

# A model for *Escherichia coli* chromosome packaging supports transcription factor-induced DNA domain formation

Miriam Fritsche<sup>1,\*</sup>, Songling Li<sup>1</sup>, Dieter W. Heermann<sup>1,2</sup> and Paul A. Wiggins<sup>3</sup>

<sup>1</sup>Institute for Theoretical Physics, University of Heidelberg, Philosophenweg 19, D-69120 Heidelberg, Germany, <sup>2</sup>Institute for Molecular Biophysics, The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 and <sup>3</sup>Department of Physics, University of Washington, 3910 15th Avenue, North East, Seattle, WA 98195-1560, USA

Received June 23, 2011; Revised and Accepted September 5, 2011

## ABSTRACT

**What physical mechanism leads to organization of a highly condensed and confined circular chromosome? Computational modeling shows that confinement-induced organization is able to overcome the chromosome's propensity to mix by the formation of topological domains. The experimentally observed high precision of separate subcellular positioning of loci (located on different chromosomal domains) in *Escherichia coli* naturally emerges as a result of entropic demixing of such chromosomal loops. We propose one possible mechanism for organizing these domains: regulatory control defined by the underlying *E. coli* gene regulatory network requires the colocalization of transcription factor genes and target genes. Investigating this assumption, we find the DNA chain to self-organize into several topologically distinguishable domains where the interplay between the entropic repulsion of chromosomal loops and their compression due to the confining geometry induces an effective nucleoid filament-type of structure. Thus, we propose that the physical structure of the chromosome is a direct result of regulatory interactions. To reproduce the observed precise ordering of the chromosome, we estimate that the domain sizes are distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.**

## INTRODUCTION

Even the simplest organisms need to physically organize their chromosomes. Bacterial chromosomes form a compact DNA–protein complex called the nucleoid (1,2)

where the interplay between compaction of the genetic material and its accessibility ensure vital cell functions such as DNA replication, segregation, gene expression and repair. In this work, we investigate an intriguing biological and physical problem: what physical mechanisms lead to organization of highly condensed and confined circular chromosomes?

One of us previously measured and analyzed the position fluctuations of 15 single genetic loci (in G and early S phase), among which are the origin of replication *OriC* as well as two loci named 'lac' and 'C4' for convenience [for more details see (3)]. Loci in the body of the nucleoid were found to show a precision of positioning of better than 10% of the cell length and the precision of interlocus distance of genomically proximate loci was found to be better than 4%. Moreover, the linear relationship between positions of the genes on the chromosome and their spatial location within the cell was confirmed. While most studies of the localization of chromosomal loci in bacteria have focused on their position along the length of the cell, a recent study reports data concerning loci positioning across the width of the *Escherichia coli* cell finding *ter*-borne loci localized at the nucleoid periphery (4).

By physicochemical approaches, it has long been shown that the circular *E. coli* DNA molecule is organized into separate chromosomal loops or superhelical domains that are relaxed independently when DNA is cut (2,6,13–15). One purpose of these topological domains appears to be the prevention of chromosome unraveling as a result of DNA damage since the loss of chromosomal supercoiling leads to cell death (16). However, cross-links between random positions on the chromosome would also protect against unwinding (2). Instead of that genomic neighbors appear to be cross-linked suggesting that the so formed domains might serve structure–function relationships with respect to gene expression, too (7,17–21).

\*To whom correspondence should be addressed. Tel: +49 6221 54 9338; Fax: +49 6221 54 9331; Email: fritsche@tphys.uni-heidelberg.de

Experimental determinations of the average sizes of these domains differ between 10 kb (6) and 100 kb (23), report intermediate values (24), and even indicate an organization into a ring compacted by four macrodomains and two less structured regions (7,25). Thus, there already is evidence of a non-trivial nucleoid structure where smaller domains are organized within higher order 'super' structures.

Several drivers responsible for the observed reliable orientation and high level of organization of the *E. coli* chromosome have been suggested, including intranucleoid interactions such as (i) macromolecular crowding (5), (ii) DNA supercoiling (6) or (iii) protein–DNA interactions (7–9) as well as explicit mechanisms of external positioning such as (iv) cellular confinement (3,10) or (v) tethering of the chromosome (11,12).

Increasing lines of evidence link the three-dimensional (3D) packaging of genes to the proper coordination of gene expression (7,17–21). Transcription factors (TFs) are the key controlling elements for appropriate gene expression in bacteria, where a functional network of regulatory interactions between TFs and target genes, which can themselves be TFs, is formed (20,22). Assuming the spatial nucleoid structure to influence transcription and vice versa, specific correlations are expected to arise in gene expression patterns (2). In fact, expression patterns correlate at short (<16 kb), medium (~100 kb) and long (~600–700 kb) distances in *E. coli* (2,26,27). The short-range correlations might result from small elementary domains of the nucleoid, while the long-range correlations could result from higher scales of organization, i.e. from long-range interactions between regions of the chromosome that are genomically distant, but spatially proximate when the chromosome is packaged within the nucleoid (2).

Underpinning the evidence for 'regulatory domains', a recent experimental study on the spatial organization of mRNA in *E. coli* shows that mRNAs display limited dispersion from their site of transcription (19). The high localization of mRNA implies that chromosome architecture might act as a spatial organizer, which compartmentalizes the cell interior such that dedicated (regulatory) proteins are produced within those subcellular regions, where their regulatory intervention is needed (8,19,28,29).

The assumption of functional domains is supported by another work on the role of transcriptional regulation in shaping the organization of genes on a chromosome (20). It was demonstrated that the more target genes a TF regulates, the higher is its need to be expressed in higher concentrations to regulate targets located dispersedly on the chromosome. In contrast, local or dedicated TFs were found to be expressed in much lower concentrations explaining the reasons for their proximity on the chromosome to their target genes (20). This aspect justifies an *a posteriori* conformational organization of DNA to produce colocalization phenomena since it is a natural way to make 3D targeting and assembly of complexes more efficient and error free.

Additionally, there is experimental evidence from other bacteria suggesting that there is a link between the final expression products and chromosome organization (30).

