

Topology-mediated organization of *Escherichia coli* chromosome in fast-growth conditions

Shreerang Pande , Debarshi Mitra , and Apratim Chatterji 

Department of Physics, *IISER-Pune*, Pune 411008, India



(Received 26 April 2023; accepted 16 September 2024; published 1 November 2024)

The mechanism underlying the spatiotemporal chromosome organization in *Escherichia coli* cells remains an open question, though experiments have been able to visually see the evolving chromosome organization in fast- and slow-growing cells. We had proposed [D. Mitra *et al.*, *Soft Matter* **18**, 5615 (2022)] that the DNA ring polymer adopts a specific polymer topology as it goes through its cell cycle, which in turn self-organizes the chromosome by entropic forces during slow growth. The fast-growing *E. coli* cells have four (or more) copies of the replicating DNA, with overlapping rounds of replication going on simultaneously. This makes the spatial segregation and the subsequent organization of the multiple generations of DNA a complex task. Here, we establish that the same simple principles of entropic repulsion between polymer segments which provided an understanding of self-organization of DNA in slow-growth conditions also explains the organization of chromosomes in the much more complex scenario of fast-growth conditions. Repulsion between DNA-polymer segments through entropic mechanisms is harnessed by modifying polymer topology. The ring-polymer topology is modified by introducing crosslinks (emulating the effects of linker proteins) between specific segments. Our simulation reproduces the emergent evolution of the organization of chromosomes as seen *in vivo* in fluorescent *in situ* hybridization experiments. Furthermore, we reconcile the mechanism of longitudinal organization of the chromosomes arms in fast-growth conditions by a suitable adaptation of the model. Thus, polymer physics principles, previously used to understand chromosome organization in slow-growing *E. coli* cells also resolve DNA organization in more complex scenarios with multiple rounds of replication occurring in parallel.

DOI: [10.1103/PhysRevE.110.054401](https://doi.org/10.1103/PhysRevE.110.054401)

I. INTRODUCTION

It is vital for the living cell to make a copy of its DNA and segregate it into two halves of the cell, before the cell division can occur [1,2]. These essential processes have been extensively studied for one of the simplest single-celled organisms, the *Escherichia coli* (*E. coli*) bacteria. As the chromosomes replicate and segregate thereafter, the mechanism of spatiotemporal organization of the chromosomes remains controversial [3–9]. Unlike in higher organisms, the bacterial cell does not have dedicated protein machinery to transfer its two daughter chromosomes to two halves of the cell [10]. *E. coli* is a rod-shaped bacterium whose chromosome occupies the central region named the nucleoid. The bacterial cell does not have a nucleus. The segregation of the daughter chromosomes occurs simultaneously as replication is in progress [2]. In contrast, in eukaryotes the mitotic spindle helps segregate the daughter chromosomes after the replication is complete. Most bacterial cells have just one chromosome, and it is a ring polymer [11]. The chromosome of the bacteria *E. coli* and *Caulobacter crescentus* consist of a single ring polymer with 4.6 million and 4 million base-pairs (BPs), respectively [12–14].

In *E. coli* and other bacteria, replication begins at a site called *oriC* to end at the *dif*-locus of the *ter* macrodomain and proceeds along the two arms of the ring DNA polymer simultaneously [5,7,15]. Approximately 1000 BPs are replicated per second by the replisome at the replication forks (RFs) [16,17]. By controlling the growth medium, the doubling time τ of the *E. coli* bacterial cells can be varied to have

values from 20 min to 3 h or more [18–20]. The doubling time is time taken for one newly born cell to divide into two. The cell cycle typically consists of three periods. The “B period” refers to the time period between the birth of the cell and the start of replication. Once replication starts, the cell enters the “C period” and lasts till the next τ_C minutes, i.e., till the time it takes for the replication to be completed. Thereafter, the cell remains in the “D period” lasting τ_D minutes, i.e., till cell division occurs [18,21]. The bacterial cells are said to be in fast growth if $\tau < \tau_C + \tau_D$.

In fast-growing cells, the B period is absent, implying that the cells are continuously replicating and segregating. The conundrum of cells doubling every 20 min, even though the time taken for a chromosome to make a copy and divide is $\tau_C + \tau_D \approx 100$ min, was resolved by Helmstetter and Cooper and others [18,22,23] who showed that a second round of replication begins even before the first round is complete. Refer to Fig. 1 for a schematic of how the multiple rounds of replication proceed with overlapping cell cycles. Thus, the chromosome in fast-growth conditions undergoes multifork replication, with the replication process occurring simultaneously at two or more pairs of RFs [18,24].

It is accepted that for *E. coli* chromosomes, entropic forces between the ring polymers play a significant role in the segregation of daughter chromosomes [12,26–32], though proteins like MukBEF also play a critical role in the process [6,33]. Moreover, researchers have used fluorescent *in situ* hybridization (FISH) experiments to track the position of multiple DNA loci at different points in the cell cycle, i.e., while the replication and segregation of the bacterial chromosome is in

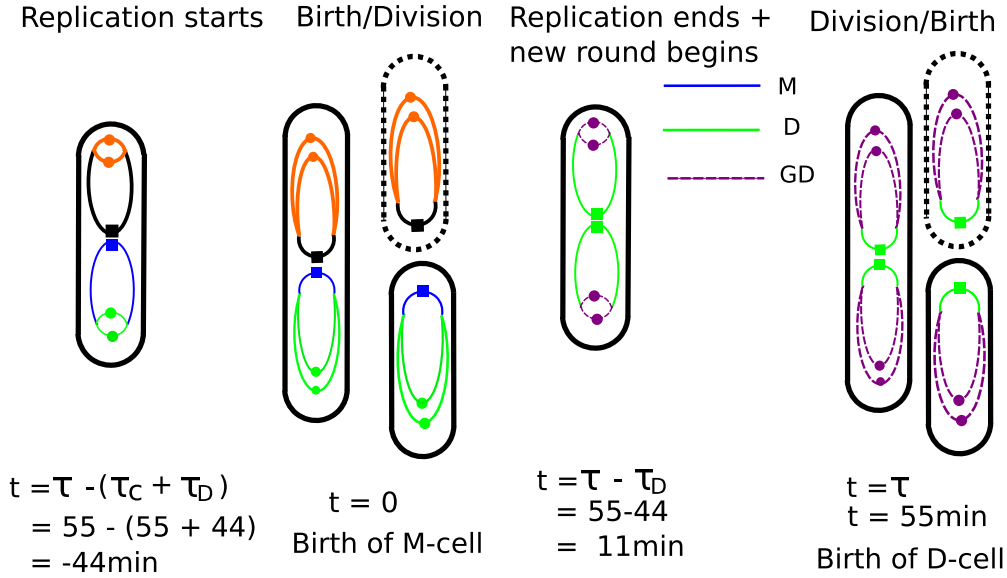


FIG. 1. Schematic of the cell cycle: Given specific growth conditions [24], the *E. coli* cells double every $\tau = 55$ min (min), the C-period is $\tau_C = 55$ min, and the D-period is $\tau_D = 44$ min. Since doubling time of $\tau = 55$ min is less than $\tau_C + \tau_D = 99$ min, we can infer that the cells are undergoing fast growth. In the schematic, cell division takes place at time $t = 0$, here the mother cell (M-cell) is born. After $\tau = 55$ min, i.e., at $t = 55$ min, another cell division takes place to form daughter cells (D-cell), as shown. However, for the pair of daughter-chromosomes [green (light gray)], which divide into the two D-cells at $t = 55$ min, their replication started $\tau_C + \tau_D = 99$ min earlier, i.e., at $t = -44$ min. We follow the “green (light-gray)” chromosomes from the start of replication. The *oriC* of blue (dark-gray thin line) chromosome in the grandmother cell (GM-cell) starts a new round of replication to form two green (light-gray) *oriC*s, and the replication forks proceed toward *dif-ter*. Meanwhile, the GM cell has divided to form two M-cells at $t = 0$, and we follow the cell that has the green (light-gray) chromosome. Since 44 min is 80% of the C-period, the mother cell is born with 80% of two daughter DNAs and 20% of mother DNA. There are two complete copies of the D-chromosome at $t = 11$ min. The other M-cell (shown in dashed outline) have the orange (light-gray thick line) (and black thick line) chromosomes, which are fully equivalent to that of blue (dark-gray thin line) [and green (light-gray)] chromosomes. But we color it differently to distinguish it from the green (light-gray) chain. From $t = 11$ to $t = 55$ min (D-period), we have two green (light-gray) *dif-ter*s connected to each other before cell division. As explained before, a round of replication starts 44 min before cell division, the purple (dashed) *oriC*s have formed from the existing green *oriC*s at $t = 11$ min, which is 44 min before cell division at $t = 55$ min. The D-cells are born with only 20% green (light-gray) D-chromosome and a pair of 80% formed purple (dashed) granddaughter chromosomes. In the legend, we show the colors of the M [blue (dark-gray)], D [green (light-gray)], GD [purple (dashed)] chromosomes. To visualize the different stages of the chromosomes in a cell cycle the reader is referred to the section titled “Movies” in the SI-1 [25] (videos “Vid-1” and “Vid-2”).

progress both in fast and slow-growth conditions [24,34,35]. For slow growth, it is observed that the *oriC* is initially found in the midcell position, and after about 20 min into the C-period, the two *oriC*s move to the quarter and three-quarter positions along the cell long axis [34,36]. The position of the *oriC*s is measured from one of the pole positions. In contrast, the *dif-ter* locus remains delocalized within the cell at the start of the C period but eventually moves to the midcell position before the end of the replication process. Other loci also move to their respective “home positions” as segregation proceeds [34]. The mechanism by which the different genomic loci identify their cellular addresses within the cell and then occupy the position at the appropriate stage of the cell cycle had remained an open question.

In our previous work in slow-growth conditions, we established that this DNA organization can be obtained by adopting a suitably modified polymer topology by having long range contacts on the chain contour, which are likely mediated by MukBEF or other linker molecules [36]. We refer to these contacts along the DNA polymer as crosslinks (CLs).

We refer the reader to Fig. 2 for a schematic of the modified polymer architectures we used for our study of chromosome organization in slow-growth conditions. We use the same

architectures for the current study of DNA organization in fast-growth conditions. We also showed that the sites of the CLs that we used in our DNA-polymer simulations, show high contact probabilities in the Hi-C map of *E. coli* chromosomes [36]. As a consequence of the introduction of CLs, internal loops of the polymer segments are formed within the ring polymers. Thus, the chromosome adopts a more complex polymer topology than a simple ring polymer. Though this choice of architecture might appear *ad hoc* to the reader at first, the architecture was arrived at using intuition developed by systematically studying the emergent segregation and loci (polymer-segment) localization properties of 12 different architectures [37].

A consequence of the the choice of introducing crosslinkers between points on the contour of the ring polymers is that we have smaller internal ring-polymer segments within the chromosome attached to each other. It is already well established by previous studies on semidilute systems and melts of ring polymers that ring polymers form compact configurations as compared to linear polymers [38–43]. Furthermore, rings have effective repulsive interactions between them and try to exclude each other in space [44,45]. Thus, the different loops of the Arc-2-2 architecture entropically repel each other and

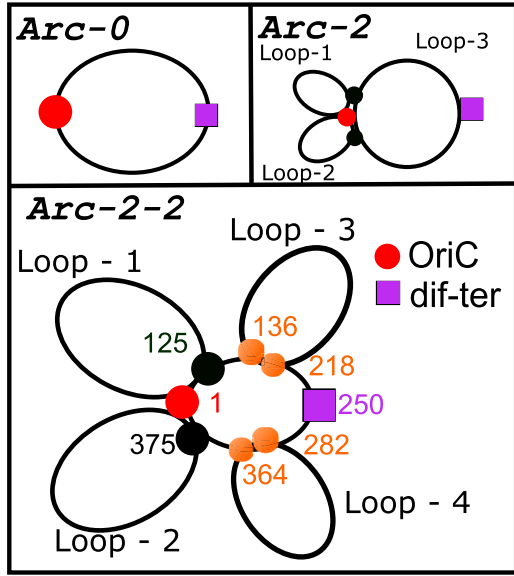


FIG. 2. Schematic of the different polymer architectures: The schematic shows the Arc-2-2 topology of the DNA polymer with 500 monomers. We start out with a ring polymer (Arc-0); thus, monomer 1 is joined to 500. We label monomer-1 as *oriC* and 250 as *dif-ter*. In addition, in our model, the monomer 125 and 375 is crosslinked to monomer 1 by harmonic springs modeling bridging proteins to create the Arc-2 architecture of the polymer. For the Arc-2-2 we have additionally crosslinked the monomers 136 and 218, as well as 282 and 364. Also refer to Fig. 3.

occupy different segments of the cell along the long axis, and thereby also localize different loci that are part of the loop as well as speed up the segregation process. An appreciation of the underlying physics of this phenomenon, viz., principles of the loci localization at different points along the long axis by choosing different sizes of internal loops can be achieved by referring to Ref. [37]. We give a brief review of this mechanism in this paper, before we present our results on modeling fast growth.

We chose two of the 12 architectures discussed in Ref. [37], which we named Arc-2 and Arc-2-2. We used their localization properties to quantitatively match our simulation results with the organization of loci as in FISH data for the chromosome of the bacterium for *E. coli* in slow-growth conditions. We also matched broad features of domain formation as seen in Hi-C maps of the *E. coli* chromosome with the calculated contact map from the simulations using the Arc-2-2 architecture and also discussed differences between the two [36]. Furthermore, using our systematic study of 12 other polymer-topologies, we showed that we can also match our model predictions with experimental Hi-C and FISH data for another bacterium, viz. *C. crescentus* [37], by suitable choice of a different polymer topology. Note that we always try to add only two to four CLs in our studies to minimally modify the topology.

In fast-growth conditions of *E. coli* with multifork replication in *E. coli*, there exists four (or more) chromosomes of different lengths at different stages of the replication process. This makes the segregation and faithful division of multiple strands of chromosomes a much more complex task. An

active machinery that could possibly direct the newly replicated chromosomes to move in opposite directions to get segregated by directed application of forces, might end up in daughter chromosomes remaining spatially overlapped in this complex lifecycle [24]. Refer to the schematic figure in SI-2 [25] for a more detailed discussion of this point. Therefore, an entropic model without any actively driven segregating machinery is a worthwhile avenue to pursue to decipher the mechanism of chromosome organization within a cell.

We establish in this article that the same simple model mechanism that was earlier developed to explain the organization of chromosomes and the replication fork in slow-growth conditions also explains the organization of chromosomes in the much more complex fast-growth conditions. We introduce appropriate modifications to our previous model to incorporate overlapping rounds of replication. We aim to obtain the spontaneous segregation of multiple DNA strands as the different rounds of replication(s) proceed. As an emergent phenomenon, the loci of partially replicated chromosomes position themselves at different sections of the long axis.

The organization of genomic loci and that of replication forks self-emerge in our model simulation as a consequence of modified topology of bead spring model ring polymer(s), which represents the bacterial chromosome. The organization we obtain is similar to that seen *in vivo*. We further show that the organization of the replication forks is a direct consequence of the entropic repulsion between the different loops of the chromosome-polymer. In the last section of the article, we suggest a plausible topology based mechanism to try to understand a long-standing problem pertaining to the organization of the chromosomal arms with respect to each other. As indicated by the data [24,35], in fast-growth conditions the chromosome arms are arranged in a longitudinal (doughnut-like) fashion. To this end, we further modify Arc-2-2 topology, and establish here that the doughnut-like organization can be obtained by introducing smaller loops along the arms of the chromosome-polymer. The introduction of smaller loops along the chain contour leads to entropic segregation of the two arms along the radial axis of the cylinder.

