

Chromatin architecture and gene expression in *Escherichia coli*

Hanni Willenbrock and David W Ussery

Address: Center for Biological Sequence Analysis, Department of Biotechnology, Building 208, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark.

Correspondence: David W Ussery. E-mail: dave@cbs.dtu.dk

Published: 1 December 2004

Genome Biology 2004, **5**:252

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/12/252>

© 2004 BioMed Central Ltd

Abstract

Two recent genome-scale analyses underscore the importance of DNA topology and chromatin structure in regulating transcription in *Escherichia coli*.

Location, location, location

Expression of a gene is in a sense a bit like purchasing a new home - the value is strongly dependent on location. This value is context-dependent: it depends on who your neighbors are and also on the larger geographical picture. Two recent studies have analyzed DNA topology and chromatin structure on a genome-wide scale in *Escherichia coli* [1,2]. Both show that an important factor in determining transcription profiles - when and to what extent a gene is expressed - is the location of the gene within the context of the *E. coli* K-12 chromosome. Whereas this is old news for those who are interested mainly in eukaryotic chromosomes, it is an important concept that has often been overlooked (in our opinion) in bacterial transcriptomics. In eukaryotes, it is well known that there are two types of chromatin: heterochromatin, which remains condensed for the most part throughout the cell cycle and contains few genes, and euchromatin, which, on the other hand, contains gene-rich regions and in some cases clusters of highly expressed genes.

Jeong *et al.* [1] analyzed similarities in the transcriptional activities of *E. coli* genes as a function of their position on the chromosome. An autocorrelation function identified three levels of spatial correlations of expressed genes: short-range (7-16 kilobase-pairs, kb), medium-range (approximately 100 kb) and long-range (over 700 kb). Figure 1 shows the gene-expression data obtained by Jeong *et al.* [1], together with that of Peter *et al.* [2], mapped onto the circular *E. coli* chromosome, with four circles (circles 3-6) corresponding to values obtained from the four experiments of

Jeong *et al.* [1]. They took into account the transcription levels of nearly all genes, although only the more highly expressed genes are visible in Figure 1. Most of the genes in *E. coli* are transcribed around the time of replication [3], and only a small fraction (typically around 10%) of the genes are highly transcribed. These 'clumps' or regions of highly expressed genes can be seen as dark bands in Figure 1, and some of these regions differ in the various experiments. The shortest level of spatial correlation found by Jeong *et al.* [1] corresponds to groups of between 7 and 15 genes that exhibit an apparently coherent transcriptional activity. These groups are larger than operons and are likely to reflect small clusters of co-regulated genes, of between roughly three and five operons (assuming about three genes per operon), including the clusters of highly expressed genes mentioned above. This is the first level of the 'bigger picture' of spatial correlations, and is also the most clearly affected by DNA supercoiling, given that correlations at this level are significantly reduced by the addition of norfloxacin, a gyrase and topoisomerase IV inhibitor (data shown in circle 5 in Figure 1). Having said that, it should also be pointed out that all the correlations, including the longer range ones, were affected by gyrase mutations (circle 6 in Figure 1).

The results reported by Jeong *et al.* [1] are slightly different from previous findings by Sousa *et al.* [4], who looked at the expression of a reporter gene when it was inserted at different positions around the chromosome. Sousa *et al.* [4] found that gene expression varies along the chromosome in a somewhat linear manner, forming a gradient in which the

expressed genes was postulated a very long time ago for the *Drosophila* polytene chromosomes [5]. More recently, there have been two studies on gene expression in human chromosomes that showed clustering of highly expressed genes [6,7]. The topic of chromatin structure and gene expression in eukaryotes has generated considerably more interest (and publications) than in bacteria. In fact, at the time of writing this article, a paper was recently published showing that the 'upstream binding factor' for RNA polymerase I causes the chromatin in mammalian cells to form a more decondensed, open structure, allowing access to the polymerase enzyme for transcription [8]. Although most animals have on the order of a thousand times as much DNA as bacteria, the level of compaction by chromatin is similar in both (about 7000-fold). But it is likely that the DNA compaction is more dynamic in bacteria, because of the higher coding density of the chromosome. Furthermore, transcription and translation are coupled in bacteria, most likely for topological reasons [9]. The long-range correlations found by Jeong *et al.* [1] are consistent with a role for chromatin structure in regulating gene expression in bacteria, showing once again that what is true for elephants can also apply to *E. coli*.

