Conserved Regions of Acinetobacter DNA Binding Regulators With Shannon Analysis

Joshua Khorsandi

Computing
jkhorsa1

December 11, 2022

Abstract

This article uses the Shannon Entropy to quantify conserved regions of Acinetobacter *DNA binding regulators. Acinetobacter are a genus of gram-negative bacteria emerging as an area of high research interest due to their increasing resistance to antibiotics[4], prevalence in hospital care facilities, and causation of nosocomial infections[9]. Via the UNIPROT database[1] and the clustal omega algorithim we compute the multi sequence alignment(MSA) for 1,879 DNA binding regulators across 479* Acinetobacter *species, then find the Shannon Entropy for the aligned regions to identify conserved residues.*

I. Introduction

Acinetobacter is widely considered one of the most successful pathogens at causing infections in modern hospital and nursing home systems, most commonly inducing hospital acquired nosocomial infection[4]. Furthermore, in recent years the accelerated use of antibiotics has lead to an increase in the prevalence of multi drug resistant *Acinetobacter* species which have proven especially challenging for hospital systems to effectively manage[9].

DNA binding response regulator proteins interact with a cells genetic material through a DNA binding effector domain in response to changes in the external environment of a cell[5]. They act as transcription factors that promote or repress the translation of specific proteins in response to environmental factors. Clearly they play an important part in bacterial antibiotic resistance mechanisms. Indeed across the literature in gram negative bacteria response regulators(which make up a part of the larger two component regulatory systems) are identified as critical parts of antibiotic resistance pathways[2].

By quantifying the degree to which DNA binding regulators are conserved across Acinetobacter species, the most essential proteins involved in the two component regulatory system can be identified[6], which is especially pertinent in the development of further treatment options.

II. Materials and Methods

We use the UNIPROT data base of protein sequences to download 1,879 proteins matching the description "DNA Binding Response Regulator" and have taxonomy ID "Acinetobacter". This data is loaded as a tsv file and processed using the following python3 packages

- Biopython - Reading biological file data
- Matplotlib, Seaborn - Graphing
- Clustalo - Multi Sequence Alignment
- Pandas, Numpy - Data manipulation

A python wrapper around the clustal omega command line tool is used to calculate the MSA, which gives consensus sequences that indicate conserved protein regions across the UNIPROT fragments. Clustal omega[8] uses a hidden Markov model to calculate progressive alignments and update an alignment tree, this tree is then used to update the alignment file recursively.

The Shannon Entropy is a measure of the amount of information contained in a sequence.

Entropy	Conservation
> 2.0	Not conserved
1.0 < H <= 2.0	Conserved
< 1.0	Highly Conserved

We calculate the entropy according to the following[7]

$$H = -\sum_{i=1}^{M} p_i \log_2(p_i) \tag{1}$$

Where M is the size of the alphabet(the number of different amino acids in any sequence) and p_i is the prevalence of each letter in the alphabet(here, each amino acid code). H gives the total entropy. Because we take a base 2 logarithm we are encoding the entropy in bits per amino acid sequence. The maximum possible amount of entropy(ie completely random) for an amino acid sequence is 4.23, the minimmum amount is 0(every element the same). We classify entropy values according to table 1 above.

III. Analysis

The UNIPROT .tsv file(included with this paper) includes the protein ID, name and sequence, the clustal omega output file(written as alignment.clustal-num file in Data/Out) contains the protein ID and alignment.

IV. Discussion

The protein most conserved according to the entropy analysis is the protein with UNIPROT ID:N9NL92, the DNA binding regulator corresponding to *Acinetobacter sp.* ANC 3862. Notably however the following DNA Response Regulators all correspond to *Acinetobacter baumanii*. This is in agreement with previous finding of *A. baumanii* as the primary driver of antibiotic resistance in the genus[4][2].

Furthermore, our model correctly shows that proteins that are part of the two signal transduction pathway fall into the highly conserved category, looking at fig. 2 shows that in the top

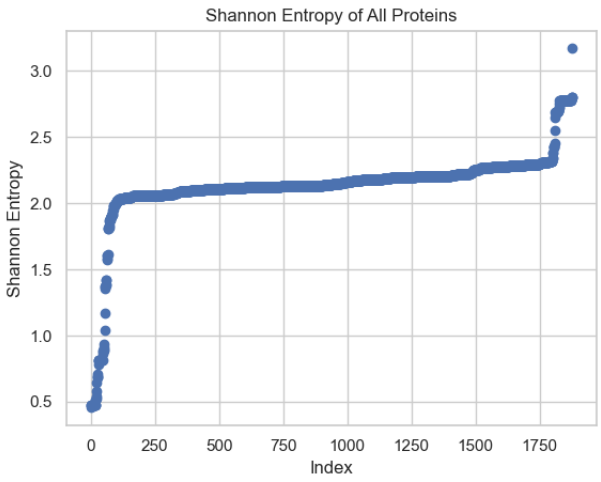

Figure 1: *The entropy of every DNA-Binding Regulator across 479 Acinetobacter species*

