Published online 26 February 2018

Nucleic Acids Research, 2018, Vol. 46, No. 8 ***3937–3952***
doi: 10.1093/nar/gky141

High-resolution 3D models of *Caulobacter crescentus* chromosome reveal genome structural variability and organization

Asli Yildirim[1] and Michael Feig[2,*]

[1]Department of Chemistry, Michigan State University, East Lansing, MI 48824, USA and [2]Department of Biochemistry & Molecular Biology, Michigan State University, MI 48824, USA

Received September 06, 2017; Revised February 06, 2018; Editorial Decision February 13, 2018; Accepted February 14, 2018

ABSTRACT

High-resolution three-dimensional models of *Caulobacter crescentus* nucleoid structures were generated via a multi-scale modeling protocol. Models were built as a plectonemically supercoiled circular DNA and by incorporating chromosome conformation capture based data to generate an ensemble of base pair resolution models consistent with the experimental data. Significant structural variability was found with different degrees of bending and twisting but with overall similar topologies and shapes that are consistent with *C. crescentus* cell dimensions. The models allowed a direct mapping of the genomic sequence onto the three-dimensional nucleoid structures. Distinct spatial distributions were found for several genomic elements such as AT-rich sequence elements where nucleoid associated proteins (NAPs) are likely to bind, promoter sites, and some genes with common cellular functions. These findings shed light on the correlation between the spatial organization of the genome and biological functions.

INTRODUCTION

Much is known about the gene organization within genomes, but the detailed three-dimensional (3D) structure of chromosomes has so far remained elusive. Gene–structure–function relationships at the DNA level are poorly understood as is the role of chromosomal structure in many cellular processes such as DNA transcription, replication and segregation. Two experimental approaches, fluorescence *in situ* hybridization (FISH) (1) and the recently introduced chromosome conformation capture (3C) techniques (2), especially the whole-genome sequencing variants (Hi-C) (3), have opened up new possibilities for understanding how chromosomes are folded inside the cell. FISH measures the spatial distance between two DNA segments in single cells thereby providing a direct visualization of the relative positioning of different *loci* in a given chromosome. Hi-C methods generate genome-wide contact probabilities between *loci* based on crosslinking from the population of many cells. Such crosslinking contact probabilities contain information about spatial proximities between genomic elements and, therefore, 3D structure ensembles of chrosomomes.

High-resolution structural insight of bacterial chromosomes has been derived from FISH (1,4–15) and 3C-based techniques (16–22). Bacterial chromosomes form highly compact structures that are induced by DNA supercoiling and further stabilized through the binding of nucleoid associated proteins (NAPs) (23,24). Further investigations of the global organization of bacterial chromosomes have revealed certain features of highly-organized chromosomal structure (6,8,25). In *Caulobacter crescentus*, the origin and terminus of replication are located at opposite poles of a longitudinally organized chromosome. This is referred to as an *ori-ter* configuration (6). This configuration has also been confirmed from 3C experiments on the *C. crescentus* genome (17,21). Chromosome Conformation Capture Carbon Copy (5C) study reported an ellipsoidal chromosome with periodically arranged arms, further confirming a longitudinal organization (21). Moreover, a Hi-C study on the same bacterium revealed that the *C. crescentus* chromosome consists of multiple chromosome interacting domains (CIDs) with highly expressed genes located at domain boundaries (17). On the other hand, the chromosome of *Escherichia coli* contains four macrodomains (Ori, Ter, Left and Right) with their localization being dependent on different stages in the cell cycle (1). Ori and Ter domains were also identified from 3C-based studies (16). In the chromosome of slow-growing *Escherichia coli* cells, the origin and terminus of replication are located at mid-cell locations whereas the right and left chromosomal arms reside in separate cell halves (5,8). This has been termed a left-*ori*-right pattern. On the other hand, the chromosome adopts an *ori-ter* configuration in fast-growing *E. coli* cells as these macrodomains localize at opposite poles of the cell

*To whom correspondence should be addressed. Tel: +1 517 432 7439; Fax: +1 517 353 9334; Email: feig@msu.edu

(9). The *Bacillus subtilis* chromosome is organized into a similar configuration alternating between *ori-ter* and left-*ori*-right patterns depending on the cell cycle phase as in *E. coli* (7,26). CIDs have also been observed in the *B. subtilis* chromosome (19,22).

NAPs play a crucial role in the observed organization of bacterial chromosomes (24,27). One of the most abundant NAPs is H-NS which is prominently found in *E. coli*. H-NS has been reported to bridge different segments of the genome (28,29) and control supercoiling of the DNA (30). This suggests an active role in nucleoid organization. HU, another ubiquitous NAP in bacteria, has been found to influence chromosome compaction by wrapping the DNA around itself analogous to histones in eukaryotes (17,27). Integration host factor (IHF) and Fis (factor for inversion stimulation) proteins also lead to chromosome compaction by introducing DNA bending (24,27). Additionally, SMC (structural maintenance of chromosomes) proteins are believed to contribute to chromosome compaction (27) but have recently been shown to affect collinearity of chromosomal arms in *C. crescentus* rather than regulating overall compaction (17).

Since bacterial genomes are not segregated into a special compartment as in eukaryotic cells, they occupy a large portion of the cell and interact extensively with the intracellular environment. Therefore, the spatial organization of bacterial genomic DNA can play a major role in the regulation of biological functions in the cell. This idea is affirmed by a recent study where the statistical analysis of genome conformation capture data for *E. coli* suggested that operons from the same regulons and genes in the same biological pathway tend to be close to each other in 3D space, thereby maximizing compactness of the genome (31). Furthermore, genes that are spatially close in 3D genomic structure tend to be co-expressed and their protein products are prone to form more protein-protein interactions (31). The relationship between gene co-regulation and the 3D organization of genes was also explored in a computational study where the chromosome was modeled as a worm-like polymer chain with interacting sites corresponding to genes that are regulated by same transcription factors (32). From this work, it was found that chromosomes form different topological structures that increase the local concentration of interaction sites and, therefore, co-localize the co-regulated genes in 3D space (32). A very recent Hi-C study on the *Mycoplasma pneumoniae* chromosome has provided the first direct evidence of a correlation between 3D chromosome organization and transcriptional regulation (20). They found that genes that are in the same CID have higher co-expression levels than genes located in different domains (20). In addition, independent of an organization into CIDs, this study more generally reported high co-expression levels for spatially close genes (20). Despite this recent progress, a more detailed understanding of the structural organization of bacterial chromosomes is still lacking.

The data generated with 3C-based methods lends itself to experiment-driven modeling of 3D chromosome structures using restraints defined from 3C studies. Initial modeling approaches have used the generally accepted assumption that 3C-based contact probabilities are inversely related to the average distances between *loci* pairs (33). Further calibration is possible by comparing 3C-based contact probabilities with average distances obtained via FISH. Such a calibration curve was obtained for *C. crescentus* by comparing contact probabilities based on Chromosome Conformation Capture Carbon Copy (5C) for 112 different *loci* from the flagellated pole of swarmer cells with FISH data (6). This allowed for a direct conversion of the 5C contacts to distances, which could then be used as constraints during modeling. Initially, such an approach based on 5C data was used to generate models for the *C. crescentus* chromosome at 13-kilobase (kb) resolution which provided first insights into its spatial organization in 3D space (21). Restraint-based modeling based on contact probabilities between *loci* pairs has also been applied to generate 3D structures of eukaryotic chromosomes or their subsections (34–41).

The generation of chromosome models simply based on satisfying restraints from 3C contact frequencies that are converted to distances via calibration against FISH data is seemingly straightforward, but it has been pointed out that this approach is problematic (33,42–46). 3C-based contacts stem from crosslinks that can only form if two *loci* come within a certain contact threshold. This means that a given contact frequency for a certain pair of *loci* only reflects in what fraction of cells those *loci* come closer than the contact threshold, instead of directly reporting on the average distance between two *loci* as FISH does when averaged over many cells. In other words, 3C-based studies only give information about the low-end of the distance distribution, while FISH considers the mean of the entire distribution. This has prompted efforts to employ population-based modeling techniques to generate ensembles that reflect cell-to-cell variability and explicitly consider the short-range sensitivity of 3C-based methods when matching the cumulative contact map from the experimental contacts (47–50).

Another issue is that the experimental data, even when Hi-C techniques are employed, is too sparse to fully determine 3D structures beyond kb resolution. This leaves an important role to computational methods to compensate for a lack of resolution in the experimental data by including general topological features and packing constraints as part of the model building protocol. The rationale for such a strategy is similar to the well-established protocols for the determination of macromolecular structures based on restraints from nuclear magnetic resonance (NMR) data. In the case of NMR, such assumptions, e.g. about peptides being polymers with a backbone and side chains with certain molecular bonding geometries are key to obtaining atomic resolution structures from data that is otherwise effectively at much lower resolutions. In the case of bacterial chromosomes, one can apply knowledge about a plectonemic structure made up of supercoiled segments with branching points and long persistence lengths to generate high-resolution models when combined with the experimental data. Le *et al.* generated 3D structures for the *C. crescentus* genome at 434 base pair (bp) resolution by modeling the chromosome as a circular polymer consisting of plectonemes (17). In that work, the resulting structure ensembles that best fit the Hi-C data were selected after varying model parameters (17). The resulting models further extended the knowledge about the 3D organization of the *C. crescentus* chromosome initially gained by Umbarger *et*

al. (21) and shed light on the presence of CIDs and their organization (17).

Using a different strategy, Hacker *et al.* recently described a model of the *E. coli* chromosome at nucleotide resolution (51). Starting with a multi-scale polymer model that captures the plectonemic topology and physical properties of double-stranded and supercoiled DNA, spatially resolved models were generated primarily based on RNA polymerase (RNAP) binding data from ChIP-chip experiments (52). RNAP binding sites obtained from the ChIP-chip data were used to identify highly-transcribed regions which were then modeled as plectoneme-free, i.e. not supercoiled, regions in the chromosomal structure. The modeling was further guided to match the distributions of the beads that mapped to RNAP binding sites to the projected 2D distribution of RNAP as well as the distribution of the rest of the beads to the 2D distribution of HU proteins obtained from single-molecule fluorescence experiments for *E. coli* (53). These models allowed, for the first time, the ground-breaking investigation of physical properties a bacterial chromosome at the nucleotide level and a direct mapping of genome sequence to structure.

In this study, we employed a similar multi-scale modeling protocol to encode the plectonemic and supercoiled topology of bacterial DNA via coarse-grained (CG) models at different resolutions up to the bp level but using 3C-based contact frequency matrices to guide the modeling. This protocol was used to generate structural ensembles for the *C. crescentus* chromosome, where extensive Hi-C data (17) is available. The Hi-C data provides direct information about relative spatial distances between *loci* under the consideration of dynamics and population-based variations. Similar to the models generated by Hacker *et al.* for *E. coli*, the models presented here for *C. crescentus* are thus believed to provide an accurate picture of where gene *loci* are located in individual chromosome structures and how such distributions vary between cells and as a result of chromosome dynamics. As the models from Hacker *et al*., the models generated here also allowed a direct mapping of the genomic sequence onto the generated 3D structures and a detailed analysis of how the mapping of genomic sequences onto the bacterial chromosome structures may be related to biological function encoded by the corresponding genes.

MATERIALS AND METHODS

Experimentally driven high-resolution models of the *C. crescentus* genome were generated as a hyperbranched polymer of supercoiled DNA segments forming plectonemic rosettes. The multi-scale modeling protocol is illustrated in Figure 1 and described in full detail in the Supplementary Data. Briefly, initial Monte Carlo (MC) sampling of a segment-based plectonemic model was guided by distance restraints based on Hi-C interaction frequencies (17) between specific pairs of *loci* (Supplementary Video 1). The distance restraints were derived by using a calibration curve obtained by Umbarger *et al.* (21) which is a polynomial function that maps between interaction frequencies from 3C-based data and expected distances. Simply put, such mapping was possible by comparing the available average spatial distances of *loci* pairs measured by fluorescence microscopy data for the *C. crescentus* chromosome with the corresponding interaction scores for pairs with similar genomic site-separations.