We emphasize that we introduce a minimal number of topological modifications in a coarse-grained minimal bead-spring polymer to obtain emergent organization of loci and the segregation of chromosomes in fast-growth conditions. These modifications are likely consequences of active processes within the cell. Thus, we incorporate the internal loops formed by linker proteins (e.g., MukBEF or other protein complexes) in our polymer model, and investigate the consequences of these loops without incorporating the details of the mechanism of formation of the same. For our model, we assume that the loops of Arc-2-2 are permanent, i.e., *effectively* long-lived *in vivo*.

We do not explicitly incorporate the effects of many other relevant nonequilibrium biological phenomena, which might contribute to the smooth functioning of the cell. Thus, we neglect the presence of organelles in the cells, which are often modeled as crowders [46], and the details of the replication bubble. We do not incorporate the effects of supercoiling [47] in our model, as models of supercoiling and its consequences are still under investigation in simpler idealized scenarios [48]. We also do not explicitly model the

process of extrusion, which creates loops of $\sim 100 - 200$ kbp. It has been shown by simulations that internal transient loops formed by extrusion indeed speed up the segregation process of DNA rings in a cylinder [49], consistent with our previous results [37] with permanent CLs. In our minimal model, we do not expect to be able to capture every aspect of the chromosome organization and dynamics. While we have identified some of the mechanisms that drive *E. coli* chromosome organization at the 100 nm to μ length scales in fast growth, it is imperative to acknowledge the inherent limitations in our ability to comprehensively capture all the intricacies of experimental observations related to chromosome organization. In our future investigations, we intend to add some of these effects to our current model.

We now describe the outline of the rest of the paper. In Sec. II, we describe the model we used to study the system. In Sec. III, we briefly review our previous understandings to introduce the basic principles of entropic repulsion between internal loops to the readers of the paper. This is to facilitate our readers to appreciate our new results on loci-localization in fast-growth conditions presented in this paper, without referring to our previous papers [37]. In Sec. IV we describe in detail the spatiotemporal organization of the different loci along the long axis in fast growth using our simulations. We also point out when we get differences with experimental observations. We not only show results for the different loci but the localization of replication forks as well. We show that the organization of the replication forks arises as a consequence of modifying the topology of the polymer. After discussing the radial organization of the loci, we summarize our results in Sec. V.

II. MODEL

We use Monte Carlo (MC) simulations of the bead-spring model of a polymer with 500 monomers to model a single *E. coli* chromosome (with 4.6 million base pairs) within a cylinder, which represents the *E. coli* cell. Thus, each coarse-grained monomer subsumes 9.2 kbp of DNA.

The equilibrium distance between two neighboring monomers (beads) along the chain contour is a , and they interact *via* the harmonic spring potential with energy $V_H = \kappa(r - a)^2$, where r is the distance between the adjacent monomers. The unit of length in our study is a . The spring constant κ is $100 k_B T / a^2$. The excluded volume (EV) interactions between monomers are modeled by the Weeks-Chandler-Anderson (WCA) potential [50] and the diameter of each monomer is given by $\sigma = 0.8a$, unless specified otherwise. This particular choice of parameters, allows us to keep the mean extension of individual springs due to thermal effects to 7% and prevent chain crossing while allowing for reasonable acceptance rates of monomer displacements in the MC simulations. These parameters are appropriate to model a normal inorganic polymer where chains do not cross each other.

The form of the WCA potential is

$$V_{\text{WCA}} = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6] + \epsilon_0, \forall r < 2^{1/6}\sigma. \quad (1)$$

The quantity ϵ_0 is needed to ensure that the potential goes smoothly to zero at the cutoff and $V_{\text{WCA}} = 0$ for values of r greater than the cutoff.

We model the chromosome replication and segregation over one doubling time τ inside an elongating cylinder, representing the growing *E. coli* cell. To model replication at the RFs, we add monomers (DNA segments) at regular intervals to the chain which represents the daughter DNA-2. A schematic figure is given in SI-3 [25]. After the RF moves to the adjacent site (monomer) to make a copy of the DNA segment of the mother (which leads to addition of a new monomer to DNA-2 in our model), we rename the monomer of the mother DNA such that it now notionally belongs to daughter DNA-1. Thus, the lengths of daughter DNA-1 and DNA-2 keep increasing as the RFs move toward *ter* along the two arms of the mother DNA. The simulation starts from the state right after cell division, equivalent to the state shown at time $t = 0$ in Fig. 1. At that stage, two new cells are just born from their parent cell. We follow the replication of chromosomes in just one cell, and the simulation ends just before the cell is ready to divide into two daughter cells, i.e., the stage shown at $t = 55$ min in Fig. 1. The newly born mother cell (M-cell) at $t = 0$ has 80% partially replicated mother M-chromosome, i.e., there are two daughter D-chromosomes, schematically marked in green with 400 monomers each, and 20% mother chromosome marked in blue with 100 monomers; refer to Fig. 1. Thereby, there are two *oriCs* at the start of simulations. The way the system is initialized is described in SI-4 [25].

MC simulations is used to update the position of the monomers, where one Monte Carlo step (MCS) consists of N attempts to update the position of the N monomers, chosen at random. Since we model replication and thereby add monomers at the RFs at regular intervals, N keeps increasing as the simulation proceeds. To update monomer positions, a trial move is made to displace the monomer in a random direction, and the move is accepted or rejected using the Metropolis criterion. The polymers explore different microstates as the monomers undergo local diffusive motion, as the simulation proceeds.

The successive Monte Carlo steps cannot and should not be interpreted as the time evolution, and we do not mention the elapsed time when we discuss DNA organization at different stages of the cell cycle. In a Monte Carlo step, the DNA polymer reaches a different microstate without following the detailed kinetic pathways, and the probability of reaching the new microstates assumes conditions of equilibrium statistical mechanics. But our simulations break detailed balance in multiple ways, which we discuss in some detail at a later stage in this paper. Hence, when we present our simulation results, we avoid the mention of time in units of minutes and instead refer to the stage of cell cycle measured in terms of the progress of replication and position of the RF on the polymer contour. However, we do provide a discussion of time scales at the end of this section.

During the course of our simulation, the two RFs reach the *dif-ter* loci such that the mother is completely replicated to form two (green) D-chromosomes and the cell enters the D-period, refer to Fig. 1. Simultaneously, a new round of replication starts such that each of the two (green) D-*oriCs* divide into two (orange) GD-*oriCs*. The simultaneous start of the second round and the end of the first round is a consequence of the values of τ , τ_C and τ_D in experiments of Ref. [24],

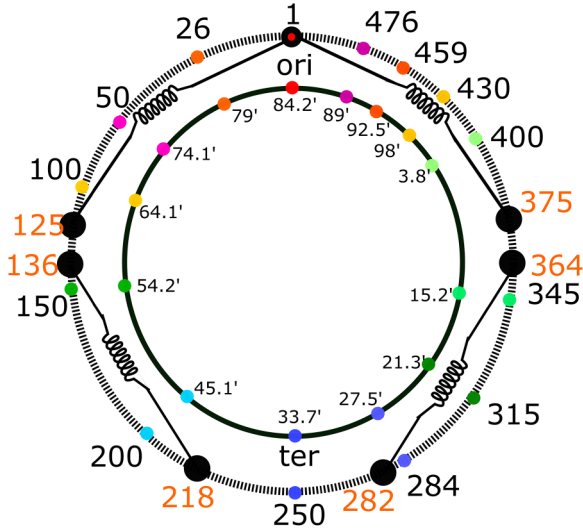


FIG. 3. Schematic of the tagged loci: The figure shows a schematic of chromosome loci tagged in experiments along the chain contour and the corresponding monomer indices for a 500 monomer chain by colored-filled small circles. Experimentally, the circular chromosome is tagged at different sections, where different loci along the chain contour are denoted in terms of minutes and seconds. The inner circle in the schematic corresponds to the loci fluorescently labeled in the experiment [24]. The outer circle denotes the monomer indices corresponding to these labels in our model system of ring polymer, where the *oriC* is denoted by monomer index 1, and *dif-ter* by monomer index 250. We have also shown the monomers which are crosslinked by springs of equilibrium length a to create the Arc-2-2 architecture, as shown in Fig. 2.

which we choose to model in this paper. For modeling replication, we add monomers at a fixed rate of 1 monomer every $f_{\text{rep}} = 2 \times 10^5$ MCS at each RF. We keep f_{rep} identical to that used in our study for slow-growth conditions [36].

As the cell cycle proceeds, the cylinder length doubles in small steps over the course of the simulation while the diameter remains fixed as observed for *E. coli* cells *in vivo*. We increase the length of the cylinder every f_{rep} MCS. The polymer is confined within a cylinder of diameter $7a$ ($\equiv 1 \mu\text{m}$, the typical diameter of the cell), and the cylinder length doubles from $21a$ ($\equiv 3 \mu\text{m}$) to $42a$ as our simulation proceeds. We consider the walls of the cylinder to be hard (with infinite potential) such that we reject any Monte Carlo trial move in which a monomer-center attempts to occupy a position located outside the cylinder. The dimensions of the confining cylinder correspond to the volume accessible to the center of the monomers.

We modify the ring-polymer architecture by introducing chromosomal loops by bridging specific loci along the chain contour of daughter DNAs after the RFs cross the corresponding loci of the mother DNA, refer to Figs. 2 and 3. These loops are created in our simulations by introducing additional springs that crosslink between specific pairs of monomers along the chromosome contour [36], using insights from Ref. [37]. Figure 3 also shows the position of tagged loci in fast-growth experiments of Ref. [24], and the corresponding monomer indices in our polymer model with 500 monomers. We have explicitly checked that a change in the choice of

monomer to be crosslinked by ≈ 5 monomers along the chain of 500 monomers will not change the global localization pattern of polymer segments significantly; refer to SI-5 [25]. Larger changes will affect the size of loops and, thereby, the relative strength of entropic interactions and hence modify the localization and organization patterns of polymer segments along the cylinder long axis.

Moreover, to mimic the role played by topoisomerase within the living cell, we allow topological constraint release (TCR) at regular intervals at rates we used previously in Ref. [36,37], i.e., every $f_{\text{TCR}} = 10^4$ MCS. We reduce the excluded volume interaction by changing the σ to $0.1a$, for the next 900 MCS. This allows the chains to cross through each other. We do not model cell division. We track the position of all the (available) *oriC*s and other monomers as the simulation proceeds. As we have previously shown in Ref. [37], topological constraint release is crucial for successful segregation to occur. If we switch off the step of TCR by chain crossing, the success rates of segregation for a pair of topologically modified polymers is low. If we decrease f_{TCR} by a factor of 10, i.e., $f_{\text{TCR}} = 1000$ (say) and thereby allow chains to cross more frequently, then we also reduce the excluded volume interaction and therefore the entropic effects. However, once segregated the localization of the loops and the loci are not affected by TCR as we show later.

Though we use Monte Carlo simulations to investigate chromosome organization as the cell goes through its life-cycle, the simulation is quintessentially a *nonequilibrium* simulation scheme. In the simulations, we (a) add effects of polymer chains crossing each other to release topological constraints, (b) add monomers to the simulation box at regular intervals at different points along the contour, i.e., at the position of the RFs to mimic replication and formation of two chains from one, (c) add crosslinks at certain stages of the simulation, and last (d) increase the length of the cylinder as the simulation proceeds. These are energy-consuming nonequilibrium active processes inside the cell, and these steps break detailed balance in the simulation. MC is used primarily to model the diffusion of monomers and explore different conformations of polymers in a confined space, assuming local equilibrium [51,52]. As we demonstrate in SI-6 [25], a polymer ring with 120 monomers relaxes in $\approx 5 \times 10^4$ MCS, whereas, we add new monomers and increase the length of the box every $f_{\text{rep}} = 2 \times 10^5$ MCS. Just as a reference, Loop-1 and Loop-2 have 125 monomers each. A single monomer of the 120 monomer polymer chain takes ≈ 400 MCS to diffuse its own diameter in confinement.

As mentioned before, we avoid mapping MCS directly to time in terms of minutes but rather mention the progress of the cell cycle in terms of the replication stage. Converting MCS into real time units by comparing diffusion rates need not be accurate also because chains can freely cross each other during the TCR step in our model, but we do not have an estimate of how many chain crossing events actually occur per minute. Incorporating the detailed process of releasing topological constraints by topo-isomerase *in vivo* is out of scope in our current coarse-grained model. This also involves the unavailability of the time scales of the release of topological constraints *in vivo*. However, we provide an estimate of the number of MCS for the completion of replication and

segregation in units of the number of MCS taken for a monomer to diffuse its own diameter to check if the ratio is compatible with experimental estimates. However, it must be remembered that a chromosome segment undergoes subdiffusive behavior *in vivo* [53,54]. While estimating the number of MCS taken for a monomer to diffuse its own size, we do not consider the presence of organelles (crowders). For the comparison, we use $a = 150$ nm, which is consistent with our choice of the cylinder diameter $D = 7a = 1050$ nm. This is shown in SI-7 [25].

III. REVIEW: MECHANISM OF ENTROPIC LOCALIZATION

We explain the mechanism of the emergence of localization of loci as a consequence of the underlying modified internal topology of the ring polymer. We show that entropic repulsion between the internal loops of the modified architectures lead to the mutual self-avoidance of the loops, and thereby loops occupy different sections of the cylinder along the long axis. The monomers belonging to different loops get localized as a consequence.

A systematic and detailed analysis of the mechanism of entropic repulsion between internal loops and their consequence on segregation and loci localization of 12 different architectures was studied in Ref. [37]. These studies started with two overlapping polymers with modified topologies, and then as the simulation proceeds the polymers segregate to two halves of the cylinder. Thereafter, we get localization of different internal loops of each polymer within the half-cylinder that they occupy. We recreate a similar simulation and analysis in this section to clarify the role of loops and their organization.

For the results presented in this section, we do not (a) add monomers as the simulation proceeds, (b) change the length of the cylinder, and (c) introduce additional CLs during the course of the simulation. Moreover, we present data (i) without topological constraint release as well as (ii) with topological constraint release at rates as used in the next sections. We hope this simpler case study will help the reader appreciate the more complex case of loci-localization with multiple rounds of replication in parallel, as happens for chromosome replication in fast-growth conditions.

A. Architecture-2: Arc-2

Two polymers, each with 500 monomers, and with modified architecture Arc-2 are confined in a cylinder of diameter $7a$ and fixed length $42a$, consistent with the dimensions used for the fast-growth model. Our previous slow-growth simulations were performed in a cylinder of length $35a$ [36]. We start our simulations with segregated conformations which were the last configuration of other runs, where we had two complete Arc-2 polymers under confinement. Thereafter, we ran the simulations for a further $n_I = 4 \times 10^7$ MCS, before we started collecting data over the next $n_P = 2 \times 10^7$ MCS. This data was used to calculate positional distributions of the COMs and other statistical quantities, which we present in this section. The 2 polymers are connected at monomer 250, which represents the *dif-ter* link that is present in the cell. From a single production run of $n_P = 2 \times 10^7$ MCS, 666 snapshots

(microstates) were collected to calculate statistical averages. We perform 50 independent Monte Carlo simulation runs, each with $n_I + n_P = 6 \times 10^7$ MCS.

