Soft Matter

Chromosome-like organization of an asymmetrical ring polymer confined in a cylindrical space

Journal:	*Soft Matter*
Manuscript ID:	SM-ART-05-2015-001286.R1
Article Type:	Paper
Date Submitted by the Author:	13-Aug-2015
Complete List of Authors:	Ha, Bae-Yeun; University of Waterloo, Jeon, Chanil; University of Waterloo, Kim, Juin; KAIST, Jeong, Hawoong; KAIST, Jung, Youngkyun; KISTI,

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Chromosome-like organization of an asymmetrical ring polymer confined in a cylindrical space†

Chanil Jeon,[a] Juin Kim,[ab‡] Hawoong Jeong,[*bc] Youngkyun Jung,[*d] and Bae-Yeun Ha[*a]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

To what extent does a confined polymer show chromosome-like organization? Using molecular dynamics simulations, we study a model *Escherichia coli* (*E. coli*) chromosome: an asymmetrical ring polymer, formed by small monomers on one side and big monomers on the other confined in a concentric-shell or simple closed cylinder. The ring polymer is organized in the way observed for the *E. coli* chromosome: if the big monomers are assumed to be localized in the inner cylinder, the two "subchains" forming the ring are spontaneously partitioned in a parallel orientation with the "body" (big-monomer) chain linearly organized with a desired precision and the crossing (small-monomer) chain residing preferentially in the peripheral region. Furthermore, we show that the introduction of a "fluctuating boundary" between the two subchains leads to a double-peak distribution of *ter*-proximate loci, as seen in experiments, which would otherwise remain single-peaked. In a simple cylinder, however, a chromosome-like organization of the ring polymer typically requires an external mechanism such as cell-wall attachment. Finally, our results clarify to what degree the spatial organization of the chromosomes can be accomplished solely by ring asymmetry and anisotropic confinement.

1 Introduction

How chromosomes are spatially organized in cells is not only important on its own merits but also benefits other biological processes (e.g., accessibility of genes to proteins as for transcription) [1–14]. Accordingly, significant effort has been made to understand the mechanism of chromosome organization in cells [1–14]. The bacterial chromosome, occupying a sub-cellular space known as the nucleoid, is of particular interest, because of the intriguing roles of cylindrical confinement in shaping chain molecules, i.e., both single chains and the way they interact [15–18]. For instance, the *Escherichia coli* (*E. coli*) chromosome is often cartooned as a 'sausage' (with a stretch connecting its ends, as shown in Fig. 1) or a 'donut', under slow-growth or fast-growth conditions, respectively [5–7]. A recent experimental study has shown that the nucleoid body (the sausage part) is overall linearly organized with loci precisely positioned in the cell [8]. In particular, the width of both single-locus and interlocus-locus distributions is estimated to be within a few or 10% of the cell length [8]. On the other hand, the spatial organization of the donut-shape chromosome is shown to be consistent with what we would expect from a symmetric ring polymer under cylindrical confinement [14]. What physical effects are responsible for the proper organization of chromosomes in a highly-confined intracellular space, the nucleoid?

A quantitative understanding of chromosome organization has benefited from polymer-chromosome models [13,14,17–19]. Indeed, the theoretical progress with polymers in a cell-like confined space is often paralleled by our effort to understand chromosome experiments [15–19]. Here we present a polymer model of the nucleoid: an asymmetrical ring polymer or a diblock copolymer ring, with a varying degree of asymmetry, trapped in a concentric-shell or simple cylinder; the asymmetric ring is formed by small monomers on one side and big monomers on the other. * Despite its relevance, this model has not been explored systematically, especially in the context of chromosome organization. In order to offer a comprehensive picture, we will explore a wide polymer-

[a] *Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. E-mail: byha@uwaterloo.ca*
[b] *Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Korea. Email: hjeong@kaist.edu*
[c] *Institute for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea; APCTP, Pohang, Gyeongbuk 790-784, Korea.*
[d] *National Institute of Supercomputing and Networking, Korea Institute of Science and Technology Information, Daejeon 305-806, Korea. Email: yjung@kisti.re.kr*
† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/
‡ Present address: Department of Physics, Republic of Korea Air Force Academy, Cheongju, Chungbuk 360-849, Korea.

* A simple cylinder can be considered as a special case of a concentric-shell cylinder, where the radii and lengths of the inner and outer cylinders coincide with each other. The concentric-shell cylinder is essentially the same as the one adopted earlier in a somewhat different context (i.e., chromosome segregation) [17,18]. The main difference is that we use a asymmetric ring polymer or a diblock copolymer ring.

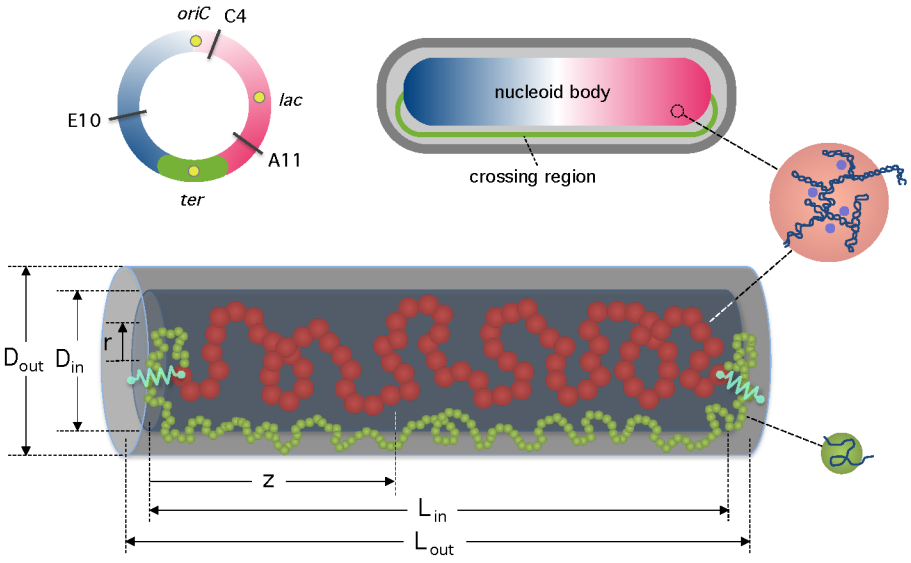


Fig. 1 *E. coli* chromosome and model nucleoid. (Top panel) "Clock" position of genomic loci and spatial organization of the *E. coli* chromosome in the cell. The densely-packed body (in a color gradient) is linearly organized and connected to a loose stretch of DNA segments (in light green). The left (in blue) and right arms (in light strawberry) of the body are located on the opposite sides along the cell length with *oriC* at midcell. (For the details of genomic loci shown on the top left corner, refer to Section 4 and Fig. 2.) (Bottom panel) Asymmetrical ring polymer trapped in a concentric-shell cylinder, where the inner cylinder is surrounded by an outer cylindrical shell. The polymer ring can be viewed as being made of two distinct linear chains (subchains) with their two ends connected to each other so as to form a ring, as shown in the figure. The subchain consisting of big monomers or the body chain is assumed to be confined to the inner cylinder, while the other one formed by small monomers or the crossing chain is free to explore the entire space bounded by the outer cylindrical shell. If the big monomers represent tightly-packed DNA segments that form the body of the chromosome, the small ones mimic loosely-packed DNA segments that make up the crossing region. A body monomer is highlighted on the right, where dots in light purple represent chromosome-associated proteins.

cylinder parameter space. As evidenced below, this effort will not only offer a physical picture of how the bacterial chromosome is organized but also clarify the applicability of polymer-chromosome models, beyond a recent attempt [19].

Here, our main focus is on clarifying to what extent the spatial organization of the chromosomes is owed to the physical effects arising from ring asymmetry and cylindrical confinement. To this end, we consider a few variations of our polymer model. First, we trap an asymmetrical ring in a closed concentric-shell cylinder. When the ring-cylinder parameters are chosen properly and the big monomers are assumed to be localized in the inner cylinder, the resulting locus distributions describe well the known data for the *E. coli* chromosome in slowly-growing cells [8]. Without cell-wall tethering or cross-links between chain segments (by protein H-NS), the two "subchains" forming the ring are spontaneously partitioned in a parallel orientation with the body (big-monomer) chain linearly organized with a desired precision and the crossing (small-monomer) chain preferentially resides in the peripheral space outside the inner cylinder, as described in Fig. 1.

Furthermore, we also examine how loci in the crossing region are spatially organized. In particular, when we introduce a "fluctuating boundary" between the two subchains, the distribution of *ter*-proximate loci becomes double-peak, as seen in experiments [8], which would otherwise remain single-peaked. A related point is that the structural units are rather conceptual entities, as also noted in Ref. [14], and their boundary are highly dynamic [20]. Further efforts along this line will be useful for advancing our understanding of how the chromosome is packed into structural units.

To further explore a polymer nucleoid model, we also consider a ring polymer with a varying degree of asymmetry in a simple cylinder. In this case, we find that a chromosome-like organization of the ring polymer typically requires an external mechanism, i.e., cylinder-wall attachment of a few chain segments; in a biological setting, this can be realized through transertion or any interaction between the chromosome and the cell wall [21]. This may imply that symmetrically organized chromosomes would reply on an external mechanism or some other effects that our coarse-grained model leaves out. A distinguishing feature of the symmetric ring chromosome in fast-growing *E. coli* cells is multi-fork replication [14,22], leading to two duplicated chromosomes topologically-constraining each other. In a cylindrical space, this will create an orientational preference of such rings, which would otherwise be rotationally symmetric along the contour.

It is worth clarifying the scope of this work. Here, we do not attempt to resolve any discrepancy between varying views on chromosome organization (e.g., biological necessity of external mechanisms) or to explore why the mechanism varies among different organisms (e.g., 'donut' vs. 'sausage') [5–7]. What intracellular machinery, if there is any, can generate such physical effects as assumed for the concentric-shell model? Some proteins that bind to specific regions in the chromosome are possibilities [3]. Because of our limited knowledge on their biological role, however, we will not elaborate further on the biological basis of our model. Instead, we will test our model nucleoid against known results and clarify its applicability [5–8,10,13].

Coarse-grained models have already been entertained in the last few years such as the elastic filament model[8] and a cross-linked polymer model[13]; in the latter, a multiply-looped chain via cross-linking is connected to a linear one to form a ring. Each model has both merits and limitations. In terms of complexity, our polymer model is an intermediate one. For this reason, it accomplishes what others may not. Our polymer model can be examined in a wide parameter space, while containing the essential details: chain connectivity, excluded-volume interactions between chain segments, and symmetrical vs. asymmetrical organization as well as a packing uncertainty assumed in our fluctuating-boundary picture. An emerging picture from this consideration is consistent with our view that the necessity of cross-linking varies with the parameter choices. Using our model, we map out a detailed physical picture of how chromosome-like organization is achieved by a confined polymer.

This paper is organized as follows: we outlines the simulation procedure in Sec. 2 and introduce a coarse-grained model of the *E. coli* chromosome in Sec. 3. We present and discuss our results in Sec. 4.

