Alyssa Hu & Yibei Jiang

Project 1: Annotation of Acinetobacter baumannii

CSE 182 Final Project Report

6/5/18

**Introduction**

*Acinetobacter baumannii* is an opportunistic pathogen and the causative agent of severe infections such as pneumonia, septicemia, meningitis, and urinary tract infections. Due to its versatile genetic machinery allowing it to quickly evolve resistance factors, survive on artificial surfaces, and persist through desiccation, *A. baumannii* has quickly evolved to become both endemic to hospitals and multidrug-resistant, thus making it a very real health threat.

In order to determine what characteristics make *A. baumannii* such a successful pathogen in hospital settings, scientists have been attempting to study which proteins are involved in mechanisms of antibiotic resistance. However, existing annotation of the *A. baumannii* genome is quite poor; thus, our project sought to develop an automated tool to query a variety of online databases in order to annotate 100 protein sequences (specifically, sequences 101 to 200) from *A. baumannii*.

**Methods & Tools**

The tool pipeline was written primarily in Python 3, and queries its data from databases such as BLAST, InterPro, Pfam, and Prosite. Each of these databases provides different information about a protein sequence. In order to make it more user-friendly, users are allowed to specify which tools to run, what their desired cutoff parameters are, and the output format (tab-, comma-, or space-separated plain text files). The tool pipeline can also take either FASTA-formatted sequence files or accession numbers as input, so that the user is not limited to only bacterial species.

1. BLAST:

BLAST (Basic Local Alignment Search Tool) is a tool to infer functional and evolutionary relationships via alignment, as well as identify members of gene families. Since we were interested in finding possible *A. baumannii* protein functions, we used the NCBI BLASTp to search against protein sequence databases. We utilized the Biopython package in order to invoke the NCBI BLAST server over the internet, as well as use the built-in parsers in order to work with the raw XML files output by BLAST. We then wrote a Python script (blast.py) which queried our 100 protein sequences of interest and returned the output (which contained up to 10 hits) in two different file formats per sequence: a raw XML file output by BLAST, and a separate plain text file which contained the parsed information such as the protein accession number, gi number, predicted/known protein function, and E-value.

1. InterPro:

As an integrated database, the InterPro consortium provides functional analysis of protein sequences by classifying them into families and predicting the presence of domains and other important sites. InterPro provides a software package known as InterProScan that allows sequences to be searched against key signatures in its database. Due to memory and processor limitations, we used InterPro’s REST service via their provided Python 3 client in order to query our sequences against their database over the internet. The drawback of this was that the InterPro REST service takes only one FASTA-formatted sequence at a time, and has a 30 sequence per batch limitation. We wrote a simple Python script (fasta\_fragment.py) to fragment our file of 100 *A. baumannii* protein sequences into individual FASTA files. We also followed an online tutorial for the InterPro REST service to write a bash script to automate the input of up to 30 files at a time.1 Raw output was given as tab-separated plain text files, which we then parsed for protein function and relevant Gene Ontology terms.

1. Pfam:

Pfam is a database of protein families and clans. It predicts the possible families the protein sequence belongs to using a HMM model. Since the web interface version of Pfam uses hmmscan to search, I downloaded the local version of HMMER 3.1b2. The database used by hmmscan tool is Pfam-A.hmm that can be downloaded from the ftp server but hmmpress should be run on the .hmm database. runPfam.py is used to query all sequences in the database. We used the default E-value of 1. It can be run independently using “python runPfam.py input\_Dir”.

1. Prosite:

Prosite is a database containing the function of proteins using regular expression. The perl script ps\_scan.pl is downloaded and used to run Prosite search. The -d flag is used to specify the database file Prosite.dat that can be downloaded from the ftp server. The -s flag is used to mask all unspecific functions, such as N-glycosylation sites that are not crucial for studying the function of our protein sequences that can result in antibiotic resistance. -l flag is used to indicate level of sensitivity. 0 indicates only output results with high sensitivity while -1 indicates output results with low sensitivity. Since while running with high sensitivity doesn’t generate much output, I first query the database with high sensitivity (-l 0) and if there is no output, I run ps\_scan again using low sensitivity (l=-1) to get more hits. runProsite.py is the code that runs Prosite to make it run, a gfortran compiler should be downloaded in the system. It can be run independently “python runProsite.py input\_Dir”.