Visualization of replicated DNA within living cells of *Bacillus subtilis* show that genes from distant chromosomal regions co-localize within a similar subcellular location for the purpose of coregulation (30).

How do bacterial cells operate such coordinated movement of specific sites within the compacted genome rapidly and faithfully? This is a place, in which physical modeling of polymers in confined space can help to relate experimental observations on *E. coli* nucleoid structure and to quantitatively test our models for chromosome organization.

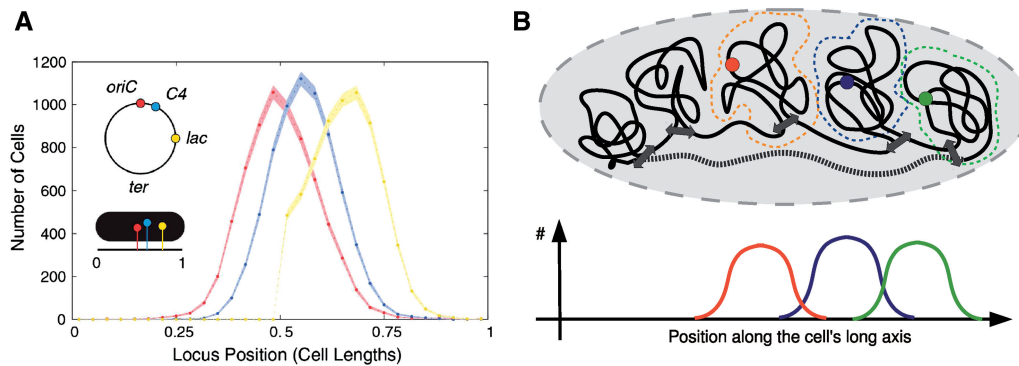
First, we demonstrate that confinement and condensation are not sufficient to spontaneously organize the chromosome. Without dividing the chromosome into topological domains, confinement-induced organization cannot overcome propensity of the chromosome to mix. The high precision of separate subcellular positioning of loci located on different chromosomal domains naturally emerges as a result of entropic demixing of these structural subunits. This concept might have important implications for chromosome segregation since the increasing topological complexity of the stacked sequence of chromosomal loops implies a stronger repulsion between replicated DNA chains, and consequently more precise organization and faithful segregation (31,32).

In a subsequent step, we propose and investigate one possible mechanism for organizing these domains: the gene regulatory network. It was demonstrated that in the gene regulatory network in *E. coli*, regulatory genes need to be expressed in different concentrations in dependence of the genomic distance from their target genes and of the number of target genes regulated (20). Additionally, it was shown that expressed mRNAs largely display limited dispersion from their sites of transcription, which suggests that translation is spatially organized by using the chromosome layout as a template (19). In light of these recent findings, we suggest that regulatory control requires the co-localization of TF genes and target genes. We do not propose a detailed mechanism generating these attractive interactions [which might result from protein–protein or from protein–RNA interactions (33,34)], but rather explore its consequences in shaping the physical structure of the *E. coli* chromosome.

In fact, we find that the DNA chain self-organizes into several topologically distinguishable domains where the interplay between the entropic repulsion of chromosomal loops and their compression due to the confining geometry induces a formation into a stacked sequence of interlinked domains. These domains are sufficient to generate the observed precision of *E. coli* chromosome structure and we estimate the domain sizes to be distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

## MATERIALS AND METHODS

Ring polymers or polymer loops do not intermingle, but entropically repel each other both in free space and even more strongly in confinement (31,35,36). This can be understood by noting that two ring polymers suffer a



**Figure 1.** (A) Histogram of long-axis locus position for IL01t C4 cells [adapted from (3)]. The genomic locus positions are shown schematically in the inset. Loci in the body of the nucleoid show a precision of positioning within the cell of better than 10% of the cell length. (B) A cartoon of the chromosomal domains being the ‘building blocks’ of the *E. coli* chromosome. The dark gray double arrows represent TF–gene interactions which ‘restrain’ the domains. The domain ‘walls’ indicated by the red, blue and green dashed lines are able to diffuse and thus subject to position fluctuations. The subcellular position distributions of the red, blue and green loci [fluorescently tagged in experiments (3,38,47,48)] exemplify the precise separate positioning of genetic sites on different domains due to their tendency of demixing. The gray dashed line represents the *ter*-proximate region acting as a linker that connects the two polymer arms. A detailed description of the polymer model can be found in section ‘Domain Topology and Confinement Shape *E. coli* Chromosome Packaging’ in Supplementary Data.

loss of conformational entropy when being brought together within a distance smaller than their gyration radius (31,32,36). This tendency to segregate, which holds to a lesser extent for linear domains, too, leads to compartmentalization. The repulsive effect of loop formation is highlighted in Supplementary Figure S1 confirming that the density clouds of adjacent polymeric loops are indeed well separated.

These results suggest that the experimentally observed separate subcellular localization of the three genetic loci in (3) and displayed in Figure 1 could emerge due to their positioning on different chromosomal domains/loops. In fact, it is the mutual entropic repulsion between chromosomal domains that constitutes an elegant, self-organized mechanism of ordering in confinement due to compartmentalization. However, without dividing the chromosome into topological domains, confinement-induced organization cannot overcome the propensity of the chromosome to mix. Additionally, the experimental observation of equal variance of the position distributions displayed in Figure 1 suggests the absence of tethering interactions.

In order to quantitatively test our assumptions, (i) we apply the Metropolis Monte Carlo method (37) to model the circular chromosome by a ring polymer being compacted due to fixed size loops within a rod-shaped geometry representing the nucleoid. The *ter*-region is represented by a stretched linker connecting the two polymer arms. This approach is based on a recent work finding the *E. coli* chromosome to be organized with a linker that connects the outer edges of the nucleoid (38). Figure 1B illustrates the employed polymer model and section ‘Domain Topology and Confinement Shape *E. coli* Chromosome Packaging’ in Supplementary Data explains further details on the polymer description. Comparing this approach to a ‘null model’, which consists of a simple ring polymer confined to the same geometry, allows for the investigation of the impact of

domain formation on the packaging of highly confined chromosomes.