2 Simulation Methods

In our molecular dynamics simulations, the bead-spring model is used for an asymmetrical or copolymer ring, consisting of two types of beads or monomers: "big" and "small." The interaction between two beads, a distance r apart, is described by Weeks-Chandler-Anderson (WCA) potential[23,24]:

$$U_{\rm WCA}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} + \frac{1}{4}\right], & \text{for } r < 2^{1/6}\sigma_{ij} \\ 0, & \text{otherwise} \end{cases} . \quad (1)$$

Here ε and σ_{ij} represent the strength and range of the WCA potential, respectively. The subscripts i and j $(=1,2)$ are used to refer to big and small beads: $\sigma_1 = \sigma_{11}$, $\sigma_2 = \sigma_{22}$, and $\sigma_{12} = \sigma_{21} = (\sigma_1 + \sigma_2)/2$. If we choose $\sigma_2 = \frac{1}{4}\sigma_1$, then $\sigma_{12} = \sigma_{21} = \frac{5}{8}\sigma_1$.

The interaction of monomers with the confining wall can be specified similarly. Here we assume that the wall is made of imaginary beads, which are the same kind as big monomers. (For physics grounds, one can argue that the choice of beads for the wall is not so critical.) As a monomer (big or small) touches the wall, it would be repelled as if there were an "image" bead of size σ_1.

Chain connectivity is ensured by the finite extensible nonlinear elastic (FENE) potential[25] between two consecutive monomers of the same kind or different kind

$$V = -\frac{1}{2}k_{ij}\left(r_{ij}^0\right)^2 \ln\left[1 - \left(r/r_{ij}^0\right)^2\right], \quad (2)$$

where $k_{ij} = k_0\left(\sigma/\sigma_{ij}\right)^2$ is the spring constant, $r_{ij}^0 = r_0\left(\sigma_{ij}/\sigma\right)$ is the range of the potential. (Recall i or j refers to the monomer type: big or small.) In our simulations, $k_0 = 30.0\varepsilon/\sigma^2$ and $r_0 = 1.5\sigma$. Note here that σ is used as length units. For a homogenous polymer, it is customary to set σ to the monomer size.

The velocity Verlet method is used to integrate Newton's equation of motion. If m is the mass of a particle subject to the WCA potential, $\tau_0 = \sigma\sqrt{m/\varepsilon}$ can be used as the unit of time. Let m_1 and m_2 be the mass of big and small monomers, respectively. In our simulations, we set $m = m_1$ and let time advance with a discrete time step $\delta t = 0.0002\tau_0$. Langevin thermostat is used with the damping constant $0.1\tau_0^{-1}$ to keep the temperature at $T = 1.0\varepsilon/k_B$, where k_B is Boltzmann constant. †

Here we explore a few variations of a polymer chromosome model. First, we trap an asymmetrical ring in a concentric-shell cylinder, as shown in Fig. 1. The concentric-shell cylinder consists of an inner cylinder (core region) surrounded with a peripheral region bounded by an outer concentric shell. Its geometry is fully specified in terms of four parameters: $D_{\rm in}$, $D_{\rm out}$, $L_{\rm in}$, and $L_{\rm out}$, i.e., the diameters and lengths of the inner and outer cylinders, respectively. On the other hand, the asymmetrical ring can be viewed as being made of two linear homogeneous "subchains" with the two ends of one chain connected to those of the other so as to form a ring; if the big-monomer subchain or the "body chain" carries N_1 monomers, the other one or the "crossing chain" consists of N_2 small monomers. It is assumed that the big monomers are confined to the inner cylinder or the core region and small ones can be anywhere inside the outer cylindrical shell.

As a variation, we also consider a ring polymer with a varying degree of asymmetry in a simple cylindrical space (i.e, $L_{\rm out} = L_{\rm in}$ and $D_{\rm out} = D_{\rm in}$). Unless the polymer is sufficiently asymmetrical, however, the ring polymer rotates along its contour. For a global organization, we use an external mechanism and connect the two ends of the body chain at the cylinder pole via a molecular spring, characterized by a harmonic potential $U = \frac{1}{2}K(z - z_0)^2$, where K is a spring constant and z is the position of the end monomer attached at z_0 at the cylinder end.

Initially, the polymer ring is confined with the two subchains arranged in parallel in a cylinder with open ends and is almost fully stretched along its long axis. We then equilibrate the ring for more than 10^7 time steps. After equilibration, the cylinder is closed with pistons and the polymer is compressed gradually over 10^7 time steps by moving the piston inward until the piston-piston distance reaches a designated value $\sim L_{\rm in}$. The system is allowed to equilibrate from its initial condition. This two-step procedure for chain equilibration is desired for our much compressed polymer, since it ensures that the initial state is free from unfavorable entanglement or kinetic trapping. It would not, however, influence equilibrium quantities. After this initial preparation, we perform the simulation run for 2×10^9 time steps and obtain data every 2000 time steps. All simulations are repeated for eight different initial chain conformations randomly chosen. To obtain an ensemble average of any relevant quantity, we first take its time average for a given initial state and then obtain an average over all realizations of the initial state randomly chosen.

We use σ defined below Eq. 2 and m_1 as units of lengths and mass in our system, respectively; in most cases considered in this work, σ_1 will be chosen to be σ.

† The choice of the damping constant is not so crucial, since we focus on equilibrium quantities. It is more crucial for dynamics[26].

3 Modeling the nucleoid

First, we introduce a coarse-grained model of the *E. coli* chromosome based on the known results: the circular chromosome is negatively supercoiled into topologically-independent domains or structural units, tightly packed in part by molecular crowding inside an overall-cylindrical space, i.e., the nucleoid[20,27–29] (see Ref.[24] for recent numerical and theoretical efforts). The structural units and their boundaries are highly dynamic[20]. Beyond this general picture, however, the details vary from reference to reference[20,27–30]. Early studies suggest that each domain or structural unit of the chromosome consists of about tens of thousand base pairs of DNA[20,27,28]. As a result, the chromosome carries 100-500 structural units of size $\sim 100\,\mathrm{nm}$ each. More recent studies suggest that the number of structural units is in the range 15-65 (per chromosome) with the size of each in the range 130 to $440\,\mathrm{nm}$[14,29].

This degree of variance is not surprising, considering that the structural unit is rather a conceptual entity; there is no clear-cut boundary between the two neighboring units or domains. It is essentially set by the collective action of various chromosome-associated proteins, especially supercoiling/cross-linking ones. Separating length scales in terms of protein action may not necessarily be unique. For instance, see Ref.[29] for a recent attempt, in which the structural unit was interpreted as a length scale, inside which the effect of cross-linking is not explicitly felt. ‡

The dimensions of *E. coli* cells/nucleoids and the global organization of the chromosome vary with growth conditions and cell ages. As noted earlier, the chromosome resembles a donut (or a branched donut)[14] under fast-growth rates or a sausage with a stretch connecting its ends under slow-growth rates[5–7]. New born-cells under slow-growth rates contain single sausage-like chromosomes; about 90% of the chromosome forms the body and the rest makes up the crossing region, as illustrated in Fig. 1. The measured nucleoid dimensions range from $1.64\,\mu\mathrm{m}\times 0.48\,\mu\mathrm{m}$[31] to $1.8\,\mu\mathrm{m}\times 0.8\,\mu\mathrm{m}$[32]. These deviate somewhat from the earlier measurement[27]: $1.39\,\mu\mathrm{m}\times 0.24\,\mu\mathrm{m}$ (this length is somewhat smaller than the population average $1.9\,\mu\mathrm{m}$[18]). Fast-growing cells are somewhat larger[14] but otherwise their sizes fall in similar ranges.

While various nucleoid-associated proteins are known in the literature[18,30], the precise mechanism of chromosome organization is still elusive. For instance, the physical effects the *local* action of proteins brings about have not been well understood at the quantitative level (see for instance Refs.[18,30]). In our approach, they are coarse-grained into a few cylinder-polymer parameters (e.g., ring asymmetry). In particular, the structural units, which are topologically constrained, are approximated as repelling monomers of size σ_1 each (defined in Sec. 2). We then employ a few variations of a confined-polymer model for exploring how the *E. coli* chromosome is spatially organized: (i) an asymmetric ring polymer formed by small monomers on one side and large ones on the other, confined (but otherwise free) in a concentric-shell cylinder (see Fig. 1), (ii) an asymmetric ring chain with the two big end monomers fixed at the two ends (poles) of a simple cylinder via molecular springs, and (iii) an asymmetric ring with a varying degree of symmetry in a simple cylinder (this includes a symmetric chain as a special case) without any cylinder-wall attachment. Note however that we primarily use model (i) and employ others for comparison purposes. Finally, we choose σ_1 in this model as length units and set σ_1 to σ, which is introduced below Eq. 2.

For the asymmetric case in a concentric-shell cylinder (model (i)), we choose $D_{\mathrm{in}}=7$, $D_{\mathrm{out}}=8$, and $L_{\mathrm{in}}=28$ (4:1) in units of σ, where (...) is the corresponding aspect ratio. Here and below, we measure lengths in units of σ and mass in units of m described in Sec. 2, unless otherwise stated. As for the number of big monomers N_1, we try a set of acceptable values: $N_1=\{80,120,160,200,...,440\}$. For $L_{\mathrm{in}}=28$ and $N_1=200$, the volume fraction of monomers is about 0.1 or 10%.

On the other hand, the crossing region constitutes about 9% of the chromosome or contains about 460K base pairs[8]. This loosely-packed DNA stretch connects two ends of the body a few micrometers apart. The persistence length of double-stranded DNA is about 150 base pairs (50nm). If we consider 300 base pair (100nm) equivalents as Kuhn lengths and the crossing region as a freely-joined chain with self-avoidance, we are tempted to choose $N_2=1800$. Considering that these 100nm subunits experience four-fold compation by supercoiling, the diameter of small monomers mimicking the crossing region is set to $\sigma_2=1/4$. The mass of monomers in the crossing region is chosen to be $m_2=0.015$, which is determined by comparing the mass of 150 base pairs of DNA and the sphere of diameter $1/4$. (As it turns out, however, the parameter choices for the crossing region is not very important, because the spatial distribution of big monomers is not so sensitive to the crossing region.)

Our parameter choices for model (ii) are similar to those for model (i), because of the similarity between the two models. Except for the simple-cylinder geometry used in model (ii), they are essentially identical to those for model (i) with the aspect ratio 4:1. As a result, $D_{\mathrm{in}}=D_{\mathrm{out}}=7$, $L_{\mathrm{in}}=L_{\mathrm{out}}=28$, $N_2=1800$, $m_2=0.015$ and $\sigma_2=1/4$. In addition, the two ends of the body chain are attached to the cylinder ends (poles) through molecular springs characterized by a spring constant K. We choose $K=\{0.1,1.0\}$ (in units of ε/σ^2) such that the energy needed to randomize the parallel orientation of the ring is $\approx 10k_BT$ for $K=0.1$ and $\approx 100k_BT$ for $K=1.0$. Furthermore, we use two different choices of σ_1: $\sigma=1,2$ (in units of σ introduced in Sec. 2). We then choose $N_1=200$ for $\sigma_1=1$ and $N_1=\{20,25,30\}$ for $\sigma_1=2$.

In models (i) and (ii), concentric-shell confinement or an external constraint is required for the proper orientation of a ring polymer. In model (iii), this complexity is removed. Instead, we examine to what extent the global organization is accomplished through the interplay between ring asymmetry and simple-cylindrical confinement. To this end, we use a range of $D=D_{\mathrm{in}}=D_{\mathrm{out}}=\{4,5,6\}$, and the length is $L=L_{\mathrm{in}}=L_{\mathrm{out}}=\{30,50,\infty\}$. Also, the size of big monomers is varied as $\sigma_1=\{1.0,1.1,\cdots,2.5\}$, while the size of small monomers is fixed at

‡ A more traditional view of a domain or a structural unit is 'the region relaxed by an interruption in the DNA'[20]. Similarly, in our picture, the effect of cross-linking does not propagate beyond a domain.

