Supplementary Data for:

Features of Genomic Organization in a Nucleotide-Resolution Molecular Model of the *Escherichia coli* Chromosome

*William C. Hacker,† Shuxiang Li,† and Adrian H. Elcock**

Department of Biochemistry, University of Iowa, Iowa City, IA 52242

e-mail: adrian-elcock@uiowa.edu

Overview

1 Rationale for construction of the chromosome models

1A Selection of growth conditions

Because, as outlined in the main text, the architecture of our chromosomal models depends on the cell's transcriptional profile, and because we are interested in representing a single, unreplicated chromosome (see below), we have selected carefully the conditions in which the putative cell harboring our chromosome is cultured. Extensive ChIP-chip and ChIP-seq data are now available, at quite high resolution, for rapidly growing *E. coli* cells grown in LB medium (1-4). While it would certainly be of interest to construct models of chromosomes in rapidly growing conditions, cells grown under these conditions contain, on average, several genomic equivalents and a "typical" chromosome (if one can even be defined) is a template both for ongoing transcription and for replication. Our models, therefore, seek to represent instead the chromosomes of *E. coli* cells in slow growing conditions (i.e., more relevant to growth in minimal medium). The slower growth rate in minimal medium reduces the number of genomic equivalents per cell (5), and crucially, at sufficiently slow growth rates, cells will undergo a "rest" phase in which they include a single, unreplicated chromosome (6): it is this state of the cell that we attempt to model.

In anticipation of future studies, in which we eventually hope to explicitly model interactions between DNA and nucleoid-associated proteins (NAPs), the minimal medium model also has the advantage that ChIP data generated under these conditions are relatively complete. Notwithstanding the subsequent improvements in ChIP-chip coverage and the advent of ChIP-seq techniques (7), the data reported by the Busby group ten years ago (for minimal medium growth with a fructose carbon source), offer what is, to our knowledge, still

the only complete condition-matched set of data for the crucially important NAPs Fis, IHF, H-NS, and CRP (1,8).

1B Design of a transcription-centric chromosome model

The Church and Laub groups have previously shown how "5C" and "HiC" data – which identify regions of close physical contact between DNA – can be used to guide the construction of structural models of the 4 Mb chromosome of *C. crescentus* (9,10). While data obtained with another "all-by-all" sequencing technique (genome conformation capture – "GCC") have recently become available for *E. coli* (11), we have not used these data in the construction of our models for two reasons. First, the data are for rapidly growing cells in LB medium, and so are likely to arise from cells containing multiple (partially replicated) chromosomes and with transcriptional profiles different from those of more slowly growing cells. Second, the cell populations whose chromosomes are analyzed are not synchronized, complicating interpretation of the contact data and obscuring potential conformational changes that might depend on progress through the cell cycle (see section S1C below). Fortunately, a key finding from the Laub group's studies of the *C. crescentus* chromosome was that the *boundaries* of domains identified by DNA-DNA contacts consistently fall at regions of active transcription (including, for example, highly active rRNA operons) (10), and data from *E. coli* as well suggest that interaction boundaries depend on RNAP activity (11,12). These important results indicate that the cell's transcriptional profile can be used to situate these boundaries.

Our model, then, assumes that regions of high transcription delimit higher-order domains in the *E. coli* chromosome. Using well-established RNAP occupancy data reported by the Busby group (see above) to identify these regions, and adapting the "bottlebrush" model

previously devised by the Laub and Mirny groups (10), we divide the genome into plectoneme-abundant regions (PARs) and plectoneme-free regions (PFRs): the latter we assume correspond to highly-transcribed regions of the chromosome (10,13). In order to divide the chromosome in this way we proceed as follows.

The data provided by the Busby group consist of signal intensities at ~20,000 ChIP-chip probes, each 60 basepairs in length, uniformly scattered throughout the *E. coli* chromosome. Since the intensities at individual probes are subject to some noise, we first smoothed the data slightly by averaging the intensities from each probe and its first neighbor to each side. To identify "hits," we then specified a cutoff (a signal intensity ratio > 5.0) above which RNAP was assumed to be bound at the probe location. We then defined PFR boundaries by beginning with each identified hit and extending it to include all successive probe locations in either direction with intensity ratios > 1.5; when a boundary determined in this way fell within a defined transcriptional unit, the region was further extended to reach the end of the gene or operon (as defined by Ecogene (14) and RegulonDB (15), respectively). Under the simplifying assumption that plectonemes do not form in regions of length < 2000 bp, all regions less than this distance from one another were consolidated into single PFRs. In rough accord with previous studies assigning a minimum of ~3000 bp to PFRs (10) in order to preserve distinct interaction domains, all highly-transcribed regions were expanded – if necessary – to include a minimum of 2000 basepairs. Use of the specified cutoff coupled with this consolidation procedure produced a total of forty PFRs. This number is somewhat higher than the Laub group specified for *C. crescentus*, which may be a reflection of the greater number of genes active in slower growth and/or a consequence of the use of RNAP-binding data rather than expression data to identify

PFRs. In any event, the behavior of the model is not expected to differ dramatically as a function of its PFR count, and the principles of its construction are readily adaptable to data that suggest the number of PFRs should be larger or smaller. In total, only 159312 base-pairs (~3%) of the chromosome are included within PFRs; this coverage falls between the ~5% suggested to be in PFRs based on contacts maps and the ~2% of the genome allocated to PFRs in the models derived from these maps (10). The full set of PARs and PFRs is listed in Table S4.

1C Rationale for constructing alternative models based on the position of *oriC*

The partition of the chromosome into PARs and PFRs allows us to describe the topology of subgenomic regions of the chromosome, but in order to build structural models we also have to consider its global arrangement and positioning. Extensive efforts have been made to determine the spatial organization of the *E. coli* chromosome *in vivo* (16); fluorescent labeling systems have been particularly useful in determining the relative locations of the origin of replication (*oriC*), the left and right replichores, and the *ter* region that is essentially half a genome distant from *oriC* (and so appears opposite it when the chromosome is drawn as a circle). These same labeling techniques have established that, in *C. crescentus* swarmer cells, the origin is confined to one cell pole, while the terminus is located at the opposite end of the cell, with left and chromosomal arms extending between them along the length of the cell (9).

But the positioning of the analogous chromosomal elements in slow-growing *E. coli* cells is more ambiguous (17). Tracking of labeled *oriC*-proximate sites in cells undergoing a pre-replicative "rest" phase suggests that the origin in a daughter cell begins at one cell pole before migrating to a position at midcell in preparation for DNA replication: before the initiation of replication, then, the cell is organized such that the left and right replichores run in parallel

along the long axis of the cell (6,16). Other studies, however, suggest that the localization of *oriC* to midcell prevails even before the onset of replication, with left and right replichores extending in either direction from the cell center to the cell poles and loci within them maintaining their position with little variance; a loosely-packed region surrounding *ter* traverses the length of the cell to connect the left and right replichores (17-19).

Because the relative positions of *oriC* and *ter* in the chromosomal structure appear to be acutely sensitive to differences in model strains and growth rates (20), and since both conformations appear to be physiologically relevant, we have designed two alternative sets of models to reflect these different locations of *oriC* (Figures 3B-C of the main text). In the first type of model (*oriC*@pole), *oriC* is located at the center of the left-hand edge of the initial "blueprint" structure (Figure 3B; see below), in a position corresponding to the cell pole in the final condensed volume of the nucleoid. The compaction procedure that we use to drive the chromosome structure into the nucleoid region of a typical cell (see section S2A below), does not change this global arrangement, so *oriC* and *ter* remain at opposite poles of the nucleoid in the final chromosome models, bridged in parallel by the left and right chromosomal arms.

In the second type of model (*oriC*@midcell), *oriC* is instead located at the center of the top edge of the initial "blueprint" structure, with the left and right replichores now extending to opposite cell poles (Figure 3C). As has been pointed out in the literature, an arrangement in which the two replichores extend in opposite directions necessitates the presence of a "crossing" region (18) that extends across the nucleoid to complete the chromosomal "circuit." The *oriC*@midcell model, therefore, allocates ~100,000 plectoneme-free base-pairs to this crossing region, with ~60% of the genome linearly arranged on either side of *oriC* and with the

remainder at the poles. The size of the crossing region in our models (100 kb) lies within the range of previous estimates, which include ~50 kb (21), 80 kb (22), and ~400 ± 200 kb (18). The conservative sizing of the crossing region reflects, in part, uncertainty about the nature of the proposed crossing "fiber," the packaging density of which is likely to be intermediate between that of PFRs and PARs. In the absence of additional data describing this lower-density fiber, the inclusion of a crossing region in our model necessitates unspooling two of the PARs, creating regions of lower RNAP density but with packaging density characteristic of PFRs. Again, the compaction procedure that we use to drive the chromosome structure into the nucleoid region does not change this overall arrangement.

Since the ChIP-chip data defining both the RNAP density and the distribution of PFRs and PARs cannot distinguish between the *oriC*@pole and *oriC*@midcell models (and may well reflect a mixed population of them), we have attempted to maintain consistent PFR/PAR boundaries in both models within the limits imposed by the presence of the crossing region in the *oriC*@midcell variant.

1D Generation of populations with varied plectoneme lengths and branching patterns

While the ChIP-chip data enable us to define PARs and PFRs, they say nothing about the physical characteristics of the plectonemes within the PARs. In particular, in order to build structural models, we need to know both the expected length distributions of plectonemes and the extent to which they are likely to be branched. Data for both come from the Cozzarelli group, who have shown: (a) that plectoneme lengths in the *E. coli* chromosome are distributed according to an exponential distribution with a mean of ~10 kb (23), and (b) that supercoiled

plasmids have branching patterns that are predictable on the basis of their lengths (24). We use these two sets of data in the construction of our models as follows.

Since the total numbers of basepairs allocated to PARs are 4482340 and 4351586 for the *oriC*@pole and *oriC*@midcell models, respectively (see above), and since the mean length of plectonemes is 10 kb (23), the total number of plectonemes to include is fixed at 448 and 436 in the *oriC*@pole and *oriC*@midcell models, respectively. Within those constraints, however, it is desirable to produce a range of structures that differ in the arrangements of their plectonemes. For this reason, we made 20 independent structures for both the *oriC*@pole and *oriC*@midcell models, each differing in the positions, lengths and branching patterns of the plectonemes assigned to each PAR.

To achieve a reasonable distribution of plectoneme lengths, we first randomly assigned lengths to each of the plectonemes within each PAR and then adjusted them iteratively – subject to the condition that no plectoneme could contain fewer than 1000 basepairs – until the lengths of the entire population of plectonemes matched the expected exponential distribution identified by the Cozzarelli group (23). One benefit of applying this randomized assignment procedure to each of the 20 structures representing each global orientation (*oriC*@pole or *oriC*@midcell), is that it allows genomic loci to sample very different local environments even while remaining part of a larger domain (i.e., PAR) whose size and position is maintained across the structures. For example, a given locus may be on a branch off a long (~40 kb) plectoneme in one structure, but on the trunk of a short (~5 kb) plectoneme in another model. Alternatively, it may be on a plectoneme at a PFR/PAR boundary in one structure, but several plectonemes distant from the boundary in another.

The assignment of branches to the plectonemes was achieved as follows. Analysis of electron micrographs by the Cozzarelli group has shown that a negatively supercoiled 3.5 kb plasmid has, on average, 1.63 ± 0.9 branches while a similarly supercoiled 7 kb plasmid has 2.94 ± 1.2 (24). These numbers translate into "branching lengths" of 2147 and 2381 basepairs per branch for the 3.5 and 7 kb plasmids, respectively. In order to obtain similar estimates for plectonemes of arbitrary lengths, we fit these two data points to a straight line, and obtained the following relationship:

$$\text{branching length} = 2147 + (\text{nbp} - 3500) \times 0.066775 \quad (1)$$

Here nbp is the number of basepairs in the plasmid or plectoneme. We then used this relationship to compute the expected number of branches for the ~450 plectonemes in each of the 20 structures generated for the two types of model. For reasons outlined in section S2A, we applied the additional condition that the total number of branches in each plectoneme must be an even number. This can be achieved, in such a way that maintains the correct average, by rounding the value calculated with Eq. 1 to one of the two even integers either side of the value with a probability weighted by the difference between the calculated value and the lower integer. Figure 4B verifies that the number of branches assigned to the plectonemes of the 20 *oriC*@pole structures using this approach (red) matches the numbers expected, on the basis of the plectoneme lengths, using Eq. 1 (blue).

An additional feature of branching revealed by the study of Boles et al. (24) is the distribution of the fractional branch lengths within a 7 kb plasmid. To parameterize an approach that reproduces this distribution, we first randomly assigned branches within a single

7 kb plectoneme such that they matched, on average, the number determined experimentally for the same-sized plasmid (see above). We then randomly positioned these branches within the plectoneme subject to the following two restrictions: (a) the minimum possible length of a branch is M_{min} bp, and (b) the minimum possible length of the trunk between two branches is N_{min} bp. As shown in Figure 4C, we achieved good agreement with the experimental fractional branch length observed by Boles et al. (24) using values of M_{min} = 350 bp and N_{min} = 200 bp. These parameters, together with the additional criterion (relevant only to very large plectonemes) that the maximum possible length of a plectoneme trunk between branches is 10,000 bp, were then used to determine the branching patterns in all plectonemes.

1E Note on the *E. coli* genome used for model generation and gene definitions

All of the models described here represent the 4641652 basepairs of the *E. coli* K12 MG1655 (U00096.3) genome. When experimental data refer to older genome builds, we have updated coordinates for comparison with our models using the Ecogene Gene Interval Updater (14).

1F Note on the role of nucleoid-associated proteins (NAPs)

The critical role played by nucleoid-associated proteins (NAPs) in controlling the structure, activity, and dynamics of the bacterial chromosome is well established (25-27). But as is clear from previous studies, some of the structuring effects of proteins like HU can be modeled approximately without their explicit representation (10). Since the models that we have constructed reflect a variety of experimental data on the behavior of the *E. coli* chromosome *in vivo*, many of the roles played by NAPs in contributing to the structural organization of the chromosome should be captured implicitly by the models: partitioning of the chromosome into

~450 plectonemic domains, for example, implicitly reflects the roles played by H-NS and Fis (28). In passing we note that, for similar reasons, the possibly significant effects of membrane protein transertion on nucleoid structure (29) should also be indirectly accounted for in the models. Given the considerable computational expense involved in producing the models described here, we have chosen to leave explicit modeling of the role of NAPs to the future. But future work will clearly need to address the roles played not only by the more generally acting NAPs such as HU, Fis, IHF, and H-NS, but also by proteins, such as MatP, whose effect on chromosome structure depends on a smaller set of high-affinity interactions (30). We note, however, that we will not be required to model the two separate foci reported for H-NS within the *E. coli* chromosome since this finding (31) was later shown to be an artifact (32) by the same group who performed the original work.

2 Details of the molecular simulations used to construct the chromosome models

2A Generation of low-resolution (500 BPB) chromosome structures

As briefly outlined in the main text, the protocol used to construct the chromosome models involved the use of multi-resolution simulation techniques. With plectoneme lengths, positions, and branches all specified for a given structure (see section S1D above), an initial, *atomic* model was constructed using in-house modeling code that assembles each chromosome, one basepair at a time, with the twist angles at basepair steps in the PARs reduced slightly from their standard B-DNA values of 34.3° in order to produce an overall supercoiling density, σ, of -0.05. The σ value chosen corresponds to estimates based on chromosomal psoralen binding (33) and falls between estimates of ~-0.025 (34) and ~-0.07 (35). In order that this initial "blueprint" model

be a fully connected circular chromosome, with no structural discontinuity where the "tail" of the chromosome meets the "head," two minor geometric constraints were introduced to simplify model construction. These are that: (a) the number of plectonemes along the left- and right-hand sides of the models must be equal (see Figures 3B-C), as must the number of plectonemes along the top- and bottom-edges of the model, and (b) the number of branches off any plectoneme trunk must be an even number, with equal numbers of branches emerging from the outward- and inward-directed arms of the plectoneme. With those geometric constraints imposed, the initial model thus obtained represents a highly idealized view of the chromosome, in which uniformly straight regions of non-plectonemic B-DNA bend smoothly upon entry to plectonemic regions, which themselves form uniformly twisted plectonemes with the right-handed super-helical arrangement expected of negatively-supercoiled DNA. The pitch of each plectoneme super-helix was set to 95.5 Å/rad (such that one complete turn required 600 Å) and the radius of each super-helix was set to 42 Å; these numbers are in rough agreement with theoretical predictions for a plectoneme with σ = -0.05 made by Marko and Siggia (36). The overall arrangement of the resulting model is largely confined to a plane, but the plectonemic branches that emerge smoothly from the main trunks of plectonemes do so in a way that means they extend above and below the plane.

In principle, it would be possible to take such an idealized atomic model of the chromosome, and subject it to a simulation protocol that seeks to confine it within progressively smaller volumes until it reaches the physical dimensions of the nucleoid (see below): such an approach would have the advantage of immediately yielding a fully connected, atomic model of the bacterial chromosome as it might look *in vivo*. In practice, however, the extraordinarily

large distances over which atoms would be required to move during compaction (thousands of Ångstroms), coupled with the computational expense associated with simulating many millions of atoms, make this approach currently infeasible.

Instead, therefore, we use each atomic chromosome model to generate three independent coarse-grained (CG) models, each with a different level of resolution, and carry out the compaction process using these cheaper alternatives. The CG models used range from a 1 nucleotide-per-bead (1 NTB) model at the finest level of detail, through a 5 basepairs-per-bead (5 BPB) model at an intermediate level of detail, to a ~500 basepairs-per-bead (500 BPB) model at the coarsest level of detail. Importantly, both of the CG model types that have been used in dynamics simulations – i.e. the 5 BPB and 500 BPB models – have been explicitly parameterized to reproduce physical properties of the DNA polymer (see section S2F below).

