Genome architecture and global gene regulation in bacteria: making progress towards a unified model?

Charles J Dorman

Department of Microbiology
Moyne Institute of Preventive Medicine
Trinity College Dublin
Dublin 2
Ireland

Tel +353 1 896 2013
Fax +353 1 679 9294
E-mail cjdorman@tcd.ie

Data obtained with advanced imaging techniques, 3C and 5C chromosome conformation capture methods, bioinformatics and molecular genetics, together with insights from polymer physics and mechanobiology, are helping to refine our understanding of the spatiotemporal organization of the bacterial nucleoid and its gene expression programmes. The proposal that gene *regulation* represents one of the organizing principles of nucleoid architecture marks an important development in our understanding of bacterial cell biology.

Introduction

The lack of a nuclear membrane is one of the defining features of prokaryotes. Instead of a nucleus, they have a nucleoid, a macromolecular complex where the genetic material and its associated molecules are located. Although the bacterial nucleoid was first described over fifty years ago[1], detailed information about its structure has become available only quite recently, with very significant progress being made in the past two years[2-10]. Advances in understanding nucleoid structure have relied historically on investigation by electron microscopy accompanied by biochemical and genetic studies. The recent application to bacteria of powerful imaging technologies and physical analytical methods developed for the study of eukaryotic cells has accelerated the rate of discovery of nucleoid structural detail.

It has been known for some time that the circular chromosome within the nucleoid is subject to compaction imposed by a combination of factors that include molecular crowding[11], DNA polymer dynamics[6], supercoiling of the DNA[12] and the interaction of the DNA with other molecules, including nucleoid-associated proteins (NAPs)[13]. In addition to underpinning nucleoid architecture, NAPs and DNA supercoiling also influence gene expression. Indeed, many NAPs were studied initially from the standpoint of gene regulation and the impact of DNA supercoiling on transcription has been recognized for decades[14-16]. Using nucleoid-structuring features such as NAPs and DNA topology to influence gene expression has intuitive appeal because it serves to integrate the process of gene expression with the very cellular structure within which it takes place. Our appreciation of the relationship between the nucleoid and gene expression has been deepened by recent findings suggesting that the processes of efficient gene *regulation* themselves represent a nucleoid structuring principle.

Nucleoid structure and superstructure

The chromosome of the model Gram-negative bacterium *Escherichia coli* is organized at a nanometer-scale structural level and at a micrometer-scale super-structural level[17]. When considering nucleoid organization it is helpful to consider the dimensions of the container: during exponential growth the *E. coli* cell measures approximately two microns in length by one micron in diameter. To put this in context, the *lac* operon, consisting of the three structural genes *lacZ, lacY* and *lacA,* measures about 1.7 microns in length. The single, circular 4,639-kb chromosome, which contains 4,496 genes[18], is itself about a millimeter and a half in circumference and, if opened fully, would have a diameter of 0.5 mm, which is 500 times greater than the diameter of an *E. coli* cell. These dimensions illustrate the extent of the packaging problem associated with the bacterial nucleoid, which is compounded by the need to arrange the DNA in ways

that make it not just compact but also readily available for replication, segregation, gene expression and gene regulation.

Landmark genetic experiments published in 2004 by Boccard and colleagues revealed that the *E. coli* chromosome is composed of six distinguishable zones: four macrodomains (Ori, Left, Right and Ter) and two additional 'non-structured' regions (NS-Left and NS-right)[19] (Fig. 1). This organization imposes certain restrictions on permitted rearrangements to the linear order sequence of the chromosome, perhaps even limiting in some respects its potential for further evolution[20].

The Ori macrodomain consists of the region around the origin of bi-directional chromosomal replication, *oriC,* and the Ter macrodomain is located at the opposite pole of the chromosome and contains the terminus of DNA replication[19,21]. The Ter macrodomain also contains the *dif* site that is crucial for the resolution of chromosome dimers by the XerC/D tyrosine integrase site-specific recombinases[22]. The other four macrodomains make up the bulk of the left and right replichores, the two 'arms' of the chromosome along which DNA polymerase moves during bidirectional chromosome replication[19] [Fig. 1]. In *E. coli,* the *oriC* locus is positioned at mid-cell, with the replichores to the left and right of this point, aligned with the long axis of the bacterium. This arrangement is dependent on MukBEF, the 'Structural Maintenance of Chromosome' (SMC) function of *E. coli*[23]. The Ter macrodomain moves from cell pole to mid-cell in newborn cells and is maintained there through an interaction between the MatP Ter-binding protein and the ZapB component of the cell division apparatus[24].

The MatP DNA binding protein has a special relationship with the Ter macrodomain. In contrast to the widespread distribution of SeqA binding sites around much of the chromosome, the MatP DNA binding protein binds to a sequence motif (*matS*) that is found uniquely within Ter, a macrodomain from which SeqA is excluded[24]. MatP has an important role in timing the separation of daughter chromosomes during cell division: it can hold two copies of the Ter macrodomain together, preventing premature chromosome segregation[25]. MatP also plays a role in forming loops in the Ter macrodomain by holding copies of the *matS* sequence together, providing a degree of compaction to that macrodomain[26,27].

There is an interesting distribution of important cell cycle DNA binding proteins among the macrodomains and these may contribute to their organization[21]. The intensively studied SeqA protein is involved in the timing of chromosome replication initiation and binds to hemi-methylated 5'-GATC-3' sites[28-30]. These sites are abundant at *oriC* but are also found elsewhere on the chromosome, though not in Ter[21,31,32].