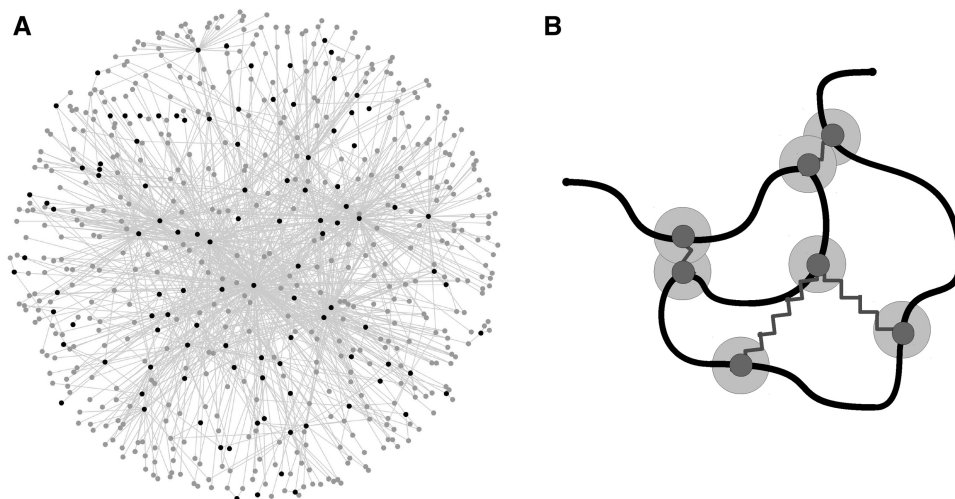
In a subsection step (ii), we propose that the *E. coli* gene regulatory network permits the genetic loci to identify and form domains with genetic neighbors. We investigate the consequences of this assumption by applying numerical simulations, where the chromosome is modeled as a self-avoiding polymer in confinement. TFs along the chromosome are associated with the respective sites along the polymer chain. These sites interact with their target sites according to an effective attractive potential, mimicking a regulatory interplay and driving TFs and their target genes to colocalize in space. Figure 2A illustrates the transcriptional regulatory network as used in this work. Additionally, Figure 2 shows a 2D cartoon of the 3D self-avoiding DNA chain, which mimics the regulatory control between TF genes and target genes by assuming a harmonic interaction between these sites. Further details on the polymer description as well as on the *E. coli* gene regulatory network are shown in section ‘Gene Expression and Colocalization’ in Supplementary Data.

## RESULTS

### Division of the chromosome into domains is key to confinement-induced organization

The position distribution of four sites along the polymer backbone with respect to the long axis of the confining cavity both for the *E. coli* polymer model as well as for the ‘null model’ are displayed in Figure 3. In case of the bare ring polymer, the absence of distinguished domains leads to a loss of spatial ordering that is reflected in the polymer’s mobility throughout the confining geometry. In contrast, the specific topology of the *E. coli* polymer model leads to an interplay between the domain’s tendency of de-mixing and the pressure exerted by confinement. As a result, the stacking of polymer loops inside the rod-shaped geometry induces a high level of spatial





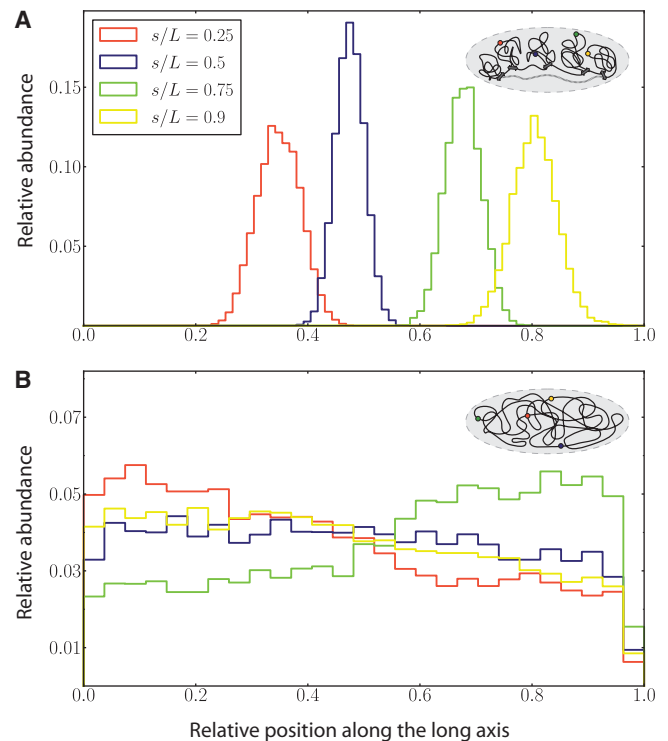
**Figure 2.** (A) Graphical illustration of the transcriptional regulatory network describing the regulatory interplay between the TFs and their target genes as applied in our polymer model. Black nodes represent TF genes, dark gray nodes represent the target genes and links represent regulatory interactions between them. A detailed description of the network can be found in section ‘Gene Expression and Co-localization in Supplementary Data’. Global regulators are those TF genes that regulate lots of target genes, while other regulatory proteins are local, dedicated regulators. (B) 2D cartoon of the 3D self-avoiding DNA chain [adapted from (39)] mimicking the regulatory control between TF genes and target genes by assuming a harmonic interaction between these sites. Sites that can interact with other sites are represented by small blue filled circles connected by blue springs. The outer blue circles define the strength of the harmonic interaction potential. See section ‘Gene Expression and Co-localization’ in Supplementary Data for further details on the polymer description and the TF-gene network.

positioning illustrated by the well-defined and separated position distributions in Figure 3.