In the initial sampling round, the general topology of branches extending from a central ring was fixed, but branches were allowed to reconnect and move. Next, 15-bp CG models were constructed by wrapping higher resolution beads around the segments in the initial plectonemic models. The 15-bp CG models were then further refined via molecular dynamics (MD) simulations using a CG interaction potential that accounts for the elastic properties of DNA. An example of a *C. crescentus* model at 15-bp resolution is shown in Figure 2. In total, 1050 models were generated covering a range of branch segment lengths and number of supercoiled loops (microdomains) (54). The resulting models were subsequently reweighted to generate an ensemble of structures where not just the average distance between two *loci* but also the distribution of contacts in the models is maximally consistent with the Hi-C scores. The reweighted ensemble was then used for all further analysis.

The final models were also used to reconstruct models at base-pair resolution by taking advantage of the long persistence length of double-stranded DNA (see Supplementary Data). The base-pair resolution model for the structure shown in Figure 2 is given in Supplementary Figure S1. Projections of the beads in the bp-resolution models and 15-bp resolution models did not show any difference (Supplementary Figure S1B versus Figure 5A, Pearson's coefficient: 1.00, Slope: 1.00, Intercept: 0.00). Therefore, 15-bp resolution models were used for further analysis throughout this study.

RESULTS

Structural characterization of *C. crescentus* chromosome models

An ensemble of chromosome models for *C. crescentus* was generated as described above and in more detail in the Supplementary Data. To characterize the ensemble, we initially compared contact maps with the experimental data. The contact map based on the number of contacts within a distance threshold (see Supplementary Data) is shown in Figure 3. This map compares to the Hi-C scores and is in good agreement with the Hi-C contact map with a Pearson's correlation coefficient of 0.88 (Slope: 1.62, Intercept: 0.04) (Figure 3). Both contact maps exhibit the same characteristic two diagonals that reflect two chromosomal arms in an ellipsoidal shape interacting with each other (17,21).

Previously, the inspection of the Hi-C interaction map revealed that the *C. crescentus* chromosome is organized into 23 CIDs, which appear as triangles along the main diagonal (17). The boundaries of the CIDs were identified by comparing the interaction preferences of *loci* from its left- and right-hand side. This definition is based on the idea that a *locus* at the border of a CID would strongly interact either with its left- or right-hand side, whereas a locus on the middle of CID interacts similarly with loci on both sides (17). Although the triangles corresponding characteristic of strong CID formation are less apparent in our contact map, the analysis of the interaction preferences of *loci* (Supplementary Data) was carried out in the same manner as was

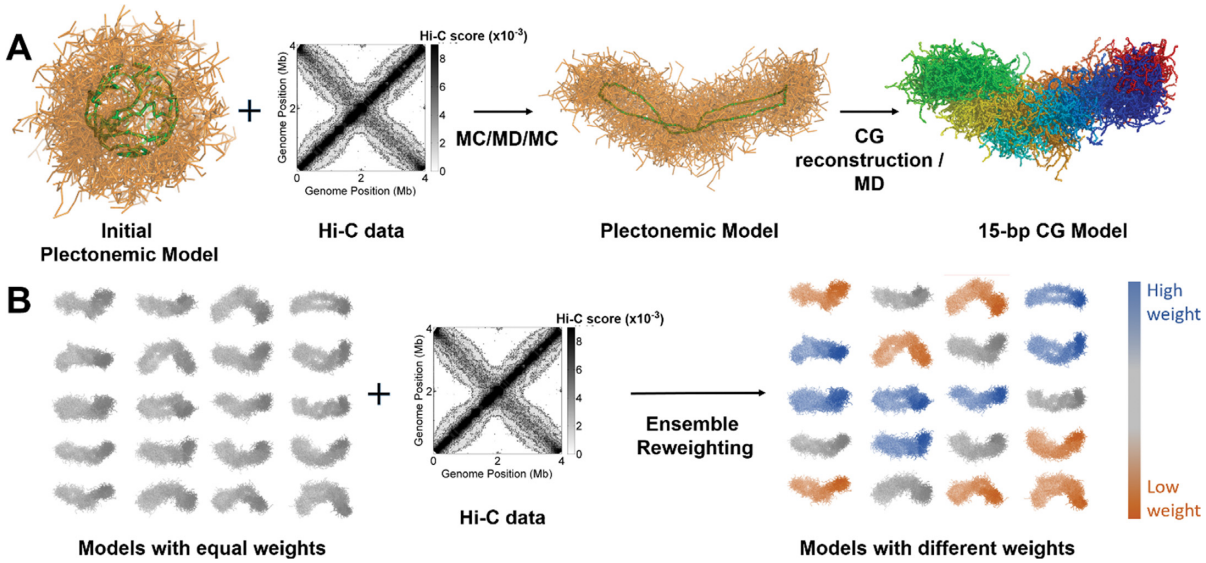

Figure 1. Multi-scale modeling procedure during model generation based on Hi-C data. (**A**) Plectonemic and CG model generation. (**B**) Model reweighting.

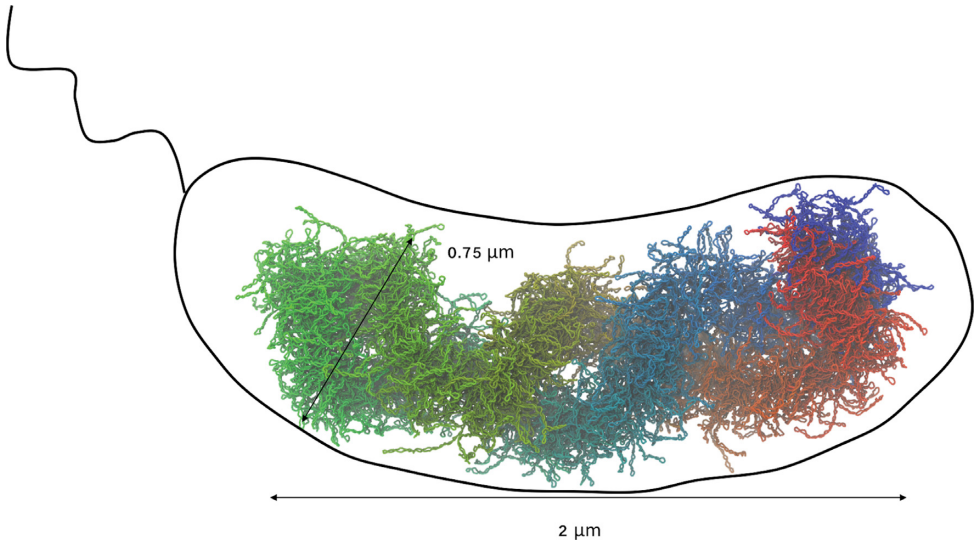

Figure 2. 3D structure of *C. crescentus* chromosome at 15-bp resolution projected onto a *C. crescentus* cell with typical dimensions.

done previously (17). The results are also shown in Figure 3. We note, that the procedure involved averaging over 100 kb sections left and right from each potential boundary site, so that some of the oscillations in the directional preferences, such as the ones between genomic positions 200–250 kb, did not lead to predicted CID boundaries. We found 21 CIDs, 18 of those match the CIDs identified previously from experiment. These results further confirm that our models are consistent with the information gained from the Hi-C experiment.

The average distance contact map at 10-kb resolution from the models is also shown in Supplementary Figure S3B and compared with the distance contact map converted from Hi-C interaction frequencies (17) by using the calibration curve derived by Umbarger *et al.* (21) (Supplementary Figure S3A). The average distance map from the models shows an excellent agreement with the Hi-C derived map with a Pearson's correlation coefficient of 0.98 (17,21). One significant difference between the two distance maps is that there are apparent contacts between the origin (at 0 Mb) and the middle of the genome (at 2 Mb) in the experimental map. In the 3D organization of the chromosome, the middle of the genome (2 Mb) resides at the opposite pole from the origin. Therefore, these interactions are unlikely. The Hi-C scores for those regions were below the cutoff for contacts to be considered significant and likely result from replication and segregation of the origin to the opposite pole in some cells during the experiment (21). These contacts were not considered in our modeling protocol, but if they had been included, a large fraction of highly bent structures would be necessary to satisfy average distances of ∼650 nm between the opposite poles of the genome structure. Such bent structures are not consistent with the typical cell dimensions of *C. crescentus* (see below).

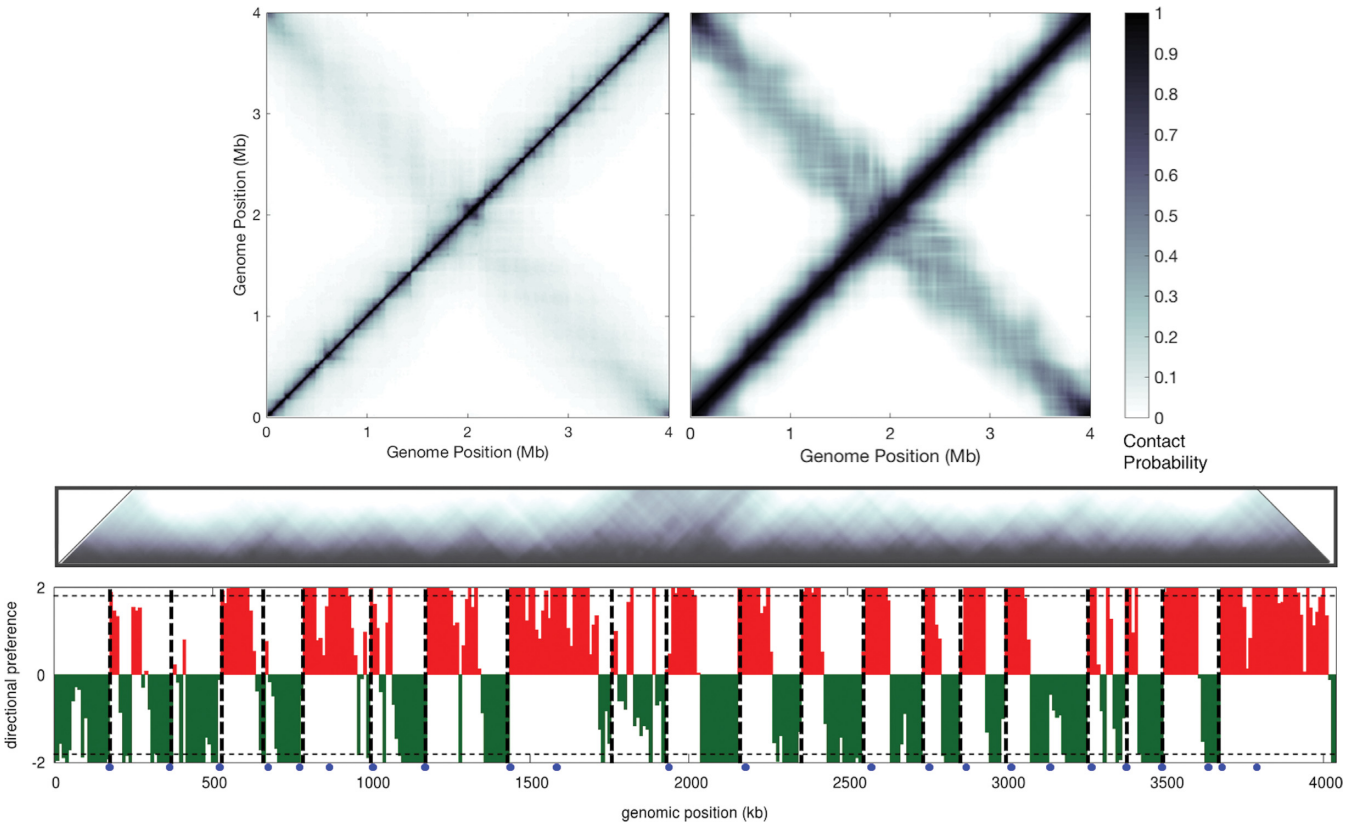

Figure 3. Contact maps. (**A**) Hi-C score map (left panel) from Le *et al.*(17), average contact map of models by using cross-linking probability function used in the reweighting procedure (right panel). Hi-C scores were normalized to 1 for comparison. (**B**) The main diagonal of the average contact map from models rotated 45° clockwise. The bottom panel shows the directional preferences of different *loci* (green: left-, red: right-hand side preference). Dashed lines show the boundaries of CIDs found from our models, blue dots indicate CID boundaries found in the study by Le *et al.* (17).

