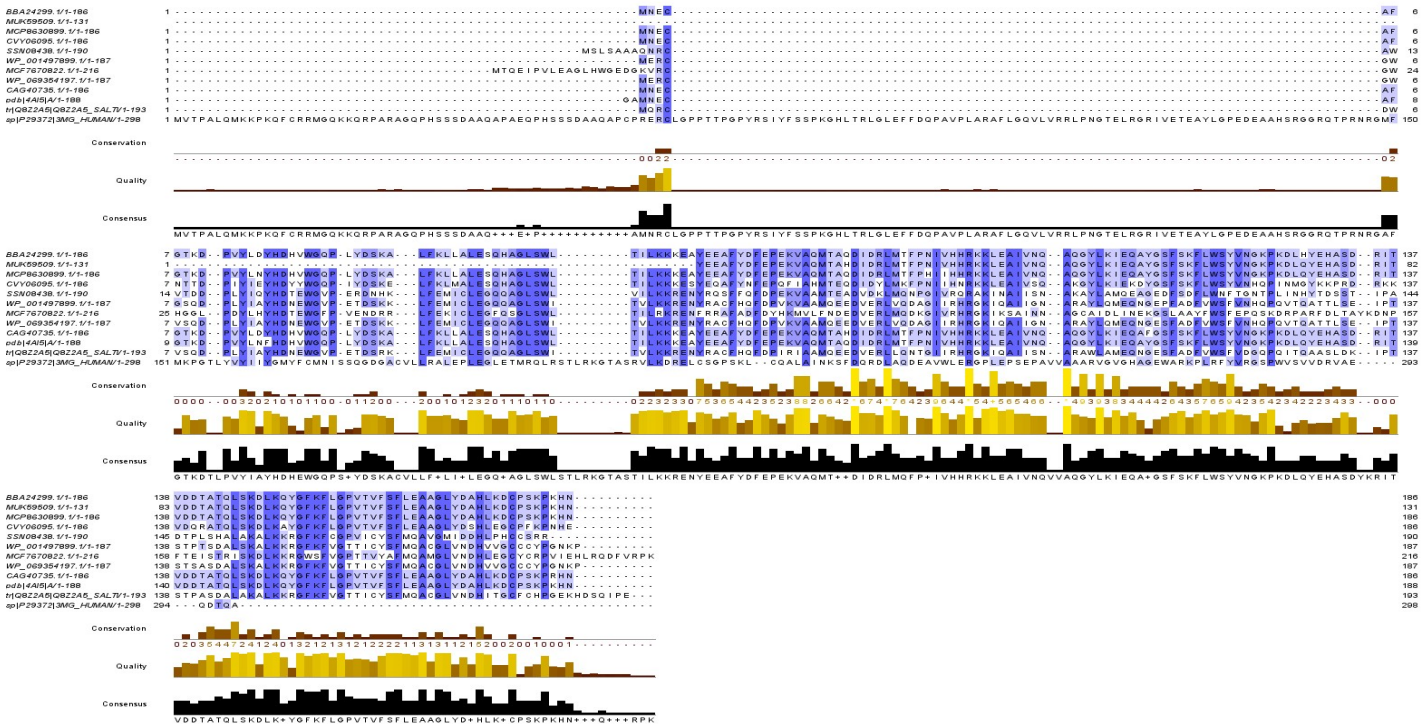
Evolutionary Analysis of 3-MPG-I

3-Methyladenine glycosylase I is purported to be ubiquitous in bacteria (Zhu et al., 2012). A widely recognized human protein known as Alkyladenine DNA Glycosylase (AAG) may be a homologous protein to the bacterial MPG. Both proteins serve the same function but are reported to share little sequence or structural homology (Drohat et al., 2002).

In order to study the homology of this enzyme across species, an evolutionary analysis was performed which included the MPG-I enzyme from : escherichia coli, klebsiella pneumoniae, acinetobacter baumannii, pseudomonas aeruginosa, staphylococcus aureus, MRSA, MSSA, streptococcus pneumoniae, bacillus subtilis, enterococcus faecalis, salmonella enterica typhi and human AAG enzyme.

A BLAST search against the reference 3-MPG-I amino acid sequence from staphylococcus aureus was run utilizing the default settings and the BLOSUM-62 matrix as it was suspected this protein could have multiple domains. A human homolog was not returned as part of the blast search, but an appropriate sequence was obtained from Uniprot to include in the study.