The most computationally expensive stage of the procedure involves the confinement of the 500 BPB model, whose initial length is nearly sixty times that of the cell in which it will be finally enclosed, to successively smaller capsule-shaped volumes over the course of a series of Brownian dynamics (BD) simulations. Although cell size estimates vary significantly depending on conditions (37), a "typical" *E. coli* cell might be considered as being composed of a cylindrical region 1.6 μm in length with a radius of 0.4 μm, capped at either end by a hemisphere of the same radius, such that its total length is 2.4 μm. Within that, the chromosome typically occupies a still-smaller region of the cell, which here is defined such that the hemisphere-capped cylinder has a total length of 1.78 μm and a radius of 0.395 μm; these values reflect the dimensions of the nucleoid under slow growth conditions determined by immunofluorescence (38-41). We note that in the absence of other cellular components we have had to *enforce* confinement of the

chromosome within the nucleoid; previous simulation work from the Yethiraj and Weisshaar groups, however, suggests that explicit inclusion of ribosomes and polysomes – which could be added to the current models in the future – can be sufficient to cause formation of a distinct nucleoid region even in the absence of enforced confinement (5).

BD simulations of the compaction process were carried out with in-house simulation code (*uiowa_bd*) that employs standard molecular mechanics terms to describe the bonded energy contributions of CG beads to the total energy of the chromosome:

$$E_{bonded} = \Sigma_{bonded} K_r(r - r_0)^2 + \Sigma_{angles} K_\theta(\theta - \theta_0)^2 \quad (2)$$

Here r and θ represent inter-bead bond lengths and angles, respectively, and r_0 and θ_0 represent the equilibrium values of the bond lengths and angles in the initial (idealized) chromosomal build. For the 500 BPB model, CG beads are placed at approximately 500 basepair intervals in plectonemic regions, with the positions of any beads that fall *near* the tips and bases of all plectonemes (including branches) being adjusted so that they are more precisely positioned *at* the tips and bases. For non-plectonemic regions – which it will be recalled account for only 3% of the basepairs in the model (see section S1B above) – CG beads are instead placed at approximately 250 basepair intervals; this ensures that the bonds connecting beads in both the PARs and PFRs are similar in length. While the equilibrium r_0 value varies somewhat throughout the chromosome, it averages ~846 Å for bonds between beads within plectonemes and ~856 Å for bonds between beads in PFRs. The equilibrium θ_0 value, on the other hand, is almost always 180° except in those cases where: (a) a bond is formed between a bead ending a PFR and a bead starting a PAR, or (b) an angle includes bonds between beads in both the trunk

and branch of a plectoneme. The force constant opposing changes in the bond length, K_r, was set to a value of 0.0005 kcal/mol/Å^2: while this might seem small, it equates to a force constant of ~160 kcal/mol/Å^2 for a (chemical) bond of equilibrium length 1.5 Å. The force constant opposing changes in the bond angle, K_θ, having been parameterized to match the persistence length of DNA (see section S2F below), was set to 0.14 and 0.37 kcal/mol/rad for bonds involving non-plectonemic and plectonemic beads, respectively.

BD simulations with the 500 BPB model were performed with the inclusion of hydrodynamic interactions using the Ermak-McCammon algorithm (42) implemented in *uiowa_bd*:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Sigma_j \mathbf{D}_{ij}\mathbf{F}_j\Delta t/k_BT + \mathbf{R}_i \quad (3)$$

Here $\mathbf{r}_i$ is the position vector of bead i, Δt is the simulation timestep, $\mathbf{D}_{ij}$ is the i-jth 3 × 3 submatrix of the diffusion tensor $\mathbf{D}$, $\mathbf{F}_j$ is the total force acting on the jth bead, and $\mathbf{R}_i$ is a random displacement applied to bead i; k_B is Boltzmann's constant and T is the temperature in Kelvin. During compaction simulations, the only forces acting on beads were those due to the bonded interactions (Eq. 2) and those acting to restrain beads within the ever-shrinking cell capsule: the latter were modeled with a one-sided harmonic potential function with a force constant initially set to 2.5 × 10^{-5} kcal/mol/Å^2 but set to 0.005 kcal/mol/Å^2 in the final stages of the protocol to ensure that all beads were safely confined to the nucleoid region. No nonbonded interactions (either steric or electrostatic) between beads were included – at this stage – since: (a) at this level of resolution, spherical beads cannot meaningfully model the shape of the ~500 basepair DNA that they are supposed to represent, and (b) it was desirable to allow

plectonemes and other regions of non-plectonemic DNA to freely pass through each other in order to crudely mimic the effects of the topoisomerases that would otherwise relieve conformational strain (43) that might accompany compaction within the cell.

While nonbonded energetic interactions between beads were not included, pairwise hydrodynamic interactions between beads were included, at the Rotne-Prager-Yamakawa level of theory (44,45). The inclusion of hydrodynamic interactions in the compaction simulations has the advantage that the correlated motions they induce between neighboring beads act to dissipate any conformational strain that builds up during compaction. The effective hydrodynamic radius assigned to each bead was 2000 Å, which, while much larger than is realistic for a 500 basepair bead, both ensures a strong hydrodynamic communication between the beads and allows a comparatively long timestep to be used in the simulations. In any case, it should be remembered that hydrodynamic interactions affect only the rates at which beads diffuse in the simulations (46), and do not play a role similar to energetic interactions in controlling the conformational energetics of the chromosome model. The timestep during the majority of stages of the compaction process was 20 ps, but this was reduced to 10 ps for later stages when the capsule restraint force constant was increased (see above).

The total number of simulation time steps allocated for the compaction of the chromosome was 2.65 million (corresponding to 51.5 μs). We note that since the compaction process has no direct biological counterpart, the timescale of the simulations is a nominal one that attempts to strike a balance between, on the one hand, allowing the DNA to "keep up" and remain relatively unstrained as the cell-shaped capsule in which it resides shrinks, and, on the other hand, allowing the simulations to be completed sufficiently quickly that multiple

chromosome models can be constructed using the allocated computational resources (recall that a total of 20 complete structures were built for both the *oriC*@pole and *oriC*@midcell models). Compaction of each chromosome structure required ~130 hours on a 64-core AMD Opteron 6272 server. It is worth noting that, while (faster) exploratory simulations that excluded hydrodynamic interactions between the DNA beads were also carried out, the DNA was unable to rapidly adjust during compaction and instead, became pressed up against the sides of the ever-shrinking capsule.

2B Optimization to reproduce spatial distributions of RNAP and DNA

The 20 structures obtained for each type of model (*oriC*@pole and *oriC*@midcell) after the initial periods of compaction represent plausible, but preliminary, low-resolution models of the bacterial chromosome. As outlined in the main text, a crucial next step in the protocol is to adjust both types of model so that their structures, as a population, reproduce the experimental distributions of: (a) transcribing RNA polymerases (RNAP), and (b) DNA, as obtained from recent single-molecule fluorescence experiments (40). Such experiments record projections of 3D objects onto a 2D plane, which are accumulated and then expressed in the form of probability distributions along the short axis of the cell. Represented in this way, however, it is difficult to use these experimental distributions to directly guide the simulation models into better agreement with experiment.

To solve this problem, we first attempted to determine the *radial* distributions of RNAP and DNA that would explain the observed *short axis* distributions when projected onto the 2D plane. This was achieved by systematically optimizing the statistical weights of points evenly distributed within a cylindrical volume equivalent to the central 10000 Å occupied by the

chromosome. Ten million points were first randomly assigned positions within the cylinder by sampling from a uniform distribution and assigned a weight of one. All points were then Boltzmann-reweighted by an energy, E, determined by applying a one-sided harmonic potential function to each point's radial distance from the cylindrical axis, where:

$$E^{RNAP} = \tfrac{1}{2} K_{rad}^{RNAP} (c - c_0^{RNAP})^2 \qquad E^{DNA} = \tfrac{1}{2} K_{rad}^{DNA} (c - c_0^{DNA})^2 \qquad (4)$$

Here, the K_{rad} entries are the force constants acting in the radial direction, c is the radial distance of the point, and the c_0 entries are the radial distances at which the reweighting energy is zero; because the potential is one-sided, the reweighting energy of all points beyond c_0 is also zero. The total populations within each 50 Å shell were determined by summing the weights of all the points residing within each shell. These populations were then projected onto the 2D plane, analyzed via the same procedure used in the experimental work, and compared with the symmetrized experimental distributions.

The energy functions used to reweight the points (Eq. 4) are sufficiently simple that their parameters (K_{rad} and c_0) can be optimized by performing a grid search in both dimensions. Performing such a search to maximize (separately) the reproduction of the experimental RNAP distributions and the DNA distributions, yields the following optimized values: $K_{rad}^{RNAP} = 1 \times 10^{-6}$ kcal/mol/ Å^2, $c_0^{RNAP} = 2750$ Å, $K_{rad}^{DNA} = 2 \times 10^{-7}$ kcal/mol/ Å^2, and $c_0^{DNA} = 900$ Å. The exact values assigned to these parameters are less important than the fact that, when used to reweight points uniformly distributed in space, they generate projected, 2D short axis distributions that closely follow the experimental distributions for both RNAP and DNA. It is, however, interesting to note that the reweighted *radial* distribution of DNA is not uniform, reaching peak

density at about 1000 Å from the nucleoid axis and declining to densities below average nucleoid density beyond 3000 Å, while that of bound RNAP is densest at ~2800 Å (Figure S15). It is these final, reweighted populations of points within each radial shell that represent the "target" populations that we next attempted to make the 500 BPB models of the chromosome reproduce.

With target radial populations defined for both RNAP and DNA, we performed a second round of BD simulations aiming to adjust the CG chromosome models to reproduce them. This was carried out in the following way. CG beads of the 500 BPB models were first assigned one of two identities: either as a RNAP-containing bead or a DNA-only bead. In accord with an estimated total RNAP copy number per cell of 2500 (47), and with single-molecule tracking data indicating that ~48% of this population is specifically bound to DNA (40,48,49), we determined that 1200 RNAPs would be subject to the density-fitting while the remaining ~1300 would be considered diffusely distributed throughout the cell and therefore not modeled.

Likely chromosomal locations at which these RNAPs would be bound were identified based on the ChIP-chip data from the Busby group (8) that we used earlier to define the positions of PFRs (see section S1B above). We selected the centers of the 1200 non-overlapping probes with the highest signal-intensity ratios from their data set, and used these as initial positions for placing the RNAPs. We then made minor adjustments based on the use of additional data describing the chromosomal locations of promoters and genes as follows. When a probe's range of coverage (200 bp) was found to encompass a σ^{70} promoter, as defined in RegulonDB (15), the RNAP associated with that probe was moved to the promoter. When, on

the other hand, a probe center was located in an intergenic region but the probe's coverage included a defined gene (14), the RNAP was resituated so as to be located within the gene body with minimal displacement from its original location. Finally, when a probe center was located in an intergenic region and no promoters or genes were found nearby, the RNAP was assumed to remain at the location initially defined by the probe center. After these refinements, final positions were adjusted when necessary to avoid clashes between pairs of RNAPs: this was carried by allowing only a single RNAP to occupy a promoter and assuming a "footprint" of 36 basepairs for each RNAP within the gene body (50).

We next determined the number of RNAP molecules associated with each CG bead in each of the 500 BPB chromosome structures by summing all of the RNAPs assigned to the set of basepairs associated with that bead. Beads with at least one RNAP assigned to them are colored red in the representative structures pictured in Figures 3B-C of the main text. Having determined the number of RNAPs associated with each CG bead, and already knowing the number of DNA basepairs associated with each CG bead, the initial radial populations of RNAP and DNA in all 20 structures of each type of model were computed and compared with the target radial populations. Using an approach identical to that used in the Iterative Boltzmann Inversion method, which is commonly used to parameterize CG force fields for macromolecular simulations (51,52), the differences between the current and target radial populations were then converted into biasing energy functions acting in the radial direction using:

$$E_0(r) = RT \ln [P(r)^{model} / P(r)^{target}] \quad (5)$$

Here $E_0(r)$ is the energy associated with each radial bin (of width 50 Å), $P(r)^{model}$ is the fractional population of CG beads in this radial bin in the 20 initial CG structures of the chromosome, and $P(r)^{target}$ is the corresponding target fractional population. Eq. 5 was used to define two separate biasing energy functions: one used for matching the RNAP distribution, and one used for matching the DNA distribution; while the basepairs assigned to RNAP-containing beads were included in bin populations used to calculate the DNA energy function, in simulations these beads were subject only to the biasing functions aimed at matching RNAP distributions. A 100-step steepest descent energy minimization of all 20 representatives of each model was performed with the two biasing functions applied, and the resulting radial populations of RNAP and DNA beads were recomputed and again compared with the target radial populations. The biasing functions were then updated to reflect the new populations using:

$$E_{i+1}(r) = E_i(r) + \alpha\ RT\ \ln [\ P(r)^{model}\ /\ P(r)^{target}\ ] \qquad (6)$$

Here the subscripts i and i+1 indicate the iteration numbers, and α is a scaling factor (<1) that acts to suppress potentially large fluctuations in the biasing functions that might occur during the iterative process (53). Having updated the biasing functions, a further energy-minimization was performed, the radial populations were recomputed, and Eq. 6 was again used to update the biasing functions. This procedure was carried out for 5000 iterations, by which time the short-axis distributions of both RNAP and DNA were found to be in excellent agreement with the experimental distributions (see Figures 5A-C). Since the use of repeated energy minimizations induced a modest, but noticeable change in bond angle distributions, after 5000

iterations, a final, short BD simulation of 50 ns duration was used to restore the initial distributions without affecting the fits to the experimental distributions.

We note that while the reported experimental data only include short axis distributions for the central cylindrical portion of the nucleoid (40), we also applied analogous potential functions to the hemispherical caps at either end of the capsule representing the final nucleoid volume. We did this because it appears likely that an effect similar to that which drives RNAP-bound DNA outwards in the radial direction (40) might also drive RNAP-bound DNA to the poles of the nucleoid. Reflecting the greater uncertainty associated with this idea, however, the biasing potential functions applied along the long axis were arbitrarily scaled down in magnitude by a factor of 2.

2C Final adjustment of 500 BPB chromosome structures

While the chromosome structures obtained by this stage remain low-resolution models, they now explicitly reproduce the experimental distributions of RNAP and DNA in the nucleoid of *E. coli*. Prior to using these structures to guide the construction of increasing-resolution structures (see section S2D below), we performed two additional stages of simulation to finalize the 500 BPB structures. First, we took steps took to correct the fact that the bond lengths connecting beads in the structures are, at this stage, unrealistically long since they were derived from the highly idealized "blueprint" model. In reality, DNA on the lengthscale of the 500 BPB model's typical bond lengths (~850 Å; see section S2A above) will exhibit significant flexibility, so the "true" bond length to apply should be considerably shorter than the idealized value. To account for this, we changed the equilibrium bond lengths, over the course of five 10,000 step steepest-descent energy-minimizations, from their initial, idealized values to the quadratic

mean of lengths predicted for a Worm-Like Chain with a persistence length, l_P, of 50 nm or 79 nm depending on whether the bond represents a pair of PFR or PAR beads, respectively. In order not to adversely affect the already excellent fits to RNAP and DNA distributions, these changes in equilibrium bond lengths were applied while CG beads representing the tips or bases of plectonemes (and their branches) were harmonically restrained to their initial (fit) positions.

Second, since at this stage the structures look unrealistically angular, we attempted to smooth out the path followed by the beads in a manner again consistent with the DNA behaving like a Worm-Like Chain. To this end, additional CG beads were added halfway along each bond connecting 500 BPB beads (Figures S16A-B). The newly added beads were displaced in a randomly selected direction normal to the original bond axis, such that the "new" bonds connecting them to the original beads were again of lengths consistent with end-to-end distances predicted by a Worm-Like Chain model of appropriate persistence length; these distances were determined using Eq. 3 of Zhou (54). If the sum of the two "new" bonds exceeded the contour length of the original bond, a different random displacement was selected. The result of this bead-adding procedure is a noticeable smoothing of the path described by the CG beads and a reduction in the local angularity of the model consistent with the persistence length of the final higher-resolution structures (Figure S16C).

2D Generation of medium-resolution (5 BPB) chromosome structures

The finalized 500 BPB structures obtained using the simulations outlined in sections S2A-S2C were used to construct higher resolution, 5 BPB structures in the following stages. In the first stage, a series of "morphing" simulations was performed with the aim of producing 5 BPB

structures that match the 500 BPB structures produced from the simulations described in sections S2A and S2B. To this end, 1330 "snapshots" were extracted from the above simulations and concatenated to form trajectories that were used as guides for the "morphing" simulations. Morphing was carried out using in-house code that: (a) applies harmonic springs (force constant of 2.5×10^{-5} kcal/mol/Å^2) to the 5 BPB beads so that they follow the pre-computed trajectory of the 500 BPB beads as closely as possible, and (b) applies additional springs (force constant of 8.4×10^{-6} kcal/mol/Å^2) between all pairs of 5 BPB beads that are separated by < 100 Å in the initial structure so that the models retain their branched-plectonemic "shapes" during the sometimes drastic changes that accompany compaction within the nucleoid. Since the structure of the chromosome is maintained by springs only – with no additional potential functions applied – the model used at this stage is essentially an elastic network model (ENM, (55)). Each of the 1330 stages of the morphing simulation involved a total of 400 steps of steepest-descent energy-minimization: over the first 300 steps, the 500 BPB beads moved with constant velocity from their positions in the "initial" snapshot to their positions in the "final" snapshot while the 5 BPB beads endeavored to "keep up." Over the next 50 steps the 5 BPB beads were allowed to continue relaxation to the positions dictated by the 500 BPB beads, and over the last 50 steps, the springs connecting the 5 BPB beads to the 500 BPB beads were removed to allow further structural relaxation.

The second stage involved making the 5 BPB structures emulate the changes that occurred to the 500 BPB structures during the simulations described in section S2C. To this end, the 5 BPB structures obtained after the morphing simulations were first subjected to 20 ns of BD simulation, during which Lennard-Jones (Gō) potential functions were gradually introduced

between only those pairs of 5 BPB beads that were: (a) part of the same plectoneme and (b) were separated by < 85 Å in the initial "blueprint" structures. The well-depth of these potential functions was set to 1 kcal/mol; they play the same role as the elastic network model used in the earlier morphing simulations, in that they act to retain the shape of the plectonemes during subsequent simulations. Next, an additional 40 ns of BD simulation was performed during which the Lennard-Jones functions were retained but during which position restraints were applied so as to induce, in selected 5 BPB beads, the same series of displacements undergone by their "parent" 500 BPB beads in the simulations described in section S2C. The force constant used in these position restraints was 2 kcal/mol/Å^2, which was sufficient, when maintained during a further BD simulation of duration 250 ns, to cause the 5 BPB model to adopt a conformation essentially identical to that originally assumed by its 500 BPB "parent" at the positions defined by the 500 BPB beads, while being structurally relaxed at all intervening points (Figure S17).