The intramolecular distances can be further compared with recently reported distance measurements between 51 fluorescent-labeled DNA segments in the *C. crescentus* chromosome by Hong *et al.* (55). We note that these distances were not used in the modeling or for generating the calibration curve between Hi-C interactions scores and spatial distances. Therefore, a comparison of the models with this data provides important independent validation (33). The comparison of the projected average distances from the models with the experimental distance measurements is shown in Figure 4A. It can be seen that the models generally reproduce the experimental distances (Pearson's correlation coefficient: 0.99, slope: 1.38, root-mean-squared-error: 39.42). For longer distances, the models have somewhat larger values than the experimental data suggesting an overall slightly too expanded shape. When only distances of DNA segments from the origin of replication are compared, the models produce similar distances as reported from FISH up to 500 nm but become more extended for larger distances compared to the available FISH data (6,55). This may be a consequence of modeling the DNA based on the Hi-C experiments that are insensitive to contacts beyond ∼485 nm (see Supplementary Data). This threshold stems from the fact that 3C-based experiments result in a flat distributions of interaction frequencies with long tails that correspond to the fragments that are in contact very frequently or infrequently. However, intermediate interaction frequencies are relatively noisy. Umbarger *et al.* have used the ∼485 nm threshold to distinguish frequent contact frequencies from less reliable intermediate contacts (21).

Another explanation for the different dimensions of the chromosome structures could be a difference in the phases of the cells that were studied in the FISH and Hi-C experiments. FISH experiments were based on swarmer cells that were not allowed to grow (6,55), whereas the Hi-C data were collected from cells that were allowed to grow for 0, 10, 30, 60 or 75 min (17). Growing cells exhibit larger cell sizes (56) and, hence, allow for more extended chromosome structures.

In order to test if our models would still be compatible with the Hi-C data if they had overall smaller dimensions consistent with the FISH data, we scaled the representative models of clusters obtained from the clustering analysis (see Figure 7) by 0.8 followed by short (50 ns) refinement via MD to relax the models. The resulting models reproduce the observed correlation between distance and site-separation of *loci* pairs from experiment (orange curve in Figure 4B) and, as expected, also match the FISH data better (Pearson's correlation coefficient: 0.98, slope: 1.09, root-mean-squared-error: 15.73) (Figure 4C). The resulting average distance map after scaling is still in good agreement with the experimental data, but the Pearson's correlation coefficient of 0.96 with respect to the Hi-C data is slightly worse than for the unscaled models (Supplementary Figure S2C).

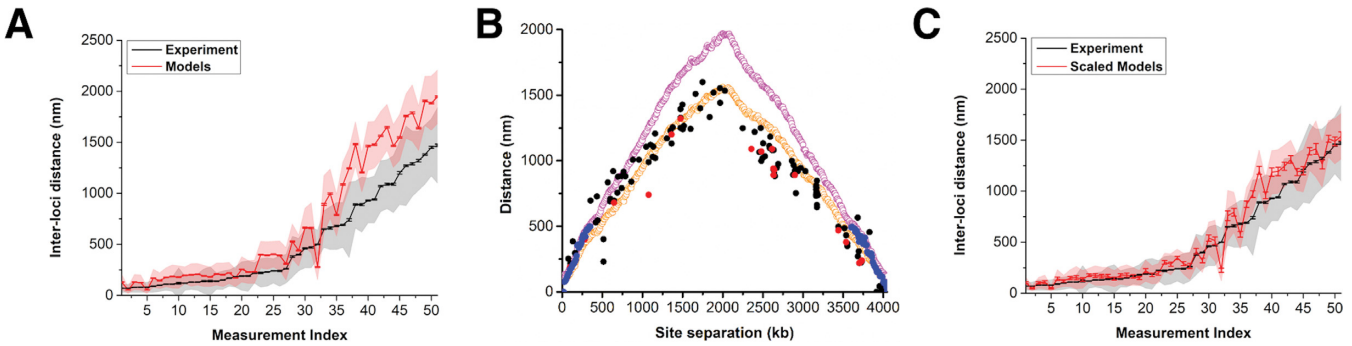

Figure 4. Compatibility of the models with experimentally measured 51 distances from FISH in the study by Hong *et al.* (55) (**A**) Distances for *loci* pairs from Hong *et al.* and from the models (Measurement index is from 1 to 51 referring to the each measured distance in the experiment). (**B**) Distance between the origin of replication and different DNA segments with different genomic distances from Viollier *et al.* (6) (black), Hong *et al.* (55) (red), models (pink), and scaled models (orange). The distances obtained from the calibration curve and used as restraints during the modeling are shown in blue. (**C**) Distances for *loci* pairs from Hong *et al.* and from models scaled by 0.8. The shaded areas indicate the standard deviations and the error bars indicate the standard errors based on the ensemble averages.

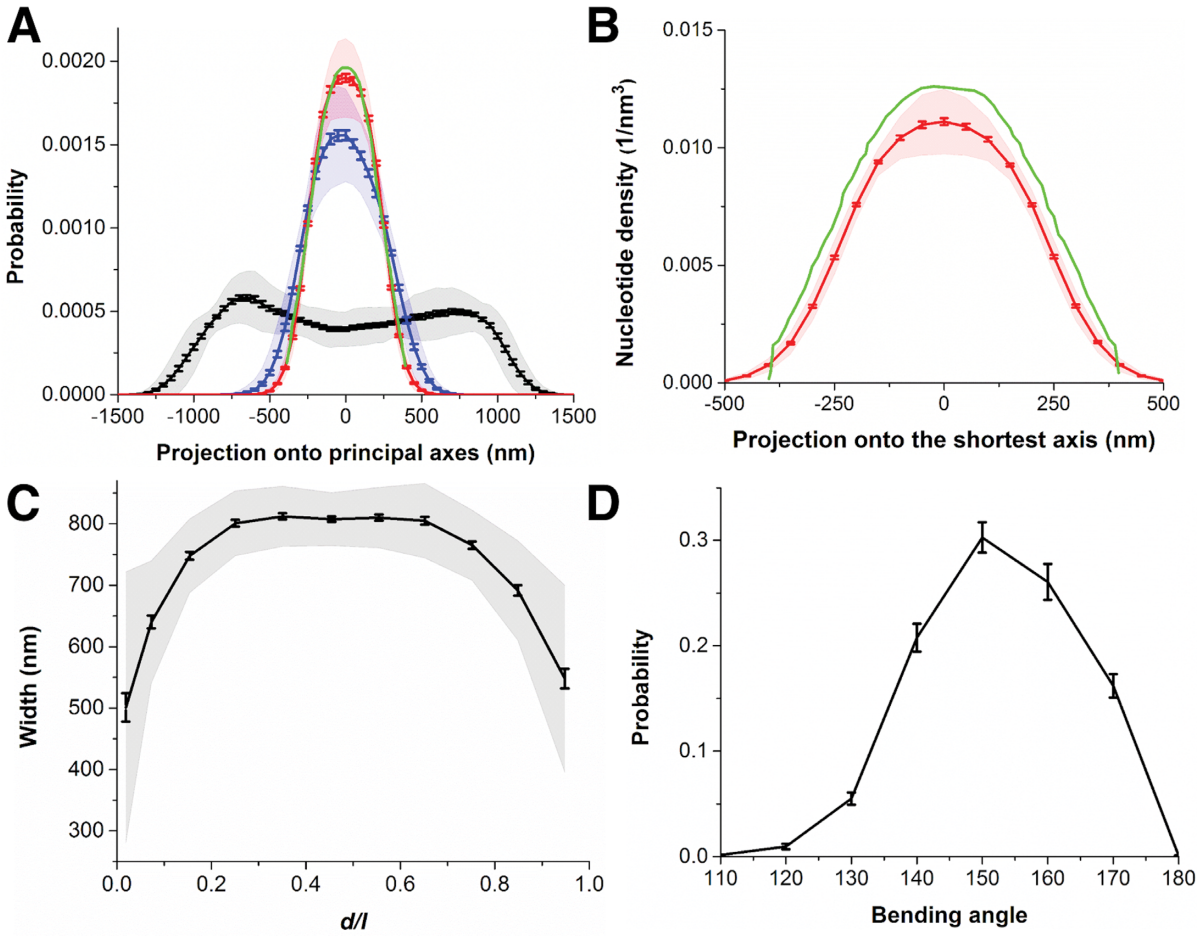

Figure 5. Dimensions of the models. (**A**) Projections of beads in the models onto their longest principal axis (black), the second principal axis (blue), the shortest principal axis (red) and the DNA distribution based on HU proteins in *E. coli* nucleoid short axis reproduced by digitizing Figure 4A-ii in the article from Stracy *et al.* (53) (green). In order to compare the *E. coli* data with our projections, the relative x-axis values given in the experimental results were multiplied by half the *C. crescentus* width (400 nm), and the curve was normalized again to keep the area under the curve 1. (**B**) Nucleotide density projection onto the shortest axis from our models (red) and 3D models of *E. coli* chromosome (51) (green) reproduced by digitizing A in the article from Hacker *et al.* (51). The probability densities of DNA in *E. coli* models in the reference were converted to nucleotide densities by multiplying the probability densities by the total number of nucleotides in *E. coli* and divided by the width and length of their models. (**C**) Width of the models vs. the positions along the longest principal axis, *p,* normalized to the length of the models along the x-axis, *l*. Here, 1 represents the pole that contains the replication of origin while 0 represents the opposite pole. (**D**) Distribution of bending angles for the models. The shaded areas indicate the standard deviations and the error bars indicate the standard errors based on the ensemble averages.