1. Analysis:

Analysis.py is the python script that runs all tools together, concatenating and outputting all the results in one file. Users will run this script using the appropriate flags.

**Biology: Antibiotic Resistance Mechanisms**

Bacteria have evolved a variety of resistance mechanisms through which they combat antibiotics. Resistance primarily comes in the form of redox mechanisms to neutralize reactive oxygen species, directly destroying or modifying a compound harmful to the bacterium. Other methods of resistance include prevention of a drug from interacting with its target via biofilm or capsule formation and efflux of the antibiotic from the bacterial cell.

In *A. baumannii*, there has been evidence of a gene cluster whose products form a polysaccharide capsule and inhibit antibiotics and their effect on the complement system when grown on ascites fluid.2Furthermore, *A. baumannii*’s ability to survive in artificial or harsh environments for extended periods of time appears to be associated with its ability to form biofilms. These biofilms can can affect the metabolism of microorganisms inside them, slowing them down and thereby reducing their sensitivity to antibiotics by preventing bacteria from taking in antibiotics quickly enough to kill them.3

Efflux pumps are another method through which bacteria resist antibiotics. They are active transporters which are responsible for moving compounds such as neurotransmitters, toxins, and antibiotics out of the cell. There exist two major efflux pumps in *A.baumannii*: AdeB and AdeDE. AdeB is primarily responsible for resistance to aminoglycosides, but both AdeB and AdeDE also provide resistance to compounds such as chloramphenicol, erythromycin, and tetracycline.

Finally, and perhaps the most relevant to our results, are the redox mechanisms which bacteria use to resist oxidative damage from compounds such as drugs or oxygen species produced by phagocytes. It has been shown that antibiotics alter cellular respiration, inducing lethal levels of intracellular hydrogen peroxide to kill bacteria. Bacteria, in turn, are able to use antioxidants such as oxidative stress defense proteins, to counter the killing by antibiotics, as antibiotics tend to be highly sensitive to the presence of molecular oxygen.4 Redox mechanisms thus appear to be one of the major pathways of bacterial resistance against antibiotic-induced death.

**Discussion**

After finishing our tool pipeline to automate the collection of annotation data relevant to our 100 *A. baumannii* protein sequences, we chose four proteins, ABO11055, ABO11102, ABO10642, and ABO10621, which seemed the most significant and had important functions that related to antibiotic resistance mechanisms. Statistical significance was determined by a score of 0 in the “Prosite Confidence” column (denoting high sensitivity), as well as very small E-values for Pfam and BLAST. Please refer to Figure 1 for an example of our output in table format. We summarize the function of each protein as follows:

1. ABO11055: Thioredoxin family

Thioredoxin family members regulate redox homeostasis in a stressed environment. Thioredoxin-1 (Trx-1) is a known agonist for chemotherapeutic drug resistance, and appears to regulate the effects of the oxidative stress response on glycolysis. While the role of thioredoxin in organisms is not fully understood, it is essential for life in mammals and appears to respond to reactive oxygen species. In mice models, mice that overexpress thioredoxin are more resistant to inflammation and live 35% longer, and there could be a similar effect in *A. baumannii* should it overexpress the encoding gene as well.5

1. ABO11102: Ferrous iron transport protein B (FeoB)

Iron uptake systems have the ability to induce biofilm formation, which subsequently induces poor antibiotic penetration into the bacterial cell. Studies in which FeoAB was overexpressed also demonstrated an increased resistance to certain antimicrobial peptides. Notably, there are a variety of protective mechanisms incorporated with the formation of biofilms; nutrient limitation and slowed growth allow the bacterium to go through key processes and uptake antibiotics more slowly, which could prevent the metabolism of a lethal amount of antibiotic as a result. As such, preventing biofilm formation could be a key component to unraveling *A.baumannii’s* defenses.6

