Whole-genome *in vivo* monitoring of protein-DNA interactions in *Escherichia coli*

Tiffany J. Vora[1] and Saeed Tavazoie[2]

***Short Abstract* — The ability to monitor the occupancy of transcription-factor binding sites *in vivo* is an important prerequisite to modeling and understanding regulatory network dynamics. We have developed a technology for whole-genome identification and monitoring of protein-DNA "footprints" in bacterial genomes. We employed the power of genome-wide analyses to detect underlying complex patterns connecting the set of DNA-protein interactions and the transcription regulatory network. In addition, we are using this technology to gain a better understanding of the structural organization of the bacterial genome in terms of nucleoid proteins and the domains they help to organize.**

***Keywords* — DNA-protein interactions, nucleoid, transcription regulatory network.**

I. Background

DESPITE the importance of DNA-protein interactions in mediating the transcriptional response to environmental signals [1], the systems-wide behavior of the genomic repertoire of these interactions has been largely inaccessible. While the set of genomic binding sites for individual proteins can be determined via chromatin immunoprecipitation (ChIP) followed by microarray hybridization [2], this now-classical technique has several limitations of resolution and global applicability.

While much is known about the organization of eukaryotic genomes, surprisingly little has been recorded about the nature of prokaryotic genomic architectures, despite many years of study. At least two major components contribute to establishing and maintaining genomic architecture in bacteria: protein-DNA interactions and DNA supercoiling. In *E. coli*, 12 conserved proteins have been identified as both binding DNA and contributing to genomic organization [3]. However, the physical mechanism combining these two aspects to create overall genomic architecture remains unknown.

II. Summary of Results

Here we present a technology for the high-resolution simultaneous *in vivo* display of the genomic repertoire of protein-DNA interactions in *E. coli*. We have developed a genome-wide *in vivo* protein occupancy display method that circumvents the need for an antibody, permitting simultaneous high-resolution monitoring of the full repertoire of DNA-protein interactions.

We present evidence for a discrete set of genomic regions that have properties consistent with possible genomic organizing centers in *E. coli*. We show through genomic protein occupancy assays that these regions are coated with proteins. These DNA sequences are conserved through evolution but do not appear to encode proteins that are expressed [4]. Binding sites for the nucleoid proteins are enriched in these regions [5], and the regions show biophysical characteristics consistent with the preferred binding regimes of these proteins [6]. We speculate that interactions among these "bacterial heterochromatin"-like organizing regions may determine the overall structure of the *E. coli* chromosome [7].

III. Conclusion

We have combined novel whole-genome experimentation with computational analysis of diverse data sources [5- to investigate the genome-wide repertoire of protein-DNA interactions in *E. coli*. Using a systems-level analysis, we speculate as to the nature of the underlying organization of the *E. coli* genome. Our experimental method can be extended to any sequenced microbe, and the established platform may be used for the genome-wide study of other molecular events, such as transcription and mutational analysis.

Acknowledgements: This work was supported in part by a NASA Graduate Student Research Program fellowship.

[1]School of Sciences and Engineering, The American University in Cairo, Cairo, Egypt. E-mail: tvora@aucegypt.edu

[2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey. E-mail: tavazoie@genomics.princeton.edu

References

[1] Lee, T. I. et al (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298**, 799-804.

[2] Ren, B. et al (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-9.

[3] Pettijohn, D. E. (1996), "The Nucleoid." In *Escherichia coli* and *Salmonella* (ed. Neidhardt, F.), ASM, Washington DC, pp. 158-166.

[4] Faith, J. J. et al (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS* **5**, e8.

[5] Grainger DC, Hurd D, Goldberg MD, & Busby SJ (2006). Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *NAR* **34**, 4642-52.

[6] Pedersen AG et al (2000). A DNA structural atlas for *Escherichia coli*. *J Mol Biol* **299**, 907-30.

[7] Vora T, Tavazoie S, "*In vivo* protein occupancy display reveals the genomic architecture of *Escherichia coli*," in preparation.