Next, we analyzed the spatial dimensions of our models. Projections of the beads in the models onto their principal axes are shown in Figure 5A. The average length of the core structure is around 2 μm, with the edges extending to ∼2.5 μm. This is slightly less than the experimentally observed *C. crescentus* cell lengths of ∼2.5–3 μm (56). The projection onto the short axes reflects the curved shape of most of our models. Therefore, we also determined the width perpendicular to the local axis of a curved line fit to the nucleoid models (see Supplementary Data). The resulting widths shown in Figure 5C are ∼0.7 μm with a maximum extent to ∼0.8 μm. For comparison, experimental cell widths are ∼0.8 μm (56). Therefore, our models fit just inside the known dimensions of *C. crescentus* cells without imposing such a requirement during the modeling protocol. If the models are scaled to better fit the FISH data (see above), the dimensions of the chromosome become somewhat smaller (Supplementary Figure S3). In either case, our results imply that the genomic DNA fills out the majority of the cellular volume with the outermost parts of the DNA able to come close to the cellular membrane. These results are consistent with previous work showing the level of nucleoid compaction in different bacteria where nucleoids have been imaged by electron microscopy (57,58). Although the morphology of the nucleoid depends on the organism as well as environmental conditions and growth rate (58,59), nucleoids are generally found to occupy a large fraction of the cell cytoplasm (57,58,60). In particular, the distribution of the nucleoid-associated HU protein in *C. crescentus* suggests that the nucleoid exhibits a diffuse morphology that extends to most of the cell volume (60). Interestingly, the distribution of HU proteins in *E. coli* obtained from single-molecule fluorescence data (53) as well as the distribution of DNA in 3D models of *E. coli* nucleoid (51) are quite similar to the DNA distribution on the shortest axis of the *C. crescentus* nucleoid models (Figure 5A, green curve). This indicates that the level of nucleoid compaction for *E. coli* and *C. crescentus* are similar. We also compared the nucleotide densities from our models and from the 3D models for the *E. coli* nucleoid (51) (Figure 5B). We report absolute nucleotide densities rather than nucleotide counts to better compare with *E. coli* chromosome models since the number of nucleotides in *C. crescentus* and *E. coli* chromosomes is different and the distribution of nucleotides along the axis could be different. Our models show slightly less DNA density for *C. crescentus* compared to *E. coli*. This is consistent with *C. crescentus* having a smaller genome by ∼600 kb compared to *E. coli* while the overall cell dimensions are very similar. However, the overall distribution patterns of the DNA density along the short axis from both our and *E. coli* models were found to be very similar.

Figure 5D shows the bending angle distribution observed for our models. The peak is between 150 and 160 degrees, which is comparable to the experimental bending angle of 162.9° with a standard deviation of 8.46° observed for the overall *C. crescentus* shape (56). Our models show a slightly wider distribution with smaller bending angles. However, we note that the average experimental bending angle we use for comparison is for swarmer cells (56); therefore the difference could again be because of the cells from different cell stages used in the Hi-C experiment.

Since the bacterial nucleoids occupy a large portion of the cell, extensive interactions with the intracellular environment are unavoidable. Proteins as large as RNAP and ribosome subunits have been found to penetrate into nucleoids to initiate co-transcriptional translation (53,59,61). On the other hand, complete ribosomes and polysomes are strongly segregated from the nucleoid (62,63). Presumably, ribosomes are excluded from the nucleoid based on their size. We analyzed our models to determine whether they are compatible with the experiments, i.e. whether the structures are porous enough for RNAP and ribosome subunits to enter while excluding assembled ribosomes. Figure 6A shows the accessible cavity volume as a function of macromolecular radius. Essentially, molecules with radii up to 10 nm are able to penetrate the chromosome structures at significant fractions, whereas molecules larger than 15 nm are almost fully excluded. Table 1 compares accessible cavity sizes for a typical size protein, RNAP, the 30S and 50S ribosomal subunits, and entire ribosomes based on their hydrodynamic radii (64). Despite its size, RNAP can access half of the volume available to small proteins and the ribosomal subunits can still access about a quarter of the volume suggesting that both RNAP and ribosomal subunits are able to penetrate and interact extensively with the DNA. However, assembled ribosomes can only access 10% of the space accessible to small proteins consistent with the experiments that find ribosomes to be largely excluded from the nucleoid (62,63). If the scaled models that better fit the FISH data are used, the overall size of the cavities is reduced but the general conclusions remain the same (Table 1, Supplementary Figure S4). Figure 6B further contrasts the much reduced volume accessible to ribosomes in our models compared to the volume that can be occupied by RNAP. Although the results reported here depend on the parameters used in the analysis, they suggest that larger molecules have more difficulties accessing the nucleoid interior and seem to agree with the finding that the porous nucleoid structure would allow mobile RNAP to diffuse relatively easily within the nucleoid in contrast to ribosomes (53). Our results can also be compared with the void distribution in the recent models for the *E. coli* chromosome (51). The *E. coli* models appear to be less compact since voids with radii up to 40 nm were identified in that study and entire ribosomes with radii of about 12 nm appear to fit comfortably within most of the structure according to the distribution of void sizes (51) despite a higher DNA density in *E. coli* compared to our *C. crescentus* models (Figure 5B). However, a direct comparison between our analysis and the work by Hacker *et al.* is complicated by likely differences in how exactly the voids are calculated as well as slightly different DNA topological parameters such as higher superhelical densities in the *E. coli* structures.

Structural variability in the ensemble

All of the nucleoid models generated by our protocol have overall similar shapes. The origin of replication is at one end of the structure. The most prominent feature consists of two arms that are wound around each other in most structures following a sinusoidal pattern with 1.5 – 2 period repeats. The overall shape is similar to lower-resolution

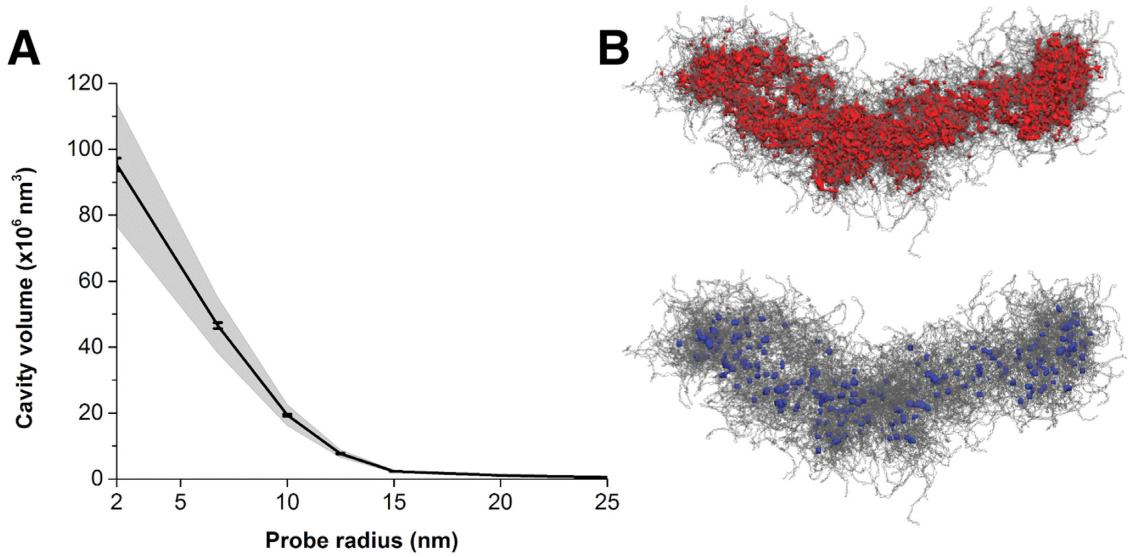

Figure 6. Cavities in the models. (**A**) Distributions of the cavities for different probe radii in the models. The shaded area indicates the standard deviations and the error bars indicate the standard errors based on the ensemble averages. (**B**) Cavities that could be occupied by RNAP (red) and ribosomes (blue).

Table 1. Nucleoid cavity volumes accessible to proteins of different sizes. Ensemble-averaged volumes with standard errors given in the parentheses

		Cavity volume (x10^6 nm^3)	
	Hydrodynamic radius (nm)	Original models	Scaled models
Average size protein	2.0	95.4 (*1.5*)	86.0 (*1.8*)
RNA polymerase	6.9	45.5 (*0.7*)	25.5 (*0.4*)
Ribosomal subunit 30S	8.7	28.6 (*0.4*)	11.1 (*0.2*)
Ribosomal subunit 50S	10.0	19.3 (*0.3*)	5.7 (*0.1*)
Ribosome	11.7	10.6 (*0.1*)	2.1 (*0.04*)

Hydrodynamic radii of the molecules were calculated by HYDRPRO (64) based on PDB structures for RNA polymerase (4KMU), the 30S subunit (5NO3), the 50S subunit (5ADY), and the ribosome (4V4Q).

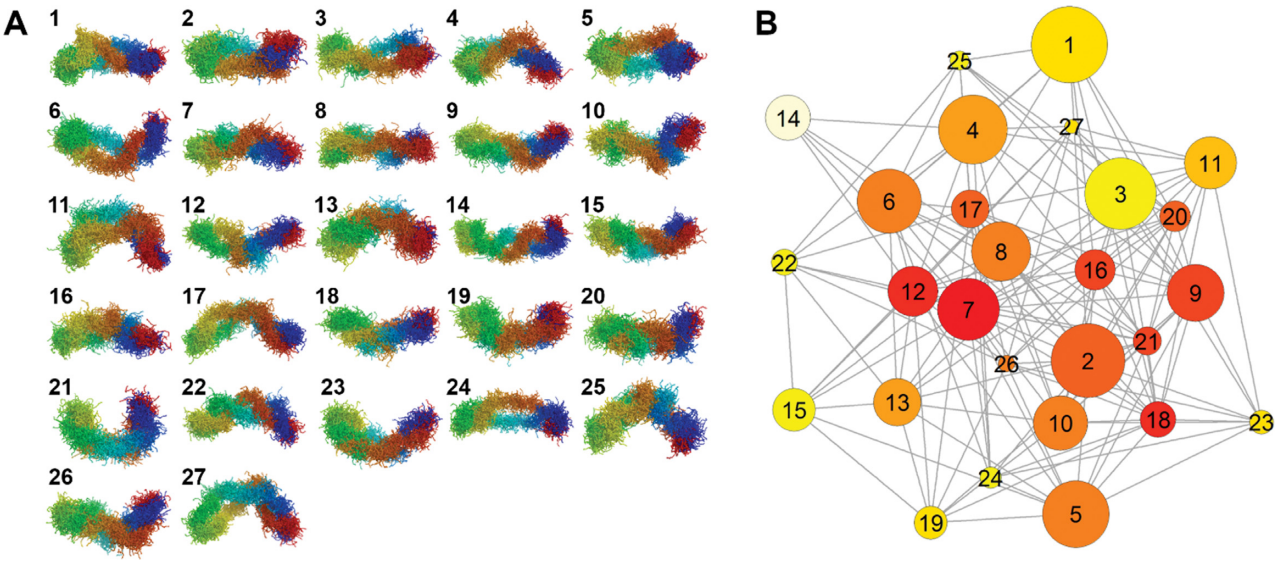

Figure 7. Clustering of our models and possible inter-conversions between clusters. (**A**) Representative structures for ensemble clusters. (**B**) Inter-conversion between clusters based on targeted molecular dynamics. Colors indicate the number of connections to other clusters (from red to white, descending order) and the size of the circles corresponds to the weights of the clusters (0.005–0.113). The indices correspond to the structures shown in (A).

models reported previously by Umbarger *et al.* (21). Many structures exhibit bending that generally follows the curved rod shape of *C. crescentus* cells (56) as already discussed above. Beyond these overall features, there are significant structural variation at the more detailed level. Clustering of the ensemble based on pairwise mutual similarity resulted in 27 different major groups. Representative structures for each cluster are shown in Figure 7A. Population percentages are given in Table 2 and individual contact maps for each cluster are shown in Supplementary Figure S5. The weights of different clusters were optimized in the final step of the ensemble generation to match the resulting contact distributions to experimental Hi-C contact frequencies as described in the Supplementary Data. Some clusters with lower population percentages ended up with higher weights. This could be because of biases and/or insufficient sampling in our initial modeling protocol. However, the initial sampling was meant to focuse just on generating individual structures that are compatible with the experimental data while the reweighting step emphasizes agreement of the average and structural variation of the entire ensemble with the experimental data. Therefore, we would not expect necessarily that the populations after the initial sampling match the weights after the ensemble-based reweighting.