Beyond this stage, the 500 BPB structures had no further role to play, and the goal was to add increasingly realistic features into the 5 BPB structures, without allowing them to deviate too far from their initial positions: since the deviations that follow were not drastic, no potential functions were included to explicitly restrain the shapes of plectonemes. First, a 1000 step steepest-descent energy-minimization was performed so that the lengths of the bonds connecting neighboring 5 BPB beads reached the values expected of B-DNA (i.e. 17 Å). Next, over the course of 25 consecutive 100 step steepest-descent energy-minimizations, steric interactions (modeled using conventional $1/r^{12}$ terms) were introduced between all pairs of 5 BPB beads that were not involved in bonding interactions. This is a crucial stage since it

represents the first time that steric effects have been introduced during the entire protocol used to generate the chromosome structures: without this stage, any claim to have generated truly realistic chromosome structures would be critically compromised. During these simulations, the σ value describing the diameters of the 5 BPB beads was increased from 0 to 25 Å in 1 Å intervals; this final value is larger than is strictly necessary (the diameter of double helical DNA is ~20 Å) but it ensures that there is no danger of steric clashes occurring at any subsequent stage of simulation.

We next introduced a more sophisticated energy model (see below) and used it to perform a 250 ns BD simulation of the 5 BPB structure in order to allow modest conformational fluctuations to occur, before performing a final 1000 step steepest-descent energy-minimization. The energy model used in these final two simulations contains both electrostatic and steric interaction terms as well as bond stretch and bond angle terms. As before, therefore, we performed a series of BD simulations to parameterize the angle bending force constant, K_θ (see section S2F below). The total charge placed on each 5 BPB bead was -5 e, which in being reduced from the "true" value of -10 e, is intended to crudely mimic the condensation of positive ions around the double-helical DNA (56). The electrostatic interaction, E_{elec}, between beads not involved in bonding interactions with each other was then modeled using the Debye-Hückel approximation:

$$E_{elec} = 332.08\, q_i\, q_j \exp(-\kappa / r_{ij}) / \varepsilon\, r_{ij} \quad (7)$$

Here q_i and q_j are the charges on the beads, κ is the Debye-Hückel screening parameter (related to the square root of the ionic strength), r_{ij} is the distance between the beads (in Ångstroms), ε is

the dielectric constant of the medium (approximated here as that of water – i.e. 78.4), and 332.08 is the conversion factor that ensures the energy is returned in units of kcal/mol.

2E Generation of high-resolution (1 NTB) chromosome structures

Once generated, the final 5 BPB chromosome structures were converted into final structures with a resolution of 1 nucleotide-per-bead (NTB) as follows. Four intermediate beads were first added to the 5 BPB structure by placing them at uniform intervals between each pair of bonded beads, before a simple 1000 step steepest-descent energy-minimization was performed to smooth out the path of the double-helical axis. During this energy-minimization, the parameters used for bonded interactions modeled with equation 2 were: r_0 = 3.4 Å, K_r = 20 kcal/mol/Å^2, θ_0 = 180°, K_θ = 2.5 kcal/mol/rad; owing to the extremely modest change to the structure that occurs during this stage no nonbonded interactions were found to be necessary. This procedure yields a 1 BPB structure in which the position of each bead represents the double-helical axis of the DNA.

The 1 NTB structure – in which each CG bead is placed at the position expected of the backbone phosphorous atom – was then obtained through a series of structural superpositions of DNA fragments, taken from the initial, "blueprint" structure, onto beads of the 1 BPB structure. The phosphates of the first 3 bp were placed by superposing their 3 helical axis beads onto the corresponding 3 beads of the 1 BPB structure; given that these 3 beads are essentially co-linear in both structures, the rotational orientation of the first 3 bp fragment around the helical axis is effectively arbitrary. From this point on, however, all additional phosphate beads are placed in a way that continues the initially defined pitch of the helical axis. Specifically, to place the phosphate beads of a "new" basepair, i+1, we simultaneously superposed the helical

axis beads of basepairs i, i+1, and the two phosphate beads of basepair i (i.e. 4 beads in total), and then applied the same coordinate transformation to the phosphate beads of basepair i+1. In order to ensure that each new basepair added to the growing 1 NTB model stayed close to the helical axis that was so laboriously defined by all the earlier simulations, the helical axis beads were assigned a weight 3 times greater than the weight attached to the phosphate beads of basepair i in the above superposition. Having placed the phosphate beads of basepair i+1, the superposition process could be repeated using helical axis beads from basepairs i+1 and i+2 in order to place the phosphate beads of basepair i+2. Continued in this way, all 4641652 pairs of phosphate beads could be placed in approximately 40 hours.

As the very final stage of the entire process, a 10,000 step energy minimization was performed to remove any slight discontinuity that might be present where the 4641652nd basepair met the 1st basepair, to remove any slight steric clashes, and to ensure that the major and minor grooves were properly formed. In this energy minimization, a so-called Gō or "native-centric" energy function, similar to that widely used to simulate protein folding events (57-59), was applied in order to encourage the structure to take on the appearance of B-DNA, while remaining subject to the local environmental constraints acting on the DNA. For this purpose, we made use of the same energy function used in our previous simulations of protein folding (46,60), with bonded interactions described using:

$$E_{bonded} = \Sigma_{bonded} K_r(r - r_0)^2 + \Sigma_{angles} K_\theta(\theta - \theta_0)^2 + \Sigma_{dihedrals} K_\phi^{(n)}[1 - \cos(n(\phi - \phi_n))] \quad (8)$$

Here, K_r, r, r_0, K_θ, θ and θ_0 have the same meanings as before (see Eq. 2), but with the following values being used: r_0 = 6.495 Å, K_r = 20 kcal/mol/Å^2, θ_0 = 149.4°, K_θ = 10 kcal/mol/rad. The final

summation over the dihedral angles (i.e. involving four consecutive phosphate beads on each DNA strand) expresses the energy as a sum of two cosine terms, with n = 1 and n = 3, respectively; in both cases, ϕ_n represents the phase angle defining the energy maximum of the cosine term (set such that the energy minimum occurs at a dihedral angle of 19.2°), and K_ϕ represents the height of the energy barrier for rotating around the dihedral. For these calculations, K_ϕ^1 = 0.5 kcal/mol/rad, and K_ϕ^3 = 0.25 kcal/mol/rad.

The formation of realistic major and minor groove dimensions (see below) was achieved through the addition of favorable Lennard-Jones potential functions acting between the DNA strands, for which the equilibrium distances were set to those found in B-DNA. Specifically, favorable interactions were included between phosphate i on strand A, and phosphates i-2 to i+3 on strand B; for all such pairs of beads, the energy well-depth, ε, was set to 0.5 kcal/mol. For all other pairs of beads, very weakly favorable Lennard-Jones interactions were included to prevent the formation of steric clashes; for such interactions we used σ = 10 Å and ε = 0.1 kcal/mol. Finally, nonbonded electrostatic interactions were again described using Eq. 7, but with each phosphate bead assigned a charge of -0.5 e in order to again crudely mimic the effects of counterion condensation. After this final energy-minimization, the absence of any steric clashes in each of the final 1 NTB chromosome structures was confirmed using in-house code that explicitly measures the shortest distances between all beads.

2F Parameterization of the coarse-grained simulation models

While the non-biological nature of the compaction process gives us some leeway in the description of its timescale and the modeling of hydrodynamic interactions (see above), one aspect that is certainly important to model correctly is the persistence length of the DNA. A CG

model that predicts a persistence length that is too small, for example, will likely allow the DNA to adopt conformations that are much more bent than is realistic; a CG model that predicts a persistence length that is too large, on the other hand, may be so stiff that compressing the DNA into the cell might prove impossible. Prior to performing the compaction simulations outlined above, therefore, a series of preliminary simulations was performed in order to properly parameterize the angle-bending force constant, K_θ, for plectonemic and non-plectonemic regions of DNA. To this end, isolated polymers representing 25 kb (i.e., 50 beads using the 500 BPB CG model) were simulated for periods of 200 μs (without confinement) with a range of values for K_θ. The "target" value of the persistence length for beads representing non-plectonemic DNA (i.e. regular B-DNA) was 50 nm (24). The target value for beads representing plectonemic DNA was estimated instead at 79 nm: this value was obtained using the relationship derived by Odijk (61): $P_s = 2P \sin \alpha$, where P is the persistence length of dsDNA and, if the plectonemic supercoil is imagined as a duplex wrapped around a central cylinder, α is the angular displacement (52°) of the rising duplex from a line perpendicular to the length of the cylinder (61,62).

Persistence lengths from BD simulations performed with the CG simulation models were calculated directly from the mean-squared value of the end-to-end length of the DNA, using the Worm-Like Chain relationship (63):

$$\langle R^2 \rangle = 2 l_p l_c - 2 l_p^2 \left(1 - e^{(-l_c / l_p)} \right) \quad (9)$$

Here R is the end-to-end distance, l_c is the contour length of the polymer, and l_p is the desired persistence length. Figure S3 shows that for both non-plectonemic (blue) and plectonemic DNA

(red), satisfactory reproduction of the target persistence lengths can be achieved using the K_θ values noted earlier (see section S2A).

A similar procedure was used to parameterize the 5 BPB simulation models. The force constant opposing changes in the bond length, K_r, was set to 20 kcal/mol/Å^2, while that opposing changes in the bond angle, K_θ, was parameterized to match the persistence length of DNA. In this case, parameterization was performed by simulating a 500 basepair double-stranded DNA (i.e. 100 5 BPB beads) for 10 μs using a timestep of 500 fs and including electrostatic interactions as noted above (see section S2D above). For K_θ, a range of values was used, with the value of 3.5 kcal/mol/rad selected for all subsequent simulations. While not exploited further in this work, it is worth noting that the inclusion of a Debye-Hückel approximation to the electrostatic interactions means that the 5 BPB model can also describe quite well the salt-dependence of the persistence length of double-stranded DNA (Figure S3C).

2G Construction of a "uniform," nonplectonemic model

To assess the role that the plectonemic arrangement of the chromosome might make in the compactness of genomic regions, including macrodomains, we constructed a chromosomal model at 1 NTB resolution consisting entirely of single-duplex DNA. The protocol for its construction was essentially as described above, moving from an extended CG representation to higher resolutions, except that minor modifications were found to be necessary owing to the fact that the structure was initially refractory to further compaction when the shrinking volume reached ~150% that of the nucleoid. To overcome this problem, the coordinates and equilibrium bond lengths of the already-substantially-contracted model, were scaled down so that it would fit within the final nucleoid volume. A series of simulations was then performed during which

the scaled-down equilibrium bond lengths were progressively increased to their original values. Although the model was then fitted to the DNA probability density distribution described above (40), allowing comparison to plectonemic models with equivalent density distributions, it was not subject to RNAP density fitting: this allowed the same model to be used to represent both of the alternative global orientations.

For purposes of comparison with *oriC*@pole and *oriC*@midcell models, basepair numberings were adjusted in the following way. To match the *oriC*@pole arrangement we searched the genome in 100 bp blocks. The mean distance of each block to the nearest nucleoid pole was calculated; then, the mean distance of a 100 bp block, displaced one half-genome (2.3 Mb) from the trial block, to the opposite nucleoid pole was calculated. The center of the block for which the sum of these distances was smallest was selected to represent basepair 3925860 (in the center of *ori*C); all other beads were numbered accordingly. To match the *oriC*@midcell arrangement, we identified all beads within 1000 Å of the cell center. For each of these beads, mean distances were then calculated between (1) a designated "left" locus (a 100 bp block at a fixed genomic displacement of -0.3 genomes from the trial bead) and the nucleoid pole closest to it, and (2) a designated "right" locus (+0.3 genomes from the trial bead) and the opposite nucleoid pole. The bead whose selection minimized the sum of these distances was selected as the central basepair in *oriC*. The apparent persistent length of DNA, calculated as in section S3A below, in regions corresponding to PFRs in these models is 475 ±129 Å.

3 Methods used to analyze physical characteristics of the chromosome models

3A Groove widths and bend angle distributions; measurement of persistence lengths

Major and minor groove widths in the final 1 NTB resolution structures were calculated from the distances between phosphate beads using the same convention used by El Hassan and Calladine (64). Specifically, the major groove width was defined using the distance between phosphate i on the first strand of the DNA and phosphate i+4 on the second strand. The minor groove width was defined using the distance between phosphate i on the first strand and phosphate i–3 on the second strand.

Although there is no reason to expect particularly good agreement, the measured groove widths from our structures (major and minor groove widths of ~16.9 ± 0.9 and ~14.2 ± 0.8 Å, respectively) are in reasonable agreement with the average values identified for NMR structures by Perez et al. (65) (major and minor groove widths of ~18 ± ~2 and ~13 ± ~2 Å, respectively), in reasonable agreement with the average values obtained from MD simulations performed with the AMBER force field with parmbsc0 parameters (66) (~19 ± ~2 and ~12-13 ± ~2 Å, respectively, based on inspection of Figure 3 of Perez et al.), and in surprisingly good agreement with the corresponding values obtained using the CHARMM27 force field (67,68) (~17 ± ~2 and ~13-14 ± ~2 Å, respectively, again based on inspection of Figure 3 of Perez et al. (65)).

Bending angle distributions of the double-helical DNA were calculated using 3 beads placed at the position of the helical axis, spaced at 15 basepair intervals: i.e., by measuring the angle made by axis-beads placed at basepair positions: i-15, i, i+15. Following the convention that an unbent DNA has a bending angle of zero, we subtract the measured angle from 180°. The bending angle defined in this way was calculated for all 4641652 possible values of i. The

persistence length was then computed from the observed distribution of bending angles assuming a Worm-Like Chain model as follows. From Mazur (69), we have that the probability distribution for bending angles is given by:

$$dP \quad \sim \quad \exp\left(- l_p \, \theta^2 / 2 \, l_c \right) \sin \theta \, d\theta \qquad (10)$$

Here θ is the bending angle, l_c is the contour length of the fragment over which bending is measured (i.e. 30 basepairs × 3.4 Å), and l_p is the persistence length. We compared the probability distribution of bend angles predicted by Eq. 10 with that obtained from direct measurement of the structures (see above), and used the Goal Seek function in Microsoft Excel to adjust l_p so as to minimize the absolute error between the two distributions.

Persistence lengths for the B-DNA in the PFRs of the 1 NTB resolution structures were also obtained using end-to-end distance measurements as described in section S2F. In this case, the end-to-end distance of the central 1 kb stretch of each PFR in each structure was measured, squared, averaged over the 20 models, and then fit to the prediction of the Worm-Like Chain (WLC) model (Eq. 4); error bars were obtained from the standard deviation of the persistence length values obtained for each of the 40-44 PFRs.

Finally, we also measured *apparent* persistence lengths from analysis of contact probabilities averaged over entire chromosome structures: such an analysis allows comparison to be made with experimental estimates of the apparent persistence length, which are derived from the effective molar concentration of interacting DNA sites in gene expression studies (70). To this end, we again assumed a Worm-Like Chain model and used the following equation taken from Zhou (54):

$$G_0(r|l_c) = 4\pi r^2(3/4\pi l_p l_c)^{3/2} \exp(-3r^2/4l_p l_c)(1 - 5l_p/4l_c + 2r^2/l_c^2 - 33r^4/80l_p l_c^3 - 79l_p^2/160l_c^2 -$$

$$329r^2 l_p/120l_c^3 + 6799r^4/1600l_c^4 - 3441r^6/2800l_p l_c^5 + 1089r^8/12800l_p^2 l_c^6) \quad (11)$$

Here, $G_0(r|l_c)$ is the probability density for end-to-end distance, r, given a contour length l_c, and l_p is the persistence length of the polymer. We determined the probability of contact (i.e., the probability that the inter-bead distance was < 100 Å) between all basepairs separated by a given genomic distance in each of the model populations. Then, we adjusted l_p for each genomic separation until the integrated probability from 0 to 100 Å calculated using Eq. 11 matched the probability of contact we had determined for that separation. We note that since, as observed in the main text, probabilities of contact scale as s^{-1}, rather than as $s^{-3/2}$, the apparent persistence lengths derived in this way decline as genomic separation increases: a similar trend can be observed in studies of DNA looping *in vivo*, where the effective molar concentration (which is proportional to contact probability) of interacting sites also declines less rapidly than $s^{-3/2}$ (71).

3B Calculation of twist, writhe, and linking number and consideration of knots

We calculated twist (Tw) and writhe (Wr) parameters for 1NTB resolution structures using either the WrLINE software released by the Noy group (University of Leeds, UK) or using in-house code that is heavily based upon it. The WrLINE Python software was downloaded from http://ccpforge.cse.rl.ac.uk/gf/project/wrline. A slight modification was made to the code to make it write out twist, which is computed as an intermediate step of the WrLINE procedure: in that code, twist is calculated using equation 6 of Sutthibutpong et al. (72). Each twist calculation on a complete chromosome structure containing 4641652 basepairs modeled at a 1 NTB resolution took less than 10 minutes. For the purposes of calculating writhe, however, (which

uses equation 5 of Clauvelin et al. (73)), the code did not complete even after three days of calculation time when applied directly to a 1NTB chromosome structure. We therefore ported the Python code into C++ and used OpenMP constructs to parallelize the most expensive parts of the calculation. To verify that our C++ code was correct, we first performed validation tests on a number of relaxed and supercoiled plasmids containing 500 or 2100 basepairs (Figure S18; Table S1) and compared the results of our code with those of the original WrLINE code. When applied to a complete 1NTB chromosome structure, our code completes the calculation of writhe in ~7 hours running on a 64-core AMD Opteron 6272 server. The linking number (Lk) was calculated using OpenMP-parallelized C++ code written entirely in-house, and using the procedure outlined by equations 14, 15 and 16 in (73). Correctness of the linking number code was verified, again using the supercoiled plasmid structures, by comparison with the sum of the twist and writhe values obtained with WrLINE, i.e. using Lk = Tw + Wr. Calculation of the linking number for a complete 1NTB chromosome structure requires ~8 days running on a 64-core server. For this reason, we made use of the relation Lk = Tw + Wr for all other chromosome structures.