The SlmA protein has an important role in coordinating chromosome positioning within the dividing cell; it too has a distribution that corresponds with the macrodomain structure of the chromosome. It binds to a specific DNA sequence that is absent from the Ter macrodomain and infrequently found in the Left and Right domains. Thus, SlmA is mainly located in the Ori macrodomain and its flanking unstructured domains[21,33]. MatP, SeqA and SlmA differ from the NAPs in that they either bind exclusively inside (MatP) or outside (SeqA, SlmA) the Ter macrodomain: most NAPs do not display such strong macrodomain specificity in their binding preferences[13]. This may imply that the NAPs are involved in chromosome organization at a different level to the macrodomain-

specific proteins. This conjecture is supported by several lines of evidence that link at least some NAPs to the maintenance of nucleoid structure at the microdomain level. Also, MatP and SlmA have not (so far) been reported to influence global gene expression, whereas many NAPs have.

Superimposed on its macrodomain structure is the organization of the chromosome at the level of looped DNA microdomains[34]. In the earliest studies of nucleoid structure, the number and size of the looped domains were respectively under- and overestimated[35,36]. Much more accurate estimates were obtained subsequently with (1) genetic methods that allowed the impact of looped domain breakage on DNA supercoiling to be measured around the chromosome[37], (2) measurements of the influence of the topological barriers at microdomain boundaries on site-specific recombination efficiency[38] and (3) through a systematic examination of looped domains in electron microscopic images of chromosomes released from lysed cells[37]. It now seems that the chromosome of *E. coli* is divided into approximately 400 looped microdomains, each with an average circumference of 10-to-12-kb[37,38]. Microdomains may be both transient and predominantly a feature of chromosomes in exponentially growing bacteria[38]. Nucleoid structure seems to be more diffuse in slow-growing bacteria[10].

Several lines of evidence from physical and genetic studies indicate that nucleoid associated proteins, especially H-NS and Fis, play a role in forming the boundaries of microdomains, where they act as insulators or 'domainins'[32]. Further support for H-NS as a domainin has come from super-resolution imaging combined with chromosome conformation capture ('3C') technology[4]. The 3C data show that H-NS has the ability to enhance the frequency of colocation of known H-NS binding sites from different positions around the chromosome. The imaging data reveal that H-NS forms, on average, two prominent foci within the nucleoid that are consistent with the clustering of many microdomain boundaries[4] (Fig. 1). The most straightforward interpretation of the imaging data is that H-NS is involved in organizing the DNA in each of the two replichores which, like the H-NS foci, align with the long axis of the *E. coli* cell [Fig. 1]. It may do this by bringing together hundreds of H-NS binding sites along an H-NS scaffold. How robust this structure is and to what extent its integrity alters with growth conditions and growth phase is currently unclear. The 3C data suggest some preference for intra-replichore contacts among H-NS binding sites, although there are also examples of contacts across replichores. This could mean that the H-NS-dependent DNA contacts introduce a degree of stochasticity into nucleoid architecture, perhaps generating variation in the gene expression patterns of individual members of the genetically identical bacterial population. This is because the DNA to which H-NS is bound is, in many cases, simultaneously in use to modulate transcription[39,40]. It is interesting to note that StpA, a closely related paralogue of H-NS, does not form these large foci but is instead scattered throughout the nucleoid[4]. The significance of these differences in distribution is unknown but may reflect a preference on the part of StpA for binding to RNA and/or its preferential interaction with a portion of the H-NS population that is not involved in focus formation.

The negatively-supercoiled nature of the chromosome itself represents an organizing influence, not least because it contributes to DNA compaction[12,41] (Fig. 1). The prokaryotic type II topoisomerase DNA gyrase introduces negative

supercoiling using the energy of ATP to drive the reaction[41-44]. In addition, the movement of the polymerases involved in transcription and DNA replication creates local domains of negatively supercoiled and relaxed DNA as they unwind the duplex[41,45]. DNA supercoiling affects transcription efficiency on several levels and so serves as a further integrating factor for gene expression and nucleoid structure[41-47]. Transcription has been proposed as a nucleoid structuring principle in its own right, based on observations that genes subject to high rates of transcription gather together to form foci[48]. Among the chromosomal genes reported to form foci are the ribosomal RNA operons whose promoters are controlled by NAPs, DNA negative supercoiling and the signaling molecule involved in the stringent response, guanosine tetraphosphate[16, 48]. However, how these *rrn* operons are organized in the absence of high transcription rates is unclear so it may be premature to conclude that high rates of transcription *per se* create *rrn* foci. On the other hand it has also been reported that plasmids with constitutive transcription units gather at the cell pole but remain randomly distributed in the absence of transcription[49]. While recognizing that plasmids lie outside the nucleoid, this work does suggest that transcription can lead to gene repositioning and has the potential to influence nucleoid structure. Perhaps it would not be surprising if the need to *regulate* transcription had the potential to influence nucleoid structure.

Gene regulation then and now

Early gene regulation studies involving model bacterial systems (e.g. *lac* in *E. coli*) strongly informed opinion about the likely mechanisms used to regulate the other genes in the cell[50]. Although a variety of mechanisms became apparent relatively quickly, the scene was dominated by the concepts of *trans*-acting protein-mediated repression or activation of transcription initiation, and the fact that regulatory genes were often located adjacent to the promoters that their protein products controlled. Placing two or more structural genes in an operon facilitated regulation of co-located genes by a single regulatory protein[51]. Work with the cyclic-AMP-dependent catabolite repressor protein, cAMP-Crp, showed that one regulatory protein could affect the expression of a multitude of genes/operons and that these could be located at many different chromosomal positions[52,53]. Insights of this type were central to our understanding of the molecular processes underlying the concept of the 'regulon', a collective of (usually geographically-dispersed) genes under the control of a common regulatory factor[51]. cAMP-Crp is an abundant protein that binds to a vast number of potential sites in the chromosome, far more than can be accounted for by binding to cAMP-Crp-dependent promoters alone[54]. This has led to the interesting suggestion that cAMP-Crp has at least as much in common with NAPs as it does with conventional transcription factors[55]. However, it is also possible that this reflects the need to express at a relatively high level those proteins that have geographically dispersed targets in the folded chromosome in the nucleoid (Fig. 2).

