Final Project Report

Jaden Anderson

May 1st, 2018

Introduction

Gram negative bacteria have a unique arrangement of membranes that help them persist in their enviornment: an outer and inner membrane. On these membranes exist a variety of proteins and structures that have a range of functions that the bacteria utilize(Remmert et al. 2010). Outer membrane proteins are used to interact with their extracellular enviornment that have a diverse array of functions like import and export of substances, cell adhesion, proteolysis, membrane protein insertion, and more. Even though proteins on the outer membrane have a variety of functions, surprisingly the secondary structure of these proteins (i.e. the structure of what the proteins look like) are very similar(Franklin et al., n.d., Gu et al. (2016)). While these secondary structures are highly similar, their sequence composition (or similarity) is highly variable which makes studying their evolution very difficult. An evolutionary analysis based on HMM profiles was carried out and distinguished roughly 8 unique groups of outer membrane beta barrels (OMBBs) (Franklin et al., n.d.).

The folding patterns of the OMBBs is controlled by the assembly group of proteins which consist of a small group of proteins as seen from Figure 1. The most notable assembly complex in gram negative bacteria is the beta barrel assembly complex or BamA (Plummer and Fleming 2015). It is responsible for folding the majority of OMBBs into the outer membrane however its evolutionary relationship with the other OMBBs its responsible for is not well understood (Plummer and Fleming 2015). Given it's important role in the folding process of OMBBs it is surprising that BamA or the assembly proteins have little to no connectedness through HMM profiles which can be used as a measure of evolutionary relationships. In this study, I will perform a sequence based analysis of BamA and the current 130 structurally characterized OMBBs. My hypothesis is that BamA sequences in different species will be highly connected or overlap with critically important beta barrels from that same species of BamA, which might give insight onto the folding preferences of the beta barrel assembly complex.

Methods

Basic Local Alignment Search Tool or BLAST

BLASTing protein sequences is a fairly standard practive when trying to determine relationships between different proteins. The BLASTing algorithm that I used is PSIBlast which is based off the NCBI blastp algorithm. Normal BLAST (or blastp) is a sequence similarity search method which identifies like protein sequences based off given thresholds like homology or sequence similiarty (Bhagwat M 2007). PSIBLAST is an iterative tool that builds protein sequence profiles from blastp results in order to refine search space and produce higher scoring alignments (Bhagwat M 2007). I PSIblasted BamA along with all 130 structurally characterized OMBBs in order to generate multiple sequence alignments. In order to make sure I was pulling similar sequences, I used a false positive rate or evalue of 10^-50.

The parent sequence, or original sequence, used to PSIBLAST similar proteins were pulled from the IDB. A PDB identifier is a 4 digit combination of numbers and letters with an optional chain identifier (_letter). I grabbed all 130 PDB files from the RSCB PDB webserver.

Determining Overlapping Species

Once I've obtained the PSIBLASTed sequences I will determine the number of overlapping sequences of BamA with a client protein (one that BamA is responsible for folding). To do this, I will exploit the format of