Following studies of plasmids indicating that, even in cells with a full complement of topoisomerases, some amount of DNA knotting occurs naturally (74,75), we also attempted to assess knotting in our structures. In principle, evaluation of knots can be performed using the Rknots package (76) – which was originally intended to analyze knots in protein structures – implemented in the statistical program R (77). Calculations using Rknots typically proceed in two stages. First, the knot diagram is projected and "reduced" in a process that decreases the number of points that a structure contains while preserving its topology. Second, to determine

the complexity of the resulting knot, a polynomial invariant (a formula defined by the knot type but independent of its particular projection) (78) is calculated from the reduced structure. We determined that the cost of the first stage rises quadratically with the number of beads in the structure. Using an entire chromosome structure represented with bead numbers ranging from 100 to 2500, for example, we found the computational time increasing from 1.153 s to 1284 s. These data fit to the following quadratic relationship: computational time = $0.00020064x^2 + 0.01147x$, where x is the number of beads in the structure, with an r^2 of 0.997. When extrapolated to the total number of beads that would be present in a 5 BPB model (928,331), we estimate that completion of the first stage of the calculation would require ~2000 days on a single CPU.

Although it might be possible to complete reduction of structures with fewer beads, it should be noted that success in the second stage – calculation of the polynomial invariant – is harder for us to be sure of since computing it can also become prohibitively expensive when a projection contains many crossings (78). In preliminary tests, we found no obvious trend connecting the time required to compute the polynomial invariant and the number of beads in the structure, and attempts to analyze larger structures generated errors. While the complexity of knotting in the chromosome structures could not be determined, visual inspection of them reveals occasional passage of one strand of a plectonemic supercoil through the helical axis of another plectoneme. We are unable to characterize the effects of this interstrand passage topologically, but we suspect that the low-frequency knotting suggested in the structures is consistent with the fact that knots can be observed in plasmids in wild-type cells (74,75).

3C Distributions of void sizes in the chromosomal interior

The distribution of void diameters accessible within the core of the nucleoid was calculated in the following way. At each spatial location of interest (see below), we attempted to determine the diameter of the largest possible sphere that could be placed without encompassing an atom of the chromosome. This maximal diameter was determined by trial and error at each location, starting with a sphere of diameter 10 Å and incrementing the diameter in intervals of 10 Å. In the coordinate frame used in our models, the chromosome extends from approximately -8000 Å to +8000 Å in the x direction, and from approximately -4000 to +4000 Å in both the y & z directions. Since we wish to focus on the steric accessibility of the nucleoid interior, the void diameter was determined only at grid points stretching from -4000 Å to +4000 Å in x, and from -2000 to +2000 Å in y & z (i.e. at a total of 81 × 41 × 41 = 136161 grid points). Histograms of the maximal diameter values were made from all 20 structures of each model type, and the mean values and standard deviations of each histogram were determined.

3D Distributions of distances between loci

2D maps showing the distances between all possible pairs of DNA segments in the chromosome models were constructed in the following way. The chromosome was first divided into 10 kb blocks, and reference points within each block were placed at the position of the helical axis for every 100th basepair. The mean distance between each pair of 10 kb blocks was then computed by averaging the distance between all pairs of reference points within the two blocks (i.e., as the average of 10,000 distances). This calculation yields a 464 × 464 map of distances for all 20 structures; these were then averaged to give the final maps shown in Figure 6B and in Figure S7.

3E Fractal globule characteristics

In several eukaryotic cell types, DNA seems to be arranged in a non-equilibrium state known as a "fractal globule" (79-81). The well-defined domains associated with PARs in contact maps derived from experiment and simulation (9,10) (Figure 6B), suggest that the organization of DNA in prokaryotic cells may also be consistent with the strong territorial definition characteristic of the fractal globule (81,82). Interestingly, while under some conditions, the physical distance between the origin and other loci in *E. coli* scales linearly with genomic distance (18) – which would not be consistent with the behavior expected of a fractal globule confined within a sphere – it has been proposed that these linear scaling regimes may nevertheless be consistent with a fractal globule confined within the elongated geometry of the bacterial nucleoid (81).

In the fractal globule model, the probability of contact between loci separated by a given genomic distance, s, is expected to scale as s^{-1} over a range of polymer and subchain lengths, while probabilities in the alternative equilibrium globule scale first as $s^{-3/2}$ before arriving at a constant value at larger separations (81,83). Similarly, root-mean-square end-to-end distances have different characteristic scaling exponents in fractal and equilibrium globules: while both states scale as $s^{1/2}$ at short genomic separations, at greater genomic separations the fractal globule scales as $s^{1/3}$ while the equilibrium globule is expected to simply plateau. Although the existence of well-defined scaling regimes is apparent in simulations of confined polymers (82,83), experimental data from eukaryotic chromosomes (as well as simulations of them) do not always resolve as cleanly into distinct regimes: in some cases, for example, the final $s^{1/3}$ scaling only appears at substantial genomic separations and is preceded by an abrupt rise in distance

and an "intermediate" scaling region where distances increase more slowly (80). In simulations in which reduced models of the *C. crescentus* chromosome are confined to a capsule, moreover, three scaling regimes are apparent, with larger scaling exponents associated with the smallest and greatest genomic separations (84). Finally, we note that although we have referred in the above to the "equilibrium globule," a highly-knotted globule may display similar scaling characteristics (83).

To compare the properties of our chromosome structures to those of the fractal globule, we proceeded as follows. To calculate the probabilities of contact between basepairs as a function of their genomic separation, we generated a set of genomic displacements according to the procedure described by Imakaev et al.: bins were of logarithmically-increasing sizes, starting with 6bp and increasing with a step size of 1.1, with displacements rounded to the nearest integer and redundant bins removed, creating the pattern …,x, 1.1x, 1.1^2x,…,N (83). The resulting displacements range from 6 bp to ~2 MB, representing approximately half the size of the *E. coli* genome. The probability of contact was determined by calculating distances between the axis of every chromosomal basepair and the axis of the basepair offset by the bin displacement in both model sets; loci were considered in contact if the distance between them was <100 Å. Root-mean-square end-to-end distances were calculated using the same sets of displacements over the entire chromosome in all structures.

3F Determination of shortest paths between loci

The shortest possible paths between all pairs of genomic loci were determined as follows under the assumption that each path consists only of: (a) linear diffusion along the DNA, and (b) jumps between genomically-distant loci at points of close DNA-DNA contact. For each

structure, we first determined all pairs of basepairs that were separated by more than 50 basepairs in sequence but whose centers were in close physical contact (i.e. closer than 50 Å). We then divided the entire chromosome into contiguous blocks of size 1 kb and identified all pairs of blocks that contained at least one basepair-basepair contact satisfying the separation and distance criteria outlined above. For all such pairs of 1 kb blocks we set the length of the path connecting them to 1000 basepairs: this is the average of the path lengths connecting all possible pairs of basepairs in the two blocks. For all pairs of blocks that did not have at least one basepair-basepair contact, we set the length of the path connecting them to be equal to their genomic distance (measured between the centers of the two blocks). Having assigned path lengths to all 4652 × 4652 block pairs, we then used in-house code implementing the Floyd-Warshall-Roy (85-87) algorithm to find the shortest possible paths between all pairs of blocks. This calculation was repeated over the 20 representative structures for each type of chromosome model in order to obtain an average path-length for all pairs of blocks. Finally, these average path lengths were grouped according to the genomic distance separating the two blocks in each pair, and the average path lengths within each group were themselves averaged.

4 Methods used to analyze genetic characteristics of the chromosome models

4A Long axis distribution of genomic loci

To compare with the fluorescence analysis of genomic loci reported by the Wiggins and Kondev groups (18), we divided the chromosome into 463 contiguous blocks of size 10 kb, determined the long axis position of the center of each block (expressed relative to the total cell length of 2.4 μm), and averaged these values over the 20 representative structures for each type of

chromosome model. We then grouped these values according to the genomic distance (expressed in genome units) of each block from the origin of replication, *oriC*, and averaged all of the values within each group. Error bars plotted in Figure S10 represent the standard deviations of the values within each group.

4B Radial distributions of EPODs and genes organized by protein subcellular location

Information on the genomic coordinates of the 272 EPODs identifed by Vora et al. (88) was downloaded from their publication. The radial positions of the centers of each EPOD were determined and averaged over the 20 representative structures for each type of chromosome model. The EPODs were then grouped into five categories according to the $\log_{10}$ of their RNA expression levels reported by Vora et al. (88). The average radial positions of all EPODs within each category were then themselves averaged and plotted versus the average $\log_{10}$ of the RNA expression levels within each category.

A similar approach was used to analyze the radial distributions of genes grouped according to the final subcellular localization of their protein products. Genes in each of the following four categories were identified from the EcoCyc server (89): cytoplasmic (2002 genes), periplasmic (209 genes), inner membrane (1175 genes), and outer membrane (123 genes). The radial positions of the centers of all genes in each category were then determined, averaged over the 20 representative structures for each type of chromosome model, before these averaged radial positions were themselves averaged over all genes in each category. As shown in Figure S9C, no discernible difference can be seen between the average radial distributions of genes in any of the four categories.

4C Analysis and attempted enforcement of macrodomain organization

Macrodomain boundaries were obtained from previous studies on chromosome organization (90,91). To assess the levels of compaction of these various domains, radii of gyration, R_g, were calculated using $R_g^2 = (1/N)\Sigma_i(r_i - r_{mean})^2$, where $r_i - r_{mean}$ is the distance between each basepair and the domain's geometric center and N is the number of basepairs in the domain. R_g values were calculated for all 20 structures in each model set and then averaged. Separate calculations of R_g were performed for the *oriC*@pole, *oriC*@midcell and non-plectonemic models.

Since our analysis found little difference between levels of compaction in macrodomains and non-structured regions (Figure 8A), we attempted to impose macrodomain structure on the chromosomal models. While previous attempts to do this with lower resolution models have involved explicit confinement of the genomic regions corresponding to macrodomains within spheres (41), the availability of structural models allowed us to adopt an alternative approach. Mean distances between recombination sites were calculated and subsequently fitted to reported recombination frequencies between the loci (91) in order to establish a relationship between physical distance and rates of recombination. Target distances between loci—i.e., distances between them that would "reproduce" their experimentally-determined recombination frequency by matching the distance associated with their recombination rate—were determined from the exponential fit. In the process designed to enforce these distances, bonds were established between potentially-combining loci in the complete coarse-grained model set. In a series of 2500 short BD simulations in which biasing potentials for fitting RNAP and DNA distribution (section S2B) continued to be applied, bond lengths, which began at the mean distance between loci in unaltered models, were progressively reduced or increased until

their equilibrium length matched the target distance for each locus-locus pair; analysis of distances between loci before and after this process tracked mean interlocus distances. Radii of gyration were then recalculated to determine the effect of enforcement on the compactness of chromosomal regions.

We note that although the correlation of target and final model distances is robust (Figure S12) and recombination frequencies derived from the final distances fit well with experimental data (Figure S11), in at least one case (attL7, for a recombining locus at ~0.89Mb) the model set is unable to realize the mean target distance between loci and the corresponding increase in recombination frequency (bottom-middle panel of Figure S11). We suspect that the models' inability to enforce this distance arises from the high density of restraints in the vicinity of 1Mb that must be enforced simultaneously in accord with experimental data for the attR17 site: within the limits of the model population, satisfaction of all distance restraints within this region does not appear to be possible. While increasing the model population might allow the greater variance across models necessary to match the calculated target distance, we think that a closer fit to this distance would not appreciably change the structuring of macrodomains.

We also note that, for the future, information incorporated in an attempt to better define macrodomains could include not only recombination rates for additional loci, but also region-specific supercoiling density data. Although work focused on these regions in *Salmonella* (92) and *E. coli* (93) has indicated differences in density between Ter and other regions it is not clear to what extent these results are consistent with a gradient of decreasing supercoiling density from Ori to Ter (94). With the exception of Ter, for example, supercoiling is reported to be approximately constant across the *Salmonella* chromosome (92), and regional differences in

supercoiling have only been described in stationary phase in *E. coli* (93). Details of the dependence of any such gradient on energy availability and local supercoiling density continue to emerge, and future iterations of the models should be able to readily accommodate these data as they become available.

4D Analysis of distributions of selected gene pairs

Gene expression data were obtained from the M3D database (95). To determine their coexpression, we adopted methods employed by Xie at al. (96), calculating Pearson correlation coefficients for expression levels of the ~9 million pairs from a set of 4297 genes whose expression was assayed in 466 experiments. Using a list of gene locations from EcoCyc (89), we measured distances between all gene pairs in the chromosome at a resolution of 10 bp. Gene pairs were rank-ordered by the Pearson correlation coefficients of their expression levels, and distances of genes separated by more than 1.5 Mb for which there was expression data were plotted: dividing the pairs into deciles by correlation coefficients generated 10 sets, each including ~315000 gene pairs. In heat maps comparing the *oriC*@pole and *oriC*@midcell models (Figures S13 and S14C), gene-pairs were binned according to their mean physical distances in the two models (using bin sizes of 100 Å) and colored according to the mean Pearson correlation coefficients of all gene-pairs sorted into each bin. Lists of genes in pairs from the bottom and top deciles were submitted to PantherDB to test for overrepresentation using the complete gene-ontological categories of molecular function, biological process, and cellular component (97) but did not reveal clear trends among the gene sets.

4E Variance of the long axis separation distance between genomic loci and *oriC*

To extend the comparison with the work of the Wiggins and Kondev groups (18), we attempted to determine the variance of the distance (expressed in units of cell length) measured from each of the 463 10 kb blocks to *oriC*. This was achieved by first determining the variance of each block's distances measured in each of the 20 representative structures for each type of chromosome model, grouping these variances according to the genomic distance of each block from *oriC*, and averaging all of the values within each group. Interestingly, when the cell length is assumed to be fixed at 2.4 μm, the distribution of variances does not follow the characteristic V-shape seen in the data of Wiggins et al (Figure S19A). When, however, the cell length is assumed to vary around a mean value of 2.4 μm but with a standard deviation of 0.5 μm, a much closer agreement with the experimental data is obtained (Figure S19B). Given that the present analysis has been carried out using only 20 representative structures, a more in-depth analysis of variance distributions should probably wait until a much larger sample of structures has been obtained.

Supplementary References

1. Grainger, D.C., Hurd, D., Harrison, M., Holdstock, J. and Busby, S.J.W. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA*, **102**, 17693-17698.

2. Prieto, A.I., Kahramanoglou, C., Ali, R.M., Fraser, G.M., Seshasayee, A.S.N. and Luscombe, N.M. (2012) Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in *Escherichia coli* K12. *Nucleic Acids Res*, **40**, 3524-3537.

3. Kahramanoglou, C., Seshasayee, A.S.N., Prieto, A.I., Ibberson, D., Schmidt, S., Zimmermann, J., Benes, V., Fraser, G.M. and Luscombe, N.M. (2011) Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res*, **39**, 2073-2091.

4. Cho, B.K., Knight, E.M., Barrett, C.L. and Palsson, B.O. (2008) Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res*, **18**, 900-910.

5. Mondal, J., Bratton, B.P., Li, Y.J., Yethiraj, A. and Weisshaar, J.C. (2011) Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* cells. *Biophys J*, **100**, 2605-2613.

6. Bates, D. and Kleckner, N. (2005) Chromosome and replisome dynamics in *E. coli*: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, **121**, 899-911.

7. Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet*, **10**, 669-680.

8. Grainger, D.C., Hurd, D., Goldberg, M.D. and Busby, S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res*, **34**, 4642-4652.

9. Umbarger, M.A., Toro, E., Wright, M.A., Porreca, G.J., Bau, D., Hong, S.H., Fero, M.J., Zhu, L.J., Marti-Renom, M.A., McAdams, H.H. *et al.* (2011) The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol Cell*, **44**, 252-264.

10. Le, T.B.K., Imakaev, M.V., Mirny, L.A. and Laub, M.T. (2013) High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, **342**, 731-734.

11. Cagliero, C., Grand, R.S., Jones, M.B., Jin, D.J. and O'Sullivan, J.M. (2013) Genome conformation capture reveals that the *Escherichia coli* chromosome is organized by replication and transcription. *Nucleic Acids Res*, **41**, 6058-6071.

12. Jin, D.J., Cagliero, C. and Zhou, Y.N. (2013) Role of RNA polymerase and transcription in the organization of the bacterial nucleoid. *Chem Rev*, **113**, 8662-8682.

13. Higgins, N.P. (2014) RNA polymerase: chromosome domain boundary maker and regulator of supercoil density. *Curr Opin Microbiol*, **22**, 138-143.

14. Zhou, J.D. and Rudd, K.E. (2013) EcoGene 3.0. *Nucleic Acids Res*, **41**, D613-D624.

15. Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muniz-Rascado, L., Garcia-Sotelo, J.S., Alquicira-Hernandez, K., Martinez-Flores, I., Pannier, L., Castro-Mondragon, J.A. *et al.* (2016) RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res*, **44**, D133-D143.

16. Kleckner, N., Fisher, J.K., Stouf, M., White, M.A., Bates, D. and Witz, G. (2014) The bacterial nucleoid: nature, dynamics and sister segregation. *Curr Opin Microbiol*, **22**, 127-137.

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

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

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

20. Cass, J.A., Kuwada, N.J., Traxler, B. and Wiggins, P.A. (2016) *Escherichia coli* chromosomal loci segregate from midcell with universal dynamics. *Biophys J*, **110**, 2597-2609.

21. Meile, J.C., Mercier, R., Stouf, M., Pages, C., Bouet, J.Y. and Cornet, F. (2011) The terminal region of the *E. coli* chromosome localises at the periphery of the nucleoid. *BMC Microbiol*, **11,** 28.

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

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

24. Boles, T.C., White, J.H. and Cozzarelli, N.R. (1990) Structure of plectonemically supercoiled DNA. *J Mol Biol*, **213**, 931-951.

25. Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L. and Dame, R.T. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *J Struct Biol*, **156**, 262-272.