In this context it is interesting to note that Montero Llopis and colleagues[3] have used fluorescence *in situ* hybridization microscopy to monitor diffusion of labelled mRNA molecules expressed by the *groESL* and *creS* genes in *Caulobacter crescentus* and by the *lacZ* gene in *E. coli*. Their analysis showed that mRNA translation occurs close to its DNA template. This discovery of *de facto*

compartmentalization in bacteria has important consequences for our view of the efficacy of *trans*-acting factors in gene regulation, not least because it raises questions about the ability of geographically dispersed genes to communicate with one another. Up-regulating *groESL* expression by heat shocking *C. crescentus* cells increases the dispersion of *groESL* mRNA, showing that transcripts from highly expressed genes can migrate from the site at which they are born. In this context it is interesting to note that a correlation exists between the level of expression of a regulatory gene and the number of genes it controls[56]. Thus, the need to bring regulatory genes and their targets within effective range for physiologically meaningful regulation to occur may be a principle driving gene co-location in the nucleoid[57]. (Perhaps small regulatory RNA [sRNA] is more diffusible and thus provides a means of regulation-at-a-distance that is more effective that protein-based mechanisms?)

Képès and colleagues have suggested that periodic arrangements of genes along a solenoidally-arranged chromosome provides a means of facilitating communications between genes and/or their products[58,59]. Bioinformatic analysis of over 100 bacterial genomes has identified statistically correlated gene pairs that tend both to co-occur and to co-locate[60]. In *E. coli,* these genes-in-a-pair are separated by multiples of 117 kb along the chromosome, leading to the suggestion that much of the chromosome is arranged in a helix-like structure with a 117-kb periodicity that facilitates the close alignment of these genes. Furthermore, these paired genes are associated with the most heavily transcribed regions of the genome[57]. Helical phasing of genes within each replichore would facilitate communication between genes and their products and the need to accommodate this spatial co-location is likely to represent a strong organizing principle within the architecture of the nucleoid [Fig. 2]. Such phasing could be achieved equally well by a solenoid-like structure or by a plectonemic (i.e. braid-like) arrangement of the chromosomal DNA[7,61] (Figs. 1, 2).

Transcriptomic signal processing analysis has been used to examine the co-expression of clusters of genes along the *E. coli* chromosome[62,63]. Based on data from their study, Khodursky and colleagues proposed three levels of transcriptional spatial organization: short range (up to 16 kb), medium range (100-125 kb) and long range (600-800 kb)[62]. It is interesting to note that these correspond in scale, respectively, to the proposed sizes of chromosomal microdomains (10-12 kb), the helix with 117-kb periodicity and macrodomains[37,38,58-60,62]. In contrast, Mathelier and Carbone[63] detected a periodicity of 33 kb in addition to the 117 kb value. These findings illustrate the importance of testing the *in vivo* importance of this periodicity issue experimentally.

Data obtained from imaging or chromosome conformation capture experiments have indicated that the arms of the nucleoid are interwound in *B. subtilis*[64]*, C. crescentus*[2] and *E. coli*[10]. Nucleoidal writhing can be expected to influence gene-gene proximity at a level above that of the periodic solenoid structure (Fig. 1). Relocating *parS* in *C. crescentus* (a feature required for chromosome segregation but lacking in *E. coli*) changes the large scale folding of the nucleoid without noticeably changing gene expression[2]. However, such alterations to the gross folding of the nucleoid may not have an impact at the small and intermediate scales where gene-gene communication might influence nucleoid architecture and *vice versa*.

The order of the genes along the *E. coli* chromosome is remarkably similar to that in *Salmonella enterica,* even though these two species separated from their common ancestor about 100 million years ago[65]. Such conservation is indicative of an underlying structure/function imperative and Sobetzko *et al*. have considered the conserved gene order in the context of the macrodomain structure of the *E. coli* nucleoid[5]. They note that genes coding for *trans*-acting transcription factors are typically found in the same replichore as their targets, an observation that is in keeping with the need for colocation of regulatory genes and their regulatory subjects. In contrast, genes contributing to the same process (e.g. ribosomes production) are distributed in both replichores but at comparable distances from *oriC*[5]. This may indicate a need to place genes at corresponding points along a putative DNA topological gradient in each replichore. ChIP-chip data reveal that gyrase binding sites are more abundant at the Ori pole than at the Ter pole of the chromosome and that this may result in a gradient of negative supercoiling extending from Ori to Ter[5,62]. However, it might simply reflect a need to maintain supercoiling set points in zones of the chromosome having different levels of transcriptionally-induced supercoiling, with average superhelical density around the chromosome being similar, at least under some growth conditions. The suggestion that the Ter region, which lies at the periphery of the nucleoid in bacteria growing in a M9 growth medium[66], might be in a more relaxed state is consistent with data from modelling that indicates that Ter has a low level of topological complexity[7,57]. On the other hand, the condition of the Ter macrodomain may be a product of the growth conditions used in the experiments were it was analyzed: rapidly-growing bacteria have a more sharply delineated nucleoid structure than bacteria in stationary phase[10]. A systematic study of nucleoid architecture in the context of growth phase would be very helpful in resolving this issue.

