Computational Analysis of Extended Protein Occupancy Domains in Bacteria

William Myers[1] and Peter L. Freddolino[2,3]

[1]Department of Biology at Rutgers University – Camden; [2]Department of Biological Chemistry and [3]Department of Computational Medicine and Bioinformatics at University of Michigan Ann Arbor

Background

In addition to being potential infectious agents, bacteria are utilized in numerous biotechnological applications across the fields of medicine, research, agriculture, and industry. Optimal bacterial performance in any of these contexts is largely dependent upon phenotypic state, which is a functional profile resulting from transcriptional regulation of the genome. DNA-binding proteins, like transcription factors and nucleoid-associated proteins (NAPs), such as histone-like nucleoid-structuring protein (H-NS), occupy genomic sites to drive transcriptional regulatory networks in bacteria. An interesting feature of large-scale protein occupancy of the bacterial genome is the presence of extended protein occupancy domains (EPODs), which are transcriptionally silent regions, at least one kilobase in length, that are crowded with NAPs such as H-NS [1,2].

The first aim of this work is to investigate the relationship between EPOD formation and DNA methylation in *Escherichia coli*. *E. coli* DNA methylation is involved in a range of processes including DNA mismatch repair, chromosomal replication and structure, and regulation of gene transcription. The DNA adenine methyltransferase (Dam) and DNA cytosine methyltransferase (Dcm), perform almost all of the methyltransferase activity in the *E. coli* genome [3]. Consideration of the potential interactions between DNA methylation and large-scale protein occupancy in the context of gene regulation encourages investigation of changes in EPOD composition of the *E. coli* genome when the genes encoding for Dam and Dcm are deleted.

The second aim of my project is to explore the contribution of EPODs to the regulation of *Vibrio cholerae* virulence genes in response to biofilm-inducing factors. Biofilms are bacterial communities that are self-encased within an extracellular matrix (ECM), comprised of polysaccharides and proteins, that in pathogenic bacteria are a virulence feature that can complicate antibiotic intervention and host immune response [4]. *V. cholerae* is the causative agent of cholera in humans, which causes "rice-water" diarrhea that can lead to severe dehydration and death [5]. A better understanding of transcriptional regulation of biofilm formation could inform new drug targets.

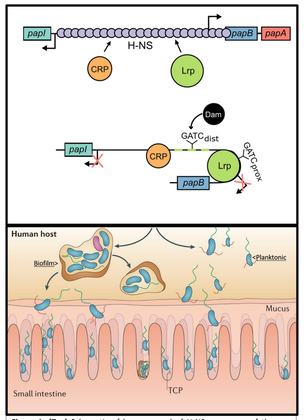

Figure 1: (Top) Schematic of how removal of H-NS occupancy of the *pap* operon in *E. coli* allows methylation of a Dam site at the promoter region of *papB*. Reproduced from [6]. **(Bottom)** Cartoon of planktonic and biofilm-associated *V. cholerae* colonizing the human gut. Reproduced from [7].

Hypothesis/Objective

We hypothesize that EPODs and global protein occupancy can be investigated to inform our understanding of how bacteria regulate transcription in response to conditions of interest.

Generating Protein Occupancy Landscape

Mapping the genomic locations of EPODs can be done through the analysis of large datasets produced by protein-DNA extraction techniques performed under relevant conditions of interest. Such techniques include high resolution *in vivo* protein occupancy display (IPOD-HR), which enriches for total protein occupancy, and RNA polymerase chromatin immunoprecipitation. RNAP occupancy is subtracted from the total protein occupancy signal because changes in RNAP occupancy can mask changes in occupancy of other DNA-binding proteins [6]. The resulting protein occupancy landscape can then be analyzed using computational tools to identify EPODs.

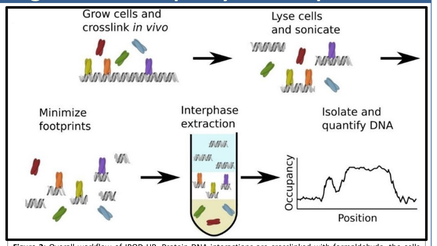

Figure 2: Overall workflow of IPOD-HR. Protein DNA-interactions are crosslinked with formaldehyde, the cells are lysed, and the DNA extract is broken-up by sonication. Interphase extraction is used to isolate protein-bound DNA. This DNA is sequenced and the reads are aligned with the genome and quantified to produce a protein occupancy landscape. Reproduced from [6].