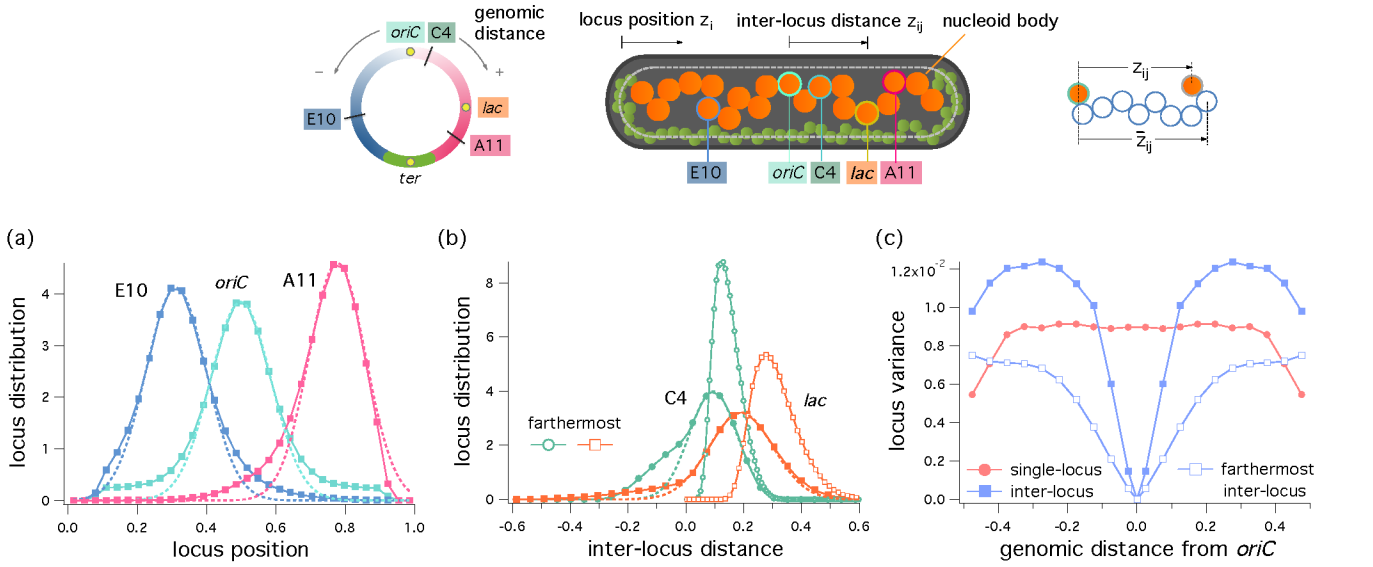


Fig. 2 Chromosome-like spatial organization of an asymmetric ring in a concentric-shell cylinder, referred to as model (i). For this, we have considered the following loci, as shown on the top left panel: *oriC*, C4, *lac*, A11, and E10 on the asymmetric *E. coli* chromosome[8]. Their relative genomic positions with reference to *oriC* are about 7%, 23%, 34%, and 73% (or -27%) for C4, *lac*, A11, and E10, respectively; the different packing level between the body and the crossing region is reflected in the corresponding contour position in our polymer model. (Also shown is *ter* at a 50.0% position from *oriC*.) Finally, we have chosen $N_1 = 200$, $N_2 = 1800$, $L_{\rm in} = 28$, $L_{\rm out} = 31$, $D_{\rm in} = 7$, and $D_{\rm out} = 8$. (a) Single-locus distributions of three loci chosen to correspond to E10, *oriC*, and A11. In this graph, the locus position is given in units of $L_{\rm in}$. The distributions of E10 and A11 are asymmetrical, because of the cap confinement at both ends of the cylinder. Note that each curve has a long-decaying tail, which is absent in the corresponding chromosome-locus distribution[8]. The Gaussian-fits to the distributions, represented by the dashed lines, are more chromosome-like. (b) Distributions of the inter-locus distances of two loci C4 and *lac* from *oriC* (curves with filled symbols). Similarly to the single-locus distributions in (a), the inter-locus distance distribution has a long-decaying tail. Also superimposed are Gaussian fits (dashed lines) and the distributions of inter-locus distances chosen to be the farthermost distances (solid lines with unfilled symbols), as illustrated on the top right corner. The latter quantities are also chromosome-like. (c) Single locus and inter-locus variances for the body vs. the genomic distance from *oriC* in units of the total genomic length of the body chain; see the top left panel for the sign convention for the genomic distance. If the curves with filled symbols are obtained with the Gaussian fits to single-locus and inter-locus distributions, the curves with open symbols are based on the farthermost distance distributions. The single-locus variance is roughly flat for loci somewhere in the middle of the cell but decreases with the genomic distance near the end of the body, as seen with the *E. coli* chromosome[8]. In contrast, the inter-locus variance is V-shaped with a decreasing or flat tail. We favor the inter-locus variance based on the farthermost-distance distribution, as it is more chromosome-like (see the Appendix for additional details).

$\sigma_2 = 1$. The mass of all monomers is $m_1 = m_2 = 1$ (in units of m), regardless of the diameter of the monomers. Our choices include $\sigma_1 = 1.0$ and $\sigma_1 \geq D/2$ as special cases: if the former represents a symmetric ring, the latter ensures that chain back-folding is forbidden.

4 Spatial organization of a model nucleoid

In this section, we first employ model (i), i.e., an asymmetric ring polymer or a diblock copolymer ring in a concentric shell cylinder, and examine how it is spatially organized. We then contrast it with other models.

4.1 Spatial distributions of loci in the body

We first test our polymer model (i) against the known distributions of several loci: *oriC*, C4, *lac*, A11, and E10, chosen from Ref.[8] (see Figs. 2 and 4 in Ref.[8] as well as our Figs. 1 and 2). § What is relevant to our modeling is their genomic position with reference to *oriC*. Their genomic positions in units of the total genomic length are about 7%, 23%, 34%, and 73% (or −27%) for C4, *lac*, A11, and E10, respectively. ¶ The corresponding contour position in our polymer model should reflect the varying packing level between the body and the crossing region, as discussed in the last section. The success of our modeling depends in part on how we choose the parameters. For this model, we have chosen the parameters $N_1 = 200$, $N_2 = 1800$, $L_{\rm in} = 28$, $L_{\rm out} = 31$, $D_{\rm in} = 7$, and $D_{\rm out} = 8$. As a result, the expect ratio of the inner cylinder is 4:1.

The graph in Fig. 2(a) shows the resulting spatial distributions of *oriC*, A11, and E10 in the body (solid lines with symbols) as well as Gaussian fits to them (dashed lines). Except for its long decaying tail, the distribution of *oriC* is essentially Gaussian. On the other hand, the distribution of the other two loci (E10 and A11) is somewhat asymmetrical, since they experience the asymmetric confinement from the cylinder poles. Similarly to the distribution of *oriC*, the other distributions also have a long-decaying tail but on one side of the peak. Excluding the long-decaying tail, the results in Fig. 2(a) are nearly the same as those reported in Ref.[8] in all respects: the location and width ($\sim 10\%$ of cell length) of their peak as well as the distortion of an otherwise Gaussian distribution near the end of the cell.

What is the origin of the long-decaying tail which is absent in the observed chromosome locus distribution? Conversely speaking, what biological effects diminish the long-decaying tail in an otherwise polymer-like object? Cross-linking is an obvious choice. Some DNA-bound proteins (e.g., MukB and H-NS) cross-link two segments (see Ref.[29,30] and references therein). Such topological constraints freeze some of the positional fluctuations of segments, and can influence the locus distribution. As a result, the long-decaying tail will be suppressed more effectively than the peak-proximate part of the distribution, because of a higher free energy cost for the long-tail fluctuation or the positional fluctuation of loci that contributes to the long tail of locus distributions. The Gaussian fitting suppresses this unfavorable fluctuation, similarly to what we would expect from cross-linking. See below in this subsection for additional discussions.

According to our results in Fig. 2(a), it appears that the proper positioning ($\sim 10\%$ of cell length), especially of *oriC* at midcell, would not require an external mechanism. With our parameter choices, the effect of cross-linking is rather quantitative and can be mimicked by suppressing the long-decaying tail of locus distributions. (Note cross-linking between proximate segments along the contour is already absorbed into monomers.) The physical effects arising from concentric-shell confinement and excluded volume interactions suffice. This does not, however, exclude the possibility of an external mechanism such as attachment of DNA segments to the cell surface. What remains unclear is the molecular origin of concentric-shell confinement, which can be clarified in future work.

Fig. 2(b) shows the distribution of inter-locus distances of two loci from *oriC*: C4 and *lac*, also chosen from Ref.[8]. Similarly to the single-locus distributions in Fig. 2(a), the inter-locus distribution (curve with filled symbols) also displays a long-decaying tail. The dashed line is a Gaussian fit to the original distribution. In parallel with what was observed with single-locus distributions, the Gaussian-fit distribution appears to be more chromosome-like[8].

Also superimposed are the distributions of inter-locus distances defined as the 'farthermost distances'[33]. Consider a subchain consisting of monomers $i, i+1, ..., j$, as illustrated on the top right corner of Fig. 2. If $z_{ij} = |z_i - z_j|$ is the original (end-to-end) distance of this subchain (in the longitudinal direction), $\bar{z}_{ij}$ represents the farthermost distance. Alternatively, imagine enclosing it with an imaginary tube. The length of the tube coincides with $\bar{z}_{ij}$. Intriguingly, the resulting distribution is in good qualitative agreement with the Gaussian fit, in the sense that the long-decaying tail is suppressed. The only difference is that the farthermost-distance distribution is somewhat narrower and its peak is a bit shifted to the right. For the reason discussed below, we favor the inter-locus distribution based on the farthermost distance.

Of related interest are the variances of the locus positions in the body: both single-locus and inter-locus variances as a function of the genomic distance from *oriC* in units of the body-chain length (see the illustration on the top left corner in Fig. 2 for the sign convention for the genomic distance from *oriC*). A natural way to estimate a single-locus variance is to use the Gaussian-fit function rather than the original distribution. Note that this will merely improve a quantitative agreement between our estimates and the chromosome data[8] by suppressing the long-decaying tail of the distribution (see the Appendix for additional details). The resulting locus variances are represented by the red (single-locus) and blue (inter-locus) line with filled symbols in Fig. 2(c).

However, we note that this method for calculating inter-locus variances is rather tedious though systematic. Furthermore, the results in Fig. 2(b) suggest that the original inter-locus distances can be coarse-grained into the farthermost distances. Also shown Fig. 2(c) is the inter-locus variance based on the farthermost-

§ Refer to the Supporting Information of Ref.[8] for the locus labeling scheme and for the details of the *E. coli* strains (IL01t, IL05, and IL06) used for chromosome-locus measurements.

¶ We estimated the genomic positions of these loci against the MG1655 *E. coli* genome sequence (4641652 base-pair long) by using the information in Ref.[8] and http://www.ecogene.org.

distance distribution.

To further test our analysis, we have obtained locus variances using a constrained asymmetric ring polymer in a simple cylinder or model (ii) for $N = 25$ and $\sigma_1 = 2\sigma$ (so as to keep the volume fraction of monomers unchanged) and presented the results in the Appendix. Recall that this choice falls in the acceptable range for the *E. coli* chromosome. The resulting "bare" variances compares favorably with the results in Fig. 2(c). In particular, the inter-locus variance appears to agree better with the farthermost-distance variance than with the one based on a Gaussian approximation. Accordingly, we favor the farthermost inter-locus variance and use it below as a representative one, unless otherwise stated. Importantly, this extra consideration validates our intuition: choosing bigger monomers is equivalent to subsuming more details (e.g., cross-linking) into each monomer, as also noted in Ref.[19]. This is consistent with our view that the degree with which cross-linking alters locus distributions varies with the parameter choices.