The fourth dimension

Bacterial physiology changes as the organism passes through successive growth stages. Following its introduction into fresh liquid medium in batch culture, the microbe spends a period of time in lag phase adjusting to its new environment before growing exponentially. Then the bacterial population expands at its maximal rate in the new environment until some vital component becomes limiting, causing a transition to stationary phase when rapid growth ceases (Fig. 3). Cell composition and nucleoid architecture change throughout these growth stages. Patterns of gene regulation are dynamic, resulting in sequential changes to global gene expression.

DNA in lag- and stationary-phase bacteria has a lower superhelical density than that of log-phase bacteria[47], reflecting shifts in the ratio of the concentration of ATP to ADP, and its impact on the ATP-dependent DNA supercoiling activity of DNA gyrase[42,44]. Increased compaction of the nucleoid in stationary phase bacteria could be achieved by binding of NAPs, especially Dps[67] (Fig. 3). The differential sensitivities of the two principal sigma factors of RNA polymerase to DNA supercoiling imposes a temporal control on their activities: that of RpoD is confined to periods of relatively high chromosomal supercoiling while RpoS becomes dominant as DNA relaxes[46] (Fig. 3).

The NAP population also shows a dynamic expression pattern[68] that is a function of the growth cycle[13,69] (Fig. 3). Fis and HU alpha are expressed early,

HU beta and the two subunits of IHF appear in exponential phase with IHF and Lrp peaking at the exponential-to-stationary-phase transition; the Dps protein is maximally expressed in stationary phase and the H-NS protein is expressed at a constant ratio to chromosome copy number throughout growth[13] (Fig. 3). There is a (very) rough correspondence between the proximity of NAP genes to OriC and the period at which they are expressed during growth[5]. Since these genes influence the expression of many others, nucleoid structural determinants have a profound and widespread impact on the global gene expression profile of the cell. This is an integrating principle that links environment, gene regulation and nucleoid structure.

Implications of a unified model

The structure of the bacterial nucleoid is subject to constraints at a number of levels because the chromosome has to be configured for optimal rates of accurate replication and segregation while accommodating the complex gene expression programmes that support the life of the cell and because any chromosomal configuration must be compatible with the volume available to house it in the bacterium. Using DNA topology and NAPs simultaneously to modulate both gene expression and nucleoid architecture allows for their integration. However, the associated folding of the genetic material in the nucleoid constrains the free movement of gene products, creating an imperative to co-locate certain genes to facilitate communication. Co-location can be achieved linearly or by exploiting periodicity in nucleoid architecture to ensure that genes remain within regulatory range of one another. Gene expression level and timing, themselves subject to regulation, can overcome the compartmentalization problem to some extent, allowing a given regulatory gene to exert influence at a distance.

To what extent is this apparently well-integrated nucleoid structure capable of further evolution? Bacterial chromosomes can undergo substantial rearrangement without incurring lethality[2,70] and the horizontal transfer and integration of novel genes is a routine event in many bacterial populations[71]. This suggests that nucleoid architecture has the scope to adapt to modifications, with the final arbiter of success being the manner in which the novel form affects the competitive fitness of the microbe.

References

1. Kellenberger, E., Ryter, A. & Séchaud, J. Electron microscope study of DNA-containing plasms. II. Vegetative and mature phage DNA as compared with normal bacterial nucleoids in different physiological states. *J. Biophys. Biochem. Cytol.* **4,** 671-678 [1958].
2. Umbarger, M. A., et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* **44,** 252-264 [2011].
3. Montero Llopis, P., *et al*. Spatial organization of the flow of genetic information in bacteria. *Nature* **466,** 77-82 (2010).
4. Wang, W., Li, G.-W., Chen, C., Xie, X. S. & Zhuang, X. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science* **333,** 1445-1449 (2011).
5. Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc. Natl. Acad. Sci. USA* **109,** E42-E50 (2012).
6. Pelletier, J. *et al.* Physical manipulation of the *Escherichia coli* chromosome reveals its soft nature. *Proc. Natl. Acad. Sci. USA* **109,** E2649-E2656 (2012).
7. Wiggins, P. A., Cheveralls, K. C., Martin, J. S., Lintner, R. & Kondev, J. Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc. Natl. Acad. Sci. USA* **107,** 4991-4995 (2010).
8. Scolari, V. F., Bassetti, B., Sclavi, B. & Lagomarsino, M. C. Gene clusters reflecting macrodomain structure respond to nucleoid perturbations. *Mol. BioSyst.* **7,** 878-888 (2011).
9. Benza, V. G. *et al*. Physical descriptions of the bacterial nucleoid at large scales, and their biological implications. *Rep. Prog. Phys.* **75,** 076602 (2012).
10. Hadizadeh Yazdi, N., Guet, C. C., Johnson, R. C. & Marko, J. Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. *Mol. Microbiol.* **86,** 1318-1333 (2012).
11. de Vries, R. DNA condensation in bacteria: interplay between macromolecular crowding and nucleoid proteins. *Biochimie* **92,** 1715-1721 (2010).
12. Trun, N. J. & Marko, J. Architecture of a bacterial chromosome. *Am. Soc. Microbiol. News* **64,** 276-283 (1998).
13. Dillon, S. C., & Dorman, C. J. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Rev. Microbiol.* **8,** 185-195 (2010).
14. Balke, V. L. & Gralla, J. D. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli. J. Bacteriol.* **169,** 4499-4506 (1987).
15. Dorman, C. J. & Corcoran, C. P. Bacterial DNA topology and infectious disease. *Nucleic Acids Res.* **37,** 672-678 (2009).
16. Travers, A. & Muskhelishvili, G. DNA supercoiling – a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* **3,** 157-169 (2005).
17. Dame, R. T., Espéli, O., Grainger, D. C. & Wiggins, P. A. Multidisciplinary perspectives on bacterial genome organization and dynamics. *Mol. Microbiol.* **86,** 1023-1030 (2012).