We refer to the two smaller loops as Loop-1 (with monomers 1–125) and Loop-2 (with monomers 375–500). We label the bigger loop as Loop-3 (125–375); also refer to Fig. 2. To identify the position of the loops along the cylinder's long axis (z axis in our simulations), we plot the probability distribution of the position of the center of mass (COM). In Fig. 4(a), we observe that the peaks of the probability distribution $p(z_{\text{COM}})$ of the COM of loops belonging to different polymers are well separated along the z axis. Moreover, the two polymers, referred to as Poly 1 and Poly 2, are spatially segregated to two different halves of the cylinder. We also observe that the COM of the loops of individual polymers occupy different sections within each half of the cylinder. We conclude that the two polymers are arranged in the fashion shown in the schematic Fig. 4(b). Probability distribution $p(z_{\text{oriC}})$ data for the particular monomer 1, which represents *oriC* in our simulations, is presented in Fig. 4(c). As *oriC* is the junction of the loops it gets localized to the quarter positions, with distinct peaks in $p(z_{\text{oriC}})$. In Fig. 4(c) we show *oriC* localization data when we allow topological constraint release. We also verified that the position of the peaks in $p(z_{\text{oriC}})$ remain unchanged even we do not have topological constraint release, refer to Fig. 4(d). Thus, we establish that the localization property of loops is also relevant for synthetic polymers, which do not have chain-crossing.

B. Architecture 2-2: Arc-2-2

We next present results of organization of loops in cylindrical confinement for a further modified topology, the Arc-2-2 polymer with two additional loops near monomer 250 (the *ter*-region). For the Arc-2-2 polymer, we name the two slightly bigger loops as Loop-1 (1–125) and Loop-2 (375–500) similarly the two smaller loops are named Loop-3 (136–218) and Loop-4 (282–364), refer to Fig. 2. The remaining monomers are clubbed into another loop named Loop-5. In Fig. 5(a), we plot the probability distribution of the position of the COMs of the 5 loops along the long axis of the cylinder. We observe from Fig. 5(a) that the two smaller loops and the two bigger loops occupy different sections of the cylinder (cell) as each pair of loops repel the other pair entropically. Moreover, the monomers of the central loop (in black) occupy the central region of the cylinder. Figure 5(b) schematically shows how the monomers of the different loops could be arranged. As a consequence of this arrangement, the *oriC* again gets localized in the quarter positions of the cylinder in Figs. 5(c) and 5(d). Another interesting thing to note is that the two smaller loops Loop-3 and Loop-4 of both the polymers show two peaks within each half of the cylinder. This will be instrumental in understanding how the replication forks also organize in the fast-growth simulation in the results section later. We have also plotted the distribution of other loci described in Fig. 3 in SI-8 [25]. In the SI we have explained in detail how the positional distribution of each loci can be explained within the framework discussed above.

In our previous study of DNA polymer in slow-growth conditions [36], DNA polymers were replicated using the

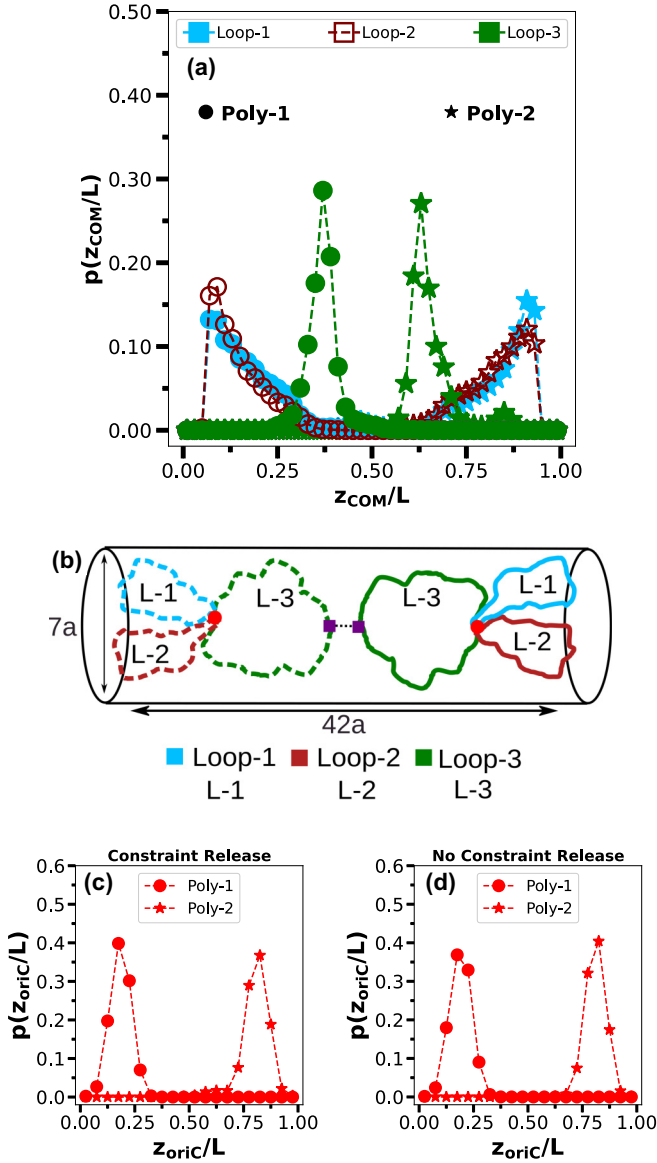


FIG. 4. Panel (a) shows the probability distribution $p(z_{\text{COM}}/L)$ of the center of mass (COM) of the different loops of a pair of Arc-2-2 polymers. The position z along the long axis is normalized by the cylinder length $L = 42a$. Loop-1 and Loop-2 of each polymer occupy the ends of the cylinder, whereas the COM of Loop-3 occupies distinct regions around the center without overlap. The monomers of different loops do partially overlap, and a detailed discussion on the degree of overlap and its relation to different architectures can be found in Ref. [37]. Panel (b) schematically shows the arrangement of the loops as indicated by the previous subfigure. We see that the *oriC* is the junction of the loops. In panel (c), we see that the *oriC* loci(monomer-1) are also localized close to the quarter positions, as it is at the junction between Loop-1 and Loop-2 and the Loop-3. Panel (d) shows *oriC* localization in the absence of topological constraint release (TCR).

same scheme as presented in the model section of this paper, and daughter DNA adopted this architecture as copies of the monomer which get crosslinked are created. As a consequence, different loci got localized at different stages of replication in our simulation, which corresponds to different stages of the lifecycle of the cell.

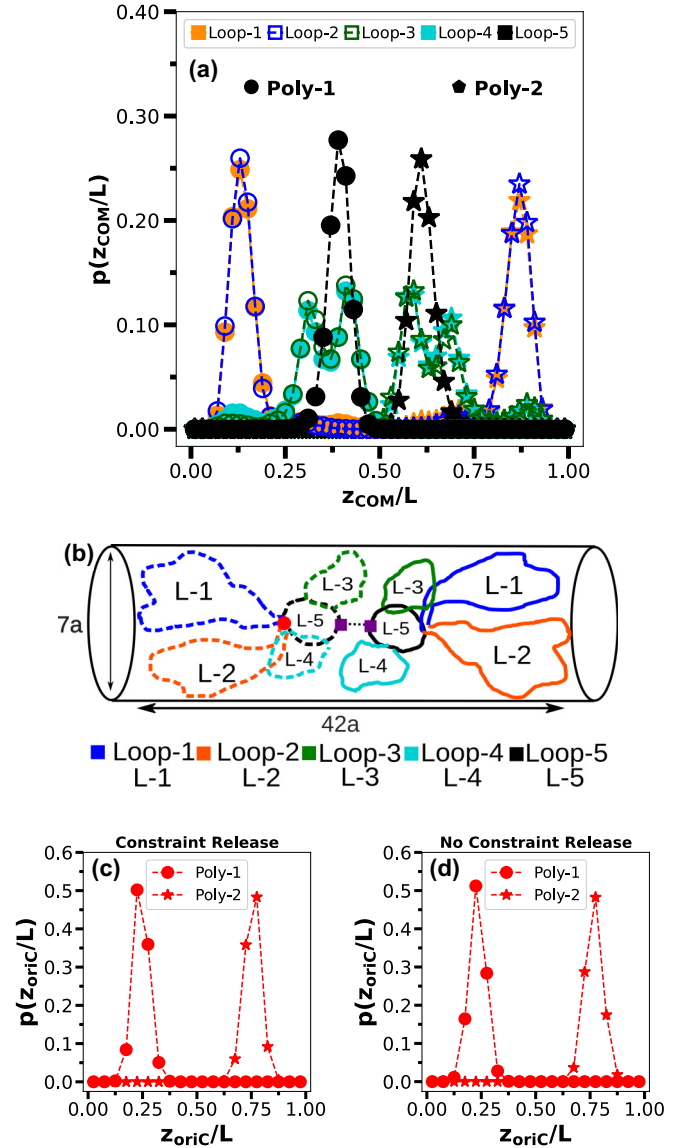


FIG. 5. Panel (a) shows the positional distribution $p(z_{\text{COM}}/L)$ of the COMs of different internal loops for a pair of Arc-2-2 polymers along the long-axis z . Loops-1, -2, -3, and -4 are defined in Fig. 2, and different colors represent $p(z_{\text{COM}}/L)$ data for COMs of different loops. We label the central loop as Loop-5, and its distribution of COM-positions is shown in black. Different loops occupy different sections of the cylinder as a consequence of entropic repulsion between loops. The Loops-3 and -4 interchange positions and remain partially overlapped with loop-5, which shows up the relatively large fluctuations in the peak. Loop-5 contains the *ter*-monomer(s). Thereby Loop-5 is positioned closer to the center of the cylinder due to the presence of *ter*-crosslink. Panel (b) shows a schematic of how the loops are likely arranged in the cylinder. Panels (c) and (d) shows the localisation data of the *oriC* when we have TCR and when we switch it off, respectively.

C. Other consequences: Predictions from Arc-2-2 topology model of the *E. coli* chromosome

A particular chosen topology (Arc-2-2) not only produced localization of loci as seen as slow-growth experiments, but also showed the emergence of domains that are

similar to macrodomains as seen in Hi-C data for *E. coli* [55]. Furthermore, we could also resolve other long-standing controversies. For example, our simulations agree with the experimental observation that *oriC*s start moving toward their quarter positions at the stage where half the mother DNA has been replicated. This corresponds to a time of roughly 20–25 min after replication starts. Simultaneously the *ter* loci moves to the midcell position: the terminus transition. Thus, we now understand the mechanism and timing of the *ter*-transition.

In addition, our studies using the Arc-2-2 architecture also illustrated the mechanism by which the RFs remain localized near the center of the cell for most of the replication cycle, despite using the train-track model of replication. Thus, our model explains the apparent disagreement between the train-track and replication factory model for the position of the RFs in the cell. As a consequence of the localization of the replication forks at the center of the cell, we now understand why the loci split always occurs at the cell center for all loci investigated [34]. Last, and most important, the particular *position* of these loops along the chain contour helps the polymer significantly increase the rate of segregation of the two replicated chromosomes to two different halves of the cell. The importance of internal loops to increase the speed of segregation has been independently validated by Molecular dynamics simulations of DNA segregation where extrusion is explicitly incorporated in the modeling [49]. These (unexpected) predictions and consistencies with other experimental observations, gives us the confidence of adopted particular topology and our modeling. All these are discussed in detail in Ref. [36,37].

As we will see in the later sections of this paper, the same adopted topology provides the mechanism of organization of loci for even fast-growth experiments, though we did not design the topology keeping the fast-growth loci-localization data in consideration. We do not currently consider the different mechanisms of loop formation, e.g., extrusion.

IV. RESULTS

Following the principles established in the previous section, we now explain in detail how the modified polymer-architecture determines the chromosome organization in fast growth. We will also plot our simulation data in a manner similar to how the experimental data was presented in Ref. [24]. But FISH experiments have limited resolution, and hence, for certain cases one has difficulty in resolving the replicated loci before they are well separated. In experiments, it is not possible to identify which foci belongs to which generation, the daughter (D) or the granddaughter (GD) chromosome. These and other issues (discussed later in this paper) has the consequences that the plots of the experimental data can be difficult to interpret for a reader.

In simulations, we can identify the number of existing loci and their positions along the long axis at any point during the simulation run without ambiguity. Hence, we present the data in a way that makes it easy to appreciate the underlying mechanism of loci localization. Thereafter, we plot our data in the way similar to how experimental data has been presented, incorporating the consequences of the inability to distinguish two foci, which are spatially close to each other. This enables

a direct comparison of modeling data in Fig. 6 (and other figures) with the experimental data in Fig. 7, which we reproduce from [24] in this paper. Our simulation run models one cell cycle, and corresponds to the time $t = 0$ to $t = \tau$ of Fig. 1. For the data presented below, we have multiple rounds of replication as detailed in the model section.

Initialization. As mentioned earlier, we start the simulations with the birth of a new M-cell corresponding to $t = 0$ of Fig. 1. Before we start our simulation, we allow the two D-chromosomes to relax over 3×10^7 MCS, where they can take different independent conformations. In this initialization process, the two D-chromosomes remain connected at the two RFs on the M-chromosome assuming completion of 80% replication of both arms. The RFs are located at monomer indices 200 and 300, equidistant from the *dif-ter* loci at the 250th monomer. Refer to figure in SI-9 [25] for a schematic to follow the initial configuration. *In vivo*, up to just before cell-division and birth of the M-cell, the two replicated *dif-ter* loci remain linked to each other in the parent cell, and are localized in the middle of the parent cell. Thus, in a newly born cell one expects the *ter*-loci to be near the new pole, which is formed in the middle of the parent (M-cell). To emulate this, we keep *dif-ter* (the 250th monomer) tethered to one of the poles of the cylinder (corresponding to the new pole of the cell) during initialization, and released the tether just at the start of the simulation. We allow chain crossing by TCR. Thus, at the beginning of the simulation, we have a polymer architecture that has the CLs that create Loop-1 and Loop-2 in each of the D-chromosomes, which we name D1 and D2. The mother M-chromosome has the CLs, which results in Loop-3 and Loop-4; refer to Fig. 2 and SI-4 [25] for other details of the initialization process.

Segregation and chromosome organization. We present a snapshot from our simulations in SI-10 [25] to show that Arc-2-2 DNA polymers from different rounds of replication remain well segregated along the cylinder long axis at the end of the cell cycle. The snapshot corresponds to a configuration at the end of the simulation run when 80% of replication of each of the D-chromosomes to 2 GD-chromosome is complete, and the cylinder has grown from $21a$ to $42a$. For comparison, we show a snapshot from the end of the simulation run with Arc-0 polymers, i.e., polymers with the unmodified ring-polymer topology, and we see that DNA polymers from different rounds of replication remain relatively more mixed as compared to the previous case.

Thereafter, we follow the positions of the same monomers that correspond to loci tagged in Ref. [24] as simulation proceeds, also refer to Fig. 3. As the simulation proceeds, the RF moves along the two arms of the chromosome. The position of the RF and amount of replicated DNA in the simulations is used to mark the evolution of the cell cycle and determine the different stages of the lifecycle of the cell, even as multiple rounds of replication are in progress. We quantify loci organization by plotting their spatial distribution in *five equally divided intervals* of their lifecycle as in experiments, refer to Fig. 6. We average our data over 50 independent runs corresponding to a lifecycle in 50 cells.

dif-ter localization. Soon after the simulation starts, the *dif-ter* monomer (i.e., the 250th monomer) will be near the end of the cylinder where it had been tethered during initialization.