The image below is the protein amino acid sequence alignment of MPG-I from all included bacterial species and one human AAG enzyme. The MSA used was the Clustal Omega alignment, and the output is colored according to percent identity at each amino acid. There is quite a bit of similarity between the bacterial enzymes (first 11 entries) and not much similarity between the human AAG and bacterial 3-MPG.



More specifically, pairwise alignments between S. Aureus 3-MPG-I and the other species resulted in more exact quantifications of percent identity and similarity. These are listed below.

List of bacterial species and human AAG enzyme with percent identity and similarity to S. Aureus 3-Methyladenine Glycosylase:

- Pseudomonas Aeruginosa: 98% shared identity and 98% similarity with S. Aureus
- Acinetobacter Baumannii: 98% shared identity and 99% similarity with S. Aureus
- Streptococcus Pneumoniae: 75% shared identity and 87% similarity with S. Aureus
- Klebsiella Pneumoniae: 46% shared identity and 64% similarity with S. Aureus
- E. Coli: 43% shared identity and 64% similarity with S. Aureus
- Bacillus Subtilis: 36% shared identity and 57% similarity with S. Aureus
- Enterococcus Faecalis: 43% shared identity and 64% similarity with S. Aureus
- MRSA-252: 98% shared identity and 98% similarity with S. Aureus

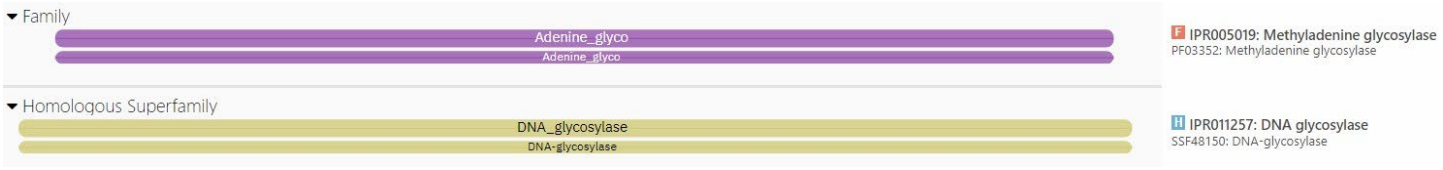
- Salmonella Enterica Typhi: 41% shared identity and 66% similarity with S. Aureus
- Homo Sapiens: 16% shared identity with S. Aureus

From this list, it is apparent that much of the bacterial enzyme may be conserved across species, while the human AAG is quite different from the bacterial form. In addition, because BLAST could not align the Homo Sapiens and S. Aureus proteins, a percent similarity was not able to be calculated, though it looks to be slightly higher than the shared identity based on the color coding of the alignment.

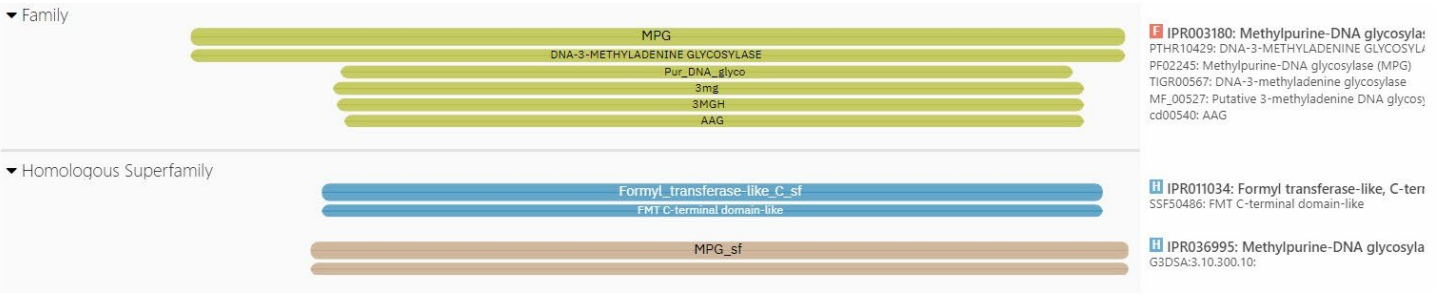
The lack of high shared identity and similarity between human and bacterial 3-MPG-I enzymes could potentially make it easier to design a drug that is specific to the bacterial form of the enzyme only. In addition, there is a range of moderate to high sequence identity and similarity across bacterial species, indicating that this target could be broad-spectrum.

