Investigating Binding Site Ratios of Transcription Factors

Center for Advanced Biotechnology and Medicine, Stock Lab
Zeyue Li
Rutgers University

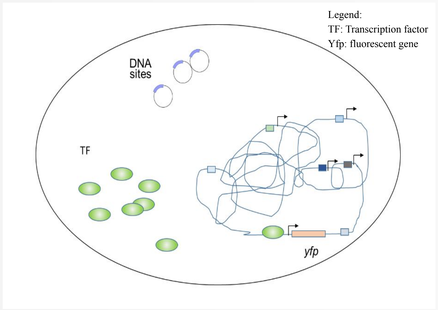
Aresty Research Center for Undergraduates

Abstract

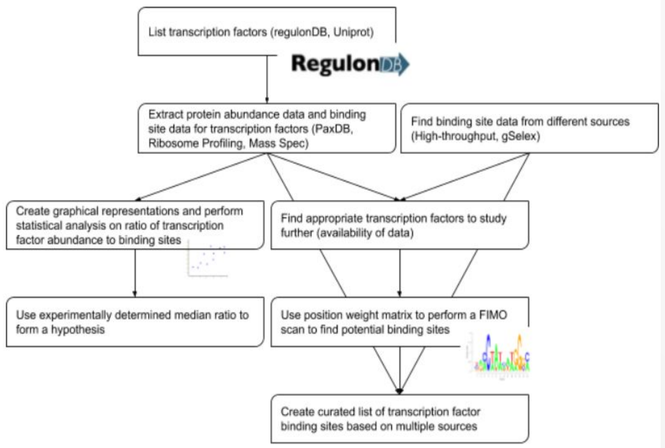
Bacterial signal transduction is controlled by transcription factors (TFs), which bind to specific binding sites on bacterial chromosomes.[3] Expression levels of TFs are often presumed to be in great excess to their binding sites, leading to high ratios of TF concentrations to the numbers of TF binding sites. Previous studies have suggested a median ratio of 10.[4] Here we used the updated binding data from regulonDB and multiple protein abundance databases to reexamine the ratios for *E. coli* TFs. We discovered that TFs are unlikely to be in great excess to their binding sites. Further, we believe that the binding site data from regulonDB is biased toward the intergenic promoter regions and there are more binding sites that have not been discovered, especially those within open reading frames. To better estimate the number of binding sites, we took a bioinformatics direction, using multiple sources, including high-throughput data, gSELEX data, and computational predictions using FIMO scan of the bacterial genome. Although different sources vary in their reports of the number of binding sites, all indicated at least two fold higher numbers of binding sites than those from regulonDB. Overall, we found that repressors tend to have a higher ratio compared to activators and dual function TFs, and we found that the originally reported ratio is higher than what studies with newer technology suggests.

Introduction

Bacterial signal transduction often relies on binding of transcription factors (TFs) to DNA to either repress or activate transcription.[3] It is often assumed that TFs are expressed in great excess to the number of their DNA binding sites (TFBS) in the chromosome.[4] However, our lab discovered that TF concentration was in balance with its DNA binding sites for the TF PhoB and additional DNA sites titrated away transcription, suggesting PhoB is not in great excess to its DNA binding sites. Here we use bioinformatics to analyze the ratio of TFs to the number of binding sites for *E. coli* TFs.



Methods



Results

Protein Abundance Data from PaxDB

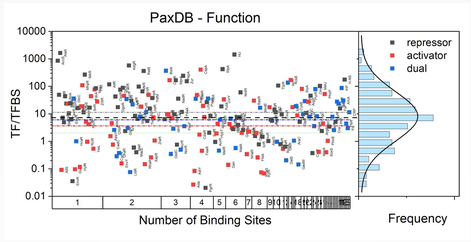


Figure 1: The median ratio of TF concentration (PaxDB) to TFBS is 7.23

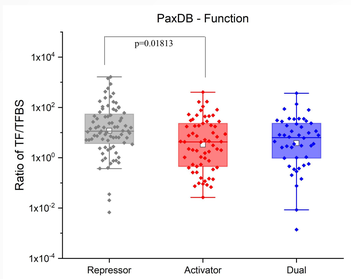


Figure 2: Box plot of PaxDB abundance (log values) based on function. Mean indicated with white square

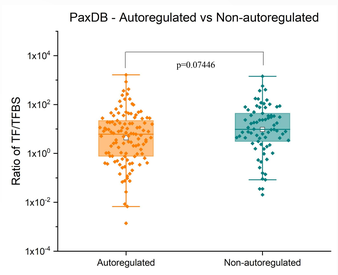


Figure 3: Box plot of PaxDB abundance (log values) based on autoregulation. Mean indicated with white square