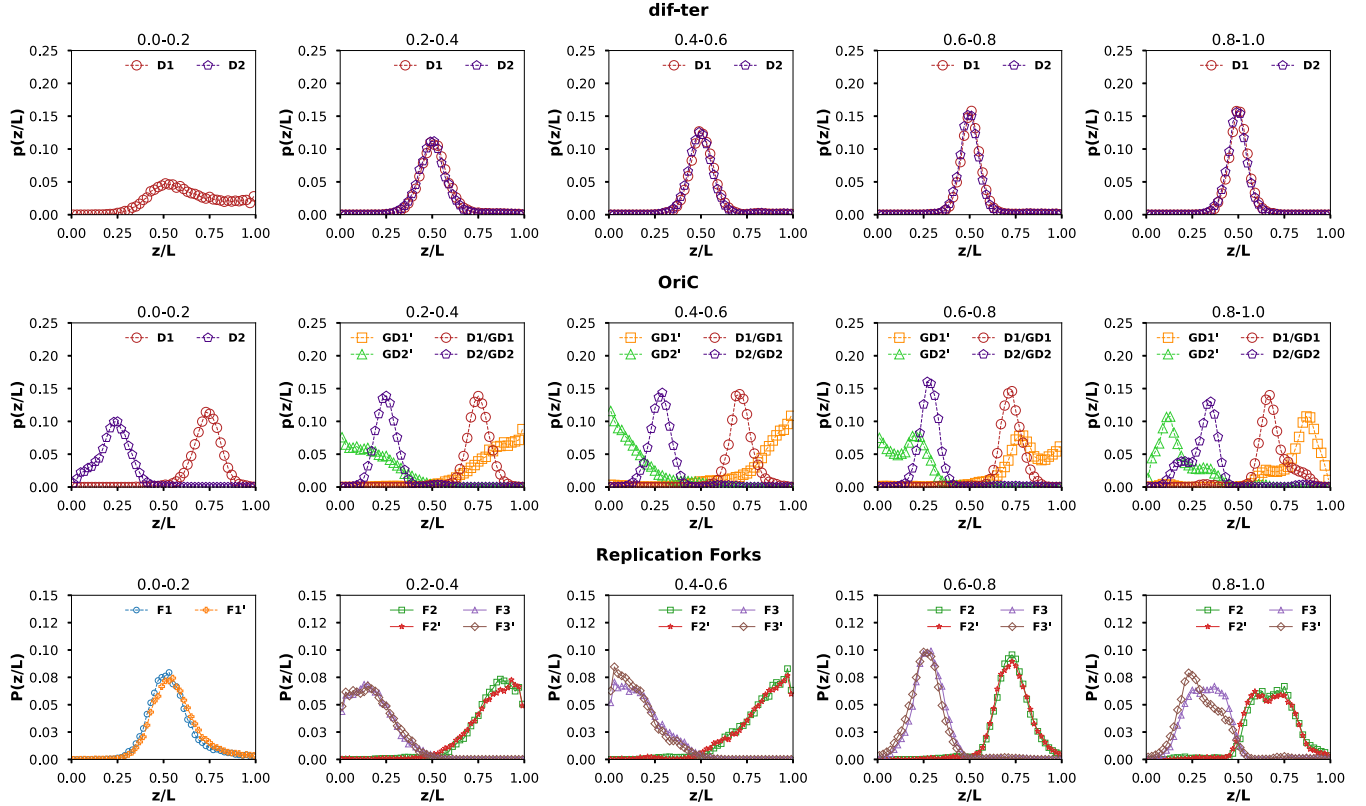


FIG. 6. Long-axis distributions for *dif-ter*, *oriC*, and the replication forks: We plot the spatial probability distributions, $p(z/L)$ of the position of different loci, where z denotes the along the long axis of the cylinder (cell), and L is the length of the cylinder at that stage of the simulation run. Data is shown for *dif-ter* locus (first row), *oriC* locus (second row) and the RFs (third row) for various intervals, as indicated at the top of each subfigure, during the lifecycle. D1 chromosome on getting replicated creates GD1 and GD1' chromosomes. Similarly, one obtains GD2 and GD2' from the replication of D2 chromosome. Once a particular locus gets replicated, then the corresponding D monomer is renamed as the GD monomer (e.g., D1 is renamed to GD1). The localization of the *oriC*s and the *dif-ter*s can be visualized from our representative simulation video “Vid-3” (refer to SI-1 [25]). In the third row we plot the spatial distributions of the RFs as they are assumed to move from one monomer to the next, along the chain contour. In the (0–0.2) interval there only two RFs which move along the contour of the M-chromosome [shown in orange (plus marker) and blue (circle marker)]. Thereafter, one has four RFs branching out from the two *oriC*s of the D-chromosomes, which get replicated to form the GD chromosomes. Note that GD1 and GD1' chromosomes occupy one-half of the cell while GD2 and GD2' occupy the other half. The RFs have been named in a specific way to clearly demarcate those which are traversing along different arms of the same chromosome. For instance, F1 and F1' are the two replication forks moving along the two arms of the mother chromosome, while F2 and F2' denote the RFs moving along D1 and corresponding F3 and F3' moving along D2.

Thus, the probability distribution shows nonzero values near one of the ends of the cylinder in the first stage (0–0.2) interval of its lifecycle; refer to top row of Fig. 6. But even as the RF-(s) move from monomer 200 (and 300) toward the *dif-ter* loci and cross the 218th monomer (and 282th monomer in the other arm), we introduce CLs between the newly introduced 218th monomer and the previously replicated 136th monomer to form a new Loop-3. Correspondingly, we introduce CLs between newly replicated 282nd and previously replicated 364th monomers to create a new Loop-4 for the D2-chromosomes. Refer to SI-11 [25] for the schematic diagram to better understand how we implement the crosslinking in our simulations.

The entropic repulsion between the loops of the two D-chromosomes ensures the segregation of the two polymers into two halves of the cylinder. This also relocates the *dif-ter* loci to the middle of the elongating cylinder. As a consequence, we also observe a peak for the probability distribution of *dif-ter* at the middle of the cylinder at all stages

of the life-cycle, refer to the first row of Fig. 6. As the *dif-ter* monomer is replicated, the two copies are connected by a spring, as the *dif-ter*s are connected *in vivo* up to cell-division. We call this the ter-CL. The peak remains unchanged in subsequent intervals as the ter-CL maintains the position of the two *dif-ter*s. For reference, we have provided experimental plots from [24] in Fig. 7. In the (0.8–1) interval in Fig. 6, we do not see two peaks in the *dif-ter* distribution as we do not model the breaking of the ter-CL and consequent separation of *dif-ter* loci before cell division; in contrast to the two-foci *dif-ter* data in Fig. 7.

oriC localization in initial stages of lifecycle. As the two D-chromosomes occupy two different halves of the growing cylinder, the *oriC*-monomer moves to the quarter positions of the cylinder, in the (0–0.2) and (0.2–0.4) intervals of the cell lifecycle, as seen *in vivo*. This is due to the repulsion between internal loops within each D-chromosome as outlined in the previous section. This organization emerges spontaneously

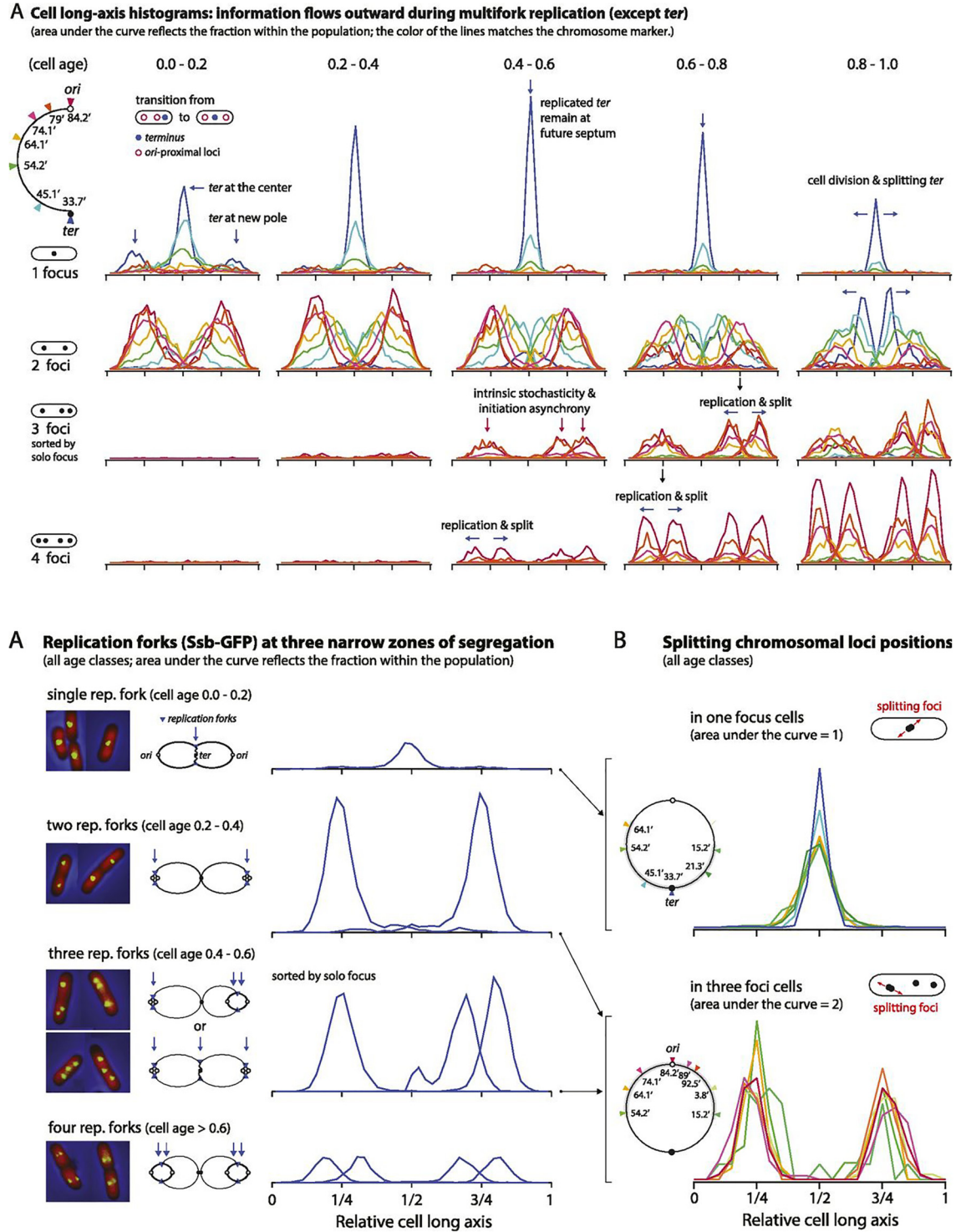


FIG. 7. Experimental data of loci positions during cell cycle: This figure has been reproduced from previously published data in Ref. [24]. We reproduce Figs. 2 and 3, respectively, from the paper of Ref. [24] (after having obtained requisite permissions) for aid of comparison with our modeling results, presented in Figs. 6, 8, 9, 11, and 12. The top panel with four rows shows spatial distributions for different fluorescently tagged loci at different stages of the cell cycle. This data has been extracted from a large number of cells at different ages, each with a different number of foci. The positional distribution of different tagged loci along the long axis is plotted for an ensemble of cells with one, two, three, or four foci in the four different rows of the figure, respectively. The figure further shows how the positional distributions of foci change as the cell ages. This is shown in the five columns. Note that in the bottom panel of the figure, we show the positional distribution of the replication forks (bottom left). The replication forks in different cells show different distributions of its positions, indicating different positions w.r.t. to each other at different intervals of the cell cycle. Using our simulation and modeling, we establish that this is a consequence of the internal topology of the ring polymer in Fig. 6 third row. In bottom right of this figure, we show the experimental plot which marks the position along the long axis at which one foci spot split into form two distinct foci.

in our simulations as a consequence of the adoption of the Arc-2-2 polymer topology, as shown in the second row of Fig. 6.

We have two *oriC*s in the (0–0.2) interval in our simulations of the cell lifecycle. At the beginning of the (0.2–0.4) interval, the D1 and D2 *oriC* replicates. According to our model of replication, the D1 *oriC* is re-named GD1 and another monomer is added, which we call GD1'. Similarly, for the D2 *oriC*, D2 is re-named to GD2, and another monomer, which we call GD2', is added.

Thus, from the (0.2–0.4) interval onwards, our simulations have four *oriC*s. We can track the four *oriC*s independently and plot their spatial probability distributions in the last four intervals of the cell cycle. However, in experiments, the two just replicated *oriC*s cannot be distinguished in FISH data in the (0.2–0.4) interval, since the newly replicated *oriC*s segregate after a certain interval of time, known as the cohesion time. Hence, for experimental data shown in Fig. 7, the spatial distribution of four *oriC*s do not appear till the (0.4–0.6) interval of the lifecycle. Consistent with the observation of cohesion time, the distributions of the *oriC*s overlap significantly in our simulations in the (0.2–0.6) interval of the lifecycle. This overlap decreases as the *oriC*s localize to new positions in later intervals.

In the simulation data, we observe that two of the four GD-*oriC*s show relatively high values in the spatial distribution near the cylinder ends for the three intervals corresponding to (0.2–0.8) intervals of the lifecycle. This is not seen in experimental data. This is because *in vivo* the bacterial chromosome is condensed to stay in a region called the nucleoid within a spherocylinder due to crowders or other mechanisms, e.g., presence of spherical end of cell [56]. These effects, which could move loci distribution away from the poles, have not been incorporated in the current study. Moreover, in experimental plots data presented in rows 2 and 3 of Fig. 7, we do observe rather broad *oriC* distributions for the two- and three-foci data in the (0.4–0.8) interval of cell cycle; and even wider distributions in (0.8–1) interval for data with two foci, i.e., when the two loci cannot be distinguished. Thus, our results are consistent with those obtained from experiments.

Absence of 3 oriC in the plots. In Fig. 6, we do not have a scenario with three *oriC* foci. But *in vivo*, one can observe three *oriC* foci in the cell because the replication starts need not be perfectly synchronous (as in simulations). Moreover, segregation of the newly replicated *oriC*s may proceed at slightly different rates due to inherent stochasticity. This can give rise to different cohesion times for different pairs of replicated *oriC*s, which can result in the observation of three *oriC*s. This difference, and the fact that in experiments, one cannot distinguish between the two to identify whether a locus belongs to D or GD-chromosomes also hold true for the data of the spatial distribution of other loci, as presented later. We observe three peaks of the *oriC* when we plot our data in Fig. 8 and SI-12 [25], similar to how the experimental data is plotted, to enable direct comparison with experiments. In experiments, one can only track the number of cells with two, three, or four fluorescent foci and plot the spatial distributions.

In Fig. 8 (and in figures shown in SI-12 [25]), we have plotted the distribution of *oriC* (other loci) using the

convention that if the distance along the long axis, between two loci is less than a cutoff a_c , the two loci will be counted as one focus. We do this to have a more direct comparison to the experimental data. This method allows us to easily have the positional distribution of *oriC* data with two, three, or four foci. As seen in experimental data for fast-growth conditions, one never has a situation when there is only one *oriC* because the cell is born with two copies of the *oriC*. We have also started with such a configuration. In Fig. 8, we have used the above convention to sort the positional distribution data into sets corresponding to the number of foci, as is done in experiments. Although multiple copies of a particular locus might exist, they will not be distinguished as separate foci if spatially close. We have chosen $a_c = 2a \approx 0.1L$ for data shown in Fig. 8. For a different choice of $a_c = a \approx 0.05L$, data is shown in SI-12 [25]. Although this allows for a direct comparison, we lose out on the resolution of our simulation.