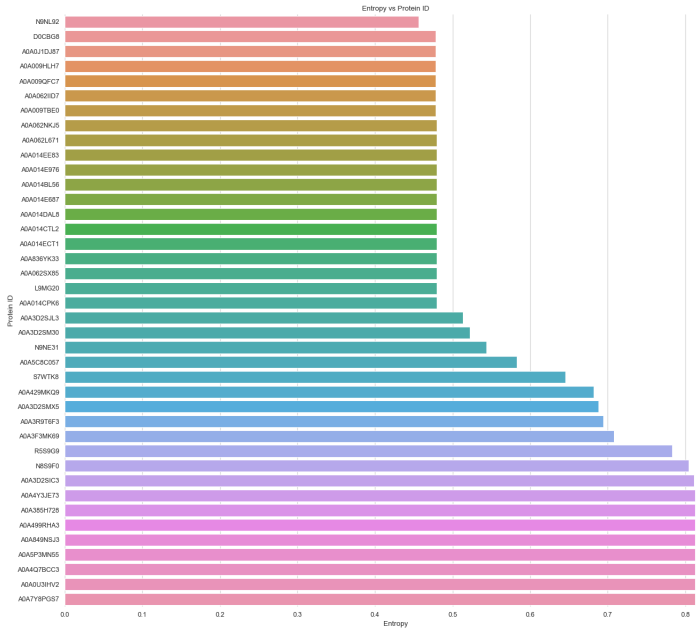

Figure 2: *The 40 most conserved proteins*

18 of highly conserved protein residues 16 of them are part of the two signal transduction pathway, those that begin with UNIPROT code

"A0A..."

As an example, consider the protein with code A0A90B602, an OmpR family transcriptional regulator with entropy score in the highly conserved range. This protein has been extensively studied as part of the two signal transduction pathway across prokaryota, indicating that the model correctly predicts that it is a highly conserved residue[3].

V. Conclusion

The use of Shannon Entropy with new and efficient Multiple Sequence Alignments to analyze conservation of protein residues across a species is clearly a beneficial and accurate tool in predicting which proteins will be conserved in a species. By identifying conserved patters in proteins we can easily identify proteins which are integral to the function of a bacteria's response to its environment.

The quantitative Shannon Entropy analysis here coincides with empirical measurement of preserved Acinetobacter DNA binding regulator proteins, and the data set of entropy's and proteins provided easily classifies proteins of possible interest for treatment. In the future, the use of Shannon Analysis can identify areas of high research interest, narrow research to the critical parts of complex signal transduction pathways and streamline treatment patterns.

References

[1] UniProt.

[2] Anjali Y. Bhagirath, Yanqi Li, Rakesh Patidar, Katherine Yerex, Xiaoxue Ma, Ayush Kumar, and Kangmin Duan. Two Component Regulatory Systems and Antibiotic Resistance in Gram-Negative Pathogens. *International Journal of Molecular Sciences*, 20(7):1781, April 2019.

[3] Karolina Jaworska, Marta Nieckarz, Marta Ludwiczak, Adrianna Raczkowska, and Katarzyna Brzostek. OmpR-Mediated Transcriptional Regulation and Function of Two Heme Receptor Proteins of Yersinia enterocolitica Bio-Serotype 2/O:9. *Frontiers in Cellular and Infection Microbiology*, 8:333, September 2018.

[4] Chang-Ro Lee, Jung Hun Lee, Moonhee Park, Kwang Seung Park, Il Kwon Bae, Young Bae Kim, Chang-Jun Cha, Byeong Chul Jeong, and Sang Hee Lee. Biology of Acinetobacter baumannii: Pathogenesis, Antibiotic Resistance Mechanisms, and Prospective Treatment Options. *Frontiers in Cellular and Infection Microbiology*, 7, March 2017.

[5] Pamela J. Mitchell and Robert Tjian. Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins. *Science*, 245(4916):371–378, July 1989.

[6] Sagnik Sen, Ashmita Dey, Sourav Chowdhury, Ujjwal Maulik, and Krishnananda Chattopadhyay. Understanding the evolutionary trend of intrinsically structural disorders in cancer relevant proteins as probed by Shannon entropy scoring and structure network analysis. *BMC Bioinformatics*, 19(S13):549, February 2019.

[7] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, July 1948.

[8] Fabian Sievers, Andreas Wilm, David Dineen, Toby J Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, Julie D Thompson, and Desmond G Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1):539, January 2011.

[9] Darren Wong, Travis B. Nielsen, Robert A. Bonomo, Paul Pantapalangkoor, Brian Luna, and Brad Spellberg. Clinical and Pathophysiological Overview of Acinetobacter Infections: a Century of Challenges. *Clinical Microbiology Reviews*, 30(1):409–447, January 2017.