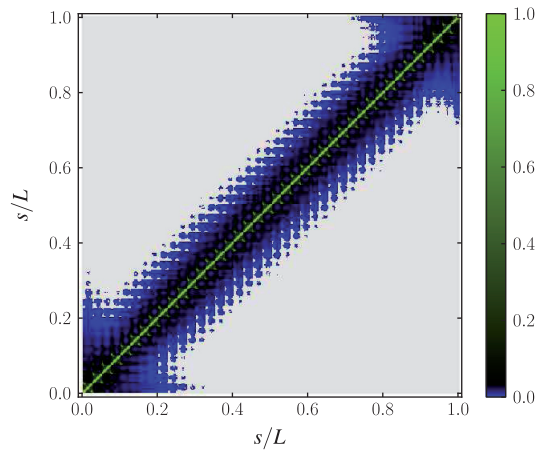
The specific spatial organization is able to explain the general linear correlation between positioning of genes on the chromosome and their location in the cellular volume: adjacent loops are displaced along the long axis of the rod-shaped confining volume in such a way to form a sequence of stacked ‘neighbor’ loops or an ‘effective nucleoid filament’ (3). The linear relationship between genomic and subcellular position is apparent in Figure 3, too.

An analysis of how the SMC-like protein MukBEF condenses DNA revealed its likely involvement in organizing the chromosome in a series of loops orthogonal to the cell axis, which was argued to account for the orderly arrangement of the chromosome (40). Here, we show that entropy might provide a purely physical driving force to support the dedicated action of specific proteins such as chromosome condensation proteins in order to create the right physical conditions for chromosome packaging. Notably, this result is confirmed by the experimental observation that cells with the MukBEF deletion are capable of almost wild-type chromosome structure.

An established measure of compartmentalization are the contact probabilities between pairs of sites along the polymer as shown in Figure 4. The sites are numbered consecutively and the contact map displays their probabilities of establishing a contact, i.e. whenever they are spatially closer than a threshold distance  $d_{\text{threshold}}$ . In fact, the very low contact probability of non-diagonal contacts in Figure 4 shows that domain formation induces contacts among sites within the same and between sites of genomically neighboring domains. Thus, the experimentally observed high precision of subcellular locus



**Figure 3.** The position distribution of four sites located at the relative positions  $s/L = 0.25, 0.5, 0.75$  and  $0.9$  along the polymer backbone  $L$  are shown with respect to the long axis of the confining cavity for (A) the domain model and (B) the ‘null model’. In case of the bare ring polymer, the absence of distinguished domains leads to a loss of spatial ordering which is reflected in the polymer’s mobility throughout the confining geometry. In contrast, the specific topology of the *E. coli* polymer model induces a high level of spatial positioning. The interplay between the domain’s tendency of de-mixing and the pressure exerted by confinement leads to a stacking of polymer loops inside the rod-shaped geometry.

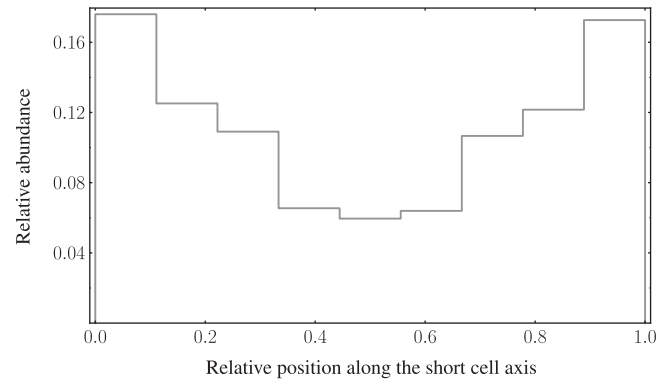


**Figure 4.** The contact probability map shows the probability for pairs of sites along the polymer backbone  $s/L$  to form a contact, i.e. whenever both sites are spatially closer than a threshold distance  $d_{\text{threshold}}$ . Domain formation induces contacts among sites within the same domain and between sites of neighboring domains. The experimentally observed high precision of subcellular locus positioning can be explained by noting that the mobility of sites belonging to different chromosomal domains is restricted to the radius of the so-defined structural subunit.

positioning of better than 10% of the cell length (3) for genetic sites lying on different chromosomal domains has to be interpreted within the concept of structural units as the building blocks of the chromosome consisting of (supercoiled) DNA loops stabilized by DNA-binding proteins (41). The size of the structural units simultaneously influences the organization of the chromosome (segregation of chromosomal domains versus mixing) and its conformation (ordered versus random) in confined space (32). Consequently, the size of the structural unit is reflected in the decay width of the contact probability backbone shown in Figure 4. While Valens *et al.* (25) provide one measure of contact probabilities, Hi-C would provide another complementary approach with the opportunity to observe the interactions in a different biochemical context (cross-linking) and with a higher resolution in order to test our predictions (42).

Moreover, while other domain organization models conclude that domain barriers are not placed stably at fixed sites on the chromosome, but are effectively randomly distributed hard walls (6), we have shown that their position need not be stochastic, since the domain walls are able to diffuse and thus subject to position fluctuations.

A recent study has evaluated the position of *E. coli* (43) chromosomal loci across the width of the cell by tagging loci with fluorescent proteins and comparing the measured distributions with simulated ones from different cell width models (4). The terminus region of the chromosome is found to be excluded from the body of the nucleoid and preferentially located at its periphery (4,44). Our model of the *E. coli* chromosome topology displays the same characteristic feature. The polymer chain region, which represents the *ter*-borne chromosome part, is preferentially located at the periphery of the confining cavity as is



**Figure 5.** The relative abundance of *ter*-proximate sites as a function of their positioning with respect to the short axis of the confining geometry (cell's small axis). Meile and coworkers (4) have determined the position distributions of loci in the terminus region with respect to the short axis of the cell. They find these *ter*-borne loci localized at the nucleoid periphery. In our model, the polymer chain region representing the stretched *ter*-proximate region connecting the two nucleoid edges is preferentially located near the confining envelope, too, thus confirming the experimental observation. Within our modeling approach, this can be understood by noting that the (mostly linearly stretched) *ter*-region has less topological complexity compared with the looped 'filament'-like structure of the remaining nucleoid. The topological complexity of the domains not only leads to an entropic repulsion between adjacent chromosomal domains, but also pushes the *ter*-region towards the envelope of the confining cavity in agreement with the experimental findings (4).

shown in Figure 5. Within our modeling approach, this result can be understood by noting that the (mostly linearly stretched) *ter*-region has less topological complexity compared with the looped 'filament'-like structure of the remaining nucleoid. The topological complexity of the domains not only leads to an entropic repulsion between adjacent chromosomal domains, but also pushes the *ter*-region towards the envelope of the confining cavity. A donut like topology is observed if the topological complexity of chromosome is uniform (45).

