

Minireview

Structural and functional map of a bacterial nucleoid

Agustino Martínez-Antonio*, Alejandra Medina-Rivera† and Julio Collado-Vides†

Addresses: *Departamento de Ingeniería Genética, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato, 36500, México. †Programa de Genómica Computacional, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México.

Correspondence: Julio Collado-Vides. Email: collado@cgc.unam.mx; Agustino Martínez-Antonio. Email: amartinez@ira.cinvestav.mx

Abstract

Genome-wide mapping of transcription factor-DNA interactions in bacterial chromosomes *in vivo* has begun to reveal global zones occupied by these factors that serve two purposes: compacting the bacterial DNA and influencing global programs of gene transcription.

In single-celled organisms such as bacteria economy is critical, including the efficient use of space in the tiny cell. Although gene density in bacterial genomes is high, the chromosomes are still long macromolecules that must be compacted by at least three orders of magnitude to fit into the space available [1,2], and the mechanism of chromosomal packing in bacteria and the proteins involved is a long-standing question. In a recent study published in *Molecular Cell*, Saeed Tavoizie and colleagues (Vora *et al.* [3]) have investigated protein binding across the complete *Escherichia coli* genome and have revealed extended regions of high protein occupancy. Together with other recent studies, this work provides valuable information on the chromosomal organization by DNA-binding proteins in bacteria and will aid understanding of their large-scale effects on gene expression.

Chromosomal size and dynamics in bacteria

A bacterial genome typically comprises a single circular DNA molecule, usually between 1.5 and 10 Mbp in free-living bacteria [4,5], which *in vivo* is packaged with proteins into a distinct structure known as the bacterial nucleoid. The information encoded in one bacterial genome directs all functions necessary to maintain a functional and self-replicating living system, from basic tasks such as nutrient and energy uptake to complex coordinated ones, such as cell division. Initial observations indicated that when DNA is released from lysed bacteria, the space it occupies is four to ten times larger than the cell itself, even though the DNA preserves supercoiled loops [6]. This implied that chromosomes are even more compacted inside the cell, probably by auxiliary proteins [2,7]. In addition to DNA gyrase and DNA topoisomerase I, which maintain supercoiling levels of DNA [6,8], the so-called

nucleoid-associated proteins (NAPs) were proposed to be in charge of most chromosomal remodeling tasks. Among others, Ishihama and colleagues have studied NAPs extensively, and at the end of the 1990s, Ali Azam *et al.* [9,10] found that in cultured cells, each NAP is maximally expressed during specific growth phases.

The regulatory regions of transcription units are located in noncoding DNA sequences where transcription factors and RNA polymerases bind to the DNA to initiate transcription. The bacterial nucleoid structure is natively able to permit transcription, despite the microscopically observed loops and predicted further levels of genome compaction. This is probably due to the fact that the level of compaction is not as restrictive as that of eukaryotic chromatin [11].

Even when transcription is permitted in bacteria, the effects of chromosomal compaction on gene expression are still not clear. Because nucleoid organization can be described on both a physical and a functional basis, these two properties should be analyzed and understood together. Nucleoid topology is strongly related to the binding patterns of NAPs. All the major NAPs, with the exception of Dps (the DNA-binding protein in starved cells), have been found experimentally to have a functional association with the regulation of gene expression. These regulatory NAPs are: Fis (factor for inversion stimulation), HU (histone-like protein), H-NS (histone-like nucleoid structuring protein), and IHF (integration host factor). The concentrations of these proteins vary in different growth phases, from 10,000 to 60,000 monomers per cell, in contrast to local regulators such as LacI, which is present at a maximum of 20 monomers per cell [12]. These observations, together with knowledge of the hierarchy of regulatory networks, have led to the hypothesis of 'analog' and 'digital' components of gene regulation in bacteria. The analog component is represented by the wide influence of superhelical and chromosomal loops (mediated by NAPs) in background regulation, and the digital component by the qualitatively more effective (almost binary) regulation exerted by DNA-binding specific transcription factors [13,14].