26. Dillon, S.C. and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol*, **8**, 185-195.

27. Browning, D.F., Grainger, D.C. and Busby, S.J.W. (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol*, **13**, 773-780.

28. Hardy, C.D. and Cozzarelli, N.R. (2005) A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Mol Microbiol*, **57**, 1636-1652.

29. Bakshi, S., Choi, H. and Weisshaar, J.C. (2015) The spatial biology of transcription and translation in rapidly growing *Escherichia coli*. *Front Microbiol*, **6**, 636.

30. Mercier, R., Petit, M.A., Schbath, S., Robin, S., El Karoui, M., Boccard, F. and Espeli, O. (2008) The matP/matS site-specific system organizes the terminus region of the *E. coli* chromosome into a macrodomain. *Cell*, **135**, 475-485.

31. Wang, W.Q., Li, G.W., Chen, C.Y., Xie, X.S. and Zhuang, X.W. (2011) Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, **333**, 1445-1449.

32. Wang, S.Y., Moffitt, J.R., Dempsey, G.T., Xie, X.S. and Zhuang, X.W. (2014) Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc Natl Acad Sci USA*, **111**, 8452-8457.

33. Sinden, R.R., Carlson, J.O. and Pettijohn, D.E. (1980) Torsional tension in the DNA double helix measured with trimethylpsoralen in living *Escherichia coli* cells - Analogous measurements in insect and human cells. *Cell*, **21**, 773-783.

34. Bliska, J.B. and Cozzarelli, N.R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism in vivo. *J Mol Biol*, **194**, 205-218.

35. Champion, K. and Higgins, N.P. (2007) Growth rate toxicity phenotypes and homeostatic supercoil control differentiate *Escherichia coli* from *Salmonella enterica serovar Typhimurium*. *J Bacteriol*, **189**, 5839-5849.

36. Marko, J.F. and Siggia, E.D. (1995) Statistical mechanics of supercoiled DNA. *Phys Rev E*, **52**, 2912-2938.

37. Volkmer, B. and Heinemann, M. (2011) Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PLoS One*, **6,** e0023126.

38. Yazdi, N.H., Guet, C.C., Johnson, R.C. and Marko, J.F. (2012) Variation of the folding and dynamics of the *Escherichia coli* chromosome with growth conditions. *Mol Microbiol*, **86**, 1318-1333.

39. Bakshi, S., Choi, H., Mondal, J. and Weisshaar, J.C. (2014) Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. *Mol Microbiol*, **94**, 871-887.

40. Stracy, M., Lesterlin, C., de Leon, F.G., Uphoff, S., Zawadzki, P. and Kapanidis, A.N. (2015) Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc Natl Acad Sci USA*, **112**, E4390-E4399.

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

42. Ermak, D.L. and Mccammon, J.A. (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys*, **69**, 1352-1360.

43. Vos, S.M., Tretter, E.M., Schmidt, B.H. and Berger, J.M. (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. *Nat Rev Mol Cell Biol*, **12**, 827-841.

44. Rotne, J. and Prager, S. (1969) Variational treatment of hydrodynamic interaction in polymers. *J Chem Phys*, **50**, 4831-4837.

45. Yamakawa, H. (1970) Transport properties of polymer chains in dilute solution - hydrodynamic interaction. *J Chem Phys*, **53**, 436-443.

46. Frembgen-Kesner, T. and Elcock, A.H. (2009) Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J Chem Theory Comput*, **5**, 242-256.

47. Bremer, H. and Dennis, P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In Neidhardt, F. (ed.), *Escherichia coli and Salmonella typhimurium.* ASM Press, Washington, DC, pp. 1553-1569.

48. Bratton, B.P., Mooney, R.A. and Weisshaar, J.C. (2011) Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. *J Bacteriol*, **193**, 5138-5146.

49. Bakshi, S., Dalrymple, R.M., Li, W.T., Choi, H.J. and Weisshaar, J.C. (2013) Partitioning of RNA polymerase activity in live *Escherichia coli* from analysis of single-molecule diffusive trajectories. *Biophys J*, **105**, 2676-2686.

50. Geszvain, K. and Landick, R. (2005) Structure of bacterial RNA polymerase. In N. Patrick Higgins, (ed.), *The Bacterial Chromosome.* ASM Press, Washington, DC, pp. 283-296.

51. Soper, A.K. (1996) Empirical potential Monte Carlo simulation of fluid structure. *Chem Phys*, **202**, 295-306.

52. Reith, D., Putz, M. and Muller-Plathe, F. (2003) Deriving effective mesoscale potentials from atomistic simulations. *J Comput Chem*, **24**, 1624-1636.

53. Andrews, C.T. and Elcock, A.H. (2014) COFFDROP: A coarse-grained nonbonded force field for proteins derived from all-atom explicit-solvent molecular dynamics simulations of amino acids. *J Chem Theory Comput*, **10**, 5178-5194.

54. Zhou, H.X. (2001) Loops in proteins can be modeled as worm-like chains. *J Phys Chem B*, **105**, 6763-6766.

55. Tirion, M.M. (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys Rev Lett*, **77**, 1905-1908.

56. Hinckley, D.M., Lequieu, J.P. and de Pablo, J.J. (2014) Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects. *J Chem Phys*, **141,** 035012.

57. Clementi, C., Nymeyer, H. and Onuchic, J.N. (2000) Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol*, **298**, 937-953.

58. Koga, N. and Takada, S. (2001) Roles of native topology and chain-length scaling in protein folding: A simulation study with a Go-like model. *J Mol Biol*, **313**, 171-180.

59. Chavez, L.L., Onuchic, J.N. and Clementi, C. (2004) Quantifying the roughness on the free energy landscape: Entropic bottlenecks and protein folding rates. *J Am Chem Soc*, **126**, 8426-8432.

60. Elcock, A.H. (2006) Molecular simulations of cotranslational protein folding: Fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Comput Biol*, **2**, 824-841.

61. Odijk, T. (1996) DNA in a liquid-crystalline environment: Tight bends, rings, supercoils. *J Chem Phys*, **105**, 1270-1286.

62. Cunha, S., Woldringh, C.L. and Odijk, T. (2001) Polymer-mediated compaction and internal dynamics of isolated *Escherichia coli* nucleoids. *J Struct Biol*, **136**, 53-66.

63. Cantor, C.R. and Schimmel, P.R. (1984) *Biophysical Chemistry: Part III*. WH Freeman, San Francicso, CA.

64. El Hassan, M.A. and Calladine, C.R. (1998) Two distinct modes of protein-induced bending in DNA. *J Mol Biol*, **282**, 331-343.

65. Perez, A., Lankas, F., Luque, F.J. and Orozco, M. (2008) Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res*, **36**, 2379-2394.

66. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., Laughton, C.A. and Orozco, M. (2007) Refinenement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J*, **92**, 3817-3829.

67. Foloppe, N. and MacKerell, A.D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem*, **21**, 86-104.

68. MacKerell, A.D. and Banavali, N.K. (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J Comput Chem*, **21**, 105-120.

69. Mazur, A.K. (2007) Wormlike chain theory and bending of short DNA. *Phys Rev Lett*, **98**, 218102.

70. Rippe, K. (2001) Making contacts on a nucleic acid polymer. *Trends Biochem Sci*, **26**, 733-740.

71. Priest, D.G., Cui, L., Kumar, S., Dunlap, D.D., Dodd, I.B. and Shearwin, K.E. (2014) Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and lambda repressors. *Proc Natl Acad Sci USA*, **111**, 349-354.

72. Sutthibutpong, T., Harris, S.A. and Noy, A. (2015) Comparison of molecular contours for measuring writhe in atomistic supercoiled DNA. *J Chem Theory Comput*, **11**, 2768-2775.

73. Clauvelin, N., Olson, W.K. and Tobias, I. (2012) Characterization of the geometry and topology of DNA pictured as a discrete collection of atoms. *J Chem Theory Comput*, **8**, 1092-1107.

74. Shishido, K., Komiyama, N. and Ikawa, S. (1987) Increased production of a knotted form of plasmid pBR322 DNA in *Escherichia coli* DNA topoisomerase mutants. *J Mol Biol*, **195**, 215-218.

75. Deibler, R.W., Rahmati, S. and Zechiedrich, E.L. (2001) Topoisomerase IV, alone, unknots DNA in *E. coli*. *Genes Dev*, **15**, 748-761.

76. Comoglio, F. and Rinaldi, M. (2012) Rknots: topological analysis of knotted biopolymers with R. *Bioinformatics*, **28**, 1400-1401.

77. R Core Team (2016) R: A language and environment for statistical comupting. R Foundation for Statistical Computing, Vienna, Austria.

78. Meluzzi, D., Smith, D.E. and Arya, G. (2010) Biophysics of knotting. *Annu Rev Biophys*, **39**, 349-366.

79. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.

80. Rosa, A. and Everaers, R. (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput Biol*, **4,** e1000153.

81. Mirny, L.A. (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res*, **19**, 37-51.

82. Tamm, M.V., Nazarov, L.I., Gavrilov, A.A. and Chertovich, A.V. (2015) Anomalous diffusion in fractal globules. *Phys Rev Lett*, **114,** 178102.

83. Imakaev, M.V., Tchourine, K.M., Nechaev, S.K. and Mirny, L.A. (2015) Effects of topological constraints on globular polymers. *Soft Matter*, **11**, 665-671.

84. Hong, S.H., Toro, E., Mortensen, K.I., de la Rosa, M.A.D., Doniach, S., Shapiro, L., Spakowitz, A.J. and McAdams, H.H. (2013) *Caulobacter* chromosome in vivo configuration matches model predictions for a supercoiled polymer in a cell-like confinement. *Proc Natl Acad Sci USA*, **110**, 1674-1679.

85. Floyd, R.W. (1962) Algorithm-97 - Shortest Path. *Commun ACM*, **5**, 345-345.

86. Warshall, S. (1962) A theorem on boolean matrices. *J ACM*, **9**, 11-12.

87. Roy, B. (1959) Transitivite Et Connexite. *Cr Hebd Acad Sci*, **249**, 216-218.

88. Vora, T., Hottes, A.K. and Tavazoie, S. (2009) Protein occupancy landscape of a bacterial genome. *Mol Cell*, **35**, 247-253.

89. Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M. *et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res*, **41**, D605-D612.

90. Espeli, O., Mercier, R. and Boccard, F. (2008) DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol Microbiol*, **68**, 1418-1427.

91. Valens, M., Penaud, S., Rossignol, M., Cornet, F. and Boccard, F. (2004) Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J*, **23**, 4330-4341.

92. Rovinskiy, N., Agbleke, A.A., Chesnokova, O., Pang, Z.H. and Higgins, P. (2012) Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLoS Genet*, **8,** e1002845.

93. Lal, A., Dhar, A., Trostel, A., Kouzine, F., Seshasayee, A.S.N. and Adhya, S. (2016) Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun*, **7,** 11055.

94. Sobetzko, P., Travers, A. and Muskhelishvili, G. (2012) Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci USA*, **109**, E42-E50.

95. Faith, J.J., Driscoll, M.E., Fusaro, V.A., Cosgrove, E.J., Hayete, B., Juhn, F.S., Schneider, S.J. and Gardner, T.S. (2008) Many Microbe Microarrays Database: Uniformly normalized Affymetrix compendia with structured experimental metadata. *Nucleic Acids Res*, **36**, D866-D870.

96. Xie, T., Fu, L.Y., Yang, Q.Y., Xiong, H., Xu, H.R., Ma, B.G. and Zhang, H.Y. (2015) Spatial features for *Escherichia coli* genome organization. *BMC Genomics*, **16,** 37.

97. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D. and Thomas, P.D. (2017) PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*, **45**, D183-D189.

98. Baumann, C.G., Smith, S.B., Bloomfield, V.A. and Bustamante, C. (1997) Ionic effects on the elasticity of single DNA molecules. *Proc Natl Acad Sci USA*, **94**, 6185-6190.

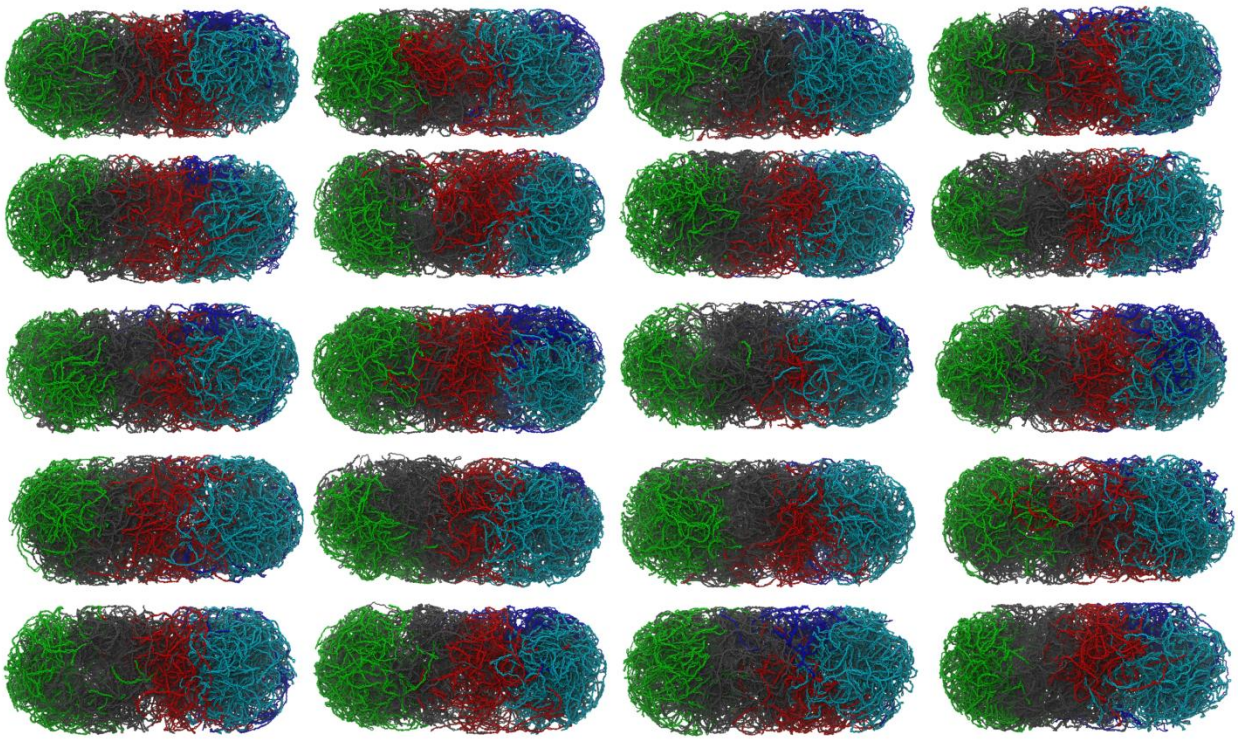

Figure S1. The *oriC*@pole model set. Twenty 1 NTB-resolution structures with the *oriC*@pole global orientation are depicted. The coloring scheme indicates genomic regions as described by Valens et al. (91): Ori (green), Ter (cyan), Right (red), Left (blue), NS-L, and NS-R (both dark gray).

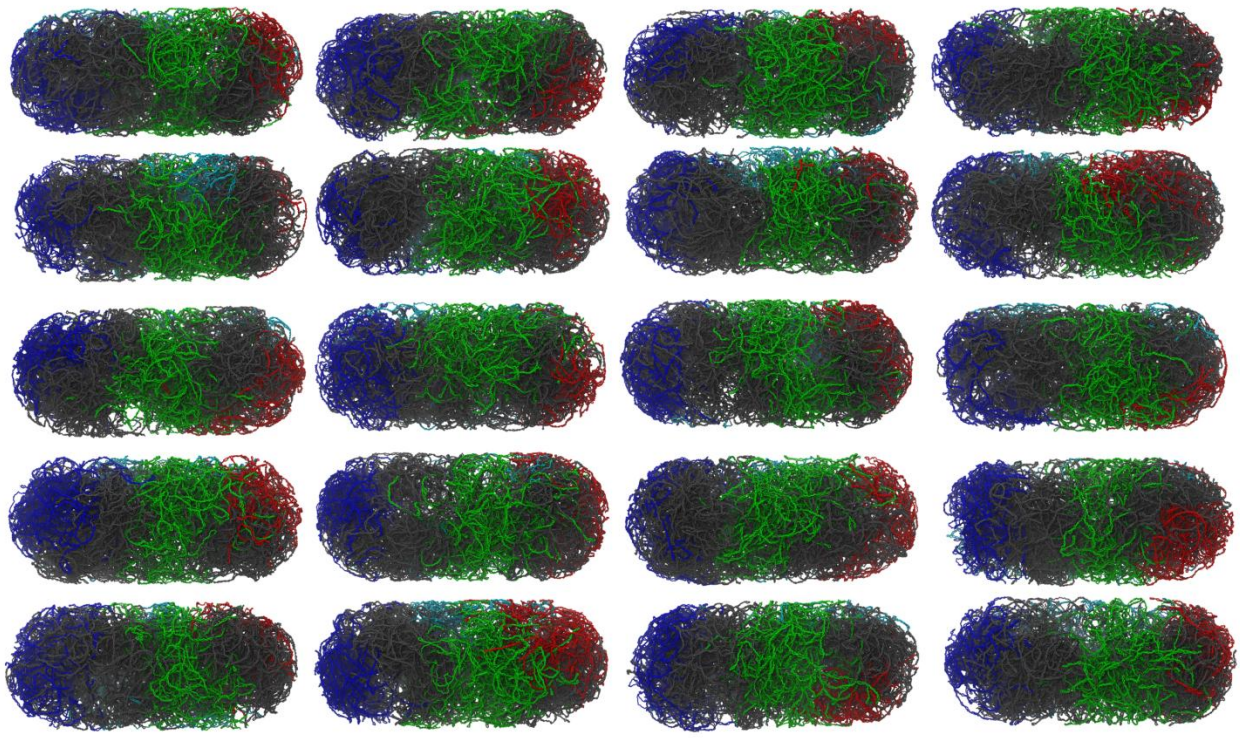

Figure S2. The *oriC*@midcell model set. Twenty 1 NTB-resolution structures with the *oriC*@midcell global orientation are depicted. The coloring scheme is as in Figure S1.