Comparative Structural Analysis

Comparative structural analysis was carried out with Staphylococcus Aureus, E. Coli, and Salmonella Enterica Typhi and Streptococcus Pneumoniae bacterial enzymes as well as the human AAG enzyme. Interpro was first utilized to identify protein classifications based on sequence. All bacterial 3-MPG-I were classified as shown below.



The human AAG enzyme was classified differently than the bacterial enzymes, and this is shown below.



CATH was then searched for structural relationships between proteins. Again, all the bacterial 3-MPG-I enzymes returned the same CATH structural classification scheme, but the human protein was classified differently as shown below (bacteria, left; human, right). These results suggest the two proteins may not be structurally conserved at all.

CATH Classification

Level	CATH Code	Description
C	1	Mainly Alpha
A	1.10	Orthogonal Bundle
T	1.10.340	Endonuclease III; domain 1
H	1.10.340.30	Hypothetical protein; domain 2

CATH Classification

Level	CATH Code	Description
C	3	Alpha Beta
A	3.10	Roll
T	3.10.300	3-methyladenine DNA Glycosylase; Chain A
H	3.10.300.10	Methylpurine-DNA glycosylase (MPG)

In a DALI confirmation search using the PDB structure for Staphylococcus Aureus 3-MPG(PDB:2JG6), all the available bacterial structures were returned as potential structures that were comparable to that of S. Aureus. This included the MSSA variant 476, E. Coli and S. Enterica Typhi. Other results of interest were the more distantly matching structures of H. Pylori, endonuclease III and 3-MPG-II. The human structure for 3-MPG-I, however, was not returned in this comprehensive list.

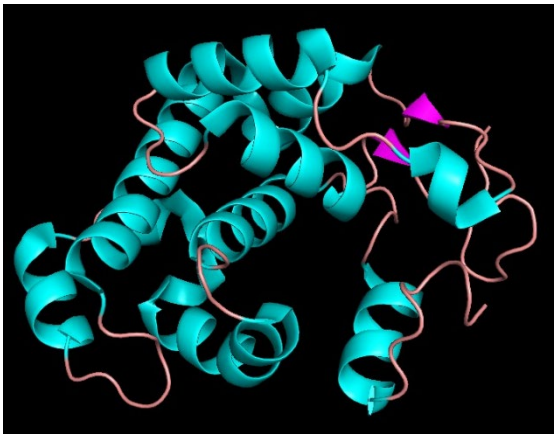
Upon further investigation into the CATH and Interpro classifications, the helix-hairpin-helix fold conformation is important in classifying the bacterial 3-MPG-I proteins separately from human AAG protein. The bacterial helix-hairpin-helix fold is conserved across bacterial species in this enzyme, and the human AAG protein lacks this conformation

(Drohat et al., 2002). This causes the two proteins to be classified separately and explains better the lack of finding any close relationship between human and bacterial methyladenine glycosylases.

Given the results from Interpro, CATH and DALI, a couple of things may be concluded. First, the bacterial 3-MPG-I enzymes have comparable structures. This may mean that any drug designed to target this enzyme may work across bacterial species. This can be said for at least those bacteria investigated here. The second conclusion is that it is also possible that the human enzyme is both sequentially and structurally different enough from the bacterial enzymes that a drug targeted to the bacterial enzyme 3-meA binding site may be specific to the bacterial sites only.

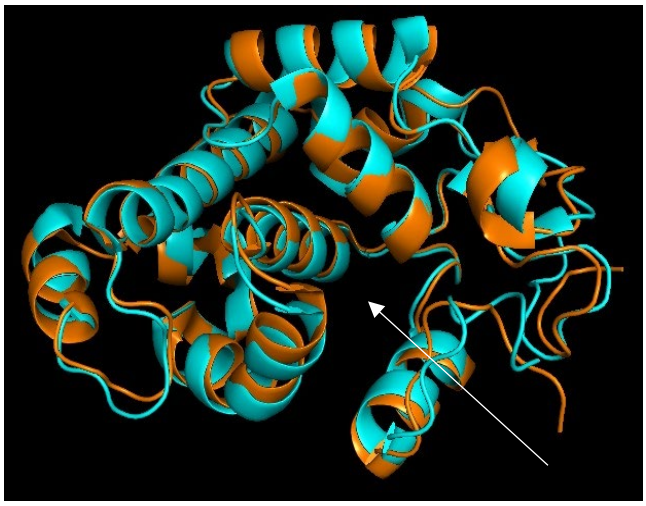
In the PDB, there are structures for the following bacterial 3-MPG-I enzymes: S. Aureus(2JG6), E.Coli(1LMZ), S. Enterica Typhi(2OFI), and MSSA-476(4AI5). The human AAG enzyme is PDB structure 1F6O. Since S. Aureus was used as the comparative species in the evolutionary analysis, each species listed here is aligned with S. Aureus.