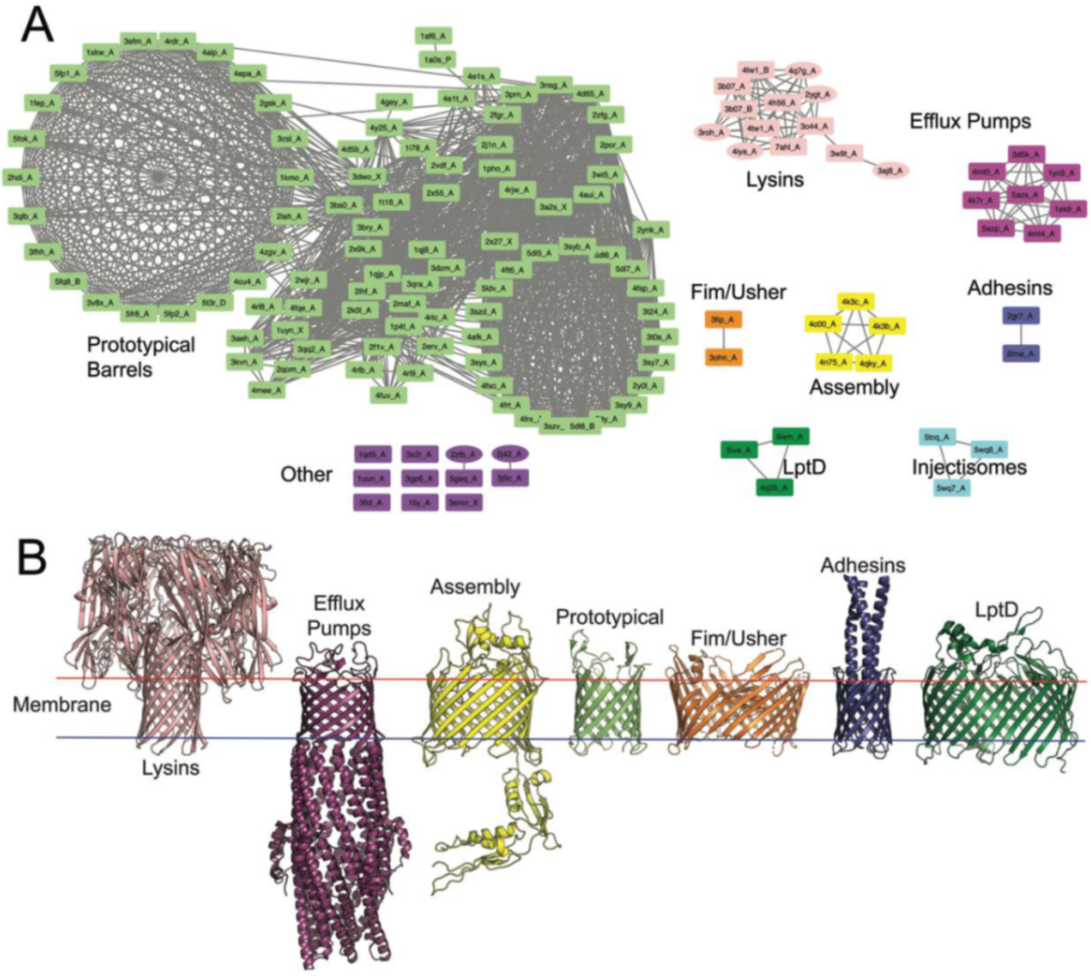

Figure 1: This figure is taken from "The Complex Evolution of Outer Membrane Proteins" which is in review. This figure illustrated the evolutionary relationships and connectedness of the OMBBs. A) The 8 groups broken up into Prototypical Barrels, Lysins, Efflux Pumps, Adhesins, Assembly proteins, Fim/Usher proteins, and LptD proteins. B) A structural representation of the conserved structure of the proteins from each group.

FASTA files and use regular expression to grab unique species identifiers. Once I have grabbed those species from BamA, I will compare that list to each of the species that resulted from the PSIBLAST alignments for each client protein. This process can be found in the python chunk get_overlapping_sp of this RMD. In order for this script to be sucessful, I must use the reformat_MSA python chunk to reformat the alignments generated from the PSIBLASTed results. This is done to make the parsing quicker.

Testing Formats

In order to ensure that the data that I've grabbed from the NCBI webserver from PSIBLAST is in the proper formats, I will be loading the Rpackage that I wrote for the last project in order to use the format and protein unit tests on my sequences. I will not need to use the same_size unit test, as this analysis does not care about the size of the proteins in the multiple sequence alignment.

Modularity

It is critically important that a script that works on one file works on all 130 OMBBs fasta files that I have in this directory. In order to accomplish that, all scripts written in this project will be modular. Sciprts were originally written in python and imported to RMD using the Knitr capabilities.

Results and Discussion

The results of the analysis are show in Figure 2. From here we can start to see some interesting trends of possible interactions that BamA might have with client proteins. Only overlapping sequences that were greater than 500 were considered to be important in this analysis. From this, we can see a variety of client proteins important for assembly (other Bam proteins, TamA), maitenence of the lipid compositions (LptDs), major transporters (BtuB), efflux pumps (1EK9) and a variety of other critically important functions.

Interestingly enough, all of these proteins are the larger proteins ($>$ 20 strands) besides one exception FadL a 14 stranded beta barrel. If we look at Figure 3 which is an HMM profile nextwork with strand size, we can see some interesting trends. FadL or 1T16 seems to be interconnected between the 14 stranded beta barrels and 22 stranded beta barrels. Given that 1T16 popped out as a highly overlapping client protein with BamA might indicate some stronger evolutionary relationship with the 22 stranded beta barrels.

Conclusion

The beta barrel assembly complex (BamA) is a crucially important protein complex that is responsible for folding OMBBs into the outer membrane (Plummer and Fleming 2015). It has been hypothesized that BamA has evolved to interact with OMBBs that are critically important for functionality (Plummer and Fleming 2015). Interestingly in my overlapping species analysis between a client protein and BamA, I find that BamA shares a large number of overlapping species with the larger OMBBs that have critically important functions in the cell, while the smaller porins fell out of my study. The smallest of the proteins that had a large overlap with BamA is FadL, a 14 stranded beta barrel. This protein has been hypothesized as a connection between the transition of 14 stranded beta barrels to 22 stranded beta barrels. My study may help solidify that argument since there is a high overlap of species between BamA and FadL (1T16).