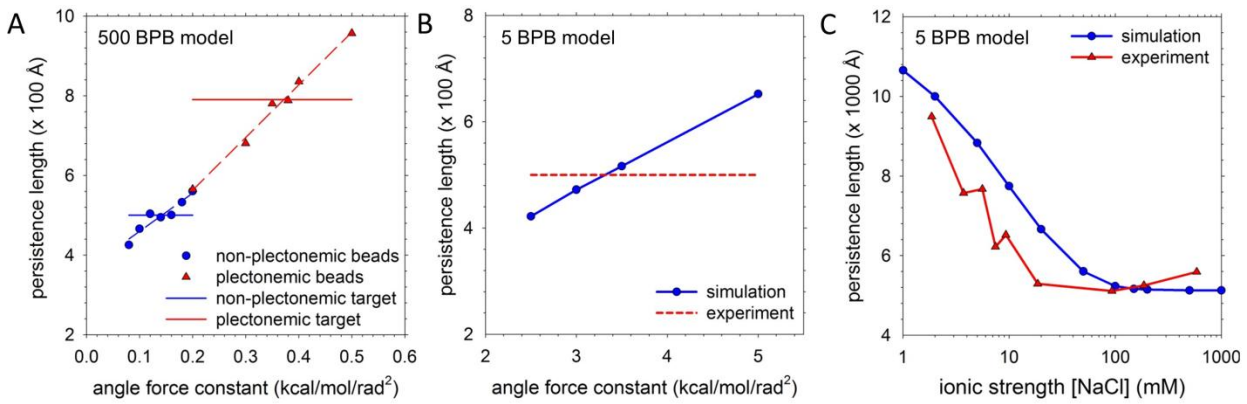

Figure S3. **Parameterization of CG simulation models to match DNA persistence length. A.** 500 BPB models match experimental estimates of persistence length. Since beads in these structures can represent either plectonemic or nonplectonemic DNA, two fittings were required. Angle force constants were adjusted in the 500 BPB model to match a predicted persistence length of 79 nm for plectonemic supercoils (red) and 50 nm for single-duplex DNA (blue). **B.** 5 BPB models match experimental estimates of persistence length. Persistence lengths of a 500 bp polymer represented at 5 BPB resolution are plotted. Since plectonemes are explicitly represented at this resolution, only a single fitting is required. **C.** Persistence length as a function of salt concentration. Simulated values (blue) using the force constant determined in **B** are compared with experimental values (red) for a range of salt concentrations (98); simulated values match experiment at ~150mM.

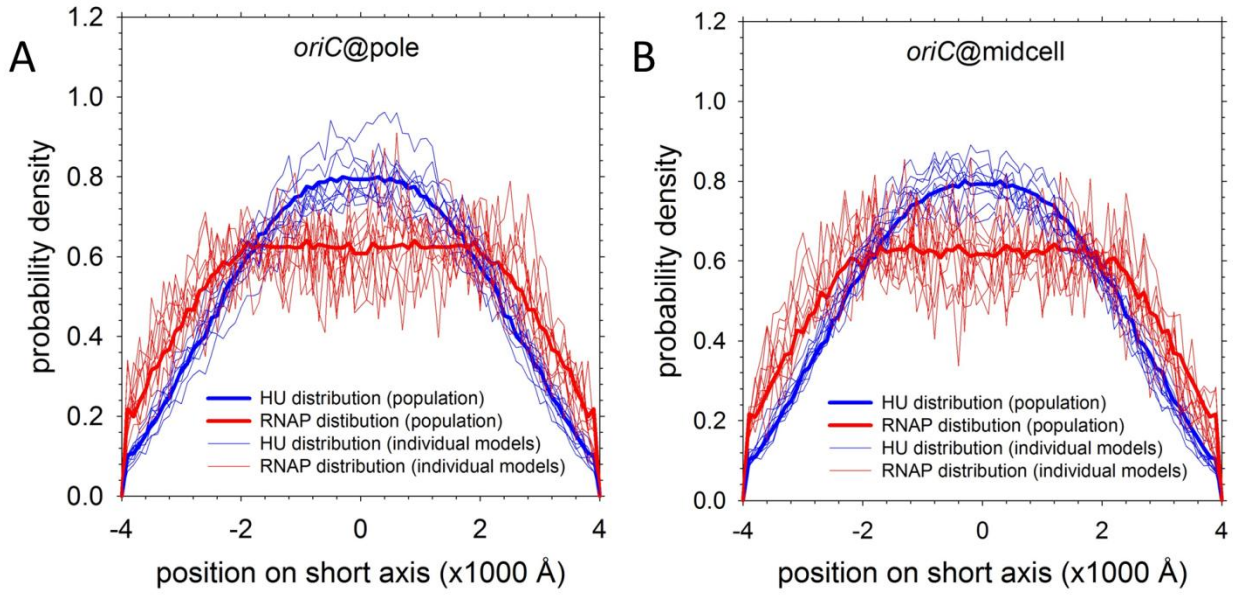

Figure S4. Model-to-model variation in DNA and RNAP densities. Short-axis probability densities are plotted for DNA (blue) and RNAP (red) as in Figure 5 of the main text. Mean densities of the model population are plotted in bold; finer lines represent densities in individual models. At left, distributions for *oriC*@pole models are plotted; at right, distributions for *oriC*@midcell models.

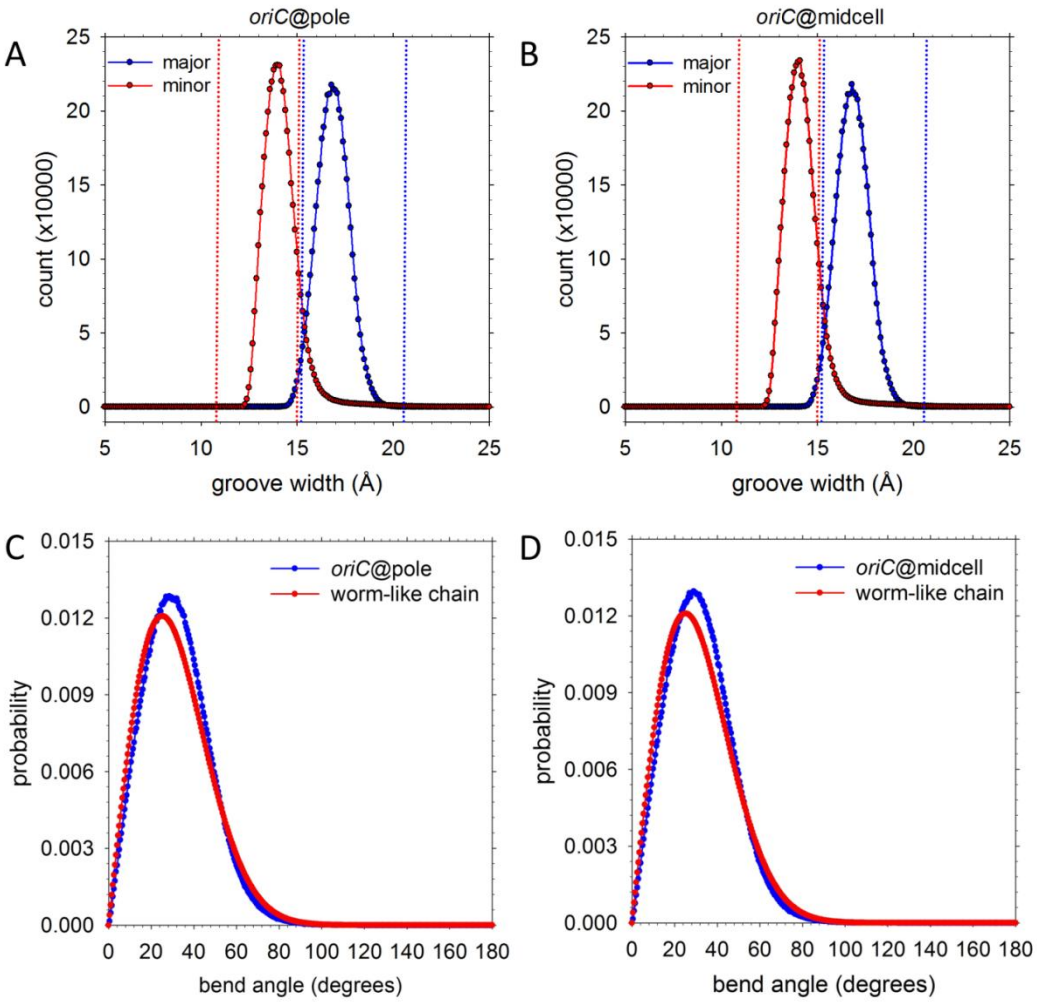

Figure S5. **Groove width and bending angle distributions. A.** Distribution of minor groove (red) and major groove widths (blue) in the most representative 1NTB structure of the *oriC*@pole models. Red dashed lines indicate the range of minor groove widths measured in NMR structures and reported by Perez et al. (65); blue dashed lines represent the same, but for major groove widths. **B.** Same as **A** but for the most representative 1NTB structure of the *oriC*@midcell models. **C.** Distribution of bending angles obtained from the most representative 1NTB structure of the *oriC*@pole models (red) compared with corresponding distribution obtained from assuming a Worm-Like Chain (WLC) model with a best-fit persistence length 496 Å (blue) (see section S3A). **D.** Same as **C** but showing results for the most representative of the *oriC*@midcell models and comparing with a WLC model distribution obtained with a best-fit persistence length of 495 Å.

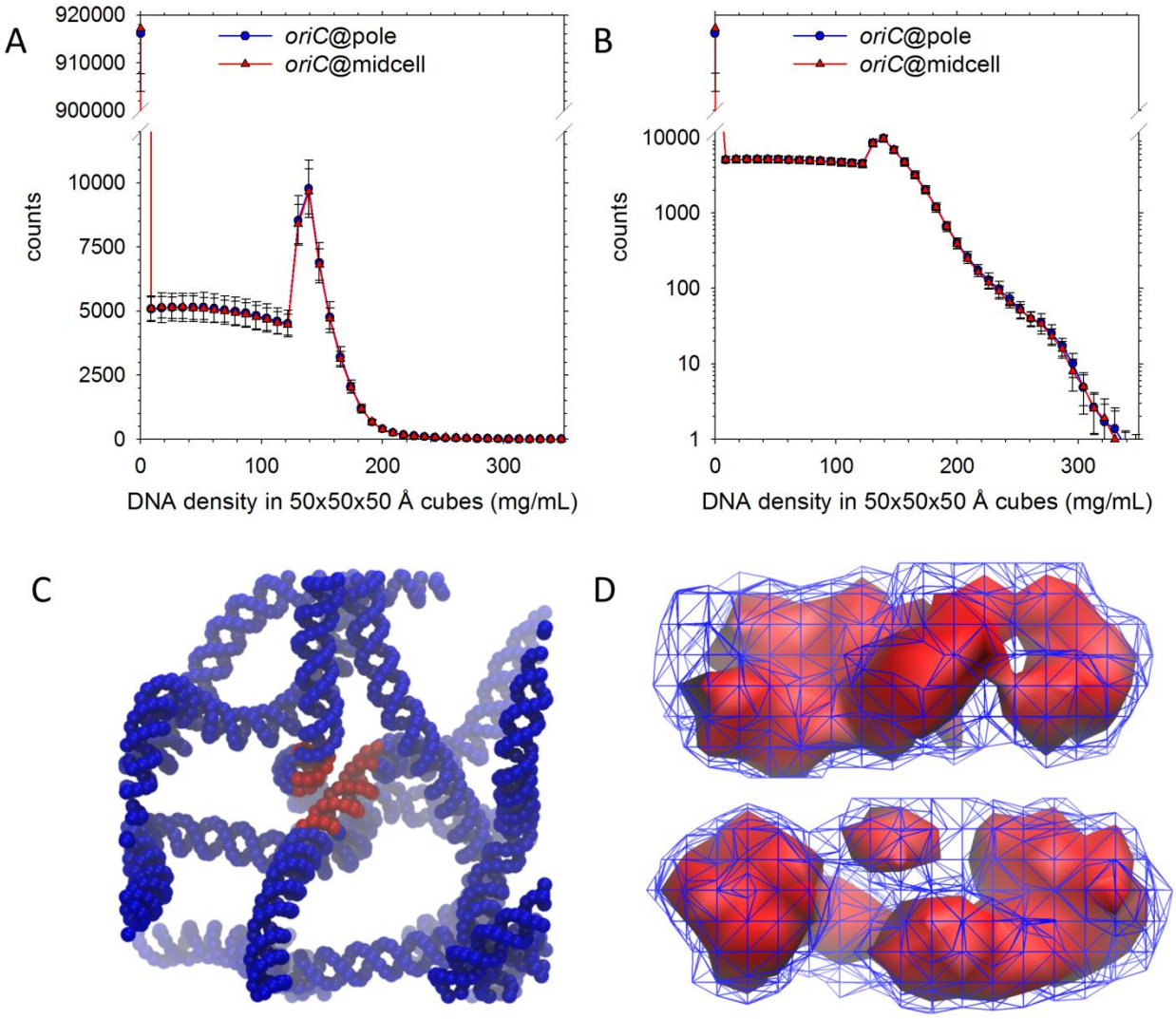

Figure S6. DNA does not fill the cell interior uniformly. A. Distribution of local DNA densities. Densities are reported for cubes of the indicated size in the *oriC*@pole (blue) and *oriC*@midcell (red) models. **B.** Same as **A**, but with frequencies plotted on a log scale. **C**. Snapshot of the region with highest density in the most representative *oriC*@midcell structure. Beads shown in red are those in the 50 Å cube with highest density (~300 mg/mL); beads shown in blue are beads occupying the neighboring cubes. **D.** (Upper panel) Isocontour plot of DNA density in the nucleoid region of the most representative *oriC*@pole structure. The blue wireframe represents regions of intermediate density while the solid red surface represents regions of high density (1.6 times more dense than the wireframe). (Lower panel): same as upper panel, but for the most representative *oriC*@midcell structure.

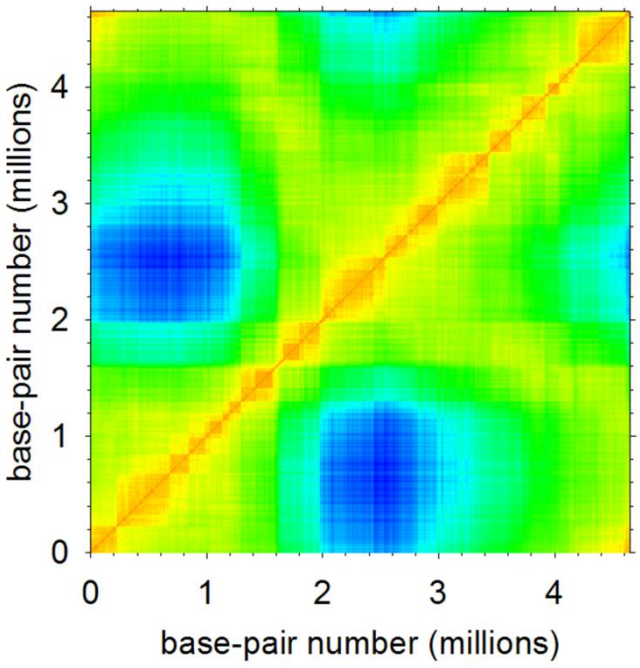

Figure S7. **Contact map for the *oriC*@midcell model**. Mean distances between chromosomal loci binned into 10 kb blocks in the *oriC*@midcell model set are plotted; coloring is as in Figure 6B of the main text, with color in bins ranging from red to blue as interlocus distance increases.

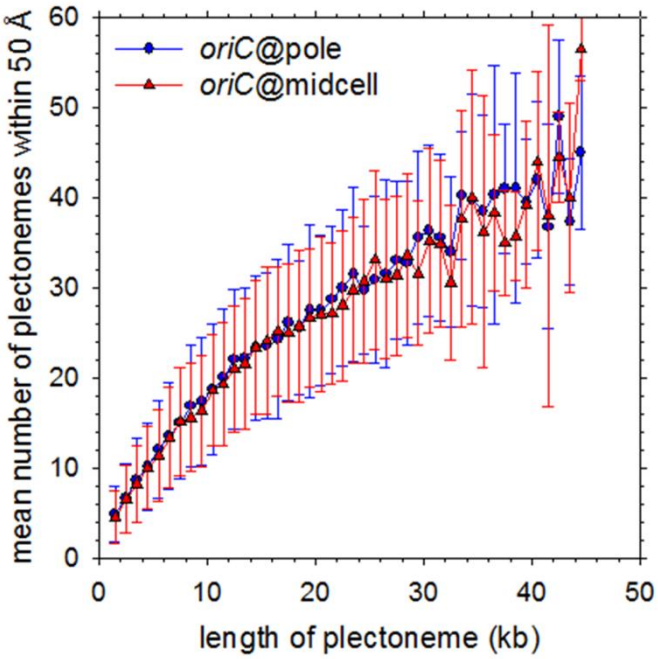

Figure S8. **Distribution of interplectonemic contacts.** The mean number of other plectonemes making close contact with a plectoneme is plotted as a function of the plectoneme's length for *oriC*@pole (blue) and *oriC*@midcell (red) models. To determine the number of contacts, basepairs closer than 50 Å and separated by more than 50 bp were identified; plectoneme affiliations of basepairs making close contact were then determined.