Table 2 summarizes the clusters along with their optimized weights and selected structural properties based on cluster averages. The different clusters are primarily distinguished by different degrees of twisting of the arms and bending. Bending angles for individual clusters range from 116.8° (for the most bent cluster 27) to 163.2° (for the least bent cluster 3). The most strongly bent structures tend to have the lowest weights and as they do not seem to fit well into reported bending angles of $162.9 \pm 8.5°$ for the *C. crescentus* cell shape (56), they are either rare outliers or modeling artifacts. The bending directions with respect to the two arms also vary in different clusters. 12 clusters show one direction, while others are bent in the opposite direction. Clusters with negative bending direction angles tend to bend towards the nucleoid arm with higher index *loci*. In contrast, the positive bending direction results in clusters with bending towards the low-index *loci* arm. In terms of the twisting patterns of the arms, all clusters exhibit partial twisting of arms around each other as observed previously (21).

We further quantified the degree of twisting by calculating the average twisting of one arm around the other and the number of crossings of the arms (Table 2) to understand differences between clusters. Positive twist values correspond to right-handed twisting and negative values indicate left-handed twisting. Both directions are present in the models. Although the overall twisting angles sum up to values near 0°, the twist angles fluctuate between –20° and 20° along the nucleoid axis with different patterns in different clusters (Supplementary Figure S6). Right-handed twisting is more dominant at the poles for all clusters, while there is more variation in the twisting pattern near the centers of the nucleoid. The number of arm-crossings also differs among the clusters. Some clusters (1 and 11) show, on average, less arm-crossing values indicating that the arms prefer to remain on different sides of the nucleoid without crossing in these structures. On the other hand, clusters 9 and 15 have larger twisting and crossing values. Therefore, these clusters tend to have their arms more extensively intertwined. Although the clusters differ in terms of bending and arm-twisting patterns, they all have similar medial axis lengths (see Supplementary Data for the medial axis definition) and widths, consistent with the reported cell size dimensions of *C. crescentus* (56) as discussed in the previous section (Table 2).

As described in the Supplementary Data, we generated models with different branch segment lengths and number of microdomains. Interestingly, each cluster contains a mixture of different internal topologies in terms of the branch length and the number of microdomains (Supplementary Table S1). This suggests that the overall structure in our models is not sensitive to the detailed topology of the DNA. An exponential distribution of microdomain sizes in *E. coli* chromosome was reported by Postow *et al*. (65). A comparison of that data with the microdomain size distributions in our models is shown in Supplementary Figure S7. Generally, the distributions are similar and the exponential distribution is largely reproduced in our models. However, our models have somewhat smaller microdomains and lack the very large microdomain sizes (50–60 kb) seen in *E. coli.* Additionally, the total lengths of the branch segments were also reported for plasmids at 3.5 and 7 kb sizes by Boles *et al.* (66). The total length of the branch segments in the microdomains with similar sizes from our models are in good agreement with the reported branch lengths except for the smallest number of microdomains (Supplementary Figure S7).

An ensemble of nucleoid models with different structures as reported here would be expected based on cell-to-cell variations. However, nucleoid structures are known to be highly dynamic. For example, studies on *E. coli* nucleoid dynamics in living cells have shown the possibility of global and local nucleoid dynamics during a cell-cycle (4,67). Therefore, it is an interesting question to what extent the different conformations for the *C. crescentus* nucleoid are interconvertible without encountering significant topological barriers that could not be overcome without unwinding and expanding the nucleoid structures. We carried out targeted MD simulations between all pairs of representative cluster structures to test which interconversions are likely feasible based on simple energetic criteria (see Supplementary Data). We find that many structures appear to be in fact interconvertible, at least at the overall topological level (see Figure 7B). There are enough pair-wise connections for all of the structures to be connected either directly or indirectly via intermediates. Some structures stand out as central hub structures with a high number of connections. It appears, however, that the number of connections is not strongly correlated (Pearson's correlation coefficient: 0.056) with the cluster weights which may suggest that many of the hub structures are less stable intermediates. As an example, the simulated interconversion of cluster 1 to cluster 5 via cluster 18 is shown in Supplementary Video 2. While the simulations do not provide meaningful insight into relative energetics or kinetics due to the biased and coarse-grained nature of the targeted MD simulations, the general conclusion is that in addition to cell-to-cell variability, the nucleoid structures could be dynamically sampling a wide variety of

Table 2. Features of major clusters of nucleoid structures

Cluster	Population %	Weight	Medial axis length (nm)	Width (nm)	Bending angle (degrees)	Bending direction (degrees)	Arm-twisting (degrees)	Arm-crossings
1	2.10	0.113	2684.7 (*52.1*)	762.7 (*12.6*)	151.5 (*2.1*)	−12.9 (*9.1*)	1.27 (*0.61*)	0.07 (*0.23*)
2	2.48	0.103	2549.9 (*33.3*)	753.8 (*9.6*)	148.3 (*2.3*)	6.9 (*17.4*)	−0.26 (*0.53*)	−0.49 (*0.25*)
3	2.00	0.095	2666.8 (*30.1*)	723.8 (*10.2*)	163.2 (*1.6*)	−110.0 (*21.3*)	3.39 (*0.92*)	−0.40 (*0.26*)
4	3.05	0.071	2648.1 (*34.3*)	745.6 (*8.9*)	153.9 (*1.9*)	−42.8 (*10.6*)	−0.92 (*0.48*)	0.26 (*0.17*)
5	1.71	0.068	2441.7 (*33.0*)	738.9 (*10.3*)	159.8 (*2.7*)	−76.1 (*14.5*)	−1.40 (*0.95*)	0.11 (*0.34*)
6	2.38	0.056	2649.9 (*37.5*)	752.8 (*10.1*)	150.8 (*1.8*)	21.5 (*19.4*)	−1.62 (*0.72*)	0.88 (*0.25*)
7	2.29	0.045	2761.9 (*30.8*)	747.5 (*9.5*)	160.0 (*1.6*)	64.8 (*21.5*)	0.49 (*0.91*)	−0.26 (*0.30*)
8	0.57	0.044	2839.1 (*28.0*)	746.7 (*21.7*)	160.5 (*3.3*)	10.8 (*14.3*)	2.55 (*0.58*)	−0.79 (*0.16*)
9	2.29	0.042	2870.2 (*31.2*)	746.7 (*9.1*)	145.5 (*1.6*)	90.9 (*8.7*)	4.38 (*0.78*)	−1.18 (*0.22*)
10	8.86	0.033	2762.6 (*14.1*)	725.8 (*5.2*)	151.6 (*0.9*)	124.2 (*4.1*)	−1.17 (*0.39*)	0.12 (*0.16*)
11	2.67	0.031	2774.0 (*36.1*)	823.6 (*14.0*)	130.8 (*2.1*)	−61.9 (*20.1*)	0.96 (*0.61*)	0.01 (*0.16*)
12	3.52	0.029	2755.8 (*28.4*)	725.4 (*9.4*)	143.0 (*1.5*)	124.8 (*8.5*)	1.52 (*0.57*)	−0.11 (*0.20*)
13	4.10	0.028	2694.0 (*27.0*)	723.7 (*7.9*)	159.2 (*1.6*)	−179.7 (*10.6*)	−0.62 (*0.54*)	0.38 (*0.17*)
14	3.33	0.028	2888.4 (*26.3*)	747.4 (*7.7*)	149.9 (*1.5*)	−65.9 (*11.3*)	−3.17 (*0.63*)	0.87 (*0.19*)
15	4.38	0.027	2791.3 (*24.3*)	721.7 (*6.9*)	144.1 (*1.5*)	−24.1 (*8.2*)	4.05 (*0.57*)	−1.28 (*0.19*)
16	8.57	0.026	2745.4 (*15.2*)	718.4 (*5.8*)	151.9 (*1.0*)	−47.3 (*6.9*)	−0.68 (*0.37*)	−0.16 (*0.14*)
17	1.05	0.025	2964.0 (*51.3*)	760.0 (*13.5*)	147.6 (*3.7*)	−100.3 (*30.9*)	2.92 (*1.07*)	−0.37 (*0.32*)
18	5.43	0.024	2729.7 (*18.3*)	734.0 (*6.5*)	147.5 (*1.4*)	−125.2 (*9.7*)	−1.31 (*0.38*)	0.64 (*0.17*)
19	4.19	0.023	2707.7 (*21.7*)	745.0 (*7.2*)	142.5 (*1.4*)	14.5 (*7.8*)	0.24 (*0.51*)	−0.53 (*0.17*)
20	2.00	0.019	2682.5 (*44.3*)	733.3 (*11.2*)	149.6 (*2.5*)	53.1 (*17.0*)	−3.29 (*0.52*)	1.00 (*0.20*)
21	4.48	0.018	2698.8 (*33.4*)	782.1 (*8.5*)	135.9 (*1.2*)	−97.8 (*13.1*)	0.56 (*0.56*)	0.14 (*0.17*)
22	8.00	0.018	2815.1 (*17.5*)	754.0 (*5.8*)	135.2 (*1.1*)	−108.1 (*9.2*)	3.10 (*0.46*)	−0.56 (*0.12*)
23	4.57	0.011	2795.9 (*23.6*)	774.2 (*7.1*)	133.8 (*1.3*)	−9.9 (*6.6*)	−3.39 (*0.50*)	0.48 (*0.15*)
24	2.57	0.007	2888.3 (*31.9*)	744.4 (*7.6*)	143.8 (*1.7*)	−133.5 (*16.6*)	−2.61 (*0.58*)	0.77 (*0.18*)
25	2.67	0.007	2749.5 (*29.5*)	731.4 (*9.8*)	145.7 (*1.5*)	35.7 (*5.9*)	−0.70 (*0.68*)	0.26 (*0.23*)
26	3.71	0.006	2763.0 (*24.9*)	787.2 (*8.2*)	132.5 (*1.8*)	15.2 (*7.8*)	−0.50 (*0.54*)	−0.36 (*0.15*)
27	4.76	0.005	2851.8 (*24.5*)	734.0 (*6.4*)	116.8 (*2.5*)	2.9 (*9.4*)	0.38 (*0.58*)	−0.22 (*0.17*)

Structural characteristics are averaged over all members in each cluster. The population percentages refer to the models in the unweighted ensemble. Standard errors are given in parentheses.

structures within a single cell. A better understanding of nucleoid dynamics is clearly an area that would benefit from further studies, both experimentally and via simulations.

Genome-structure mappings

Our models of the *C. crescentus* nucleoid are of sufficiently high-resolution to directly map the genetic sequence onto the 3D structure. This allows an investigation into possible correlations between the DNA structure and the genome sequence that it represents. In order to facilitate the analysis of projections onto variable 3D structures, we will primarily discuss projections of sequence features onto the long axis of the nucleoid models. All of the results were further averaged over the entire ensemble of models taking into account the weights of the clusters that different models belong to.

We began by analyzing the spatial locations of basic sequence features. The distribution of AT- and GC-rich sections is shown in Figure 8 in comparison with the positions of all base pairs in the model. The distributions of AT-rich sites are significantly different from the distribution of all base pairs with enrichment in the central region towards the origin of replication (*Cori*) and depletion opposite *Cori*. In contrast, no preference was found for GC-rich sites. Most NAPs are known to bind to AT-rich sites (24), therefore an enrichment of AT-rich sites from the center to *Cori* may imply enhanced binding of NAPs in that region. Integration host factor (IHF), another NAP, is also reported to assist in maintaining a compact genome by introducing U-turns to DNA (24). We find a statistically significant preference for binding near the center of the nucleoid (Figure 8B). Many of our models are bent and such bending is primarily facilitated by kinking in the central region. Therefore, it could be that IHF is involved in stabilizing a bent nucleoid structure to better fit into the curved bacterial envelope of *C. crescentus*. If this hypothesis is correct, we would expect straightening of the DNA when the IHF gene is disabled.