### The gene regulatory network as a mechanism for domain formation

We have shown that the interplay between entropic repulsion of chromosomal domains and pressure exerted by the envelope of the confining cavity can be one driving force for nucleoid organization (and segregation) supporting the action of dedicated cellular machinery (46,47) or internal pushing forces (31,32,48). However, while the existence of chromosomal domains is widely accepted, their size and the mechanism that gives rise to them is still under debate. In this work, we investigate one possible mechanism for organizing chromosomal domains: the gene regulatory network.

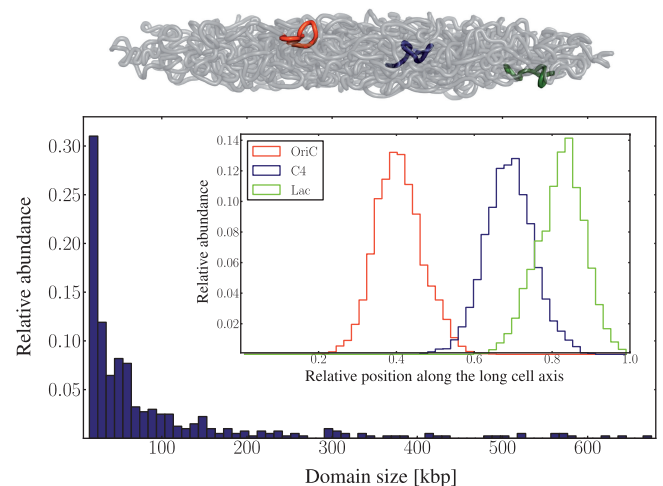
The decision about gene expression or repression is controlled by TFs, which use metabolic or environmental signals to trigger a transcriptional response (22) within a functional network of regulatory interactions between TFs and target genes (21,49–51). Additionally, the proper genome-wide coordination of gene expression has been shown to be linked to the spatial organization of the chromosome within the nucleoid (8,22,52,53). In this

respect, one can distinguish between ‘analog’ control, i.e. regulatory action by chromosome topology, and ‘digital’ control, i.e. regulation mediated by transcription factors (21,50,51,54).

Regarding the observed existence of chromosome domains in *E. coli*, we propose that one-dimensionally distant target genes, i.e. genes that are genomically far away from their regulative TF, colocalize with it in order to facilitate transcription. This assumption is supported by a recent work on the role of transcriptional regulation in shaping the organization of genes on a chromosome (20). It was demonstrated that the higher a TF is in the transcriptional hierarchy, the higher is its need to be expressed in higher concentrations to regulate target genes located dispersedly on the chromosome. In contrast, local or dedicated TFs, that are lower in the network hierarchy, were found to be expressed in much lower concentrations explaining the reasons for their proximity on the chromosome to their target genes. This aspect justifies an *a posteriori* conformational organization of DNA to produce colocalization phenomena since it is a natural way to make 3D targeting and assembly of complexes more efficient and error free. Thus, the formation of a specific chromosome topology (chromosomal domains) stabilized by nucleoid-associated proteins (NAPs) who are in charge of most chromosomal remodeling tasks could be seen as a feature of analog control.

Since a model for *E. coli* domain organization involves the recognition of a domain-specific pattern by a protein which would isolate it from other domains, various experimental groups have been looking for dedicated proteins that bind specifically to a single domain to organize the chromosome. A few examples of domain forming proteins have been identified. In *B. subtilis*, the DNA-binding ParB-like protein Spo0J appears to ensure proper arrangement and partitioning of chromosomal DNA by recruiting the condensin structural maintenance of chromosomes (SMCs) complex to the replication origin region (55,56). In *E. coli*, SMC-like complex MukBEF appears to colocalize with the origin of replication, but the mechanism is unknown. A recent study has identified a protein MatP that structures the terminus macrodomain. ChIP experiments have revealed that proteins SeqA and SlmA both appear to be excluded from the terminus region (7,57,58). However, the existence of these known examples is insufficient to explain the precision of structure observed on the chromosome (7). Here, we propose that structure proteins have been hiding in plain sight. They are the TFs that are already known to target small regions of the chromosome specifically.

We mimic the regulatory control between TF genes and target genes in the transcriptional regulatory network by assuming a harmonic interaction between these sites as illustrated in Figure 2. Within this rather general framework, we find that the DNA chain indeed self-organizes into several topologically distinguishable domains. Figure 6 displays a snapshot of the genome organization as obtained after the equilibration of the self-interacting DNA chain, i.e. when the system fluctuates around its global energy minimum.



**Figure 6.** The 3D chromosome organization is obtained after the equilibration of the interacting self-avoiding polymer chain in a rod-shaped geometry. Modeling the regulatory interplay between TFs and their targets based on the *E. coli* transcriptional regulatory network, we find that genes on a DNA chain self-organize into several topologically distinguishable domains of different sizes. The chromosomal loop to which *OriC* is associated with is displayed in red, while the blue and green marked chromosomal regions refer to the chromosome domains that contain the genetic loci *C4* and *lac*, respectively. It becomes clear that the three genetic loci are well separated with respect to projections on the long axis of the confining envelope. The formed domains are sufficient to generate the observed precision of *E. coli* chromosome structure and we estimate the domain sizes to be distributed between 10 and 700 kb.

Local TFs regulating a small number of target genes tend to localize peripherally, while global TFs which regulate a large number of genes assume more central localizations within the confining geometry. Several models have been proposed that lead to such a macro-arrangement of the nucleoid (59,60). In particular, the microarray experiments by Jeong *et al.* (27) show a high degree of correlation between the transcriptional signal of genes close together on the chromosome, where the observed stability and range of correlations extend far beyond the expected size of the average operon. Such dependence offers an intriguing hypothesis about the physical basis of the short-range transcriptional correlations: the transcription of the genes within a chromosomal domain is more similar to each other than to genes in other domains (27).