In Fig. 8, in the interval (0.2–0.4), we observe that the rather quick separation of *oriC*s in simulations contrasts with the experimental data. In the experimental data, one can primarily distinguish positions of only two foci in this particular interval, though there is likely to be four *oriC*s in the cell since the beginning of the interval. As the GD1' monomers get added due to replication, the GD1' chain remains connected at the RF's to the D1 chain. This ensures that the two *oriC*s of GD1 and GD1' are spatially proximal post replication; this contributes to the two foci data in the plot. As the length of the GD1' chromosome increases the entropic repulsion between the existing Loop-1 and Loop-2 and the GD1' strand increases. This pushes the *oriC* of GD1' toward the cell pole, giving rise to the four-foci distributions in the interval (0.2–0.6) of the lifecycle. A reminder to the reader is that the monomers which get crosslinked to form Loop-1 and Loop-2 for GD1' chromosome do not replicate until the middle of the (0.6–0.8) interval. Also, *in vivo* the chromosome occupies only 60% of the volume of the cell. If we incorporate this aspect in our simulation, one could expect these distributions to be shifted closer to the centre. This would also lead to low probabilities of occupation at the cylinder ends, as seen from our calculations presented in the third row of Fig. 8. In the (0.6–0.8) interval, we mostly see four segregated foci, but we also see a contribution to probability distribution corresponding to data when only two foci are seen. This is because the 125th monomer (and 375th monomers) of the D1 and D2 chain get replicated, and two additional crosslinks between these and the *oriC*s of GD1' and GD2' are created. This brings the pair of *oriC*s close to the quarter positions in each half of the cylinder. In the last interval of the cell cycle (0.8–1), there are mostly only 4 peaks, as seen in the plots above, which match well with experimental data. The mutual repulsion between the Loop-1 and Loop-2 of GD1 and GD1', and similarly the GD2 and GD2', keep all *oriC*s well separated.

oriC localization in the later stages of lifecycle. In simulations, as the RFs move from monomer 1 (*oriC*) toward monomers 125 and 375, the monomers belonging to D1 and D2 chains get re-named to GD1 and GD2 monomers. In addition, monomers are added to create the GD1' and GD2' chains, whose lengths keep increasing as RFs keep

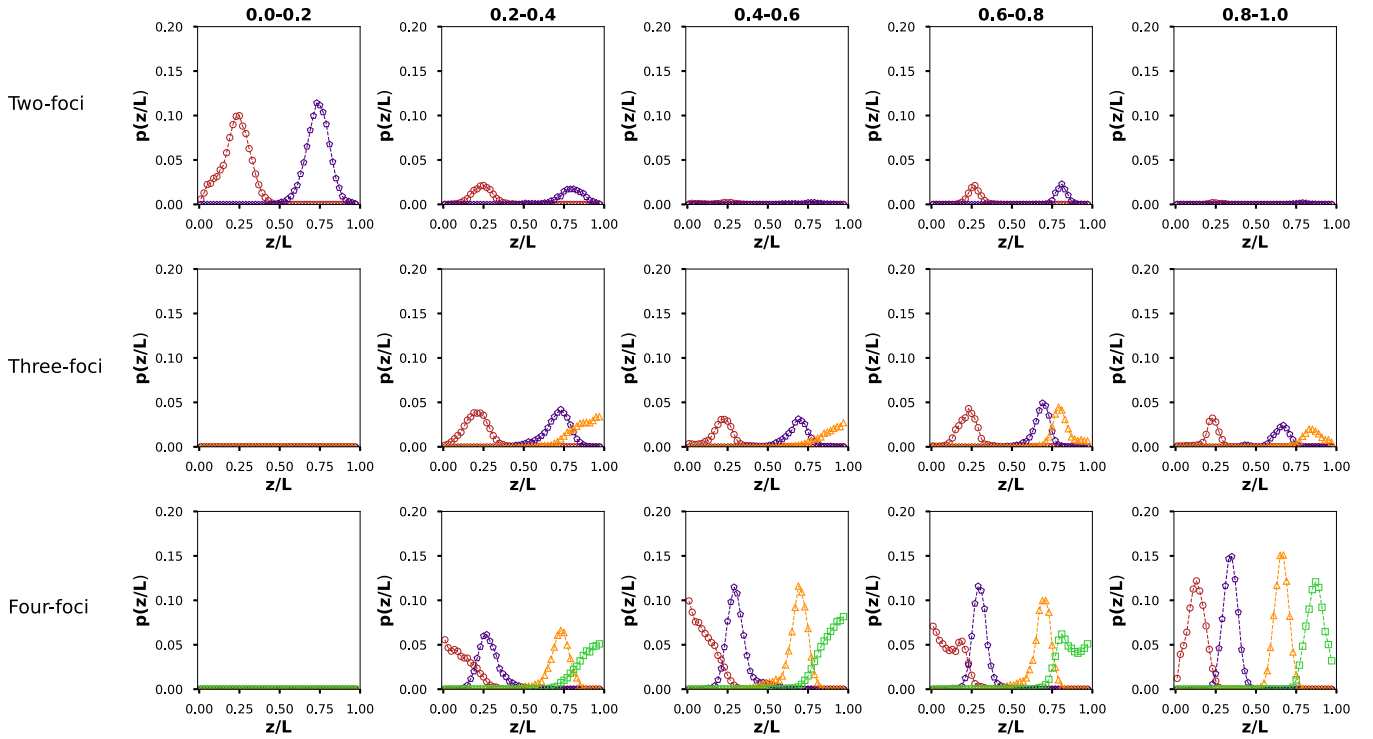


FIG. 8. In the figure, we have plotted the positional distribution of *oriC* loci with a convention similar to that used for plotting data from experiments. If two loci cannot be distinguished due to spatial proximity, the two loci appear as one focus under the microscope. In steady state fast growth, one does not have a cell with a single focus for *oriC*, and hence we show spatial distribution data for only two foci, three foci, and four-foci cases in the different rows of the figure. The different columns refer to the different intervals in the cell's lifecycle. We see that in the interval (0–0.2), the two *oriC* loci are localized to the quarter positions, as can be deduced from the peaks in the probability distributions. In the second interval, i.e., (0.2–0.4), there are 4 *oriC*s, and hence, in simulations, there are four distinct peaks indicating 4 distinct foci as seen in the third row. We also plot two foci peaks corresponding to the set of configurations, where the distance between *oriC*s is less than $2a$. The sum of probability distribution in individual subfigures is not 1, but the sum of probabilities summed over subfigures in a column will be 1.

moving away from *oriC*s. The CLs between the 125th and 375th monomer and the *oriC* of D1 (and also for D2) remain present, as D1 gets converted to GD1 chromosomes (and similarly from D2 to GD2 chromosome) in the (0.2–0.4) and (0.4–0.6) intervals. We refer to them as GD1 and GD2 only for ease of communication; otherwise, there is nothing to distinguish between GD1 and GD1' in our simulations. Additional CLs are introduced to create Loop-1 and Loop-2 in the GD1' and GD2' chains in the middle of (0.6–0.8) interval. It is only then that the additional 125' and 375' monomers get introduced in the cylinder, i.e., after the RF's pass these points on the chain contour of the D-chromosomes. Thereafter, one has four Loop-1 and four Loop-2, one pair on each of the four GD-chains. Thus, from the middle of (0.6–0.8) interval of the lifecycle, the two *oriC*s belonging to GD1 and GD1' start to occupy the 1/8 and 3/8 positions in one half of the cell. Correspondingly, the other two *oriC*s (belong to GD2 and GD2', respectively) occupy 5/8 and 7/8 positions in the other half of the cell, leading to the appearance of four peaks at these position distribution plots for (0.8–1) interval in Fig. 6 and Fig. 8. This is seen *in vivo* and in our simulations, and the localization occurs due to entropic repulsion between GD-loops. The peaks are enhanced in the last interval of the lifecycle.

Position of replications forks (RFs). The confidence in our model is further strengthened by the reconciliation of

the spatial distribution of the RFs from our model and *in vivo* results. We have access to the spatial coordinates of the monomers on which the RFs are located throughout our simulation, corresponding to a cell's lifecycle. refer 3-rd row of Fig. 6. In the (0–0.2) interval, the two RFs are situated on the M-chromosome arms, and move from the 200th (and the 300th) monomer to the *dif-ter* position. The cell center gets occupied also by the *dif-ter* loci, soon after cell division due to reasons already explained previously. Thereby, there is a peak in the spatial distribution $p(z/L)$ of RFs at the center of the cell. At the end of the (0–0.2) interval, the replication of the M-chromosome is complete and one has two complete D-chromosomes connected at the *dif-ter* in the Arc-2-2 architecture. The Arc-2-2 architecture ensures that the *oriC* are in the quarter positions at this stage of the lifecycle. At the start of (0.2–0.4) interval, replication of the two D-chromosomes begins from the two *oriC*s. Thus, four new RFs start at the position of the *oriC*s, i.e., at quarter positions, and start creating the GD-polymer segments. Thus, the RFs are found at the quarter positions in our simulations in the (0.2–0.4) interval, and are consistent with the experimental data reported. Thereafter, the RFs start moving along the two arms of each daughter chromosome assuming the train track model of RF-movement. The RFs move toward the poles because the loci that are being replicated, are also closer to the poles due to Arc-2-2 architecture, as explained before. As

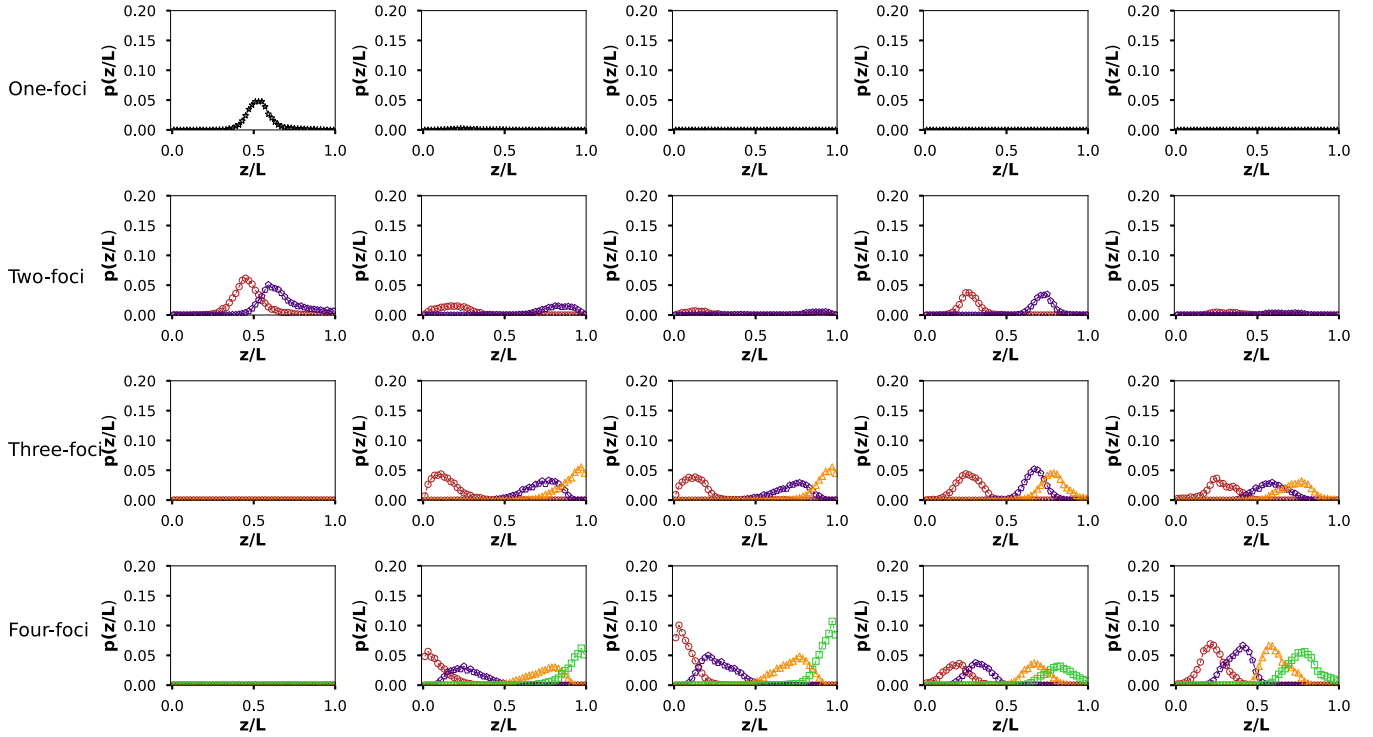


FIG. 9. The probability distribution of the spatial position of the replication forks (RFs) at different intervals of the lifecycle. The plot follows the conventions of Fig. 8 which enables direct comparison with the RF positions observed experimentally and presented in Fig. 7. In interval (0–0.2), the two RFs of the D1 and D2 chromosomes can be distinguished, and hence their spatial distributions is shown in the two-foci data: second row. However, toward the end of the interval as the two RFs reach close to the *dif-ter* loci, the spatial separation is often less than $2a$, and for those cases, spatial distribution of the two RFs is shown in the one-foci row as a single peak near the cylinder center, where the *dif-ter* is located. In the intervals (0.2–0.6), the pair of RFs in each half of the cylinder, can be distinguished at times and hence RFs show up their distributions in the two-foci, three-foci, and four foci rows. In the middle of the (0.6–0.8) interval, the RFs are spatially close to each other near the quarter positions, as they replicate the 125th and the 375th monomer and start moving toward the *dif-ter*. Depending on the distance between the RFs in different microstates, the spatial distribution can contribute to the two-foci, three-foci, and four-foci rows. In the last (0.8–1) interval, the RFs can mostly be distinguished due to the repulsion between internal loops in the GD chromosomes, and thus one sees contribution in from four foci for the four loci in the last row. The data is suitably normalized depending on the number of configurations for the four different foci cases in a particular interval.

an example, at the end of (0.2–0.4) interval the RF reached locus 50 and 450 on the left arm, which is located near the poles. Also refer the review section and SI-8 [25], where we show loci localization of monomer 50 when replication is switched off. Thereby RFs show a higher propensity to be in cylinder-ends in the (0.2–0.6) stage, though it peaks near the quarter positions as in experiments. Experimentally, the presence of the nucleoid ensures that the probability of finding the RFs at the cell poles is zero.

The peaks in the spatial distribution of RFs near the quarter position can be observed for the (0.6–0.8) interval in Fig. 6, and can be understood as follows. In this stage, the RF's traverse the D chromosomes from the monomer 100 (400 on the other arm) to 150 (350 on the other arm), and reach 125 (375) at 0.7τ . We have crosslinks at 125 and 375 on D1 and D2. These monomers are connected to *oriC*. As shown before *oriC* is at the quarter positions. Any monomer connected to *oriC* will also be at the quarter positions. This implies that 125(375) are at quarter positions. Since the RF's are traversing around these monomers, the RF's also localize at the quarter positions. The same raw data has been analyzed and presented in the convention, which

allows for a more direct comparison with experimental data in Fig. 9.

Role of Loop-3 and Loop-4 in RF positioning. In the last (0.8–1) interval of the cell cycle, the RFs move toward *dif-ter*s of the two D-chromosomes, starting out from the CL sites, as mentioned in the previous paragraph. Thus, there is a higher propensity for them to be near the cell center, but there are indications of some spatial separation along the long axis in the position of the peaks of the two RFs in one-half of the cell in the data from our simulations as seen in Fig. 6 (third row). In contrast, the experimental data for cell-age >0.6 in Fig. 7 clearly shows a prominent separation of peaks in the spatial distribution of RFs.