The results for the single-locus variances in Fig. 2 agree well with the chromosome data in Ref.[8]. In particular, the width of the single-locus distribution or the standard deviation of the locus position, estimated to be $\approx 0.09 L_{\rm in}$, is essentially identical to the experimental finding. Also the inter-locus variance relative to *oriC* (the blue solid line with filled or open symbols) is initially V-shaped but decreases or becomes flat toward the body ends, because of the pole-confinement effect they experience, similarly to what we expect from a mechanical-spring analogy (with the two ends held fixed) and in accord with Ref.[8]. The difference between the two sets of inter-locus variances can be attributed to the way they are coarse-grained into chromosome-like ones – so as to mimic the effect of cross-linking as discussed earlier. Because of the uncertainties in locus measurements[8] (see the Supporting Information), further clarification of our polymer model would necessitate more accurate data for locus variances.

The good agreement between our results and experimental data[8] is rather unexpected in light of a recent computational approach, in which the significance of cross-linking is highlighted[13]. As noted recently[19], however, some discrepancy between polymer models is only deceptive one. Also as noted above, the extent to which cross-linking modifies variances depends on the parameter choices.

Why is an overall rotation of the ring polymer along its contour discouraged in model (i) (in the absence of an external mechanism), as implied by our results in Fig. 2? For the parameter choices relevant for *E. coli*, a linear chain in a cylindrical space is linearly organized (see Ref.[19] and references therein). It is thus unlikely that the chain is reoriented with the two ends switched in the opposite directions. This picture does not necessarily applies to a ring polymer. In our view, the global organization in model (i) is a combined effect of the nucleoid (concentric-shell) geometry and the asymmetrical-ring nature. Unlike a linear chain, a symmetrical ring can rotate along its contour in a "simple" cylindrical space with no free energy cost, because of its symmetry. In contrast, the rotatability of an asymmetrical ring depends on the ratio σ_1/σ_2 as well as D, among others. In the limit of $\sigma_1/\sigma_2 \gg 1$ and $\sigma_1 \to D$, rotation is discouraged. In this case, the free energy cost for overlapping sections of the body chain into a "hairpin"[15] disfavors a rotation of the ring.

For the values of D, σ_i, and N_i $(i = 1,2)$ relevant for the *E. coli* chromosome, however, the asymmetrical ring can rotate in a simple cylinder but *not necessarily* in the concentric shell (see Fig. 6). When the body chain is localized in the inner cylinder, the crossing chain is better excluded from the inner cylinder. As a result, the linear organization of the body chain is better preserved in the concentric-shell cylinder than in an equivalent simple cylinder. This is responsible for the observed global organization of the ring polymer. As detailed in the next subsection, the choice of N_1 is also implicated in the overall orientation of the ring.

As mentioned earlier, our results do not necessarily exclude the possibility of an external mechanism for nucleoid positioning. They rather imply that an external or internal constraint would be required for the precise single-locus positioning of a symmetrical ring (donut-shape chromosome). For instance, the donut-shape chromosome in *Caulobacter crescentus* is known to be anchored to the cell wall at *ParS* sites[11].

4.2 Partitioning of subchains

A recent experimental study suggests that the terminal region resides preferentially in the peripheral region of the nucleoid[10]. To test our model against this observation, we have calculated monomer distributions and plotted our results in Fig. 3. In all graphs, tangerine is used for the body chain and green for the crossing chain. We first use model (i) with the same parameters used in Fig. 2 (e.g., $N_1 = 200$). Fig. 3(a) shows the resulting volume fraction of monomers in the radial (r) and longitudinal (z) directions, as illustrated on the top. This graph suggests that the two subchains are partitioned in a parallel orientation with the crossing chain residing preferentially in the peripheral region and its two ends localized at the poles, similarly to what is known about the asymmetric *E. coli* chromosome[10]. Linear organization of the body along the long axis of the cylinder (Fig. 2) means that on average it fills the inner cylinder uniformly and crowds out the crossing chain, leading to the observed sub-chain organization in Fig. 3(a).

What is less clear is the biological basis of the concentric-shell cylinder model. But the consequence is obvious. In the corresponding simple-cylinder model referred to as model (ii) (see the illustration on the top panel in Fig. 3(b)), in which the body-chain ends are attached to the cylinder poles, the general trend observed with the concentric-shell cylinder model is somewhat reversed, as shown in Fig. 3(b). The body chain tends to be excluded from the central part of the cylinder, where the crossing chain preferentially resides. This can be understood based on the following physical picture. First, ignore the crossing chain. The big monomers will be distributed so as to maximize the entropy (energy is irrelevant in an athermal case as in our consideration here). Earlier, it was shown that monomers in a confined space show wall-layering, i.e., leading to their volume fraction peaked near the cylinder wall[33], as seen in the results in Fig. 3(b).

While the general picture in Fig. 3 is relevant for the asymmetric *E. coli* chromosome (from a slowly-growing cell), it should

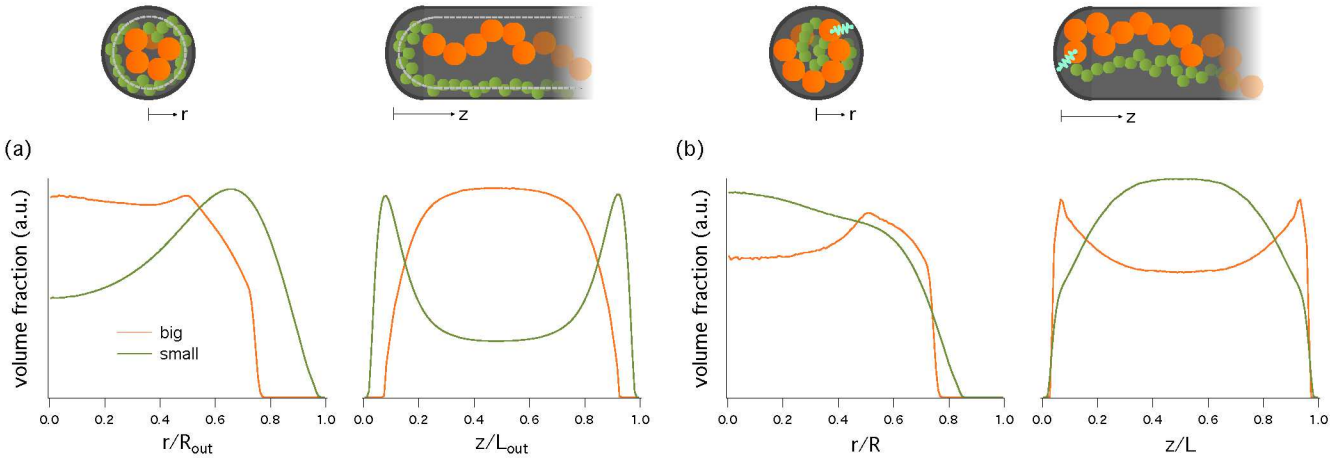


Fig. 3 Monomer distributions in the radial (r) and longitudinal (z) direction (in arbitrary units), as illustrated on the top. In all graphs, the curve in tangerine corresponds to the body chain and the one in green represents the crossing chain. If the chain in (a) is trapped in a concentric-shell cylinder, the one in (b) is in a simple cylinder with the end monomer attached to the cylinder pole. (a) In the concentric-shell cylinder, the body chain is confined in the inner cylinder. The crossing chain thus tends to be excluded from the inner cylinder and preferentially resides in the peripheral region bounded by the two concentric shells. Our estimated cumulative density suggests that about half of small monomers are in the pole region. The general tendency shown in (a) is consistent with the known results [10], even though the biological basis of the concentric-shell cylinder model is not so clear. (b) The trend seen in (a) is now somewhat reversed in the simple cylinder. The body chain tends to be excluded from the central part of the cylinder, where the crossing chain preferentially resides. Using the results in this figure, the relevance of the concentric-shell cylinder can be further tested against experiments.

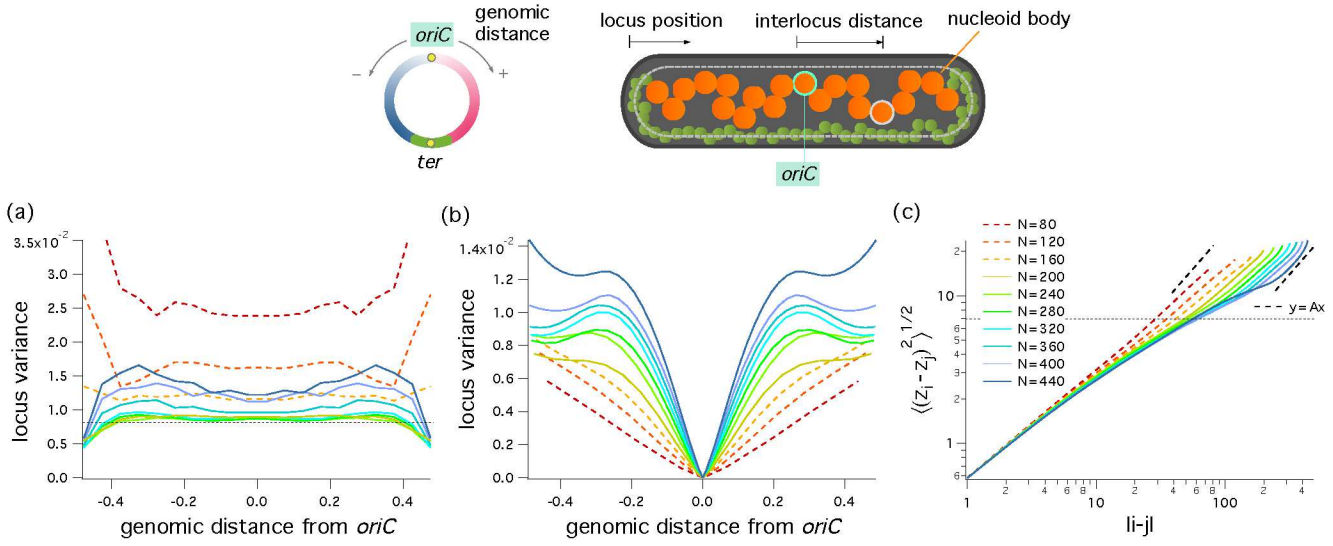


Fig. 4 Single-locus (a), inter-locus (b) variances, and linear ordering (c) for different N (or N_1) values; if the dashed horizontal line in (a) corresponds to an accuracy (standard deviation) of $0.09 = 9\,\%$ as seen for *E. coli* chromosomes [8], the one in (c) coincides with the value of $D_{\text{in}} = 7$. The results in this figure show how to choose N for chromosome-like organization. There appears to be a qualitative difference between the small and large-N cases, i.e., $N < 200$ and $N \geq 200$, respectively. While the results for $N \geq 200$ compare favorably with *E. coli* data, those obtained for $N < 200$ deviate from the data. In particular, the plateau in the large-N single-locus results disappears (a), and the inter-locus variance increases monotonically for small N (b); also, the chain loses its linear ordering along the long axis of the cylinder for small N (c). This can be attributed to the rotation of the ring along its contour. The results here imply how crucial the parameter choices are.

not be over-interpreted. For the symmetric-ring chromosome (under fast-growth conditions), there is now much evidence that the two arms or subchains of the ring are well-separated in the radial direction[14]. This can be attributed to two complementary effects: excluded-volume repulsion between the two subchains and crowding effects[14,29]; the crowding effect can induce a depletion force not only between two monomers and but also between a monomer and the cylinder wall. A recent study shows that under certain conditions, the latter dominates, leading to cylinder-wall adsorption of chain segments and a separation of the two opposing arms[34].