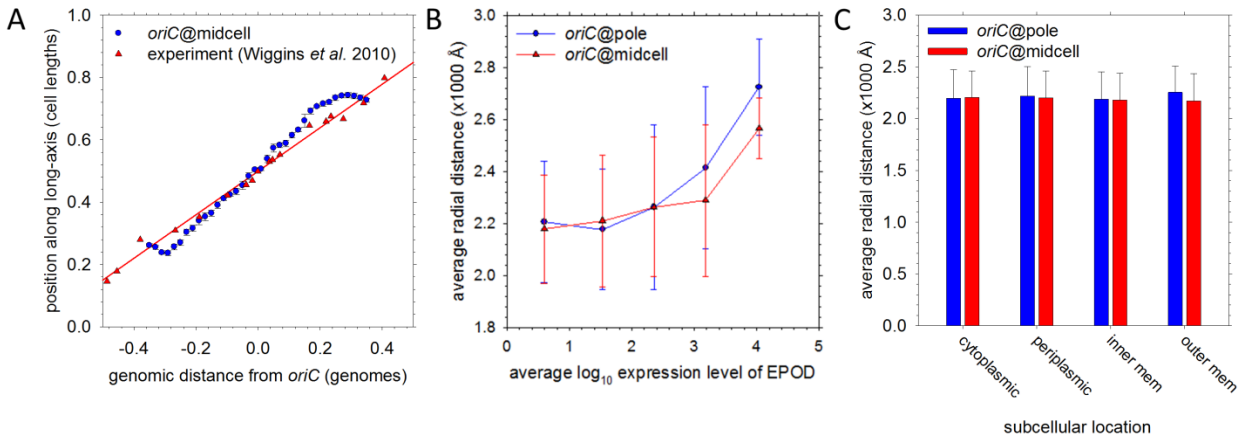

Figure S9. **Radial and longitudinal positioning of genomic loci. A.** Long-axis position varies as a function of genomic distance from *oriC* in *oriC*@midcell structures. Positions as a fraction of cell length plotted as a function of distance from *oriC* in the 20-structure set (blue) are compared with experiment ((18); red). **B.** Highly-expressed protein-rich domains are biased toward the nucleoid periphery. Radial positions of extended protein-occupancy domains (EPODs) are plotted as a function of their expression level (88); highly-expressed domains, which have not been explicitly built in to the structures, tend to be located at greater radial distances. **C.** Subcellular location of gene products does not affect the genes' radial positioning. Genes were classed by the location of their products according to EcoCyc. Mean radial distances of genes within each class were determined in the *oriC*@pole and *oriC*@midcell populations.

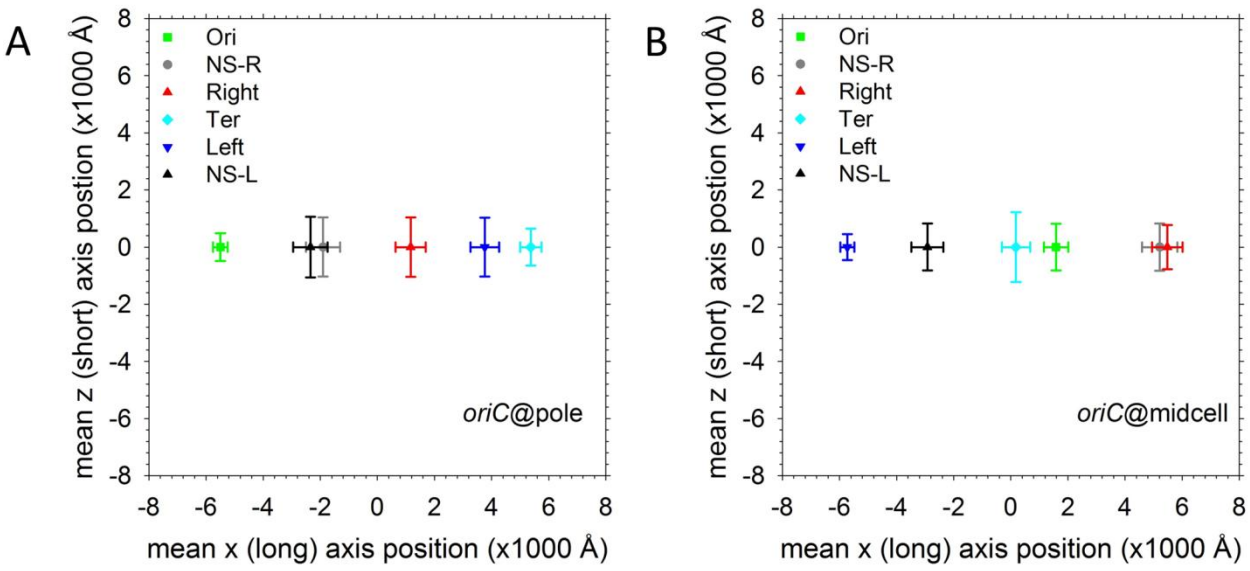

Figure S10. **Cellular positioning of macrodomains. A**. Location by region in *oriC*@pole. Mean positions of genomic regions as defined by Valens et al. (91) were calculated. To determine mean short-axis position, a mean radial position was first calculated for each model by rotating a structure through 360 degrees and projecting its radial position onto the short axis for each increment of the rotation; radial positions for all models were then averaged. Macrodomains are stably positioned across the model set, but are no more strongly positioned than non-structured regions. **B.** Location by region in *oriC*@midcell. Same as **A**, but describing *oriC*@midcell models.

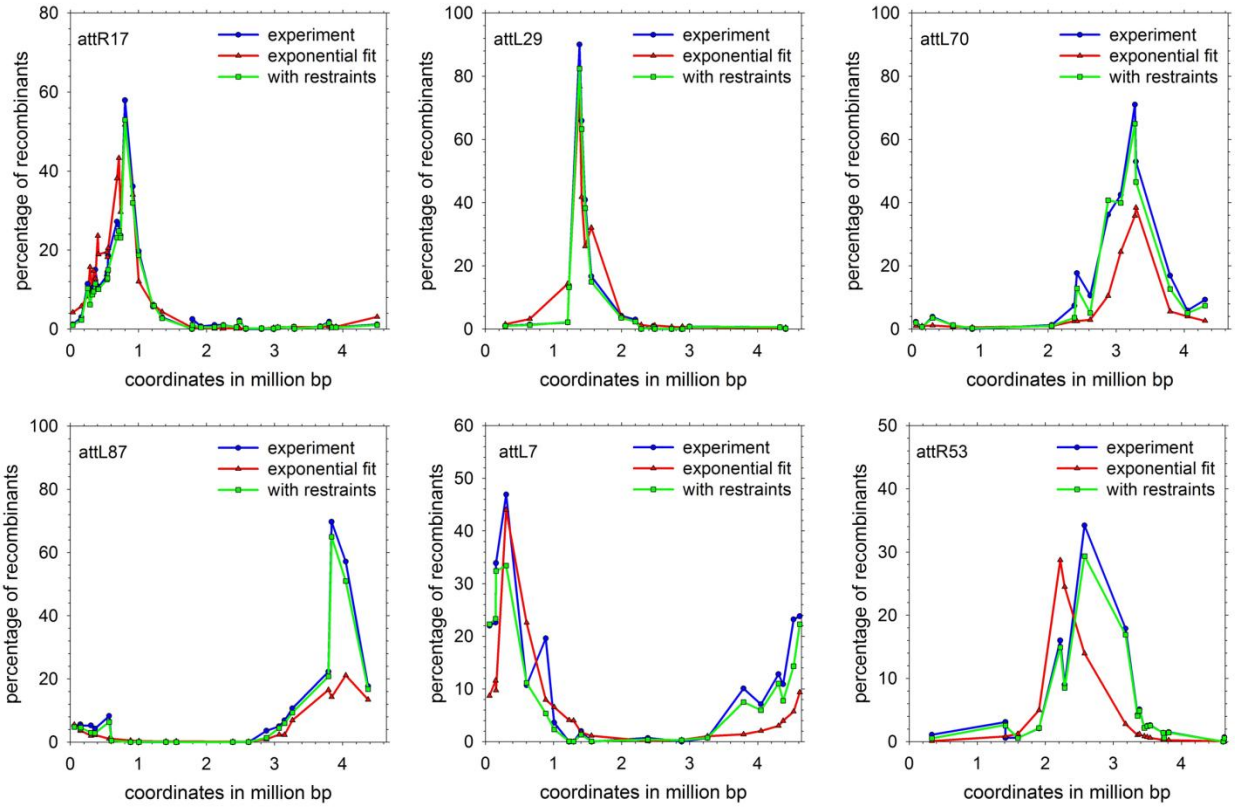

Figure S11. **Comparison of experimentally-determined and fitted recombination frequencies**. Experimental recombination frequencies appear in blue; data are reproduced from (91). Recombination frequencies for the model sets were calculated using the exponential fit described in the main text and section S4C: frequencies calculated from model distances measured *before* enforcement of interlocus distances appear in red; frequencies calculated *after* the imposition of distance restraints to match experiment appear in green. The restraints ensure that distances will be brought into line with experimental recombination rates for all tested locations of recombining elements.

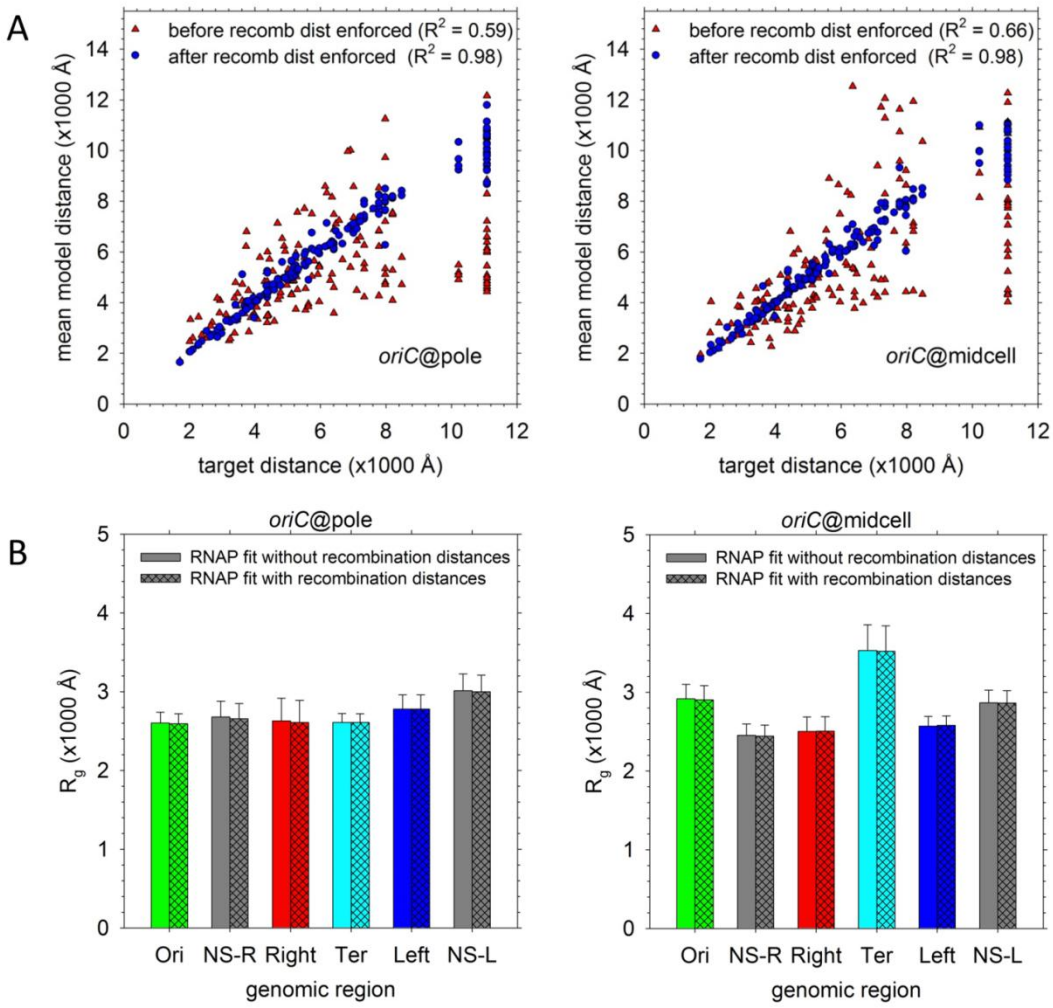

Figure S12. **Enforcing distances between recombination loci does not structure macrodomains. A**. Enforcement of distances determined by experimental recombination frequency. Target distances based on the exponential fit depicted in Figure 8B of the main text were enforced over 2500 short BD simulations. Distances before enforcing simulations appear in red; distances measured after enforcement appear in blue. Initial and final distances for *oriC*@pole are pictured in the left panel; *oriC*@midcell distances are at right. **B.** Fitting physical distances to recombination frequency does not appreciably alter structuring of macrodomains. Radii of gyration for each of the designated regions in 500 BPB model sets were calculated before (solid bars) and after (hashed bars) enforcement of recombination distances. Radii of gyration in both *oriC*@pole (left) and *oriC*@midcell (right) were essentially unchanged.

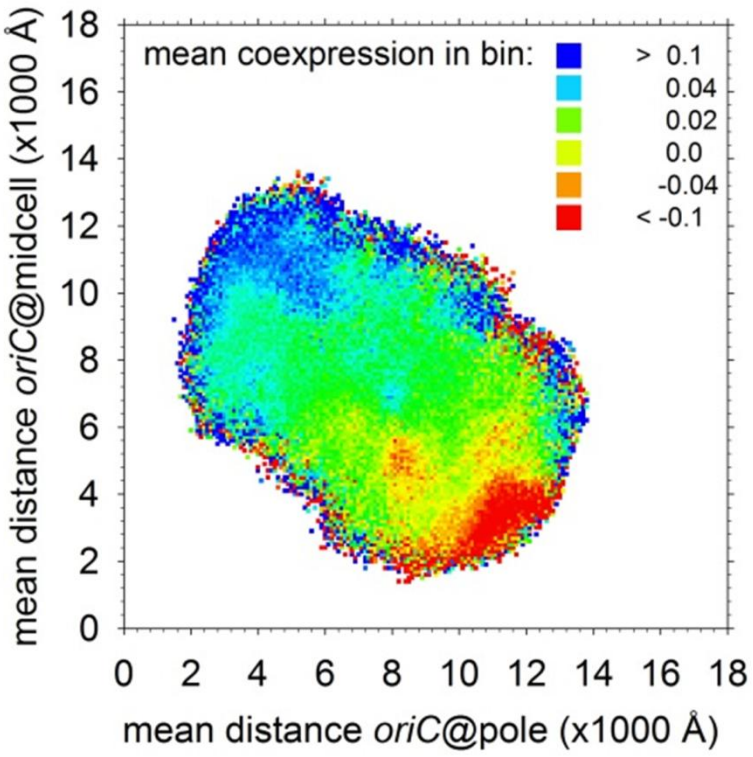

Figure S13. Expression levels of closely-apposed genes tend to be correlated in *oriC*@pole and anti-correlated in *oriC*@midcell. Genomically-distant gene pairs (separated by >1.5Mbp) were sorted into 100 Å bins by mean distance in each model set; the plot depicts the mean correlation coefficient in each xy (*oriC*@pole distance-by-*oriC*@midcell distance) bin. Gene pairs are partitioned into well-defined zones of high co-expression (blue), where genes in a pair are proximate in *oriC*@pole but distant in *oriC*@midcell, and low co-expression (red), where genes in a pair are distant in *oriC*@pole, but proximate in *oriC*@midcell.

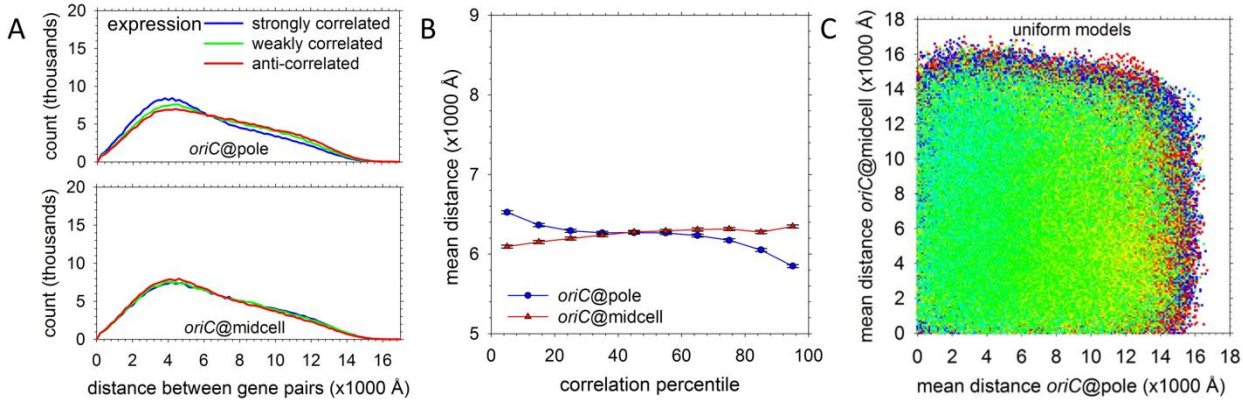

Figure S14. **Differences in relative gene positioning in *oriC*@pole and *oriC*@midcell depend on organization of the chromosome into plectonemes. A.** Mean distances between gene pairs in the top, middle, and bottom 10% of coexpression rankings, as in Figure 9A of the main text, are plotted for models in which the entire chromosome is represented as a single duplex without plectonemic supercoiling. Separation of anti-correlated genes does not occur in the nonplectonemic *oriC*@pole model (upper panel), and differences in the positioning of strongly and weakly coexpressed gene pairs are substantially diminished in both nonplectonemic model types. **B.** Same as Figure 9B of the main text, but for entirely non-plectonemic reference models. **C.** Same as Figure S13, but for entirely non-plectonemic reference models.

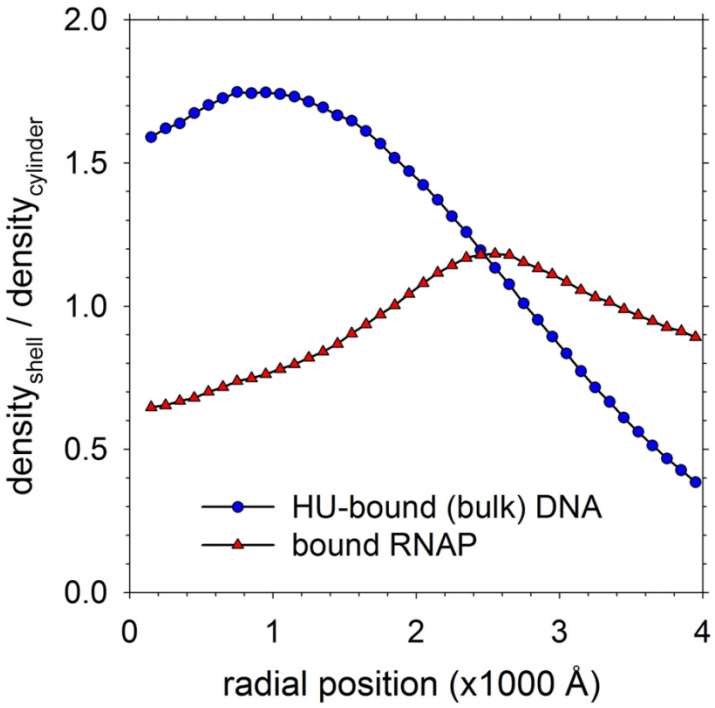

Figure S15. DNA and RNAP densities depend on radial position. DNA density (blue) and RNAP density (red) in radial shells 50 Å thick are plotted for the central 10000 Å cylinder of the nucleoid. Densities were calculated on the basis of radial target populations derived to fit experimental short-axis probability densities for DNA and RNAP (see section S2B of the Supplementary Data); density within each shell is normalized by the mean density within the central cylinder.