18. Blattner, F. R., *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453-1462 (1997).
19. Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.* **23,** 4330-4341 (2004).
20. Esnault, E., Valens, M., Espéli, O. & Boccard, F. Chromosome structuring limits genome plasticity in *Escherichia coli. PLoS Genet.* **3,** e226 (2007).
21. Dame, R. T., Kalmykowa, O. J. & Grainger, D. C. Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in Gram-negative bacteria. *PLoS Genet.* **7,** e1002123 (2011).
22. Lesterlin, C., Barre, F. X. & Cornet, F. Genetic recombination and the cell cycle: what we have learned from chromosome dimers. *Mol. Microbiol.* **54,** 1151-1160 (2004).
23. Badrinarayanan, A., Lesterlin, C., Reyes-Lamothe, R. & Sherratt, D. The *Escherichia coli* SMC complex, MukBEF, shapes nucleoid organization independently of DNA replication. *J. Bacteriol.* **194,** 4669-4676 (2012).
24. Espéli, O. *et al.* A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli. EMBO J.* **31,** 3198-3211 (2012).
25. Thiel, A., Valens, M., Vallet-Gely, I., Espéli, O. & Boccard, F. Long-range chromosome organization in *E. coli*: a site-specific system isolates the Ter Macrodomain. *PLoS Genet.* **8,** e1002672 (2012).
26. Mercier, R., Petit, M. A., Schbath, S., Robin, S., El Karoui, M., Boccard, F. & Espéli, O. The MatP/*matS* site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell* **135,** 475-485 (2008).
27. Dupaigne, P., Tonthat, N. K., Espéli, O., Whitfill, T., Boccard, F. & Schumacher, M. A. Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chromosome. *Mol. Cell* **48,** 560-571 (2012).
28. Bach, T., Krekling, M. A. & Skarstad, K. Excess SeqA prolongs sequestration of *oriC* and delays nucleoid segregation and cell division. *EMBO J.* **22,** 315-323 (2003).
29. Lu, M., Campbell, J. M., Boye, E. & Kleckner, N. SeqA: a negative modulator of replication initiation in *E. coli. Cell* **77,** 413-426 (1994).
30. von Friesleben, U., Rasmussen, K. V. & Schaechter, M. SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli. Mol. Microbiol.* **14,** 763-772 (1994).
31. Sánchez-Romero, M. A. *et al.* Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12*. mBio* **1,** e00012-10 (2010).
32. Hardy, C. D. & Cozzarelli, N. R. A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol. Microbiol.* **57,** 1636-1652 (2005).
33. Tonthat, N. K., *et al.* Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *EMBO J.* **30,** 154-164 (2011).
34. Pettijohn, D. E. Structure and properties of the bacterial nucleoid. *Cell* **30,** 667-669 (1982).

35. Sinden, R. R. & Pettijohn, D. E. Chromosomes in living *Escherichia coli* are segregated into domains of supercoiling. *Proc. Natl. Acad. Sci. USA* **78,** 224-228 (1981).
36. Worcel, A. & Burgi, E. On the structure of the folded chromosome of *Escherichia coli. J. Mol. Biol.* **71,** 127-147 (1972).
37. Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* **18,** 1766-1779 (2004).
38. Deng, S., Stein, R. A. & Higgins, N. P. Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.* **57,** 1511-1521 (2005).
39. Noom, M. C., Navarre, W. W., Oshima, T., Wuite, G. J. & Dame, R. T. H-NS promotes looped domain formation in the bacterial chromosome. *Curr. Opin. Biol.* **17,** R913-914 (2007).
40. Dillon, S. C., Cameron, A. D. S., Hokamp, K., Lucchini, S., Hinton, J. C. & Dorman, C. J. Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Mol. Microbiol.* **76,** 1250-1265 [2010].
41. Sinden, R. R. *DNA Structure and Function*. (Academic Press, San Diego, 1994).
42. Basu, A., Schoeffler, A. J., Berger, J. M. & Bryant, Z. ATP binding controls distinct structural transitions of *Escherichia coli* DNA gyrase in complex with DNA. *Nat. Struct. Mol. Biol.* **19,** 538-546 (2012).
43. Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z. & Higgins, N. P. Rates of DNA gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet.* **8,** e1002845 (2012).
44. van Workum, M. *et al.* DNA supercoiling depends on the phosphorylation potential in *Escherichia coli. Mol. Microbiol.* **20,** 351-360 (1996).
45. Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* **84,** 7024-7027 (1987).
46. Bordes, P., Conter, A., Morales, V., Bouvier, J., Kolb, A. & Gutierrez, C. DNA supercoiling contributes to disconnect σ^S accumulation from σ^S-dependent transcription in *Escherichia coli. Mol. Microbiol.* **48,** 561-571 (2003).
47. Cameron, A. D. S. & Dorman, C. J. A fundamental regulatory mechanism operating through OmpR and DNA topology controls expression of *Salmonella* pathogenicity islands SPI-1 and SPI-2. *PLoS Genet.* **8,** e1002615 (2012).
48. Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L. & Jin, D. J. Active transcription of RNA operons condenses the nucleoid in *Escherichia coli*: examining the effect of transcription on nucleoid structure in the absence of transertion. *J. Bacteriol.* **191,** 4180-4185 (2009).
49. Sánchez-Romero, M. A., Lee, D. J., Sánchez-Morán, E. & Busby, S. J. Location and dynamics of an active promoter in *Escherichia coli* K-12. *Biochem. J.* **441,** 481-485 (2012).
50. Brock, T. D. *The Emergence of Bacterial Genetics* (Cold Spring Harbor Laboratory Press, New York, 1990).
51. Maas, W. K. & McFall, E. Genetic aspects of metabolic control. *Annu. Rev. Microbiol.* **18,** 95-110 (1964).