4.3 The parameter N_1 - how to choose

The parameter choices have so far been motivated by the existing *E. coli* data. Nevertheless, the parameter N_1 (or the number of structural units) is more variable than others. (In this subsection, we drop the subscript 1.) To understand how crucial the choice of N is, we have calculated and plotted locus-variances for a sizable range of N values and the mean internal distance $\langle (z_i - z_j)^2 \rangle^{1/2}$ in Fig. 4(a)(b) and Fig. 4(c), respectively. For the reason described above, we have obtained single-locus variances using Gaussian-fit approximations and inter-locus ones using the farthermost-distance distributions. On the dashed horizontal line in (a), the accuracy or standard deviation of locus positions is $0.09 = 9\%$, as seen with *E. coli* chromosomes[8]; the dashed horizontal line in (c) is to denote the value of $D_{\rm in} = 7$. Different colors correspond to different N values and the color scheme is the same in all graphs in Fig. 4. (The illustration on the top left corner includes the sign convention for the genomic distance from *oriC*.)

Let us first consider the locus variances as a function of the genomic or contour distance in units of the body-chain length in Fig. 4(a)(b). The difference between the small and large-N cases, i.e., $N < 200$ and $N \geq 200$, appears to be qualitative. While the results for $N \geq 200$ are chromosome-like, those obtained for $N < 200$ deviate from what one would expect from chromosomes. Notably, the effect of cylinder-end confinement is felt differently: for large N, this effect tends to suppress single-locus variances, as seen with a constrained mechanical spring with the two ends held fixed. The results for large N ($N = 200, 240, 280, \ldots$) in both (a) and (b) compare favorably with *E. coli* data[8].

The large-N behavior of the locus variances in Fig. 4 is correlated with what the results in Fig. 2 suggest: the body chain compressed longitudinally but remaining linearly organized (see Fig. 4(c) for this), effectively crowding out the crossing chain from the inner cylinder. As a result, the confined ring is globally organized with the two subchains in parallel. This explains the main feature of the results in Fig. 4(a)(b) for ($N \geq 200$), in particular the decreasing tail of single-locus variances toward the body-chain ends, preferentially localizing at the cylinder ends.

The maximum value of the single-locus variances is larger for larger N as long as $N \geq 200$, as shown in Fig. 4(a). This observation can be explained in terms of the N dependence of $\zeta_\parallel$, the length scale beyond which a confined chain is linearly organized[16]. Within this length, the chain is disordered. The emergence of this additional length is unique to the case of closed confinement or a compressed chain[16]. In this case, $\zeta_\parallel$ increases as N increases. ‖ This is consistent with the observed N dependence of single-locus variances for $N \geq 200$. **

For small N, however, the crossing chain is less excluded from the inner cylinder and mix better with the body chain. This can lower the free energy cost for chain rotation, as if both subchains *were* in the inner cylinder. This enhances the positional fluctuation of monomers near the end of the body chain, which would otherwise be confined to the pole region. As a result, the trend observed with the large-N single-locus variance is reversed in this case. This also explains the monotonically increasing inter-locus variance for small N in Fig. 4(b). This drawback can be remedied by attaching chain segments to the cylinder wall as discussed in the Appendix. On the other hand, the inter-locus variances are less sensitive to the degree of a global organization. This explains why they change more or less uniformly with N, even though there is still a noticeable difference between the two sets of data (large N and small N).

4.4 Spatial Organization for the Crossing Region: fluctuating boundary model

We have further explored our model nucleoid and examined how loci in the crossing region are spatially arranged along the long axis of the cell. Less is, however, known about the crossing region compared to the nucleoid body[8,10]. In this section, we content ourselves by mapping out a few possible scenarios that can be tested experimentally in the future. Nevertheless, we will attempt to put this effort in a biological context and use our predictions to understand quantitatively relevant biological data.

For this consideration, we use the concentric-shell cylinder model (i). First, we have obtained and plotted in Fig. 5(a) the locus distributions of three loci in the crossing region: $1/2$ and $2/3$ from *ter* in units of the genomic length of the crossing region as well as *ter* itself, assumed to be located opposite *oriC*. We have chosen $N = 200$ and the same cylinder parameters used in the previous subsections. As shown in the graph, the locus near the body ($2/3$ from *ter*) tends to be at the cylinder end. On the other hand, the locus *ter* has a broad distribution centered in the middle. The middle one ($1/2$ from *ter*) seems to have a distribution that falls between the other two.

However, these results are qualitatively different from the experimental data that indicate 'M'-shape or double-peak distributions of *ter* or *ter*-proximate sites (see the Supporting Information of Ref.[8]). In our original model, the boundary between the body and the crossing region is fixed. We now relax this constraint and introduce a "fluctuating-boundary" model, where the boundary is defined with some precision (see the illustration on the top of

‖ This trend will not be reversed by Gaussian-fit approximations or coarse-graining of lengths through $z_{ij} \to \bar{z}_{ij}$.

** The N-dependence of locus variances shown in Fig. 4 is not so conclusive for chromosome organization, since cross-linking or other biological effects are implicated in the latter case. These effects are only approximately captured in our results. In particular, there is evidence that the bacterial chromosome is length-wise folded[35], i.e., proximate segments along the contour are more closely clustered. As discussed recently[19], this can be considered as shortening $\zeta_\parallel$, more so for larger N.

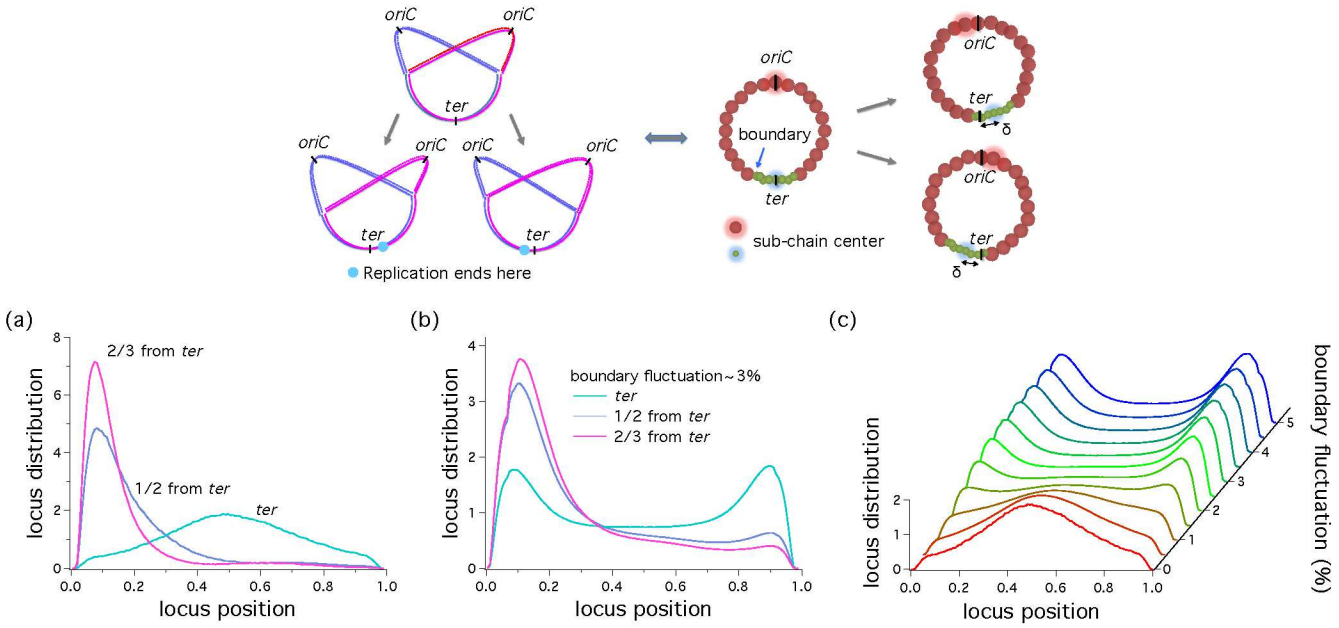


Fig. 5 Fixed vs. fluctuating boundary between the body and crossing chains, and the spatial organization of the crossing region. Shown on the top are the biological origin of the fluctuating boundary (Left) and the corresponding asymmetric-polymer model (Right). (a) Single-locus distributions of three loci in the crossing region with our original model (i): a fixed boundary between the body and the crossing region. Here the number like 2/3 stands for the contour or genomic distance from *ter* in units of the total contour length of the crossing region. The *ter*-distant monomer described by the blue or magenta curve tends to reside in the cylinder end. The locus at the center of the crossing-region, i.e., *ter*, has a broad distribution at the center of the long axis of the cylinder. These results are in conflict with the experimental data indicating M-shape distributions[8]. (b) In our "fluctuating-boundary" model, the boundary between the body and crossing region is defined with some precision: the genomic position of the boundary is assumed to follow a Gaussian distribution with a standard deviation δ. When δ is chosen to be 3% of the genomic length, the distributions of the loci in the crossing region become M-shaped and agree well with the data[8]. (c) This graph shows how the distribution becomes M-shaped or double-peaked as δ increases; a δ value somewhat larger than 2% is required for a noticeable M shape.

Fig. 5); the genomic position of the boundary is assumed to follows a Gaussian distribution with a standard deviation δ. When $\delta = 3.3\,\%$ is used, the distribution of the *ter* site in Fig. 5(b) becomes M-shaped, similarly to what was seen with the asymmetrical *E. coli* chromosome[8].

If taken together with our results in Fig. 5, the observation of M-shaped distributions[8] appears to be a result of molecular uncertainties in chromosome packing, as reflected in our fluctuating-boundary model. This idea is *in part* (but not solely) inspired by a recent experimental observation regarding DNA replication in slowly-growing *E. coli* cells[36]. As illustrated on the top of Fig. 5, the circular DNA in the cell is replicated bi-directionally from the single replication origin *oriC*. It was shown that replication in one direction (in the left replichore) proceeds a bit faster, i.e., by $\sim 7\,\mathrm{min}$[36]. Since replication and packing occur concomitantly, this "asymmetry" between left and right replichore replication can then lead to packing asymmetry as assumed in the fluctuating-boundary model and as illustrated in Fig. 5. Considering the time for replicating the entire genome is $\sim 61\,\mathrm{min}$, this difference translates into $\delta = 5.7\,\%$ [††], which is comparable to what we used.

A related (possibly complementary) point is the molecular mechanism by which DNA replication ends. The terminus region contains ten 23 bp *ter* sites: five on the left replichore and five on the right replichore[37,38]. As the *E. coli* chromosome is replicated bi-directionally from *oriC*, the two replicating forks proceed in the opposite directions and meet about halfway from *oriC* in the terminus region; replication stops there. More precisely, the determination site falls between the two closest *ter* sites on the opposite replichores, which are about $5.8\,\%$ of the entire genome apart[37,38]. This is comparable to 2δ we chose.