The folding of the nucleoid in domains due to gene colocalization puts into spatial contact distant chromosomal regions. The inset of Figure 6 shows the position distributions for the three genetic loci *oriC*, *C4* and *lac* concurrently visualized in (3) as obtained by our model of the *E. coli* nucleoid. We find the three genetic loci to be located on different domains emerging due to TF–gene colocalization. In agreement with this finding, the genomic distances between the loci *oriC* and *C4* as well as between the loci *C4* and *lac* being ~300 and 600 kb, respectively, formally confirm the sites’ positioning on different domains even when an upper loop size limit of 700 kb (emerging from our numerical calculations) is assumed. Additionally, we find the terminus region to



self-organize at the mid-cell position connecting the two chromosome arms in agreement with Ref. (3)

Moreover, the inset of Figure 6 also shows that the assumption of TF–gene interactions as drivers of domain formation (among other possible mechanisms) are able to reproduce the linear correlation between the position of a gene on the chromosome (i.e. site along the polymer) and its subcellular position inside the nucleoid (i.e. the long axis of the confining cavity). In fact, the strong linear correlation is a direct consequence of the fact that DNA is compacted and confined. Notably, there is only a linear correlation for sufficiently large domain sizes since the absence of domains leads to a loss of precision of positioning and consequently linear ordering (61).

Thus, we study the distribution of loop sizes in Figure 6 finding them to range between 10 and 700 kb in agreement with the size of topological domains identified in the context of DNA supercoiling (27,62). We find a mean domain size of  $\langle l_{\text{domain}} \rangle = 86$  kb. The long tail of large loop sizes (>150 kb) can be explained by noting that the larger domains are themselves built up by smaller subdomains in agreement with the observed long-range correlations in gene expression patterns (2,26,27).

## DISCUSSION

In this work, we have (i) proposed a mechanism by which chromosomal domains are formed and (ii) quantitatively investigated this model applying numerical simulations. We assume the *E. coli* transcriptional regulatory network to give rise to colocalization of TFs and their target sites due to attractive interactions between genetic loci. Under these conditions, we find that the DNA chain self-organizes into several topologically distinguishable domains where the interplay between the entropic repulsion of chromosomal loops and their compression due to the confining geometry induces a formation into a stacked sequence of interlinked domains or ‘rosettes’ (63). Thus, the experimentally observed high precision of separate subcellular positioning of genetic loci located on different chromosomal domains naturally emerges as a result of entropic de-mixing of loops where the precision of localization is related to the position fluctuations of each chromosomal domain. To recover the precision of organization observed in *E. coli*, we estimate the domain sizes to be distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

However, the question has to be raised whether the circular chromosome behaves as an equilibrated polymer (64). Calculation of the rate of uncoiling of the DNA molecule can be computed by taking into account the increase of entropy on unwinding as well as the viscous resistance of the surrounding medium (65). Thus, at ~300 K the bidirectional replication of a 4-Mb chromosome would require ~20 min (65), which might leave a substantial amount of time for the non-replicating phase where chromosome domains could be rearranged. Notably, the study of *E. coli* chromosome organization is influenced by the experimental conditions, which

might not be judicious due to different genetic background and growth conditions. Due to the high packing density of the nucleoid, the bacterial chromosome might not be able to explore the whole configuration space of possible conformations but rather a restricted subspace starting from similar initial configurations emerging due to the progressive segregation and DNA compaction after each replication cycle. In fact, numerical calculations on time scales that are small compared with the relaxation time of our polymer system indicate even more precise locus positioning due to ‘frozen-in’ configurations.

Our concept could have important implications for chromosome segregation, too. It was previously shown that compaction of the bacterial chromosome and conformational entropy alone could direct and facilitate the segregation of newly replicated daughter strands of DNA (31,32). In this work, the sequence of stacked domains has higher internal topological complexity compared with linear or circular chains leading to even stronger repulsive interactions. Thus, segregation by entropic forces induced by chain topology in strong confinement might constitute a reliable mechanism, where no additional drivers such as a mitotic spindle-like machinery or dedicated proteins may be needed (32).

We have not taken into account the effect of the chromosomal domains being negatively supercoiled by either plectonemic or toroidal supercoils (62) since we have focused on the global physical properties of the chromosome on a more coarse-grained level. However, DNA gyrases, which cause branched supercoils and thus increase domains topological complexity, further promote entropic repulsion between the supercoiled chromosomal domains as pointed out above.

Summarizing, our approach offers a robust framework for understanding the basic physical principles underlying *E. coli* chromosome organization. Its advantage is that it does not depend on the microscopic details of the DNA chain or on specific DNA–protein interactions. In light of the difference in length scales between proteins and chromosomes, the question has to be raised whether local actions of specific proteins alone are able to globally shape chromosome organization (7,32). Thus, our model is based on the idea that nature exploits entropy, excluded volume, specific chromosome topologies and confinement as a driving force to create the right physical conditions for chromosome packaging eventually fine-tuned by the dedicated action of proteins such as NAPs that bridge and bend DNA (9) or chromosome condensation proteins [SMC (51) and MukBEF (66)]. Future experiments investigating the spatial distribution of TF genes and target genes would be insightful to test our predictions.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figure 1 and Supplementary References [1–6].

## ACKNOWLEDGEMENTS

The authors would like to thank Jané Kondev, Suckjoon Jun, Jonathan Machta, Lindsay Shopland and Benót Knecht for their comments and suggestions. For this work, the bwGRiD parallel computing facilities (<http://www.bw-grid.de>), member of the German D-Grid initiative, funded by the Ministry for Education and Research and the Ministry for Science, Research and Arts Baden-Wuerttemberg, as well as the high performance PC cluster HELICS II at the Interdisciplinary Center for Scientific Computing (IWR) were used.