To check the reason for this discrepancy of our data (averaged over 50 runs) from experimental data, we plot the spatial distribution of the RFs from individual runs in the (0.8–1) interval of the lifecycle: refer to SI-13 [25] for data on spatial positions of RFs from 50 individual runs. Many of the data from individual runs clearly show 3 to 4 peaks, consistent with experiments. Note that when we plot the averaged data shown in Fig. 6 (third row), there is no clear separation in the positions of different peaks of the distribution. This because the

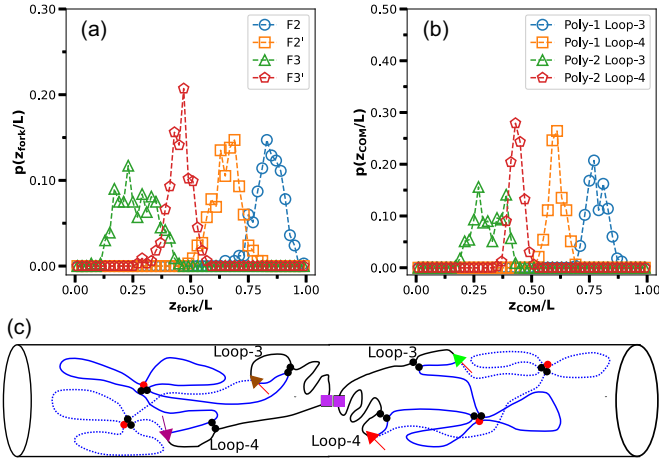


FIG. 10. The plots show data spatial probability distribution data for (a) the RFs and (b) the center of mass (COM) of loops Loop-3 and Loop-4 from a **single** independent run for the interval (0.8–1). We draw attention to the similarity in position of peaks in the spatial distribution of the quantities plotted in the two graphs. The replication forks have higher probability of occupying the same positions occupied by the COM of Loop-3 and Loop-4. The F2 and F2' denote the RFs of D1 which create GD1 and GD1', whereas F3 and F3' are the RFs of D2. Below we show a schematic on how the loops could be arranged in this complex scenario.

RFs can be positioned differently relative to each other, across different simulation runs. In a particular run, the RF from the left arm (of say D1) can be closer to the middle of the cylinder, whereas the RF from the right arm of D1 might be closer to center in another run. We compared the spatial distribution of RF data obtained with Arc-2-2 with that obtained using the Arc-2 architecture in SI-13 and SI-14 [25], respectively. In comparison, the data obtained using Arc-2 architecture shows lesser distinct separation between the peaks, which indicates that Loops-3 and Loops-4 (absent in Arc-2 architecture) play a role in the separation of peaks, seen in the spatial distribution of RFs.

For further investigation of the above observations, we plot the spatial distributions of the center of mass (COM) of Loop-3 and Loop-4 in interval (0.8–1) for each of the individual 50 independent runs, refer to data provided in SI-15 [25]. Furthermore, we compare this distribution with the distribution of RFs for Arc-2-2 architecture for each independent runs. We do this because at this stage of the lifecycle, the RFs are traversing along the contours of Loop-3 and Loop-4. We find reasonable one-to-one correspondence in the position of the peaks of the distribution for the COM of loops and the RFs. The representative data for one particular run is shown in Figs. 10(a) and 10(b). Refer to SI-16 [25] for one-to-one comparison for 5 different independent runs out of 50, which are shown individually in SI-13 and SI-15 [25]. This implies that the RFs remain separated along the cell's long axis in the region between the quarter position and the center of the cylinder. The Loops-3 and -4 can exchange positions as they occupy positions along the long axis with only partial overlap. This is a consequence of mutual entropic repulsion between Loops-3 and -4. Entropic interactions with Loops-1 and -2 of both the replicated chromosomes position the Loops-3 and

-4 away from the cylinder poles. All loops jostle for space to avoid each other and often end up interchanging positions along the long axis.

Difference in the behavior of Loop-1 and Loop-2 versus Loop-3 and Loop-4. Why do the spatial distributions of the RFs in the (0.2–0.4) and (0.4–0.6) intervals not show four peaks in contrast to the distribution in the (0.8–1) interval? This is because in the earlier stages of the lifecycle, the RFs traverse along the chain contours of Loop-1 and Loop-2 of DNA-1 and DNA-2. They reach Loops-3 and -4 only in the last stage of the lifecycle. The Loop-3 and Loop-4 behave differently from Loop-1 and Loop-2, as they are closer to the *dif-ter* CL, and thereby try to avoid overlaps not only with each other but also Loop-3,4 of the other polymer and the *ter*-segments (Loop-5) from both polymers. As a consequence Loop-3 and Loop-4 have a greater propensity to interchange positions along the cell long axis, as compared to that of Loops-1 and -2. The interchanging of loops are better visualised in Fig. 10. To have a improved understanding of how different polymer architectures affect the organization of loops with respect to each other, refer to our article [37].

Position of loci-split. The location of each locus as they get separated from its copy (post-replication) was studied in experiments [24]. The distribution of these positions along the long axis for each locus is reproduced in the bottom right panel of Fig. 7 for the aid of the reader. From our simulations, we can also obtain this distribution of “position of split” of the relevant replicated monomers corresponding to the tagged loci in experiments; and we show our data from simulations in Fig. 11. We have more resolution than in experiments, but while comparing our modeling results with the experimental distributions one also needs to account for the fact that we do not have nucleoid or the spherical ends of the cell. The experimental data divides the set of loci into two sets, viz., the ones which split at the cell center and those who primarily split at the quarter positions. When we present data from our simulations, we provide the distribution of the position of split in three subfigures, i.e., (a) the ones which are located near the *dif-ter* loci along the chain contour, and hence spatially close to the position of the *dif-ter*, i.e., near the cell center (b) the loci which are located closer to the *oriC* loci (along the chain contour) or the monomers (125 and 375) crosslinked to it, which are spatially located near the quarter positions, and (c) the monomers which are located on the Loop-1 and Loop-2 but away from CLs and these get replicated closer to the cell poles. As a consequence the position of split is also spatially closer to these positions.

Data normalization: Experiments versus Simulations. There are some other caveats that the reader must take into account while comparing data from simulations with that of experiments. In simulations, we never see a pair of distinctly separated peaks of the *ter* loci distributions, as we do not model cell division. However, the experimental data shows finite probability for two *dif-ters*, even when the cell is in its (0.6–0.8) interval of lifecycle (as deduced from the length of the cell) as well as for the interval (0.8–1). In experiments, the cell lengths are used as a proxy for the age of the cell. These could give rise to discrepancies when analyzing data by image processing software. Moreover, there are differences in the methods of collecting data and normalizing the spatial

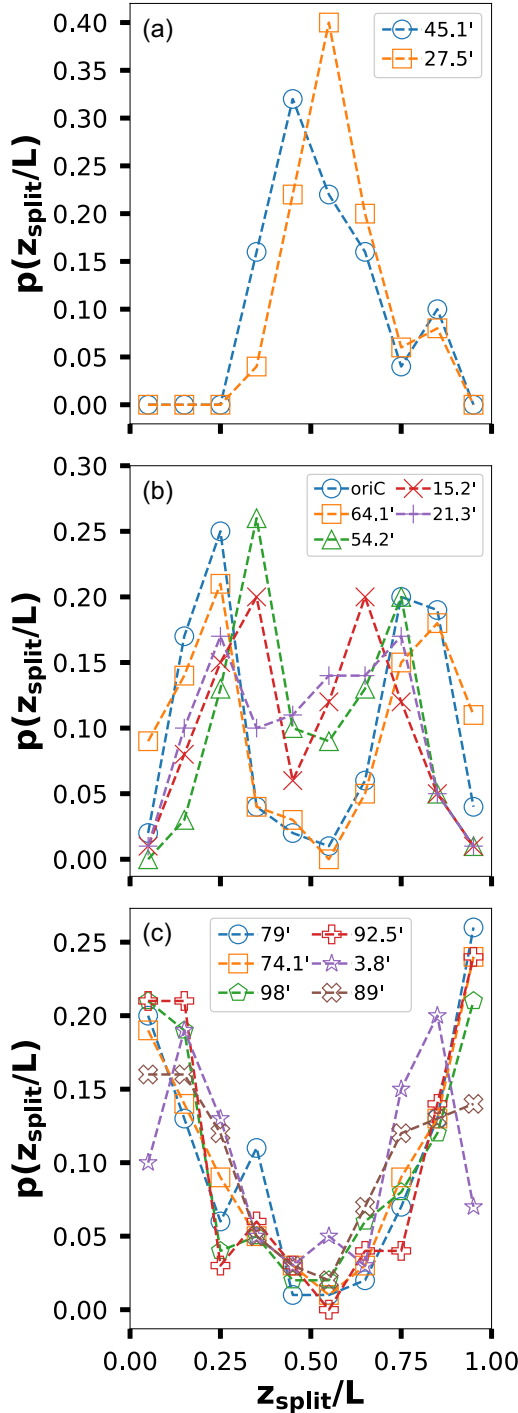


FIG. 11. The probability distribution of the position of split for pair of replicated loci along the long axis. There is a distinctive two peak distribution for the loci initially having two copies, i.e., for D1 (and D2) loci replicating to GD1-GD1' (to GD2-GD2') loci and splitting thereafter. There is a single peak distribution for the loci that only have a single copy initially, i.e., for M-loci replicating to D1 and D2. We specify that the loci are spatially segregated if their axial distance is greater than $2a$. When the axial distance becomes $>2a$, we calculate the midpoint of the two loci and identify the point as the point of split. Since we have 50 independent runs, there are only 100 replication events (data points) to calculate the position of “split” for the two-peaked distributions. Moreover, we have 50 points for single peaked distribution.

distributions. In the given experimental data, the cells were first differentiated by their cell size to categorize the age of the cell. For each such category, the number of distinguishable foci in each cell is observed and spatial distribution data corresponding to the number of foci observed, they are normalized with respect to the number of cells in each subcategory. The experimental images cannot discern if the foci belong to the D-chromosome or GD-chromosome. Furthermore, if the foci cannot be spatially resolved, a four-loci cell may be erroneously categorized as a three-foci or a two-foci cell. The three foci scenario may also arise due to asynchrony in the replication initiation process or stochasticity in the cohesion times. In the simulations, we do not have asynchrony in the replication initiation process, however, there may be stochasticity in the cohesion times of the loci. But, we have access to the positions of each monomer at all stages of the simulation run. Thus, our normalization protocol is different from that adopted in experiments.

In simulations, we have 50 independent runs to collect data over the entire lifecycle. We precisely know the stage of the life-cycle of the model cell, and the stage when a loci is replicated to two loci of the next generation. The RF moves to the next monomer on the contour every 2×10^5 MCS. We thereby normalize by the precise number of microstates relevant for a particular monomer, depending on when that monomer has been introduced within that interval of the lifecycle. For each independent run having a total of 5×10^7 MCS, we store data to calculate distributions every 3×10^4 MCS. Thereby, we know the number of contributing microstates in each stage of the lifecycle for each loci.

Spatial distribution of other loci. We now calculate the spatial distributions of the other loci that were tagged in experiments and compare the distributions obtained by our simulations to those obtained by [24], refer to Fig. 7 for comparison. Here, we provide data for five such loci from the left arm in Fig. 12. The data for the loci on the right arm are given in SI-17 [25].

For the locus marked as 54.2' (monomer 150), we see only two distributions for (0–0.8) intervals of the lifecycle, as the monomers of the daughter chromosomes get replicated only at the end of the fourth interval. Our modeling data is in fair agreement with the experimental data. We find that there are four peaks toward the end of the lifecycle. Experimentally, it is not possible to distinguish between the two GD loci near the end of the lifecycle, as they cannot be spatially resolved if are spatially proximal. In our simulations, we can uniquely identify the loci of each GD chromosome, and thereby, we obtain four distinct spatial distributions, albeit they overlap. In experiments, the loci distributions (that we obtain) will appear as broad distributions having only two distinct peaks. In this sense, our spatial distribution of loci is in agreement with the experimentally obtained distributions [24].

The data obtained from simulations for the locus marked as 45.1' (monomer 200) also shows good agreement with the experimental data for two foci. We do not have four peaks for this locus in the (0.8–1) interval (as in experiments) since the simulations (cell cycle) are stopped just as this specific locus gets replicated. This is because cell division takes place at this stage of the cell cycle. Correspondingly, the experimental data for four foci also does not have any contribution from

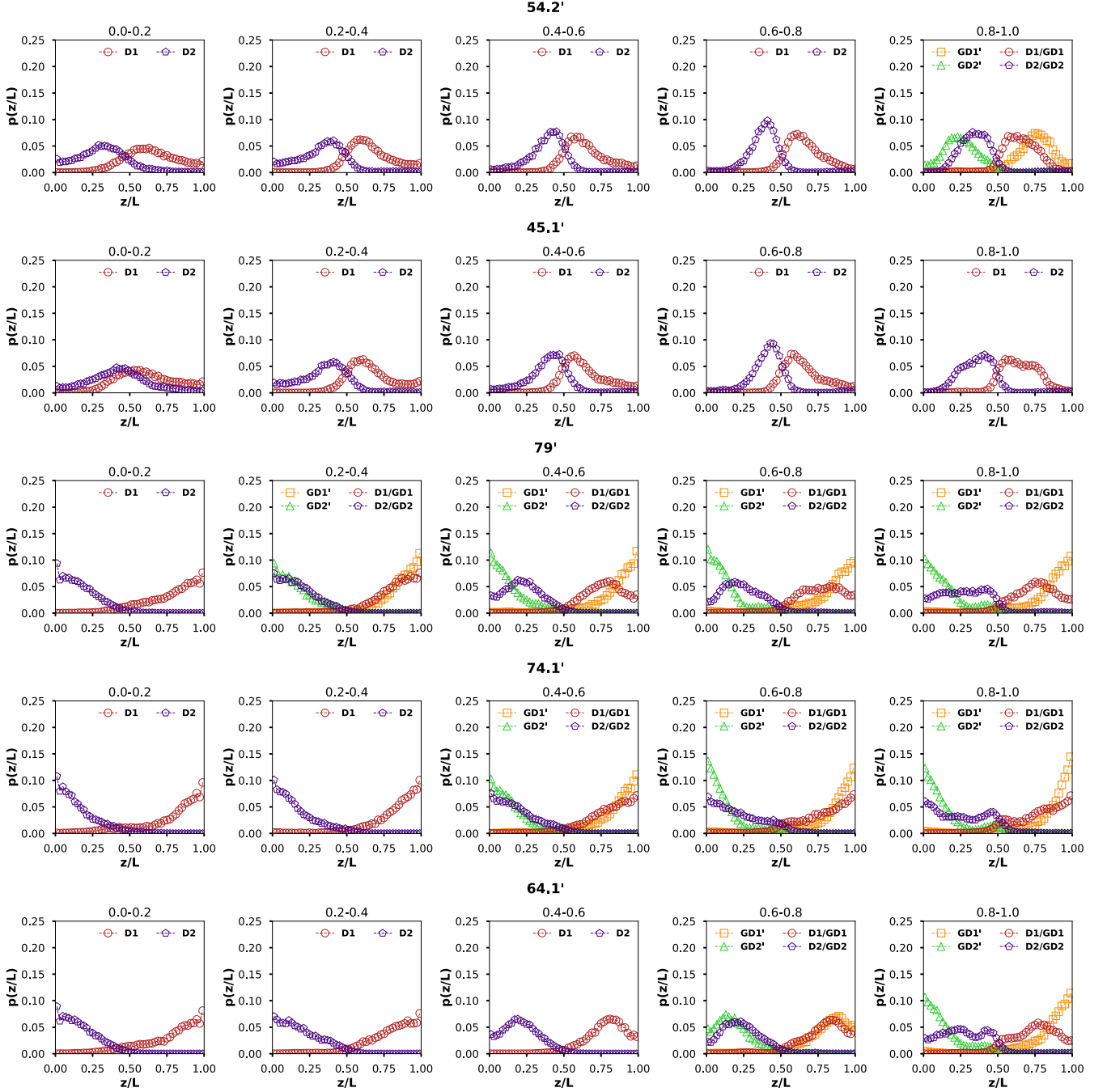


FIG. 12. Long-axis distribution for other tagged loci: We plot the spatial probability distributions $p(z/L)$ of the position of different loci, where z denotes the position along the long axis of the cylinder (cell), and L is the length of the cylinder at that stage of the simulation run. Data is shown for 54.2' locus corresponding to monomer 150 in our simulations (first row), for 45.1' locus corresponding to monomer 200 (second row), 79' locus corresponding to monomer 26 (third row), 74.1' locus corresponding to monomer 50 (fourth row) and 64.1' locus corresponding to monomer 100 (fifth row), during the lifecycle. The corresponding monomer indices are at the top of each row. The other plotting conventions are same as in Fig. 6.

this locus in the (0.8–1) interval. Furthermore, comparing our data in other intervals with the two foci data in Fig. 7, the distribution of the locus 45.1' is peaked near the center as this locus is close to *dif-ter* at initial intervals. As the cell ages, the loci move away from each other, hence delocalizing from the center of the cell due to the presence of mutual repulsion between Loop-3 and Loop-4. But once the Loop-1

and Loop-2 of the GD1' and GD2' chains are formed, they are pushed away from the poles in the (0.6–0.8) interval. Thus, our data are consistent with the experimentally obtained two-foci data shown in Fig. 7.