Because of the coarse-grained nature of our approach and biological uncertainties, especially about chromosome packing, the value of δ is not to be taken too literally. Depending on how this is taken into account, a *ter*-proximate locus can be localized at one of the poles or elsewhere. To map out a general picture of how the choice of δ influences *ter* distributions, we have obtained *ter* distributions with various choices of δ and plotted them in the graph in Fig. 5(c). The appearance of an M or double-peak distribution is obvious for $\delta \approx 3\%$.

Despite some success in accounting for the distributions of *ter*-proximate loci, our discussion in this subsection remains open for further exploration. To the best of our knowledge, the mechanism for packing the bacterial chromosome into the body and crossing region has not been well understood[4]. A related observation is that the boundary between structural domains is dynamic rather than static[20]. How these uncertainties and stochasticity enter into the picture is unclear. Further considerations along this line will be desirable.

†† Recall how δ is defined in Fig. 5. This is the "width" of the Gaussian *ter* distribution measured from the peak location.

4.5 Symmetrical vs. asymmetrical rings

So far, we have used a concentric-shell cylinder or an external mechanism so as to partition the subchains properly for our parameter choices. For the typical parameters used in Fig. 2 and Fig. 3, the asymmetrical ring tends to rotate along its contour in a cell-mimicking simple cylinder, but not in a corresponding concentric-shell cylinder. To further exploit our ring polymer model and map out a more complete polymer picture of the nucleoid, especially under different growth conditions, we have also considered a ring polymer with a varying degree of ring asymmetry trapped in a simple cylinder. In particular, we have examined the necessity of an external mechanism for the global organization of the confined polymer, e.g., *oriC* localized at midcell. To this end, we have examined the degree with which the two subchains are organized in a parallel orientation. It is measured in terms of $\langle\lambda\rangle/L_{\mathrm{ring}}$ and the distribution of λ, where λ is the overlap distance between the two subchains, L_{ring} is the average length of the entire ring, as illustrated in Fig. 6, and $\langle\ldots\rangle$ is an ensemble average. Note that L_{ring} is the maximum value that λ can take on.

On physics grounds, the quantity $\langle\lambda\rangle/L_{\mathrm{ring}}$ is expected to change from $1/2$ (randomly organized or freely rotatable along the contour) to 1 (perfectly aligned) with the standard deviation varying from $\sqrt{1/12}$ to 0, as marked by the dashed lines on the graph in Fig. 6(a). The graph clearly shows how ring asymmetry influences subchain orientation. As σ_1/D increases (i.e., the ring becomes more asymmetrical), parallel organization becomes preferred: $\langle\lambda\rangle \to L_{\mathrm{ring}}$. For this, we have chosen $\sigma_1 = \{1.0, 1.1, \cdots, 2.5\}$, as small as $\sigma_2 = 1$ and as big as half of the cylinder diameter, and $D = 4, 5, 6$ (all in units of σ introduced in Sec. 2). Furthermore, N_1 is chosen to be the biggest integer satisfying $(4/3)\pi(\sigma_1/2)^3 N_1 \leq (4/3)\pi(\sigma_2/2)^3 N_2$, while N_2 is fixed at 100. As a result, some values of N_1 turn out to be too small to represent the *E. coli* chromosome. Here our main focus is on extracting general trends, regarding the global organization of rings, but not on mimicking the chromosome. The meaning of this choice is most obvious for $L = \infty$. In this case, the total volume fraction of monomers remains almost constant, as σ_1 or N_1 varies, allowing us to focus on the "shape" parameters. Additionally, for the symmetric case of $\sigma_1 = \sigma_2$, $N_1 = N_2$.

The transition from the freely rotatable to parallel state occurs at $\sigma_1/D \approx 0.4$-0.45. The main effect of longitudinal compression (squares) appears to make the transition somewhat steeper, possibly by delaying the onset of the transition. With the typical parameter choices used in Figs. 2 and 3, the asymmetric ring rotates in a simple cylinder, but not in a concentric cylinder.

Fig. 6(b) shows the probability distribution of λ for a few choices of σ_1 with $\sigma_2 = 1$. The distribution for a freely rotatable state would be uniform, as described by the red line obtained for $\sigma_1 = \sigma_2 = 1$. The decaying tail of this curve at $\lambda \approx 40$ can be attributed to chain-end fluctuations. When the subchains are organized in parallel conformations, λ tends to L_{ring}, resulting in a narrower distribution, as is particularly the case for the blue line obtained for $\sigma_1/D = 5/12$; this curve corresponds to the blue circle marked by an arrow in Fig. 6(a). Even for as small σ_1 as

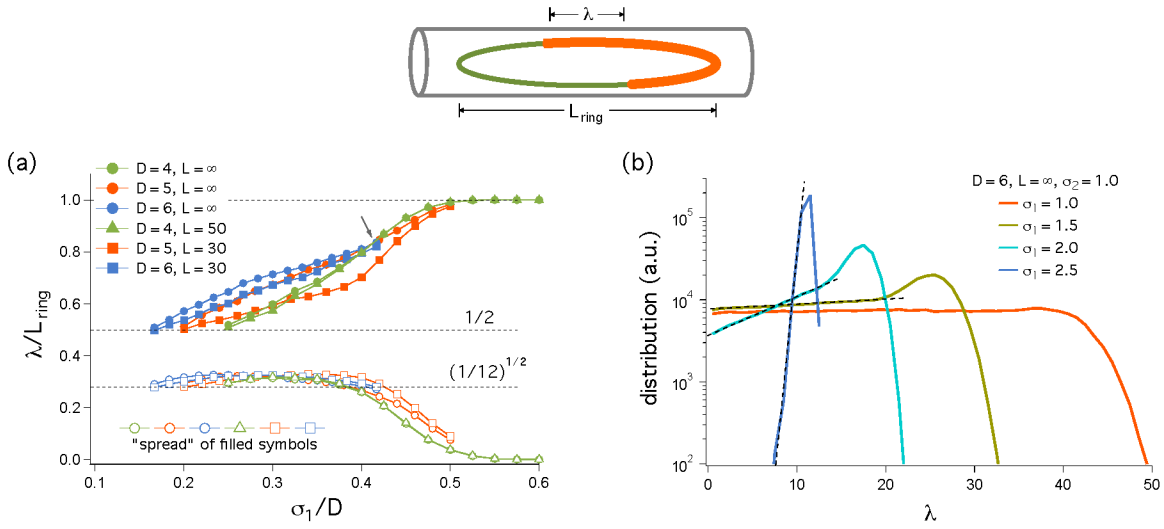


Fig. 6 Subchain organization in a "plain" cylindrical space: parallel vs. random. Here we measure the degree of subchain's parallel organization by $\langle\lambda\rangle/L_{\text{ring}}$ (a) as well as by the distribution of λ/L_{ring} (b), where λ is the overlap distance between the two subchains, L_{ring} is the length of the entire ring, and $\langle ...\rangle$ is an ensemble average. (a) The quantity $\langle\lambda\rangle/L_{\text{ring}}$ is expected to change from $1/2$ (randomly organized or freely rotatable along the contour) to 1 (perfectly aligned) with the standard deviation varying from $\sqrt{1/12}$ to 0, as marked by the dashed lines. The graph in (a) shows that as σ_1/D increases (i.e., the ring becomes more asymmetrical) parallel organization becomes preferred. For this, we have chosen $\sigma_1 = \{1.0, 1.1, \cdots, 2.5\}$ and $D = 4, 5, 6$. Furthermore, N_1 is the biggest integer satisfying $4/3\pi(\sigma_1/2)^3 N_1 \leq 4/3\pi(\sigma_2/2)^3 N_2$, while N_2 is fixed at 100. As a result, for $L = \infty$, the total volume fraction of monomers remains almost constant, allowing us to focus on the "shape" parameters. There appears to be a transition from the freely rotatable to parallel state for $\sigma_1/D \approx 0.4$-0.45. The main effect of longitudinal compression (squares) is to make the transition a bit steeper. (b) Probability distributions of λ. The distribution for a freely rotatable state would be uniform (the red line for $\sigma_1 = \sigma_2 = 1$). When the subchains are organized in parallel, λ tends to L_{ring}, resulting in a narrower distribution (see the blue line obtained for $\sigma_1/D = 5/12$, which corresponds to the blue circle on the right top corner on the left figure, marked by an arrow). Even for $\sigma_1 = 1.5$, the distribution deviates somewhat from the uniform distribution. The deviation on the left side of the peak is well characterized by an exponential fit (the dashed line), meaning that the free energy cost of overlapping two subchains is proportional to the distance $L_{\text{ring}} - \lambda$.

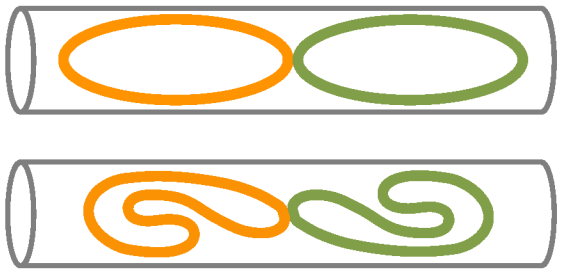


Fig. 7 Hypothetically organized symmetric rings that form an 8-shape structure. Because of the topological constraint, each symmetric ring cannot readily rotate along the contour without any unfavorable consequence, as illustrated in the lower panel; the lower panel shows a possible rearrangement, which is a high-free energy state.

$\sigma_1 = 1.5$, the distribution deviates somewhat from the uniform distribution. The deviation on the left side of the peak is well characterized by an exponential fit (the dashed line), meaning that the free energy cost of overlapping two subchains is proportional to the distance $(L_{\rm ring} - \lambda)$, which can be interpreted as the overlap distance within the subchain with larger monomers. This observation offers a physical explanation for the spontaneous organization of an asymmetrical ring polymer into a parallel connection of its subchains. In the crudest picture, we can ignore the subchain with smaller monomers. As the big-monomer subchain folds up into a hairpin-like structure along the cylinder axis, the free energy cost will be proportional to $(L_{\rm ring} - \lambda)$, as seen in Fig. 6(b) (the dashed lines).

What do our results in Fig. 6(a) suggest for the *E. coli* chromosome? For an asymmetric chromosome in a slowly-growing cell, ring asymmetry and any physical effects that mimic the concentric-shell cylinder contribute toward its global organization and may compliment an external mechanism, if there is any. Besides DNA-organizing proteins, some others act as crowing agents. How these molecules organize such an asymmetric chromosome is largely unknown but can be clarified in the future.

In contrast, in fast-growing *E. coli* cells[14,22], two duplicated chromosomes continue to constrain each other topologically, a distinct feature arising from multi-fork replication. The rotational symmetry along the contour of an otherwise symmetric ring is no longer preserved. Fig. 7 shows two such rings that form an 8-shape structure, a particular realization of topologically-constraining rings. The consequence of rearranging the structure in the upper panel into the one in the lower panel is obvious: the lower one containing a hairpin presents a high-free energy state (see above and Ref.[15]). This physical effect will contribute toward positioning the chromosomes in fast-growing *E. coli* cells. This internal mechanism will not necessarily exclude the possibility of an external one but it rather compliments the latter.

5 Conclusions

In this work, we have examined to what extent a confined polymer shows chromosome-like organization. In particular, our results suggest that an asymmetric ring polymer trapped in a concentric-shell cylinder describes several features of what was observed with the asymmetric chromosome in *E. coli* under slow growth conditions. If the parameters are chosen properly, without cross-linking, the confined ring polymer is organized similarly to what was seen with the *E. coli* chromosome[8]. For instance, our results for locus distributions and variances agree well with the existing experimental data; the agreement is even quantitative, when biologically-inspired coarse-graining schemes (e.g., Gaussian fitting) were used for computing these quantities. Furthermore, our studies with the fluctuating-boundary model show how the biological uncertainty in chromosome packing modifies the spatial distributions of *ter*-proximate loci. The observed M-shape distributions[8] are indirect evidence of this uncertainty.