52. Busby, S. J. Positive regulation in gene expression. In *Regulation of Gene Expression – 25 Years On* (ed. Booth, I. R. & Higgins, C. F.) 51-77 (Cambridge University Press, Cambridge, 1986).
53. Magasanik, B. & Neidhardt, F. C. Regulation of carbon and nitrogen utilization. In Escherichia coli *and* Salmonella, *Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1318-1325 (American Society for Microbiology Press, Washington D. C. 1987).
54. Grainger, D. C., Hurd, D., Harrison, M., Holdstock, J. & Busby, S. J. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc. Natl. Acad. Sci. USA* **102,** 17693-17698 (2005).
55. Browning, D. F., Grainger, D.C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* **13,** 773-780 (2010).
56. Janga, S. C., Salgado, H. & Martinez-Antonio, A. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res.* **37,** 3680-3688 (2009).
57. Fritsche, M., Li S., Heermann, D. W. & Wiggins, P. A. A model for *Escherichia coli* chromosome packaging supports transcription factor-induced DNA domain formation. *Nucleic Acids Res.* **40,** 972-980 (2011).
58. Képès, F. Periodic transcriptional organization of the *E. coli* genome. *J. Mol. Biol.* **340,** 957-964 (2004).
59. Junier, I., Hérison, J. & Képès, F. Genomic organization of evolutionarily correlated genes in bacteria: limits and strategies. *J. Mol. Biol.* **419,** 369-386 (2012).
60. Wright, M. A., Kharchenko, P., Church, G. M. & Segrè, D. Chromosomal periodicity of evolutionarily conserved gene pairs. *Proc. Natl. Acad. Sci. USA* **104,** 10559-10564 (2007).
61. Hong, S.-H. et al. *Caulobacter* chromosome *in vivo* configuration matches model predictions for a supercoiled polymer in a cell-like confinement. *Proc. Natl. Acad. Sci. USA* **110,** 1674-1679 (2013).
62. Jeong, K. S., Ahn, J. & Khodursky, A. B. Spatial patterns of transcription activity in the chromosome of *Escherichia coli. Genome Biol.* **5,** R86.1-R86.10 (2004).
63. Mathelier, A. & Carbone, A. Chromosomal periodicity and positional networks of genes in *Escherichia coli. Mol. Syst. Biol.* **6,** 366 (2010).
64. Berlatzky, I.A., Rouvinski, A. & Ben-Yehuda, S. Spatial organization of a replicating bacterial chromosome. *Proc. Natl. Acad. Sci. USA* **105,** 14136-14140 [2008].
65. Groisman, E.A. & Ochman, H. How *Salmonella* became a pathogen. *Trends Microbiol.* **5,** 343-349 (1997).
66. Meile, J.C. *et al.* The terminal region of the *E. coli* chromosome localizes to the periphery of the nucleoid. *BMC Microbiol.* **11,** 1-28 (2011).
67. Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimoni, E., Wolf, S. G. & Minsky, A. Nucleoid restructuring in stationary phase bacteria. *Mol. Microbiol.* **2,** 395-405 (2004).
68. Luijsterburg, M. S., Noom, M. C., Wuite, G. J. & Dame, R. T. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.* **156,** 262-272 (2006).

69. Martínez-Antonio, A., Lomnitz, J. G., Sandoval, S., Aldana, M. & Savageau, M. A. Regulatory design governing progression of population growth phases in bacteria. *PLoS One* **7,** e30654 (2012).
70. Andersson, D. I. & Hughes, D. Gene amplification and adaptive evolution in bacteria. *Annu. Rev. Genet.* **43,** 167-195 (2009).
71. Syvanen, M. Evolutionary implications of horizontal gene transfer. *Annu. Rev. Microbiol.* **46,** 341-358 (2012).

Acknowledgements
I am grateful to RL Gourse and RT Dame for very helpful discussions and to the three anonymous reviewers for insightful comments. I also thank N Ní Bhriain for comments on the manuscript and MJ Dorman for assistance with computation. This work was supported by a grant from Science Foundation Ireland.

Competing interests statement
The author declares no competing financial interests.