Promoter sites show only a slight preference for the middle part of the nucleoid along the longest axis (Figure 8C). Recently, the distribution of RNAP in *E. coli* nucleoid showed that RNAPs that are specifically bound to DNA tend to stay closer to the edge of the nucleoid (53). The positions of the bound RNAPs on the nucleoid were later used in the *E. coli* nucleoid modeling and the generated 3D models reproduced the experimental distribution (51). Here, we also analyzed the radial distributions of the promoters along the medial axis of the models (see Supplementary Data) in order to compare with the bound RNAP distribution found for *E. coli* (53) (Figure 8D). We note that in our modeling protocol we did not impose the RNAP distribution as a constraint. The promoter distribution is quite similar and a slight depression in the bound-RNAP distribution in the center relative to the distribution of nucleotides (see Figure 5A) was also observed for the promoter distribution of our models. However, a similar distribution was found for the radial distribution of all beads. This suggests that there is not a clear special pre-arrangement of promoters that would favor positions on the 'outside' of the chromosome structure.

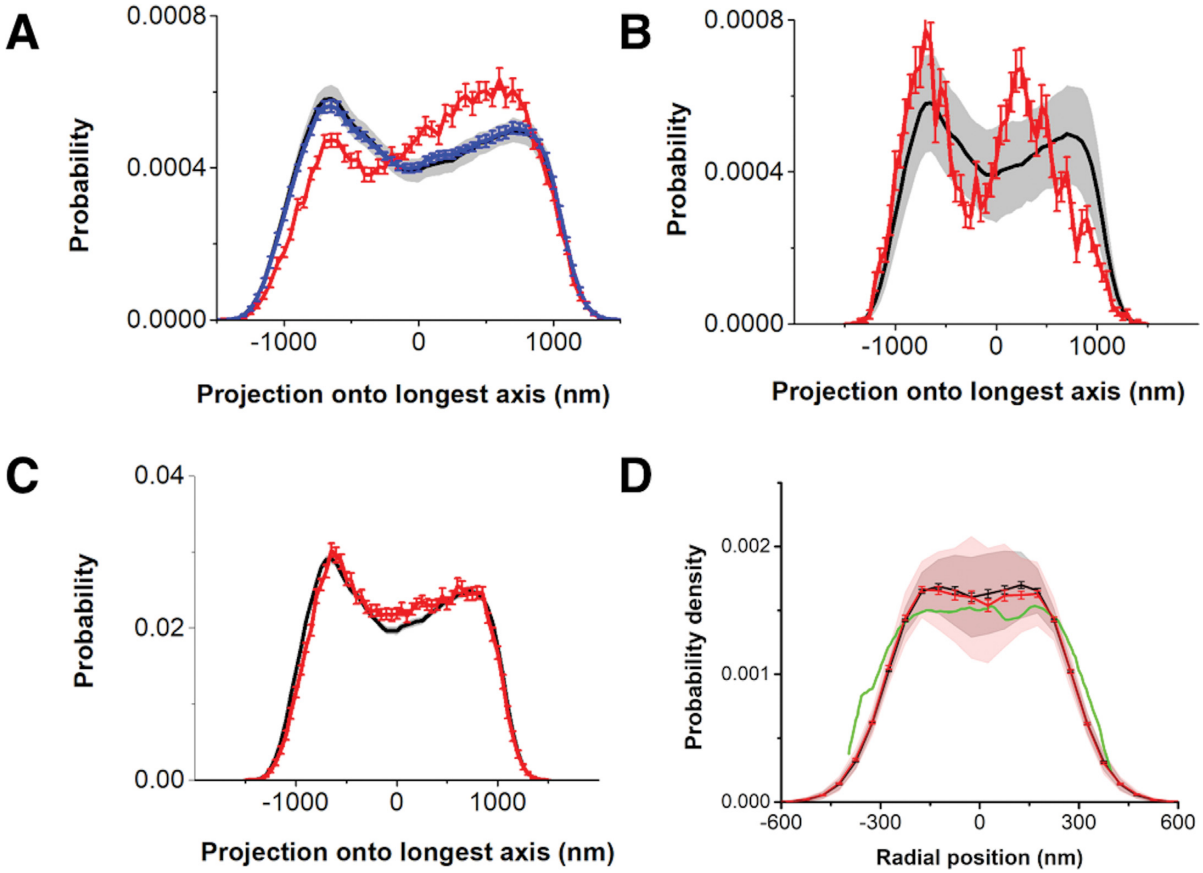

Figure 8. Projections of genomic sequence features onto the 3D nucleoid structures: (**A**) Sites with least 70% in AT-sites (red) or GC-sites (blue) within a 20-bp window compared with the positions of all base pairs (black), (**B**) IHF-binding sites (red), (**C**) Promoter sites (red). (**D**) Radial positions of the promoter sites (red) and all sites (black) in the models from the nucleoid center, and the distribution of bound-RNAP in *E. coli* nucleoid short axis reproduced by digitizing the Figure 4A-ii in the article from Stracy *et al.* (53) (green). In order to compare the *E. coli* data with our projections, the relative x-axis values given in the experimental results were multiplied by half the *C. crescentus* width (400 nm), and the curve was normalized again to keep the area under the curve 1. Gray shaded areas in A, B, and C indicate standard deviations obtained from distributions for the same number of *loci* that were randomly selected 200 times. The shaded area in D indicates the standard deviations and all the error bars indicate the standard errors based on the ensemble averages.

A study by Fang *et al.* analyzed groups of genes in *C. crescentus* that are co-expressed (68). 76 different gene modules were clustered according to their expression profiles (68). In order to test whether co-expression correlates with spatial co-localization, we mapped the beginning of the operon for the corresponding genes in each module onto our models and calculated spatial proximity from the average pair-wise distances of all genes in a given module. We note that we only included the genes with unique operons for each module in the analysis. The results were then compared with pair-wise distances from a distribution of the beginning of randomly selected operons on the same structure. The same analysis was also performed considering genomic distances rather than spatial distances. The resulting z-scores for the modules for which genomic separations of operons are not different than the genomic separations of randomly selected ones ($-1<$z-score<1) are shown in Figure 9A. Out of total 43 modules, z-scores of three modules are skewed towards negative values indicating that a subset ($\sim$7%) of the modules of co-expressed genes is also co-localized. Additionally, in order to avoid any linear sequence effects, we also analyzed only the gene-pairs that are at least 500 kb apart in genomic sequence and we found that the percentage of the modules that have negative z-scores is increased to $\sim$20%. However, we also observed that a similar fraction of co-expressed gene modules resulted in positive z-scores. Overall, the results show that although the genes that are co-expressed seem to have some non-random distributions, a strong tendency for co-expressed gene pairs to be spatially closer was not found. z-scores of the models are listed in Supplementary Table S2. While co-expression is expected to involve the regulatory elements at the beginning of operons, we also examined whether the observed co-localization remains valid if we analyze the positions of the end of each gene in a given module. Figure 9B shows that there is no significant difference to the analysis of the beginning of the operon for each gene.

To further examine a possible relation between gene co-localization and protein product co-localization, we compared with another experimental study on *C. crescentus* where the localization of $\sim$300 proteins was analyzed (69). We mapped the corresponding genes according to their protein product localization onto our models. We found weak evidence that genes whose products are near the poles may be located closer to *Cori*, on one end of the structure (Figure 10), but we could not correlate proteins localized in

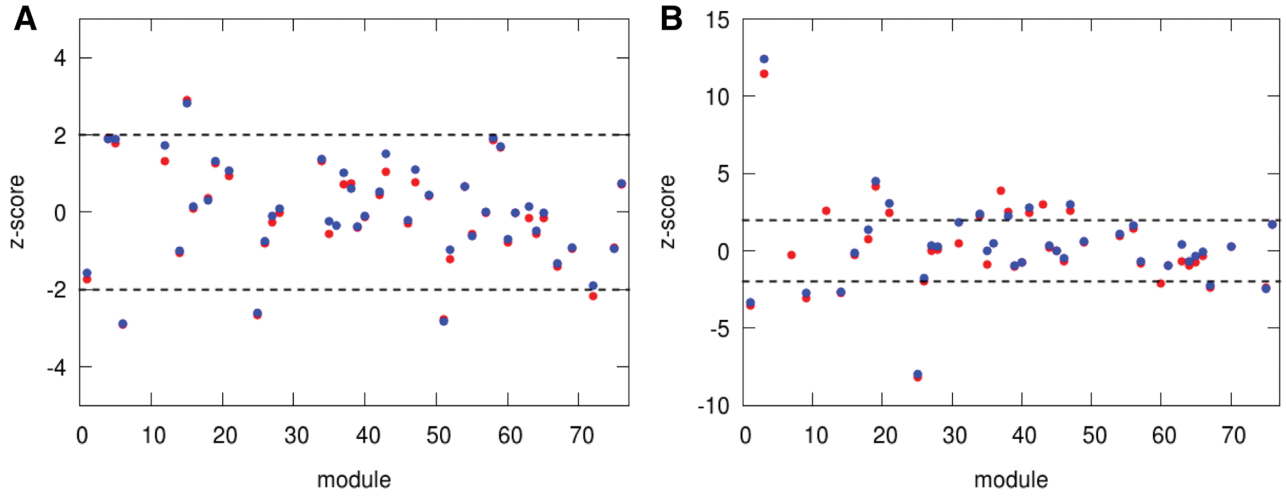

Figure 9. Correlation between co-localized and co-expressed genes: (**A**) z-scores for each module of intra-gene distances for co-expressed genes versus random genes for the end of genes (red) or the beginning of the corresponding operons (blue). (**B**) Same plot as in (A) but only for gene or operon pairs that are separated at least by 500 kb in the genomic sequence.

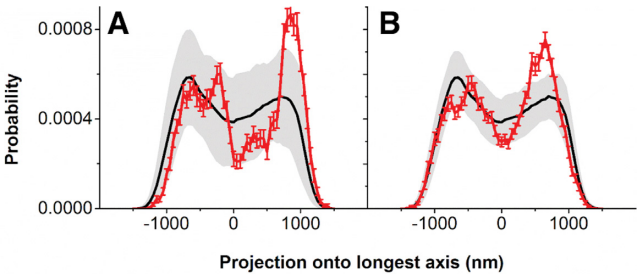

Figure 10. Correlation between gene co-localization and protein product co-localization. Distributions of the genes of the proteins which are experimentally found to be at central (**A**) and polar (**B**) locations (69). Central genes include 'Midband' and 'Central Focus' genes whereas polar genes correspond to 'Polar' and 'Bipolar' genes in the experimental analysis (69). Gray areas indicate the standard deviations as calculated in Figure 8. The error bars indicate the standard errors based on the ensemble averages.

the center of the cell with an enhanced localization of their genes in the center of the nucleoid (Figure 10). These results are again consistent with the findings from *E. coli* chromosome models where no localization was found for the genes for which protein products are co-localized (51). While this does not contradict the idea that proteins are initially synthesized close to where the gene is located, it suggests that the memory of the synthesis site is largely lost due to protein diffusion and/or other cellular transport processes in the experimental study.