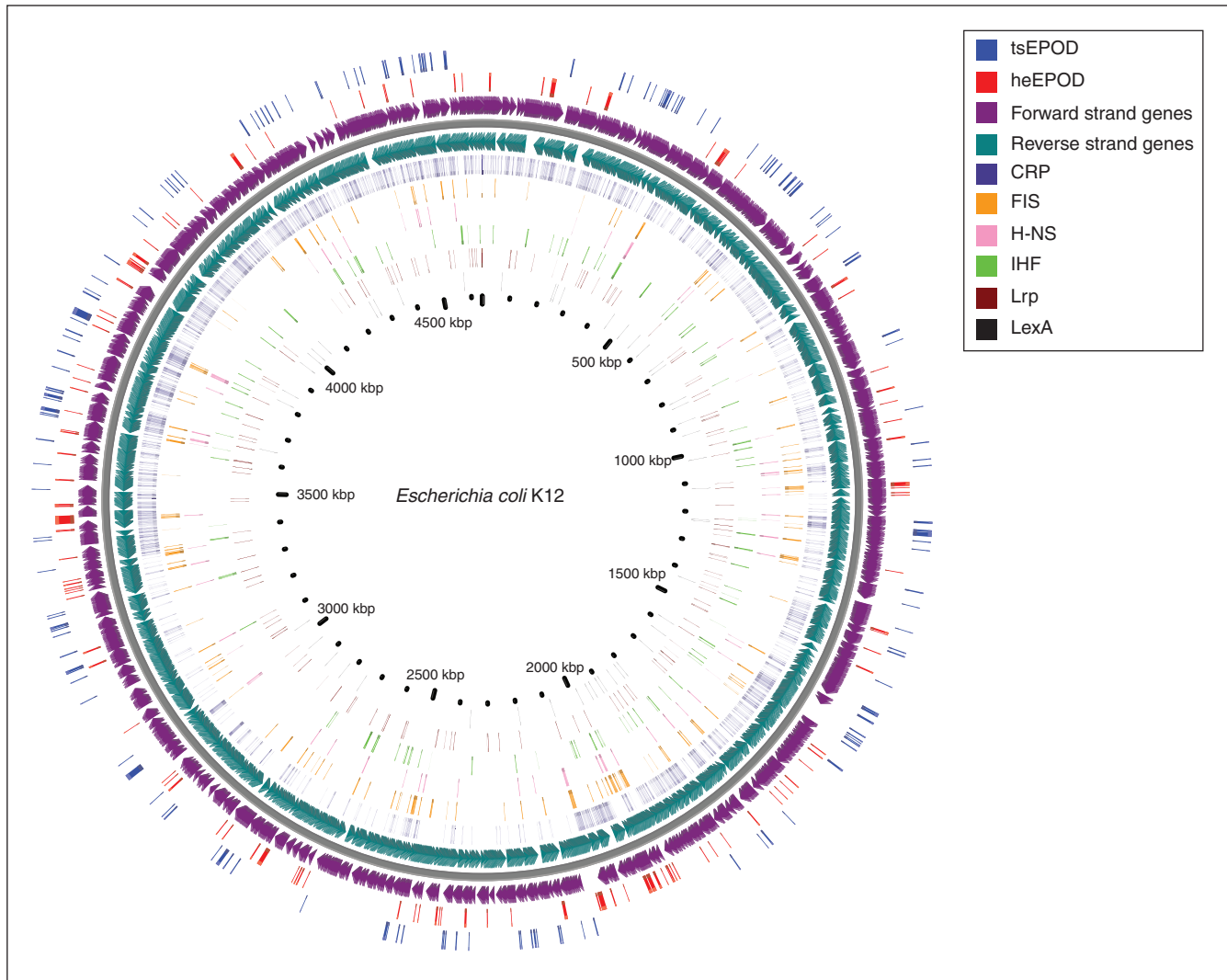


Figure 1

Occupancy profiles of DNA-binding factors on the *Escherichia coli* K12 chromosome (gray line), based on integrated results from three different laboratories. Genes are represented by purple and turquoise arrows on forward and reverse strands, respectively. Outer circles represent the EPODs (domains greater than 1 kb with high protein occupancy) reported by Vora *et al.* [3]. EPODs were clustered using the median expression level across domains obtained from the supplementary material to [3]: heEPODs are red, and tsEPODs are blue. Inner circles represent individual ChIP-chip binding patterns of known transcription factors: CRP [17]; Fis, HN-S, IHF, and Lrp [18]; and LexA [19]. All ChIP-chip data were obtained from the respective supplementary material. The figure was created using the CGview program [24]. CRP, cAMP receptor protein; FIS, factor for inversion stimulation; heEPOD, highly expressed extended protein occupancy domain; H-NS, histone-like nucleoid structuring protein; IHF, integration host factor; Lrp, leucine-responsive protein; LexA, SOS regulatory protein; tsEPOD, transcriptionally silent extended protein occupancy domain.