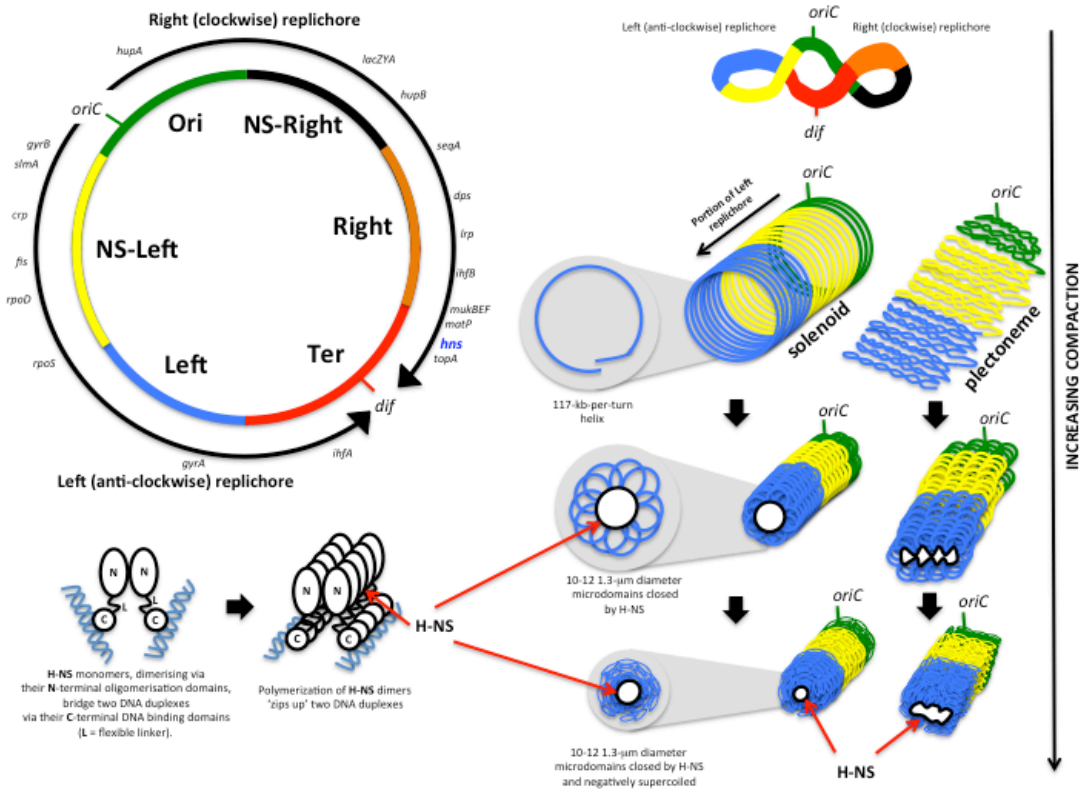

Figure 1 | **Organization of the *E. coli* nucleoid.** The circular chromosome is shown at top left, with its macrodomains[19,21] colour-coded as follows: Ori (green), NS-Right (black), Right (orange), Ter (red), Left (blue) and NS-Left (yellow). The *oriC* locus is the origin of chromosome replication and *dif* is the site where the XerC/D site-specific recombinases resolve chromosome dimers[22]. The directions of DNA replication are shown by the black arrows and these constitute the Right (clockwise) and Left (anticlockwise) replichores. The positions of key genes that are named in the text are indicated. The chromosome is shown as a writhed structure (top right), reflecting imaging data that suggest that it adopts a conformation of this type[2,10,64], at least in rapidly growing bacteria[10]. Its thickness is indicative of the underlying layers of structure. Below this, a portion of the left replichore is illustrated alternatively as a solenoidal and as a plectoneme of periodicity 117 kb. The DNA is next compacted by introducing 10 to 12 microdomains into each of its 117-kb units. These 10-to-12-kb circles have a diameter of approximately 1.3 µm giving the nucleoid a cross section of about 2.6 µm. Supercoiling these small circles (bottom right) compacts them approximately 2-fold[12]. The H-NS nucleoid-associated protein[4,13,32] is shown as playing a core role within the two replichores where it acts to hold together the ends of the 10-to-12-kb microdomain loops. The DNA bridging and polymerization properties of H-NS are illustrated at bottom left.

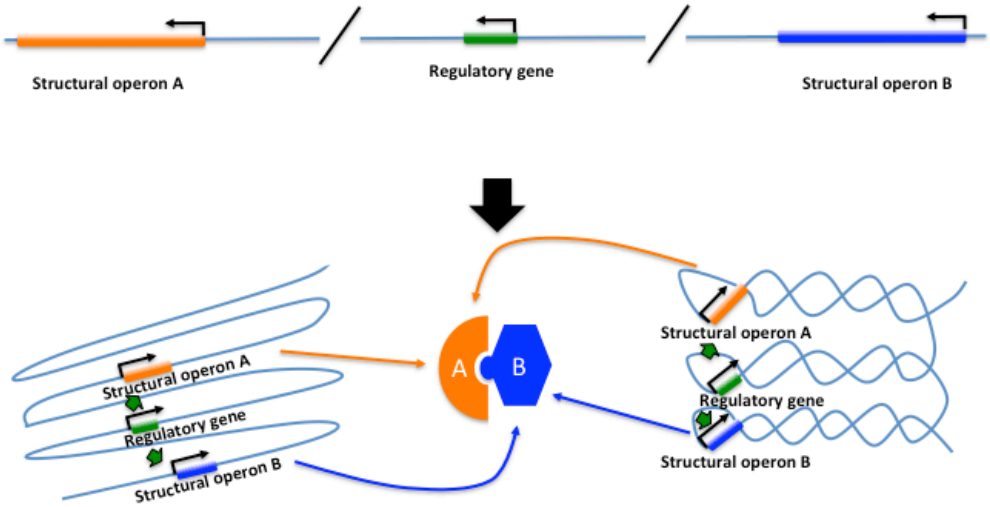

Figure 2| **Nucleoid folding and gene regulation.** A simple regulon consisting of a regulatory gene and two structural operons, A and B, is illustrated. When the chromosome is represented in a 1-dimensional form, the three genetic loci are separated by large distances in space. When the chromosome is reorganized as a solenoid (left) or as a plectoneme (right), its periodicity brings the three genes close together, facilitating communication between the regulatory gene and its two targets. Moreover, the products of the A and B operons are produced in close proximity, favouring their interaction.