1. ABO10642: Glutathione peroxidase

Glutathione peroxidase is an enzyme family that protects organisms from oxidative damage. Glutathione (GSH) is one of the main antioxidants in living organisms, and reduces lipid hydroperoxides into alcohol and water. Furthermore, phagocytes use oxygen species such as superoxides and hydrogen peroxide in order to kill bacteria. Thus, up-regulation of the gene responsible for glutathione peroxidase could potentially allow *A. baumannii* to have increased resistance to the natural defenses of the human immune system.7

1. ABO10621: Glutathione S-transferase (GST)

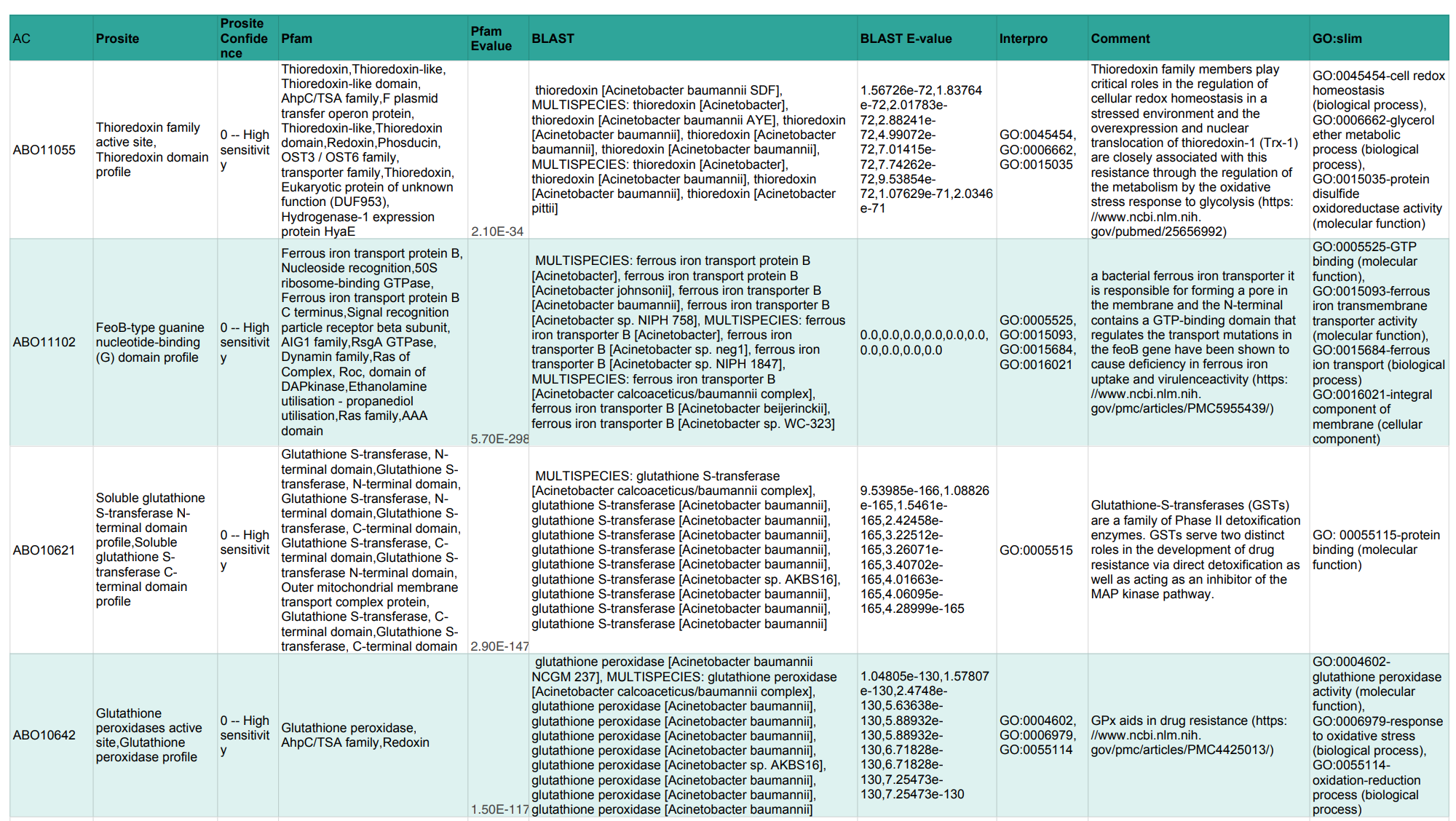
Glutathione S-transferase is a detoxification enzyme which catalyzes the conjugation of GSH to make compounds more water-soluble. It also has the ability to bind toxins, functioning as a transport protein to remove them from the organism, and detoxify compounds such as peroxidized lipids and enable the breakdown of xenobiotics. GST is also an inhibitor to the MAP kinase pathway. As the MAPK pathway regulates cell proliferation and death, high levels of GST are thus associated with resistance to apoptosis induced by a range of substances, and could be a mechanism through which *A. baumannii* achieves its antibiotic resistance.8