DNA Methylation in *Escherichia coli*

To explore this question, the Freddolino Lab performed IPOD-HR and RNAP-ChIP on wild-type *E. coli* K-12 MG1655 in addition to Δ*dam*, Δ*dcm*, and Δ*dam*Δ*dcm* mutant strains. The resulting reads were aligned and quantified with the *E. coli* genome and then scanned for EPOD locations. Following this, the total protein occupancy, minus RNAP occupancy, at Dam and Dcm target motifs between each strain were compared to determine if the different methylation states present in each strain would affect the protein occupancy at these methylation sites (Fig 3Left). A similar comparison was made based on the representation of methylation sites in EPODs (Fig 3Right).

No Occupancy Changes at Methylation Sites

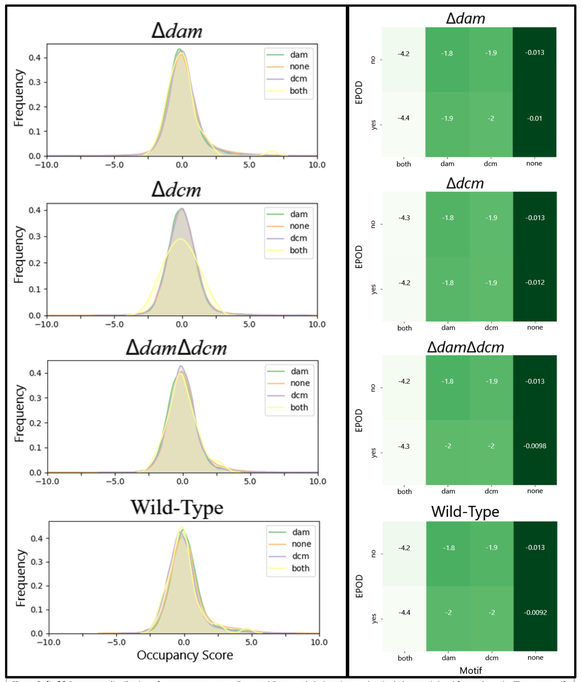

Figure 3: (Left) Frequency distribution of occupancy scores at Dam and Dcm methylation sites at the single base pair level for each strain. The two motifs for each methyltransferase can sometimes overlap, these instances are represented by the "both" distribution. The "none" distribution indicates the frequency of the protein occupancy scores at base pairs associated with neither methylation site. **(Right)** Contingency tables of the presence of EPODs at each methylation site at the single base pair level for each strain. Each value in each row was divided by the sum of the row and then $\log_{10}$ normalized.

Taken together, these results demonstrate that different methylation states of the genome is not globally impacting protein occupancy at the Dam and Dcm methylation sites. This holds to be true even when looking specifically at EPODs. These findings strongly suggest that DNA methylation states in *E. coli* do not impact large-scale protein occupancy.

Biofilm-Inducing Factors in *Vibrio cholerae*

The shift from planktonic to biofilm-associated cells in *V. cholerae* is characterized by global transcriptional changes such as repression of motility genes and induction of genes involved in the synthesis of ECM components [8]. These changes are in response to extracellular and intracellular signals such as high local cell density and the second messenger cyclic-di-GMP [7,8,9]. To investigate how EPODs regulate *V. cholerae* virulence and biofilm-associated genes in response to these factors, wild-type El Tor and a deletion mutant of a transcription factor in the regulon of cyclic-di-GMP (Δ*vpsR*) were grown in M9 defined rich media were exposed to two plasmids. One plasmid induced cyclic-di-GMP overexpression while the other acted as a control to represent the effects of increased cell density. A nitrocellulose filter binding assay was performed to generate total protein occupancy and RNAP-ChIP was performed as well.

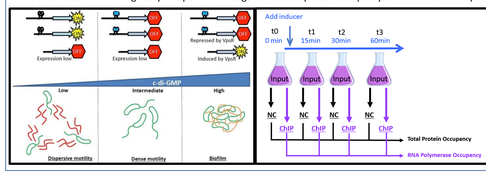

Figure 4: (Left) Schematic demonstrating cyclic-di-GMP levels associated with motility/biofilm state. Reproduced from [9]. **(Right)** Cartoon of experimental design. "Inducer" represents plasmid introduction. "NC" represents nitrocellulose assay. Figure courtesy of Rebecca Hurto, PhD.