Protein Production Rate from Ribosome Profiling

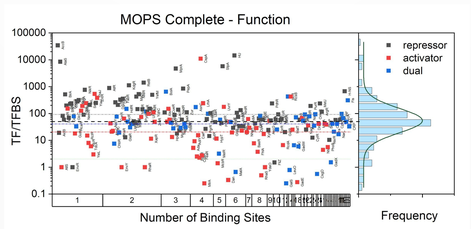


Figure 4: The median ratio of TF concentration (Ribosome Profiling) to TFBS is 49.8

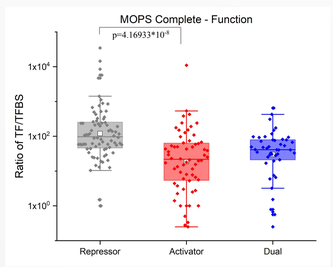


Figure 5: Box plot of ribosome profiling data (log values) grouped by function

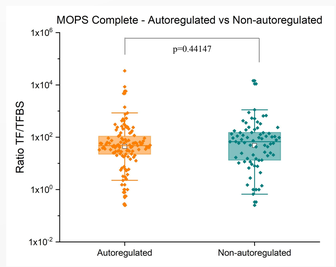


Figure 6: Box plot of ribosome profiling data (log values) grouped by autoregulation

TFBS Identified by High Throughput Methods

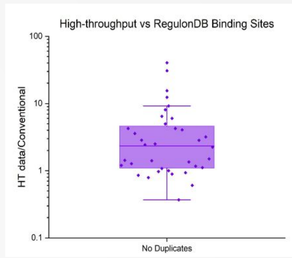


Figure 7: Box plot of the ratio of high throughput data to regulonDB data list. Graph has duplicated transcription factors removed (median used for those with odd numbers of data sets and maximum used for those with 2)

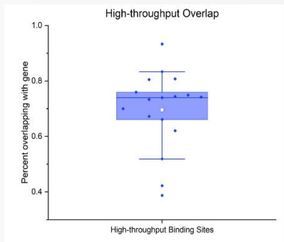


Figure 8: Box plot showing the percent of sites that HT data found that overlapped with the gene. Data are limited since some methods did not produce position results or were not precise enough to give an exact binding site.

Conclusions

- Based on regulonDB data, our data indicate the median ratio of transcription factors to their binding sites is 7.23 for PaxDB and 49.8 for Ribosome Profiling.
- Statistical tests indicate that repressors tend to have a significantly larger ratio compared to either activators or dual function TFs.
- The number of binding sites found in high-throughput methods is double that identified in regulonDB.
- FIMO scans using position-weight matrices of the *E. coli* genome predicted a large number of binding sites, with many located in the ORF region.

References

1. Alberto Santos-Zavaleta, Heladia Salgado, Socorro Gama-Castro, Mishael Sánchez-Pérez, Laura Gómez-Romero, Daniela Ledezma-Tejeida, Jair Santiago García-Sotelo, Kevin Alquicira-Hernández, Luis José Muñiz-Rascado, Pablo Peña-Loredo, Cecilia Ishida-Gutiérrez, David A Velázquez-Ramírez, Víctor Del Moral-Chávez, César Bonavides-Martínez, Carlos-Francisco Méndez-Cruz, James Galagan, Julio Collado-Vides
Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D212?D220, doi: https://doi.org/10.1093/nar/gky1077
2. Charles E. Grant, Timothy L. Bailey, William Stafford Noble, FIMO: scanning for occurrences of a given motif, Bioinformatics, Volume 27, Issue 7, 1 April 2011, Pages 1017–1018, https://doi.org/10.1093/bioinformatics/btr064
3. Ishihama, A., Shimada, T., & Yamazaki, Y. (2016). Transcription profile of Escherichia coli: genomic SELEX search for regulatory targets of transcription factors. Nucleic acids research, 44(5), 2058–2074. https://doi.org/10.1093/nar/gkw051
4. Li, G. W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell, 157(3), 624–635. https://doi.org/10.1016/j.cell.2014.02.033
5. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D., & von Mering, C. (2015). Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. Proteomics, 15(18), 3163–3168. https://doi.org/10.1002/pmic.201400441

Acknowledgements

- Dr. Ann Stock - Principal Investigator
- Dr. Rong Gao - Lab Mentor
- ARESTY leaders and program directors