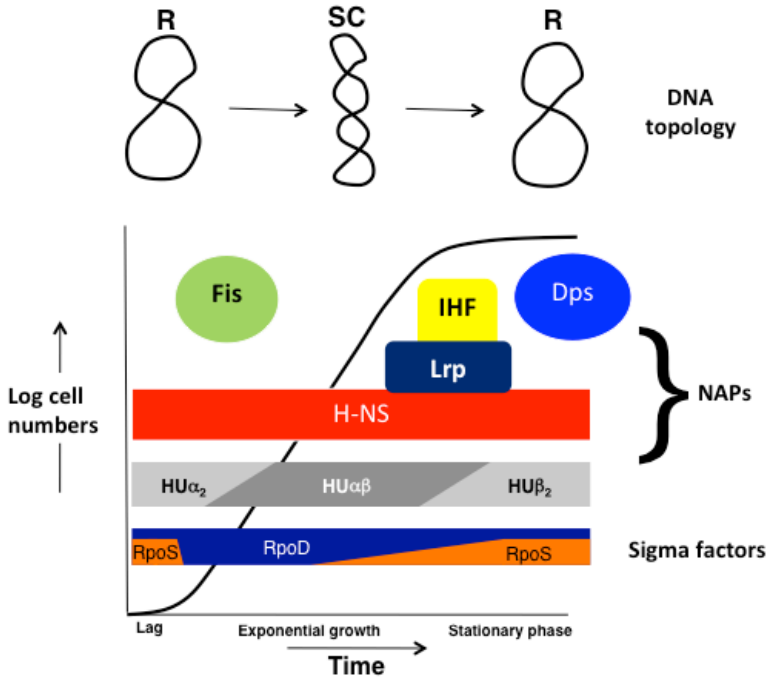

Figure 3| **Growth phase and nucleoid structural elements.** A typical growth curve is presented for *E. coli* growing in batch culture, with time and the three major phases of growth on the x-axis and population size on the y-axis. The approximate periods in growth at which significant changes in DNA topology are detectable, at which important NAPs named in the text appear and at which the balance between the RpoD and RpoS sigma factors of RNA polymerase occurs is indicated. The figure gives a purely qualitative impression of these events. R = relaxed, SC = (negatively) supercoiled.

Glossary

Chromatin immunoprecipitation on chip (ChIP-chip). This method allows the binding sites for a specific protein to be identified across a genome *in vivo*. The protein of interest is cross-linked to DNA with formaldehyde in living bacteria and the genomic DNA is extracted and then sonicated to achieve a desired average DNA fragment length. An antibody specific for the protein of interest (or for an epitope tag that has been attached to the protein by genetic engineering) is used to precipitate the protein-DNA complex. The cross links are then reversed, the released DNA is fluorescently tagged and its genomic location is then identified using a DNA microarray.

Chromosome conformational capture ('3C'). This technique[4] identifies physical interactions between parts of the genome that would not be predictable from a survey of its DNA sequence alone. Macromolecules are chemically cross-linked in living cells, the DNA is extracted, digested with a restriction enzyme and subjected to intramolecular ligation. The polymerase chain reaction is used to detect novel junctions in the ligated DNA that are predicted to arise from the close proximity of the now-joined sequences in the folded nucleoid. A chromatin immunoprecipitation step can be added to study specifically novel interactions that depend on a specific protein, such as a nucleoid-associated protein.

Chromosome conformational capture carbon copy ('5C'). A 3C library is first constructed as described above and then multiplex primers with universal primer extensions are annealed to the novel junctions in the library and then ligated together. Here the 3C junctions serve as templates to guide the perfect ligation of the primers. These can then be used in microarrays or subjected to high throughput sequencing to identify the DNA forming the junctions[2].

Dps. A NAP expressed in stationary phase cultures (or in cultures undergoing oxidative stress) that is thought to protect DNA from damage[67].

Fis. The Factor for Inversion Stimulation[13], a NAP expressed in early exponential phase cultures that organizes local DNA topology and modulates transcription.

H-NS. The histone-like, nucleoid structuring protein[13]. A NAP with a preference for binding to A+T-rich DNA, H-NS is expressed at all stages of growth, silences transcription of hundreds of genes and organizes nucleoid structure.

HU. A NAP with general DNA binding and compacting activity[13].

IHF. The Integration Host Factor[13], a paralogue of HU with site-specific DNA binding and bending activity.

Macrodomain. This is a genetically-defined large-scale chromosomal segment which is unlikely to undergo recombination with another macrodomains because the resulting rearrangement is detrimental to the survival of the bacterium[19]. The *E. coli* chromosome has four macrodomains (Ori, Ter, Left and Right) and two so-called non-structured regions (NS-Left and NS-Right) (Fig. 1).

Microdomain. This refers to the approximately 400 topologically independent loops of between 10 and 12 kb that co-exist within the macrodomain superstructure of the *E. coli* genome[34-38].

Nucleoid-associated proteins (NAPs). These are the low molecular mass, abundant DNA-binding proteins that are thought to act as architectural components within the nucleoid and to modulate gene expression. *E. coli* has at least twelve distinct NAPs[13].

Plectoneme. Interwound or braided DNA strands adopt a plectonemic structure. A common example is the writhing that is associated with negatively supercoiled DNA.

Replichore. The two arms of the circular chromosome along which DNA replication occurs are called replichores. There are two, the Right (or clockwise) and the Left (or anticlockwise) replichores, and these extend from the origin of chromosome replication, *oriC,* to the terminus of replication within the Ter macrodomain[19] (Fig. 1).

Topoisomerases. These are the enzymes that alter the linking number of DNA by cutting, strand passage and religation[41-45].