Genome-wide chromosomal occupancy by DNA-binding proteins

Chromatin immunoprecipitation followed by DNA microarray (ChIP-chip) was developed 10 years ago as a technique for identifying all those sites on the chromosome occupied by a particular DNA-binding protein at a given time [15]. Protein-DNA complexes are purified by precipitation with antibodies against the protein, and the DNA fragments are then separated and analyzed by microarray

to identify the binding sites. In *E. coli*, this technique has been used to determine the binding sites for RNA polymerase, for global transcriptional regulators such as CRP (cAMP receptor protein), Fis, H-NS, IHF and Lrp (leucine-responsive protein), and for some local regulators, such as MelR (melibiose metabolism regulator) and LexA (SOS regulatory protein) (Figure 1) [16–19]. In this way a genome-wide profile of binding sites for transcription factors in DNA is beginning to emerge for *E. coli*.

In their recent study Vora *et al.* [3] aimed at obtaining all the protein-DNA complexes present in *E. coli* at early and late exponential growth phases, respectively. This genome-wide screening methodology is known as *in vivo* protein occupancy display (IPOD). To recover occupied DNA sequences at a high resolution, they obtained short fragments (50 bp) of DNA protected by proteins and then used a high-density tiling array to analyze the DNA. In order to cover the entire *E. coli* genome, the array was composed of overlapping oligomers of 25 bp, designed to locate a DNA fragment at a resolution of 4 bp of genomic DNA.

Vora *et al.* [3] detected 2,063 individual protein-occupied sites, some of which were found in close proximity to each other - forming what the authors call extended protein occupancy domains (EPODs) with lengths ranging from 1 to 14 kbp (Figure 1). They then determined the transcriptional profiles of the EPODs by DNA microarray analysis and found that they fell into two groups - highly expressed (heEPODs) and transcriptionally silent (tsEPODs). Using previous data of Grainger *et al.* [17], who had determined DNA polymerase occupancy in the same growing conditions, Vora *et al.* found that the 121 heEPODs showed high polymerase occupancy whereas the 151 tsEPODs showed lower occupancy. The 121 highly occupied zones included highly expressed genes such as those for ribosomal proteins, while the 151 tsEPODs had a high content of predicted or hypothetical open reading frames that, interestingly, corresponded to transcriptionally silent genes (Figure 1). An extensive search for putative H-NS-, Fis- and IHF-binding sites (available from RegulonDB [20]) in the EPOD sequences indicated that binding sites for these proteins are overrepresented in tsEPODs, whereas only Fis showed overrepresentation of binding sites within heEPODs. This was as expected, as Fis is maximally expressed at the beginning of the exponential growth phase and regulates the transcription of the ribosomal genes, among others. On this basis, Vora *et al.* [3] hypothesize that tsEPODs may comprise the predicted structural organizational center of the bacterial nucleoid, potentially also carrying out the important functional task of repression of silent DNA sequences by H-NS [21,22].

Taking it further

The work of Tavazoie and colleagues [3] opens up the possibility of studying, at a high resolution, the zones of the nucleoid occupied by the entire repertoire of transcription factors. The next step should be to obtain chromosomal occupancy profiles at different growing phases - that is, lag, early, mid, and late exponential and early and late stationary phases. With these data, investigators should be able to obtain a dynamic picture of protein occupancy for NAPs along the different growth phases of a bacterial culture. As each NAP is produced maximally at different growth phases, one would expect that the nucleoid dynamics would be different, influencing the running of global

transcriptional programs within each growth phase - that is, the analog programs [13,14]. In parallel, computational efforts should be made to find all putative binding sites in DNA for the approximately 81 transcription factors in RegulonDB [20] that currently have experimentally annotated binding sites. This will enable determination of the digital control exerted in each growth phase. To reveal the complete picture of the dynamic nucleoid, efforts should be made to characterize the binding sites for the complete repertoire of around 300 transcription factors in the *E. coli* genome. It is intriguing that chromosomal loops, EPODs and the maximal operon size are all around 10 kbp. If not a coincidence, this could reflect the presence in *E. coli* of local supercoiling domains whose boundaries limit coordinated transcription, by analogy with observations in eukaryotes [23].