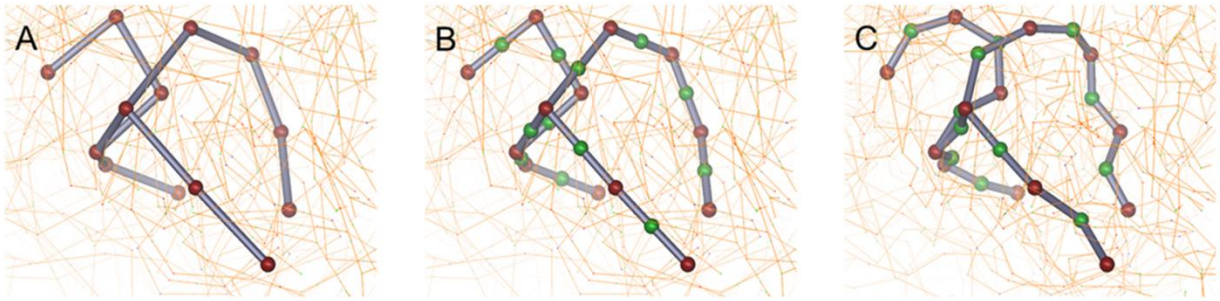

Figure S16. **Addition of intermediate beads in preparation for generating finer-grained structural models. A.** Snapshot from a 500 BPB model. **B.** Introduction of intermediate beads. Beads are added at half bond-length across the chromosome; while original 500 BPB beads at plectoneme and branch tips and bases are held fixed, bonds between the original beads are contracted to fit the mean end-to-end length of 500bp of supercoiled DNA, effectively adding "slack" to the initially fully-extended DNA represented by the span between original beads, in conformity with the persistence length of the span. **C.** Displacement of intermediate beads. Subject to available "slack" in the contracted bond, added beads are displaced by lengths sampled from a distribution determined by the persistence length of single-duplex or plectonemic DNA (54), in a random direction normal to the bond axis.

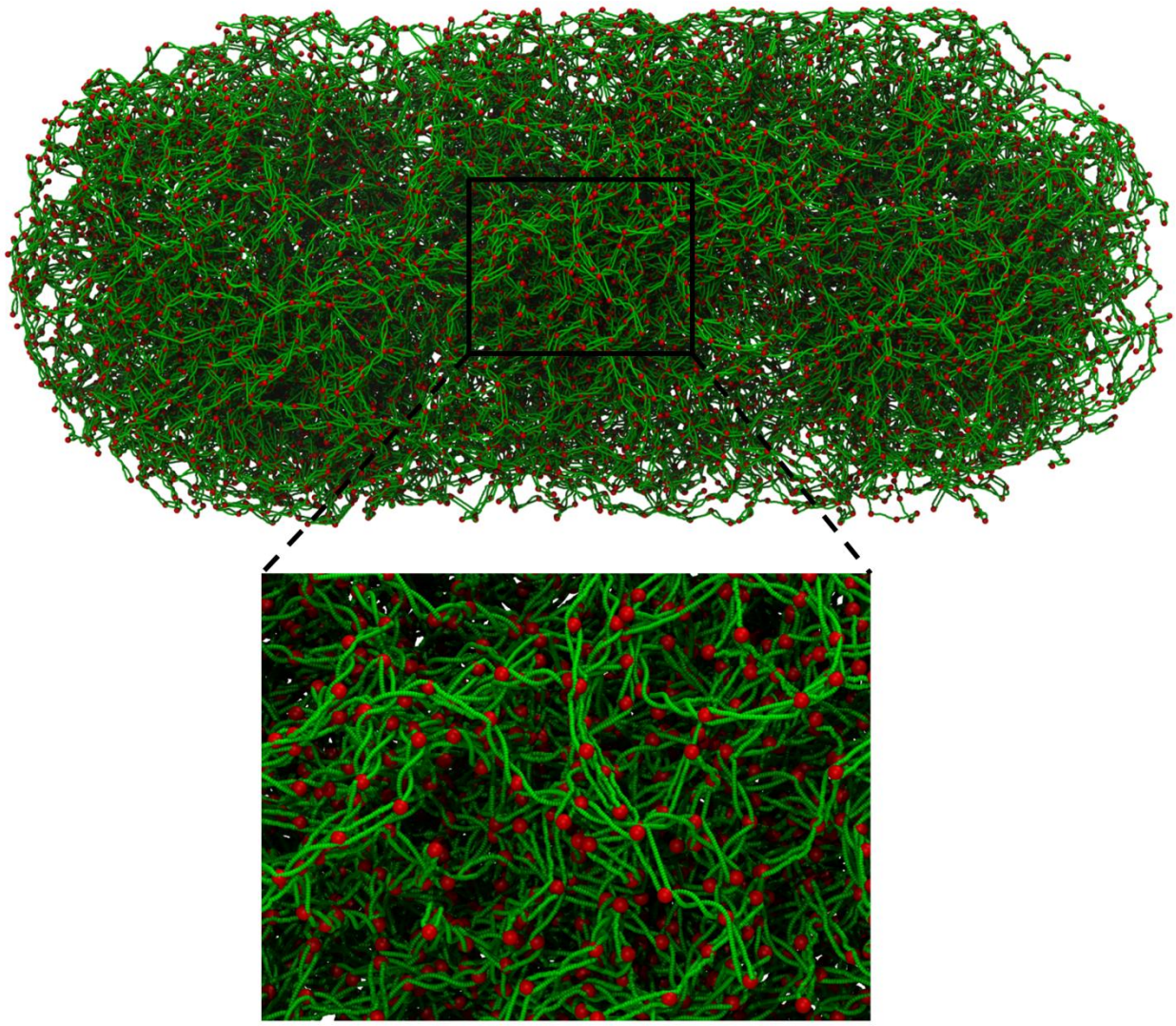

Figure S17. Correspondence of 5 BPB and (500 BPB + Intermediate bead) models. A 5 BPB *oriC*@pole structure (green) is depicted along with a coarse-grained structure (red). The coarse-grained structure is the result of the processes detailed in sections S2A-C of the Supplementary Data, in which a 500 BPB model is compacted, fitted to density distributions, and refined by the addition of intermediate beads. The close correspondence of the 5 BPB structure and the coarser-grained model indicates that morphing simulations have successfully followed the trajectory established by the 500 BPB model during compaction and that displacements analogous to those of the intermediate beads in the 500 BPB model have been accomplished at higher resolution (see section S2D for details).

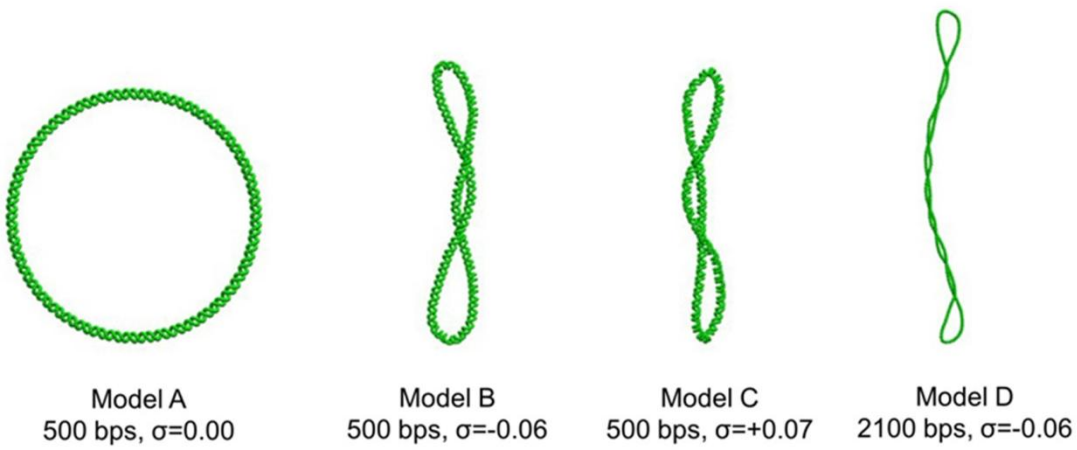

Figure S18. **Plasmid models used to verify twist and writhe calculations**. Size and specific linking difference (supercoiling density, σ) are listed for each of the plasmids used. See section S3B of the Supplementary Data for additional details.

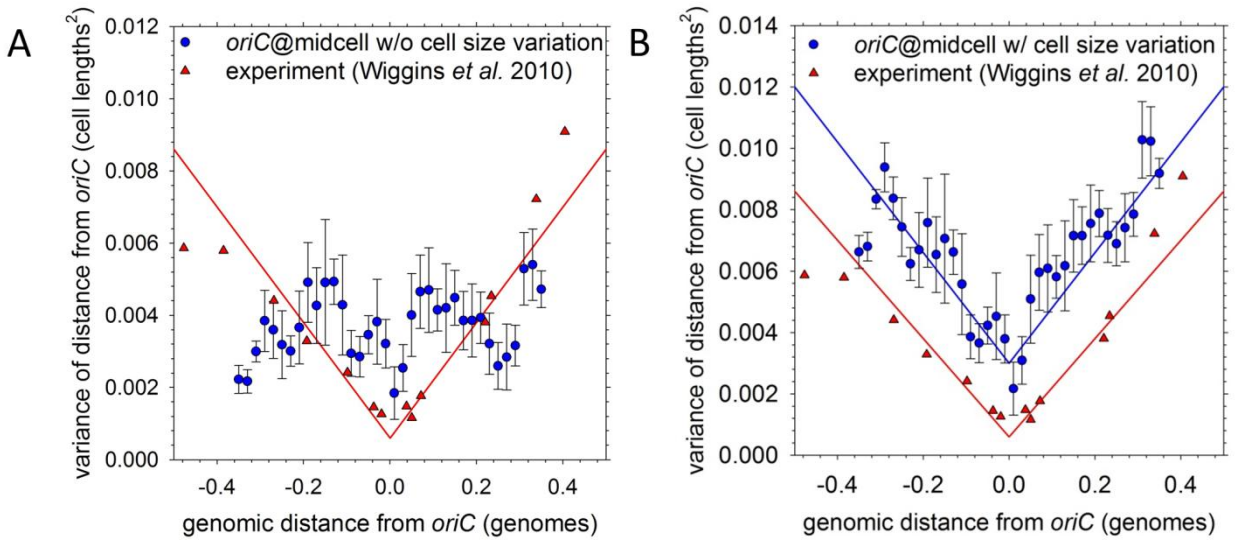

Figure S19. **Dependence of the variance of locus positioning on genomic distance from *oriC*.** **A.** Variances in the physical distances of 10 kb blocks from *oriC* (expressed in units of cell-length) in the *oriC*@midcell model, grouped and averaged according to their genomic distance from *oriC*. The plot assumes no variation in the length of the cell (2.4 μm) enclosing the chromosome (see section S4E of the Supplementary Data for additional details). Red data points and line are taken from the work of the Wiggins and Kondev groups (18). **B.** Same as A but allowing variation in the length of the cell with standard deviation of 0.5 μm; blue lines are added to guide the eye.

Model	Size (bp)	Supercoil density(σ)	Twist (Tw)	Writhe (Wr)	Tw+Wr	Linking (Lk)	Difference (%)
A	500	0.00	48.0710	0.0000	48.0710	48.0000	0.15
B	500	-0.06	46.3820	-2.4228	43.9592	44.0000	-0.09
C	500	+0.07	49.5160	2.4531	51.9691	52.0000	-0.06
D	2100	-0.06	195.7070	-7.7174	187.9896	188.0000	-0.01

Table S1. Topological properties of tested plasmids. Topological properties calculated for plasmids depicted in Figure S18 as described in section S3B above. Properties were calculated from equations reported by Clauvelin et al. ((73): twist, according to equation (6); writhe, according to equation (5); and linking number, according to equations (14), (15), and (16)).

Model	Tw	Wr	Lk	Supercoil density	Wr/ΔL
1	434346.11	-13880.20	420465.9	-0.04885	0.643
2	434387.91	-14173.20	420214.7	-0.04942	0.649
3	434329.52	-14276.30	420053.2	-0.04979	0.649
4	434312.65	-14131.30	420181.4	-0.04950	0.646
5	434280.68	-14006.20	420274.5	-0.04929	0.643
6	434323.51	-14195.40	420128.1	-0.04962	0.647
7	434263.89	-13706.70	420557.2	-0.04865	0.637
8	434326.61	-14189.70	420136.9	-0.04960	0.647
9	434356.49	-14148.80	420207.7	-0.04944	0.647
10	434359.25	-14112.80	420246.4	-0.04935	0.647
11	434287.21	-14185.00	420102.2	-0.04968	0.646
12	434304.06	-14349.00	419955.1	-0.05001	0.649
13	434328.52	-14266.40	420062.1	-0.04977	0.648
14	434278.48	-13946.50	420332.0	-0.04916	0.642
15	434312.48	-14071.60	420240.9	-0.04936	0.645
16	434334.319	-14176.30	420158.0	-0.04955	0.647
17	434304.11	-13847.90	420456.2	-0.04888	0.641
18	434334.48	-14457.50	419877.0	-0.05019	0.652
19	434307.55	-14087.20	420220.3	-0.04941	0.645
20	434258.32	-14156.10	420102.2	-0.04968	0.645
mean	434316.81	-14118.21	420198.6	-0.04946	0.646
std dev	32.14	170.89	162.3	0.00037	0.003

Table S2. **Topological properties of the final *oriC*@pole structures.** Twist (Tw), writhe (Wr), linking number (Lk), and supercoiling density (σ) are reported for each of the *oriC*@pole structures. The fraction of the change in linking number (ΔLk) contributed by writhe is listed in column 6.

Model	Tw	Wr	Lk	Supercoil density	Wr/ΔL
1	433774.31	-13424.20	420350.1	-0.04912	0.618
2	433725.05	-13071.40	420653.6	-0.04843	0.611
3	433684.69	-13026.50	420658.2	-0.04842	0.609
4	433712.60	-13177.30	420535.3	-0.04870	0.612
5	433682.11	-13069.80	420612.3	-0.04852	0.609
6	433659.29	-13003.20	420656.1	-0.04842	0.607
7	433696.36	-13161.30	420535.1	-0.04870	0.611
8	433695.72	-13290.00	420405.7	-0.04899	0.614
9	433718.85	-13228.20	420490.6	-0.04880	0.613
10	433676.85	-13039.90	420637.0	-0.04847	0.609
11	433707.57	-13236.70	420470.9	-0.04884	0.613
12	433683.47	-12847.30	420836.2	-0.04802	0.605
13	433657.98	-12688.90	420969.1	-0.04772	0.602
14	433705.50	-12823.50	420882.0	-0.04791	0.605
15	433685.36	-13296.80	420388.6	-0.04903	0.614
16	433665.56	-12865.90	420799.7	-0.04810	0.605
17	433652.55	-13091.80	420560.7	-0.04864	0.609
18	433730.67	-13228.60	420502.1	-0.04877	0.614
19	433723.91	-13312.80	420411.1	-0.04898	0.615
20	433663.84	-13114.70	420549.1	-0.04867	0.610
mean	433695.11	-13099.94	420595.2	-0.04856	0.610
std dev	29.63	183.18	166.3	0.0008	0.004

Table S3. **Topological properties of final *oriC*@midcell structures.** Twist (Tw), writhe (Wr), linking number (Lk), and supercoiling density (σ) are reported for each of the *oriC*@midcell structures. The fraction of the change in linking number (ΔLk) contributed by writhe is listed in column 6.

Region	PAR start	PAR stop	PFR start	PFR stop
1	4606402	223771	223772	230278
2	230279	331458	331459	333459
3	333460	848965	848966	850965
4	850966	984519	984520	986982
5	986983	1029418	1029419	1031418
6	1031419	1129412	1129413	1140986
7	1140987	1197588	1197589	1199588
8	1199589	1290922	1290923	1292922
9	1292923	1299566	1299567	1302812
10	1302813	1620914	1620915	1622915
11	1622916	1636335	1636336	1639335
12	1639336	1969383	1969384	1972691
13	1972692	1975329	1975330	1977330
14	1977331	1990954	1990955	1994707
15	1994708	1999585	1999586	2007643
16	2007644	2011223	2011224	2023680
17	2023681	2101396	2101397	2104396
18	2104397	2174282	2174283	2176321
19	2176322	2428719	2428720	2430763
20	2430764	2521593	2521594	2523985
21	2523986	2533763	2533764	2537237
22	2537238	2563331	2563332	2565332
23	2565333	2626694	2626695	2628938
24	2628939	2726068	2726069	2731157
25	2731158	2817783	2817784	2819783
26	2819784	2991267	2991268	2993855
27	2993856	3054865	3054866	3056988
28	3056989	3319986	3319987	3322505
29	3322506	3423423	3423424	3428762
30	3428763	3440038	3440039	3448148
31	3448149	3623398	3623399	3625515
32	3625516	3719477	3719478	3722477
33	3722478	3915553	3915554	3922051
34	3922052	3941807	3941808	3948034
35	3948035	4035530	4035531	4041637
36	4041638	4115713	4115714	4118090
37	4118091	4166658	4166659	4175026
38	4175027	4178446	4178447	4181446
39	4181447	4208146	4208147	4214433
40	4214434	4604401	4604402	4606401

Table S4. PFR and PAR locations in *oriC*@pole structures. Coordinates in basepairs are given for the forty PARs and forty PFRs in the *oriC*@pole model set.