References

Bhagwat M, Aravind L. 2007. *Comparative Genomics: Volume 1 and 2.* Edited by Bergman NH. 1st ed. Totowa, NJ. https://www.ncbi.nlm.nih.gov/books/NBK2590/.

Franklin, Meghan, Sergey Nepomnyachiy, Ryan Feehan, Nir Ben-tal, Rachel Kolodny, Joanna S G Slusky, Molecular Biochemistry, and Mount Carmel. n.d. "1,4 . 1 2."

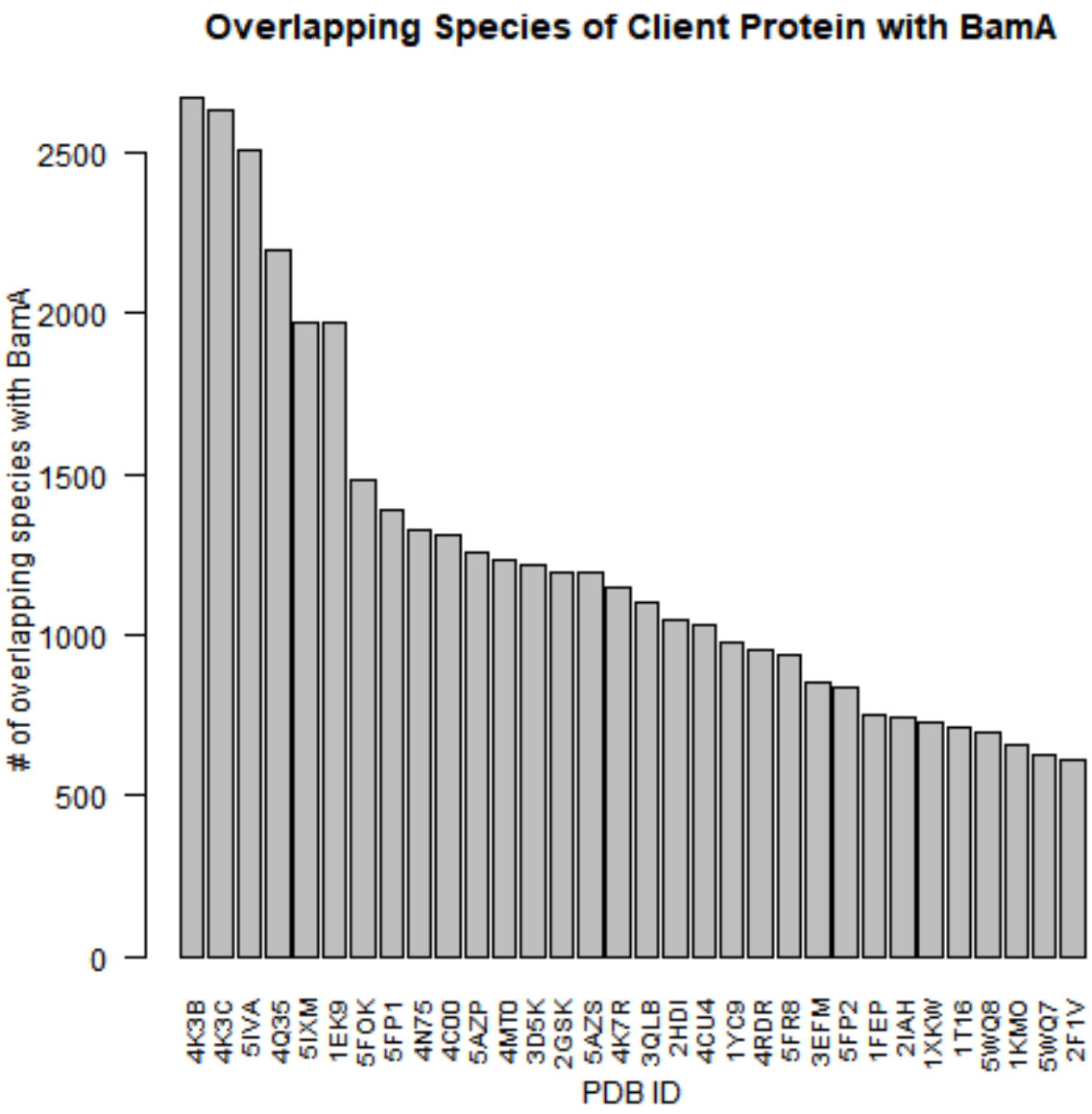

Figure 2: Figure 2: Barplot of overlapping species of client protein with BamA. The client proteins with an overlap of at least 500 sequences was plotting. From here, we can see the majority of the proteins that have the most sequences are critically important proteins for assembly of OMBBs, maintenance of the lipid composition of the membranes, major transporters, efflux pumps and more.