Occupancy Changes at Virulence-Associated Operons

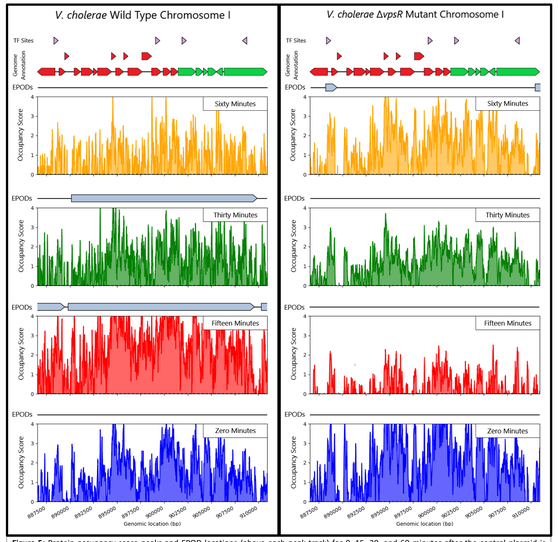

Figure 5: Protein occupancy score peaks and EPOD locations (above each peak track) for 0, 15, 30, and 60 minutes after the control plasmid is introduced. The peaks shown here are from chromosome I of **(Left)** wild-type and **(Right)** Δ*vpsR* mutant *V. cholerae*. The genome annotation track at the top shows the *tcp* (red) and *acf* (green) operons. Transcription factor binding sites (pink) are also shown.

Changes in Global EPOD Locations Between Conditions

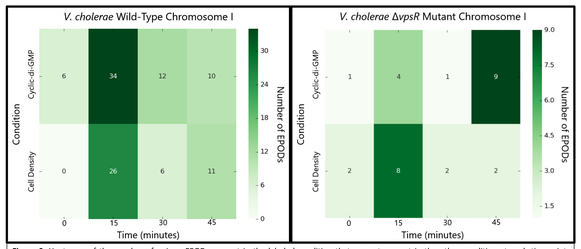

Figure 6: Heatmaps of the number of unique EPODs present in the labeled condition that are not present in the other condition at each timepoint. Shown here is data from chromosome I of **(Left)** wild-type and **(Right)** Δ*vpsR* mutant *V. cholerae*. An example interpretation of these values: **(Left)** at 15 minutes there are 34 EPODs present at various locations when cyclic-di-GMP production is induced that are *not* present at these locations when the control plasmid is introduced.

Discussion for *V. cholerae*

These findings show that the composition of EPODs across the *V. cholerae* genome changes in response to the imposed conditions. Broadly, the EPOD landscape of *V. cholerae* chromosome I differs drastically after 15 minutes of the cells overproducing cyclic-di-GMP compared to when overproduction is not induced (Fig 6). More specifically, EPODs otherwise present at the location of the toxin co-regulated pilus (tcp) operon, which encodes for a type IV pilus that mediates *V. cholerae* adherence to human intestinal cells, and the accessory colonization factor (acf) operon, which encodes for a periplasmic protein that may mediate V. cholerae chemotaxis toward mucin, are not present at this location 15 and 30 minutes after cyclic-di-GMP production is induced (Fig 5) [8,10]. This example of our findings, in addition to other findings not shown here, demonstrates that *V. cholerae* EPOD composition relative to virulence and biofilm-associated genes changes in response to production of the biofilm-inducing second messenger cyclic-di-GMP. This difference being most apparently present at the 15-minute timepoint encourages focus on this timepoint as when *V. cholerae* EPOD composition is changing in response to virulence-associated conditions such as increased production of cyclic-di-GMP.

Future Directions for *V. cholerae*

Similar analysis performed on an IPOD-HR protein occupancy signal should further support my initial findings. In addition to wild-type IPOD-HR data, data on several virulence-associated regulator deletion mutants has also been collected. Critical to supporting my assertion that the presence of EPODs is a silencing feature will be the analysis of RNA-Seq expression data that is currently being produced. For key virulence and biofilm-associated genes that are noticed to be under EPOD regulation, it may prove informative to demonstrate EPOD silencing by inserting a reporter with its own promoter into the EPOD region [1]. Additional follow-up experimentation that would further inform this investigation would be to delete V. cholerae production of NAPs, namely H-NS which comprises most EPODs, and evaluate the EPOD compositions of such NAP mutants under the same experimental conditions [1,2]. Qualitative and quantitative analysis of *V. cholerae* biofilms in such deletion mutants could also support the claim that EPODs are an important regulatory feature of *V. cholerae* biofilm formation.

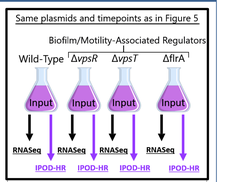

Figure 7: Cartoon of follow-up experimentation that was recently performed. The data from this experimentation will soon be available for analysis.

Conclusion

Overall, this project demonstrates that computationally analyzing features of large-scale protein occupancy such as EPODs can inform investigations into changes in bacterial regulatory logic in response to conditions of interest.