Our work also clarifies the necessity of cell-wall tethering for the global organization of a ring polymer (e.g., bacterial chromosomes). In a simple cylinder, as the ring becomes more symmetric, a transition occurs from non-rotatable (subchains-in-parallel) to rotatable states (allowing the subchain to form a hairpin near one of the poles). This observation is not conclusive for symmetrically-packed chromosomes as in fast-growing *E. coli* cells, since the resulting rotatability can be opposed by any topological constraints between duplicated chromosomes, forming a 8-shape structure or others with nontrivial topology[14,22].

Because of its simplicity, our model and its variations can be tested and exploited against new experimental data. Conversely, one can design experiments for testing our predictions. This effort will be beneficial for further unraveling how various effects, arising from confinement, chain topology (e.g., branched, 8-shape or θ-shape structure), and cross-linking, are intertwined in organizing the bacterial chromosome. Also how crowding particles (e.g., water-soluble cytoplasmic proteins) are implicated in chromosome packing and organization merits some consideration. Recent experiments suggest that they can alone fit the *E. coli* chromosome into the nucleoid[29]. What is less obvious is how they influence chromosome-locus distributions. Along this line, a polymer model was used to account for the locus distribution of a symmetric *E. coli* chromosome in the radial direction[14]. Inclusion of crowding particles will be useful for further clarifying the applicability of the polymer model.

Use of a coarse-grained model (e.g., polymer models used here) has merits: first, one can explore a wide parameter space, even outside what can be realized in a biological setting, and focus on one effect at a time by "turning off" others. This will help us understand systematically their (relative) significance. To improve such a model, a better understanding of domain (body and crossing)-specific proteins, especially the physical effects they bring about, will be desirable. Importantly, chromosomes are dynamic and heterogeneous structures, which can be reshaped by chromosome-associated proteins as well as by "free" molecules via molecular crowding. The way bacterial chromosomes are organized in crowded cells will remain as an inspiring source for physical modeling.

Acknowledgement

BYH acknowledges useful discussions with P. Wiggins, C. L. Woldringh, and S. Jun. This work was supported by the collaborative research contract funded by Korea Institute of Science and Technology Information (KISTI) and NSERC (Canada) (B-YH) as

well as by the NRF grant funded by the Korean Government (NRF-2013R1A6A3A03064965 (CJ), NRF-2012R1A1A2007488 (YJ), and NRF-2011-0028908 (HJ)).

References

1 T. Cremer and C. Cremer, "Chromosome territories, nuclear architecture and gene regulation in mammalian cells," *Nat. Rev. Genetics*, 2001, **2**, 292-301.

2 C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli and T. Cremer, "Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions," *Nat. Rev. Genetics*, 2007, **8**, 104-115.

3 R. T. Dame, O. J. Kalmykowa, and D. C. Grainger, "Chromosomal Macrodomains and Associated Proteins: Implications for DNA Organization and Replication in Gram Negative Bacteria," *PLoS Genet.*, 2011, **7**, e1002123-1-5.

4 R. Mercier, M. A. Petit, S. Schbath, S. Robin, M. El Karoui, F. Boccard, and O. Espéli, "The MatP/matS Site-Specific System Organizes the Terminus Region of the *E. coli* Chromosome into a Macrodomain," *Cell*, 2008, **135**, 475-485.

5 X. Wang, X. Liu, C. Possoz, and D. J. Sherratt, "The two *Escherichia coli* chromosome arms locate to separate cell halves," *Genes Dev.*, 2006, **20**, 1727-1731.

6 X. Liu, X. Wang, R. Reyes-Lamothe, and D. Sherratt, "Replication-directed sister chromosome alignment in *Escherichia coli*," *Mol. Microbiol.*, 2010, **75**, 1090-1097.

7 H. Niki, Y. Yamaichi, and S. Hiraga, Dynamic organization of chromosomal DNA in *Escherichia coli*, *Genes & Dev.*, 2000, **14**, 212-223.

8 P. A. Wiggins, K. C. Cheveralls, J. S. Martin, R. Lintner, and J. Kondev, "Strong intranucleoid interactions organize the *Escherichia coli* chromosome into a nucleoid filament," *Proc. Natl. Acad. Sci., USA*, 2010, **107**, 4991-4995, and the Supporting Information.

9 V. G. Benza, B. Bassetti, K. D. Dorfman, V. F. Scolari, K. Bromek, P. Cicuta, and M. C. Lagomarsino, "Physical descriptions of the bacterial nucleoid at large scales, and their biological implications." *Rep. Prog. Phys.*, 2012, **75**, 076602-1–076602-20.

10 J.-C. Meile, R. Mercier, M. Stouf, C. Pages, J.-Y. Bouet, and F. Cornet, "The terminal region of the *E. coli* chromosome localises at the periphery of the nucleoid," *BMC Microbiol.*, 2011, **11**, 1-10.

11 M. A. Umbarger et al., "The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation," *Mol. Cell*, 2011, **44**, 252-264.

12 R. Reyes-Lamothe, X. Wang, and D. Sherratt, "*Escherichia coli* and its chromosome," *Trends in Microbiology*, 2008, **16** , 238-245.

13 M. Fritsche, S. Li, D. W. Heermann, and P. A. Wiggins, "A model for *Escherichia coli* chromosome packaging supports transcription factor-induced DNA domain formation," *Nucl. Acids Res.*, 2012, **40**, 972-980.

14 B. Youngren, H. J. Nielsen, S. Jun, and S. Austin, "The multifork *Escherichia coli* chromosome is a self-duplicating and self-segregating thermodynamic ring polymer," *Genes & Dev.*, 2014, **28**, 71-84.

15 Y. Jung, C. Jeon, J. Kim, H. Jeong, S. Jun, and B.-Y. Ha, "Ring polymers as model bacterial chromosomes: confinement, chain topology, single chain statistics, and how they interact," *Soft Matter*, 2012, **8**, 2095-2102.

16 Y. Jung, J. Kim, S. Jun, and B.-Y. Ha, "Intrachain Ordering and Segregation of Polymers under Confinement," *Macromolecules*, 2012, **45**, 3256-3262.

17 S. Jun and B. Mulder, "Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome," *Proc. Natl. Acad. Sci., USA*, 2006, **103**, 12388-12393.

18 S. Jun and A. Wright, "Entropy as the driver of chromosome segregation," *Nat. Rev. Microbiol.*, 2010, **8**, 600-607.

19 B.-Y. Ha and Y. Jung, "Polymers under confinement: single polymers, how they interact, and as model chromosomes," *Soft Matter*, 2015, **11**, 2333-2352.

20 L. Postow, C. D. Hardy, J. Arsuaga, and N. R. Cozzarelli, "Topological domain structure of the *Escherichia coli* chromosome," *Genes Dev.*, 2004, **18**, 1766-1779.

21 C. L. Woldringh, "The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation," *Mol. Microbiol.*, 2002, **45**, 17-29.

22 S. Cooper and C. E. Helmstetter, "Chromosome replication and the division cycle of *Escherichia coli* B/r," *J. Mol. Biol.*, 1968, **31**, 519-540.

23 J. D. Weeks, D. Chandler, and H. C. Andersen, "Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids," *J. Chem. Phys.*, 1971, **54**, 5237-5247.

24 J. Kim, C. Jeon, H. Jeong, Y. Jung, and Bae-Yeun Ha, "A polymer in a crowded and confined space: effects of crowder size and poly-dispersity," *Soft Matter*, 2015, **11**, 1877-1888.

25 K. Kremer, and G. S. Grest, "Dynamics of entangled linear polymer melts- A molecular-dynamics simulation," *J. Chem. Phys.*, 1990, **92**, 5057- 5086; G. S. Grest and K. Kremer, "Molecular dynamics simulation for polymers in the presence of a heat bath," *Phys. Rev. A*, 1986, **33**, 3628-3631.

26 Y. Jung, C. Jeon, M. Ha, and B.-Y. Ha, "Expansion dynamics of a self-avoiding polymer in a cylindrical pore," *Europhys. Lett.*, 2013, **104**, 68003-p1–p6.

27 See C. L. Woldringh and T. Odijk in "Organization of the Prokaryotic Genome," edited by R. L. Charlebois (ASM Press, Washington, D.C. 1999).

28 T. Romantsov, I. Fishov, and O. Krichevsky, "Internal Structure and Dynamics of Isolated *Escherichia coli* Nucleoids Assessed by Fluorescence Correlation Spectroscopy," *Biophys. J.*, 2007, **92**, 2875-2884.

29 J. Pelletier, K. Halvorsen, B.-Y. Ha, R. Paparcone, S. J. Sandler, C. L. Woldringh, W. P. Wong, and S. Jun, "Physical manipulation of the *Escherichia coli* chromosome reveals its soft nature," *Proc. Natl. Acad. Sci., USA*, 2012, **109**, E2649-E2656.

30 J. Stavans and A. Oppenheim, "DNA-protein interactions and bacterial chromosome architecture," *Phys. Biol.*, 2006,

3, R1-R10.

31 J. K. Fisher, A. Bourniquel, G. Witz, B. Weiner, M. Prentiss, and N. Kleckner, "Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells," *Cell*, 2013, **153**, 882-895.

32 I. Junier, F. Boccard, and O. Espéli, "Polymer modeling of the *E. coli* genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation," *Nucleic Acids Res.*, 2014, **42**, 1461-1473.

33 J. Kim, C. Jeon, H. Jeong, Y. Jung, and B.-Y. Ha, "Elasticity of flexible polymers under cylindrical confinement: appreciating the blob scaling regime in computer simulations," *Soft Matter*, 2013, **9**, 6142-6150.

34 C. Jeon et al., private communication, 2015.

35 X. Wang, P. M. Llopis and D. Z. Rudner, "Organization and segregation of bacterial chromosomes," *Nat. Rev. Genet.*, 2013, **14**, 191-203.

36 M. C. Joshia, A. Bourniquel, J. Fisher, B. T. Ho, D. Magnan, N. Kleckner, and D. Bate, "*Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps," *Proc. Natl. Acad. Sci., USA*, 2010, **108**, 2765-2770.

37 C. Neylon, A. V. Kralicek, T. M. Hill, and N. E. Dixon, "Replication Termination in *Escherichia coli*: Structure and Antihelicase Activity of the Tus-Ter Complex," *Microbiol. Mol. Biol. R.*, 2005, **69**, 501-526.

38 M. D. Mulcair, P. M. Schaeffer, A. J. Oakley, H. F. Cross, C. Neylon, T. M. Hill, and N. E. Dixon, "A molecular mousetrap determines polarity of termination of DNA replication in *E. coli*," *Cell*, 2006, **125**, 1309-1319.

Appendix

In the Appendix, we examine how much locus variances depend on the way they are obtained, the model used, and the parameter choices.

Fig. 8(a) shows both single-locus (red) and inter-locus (blue) variances obtained with our concentric-shell cylinder model or model (i) for $D_{\rm in} = 7$, $L_{\rm in} = 28$, and $N = N_1 = 200$. In this graph, three sets of data are compared. The solid lines with filled symbols are Gaussian-fit approximations and the solid line with open symbols is based on the distribution of the farthermost distance $\bar{z}_{ij}$; the dashed lines obtained without any of these approximations are plotted on the right axis. The consequence of using a Gaussian fit or $\bar{z}_{ij}$ is obvious: variances obtained this way are smaller than the original ones (dashed lines). Which set of data is the most chromosome-like?