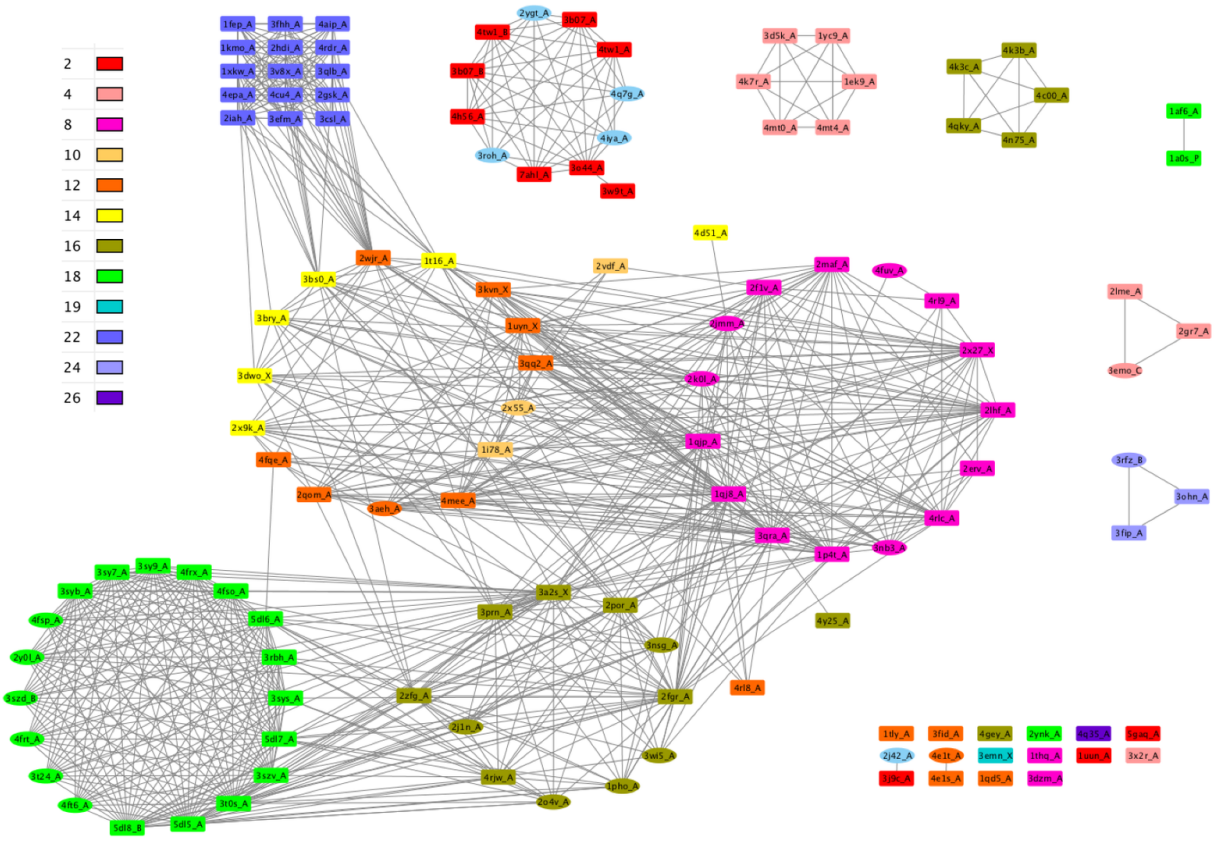

Figure 3: Figure 3: HMM profile of OMBBs with an evalue cutoff of 10ˆ-2. The evalue must be kept that high due to the low sequence similarities that OMBBs naturally have. Here we can see high connectedness of FadL (1T16) a 14 stranded beta barrel with the 22 stranded beta barrels and the 12 stranded beta barrels. It's high overlapping as a client protein with BamA might indicate a stronger evolutionary relationship depicted in this figure. This figure is also from "The Complex Evolution of Outer Membrane Proteins".

Gu, Yinghong, Huanyu Li, Haohao Dong, Yi Zeng, Zhengyu Zhang, Neil G. Paterson, Phillip J. Stansfeld, et al. 2016. "Structural basis of outer membrane protein insertion by the BAM complex." *Nature* 531.

Plummer, Ashlee M, and Karen G Fleming. 2015. "through a Catalytic Mechanism," 6009–11. https://doi.org/10.1021/acs.biochem.5b00950.

Remmert, M, A Biegert, D Linke, A N Lupas, and J So. 2010. "Evolution of Outer Membrane b -Barrels from an Ancestral bb Hairpin Research article" 27 (6):1348–58. https://doi.org/10.1093/molbev/msq017.