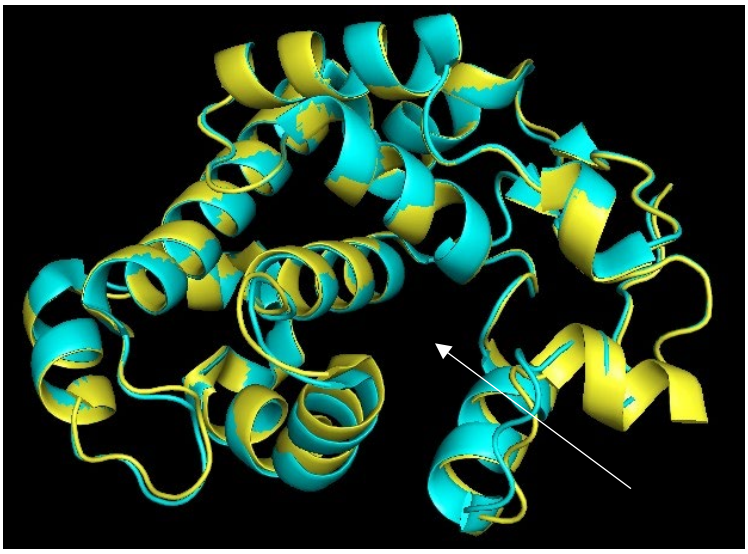
The 3-MPG-I enzyme from S. Aureus (2JG6) is shown below. It includes the canonical helix-hairpin-helix structure common to proteins found in the DNA glycosylase superfamily (Drohat et al., 2002). Two beta sheets are also seen in the S. Aureus variant of 3-MPG-I. Helices: cyan, hairpin loops: light pink, Beta-sheet: dark pink.



3-MPG-I of S. Aureus(cyan) was aligned with the same enzyme from E. Coli(pink), shown below. The RMSD score for this alignment was 1.879 which is an excellent alignment score and demonstrates the potential for cross-species druggability. The DNA binding pocket is indicated with an arrow. The alignment reveals that the size and shape of the pockets are similar, but that the E. Coli pocket is potentially smaller.

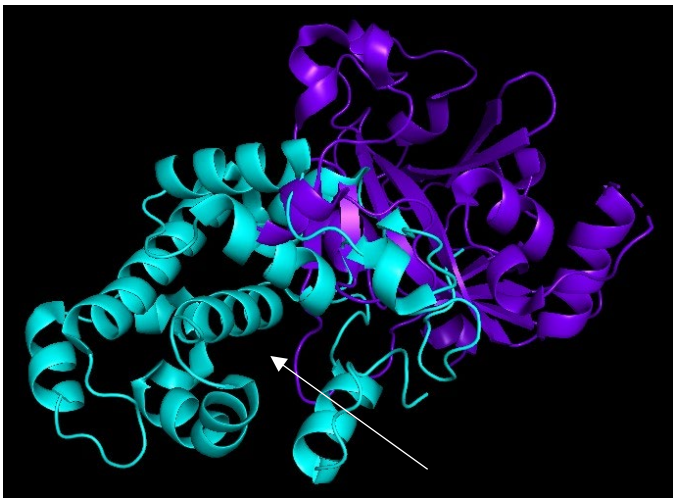
3-MPG-I of S. Aureus(cyan) was aligned with the same enzyme from S. Enterica Typhi(orange). The RMSD score for this alignment was 1.048 which is an excellent alignment score, even stronger than the E. Coli alignment, and demonstrates the potential for cross-species druggability. The DNA binding pocket is indicated with an arrow. The alignment reveals that the size and shape of the pockets are very similar with little variation between them.