**Conclusion**

There is a large variety of mechanisms through which *A. baumannii* may have evolved its remarkable resistance to antibiotics. We have only performed analysis on comparatively small number of protein sequences from the *A. baumannii* genome; however, we did obtain promising results simply from picking four proteins sequences to investigate more thoroughly. Via these four proteins we chose to investigate, we found that redox mechanisms and biofilm formation appeared to be prime suspects in contributing to the ability of *A. baumannii* to resist antibiotics.

In regards to future considerations to improve our tool, we can speed up our analysis by using parallel processing that runs all input sequences in parallel to speed up the process. Moreover, annotating as well as getting the GO:slim term is very tedious. We can automate annotation by including other tools. We also hope to keep the basic configuration as well as package dependencies to a minimum for user-friendliness. We can write bash scripts to check for all necessary dependencies and download them.

**Figure 1: Sample output data from tool pipeline**



**Citations:**

1. Mun, T. (2016, November 27). Using InterProScan like a pro. Retrieved June 5, 2018, from <https://lotus.au.dk/blog/bioinformatics/2016/11/28/interproscan>
2. Luke, N. R., Sauberan, S. L., Russo, T. A., Beanan, J. M., Olson, R., Loehfelm, T. W., … Campagnari, A. A. (2010). Identification and Characterization of a Glycosyltransferase Involved in Acinetobacter baumannii Lipopolysaccharide Core Biosynthesis. Infection and Immunity, 78(5), 2017–2023. <http://doi.org/10.1128/IAI.00016-10>
3. Effect of biofilm formation on the survival of Acinetobacter baumannii on dry surfaces. Espinal, P. et al. Journal of Hospital Infection , Volume 80 , Issue 1 , 56 - 60
4. Dwyer, D. J., Belenky, P. A., Yang, J. H., MacDonald, I. C., Martell, J. D., Takahashi, N., … Collins, J. J. (2014). Antibiotics induce redox-related physiological alterations as part of their lethality. Proceedings of the National Academy of Sciences of the United States of America, 111(20), E2100–E2109. <http://doi.org/10.1073/pnas.1401876111>
5. Li Zhao, Wei Li, Yuxin Zhou, Yi Zhang, Shaoliang Huang, Xuefen Xu, Zhiyu Li, Qinglong Guo,The overexpression and nuclear translocation of Trx-1 during hypoxia confers on HepG2 cells resistance to DDP, and GL-V9 reverses the resistance by suppressing the Trx-1/Ref-1 axis,Free Radical Biology and Medicine,Volume 82,2015,Pages 29-41,ISSN 0891-5849,, https://doi.org/10.1016/j.freeradbiomed.2015.01.014., (<http://www.sciencedirect.com/science/article/pii/S0891584915000210>)
6. Cheryl K. Y. Lau, Karla D. Krewulak, Hans J. Vogel; Bacterial ferrous iron transport: the Feo system, FEMS Microbiology Reviews, Volume 40, Issue 2, 1 March 2016, Pages 273–298, <https://doi.org/10.1093/femsre/fuv049>
7. Kominkova M, Michalek P, Cihalova K, et al. Study of Linkage between Glutathione Pathway and the Antibiotic Resistance of Escherichia coli from Patients’ Swabs. Yan B, ed. International Journal of Molecular Sciences. 2015;16(4):7210-7229. <https://doi.org/10.3390/ijms16047210>.
8. Longkumer T, Parthasarathy S, Vemuri S, Siddavattam D. Microbiology 160(1):102-112 <https://doi.org/10.1099/mic.0.070664-0>.