Finally, we analyzed the spatial organization of the genes as a function of their cellular functions to test the hypothesis that the localization of certain genes on the nucleoid may be correlated to where they are needed within the cell. Groups of functionally related genes that were distributed significantly different from a random distribution are shown in Figure 11. Genes for regulatory functions, and proteins with unknown functions ('hypothetical proteins') were found to localize in the middle of the nucleoid. In contrast, translation, metabolism and replication genes appear to have a tendency to localize near the poles of the nucleoid while transcription genes tend to be clustered around the *Cori* region. Genes involved in cell division are both near the center and near *Cori*. One can potentially rationalize why cell division genes are near the center since Fts genes form rings attached to the membrane to constrict the cell during cell division. On the other hand, parA/parB accumulate near the poles before cell division in *C. crescentus* swarmer cells (70–73) that may be facilitated by a location of the corresponding genes near *Cori*. The preferential localization of metabolic genes opposite *Cori* could contribute to enhanced metabolic efficiency by bringing proteins involved in metabolic cascades closer together. Based on our data, we generated a diagram in Figure 11B to summarize relative preferences for genes with different functions along the nucleoid. We assume that the resulting protein products would also be enhanced or suppressed accordingly given that transcription and translation is local to a given gene location and that diffusion in crowded environments is relatively slow (74). In general, our findings based solely on the distribution of genes on the nucleoid structure are consistent with findings for *E. coli* where a special organization of genes with similar biological pathways was found (31) and also with the study by Junier *et al.* that found a correlation between co-localization of genes with similar transcriptional regulations and the 3D structure of the chromosome (32). However, it is clear that further experimental validation is necessary to fully understand a possible correlation between gene location on the nucleoid and its function. Experimentally, this could be accomplished for example by analyzing phenotypes for bacteria with shuffled gene distributions.

In order to ensure that the localizations of genomic elements observed in our models are really an effect of the 3D organization of the chromosome rather than the relative locations of genomic features in linear sequence, we also checked their linear sequence localization. In general, it is difficult to distinguish specific patterns in the distributions of the gene *loci* along the linear sequences. (Supplementary Figures S8–S10)

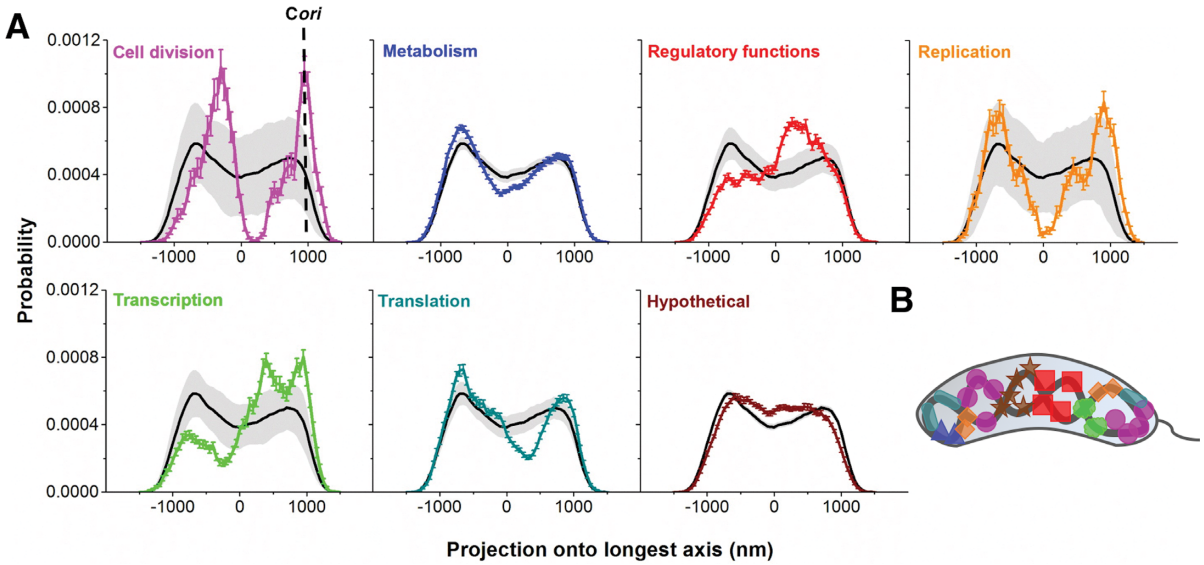

Figure 11. Projections of functionally related genes onto the 3D nucleoid structures. (**A**) Localization of functionally related genes compared to the distribution of all the genes (black) with the standard deviations in gray as calculated in Figure 8. The error bars indicate the standard errors based on the ensemble averages. (**B**) Schematic representation for the proposed functional localization of genes in *C. crescentus*. Colors are as in (A).

We repeated the mapping analysis for the central hub clusters in Figure 7B (clusters 2, 7, 8, 9, 12, 16, 17, 18 and 21) as well as all of the individual clusters to see if there is any difference in the localization of the genomic elements by nucleoid structure and/or connectivity. However, we did not find significantly different results from the results for the overall ensemble (Supplementary Figures S11–S13). This suggests that despite differences in the nucleoid structure, the overall genome localization is largely preserved.

DISCUSSION

In this study, we developed an advanced multi-scale modeling method that is based on data extracted from 3C-based experiments for generating an ensemble of 3D structures of a bacterial chromosome at high-resolution. The key idea was to use information about the plectonemic and supercoiled organization of bacterial chromosomes in combination with the specific experimental restraints that allowed us to generate higher-resolution models compared to models that only use the experimental data. This strategy is similar to the refinement of proteins via NMR where intermediate resolution restraints from the experiment are combined with the knowledge that all proteins are polypeptide chains with certain limited topologies to generate atomistic models. We applied this approach to the *C. crescentus* chromosome where extensive data sets from Hi-C experiments are available (17). Previously, low-resolution models at 13-kb and 434-bp were constructed for *C. crescentus* chromosome using the 5C (21) and Hi-C (17) experimental data, respectively. The 434-bp models generated by Le *et al.* (17) have been the starting point for understanding the organization of genomic DNA based on the Hi-C based experimental data. Here, we further extend this work by developing models at higher resolution that includes DNA supercoiling. We are primarily discussing models with 15-bp resolution, slightly more than one helical turn, but because the DNA structure is fairly stiff on such short length scales, we were also able to generate base-pair resolution models with a reasonable degree of accuracy to the extent that the 15-bp resolution model is accurate.

Another important aspect of our modeling involves the application of experimental restraints by considering the population-based nature of 3C-based experiments. In the first step of modeling, we used soft distance restraints to bring a random polymer model to an average possible structure from where it can deviate to other alternative structures. After the initial generation of plectonemic models and the reconstruction of 15-bp CG models, all the simulations were performed without applying any restraints which allowed models to have fluctuations in their local topologies. Further reweighting of the generated models with significant structural variability enabled obtaining an overall ensemble of chromosomal structures that possibly reflects the possible cell-to-cell structural variability observed in 3C-based experiments. Our modeling protocol resulted in 27 different model clusters with different weights. We found that the average distance map of the reweighted ensemble is in very good agreement with the distance map derived from experiment, on the other hand, individual clusters do not show that much consistency indicating that we were able to generate an ensemble that contains structural variability but also is compatible to the experimental distances in average. Although the details of our modeling protocol differ from previous work, the emphasis on an ensemble approach is similar to previous studies of the *C. crescentus* chromosome (17,21,51).

Investigation of the physical properties of the chromosome models showed that our models are compact enough to fit within the cell envelope of *C. crescentus.* At the same time, the structures are porous enough for proteins as large as RNAP and ribosome subunits to penetrate and diffuse through the nucleoid as previously reported (53,59,61). In

addition, we explored if the models are compatible with the experimental data reported for *C. crescentus* that was not used while building the models (55). The distances calculated from the models showed agreement with the experimental distances for the given DNA segment pairs but with some deviations for distances larger than ∼1000 nm. We argued that this deviation might result from the fact that the Hi-C experiment were performed on cells at different cell phases (17) whereas FISH data was collected only from swarmer cells that were stopped from cell growth (6,55). However, if we scale our models and reanalyze their compatibility to the FISH distances showed that the scaled models still agree with the experimental data and preserve their global topologies. This suggests that there may be significant room for the chromosome to expand and shrink as the cell size varies during the cell cycle.

Our 3D models are similar to previous shape outlines for the *C. crescentus* genome derived based on the 5C data (21), but we also see significant differences in topology. We found that there is a different degree of arm-twisting and bending in different clusters. Some of the structures exhibit extreme bending that is not fully compatible with previously reported data for *C. crescentus* (56). These models may be artifacts of our modeling procedure, however as the modeling suggests relatively low weights, these structures could also be rare outliers that are difficult to detect experimentally if they occurred.

We found significant structural variability in our ensemble. This is expected and interpreted generally as cell-to-cell variation of chromosome structure. However, our analysis suggests that different structures may interconvert to a significant degree within the same cell. In fact, it appears that all of the clusters we observe are inter-convertible to each other either directly or through intermediate structures. Such remodeling of chromosome structures would require cooperation and activity by NAPs which was not considered here. It is clear that there is much room for further studies and the nucleoid structures presented here are an ideal starting point to further examine nucleoid dynamics via simulations.

The high-resolution models enabled mapping the genomic sequence onto the 3D structure. This allowed us to analyze the distribution of generic sequence elements as well as specific genes. We found non-random distributions of AT-rich sequences and other NAP-binding motifs that hint at the role of NAPs in stabilizing the nucleoid structure. We also observed an apparent spatial organization of genes involved in certain functions and we hope that these results will further motivate new research. Experimental tests for a proposed non-random distribution of genes could involve gene or operon randomizations with the hypothesis that randomly distributed genes could affect metabolic efficiency and/or regulatory processes.

High-resolution nucleoid models were also recently generated by Hacker *et al.* for the *E. coli* chromosome (51). They generated chromosome models at one-nucleotide resolution for the first time. Here we apply similar approach to generate *C. crescentus* chromosome initially at 15-bp resolution but then extended the resolution to the one-base pair level. One difference between the modeling protocols by Hacker *et al.* and us is that our models rely primarily on Hi-C based contacts. In the work by Hacker *et al.*, the modeling was guided by experimental 2D radial distributions of RNAP to generate models that reproduce these experimental distributions (53). This approach is expected to give nucleoid models that are physically realistic, but since RNAP radial distributions are projected onto 2D, the specific positions of the genomic *loci* in 3D space might not be as well defined as there were no constraints guiding those positions, or their relative distances, in the modeling protocol. On the other hand, the *E. coli* chromosome models include plectoneme-free (relaxed DNA) and plectoneme-abundant (supercoiled DNA) regions based on the ChIP-chip data on RNAP for *E. coli* (52). In our *C. crescentus* chromosome models, we did not specifically identify supercoiled-free regions. However, we assume that some parts may be unwound for active transcription and there is sufficient space for supercoiled regions to unwind. Because we only modeled the segments lying between plectoneme rosettes that form the central ring of the model as non-supercoiled double-stranded DNA strands, it is possible that our model is topologically less accurate. However, despite the significant differences in the modeling protocols, it is reassuring that the overall structural and mapping analyses showed consistent conclusions in both *C. crescentus* and *E. coli* chromosomes.

Our methodology is easily applicable to other bacteria as additional 3C-based data set are becoming available and expand the ability to study the direct connection between the 3D distribution of genes and their function that should be part of a complete analysis of genotype-phenotype relationships.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Aaron Dinner for sharing experimentally observed bending angles for the *C. crescentus* cell shape and Dr Aleksei Aksimentiev for sharing potential mean force data for DNA–DNA interactions at different ion concentrations. Authors would also like to thank Dr Liang Fang for initial contributions in developing the modeling protocol. Computer resources were used at XSEDE facilities [TG-MCB090003] and at Michigan State University's High-Performance Computing Center.

FUNDING

National Institutes of Health [GM092949]; National Science Foundation [MCB 1330560]. Funding for open access charge: [MCB 1330560].
Conflict of interest statement. None declared.

REFERENCES

1. Niki,H., Yamaichi,Y. and Hiraga,S. (2000) Dynamic organization of chromosomal DNA in Escherichia coli. *Genes Dev.*, **14**, 212–223.
2. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.