3-MPG-I of S. Aureus(cyan) was aligned with the same enzyme from S.Aureus variant MSSA-476 (yellow). The RMSD score for this alignment was .347 which is an excellent alignment score but not surprising since MSSA-476 likely has only a few amino acid changes from native S. Aureus in its sequence. The DNA binding pocket is indicated with an arrow. The alignment reveals that the size and shape of the pockets are very similar with little variation between them. This particular MSSA variant has a Y16F change in the binding pocket of the enzyme, but the pocket still appears to adopt a very similar shape and size to the native S. Aureus protein and would likely be susceptible to a drug that inhibits the native binding site.

3-MPG-I of S. Aureus(cyan) was aligned with the AAG enzyme from human (purple). The RMSD score for this alignment was 10.75 which is a poor alignment score and demonstrates the possibility of achieving binding specificity for only the bacterial enzyme binding sites. The DNA binding pocket is indicated with an arrow. The alignment reveals that the human enzyme does not structurally align at all with the DNA binding pocket of 3-MPG-I of S. Aureus.

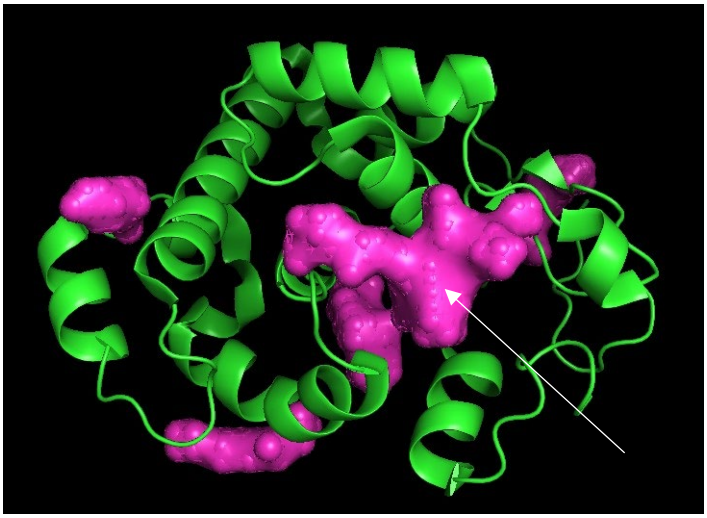


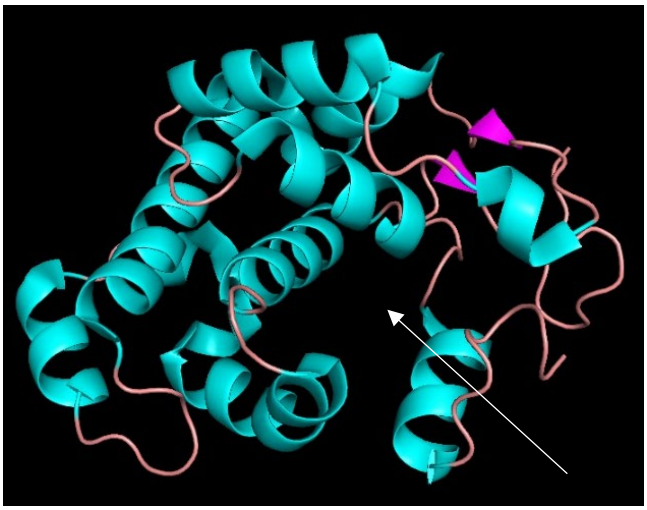
Binding Site and Druggability Assessment

A crystal structure for S. Aureus with 3-meA bound is available in the PDB and shown to the right (ID:4AIA). Prior to exploring the binding site further, pocket finders were used in order to be sure the binding site would be considered a druggable pocket.

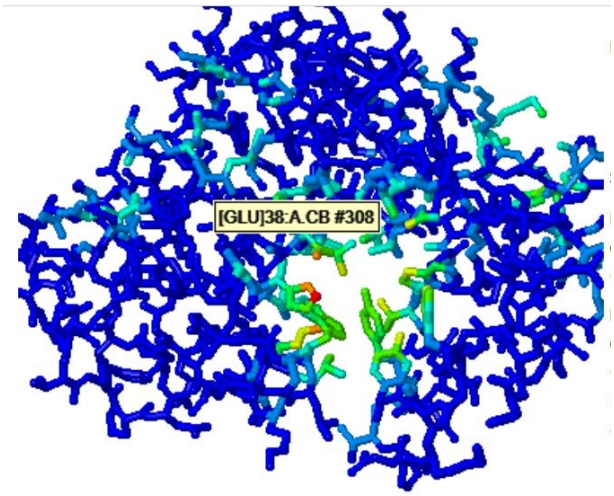
CASTp failed to identify the meA site as a druggable pocket however two other pocket finders, POCASSA and GHECOM, identified the meA binding site as the highest potential for druggable pocket.