DNA supercoiling and gene expression

More than 20 years ago, it was postulated that supercoiling could be used to regulate gene expression in *E. coli* [10], and about a decade later (before microarray technology was readily available) the influence of supercoiling on the concentration of 88 proteins in *E. coli* was demonstrated [11]. In the recent article by Peter *et al.* [2], the influence of DNA supercoiling on transcription was studied using DNA microarrays to systematically probe the expression profiles of all *E. coli* genes. The authors [2] demonstrated that supercoiling may act as a 'transcription factor' and that it can have either a negative or a positive effect on the transcription of a specific gene. They identified 306 'supercoiling-sensitive genes', and the expression of most of these genes correlates very well with the amount of chromosomal relaxation in each experiment. The fact that most of these supercoiling-sensitive genes were localized in regions of high density 'clumps' that were affected by DNA relaxation agrees well with the findings by Jeong *et al.* [1] that short-range correlations are dependent on negative supercoiling.

The outermost two circles in Figure 1 are based on the data of Peter *et al.* [2] and show the locations of supercoiling-sensitive genes (log *p* values; circle 1) and the correlation with chromosomal relaxation (circle 2). Anti-correlations corresponding to regions where expression decreases upon DNA relaxation were also found. As reported by Peter *et al.* [2], chromosomal regions with significant numbers of supercoiling-sensitive genes generally overlap with regions that are more correlated or anti-correlated with the level of chromosomal relaxation than regions with no supercoiling-sensitive genes.

Some of the chromosomal regions that are mostly correlated with supercoiling overlap with regions showing differential expression patterns among the experimental conditions used by Jeong *et al.* [1]. For example, *gyrA* and *gyrB* at 2.33 megabases and 3.88 megabases on the chromosome, respectively, are highly expressed in DNA-relaxed cells (wild-type cells grown with norfloxacin; circle 6 in Figure 1) but hardly expressed in wild-type cells grown in rich (LB; circle 3) or minimal (M9; circle 4) media. Because of the experimental conditions used in the two studies, however, this picture is expected for the gyrase genes. These genes are known to be sensitive to supercoiling and are involved in maintaining a precise level of supercoiling in the cell. Thus, the inhibition of these proteins is very likely to increase their mRNA expression. Surprisingly, a substantial number of additional genes were also affected by gyrase inhibition, indicating that this change in expression has to be due to the effect that gyrase inhibition has on DNA supercoiling - that is, chromosomal relaxation.

Peter *et al.* [2] also found that supercoiling-sensitive genes whose expression increased upon DNA relaxation were significantly more AT-rich in their upstream and coding regions compared with the corresponding regions of genes not sensitive to supercoiling; the opposite was true for supercoiling-sensitive genes whose expression decreased upon DNA relaxation. This may, however, be due to the fact that AT-rich regions tend to be more curved than AT-poor regions. Supercoiling-sensitive genes may, therefore, be expected to be more AT-rich in upstream regions than genes that are regulated by means other than supercoiling. Nonetheless, these small local variations in upstream regions are not visible on the genome-scale atlas plot (Figure 1, circle 9). Because these supercoiling-sensitive genes are localized to specific regions, one would expect that in some cases a region would appear AT-rich if all of its supercoiling-sensitive genes were significantly AT-rich in their upstream regions.

A bit more context is needed here - at the risk of complicating the picture, there are two additional pieces of information that can help build a clearer picture of what is going on in terms of chromatin structure. The first is DNA curvature and the second is a bit more detail about DNA supercoiling. DNA has sequence-dependent structures, just like proteins, and certain sequences tend to coil in three-dimensional space. These 'DNA curves' are correlated with phased tracts of A residues, and have been found to be localized at the tips of supercoils [12]. The DNA in *E. coli* is known to be supercoiled, and curved DNA (which tends to be AT-rich) can result in the placement of certain DNA sequences at the apical tips of supercoils, as shown in Figure 2. The supercoils can be divided into two types: plectonemic and toroidal, depending on the shape (Figure 2). Roughly half of the supercoils in *E. coli* are toroidal - the DNA is wrapped around proteins and it is 'restrained', although this is transient in bacteria (but permanent in the form of stable nucleosomes in eukaryotes). The other half of the supercoils