## FUNDING

Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences (HGS MathComp to M.F.); Heinz-Goetze-Foundation (to S.L.); National Science Foundation (Grant PHY-084845 to P.A.W.). Funding for open access charge: National Science Foundation (Grant PHY-084845 to P.A.W.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Toro, E. and Shapiro, L. (2010) Bacterial chromosome organization and segregation. *Cold Spring Harb. Perspect. Biol.*, **2**, a000349.
- Rocha, E.P.C. (2008) The organization of the bacterial genome. *Annu. Rev. Genet.*, **42**, 211–233.
- Wiggins, P.A., Cheveralls, K.C., Martin, J.S., Lintner, R. and Kondev, J. (2010) Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament. *Proc. Natl Acad. Sci. USA*, **107**, 4991–4995.
- Meile, J.-C., Mercier, R., Stouf, M., Pages, C., Bouet, J.-Y. and Cornet, F. (2011) The terminal region of the *E. coli* chromosome localises at the periphery of the nucleoid. *BMC Microbiol.*, **11**, 1–10.
- Marenduzzo, D., Micheletti, C. and Cook, P.R. (2008) Entropy-driven genome organization. *Biophys. J.*, **90**, 3712–3721.
- Postow, L., Hardy, C.D., Arsuaga, J. and Cozzarelli, N.R. (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes. Dev.*, **18**, 1766–1779.
- Dame, R.T., Kalmykova, O.J. and Grainger, D.C. (2011) Chromosomal macrodomains and associated proteins: implications for DNA organization and replication in Gram negative Bacteria. *PLoS Genet.*, **7**, e1002123.
- Vora, T., Hottes, A.K. and Tavazoie, S. (2009) Protein Occupancy Landscape of a Bacterial Genome. *Mol. Cell*, **35**, 247–253.
- Luijsterburg, M.S., Noom, M.E., Wuite, G.J.L. and Dame, R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.*, **156**, 262–272.
- Fritsche, M. and Heermann, D.W. (2011) Confinement driven spatial organization of semiflexible ring polymers: implications for biopolymer packaging. *Soft Matter*, **7**, 6906–6913.
- Bowman, G.R., Comolli, L.R., Zhu, J., Eckart, M., Koenig, M., Downing, K.H., Moerner, W.E., Earnest, T. and Shapiro, L. (2008) A self-associating protein critical for chromosome attachment, division, and polar organization in *caulobacter*. *Cell*, **134**, 945–955.
- Ebersbach, G., Briegel, A., Jensen, G.J. and Jacobs-Wagner, C. (2008) A self-associating protein critical for chromosome attachment, division, and polar organization in *caulobacter*. *Cell*, **134**, 956–968.
- Griffith, J.D. (1976) Visualization of prokaryotic DNA in a regularly condensed chromatin-like fiber. *Proc. Natl. Acad. Sci. USA*, **73**, 563–7.
- Meyer, M., De Jong, M.A., Woldringh, C.L. and Nanninga, N. (1976) Factors affecting the release of folded chromosomes from *Escherichia coli*. *Eur. J. Biochem.*, **63**, 469–75.
- Kavenoff, R. and Ryder, O.A. (1976) Electron microscopy of membrane-associated folded chromosomes of *Escherichia coli*. *Chromosoma*, **55**, 13–25.
- Deng, S., Stein, R.A. and Higgins, N.P. (2005) Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.*, **57**, 1511–1521.
- Frenster, J.H. (1965) Correlation of the binding to DNA loops or to DNA helices with the effect on RNA synthesis. *Nature*, **208**, 1093.
- Danchin, A., Guerdoux-Jamet, P., Moszer, I. and Nitschké, P. (2000) Mapping the bacterial cell architecture into the chromosome. *Philos. Trans. Roy. Soc. Lond. B Biol. Sci.*, **355**, 179–190.
- Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T. and Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature*, **466**, 77–82.
- Janga, S.C., Salgado, H. and Martínez-Antonio, A. (2009) Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res.*, **37**, 3680–3688.
- Marr, C., Geertz, M., Htt, M.-T. and Muskhelishvili, G. (2008) Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.*, **2**, 18.
- Martinez-Antonio, A., Medina-Rivera, A. and Collado-Vides, J. (2009) Structural and functional map of a bacterial nucleoid. *Genome Biol.*, **10**, 247–250.
- Worcel, A. and Burgi, E. (1972) On the structure of the folded chromosome of *Escherichia coli*. *J. Mol. Biol.*, **71**, 127–147.
- Romantsov, T., Fishov, I. and Krichevsky, O. (2007) Internal structure and dynamics of isolated *Escherichia coli* nucleoids assessed by fluorescence correlation spectroscopy. *Biophys. J.*, **92**, 2875–2884.
- Valens, M., Penuad, S., Rossignol, M., Cornet, F. and Boccard, F. (2004) Macrodome organization of the *Escherichia coli* chromosome. *EMBO J.*, **23**, 4330–4341.
- Carpentier, A.S., Torresani, B., Grossmann, A. and Henaut, A. (2005) Decoding the nucleoid organization of *Bacillus subtilis* and *Escherichia coli* through gene expression data. *BMC Genomics*, **6**, 84.
- Jeong, K.S., Ahn, J. and Khodursky, A.B. (2004) Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol.*, **5**, R86.
- Bailly-Bechet, M., Danchin, A., Iqbal, M., Marsili, M. and Vergassola, M. (2006) Codon usage domains over bacterial chromosomes. *PLoS Comput. Biol.*, **2**, e37.
- Cho, B.-K. and Palsson, B.Ø. (2009) Can the protein occupancy landscape show the topologically isolated chromosomal domains in the *E. coli* genome?: an exciting prospect. *Mol. Cell*, **35**, 255–256.
- Berlatzky, I.A., Rouvinski, A. and Ben-Yehuda, S. (2008) Spatial organization of a replicating bacterial chromosome. *Proc. Natl Acad. Sci. USA*, **105**, 14136–14140.
- Jun, S. and Mulder, B. (2006) Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proc. Natl Acad. Sci. USA*, **103**, 12388.
- Jun, S. and Wright, A. (2010) Entropy-driven chromosome segregation. *Nat. Rev. Microbiol.*, **8**, 600–607.
- Ohniwa, R.L., Morikawa, K., Takeshita, S.L., Kim, J., Ohta, T., Wada, C. and Takeyasu, K. (2007) Transcription-coupled nucleoid architecture in bacteria. *Genes Cells*, **12**, 1141–1152.
- Foley, P.L., Wilson, D.B. and Shuler, M.L. (2010) Macromolecular crowding can account for RNase-sensitive constraint of bacterial nucleoid structure. *Biochem. Biophys. Res. Commun.*, **395**, 42–47.
- Bohn, M., Heermann, D.W., Lourenço, O. and Cordeiro, C. (2010) On the influence of topological catenation and bonding constraints on ring polymers. *Macromolecules*, **43**, 2564–2573.
- Bohn, M. and Heermann, D.W. (2010) Topological interactions between ring polymers: implications for chromatin loops. *J. Chem. Phys.*, **132**, 044904–044914.
- Binder, K. and Heermann, D.W. (2010) *Monte Carlo Simulation in Statistical Physics: An Introduction*. Springer, Berlin.