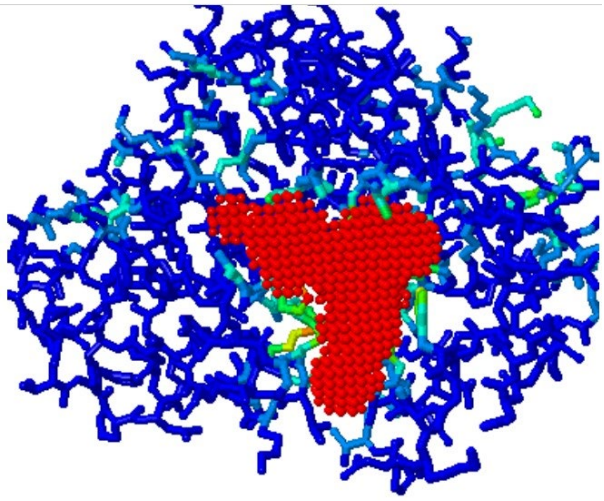
POCASSA identified several pockets and all are shown in the image below. Notably, the meA binding pocket is identified.





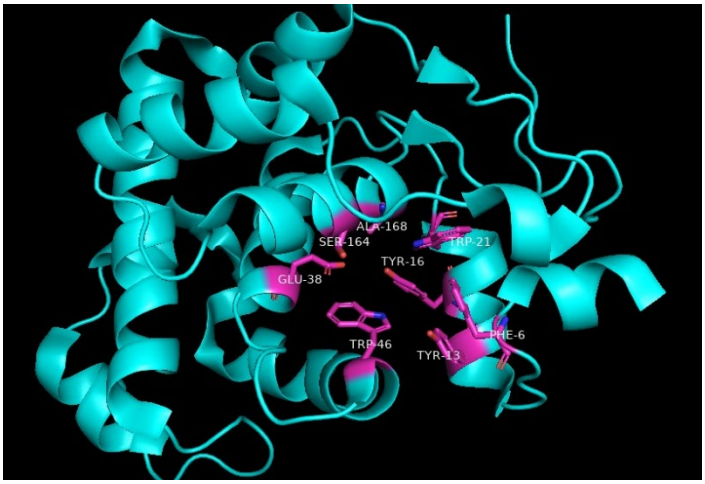
GHECOM was also able to identify the binding site as a potentially druggable pocket. In the empty pocket image, GLU-38 is identified to orient the viewer for comparison to the Pymol binding site image. The second image shows the pocket identified and filled.