3. 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.
4. 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.
5. Nielsen,H.J., Ottesen,J.R., Youngren,B., Austin,S.J. and Hansen,F.G. (2006) The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. *Mol. Microbiol.*, **62**, 331–338.
6. Viollier,P.H., Thanbichler,M., McGrath,P.T., West,L., Meewan,M., McAdams,H.H. and Shapiro,L. (2004) Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 9257–9262.
7. Wang,X., Montero Llopis,P. and Rudner,D.Z. (2014) Bacillus subtilis chromosome organization oscillates between two distinct patterns. *Proc. Natl Acad. Sci. U.S.A.*, **111**, 12877–12882.
8. 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.
9. Youngren,B., Nielsen,H.J., Jun,S. and Austin,S. (2014) The multifork Escherichia coli chromosome is a self-duplicating and self-segregating thermodynamic ring polymer. *Genes Dev.*, **28**, 71–84.
10. Weng,X. and Xiao,J. (2014) Spatial organization of transcription in bacterial cells. *Trends Genet.*, **30**, 287–297.
11. Lau,I.F., Filipe,S.R., Soballe,B., Okstad,O.A., Barre,F.X. and Sherratt,D.J. (2003) Spatial and temporal organization of replicating Escherichia coli chromosomes. *Mol. Microbiol.*, **49**, 731–743.
12. Fekete,R.A. and Chattoraj,D.K. (2005) A cis-acting sequence involved in chromosome segregation in Escherichia coli. *Mol. Microbiol.*, **55**, 175–183.
13. Robinett,C.C., Straight,A., Li,G., Willhelm,C., Sudlow,G., Murray,A. and Belmont,A.S. (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.*, **135**, 1685–1700.
14. Teleman,A.A., Graumann,P.L., Lin,D.C.H., Grossman,A.D. and Losick,R. (1998) Chromosome arrangement within a bacterium. *Curr. Biol.*, **8**, 1102–1109.
15. Webb,C.D., Teleman,A., Gordon,S., Straight,A., Belmont,A., Lin,D.C.H., Grossman,A.D., Wright,A. and Losick,R. (1997) Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of B-subtilis. *Cell*, **88**, 667–674.
16. 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.
17. 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.
18. Marbouty,M., Cournac,A., Flot,J.F., Marie-Nelly,H., Mozziconacci,J. and Koszul,R. (2014) Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. *eLife*, **3**, e03318.
19. Marbouty,M., Le Gall,A., Cattoni,D.I., Cournac,A., Koh,A., Fiche,J.B., Mozziconacci,J., Murray,H., Koszul,R. and Nollmann,M. (2015) Condensin- and replication-mediated bacterial chromosome folding and origin condensation revealed by Hi-C and super-resolution imaging. *Mol. Cell*, **59**, 588–602.
20. Trussart,M., Yus,E., Martinez,S., Bau,D., Tahara,Y.O., Pengo,T., Widjaja,M., Kretschmer,S., Swoger,J., Djordjevic,S. *et al.* (2017) Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat. Commun.*, **8**, 14665.
21. 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.
22. Wang,X.D., Le,T.B.K., Lajoie,B.R., Dekker,J., Laub,M.T. and Rudner,D.Z. (2015) Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis. *Genes Dev.*, **29**, 1661–1675.
23. Browning,D.F., Grainger,D.C. and Busby,S.J. (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.*, **13**, 773–780.
24. Dillon,S.C. and Dorman,C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, **8**, 185–195.
25. 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. U.S.A.*, **107**, 4991–4995.
26. Wang,X. and Rudner,D.Z. (2014) Spatial organization of bacterial chromosomes. *Curr. Opin. Microbiol.*, **22**, 66–72.
27. Badrinarayanan,A., Le,T.B. and Laub,M.T. (2015) Bacterial chromosome organization and segregation. *Annu. Rev. Cell Dev. Biol.*, **31**, 171–199.
28. Dame,R.T., Noom,M.C. and Wuite,G.J. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, **444**, 387–390.
29. Dame,R.T., Wyman,C. and Goosen,N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res.*, **28**, 3504–3510.
30. Higgins,C.F., Dorman,C.J., Stirling,D.A., Waddell,L., Booth,I.R., May,G. and Bremer,E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in S. typhimurium and E. coli. *Cell*, **52**, 569–584.
31. Xie,T., Fu,L.Y., Yang,Q.Y., Xiong,H., Xu,H., Ma,B.G. and Zhang,H.Y. (2015) Spatial features for Escherichia coli genome organization. *BMC Genomics*, **16**, 37.
32. Junier,I., Martin,O. and Kepes,F. (2010) Spatial and topological organization of DNA chains induced by gene co-localization. *PLoS Comp. Biol.*, **6**, e1000678.
33. Dily,F.L., Serra,F. and Marti-Renom,M.A. (2017) 3D modeling of chromatin structure: is there a way to integrate and reconcile single and population experimental data? *Wiley Interdiscp. Rev.: Comput. Mol. Sci.*, **7**, e1308.
34. Bau,D., Sanyal,A., Lajoie,B.R., Capriotti,E., Byron,M., Lawrence,J.B., Dekker,J. and Marti-Renom,M.A. (2011) The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.*, **18**, 107–114.
35. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y.J., Lee,C., Shendure,J., Fields,S., Blau,C.A. and Noble,W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
36. Hu,M., Deng,K., Qin,Z., Dixon,J., Selvaraj,S., Fang,J., Ren,B. and Liu,J.S. (2013) Bayesian inference of spatial organizations of chromosomes. *Plos Comp. Biol.*, **9**, e1002893.
37. Jhunjhunwala,S., van Zelm,M.C., Peak,M.M., Cutchin,S., Riblet,R., van Dongen,J.J., Grosveld,F.G., Knoch,T.A. and Murre,C. (2008) The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*, **133**, 265–279.
38. Lesne,A., Riposo,J., Roger,P., Cournac,A. and Mozziconacci,J. (2014) 3D genome reconstruction from chromosomal contacts. *Nat. Methods*, **11**, 1141–1143.
39. Segal,M.R., Xiong,H., Capurso,D., Vazquez,M. and Arsuaga,J. (2014) Reproducibility of 3D chromatin configuration reconstructions. *Biostatistics*, **15**, 442–456.
40. Varoquaux,N., Ay,F., Noble,W.S. and Vert,J.P. (2014) A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*, **30**, i26–i33.
41. Zhang,Z., Li,G., Toh,K.C. and Sung,W.K. (2013) 3D chromosome modeling with semi-definite programming and Hi-C data. *J. Comput. Biol.*, **20**, 831–846.
42. Dekker,J. (2016) Mapping the 3D genome: Aiming for consilience. *Nat. Rev. Mol. Cell Biol.*, **17**, 741–742.
43. Giorgetti,L. and Heard,E. (2016) Closing the loop: 3C versus DNA FISH. *Genome Biol.*, **17**, 215.
44. Imakaev,M.V., Fudenberg,G. and Mirny,L.A. (2015) Modeling chromosomes: Beyond pretty pictures. *FEBS Lett.*, **589**, 3031–3036.
45. Trussart,M., Serra,F., Bau,D., Junier,I., Serrano,L. and Marti-Renom,M.A. (2015) Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. *Nucleic. Acids. Res.*, **43**, 3465–3477.
46. Williamson,I., Berlivet,S., Eskeland,R., Boyle,S., Illingworth,R.S., Paquette,D., Dostie,J. and Bickmore,W.A. (2014) Spatial genome organization: contrasting views from chromosome conformation

capture and fluorescence in situ hybridization. *Genes Dev.*, **28**, 2778–2791.
47. Di Pierro,M., Zhang,B., Aiden,E.L., Wolynes,P.G. and Onuchic,J.N. (2016) Transferable model for chromosome architecture. *Proc. Natl Acad. Sci. U.S.A.*, **113**, 12168–12173.
48. Zhang,B. and Wolynes,P.G. (2015) Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl Acad. Sci. U.S.A.*, **112**, 6062–6067.
49. Kalhor,R., Tjong,H., Jayathilaka,N., Alber,F. and Chen,L. (2011) Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.*, **30**, 90–98.
50. Tjong,H., Li,W., Kalhor,R., Dai,C., Hao,S., Gong,K., Zhou,Y., Li,H., Zhou,X.J., Le Gros,M.A. *et al.* (2016) Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl Acad. Sci. U.S.A.*, **113**, E1663–E1672.
51. Hacker,W.C., Li,S. and Elcock,A.H. (2017) Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. *Nucleic. Acids. Res.*, **45**, 7541–7554.
52. Grainger,D.C., Hurd,D., Goldberg,M.D. and Busby,S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic. Acids. Res.*, **34**, 4642–4652.
53. Stracy,M., Lesterlin,C., Garza de Leon,F., 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. U.S.A.*, **112**, E4390–E4399.
54. Dorman,C.J. (2013) Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.*, **11**, 349–355.
55. 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. U.S.A.*, **110**, 1674–1679.
56. Wright,C.S., Banerjee,S., Iyer-Biswas,S., Crosson,S., Dinner,A.R. and Scherer,N.F. (2015) Intergenerational continuity of cell shape dynamics in Caulobacter crescentus. *Sci. Rep.-UK*, **5**, 9155.
57. Robinow,C. and Kellenberger,E. (1994) The Bacterial Nucleoid Revisited. *Microbiol. Rev.*, **58**, 211–232.
58. Wang,X., Montero Llopis,P. and Rudner,D.Z. (2013) Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.*, **14**, 191–203.
59. 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.
60. Lee,S.F., Thompson,M.A., Schwartz,M.A., Shapiro,L. and Moerner,W.E. (2011) Super-resolution imaging of the nucleoid-associated protein HU in Caulobacter crescentus. *Biophys. J.*, **100**, L31–L33.
61. Sanamrad,A., Persson,F., Lundius,E.G., Fange,D., Gynna,A.H. and Elf,J. (2014) Single-particle tracking reveals that free ribosomal subunits are not excluded from the Escherichia coli nucleoid. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 11413–11418.
62. Mondal,J., Bratton,B.P., Li,Y., Yethiraj,A. and Weisshaar,J.C. (2011) Entropy-based mechanism of ribosome-nucleoid segregation in E. coli cells. *Biophys. J.*, **100**, 2605–2613.
63. 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.
64. Ortega,A., Amoros,D. and Garcia de la Torre,J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.*, **101**, 892–898.
65. 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.
66. Boles,T.C., White,J.H. and Cozzarelli,N.R. (1990) Structure of plectonemically supercoiled DNA. *J. Mol. Biol.*, **213**, 931–951.
67. Liu,Y., Xie,P., Wang,P., Li,M., Li,H., Li,W. and Dou,S. (2015) A model for chromosome organization during the cell cycle in live E. coli. *Sci. Rep.-UK.*, **5**, 17133.
68. Fang,G., Passalacqua,K.D., Hocking,J., Llopis,P.M., Gerstein,M., Bergman,N.H. and Jacobs-Wagner,C. (2013) Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. *BMC Genomics*, **14**, 450.
69. Werner,J.N., Chen,E.Y., Guberman,J.M., Zippilli,A.R., Irgon,J.J. and Gitai,Z. (2009) Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl Acad. Sci. U.S.A.*, **106**, 7858–7863.
70. Goley,E.D., Iniesta,A.A. and Shapiro,L. (2007) Cell cycle regulation in Caulobacter: location, location, location. *J. Cell Sci.*, **120**, 3501–3507.
71. Quardokus,E.M., Din,N. and Brun,Y.V. (2001) Cell cycle and positional constraints on FtsZ localization and the initiation of cell division in Caulobacter crescentus. *Mol. Microbiol.*, **39**, 949–959.
72. Thanbichler,M. and Shapiro,L. (2006) MipZ, a spatial regulator coordinating chromosome segregation with cell division in Caulobacter. *Cell*, **126**, 147–162.
73. Mohl,D.A. and Gober,J.W. (1997) Cell cycle-dependent polar localization of chromosome partitioning proteins in Caulobacter crescentus. *Cell*, **88**, 675–684.
74. Mika,J.T. and Poolman,B. (2011) Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotech.*, **22**, 117–126.