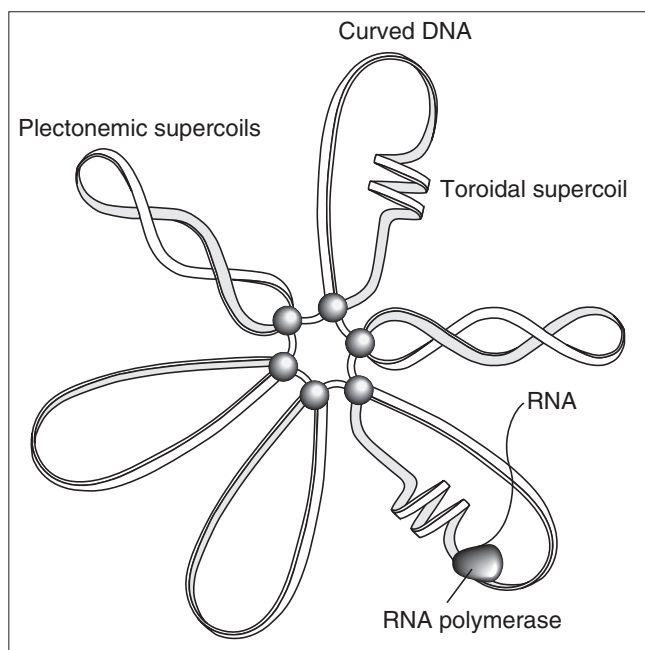


Figure 2
An illustration of DNA supercoiling domains in the *E. coli* chromosome. This is a cartoon of the chromosome; in real life there are perhaps as many as 400 different domains. Plectonemic (unrestrained) and toroidal (restrained, for example by wrapping around a protein) supercoiling is indicated. Curved DNA tends to be localized at the tips of supercoils. The illustration is modified with permission from [23].

are plectonemic (unrestrained) and are under torsional stress, which can be relieved by formation of a bubble in the DNA helix. The ratio between plectonemic and toroidal supercoiling might vary along the chromosome and also with time, because, for example, an RNA polymerase can wrap DNA around it (a restrained toroidal supercoil) and then release the DNA later, creating an unrestrained supercoil. Furthermore, a region that in one set of experimental conditions contains mainly restrained supercoils can suddenly have most of the supercoils become 'free' (plectonemic) in the absence of chromatin proteins.

From a DNA topology perspective, the plectonemic supercoils contain more potential energy, in terms of driving superhelical-dependent transitions (such as melting of the DNA helix). Thus, if there were regions along the chromosome that contained lots of binding sites for proteins involved in chromatin structure, most of the supercoiling would be transiently restrained, and hence less free energy would be available for transcription. In addition, the chromatin proteins can physically block the RNA polymerase from binding to the DNA. Because the *E. coli* chromatin proteins Ihf and Fis show some sequence specificity, it is possible to predict binding sites throughout the chromosome. On a global scale, there tends to be an anti-correlation between these chromatin-binding sites and regions of highly expressed genes [13]. Finally, on the

more local level of a few kilobases (for example, an operon), it is possible to predict regions that tend to exclude chromatin proteins and hence might potentially be highly expressed [14]. In Figure 1, this 'nucleosomal position preference' measure is plotted in circle 8. As expected, regions of low position preference tend to correspond to the regions with highly expressed genes found by Jeong *et al.* [1]. However, the majority of cellular DNA is compacted transiently by chromatin proteins, and there are many regions that are not highly expressed but are nonetheless regulated, with their relative expression levels dependent on supercoiling.

Originally, it was postulated that the chromosome was divided into 12-80 topologically isolated loops, so-called domains, in which chromatin could be relaxed independent of supercoiling in nearby domains [15]. Later this number was estimated more exactly at around 50 domains corresponding to a domain size of approximately 100 kb [16]. Recently, Postow *et al.* [17] presented evidence of an even smaller domain size of approximately 10 kb on average, corresponding to as many as 400 distinct topological domains in *E. coli*. This result corresponds very well with the finding of Jeong *et al.* [1] that up to 16 genes exhibit apparent coherent transcriptional activity and the idea that genes may be organized into confined supercoiled domains with a size of up to 16 kb.

The fact that the genes identified as sensitive to supercoiling have a variety of functions supports the hypothesis that supercoiling may act as a global transcriptional regulatory mechanism and that the cell may use this mechanism as an environmental sensor because the topology of the chromosome may be affected by the surrounding environment. The chromatin protein H-NS regulates many environmental genes, probably through topological changes to DNA [18].