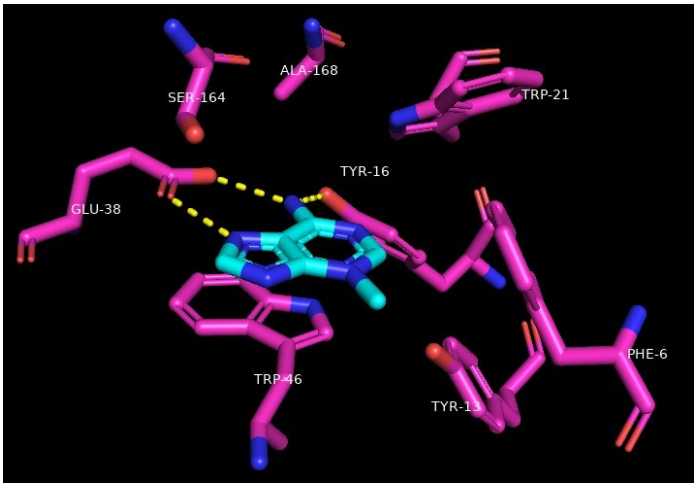


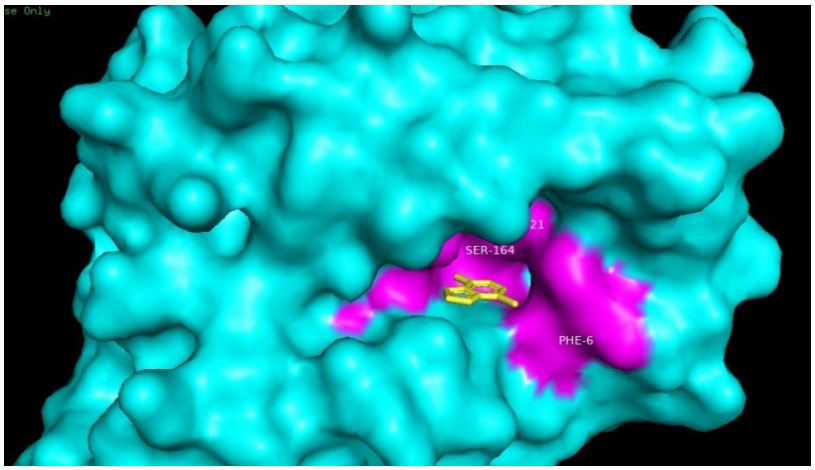


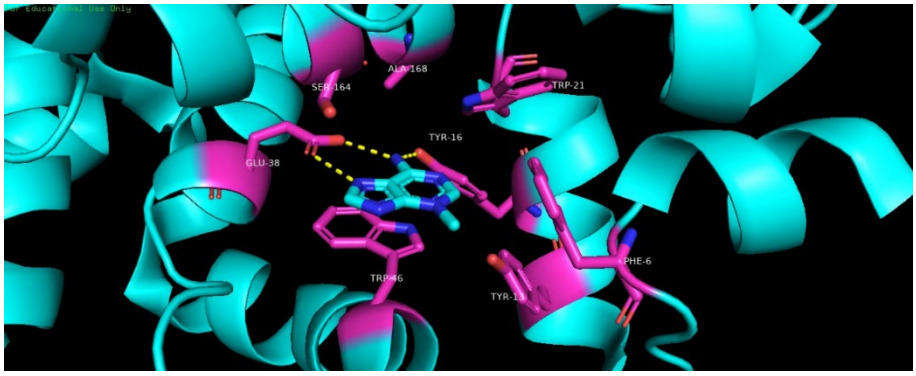
Since two out of three pocket finders were able to identify the binding site as a potential druggable pocket, it is quite possible that the binding site would work for an inhibitor of 3-MPG-I.

Finally, a more detailed analysis of the binding site for the S. Aureus 3-MPG-I was undertaken. The following images show the pocket with and without meA in the binding site(top; left& right), a detailed close- up view of the residues found in the binding site around 4meA(lower left&bottom) and a surface representation of the binding site pocket(lower right).









Looking closely at the binding site, there are 3 hydrogen bonds that are crucial for binding between the enzyme and the 3-meA ligand. Two of those bonds occur between GLU-38 and meA and the final H bond occurs between TYR-16 and meA. Zhu et al., 2012 discovered these same binding characteristics for the S. Aureus 3-MPG-I. Several hydrophobic residues exist in the pocket as well, with phenylalanine, tryptophan and alanine all within 4A of the meA ligand. These could all be potentially involved in van der Waals interactions with meA. Overall, what we find is a nice-sized primarily hydrophobic pocket that has access to solvent and is not too deep into the 3-MPG-I protein. The pocket also has polar residues that lend themselves well to participating in polar/H-bonding with a potential drug. The S. Aureus pocket appears to be a good candidate for inhibitor design.

Discussion and Final Recommendation

3-MPG-I is an essential enzyme in the base excision repair pathway, and inhibition of 3-MPG-I would likely result in inhibition of bacterial replication. This alone makes 3-MPG-I worthy of consideration, and this study supports the idea that 3-methyladenine glycosylase I is in a strong potential antibiotic target.

The comprehensive evolutionary analysis revealed that this enzyme is likely conserved across both gram-positive and gram-negative bacterial homologs. Therefore, a drug targeting 3-MPG-I could be a broad-spectrum antibiotic. Further, the evolutionary analysis suggests that if the human AAG enzyme is a homolog of 3-MPG-I, it is a distant one. This suggests that drugs targeting the bacterial form of 3-MPG-I may be able to achieve specificity in binding only to the bacterial enzyme.