Acknowledgments

The authors are grateful for the comments of colleagues and reviewers which helped improve the article. AM-R was supported during her PhD studies (Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México) by a fellowship from the Consejo Nacional de Ciencia y Tecnología (Mexico). This work was partially supported by the "Consejos de Ciencia y Tecnología Nacional (102854) y del Estado de Guanajuato" (Young Researcher grants) given to AM-A and CONACYT (103686) and NIH grant number GM071962-06 given to JC-V.

References

1. Krawiec S, Riley M: **Organization of the bacterial chromosome.** *Microbiol Rev* 1990, **54**:502-539.
2. Kellenberger E: **Functional consequences of improved structural information on bacterial nucleoids.** *Res Microbiol* 1991, **142**:229-238.
3. Vora T, Hottes AK, Tavazoie S: **Protein occupancy landscape of a bacterial genome.** *Mol Cell* 2009, **35**:247-253.
4. Perez-Rueda E, Janga SC, Martinez-Antonio A: **Scaling relationship in the gene content of transcriptional machinery in bacteria.** *Mol Biosyst* 2009, **5**:1494-1501.
5. Ochman H, Moran NA: **Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis.** *Science* 2001, **292**:1096-1099.
6. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR: **Topological domain structure of the *Escherichia coli* chromosome.** *Genes Dev* 2004, **18**:1766-1779.
7. Woldringh CL, Jensen PR, Westerhoff HV: **Structure and partitioning of bacterial DNA: determined by a balance of compaction and expansion forces?** *FEMS Microbiol Lett* 1995, **131**:235-242.
8. Travers A, Muskhelishvili G: **DNA supercoiling - a global transcriptional regulator for enterobacterial growth?** *Nat Rev Microbiol* 2005, **3**:157-169.
9. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A: **Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid.** *J Bacteriol* 1999, **181**:6361-6370.
10. Azam TA, Ishihama A: **Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity.** *J Biol Chem* 1999, **274**:33105-33113.
11. Struhl K: **Fundamentally different logic of gene regulation in eukaryotes and prokaryotes.** *Cell* 1999, **98**:1-4.
12. Elf J, Li GW, Xie XS: **Probing transcription factor dynamics at the single-molecule level in a living cell.** *Science* 2007, **316**:1191-1194.
13. Marr C, Geertz M, Hutt MT, Muskhelishvili G: **Dissecting the logical types of network control in gene expression profiles.** *BMC Syst Biol* 2008, **2**:18.

14. Janga SC, Salgado H, Martínez-Antonio A: **Transcriptional regulation shapes the organization of genes on bacterial chromosomes.** *Nucleic Acids Res* 2009, **37**:3680-3688.
15. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA: **Genome-wide location and function of DNA binding proteins.** *Science* 2000, **290**:2306-2309.
16. Grainger DC, Hurd D, Goldberg MD, Busby SJ: **Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome.** *Nucleic Acids Res* 2006, **34**:4642-4652.
17. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ: **Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome.** *Proc Natl Acad Sci USA* 2005, **102**:17693-17698.
18. Cho BK, Barrett CL, Knight EM, Park YS, Palsson BO: **Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*.** *Proc Natl Acad Sci USA* 2008, **105**:19462-19467.
19. Wade JT, Reppas NB, Church GM, Struhl K: **Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites.** *Genes Dev* 2005, **19**:2619-2630.
20. Gama-Castro S, Jiménez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Peñaloza-Spinola MI, Contreras-Moreira B, Segura-Salazar J, Muñoz-Rascado L, Martínez-Flores I, Salgado H, Bonavides-Martínez C, Abreu-Goodger C, Rodríguez-Penagos C, Miranda-Ríos J, Morett E, Merino E, Huerta AM, Treviño-Quintanilla L, Collado-Vides J: **RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation.** *Nucleic Acids Res* 2008, **36**(Database issue):D120-D124.
21. Lang B, Blot N, Bouffartigues E, Buckle M, Geertz M, Gualerzi CO, Mavathur R, Muskhelishvili G, Pon CL, Rimsky S, Stella S, Babu MM, Travers A: **High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes.** *Nucleic Acids Res* 2007, **35**:6330-6337.
22. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC: **Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*.** *Science* 2006, **313**:236-238.
23. Cremer T, Cremer C: **Chromosome territories, nuclear architecture and gene regulation in mammalian cells.** *Nat Rev Genet* 2001, **2**:292-301.
24. Stothard P, Wishart DS: **Circular genome visualization and exploration using CGView.** *Bioinformatics* 2005, **21**:537-539.

Published: 11 December 2009
doi:10.1186/gb-2009-10-12-247
© 2009 BioMed Central Ltd