38. Liu, X., Wang, X., Reyes-Lamothe, R. and Sherratt, D. (2010) Replication-directed sister chromosome alignment in *Escherichia coli*. *Mol. Microbiol.*, **75**, 1090–1097.
39. Junier, I., Martin, O. and Képès, F. (2010) Spatial and topological organization of DNA chains induced by gene co-localization. *PLoS Comput. Biol.*, **6**, e1000678.
40. Breier, A.M. and Cozzarelli, N.R. (2004) Linear ordering and dynamic segregation of the bacterial chromosome. *Proc. Natl Acad. Sci. USA*, **101**, 9175–9176.
41. Stavans, J. and Oppenheim, A. (2006) DNA-protein interactions and bacterial chromosome architecture. *Phys. Biol.*, **3**, R1.
42. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, J., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
43. Espéli, O., Mercier, R. and Boccard, F. (2008) DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol. Microbiol.*, **68**, 1418–1427.
44. Mercier, R., Petit, M., Schbath, S., Robin, S., El Karoui, M., Boccard, F. and Espéli, O. (2008) The MatP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, **135**, 475–485.
45. Niki, H., Yamaichi, Y. and Hiraga, S. (2000) Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Dev.*, **14**, 212–223.
46. Reyes-Lamothe, R., Wang, X. and Sherratt, D. (2008) *Escherichia coli* and its chromosome. *Trends Microbiol.*, **16**, 238–245.
47. Bates, D. and Kleckner, N. (2005) Chromosome and replisome dynamics in *E. coli*: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, **121**, 899–911.
48. Joshi, M.C., Bourniquel, A., Fisher, J., Brian, T.H., Kleckner, N. and Bates, D. (2011) *Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc. Natl Acad. Sci. USA*, **108**, 2765–2770.
49. Hershberg, R., Yeger-Lotem, E. and Margalit, H. (2005) Chromosomal organization is shaped by the transcription regulatory network. *Trends Genetic*, **21**, 138–142.
50. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, **14**, 283–291.
51. Janga, S.C. and Collado-Vides, J. (2007) Structure and evolution of gene regulatory networks in microbial genomes. *Res. Microbiol.*, **158**, 787–794.
52. Sonnenschein, N., Geertz, M. and Muskhelishvili, G. (2011) Analog regulation of metabolic demand. *BMC Microbiol.*, **5**, 1–13.
53. Sexton, T., Schober, H., Fraser, P. and Gasser, S. (2007) Gene regulation through nuclear organization. *Natl. Struct. Mol. Biol.*, **14**, 1049–1055.
54. Marr, C., Geertz, M., Hütt, M.T. and Muskhelishvili, G. (2008) Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.*, **2**, 18.
55. Sullivan, N.L., Marquis, K.A. and Rudner, D.Z. (2009) Recruitment of condensin to replication origin regions by ParB-parS organizes the origin and promotes efficient chromosome segregation. *Cell*, **137**, 697–707.
56. Gruber, S. and Errington, J. (2009) Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*. *Cell*, **137**, 685–696.
57. Sánchez-Romero, M.A., Busby, S.J., Dyer, N.P., Ott, S., Millard, A.D. and Grainger, D.C. (2010) Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12. *mBio.*, **1**, e00012–e00010.
58. Tonthat, N.K., Arold, S.T., Pickering, B.F., Van Dyke, M.W., Liang, S., Lu, Y., Beuria, T.K., Margolin, W. and Schumacher, M.A. (2011) Molecular mechanism by which the nucleoid occlusion factor, SlnA, keeps cytokinesis in check. *EMBO J.*, **30**, 154–164.
59. Wright, M.A., Kharchenko, P., Chrch, G.M. and Segre, D. (2007) Chromosomal periodicity of evolutionary conserved gene pairs. *Proc. Natl Acad. Sci. USA*, **104**, 10559–10564.
60. Kepes, F. (2004) Periodic transcriptional organization of the *E. coli* genome. *J. Mol. Biol.*, **340**, 957–964.
61. Buenemann, M. and Lenz, P. (2010) A geometrical model for DNA organization in bacteria. *PLoS ONE*, **5**, e13806.
62. Willenbrock, H. and Ussery, D.W. (2004) Chromatin architecture and gene expression in *Escherichia coli*. *Genome Biol.*, **5**, 252.
63. Bates, D. (2008) The bacterial replisome: back on track? *Mol. Microbiol.*, **69**, 1341–1348.
64. Rosa, A. and Everaers, R. (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.*, **4**, e1000153.
65. Longuet-Higgins, H.C. and Zimm, B.H. (1960) Calculation of the Rate of Uncoiling of the DNA Molecule. *J. Mol. Biol.*, **2**, 1–4.
66. Danilova, O., Reyes-Lamothe, R., Pinskaya, M., Sherratt, D. and Possoz, C. (2007) MukB colocalizes with the oriC region and is required for organization of the two *Escherichia coli* chromosome arms into separate cell halves. *Mol. Microbiol.*, **65**, 1485–1492.