In the third row of Fig. 12, we show data for the locus 79' (monomer 26). We note that the distributions from our simulations peak around the cell poles at the (0–0.2)

interval. This differs from two foci data from experiments in the (0–0.4) intervals, where they find that this locus is localized around the quarter positions. However, if one looks carefully at the two-foci data of Fig. 7, one notices that the distribution has more contribution from regions between the poles and the quarter position. In simulations, we have not incorporated the effects of the nucleoid, the hemispherical poles at the end of the cylinder in our current work. If we introduce these effects in our simulations, we presume that this locus will stay away from the poles. In later parts (0.6–1 interval) of the cell cycle and for data with two or more foci, the experimental distribution is either broad or four peaks are observed for the four-foci case. In simulations, we do have distributions that are spread out over the length of the cell. We observe the spatial distribution of the four loci, where two of the four distributions are closer to the center, though we do not observe distinct peaks in the distribution.

We may reason along similar lines as to why the distributions obtained for the 74' (50th monomer) locus appear quantitatively different to those seen in experiments. This locus again lies within the Loop-1, and would remain away from the cell-center which is the position of the *dif-ter*. While there are broad similarities between the experimental and modeling data, we still have high probability to obtain the loci right at the poles. This differs from the experimental data which has 4 distinct peaks for the four-foci data at the intervals 0.6–1 interval. In our case Loop-1, which carries monomer 50, from GD1 and GD1' (and correspondingly GD2 and GD2') may keep interchanging positions along the long axis. In our ongoing studies without replication (unpublished data), but in the presence of a rosette of small loops in the region 1–125 and 1–375, we observe that the loops repel each other and rarely interchange positions, which would lead to the sharp peaks observed in experiments.

The 64.1' loci (100th monomer) is again on Loop-1 but closer to the position of CLs along the chain contour. Hence we expect this loci to primarily occupy the region between the quarter positions and the poles in the 0–0.6 interval in the two-foci data. This foci gets replicated at the end of the 0.4–0.6 interval, and hence four foci do not appear till after this stage. This is exactly what we see in Fig. 7. After the middle of the 0.6–0.8 intervals, the Loop-1 and Loop-2 of all four GD chromosomes are formed. Thereby, this loci will be near the 1/8 and 3/8 position but relatively away from the cell center. The Loop-1 from GD1 and GD1' may keep interchanging positions along the long axis. Consequently, we see broad distributions in simulations and in experiments. Since the loci 64.1' and 79' are equidistant from the position of the CLs which create Loop-1 and Loop-2, the reader can observe that the spatial distribution of loci is nearly identical in the 0.8–1 interval. However, 79' is replicated at the middle of the 0.2–0.4 interval, whereas 64.1' is replicated at the end of 0.4–0.6 interval. So the four-foci data appear at a later stage for the 64.1' loci.

Organization of chromosomal arms. The organization of the arms of the *E. coli* chromosome has been a topic of considerable interest in the literature. It has been suggested that in slow-growth conditions, the arms occupy different sections of the cylinder along the cell long axis, while in fast-growth conditions, the arms occupy different halves along the short

axis [24,35], i.e., if the arms would occupy different halves if the cylinder were sliced into two halves parallel to the long axis. In this context, the work of Ref. [24] also provides experimental data for the radial distributions of loci along the short axes. They find that many loci show double-peaked distributions, indicating that the two arms lie preferentially on opposite sides along the short axis.

We plot spatial distributions using the loci's x or y coordinate, as both can be used to measure “radial” distance from the central axis. Data from simulations using the Arc-2-2 polymer topology has been plotted in figures shown in SI-18 and SI-19 [25] to estimate the preferential radial distance of loci. We find that these distributions show, in general, broad distributions that peaked at the middle of the cylinder and thus do not agree with what is observed experimentally. Experiments show doubled peaked radial distribution for some of the loci.

Thus, we propose modifications to a simpler version of our model with replication switched off and discuss the loop-based mechanism to obtain the longitudinal organization of chromosomal arms. The longitudinal organization of the arms can be inferred from the bimodal radial distributions of loci. We introduce six smaller subloops in each of Loop-1 and Loop-2 of size 10 monomers each. These subloops are equally spaced with 10 monomers in between. Introducing these subloops enhances the entropic repulsion between Loop-1 and Loop-2 along the short axis. Consequently, the monomers belonging to those loops show a bimodal distribution. We establish this emergence of a bimodal distribution through simulations of two (modified) Arc-2-2 polymers that have segregated along the long axis, with additional subloops of size 10 monomers in Loop-1 and Loop-2. These subloops are created by introducing extra CLs between monomers separated by 10 monomers, i.e., between 11 and 20, 31 and 40, and so on. We do not incorporate replication since we only outline the mechanism by which one may obtain such bimodal distributions. The entropic repulsion will likely act through transient loops *in vivo* unlike the long-lived permanent loops we consider for our model.

To establish that Loop-1 and Loop-2 lie in different halves of the cylinder, along the radial axis, we carry out the following calculation. We construct a vector $\vec{l}_1$ joining the mid-point of the cylinder and the COM (center of mass) of Loop-1. Similarly, we construct another vector $\vec{l}_2$ joining the center of the cylinder and the COM of Loop-2. Note that we only consider the $\hat{x}$ and $\hat{y}$ components of the vectors. The angle between the two vectors is denoted by θ . Then, if $\cos\theta \approx -1$, the two vectors are antiparallel to each other. This implies that the two loops, Loop-1 and Loop-2 (belonging to different arms of the chromosome) lie in different cell halves along the short axis. As can be inferred from Fig. 13(a), introducing smaller additional loops leads to the separation of arms along the short axis. We further note in Figs. 13(b)–13(h) that the introduction of subloops also leads to bimodal distributions of some loci along the short axis, similar to what is seen in Ref [24]. Other loci show broader radial distributions as compared to radial distributions for Arc-2-2 polymers without smaller loops (refer to SI-18 and SI-19 [25]), even if they do not show bimodal distributions. To check for the robustness of our conclusions, we conduct similar simulations with five smaller subloops

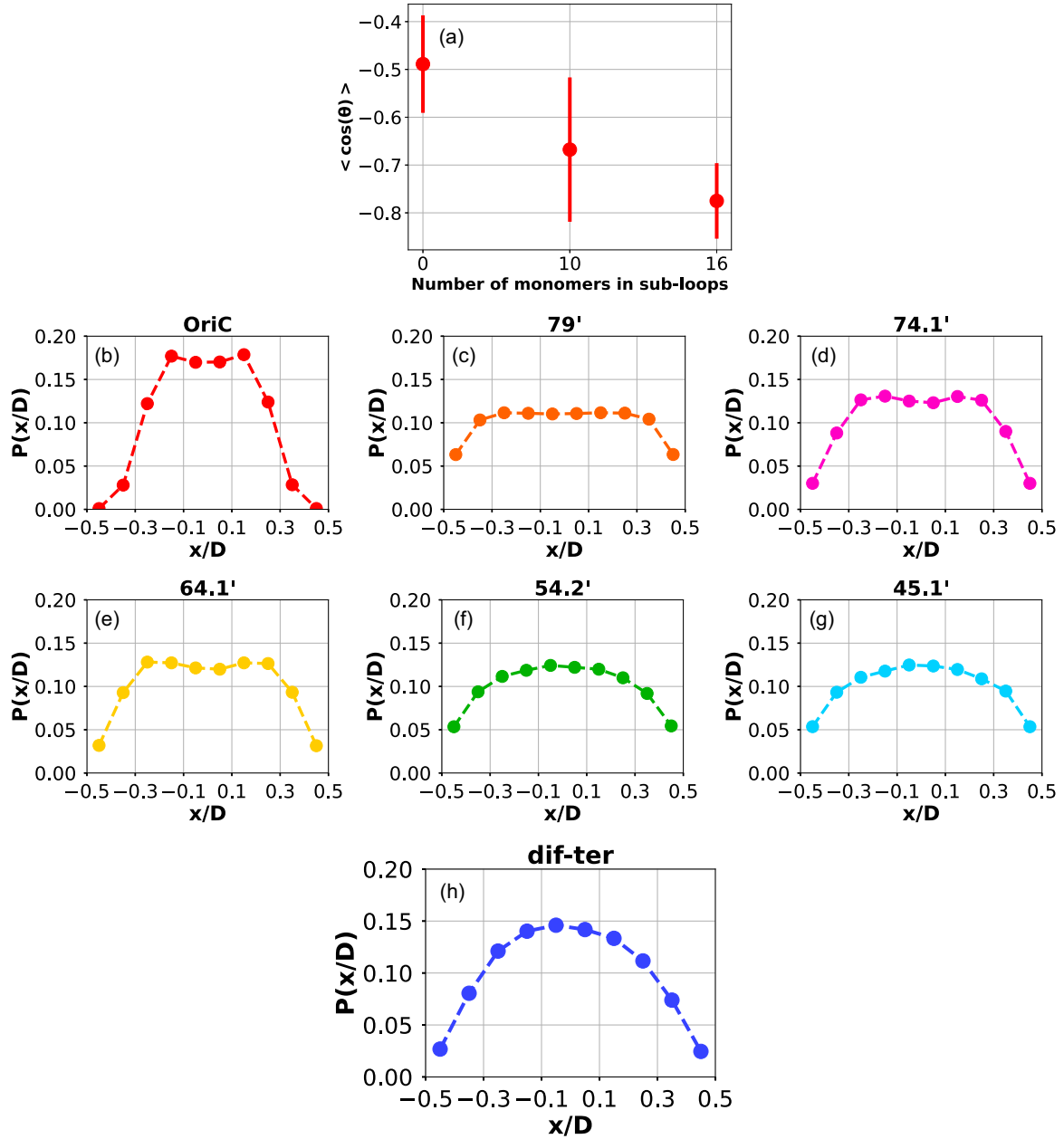


FIG. 13. Organization of chromosomal arms and radial distribution of loci: Panel (a) shows $\langle \cos(\theta) \rangle$, where θ denotes the angle between vectors $\vec{l}_1$ and $\vec{l}_2$ (refer text). A high negative value of $\cos(\theta)$ indicates that the two loops (belonging to the two arms of the chromosome) lie on different cell halves along the radial axis. We observe that the average $\cos\theta$ value is more negative in the cases with smaller loops (within Loop-1 and Loop-2) as compared to the case without smaller loops. Panels (b)–(h) show the radial probability distributions of the monomers for several loci in the presence of smaller loops. These smaller loops are of size 10 monomers each and are placed along loop-1 and loop-2. We show here the data for the loci of the left arm, while the corresponding data for the right arm loci have been provided in the Supplemental Material (SI-20) [25]. We note that we obtain a bimodal distribution for some loci by introducing these smaller loops. The distributions we obtain match those found in the experiments for the loci *oriC*, 79', 74', 45.1' *dif-ter* along the left arm only. The experimentally obtained radial distributions for the loci have also been reproduced in SI-21 [25] for aid of comparison. We do not obtain a match for the locus 54.2' and data for some other loci on the right arm (presented in SI-20) [25]. We also notice that we obtain a double peaked distribution of the *oriC* while the experimentally obtained distribution has a single peaked distribution.

within Loop-1 (and Loop-2) with 16 monomers in each of the subloops. This also shows the separation of these arms as can be inferred from values of $\langle \cos(\theta) \rangle$, refer to Fig. 13(a). This, however, fails to show bimodal radial distribution.

Furthermore, one may also introduce smaller loops along the rest of the Arc-2-2 polymer (outside loop-1 and loop-2)

to obtain the bimodal distribution of other monomers. In the experimental data of Ref. [24], some loci show single peaked distributions while others show bimodal distribution. Even in our simulations, some monomers show single peaked distributions while others show bimodal distributions, although an exact match is not obtained with experiments. To obtain

an exact match, one needs to optimize the size and location of loops along the contour, which is outside the scope of the current manuscript. We have also shown in SI-22 [25] that the localization of the tagged loci along the long axis which was discussed previously, remains unaffected with the introduction of smaller loops.

V. DISCUSSION

We establish that entropic repulsion between internal loops is a viable mechanism through which the chromosomes segregate and organize themselves. We show that the organization of *oriC*, *dif-ter*, and other loci emerge spontaneously in our model, both along the longitudinal and the radial axis. Moreover, we outline the key mechanisms governing the localization of loci both along the radial and long axis of the *E. coli* cell. Our model also successfully reconciles other experimental observations, such as the spatial organization of replication forks. Though we have implemented the replication process following the train-track model, our simulations show the localization of the RFs, which supports (in spirit) the replication-factory model. Thus, the hypothesis proposed by the authors of Ref. [24]: “The position and dynamics of the replication forks are likely the consequence of the spatial organization of the chromosomes rather than vice versa” is supported by our simulations. With this work and our past paper on the organization of *E. coli* chromosomes in slow-growth conditions using the Arc-2-2 architecture [36], we propose that this model of the *E. coli* chromosome provides a viable mechanistic understanding of chromosome organization in all growth conditions. To the best of our knowledge, we are the first to explicitly model the replication and evolution of the organization of the chromosomes in the complex case of overlapping rounds of replication.

Despite the many successes of the model in reconciling the broad features of chromosome organization in fast-growth conditions, future studies may improve upon this model in some areas. We were not able to match all the details presented in Ref. [24]. For instance, we note in the data presented in Ref. [24] that the probability distributions of loci keep shifting toward the quarter positions along the long axis. The time taken by each locus to move to the quarter positions is a function of distance from *oriC* along the contour of the polymer. Our simulations fail to capture this aspect. Moreover, some loci in our simulations have a higher probability of being at the cell poles, which is absent in the experimental data. We attribute this to not incorporating the presence of crowders which are known to condense the chromosome through depletion interactions [28,46,57–61].

We remind the reader that we constructed a minimalist model with only four additional CIs to obtain a mechanistic understanding of the localization patterns of genomic segments as seen *in vivo*, from FISH experiments. The aim was to provide an underlying mechanistic explanation for the experimental observations seen in both slow [36] and fast-growth conditions. We are fully aware that this minimal model cannot be a complete and accurate description of all the phenomena associated with the *E. coli* chromosome. The *E. coli* chromosome is significantly more complex than our description of it with a minimalist model, with a variety of

different subcellular phenomena affecting its properties. Thus, the model has ample scope for extension to incorporate relevant biological processes more accurately and make more improved predictions thereafter.

The CLs in our simulations are likely mediated by linker proteins, such as MukBEF or H-NS, *in vivo* [6,8,62,63]. MukBEF complexes, dissociating every 60 s *in vivo*, are observed in clusters at ori proximal regions. Despite the dissociation of individual complexes, replacements within the cluster may create new crosslinks. Continuous loading and unloading of MukBEF complexes may lead to slight variations in crosslink positions. But we have shown that altering positions of CLs in our simulations by five monomers (i.e., ≈ 45 kbp) on either side does not impact *oriC* localization.

Extrusion due to MukBEF can also form smaller transient loops *in vivo* [6]. A series of smaller loops can bring distant DNA segments closer spatially. This can result in a scenario that resembles “effective” CLs between distant monomers along the polymer chain, such as those we have considered in this paper. Other proteins, like H-NS, potentially mediate long-range interactions [62,63] and may act as crosslinks.