Analysis in the context of *E. coli* DNA methylation concluded that there is no global response in protein occupancy at methylation sites when *E. coli* methyltransferases are deleted. This suggests that DNA methylation likely does not regulate large-scale protein occupancy.

The protein occupancy landscape of the *V. cholerae* genome changed drastically in response to biofilm-inducing conditions, namely increasing cell density and induced production of cyclic-di-GMP. The most notable difference in occupancy peak changes and EPOD location changes happened 15 minutes after the introduction of the plasmid conditions, and some of these changes were associated with virulence operons. Broadly, this investigation into the role of large-scale protein occupancy in regulating the *V. cholerae* genome in response to biofilm-forming conditions could identify the importance of previously overlooked regulators of *V. cholerae* pathogenicity.

Acknowledgments

This work is made possible by support of the Freddolino Lab at University of Michigan (affiliated with Department of Biological Chemistry and Department of Computational Medicine and Bioinformatics), which is funded primarily by National Institutes of Health (NIH).

The Summer Research Opportunity Program (SROP) at University of Michigan Rackham Graduate School supported my work with Dr. Freddolino.

The MARC*USTAR training grant (NIH T34 GM127154) at Rutgers University – Camden also supports this project.

The *Vibrio cholerae* work was in collaboration with the Waters Lab at Michigan State University (Department of Microbiology and Molecular Genetics).

Protein-DNA extractions on *V. cholerae* were performed by Rebecca Hurto, PhD of the Freddolino Lab.

Experimental data on wild-type *E. coli* and the methyltransferase deletion mutants was generated by Haley Amemiya of the Freddolino Lab.

I would like to further recognize Rutgers–Camden MARC Program Directors Dr. Kwangwon Lee and Dr. Nathan Fried as well as my MARC cohorts Harjit Khaira, Shariq Khan, and "Anna" Xiao Hua Liang for their support.

References

1. Freddolino PL, Goss TJ, Amemiya HM, Tavazoie S. Dynamic landscape of protein occupancy across the Escherichia coli chromosome. bioRxiv. 2020 Aug 10;2020.01.29.924811. DOI: https://doi.org/10.1101/2020.01.29.924811
2. Vora T, Hottes AK, Tavazoie S. Protein occupancy landscape of a bacterial genome. Mol Cell. 2009 Jul 31;35(2):247–53. DOI: 10.1016/j.molcel.2009.06.035
3. Marinus MG, Løbner-Olesen A. DNA Methylation. Ecosal Plus. 2014 May;6(1). DOI: 10.1128/ecosalplus.ESP-0003-2013
4. Høiby N, Bjarnsholt T, Givskov M, Molin S, Ciofu O. Antibiotic resistance of bacterial biofilms. International Journal of Antimicrobial Agents. 2010 Apr 1;35(4):322–32. DOI: https://doi.org/10.1016/j.ijantimicag.2009.12.011
5. Clemens JD, Nair GB, Ahmed T, Qadri F, Holmgren J. Cholera. Lancet. 2017 Sep 23;390(10101):1539–49. DOI: 10.1016/S0140-6736(17)30059-7
6. Zamora M, Ziegler CA, Freddolino PL, Wolfe AJ. A Thermosensitive, Phase-Variable Epigenetic Switch: pap Revisited. Microbiology and Molecular Biology Reviews. Jul 2020, 84 (3) e00030-17; DOI: 10.1128/MMBR.00030-17
7. Teschler, J., Zamorano-Sánchez, D., Utada, A. et al. Living in the matrix: assembly and control of Vibrio cholerae biofilms. Nat Rev Microbiol 13, 255–268 (2015). https://doi.org/10.1038/nrmicro3433
8. Silva AJ, Benitez JA. Vibrio cholerae Biofilms and Cholera Pathogenesis. PLoS Negl Trop Dis. 2016 Feb;10(2):e0004330. DOI: 10.1371/journal.pntd.0004330
9. Pursley BR, Maiden MM, Hsieh M-L, Fernandez NL, Severin GB, Waters CM. Cyclic di-GMP Regulates TfoY in Vibrio cholerae To Control Motility by both Transcriptional and Posttranscriptional Mechanisms. J Bacteriol. 2018 Jan 8;200(7):e00578-17; DOI: 10.1128/JB.00578-17
10. Valiente E, Davies C, Mills DC, Getino M, Ritchie JM, Wren BW. Vibrio cholerae accessory colonisation factor AcfC: a chemotactic protein with a role in hyperinfectivity. Sci Rep. 2018 30;8(1):8390. DOI: 10.1038/s41598-018-26570-7

Contacts

William Myers: wjm141@scarletmail.rutgers.edu

Dr. Peter Freddolino: petefred@umich.edu