One final aspect of this global view of regulation of transcription at the level of chromatin structure is that some of these environmentally regulated and supercoiling sensitive genes are involved in bacterial pathogenesis. For example, in *Salmonella* it has been shown that expression of genes involved in invasion is regulated by DNA supercoiling [19]. Thus, the global regulation of gene expression by DNA topology could prove to be an important aspect of understanding the mechanisms of bacterial virulence [20].

Acknowledgements

This work was supported by a grant from the Danish Center for Scientific Computing.

References

1. Jeong KS, Ahn J, Khodursky AB: **Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*.** *Genome Biol* 2004, **5**:R86.
2. Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR: **Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*.** *Genome Biol* 2004, **5**:R87.

3. Dworkin J, Losick R: **Does RNA polymerase help drive chromosome segregation in bacteria?** *Proc Natl Acad Sci USA* 2002, **99**:14089-14094.
4. Sousa C, de Lorenzo V, Cebolla A: **Modulation of gene expression through chromosomal positioning in *Escherichia coli*.** *Microbiology* 1997, **143**:2071-2078.
5. Ananiev EV, Gvozdev VA: **Changed pattern of transcription and replication in polytene chromosomes of *Drosophila melanogaster* resulting from eu-heterochromatin rearrangement.** *Chromosoma* 1974, **45**:173-191.
6. Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH: **The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes.** *Genome Res* 2003, **13**:1998-2004.
7. Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA: **Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers.** *Cell* 2004, **118**:555-566.
8. Chen D, Belmont AS, Huang S: **Upstream binding factor association induces large-scale chromatin decondensation.** *Proc Natl Acad Sci USA* 2004, **101**:15106-15111.
9. Gowrishankar J, Harinarayanan R: **Why is transcription coupled to translation in bacteria?** *Mol Microbiol* 2004, **54**:598-603.
10. Smith GR: **DNA supercoiling: another level for regulating gene expression.** *Cell* 1981, **24**:599-600.
11. Steck TR, Franco RJ, Wang JY, Drlica K: **Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins.** *Mol Microbiol* 1993, **10**:473-481.
12. Pavlicek JW, Oussatcheva EA, Sinden RR, Potaman VN, Sankey OF, Lyubchenko YL: **Supercoiling-induced DNA bending.** *Biochemistry* 2004, **43**:10664-10668.
13. Ussery D, Larsen TS, Wilkes KT, Friis C, Worning P, Krogh A, Brunak S: **Genome organisation and chromatin structure in *Escherichia coli*.** *Biochimie* 2001, **83**:201-212.
14. Dlakic M, Ussery D, Brunak S: **DNA bendability and nucleosome positioning in transcriptional regulation.** In: *DNA Conformation in Transcription*. Edited by Ohyama T. Georgetown: Landes Bioscience; 2004.
15. Worcel A, Burgi E: **On the structure of the folded chromosome of *Escherichia coli*.** *J Mol Biol* 1972, **71**:127-147.
16. Sinden RR, Pettijohn DE: **Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling.** *Proc Natl Acad Sci USA* 1981, **78**:224-228.
17. Postow L, Hardy CD, Arsuaga J, Cozzarelli NR: **Topological domain structure of the *Escherichia coli* chromosome.** *Genes Dev* 2004, **18**:1766-1779.
18. Rimsky S: **Structure of the histone-like protein H-NS and its role in regulation and genome superstructure.** *Curr Opin Microbiol* 2004, **7**:109-114.
19. Leclerc GJ, Tartera C, Metcalf ES: **Environmental regulation of *Salmonella typhi* invasion-defective mutants.** *Infect Immun* 1998, **66**:682-691.
20. Dorman CJ: **DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria.** *Infect Immun* 1991, **59**:745-749.
21. Pedersen AG, Jensen LJ, Brunak S, Staerfeldt HH, Ussery DW: **A DNA structural atlas for *Escherichia coli*.** *J Mol Biol* 2000, **299**:907-930.
22. Satchwell SC, Drew HR, Travers AA: **Sequence periodicities in chicken nucleosome core DNA.** *J Mol Biol* 1986, **191**:659-675.
23. Sinden RR: **DNA Structure and Function:** San Diego: Academic Press; 1994.