Simulations by other groups have shown that entropic effects persist even with transient extruded loops. Moreover, internal loops within a DNA ring polymer aid the segregation of daughter chromosomes [49]. However, the authors have not shown the organization or localization of loci of the bacterial chromosome with transient loops at random positions. In contrast, we show that introducing smaller internal permanent loops within Loop-1 and Loop-2 in our Arc-2-2 topology keeps the organization of the chromosome along the long axis relatively unchanged. Although in reality, there are likely multiple smaller, transient extruded loops organized in an hierarchical fashion within the cell, we currently have a simplified description of the genome using just four permanent (likely to be long lived *in vivo*) crosslinks. Future experiments can try to identify the protein complexes which mediate the long-lived links at the positions of CLs of the Arc-2-2 model, proposed by us. These connect the *oriC* with points that are nearly halfway along the contour of the left arm and right arm of the DNA. Future work aims to incorporate transient small loops by introducing additional crosslinks within Loop-1 and Loop-2, and systematically investigate its consequences, using appropriate coarse-graining techniques [64] to reproduce the same results we obtain in the manuscript. Note that we keep longer loops, viz., Loop-3 and Loop-4, of length ~ 800 kbp, consistent with what is reported in Refs. [33,65].

Many energy-consuming active processes, such as topological constraint release, motion of the RF, and formation of loops due to various proteins occur within the cell. The consequences of some of these processes are incorporated in an effective manner in the current model. Though the energy consuming cellular processes drive the system out of equilibrium, we use Monte Carlo simulations to realize local diffusion of polymer segments. We do not have an estimate of the time scales involved in our simulation in real units. An estimate of time scales would require using Langevin dynamics simulations. At our current stage of understanding, incorporating (a) replication, (b) topological constraint release due to the presence of topo-isomerase, (c) addition of crosslinks at

specific stages in the cell cycle, especially between monomers that might be spatially far apart, and (d) changing the size of the cylinder systematically is significantly more difficult in Langevin dynamics simulations than in Monte Carlo simulations.

We have outlined principles by which one may obtain the experimental data of Ref. [24] through our model of the DNA polymer with a modified topology. The experiments were conducted on a specific growth medium of the bacterial cells, which determines the value of the doubling time. We have adapted our model similarly to establish a correspondence to the experiments of Ref. [24]. In future manuscripts, we shall communicate our results for a different choice of τ_C , τ_D , and the doubling time as realized experimentally using different strains and growth conditions. We hope our theoretical predictions can be validated by experiments conducted using different growth media.

ACKNOWLEDGMENTS

Authors acknowledge useful discussions with Arie Zaritsky, Conrad Woldringh, Tejal Agarwal, and Suckjoon Jun. A.C., with DST-SERB (IN) Identification No. SQUID-1973-AC-4067, acknowledges funding by DST-SERB(India), Projects No. MTR/2019/000078 and No. CRG/2021/007824. A.C. also acknowledges discussions in meetings organized by ICTS, Bangalore, and use of the computing facilities by PARAM-BRAHMA. We acknowledge the support from women in STEM fellowship from DST and IUSSTF, which enabled Tejal Agarwal to visit the Suckjoon Jun group. The authors thank the anonymous referees for their detailed comments, which helped in the significant improvement of the paper.

S.P. implemented the model and performed calculations and analysis. The research plan was designed and discussed by S.P., D.M., and A.C. S.P., D.M., and A.C. wrote the paper.

- [1] R. Phillips, J. Kondev, J. Theriot, H. G. Garcia, and N. Orme, *Physical Biology of the Cell* (Garland Science, New York, NY, 2012).
- [2] A. Kuzminov, The chromosome cycle of prokaryotes, *Mol. Microbiol.* **90**, 214 (2013).
- [3] 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* **153**, 882 (2013).
- [4] L. Dewachter, N. Verstraeten, M. Fauvart, and J. Michiels, An integrative view of cell cycle control in *Escherichia coli*, *FEMS Microbiol. Rev.* **42**, 116 (2018).
- [5] A. Japaridze, C. Gogou, J. W. J. Kerssemakers, H. M. Nguyen, and C. Dekker, Direct observation of independently moving replisomes in *Escherichia coli*, *Nat. Commun.* **11**, 3109 (2020).
- [6] J. Mäkelä and D. J. Sherratt, Organization of the *Escherichia coli* chromosome by a MukBEF axial core, *Mol. Cell* **78**, 250 (2020).
- [7] S. M. Mangiameli, J. A. Cass, H. Merrikh, and P. A. Wiggins, The bacterial replisome has factory-like localization, *Curr. Genet.* **64**, 1029 (2018).
- [8] A. Badrinarayanan, T. B. K. Le, and M. T. Laub, Bacterial chromosome organization and segregation, *Annu. Rev. Cell Dev. Biol.* **31**, 171 (2015).
- [9] C. L. Woldringh and N. Nanninga, Structural and physical aspects of bacterial chromosome segregation, *J. Struct. Biol.* **156**, 273 (2006).
- [10] S. Ben-Yehuda, D. Z. Rudner, and R. Losick, RacA, a bacterial protein that anchors chromosomes to the cell poles, *Science* **299**, 532 (2003).
- [11] C. Gogou, A. Japaridze, and C. Dekker, Mechanisms for chromosome segregation in bacteria, *Front. Microbiol.* **12**, 685687 (2021).
- [12] S. Jun and A. Wright, Entropy as the driver of chromosome segregation, *Nat. Rev. Microbiol.* **8**, 600 (2010).
- [13] J. J. B. Messelink, M. C. F. van Teeseling, J. Janssen, M. Thanbichler, and C. P. Broedersz, Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales, *Nat. Commun.* **12**, 1963 (2021).
- [14] P. H. Viollier, M. Thanbichler, P. T. McGrath, L. West, M. Meewan, H. H. McAdams, and L. Shapiro, Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication, *Proc. Natl. Acad. Sci. USA* **101**, 9257 (2004).
- [15] S. M. Mangiameli, B. T. Veit, H. Merrikh, and P. A. Wiggins, The replisomes remain spatially proximal throughout the cell cycle in bacteria, *PLoS Genet.* **13**, e1006582 (2017).
- [16] H. J. Nielsen, Y. Li, B. Youngren, F. G. Hansen, and S. Austin, Progressive segregation of the *Escherichia coli* chromosome, *Mol. Microbiol.* **61**, 383 (2006).
- [17] J. F. Marko, Linking topology of tethered polymer rings with applications to chromosome segregation and estimation of the knotting length, *Phys. Rev. E* **79**, 051905 (2009).
- [18] C. Helmstetter, S. Cooper, O. Pierucci, and E. Revelas, On the bacterial life sequence, *Cold Spring Harbor Symp. Quant. Biol.* **33**, 809 (1968).
- [19] A. Zaritsky, N. Vischer, and A. Rabinovitch, Changes of initiation mass and cell dimensions by the “eclipse”, *Mol. Microbiol.* **63**, 15 (2007).
- [20] A. Zaritsky, P. Wang, and N. O. E. Vischer, Instructive simulation of the bacterial cell division cycle, *Microbiology* **157**, 1876 (2011).
- [21] A. Zaritsky, W. Vollmer, J. Männik, and C. Liu, Does the nucleoid determine cell dimensions in *Escherichia coli*? *Front. Microbiol.* **10**, 1717 (2019).
- [22] H. Bremer and G. Churchward, An examination of the Cooper-Helmstetter theory of DNA replication in bacteria and its underlying assumptions, *J. Theor. Biol.* **69**, 645 (1977).
- [23] K. Skarstad, H. B. Steen, and E. Boye, *Escherichia coli* DNA distributions measured by flow cytometry and compared with theoretical computer simulations, *J. Bacteriol.* **163**, 661 (1985).
- [24] 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.* **28**, 71 (2014).
- [25] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.110.054401> for more information, which includes Refs. [12,24,39,54].

- [26] S. Jun and B. Mulder, Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome, *Proc. Natl. Acad. Sci. USA* **103**, 12388 (2006).
- [27] S. Jun, A. Arnold, and B.-Y. Ha, Confined space and effective interactions of multiple self-avoiding chains, *Phys. Rev. Lett.* **98**, 128303 (2007).
- [28] 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* **109**, E2649 (2012).
- [29] Y. Jung and B.-Y. Ha, Overlapping two self-avoiding polymers in a closed cylindrical pore: Implications for chromosome segregation in a bacterial cell, *Phys. Rev. E* **82**, 051926 (2010).
- [30] 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* **8**, 2095 (2012).
- [31] Y. Jung, J. Kim, S. Jun, and B.-Y. Ha, Intrachain ordering and segregation of polymers under confinement, *Macromolecules* **45**, 3256 (2012).
- [32] B.-Y. Ha and Y. Jung, Polymers under confinement: Single polymers, how they interact, and as model chromosomes, *Soft Matter* **11**, 2333 (2015).
- [33] V. S. Lioy, I. Junier, and F. Boccard, Multiscale dynamic structuring of bacterial chromosomes, *Annu. Rev. Microbiol.* **75**, 541 (2021).
- [34] J. A. Cass, N. J. Kuwada, B. Traxler, and P. A. Wiggins, *Escherichia coli* chromosomal loci segregate from midcell with universal dynamics, *Biophys. J.* **110**, 2597 (2016).
- [35] C. L. Woldringh, F. G. Hansen, N. O. E. Vischer, and T. Atlung, Segregation of chromosome arms in growing and non-growing *Escherichia coli* cells, *Front. Microbiol.* **6**, 448 (2015).
- [36] D. Mitra, S. Pande, and A. Chatterji, Polymer architecture orchestrates the segregation and spatial organization of replicating *E. coli* chromosomes in slow growth, *Soft Matter* **18**, 5615 (2022).
- [37] D. Mitra, S. Pande, and A. Chatterji, Topology-driven spatial organization of ring polymers under confinement, *Phys. Rev. E* **106**, 054502 (2022).
- [38] A. Hofmann and D. W. Heermann, The role of loops on the order of eukaryotes and prokaryotes, *FEBS Lett.* **589**, 2958 (2015).
- [39] J. D. Halverson, W. B. Lee, G. S. Grest, A. Y. Grosberg, and K. Kremer, Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. II. Dynamics, *J. Chem. Phys.* **134**, 204905 (2011).
- [40] A. Rosa and R. Everaers, Ring polymers in the melt state: The physics of crumpling, *Phys. Rev. Lett.* **112**, 118302 (2014).
- [41] I. Chubak, C. N. Likos, and S. A. Egorov, Multiscale approaches for confined ring-polymer solutions, *J. Phys. Chem. B* **125**, 4910 (2021).
- [42] J. D. Halverson, W. B. Lee, G. S. Grest, A. Y. Grosberg, and K. Kremer, Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. I. Statics, *J. Chem. Phys.* **134**, 204904 (2011).
- [43] S. M. Pachong, I. Chubak, K. Kremer, and J. Smrek, Melts of nonconcatenated rings in spherical confinement, *J. Chem. Phys.* **153**, 064903 (2020).
- [44] A. Narros, A. J. Moreno, and C. N. Likos, Influence of topology on effective potentials: Coarse-graining ring polymers, *Soft Matter* **6**, 2435 (2010).
- [45] A. Narros, C. N. Likos, A. J. Moreno, and B. Capone, Multi-blob coarse graining for ring-polymer solutions, *Soft Matter* **10**, 9601 (2014).
- [46] C. L. Woldringh, Compaction and segregation of DNA in *Escherichia coli*, *Life* **14**, 660 (2024).
- [47] Z. Fu, M. S. Guo, W. Zhou, and J. Xiao, Differential roles of positive and negative supercoiling in organizing the *E. coli* genome, *Nucleic Acids Res.* **52**, 724 (2024).
- [48] I. Junier, E. Ghobadpour, O. Espeli, and R. Everaers, DNA supercoiling in bacteria: State of play and challenges from a viewpoint of physics based modeling, *Front. Microbiol.* **14**, 1192831 (2023).
- [49] J. Harju, M. C. F. van Teeseling, and C. P. Broedersz, Loop-extruders alter bacterial chromosome topology to direct entropic forces for segregation, *Nat. Commun.* **15**, 4618 (2024).
- [50] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, (Oxford University Press, Oxford, UK, 2017).
- [51] K. Kikuchi, M. Yoshida, T. Maekawa, and H. Watanabe, Metropolis Monte Carlo method as a numerical technique to solve the Fokker-Planck equation, *Chem. Phys. Lett.* **185**, 335 (1991).
- [52] S. C. Glotzer, D. Stauffer, and N. Jan, Monte Carlo simulations of phase separation in chemically reactive binary mixtures, *Phys. Rev. Lett.* **72**, 4109 (1994).
- [53] A. Amitai and D. Holcman, Polymer physics of nuclear organization and function, *Phys. Rep.* **678**, 1 (2017).
- [54] A. Amitai, A. Seeber, S. M. Gasser, and D. Holcman, Visualization of chromatin decompaction and break site extrusion as predicted by statistical polymer modeling of single-locus trajectories, *Cell Reports* **18**, 1200 (2017).
- [55] V. S. Lioy, A. Cournac, M. Marbouty, S. Duigou, J. Mozziconacci, O. Espeli, F. Boccard, and R. Koszul, Multiscale structuring of the *E. coli* chromosome by nucleoid-associated and condensin proteins, *Cell* **172**, 771 (2018).
- [56] F. Wu, P. Swain, L. Kuijpers, X. Zheng, K. Felter, M. Guurink, J. Solari, S. Jun, T. S. Shimizu, D. Chaudhuri, B. Mulder, and C. Dekker, Cell boundary confinement sets the size and position of the *E. coli* chromosome, *Curr. Biol.* **29**, 2131 (2019).
- [57] M. Joyeux, Compaction of bacterial genomic DNA: Clarifying the concepts, *J. Phys.: Condens. Matter* **27**, 383001 (2015).
- [58] M. Joyeux, A segregative phase separation scenario of the formation of the bacterial nucleoid, *Soft Matter* **14**, 7368 (2018).
- [59] J. Kim, C. Jeon, H. Jeong, Y. Jung, and B.-Y. Ha, A polymer in a crowded and confined space: Effects of crowder size and polydispersity, *Soft Matter* **11**, 1877 (2015).
- [60] T. Odijk, Osmotic compaction of supercoiled DNA into a bacterial nucleoid, *Biophys. Chem.* **73**, 23 (1998).
- [61] D. Yang, J. Männik, S. T. Retterer, and J. Männik, The effects of polydisperse crowders on the compaction of the *Escherichia coli* nucleoid, *Mol. Microbiol.* **113**, 1022 (2020).

- [62] W. Wang, G.-W. Li, C. Chen, X. S. Xie, and X. Zhuang, Chromosome organization by a nucleoid-associated protein in live bacteria, [*Science* **333**, 1445 \(2011\)](#).
- [63] V. Norris, C. Kayser, G. Muskhelishvili, and Y. Konto-Ghiorgi, The roles of nucleoid-associated proteins and topoisomerases in chromosome structure, strand segregation, and the generation of phenotypic heterogeneity in bacteria, [*FEMS Microbiol. Rev.* **47**, fuac049 \(2023\)](#).
- [64] S. Kadam, K. Kumari, V. Manivannan, S. Dutta, M. K. Mitra, and R. Padinhateeri, Predicting scale-dependent chromatin polymer properties from systematic coarse-graining, [*Nat. Commun.* **14**, 4108 \(2023\)](#).
- [65] A. Thiel, M. Valens, I. Vallet-Gely, O. Espéli, and F. Boccard, Long-range chromosome organization in *E. coli*: A site-specific system isolates the *Ter* macrodomain, [*PLoS Genet.* **8**, e1002672 \(2012\)](#).