The results in Fig. 4 suggest that the concentric-shell cylinder model remains applicable for $N \geq 200$. As described earlier, the value of N is to coincide with the number of structural units or topological domains n_d. The acceptable values of N for this model, as suggested by Fig. 4, fall in the rather larger range of n_d. To remedy this limitation and to further exploit our polymer model, we also consider a somewhat distinct model referred to as model (ii): a constrained asymmetrical ring polymer in a simple cylinder with the two monomers at the end of the body chain attached to the cylinder pole via a spring, as illustrated in Fig. 8 (top right corner). The main advantage of this model is that it allows us to choose N smaller than required for model (i). (See Fig. 3 for the drawbacks of this model.) For sufficiently small N, a Gaussian approximation or coarse-graining of lengths via $z_{ij} \to \bar{z}_{ij}$ is not required as evidenced below.

Fig. 8(b) summaries our results obtained with model (ii). For this, we have chosen the parameters as follows: $K = \{0.1, 1.0\}$ (in units of ε/σ^2), $\sigma_1 = 2\sigma$, $\sigma_2 = \sigma/4$, $N = N_1 = \{20, 25, 30\}$ (so as to keep the volume fraction roughly the same as in Fig. 2), and $N_2 = 1800$. Without a Gaussian approximation or coarse-graining of lengths, this model with $K = 0.1$ shows a chromosome-like organization and reproduces desired variances (see the graph on the left). With the choice $K = 1.0$, however, the calculated variances are a bit too small (see the graph on the right). The comparison between the cases $K = 0.1$ and $K = 1.0$ suggests that the former choice is more appropriate.

We favor the Gaussian approximation for single-locus variances and the coarse-grained inter-locus variances (via $z_{ij} \to \bar{z}_{ij}$) over others in that they compare more favorably with chromosome data[8] or those in Fig. 8(b). For the parameters chosen, there is a noticeable quantitative difference between the coarse-grained and bare inter-locus variances (see Fig. 8(a)). But the difference between single-locus variances appears to be minor.

What are the limitations of model (ii)? In this model, the chain is only sightly compressed; if relaxed in a corresponding open cylinder, the chain is about 150% of $L_{\rm out} = L_{\rm in}$. At first glance, this model appears to be somewhat artificial in the sense that the bacterial chromosome is in a much compressed state within the nucloeid. For instance, a recent experiment suggests that the *E. coli* chromosome should undergo about 10 fold-compaction by crowding effects so as to fit the cell[29] (note that the chromosome was already partially condensed by chromosome-bound proteins, which would have to be further compressed). We believe this is rather a seeming discrepancy. First, recall that each structural unit is a dynamic structure (Sec. 3) and remains rather constrained in the cell. Each monomer in this model represents this constrained unit, a feature our coarse-grained model does not take into account.

Despite its potential limitation, the main advantage of this model (ii) is that it produces expected locus distributions. Importantly, this validates our intuition: choosing bigger monomers is equivalent to subsuming more details (e.g., coarse-graining) into each monomer, as noted in Ref.[19].

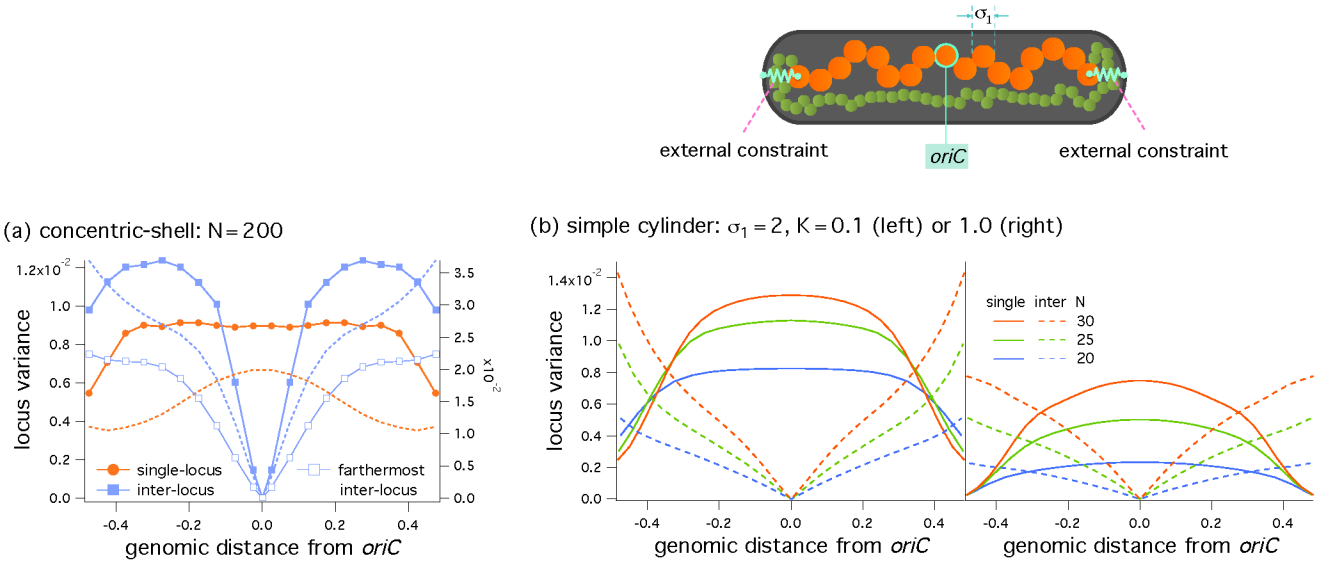
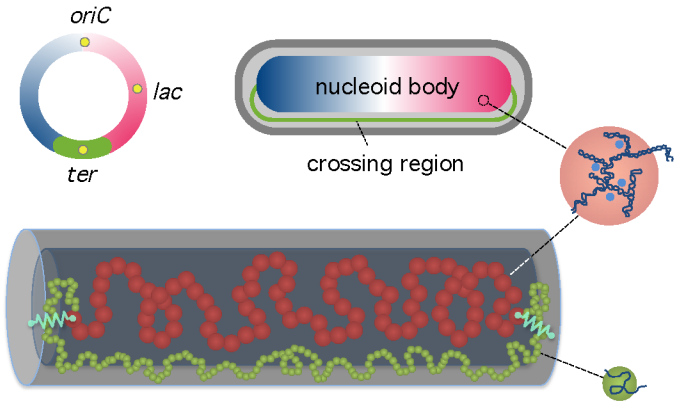
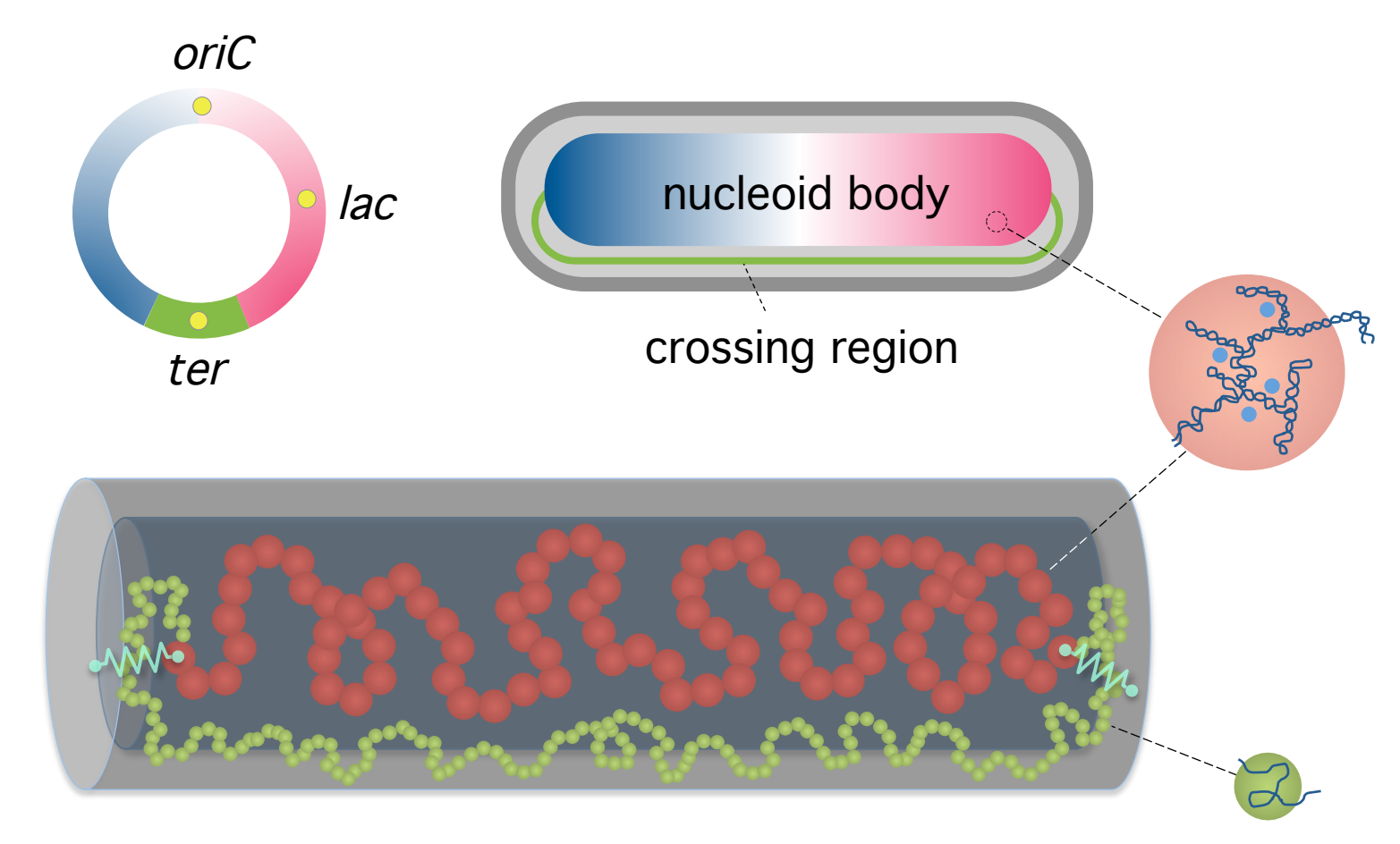


Fig. 8 Single- and inter-locus variances for the concentric-shell cylinder model and its variation: a constrained chain in a simple closed cylinder (model (ii)). Thanks to the cylinder attachment of chain segments, in model (ii), one can use smaller N values that are not suitable for model (i). (a) This graph compares single- and inter-locus variances, obtained with different methods ($N = 200$). If the solids lines with filled symbols are Gaussian approximations, the blue solid line with open symbols is based on the farthermost-distance distribution. The dashed lines obtained without any of these approximations are a few times larger. Among them, the Gaussian-fit single locus variance and the inter-locus from the farthermost distance distribution are more chromosome-like, when judged from Ref.[8]. (b) In model (ii), we double σ_1, compared to that in model (i), and adjust N so as to keep the volume fraction roughly unchanged; we have chosen $N = \{20, 25, 30\}$. As a result, the approximation used in (a) is not needed for a good quantitative agreement with the *E. coli* data[8], especially with the combination $N = 25$ and $K = 0.1$.

TOC Graphic



An asymmetric ring polymer in a concentric-shell cylinder shows chromosome-like spatial organization.


oriC
lac
ter
nucleoid body
crossing region