Structural analysis further supports the hypothesis that 3-MPG-I is a possible bacterial specific broad-spectrum target. Comparing both the structural evolution and classifications demonstrated that the bacterial 3-MPG-I were all closely related and classified into the exact same CATH categories. The human AAG enzyme was differentially classified from the bacterial enzyme. Further, the human enzyme was not considered to be a close or even distant structural homolog to the S. Aureus 3-MPG-I enzyme by DALI. Again, this suggests that the human and bacterial structures may not be homologous and that their structures are so different that any drug targeting the binding site of bacterial 3-MPG-I would not bind to the human enzyme. 3D structural alignment in PyMol between the S. Aureus enzyme and other bacterial enzymes returned strong alignment scores and excellent visual alignment such that the 3meA binding site was obvious and appeared conserved in all bacterial species studied. In contrast, the S. Aureus enzyme did not align with the human 3-MPG-I enzyme almost at all, and the binding sites of each respective enzyme did not align. The 3D structural alignments support the idea that antibiotics designed for bacterial 3-MPG-I may be both specific to bacteria and could potentially affect multiple different types of bacteria.

Finally, despite the conflicting results between pocket finders, two out of three of the tools used were able to identify the 3meA binding site in 3-MPG-I as a druggable pocket. Additionally, the binding site pocket is quite suitable for drugs and small molecule inhibitors. The pocket is primarily hydrophobic with some polar residues that could allow for H bonding. It is also easily accessible from the solvent and not too deep into the 3-MPG-I enzyme. Many of the residues within the pocket also appear to be conserved when looking at the evolutionary analysis and most of them are not conserved in human. However, the two H-bonding residues, Tyr16 and Asp38, are conserved in human and could potentially be problematic for off target effects if these residues are also found in the human enzyme binding site. Further research should be done in order to better understand the human AAG binding site to determine if there is sufficient difference in residues to achieve high antibiotic specificity. Additionally, the human 8-oxoguanine DNA glycosylase was returned as part of the DALI distant homologs search. The binding site of this protein should also be investigated further for potential issues with drug specificity. Overall, though, the evidence provided in this study supports the idea that high specificity is possible.

Prior to proceeding with drug design against the bacterial 3-methyladenine glycosylase I enzyme, it is recommended that more basic research into the effects of knock-down or knock-out of this protein be explored. It would be helpful to first confirm the essentiality (and non-redundancy) of this protein to bacterial replication and survival. The evidence in the present analysis supports moving forward with functional experiments as 3-MPG-I appears to be a potential antibiotic target that may have both broad-spectrum and specific effects.

REFERENCES

Drohat, A. C., Kwon, K., Krosky, D. J., & Stivers, J. T. (2002). 3-Methyladenine DNA glycosylase I is an unexpected helix-hairpin-helix superfamily member. nature structural biology, 9(9), 659-664.

Grin IR, Mechetin GV, Kasymov RD, Diatlova EA, Yudkina AV, Shchelkunov SN, Gileva IP, Denisova AA, Stepanov GA, Chilov GG, et al. (2021). A New Class of Uracil–DNA Glycosylase Inhibitors Active against Human and Vaccinia Virus Enzyme. Molecules. 26(21):6668. https://doi.org/10.3390/molecules26216668

Mechetin, G.V., Endutkin, A.V., Diatlova, E.A., Zharkov, D.O. (2020). Inhibitors of DNA Glycosylases as Prospective Drugs. International Journal of Molecular Sciences. 21(9):3118. https://doi.org/10.3390/ijms21093118

Metz, A. H., Hollis, T., & Eichman, B. F. (2007). DNA damage recognition and repair by 3-methyladenine DNA glycosylase I (TAG). The EMBO journal, 26(9), 2411–2420. https://doi.org/10.1038/sj.emboj.7601649

Nayak, S., Pradhan, D., Singh, H., & Reddy, M. S. (2019). Computational screening of potential drug targets for pathogens causing bacterial pneumonia. Microbial pathogenesis, 130, 271–282. https://doi.org/10.1016/j.micpath.2019.03.024

Wyatt, M. D., Allan, J. M., Lau, A. Y., Ellenberger, T. E., & Samson, L. D. (1999). 3-methyladenine DNA glycosylases: structure, function, and biological importance. BioEssays : news and reviews in molecular, cellular and developmental biology, 21(8), 668–676. https://doi.org/10.1002/(SICI)1521-1878(199908)21:8<668::AID-BIES6>3.0.CO;2-D

Zhu, X., Yan, X., Carter, L. G., Liu, H., Graham, S., Coote, P. J., & Naismith, J. (2012). A model for 3-methyladenine recognition by 3-methyladenine DNA glycosylase I (TAG) from Staphylococcus aureus. Acta crystallographica. Section F, Structural biology and crystallization communications, 68(Pt 6), 610–615. https://doi.org/10.1107/S1744309112016363