Computational Elucidation of self-organization of *E. coli* chromosome underlying HI-C data

Abdul Wasim[1], Ankit Gupta[1], and Jagannath Mondal *[1]

[1]Tata Institute of Fundamental Research, Centre for Interdisciplinary Sciences, Hyderabad 500046, India

June 30, 2020

Abstract

The chromosome of archetypal bacteria *E. coli* is riddled with multifaceted complexity and its multiscale organization is slowly getting recognised. The emergence of chromosome conformation capture techniques and super-resolution microscopy are providing newer ways to explore chromosome organisation, chromosome dynamics and its effect on gene expression. Here we combine a beads-on-a-spring polymer model and recently reported high-resolution Hi-C data of *E. coli* chromosome to develop a comprehensive model of *E. coli* chromosome at 5 kbp resolution. The Hi-C data-integrated chromosome model elucidates a self-organised structure of *E. coli* chromosome into multiple macrodomains within ring-like architecture, with oriC loci located at the mid-cell position. The model also predicts that a majority of the genetic loci are linearly organised, except Ter macrodomain. The distance profiles, extracted from the model are in quantitative agreement with data from FISH and DNA-recombination assay experiments. Additionally, a genome-wide fine-grained radius of gyration map captures multiple chromosomal interaction domains (CIDs) and identifies the location of rrn operons. Finally, via a mutation-based analysis, the model quantifies the role of multiple nucleotide-associated proteins (NAPs) like HU, Fis and MatP in controlling the chromosome architecture.

Introduction

The folding of the 4.64 Mbp circular chromosome, with a contour length of 1.6 mm, inside 2-4 μm long spherocylindrical cell[1, 2] of *Escherichia coli* (*E. coli*) ($\sim 1.5\mu m^3$ in volume) is a complex process, mediated by numerous factors and cues. Several decades' investigations on chromosomal DNA of *E. coli* have rendered a picture of a highly condensed form called the nucleoid which is a dynamic macromolecular complex of the genetic material and nucleotide associated proteins (NAP) along with proteins such as RNA Polymerases(RNAPs)[3]. The multitude of experiments on *E. coli* chromosomes have highlighted a ring-like architecture of nucleoid whose organization results from a combination of processes including DNA supercoiling[4], NAP-induced condensation of DNA[5], crowding and non-equilibrium processes like transcription[3]. In rapidly growing *E. coli* cells, nucleoid is severely compacted at the centre of cytoplasm with ribosomes being strongly concentrated at the periphery of the nucleoid[6]. The multi-scale organization underlying the *E. coli* nucleoid is only slowly getting recognised.

*jmondal@tifrh.res.in

A set of classic experiments, including fluoroscence in situ hybridisation (FISH) techniques [7, 8] and site-specific recombination assays[9] had indicated spatial proximity and increased interactions among genetically distant DNA sites, giving birth to the idea of organization into a ring-like chromosomal architecture comprised of four large macrodomains, namely Ori, Ter, Left, Right and two Non-Structured regions- Right (NS-R) and Left (NS-L). These macrodomains are considered to be spatially segregated from each other[7, 8]. Another investigation via fluoroscence labelling of the genetic loci had proposed a model in which *E. coli* nucleoid displays a linear order of the locus distribution along the axial dimension of spherocylinder[10]. As key features, these investigations have alluded to an intricate role that specific and non-specific interactions of NAPs with the DNA play in maintaining the short-range and long-range organization in the bacterial chromosome.

The complexity underlying bacterial chromosome has motivated a series of computer simulation studies and models to describe the architecture of *E. coli* chromosome. Some of them are based on coarse-grained models[6, 11, 12, 13, 14], while others used a more fine-grained approach for bacteria with base pair or approaching base-pair resolutions[15, 16, 17, 18]. While these models significantly contribute to our current understanding on *E. coli* chromosome, most of these models are phenomenological in nature and lack attempts in integrating experimental Hi-C data.

In the first data driven model for *E. coli* chromosome, ChIP-chip data for RNAP was integrated into a fine grained polymer model[19]. On the other hand, the emergence of high resolution chromosome conformation capture data in bacterial cells [20, 21, 22] has helped to detect the interaction frequency between any two genomic loci in the whole genome of an organism and build data-informed models of bacterial chromosomes. Hi-C has the unique ability to determine the interaction frequency map of whole genome of an organism at high resolution[23]. The ensemble nature of the resulting contact frequency matrices automatically incorporate the inherent stochasticity of the chromosome's interactions. Using Hi-C and super-resolution microscopic imaging the 3D chromosome conformation of *Mycoplasma pneumoniae* was determined[22]. Similarly a Hi-C driven model was proposed for investigation of conformational dynamics of *Caulobacter cresentus* chromosome[24]. All of the above data driven approaches show that with emerging experimental techniques, one can fine-tune one's computational model using experimental data. Such theoretical studies can introduce specificity such as macrodomain formation, their locations and relative sizes for a bacterium into its model(s).

In this regard, the recent report of high resolution Hi-C interaction maps of *E. coli* chromosome, has been a key breakthrough[20]. It brings out the salient features of NAP-mediated multiscale organization underlying the *E. coli* chromosome architecture. It also opens up promising opportunities for developing high resolution and quantitative model for the same. In the current work, we present computer simulations of the *E. coli* chromosome by integrating beads-on-a-spring polymer model with recently reported Hi-C interaction matrix of *E. coli* chromosome[20]. As would be detailed in the article, the model, developed at a 5 kbp resolution, unifies multiple existing hypothesis related to *E. coli* chromosome's architecture. It compares a wide array of experimental data such as FISH data[7], recombination assay[9] and ChIP-chip data based simulations[19]. As an important application, by comparing with mutation-based Hi-C data, the model finally quantifies the role of a number of nucleotide associated proteins in maintaining structural organization in the chromosome.

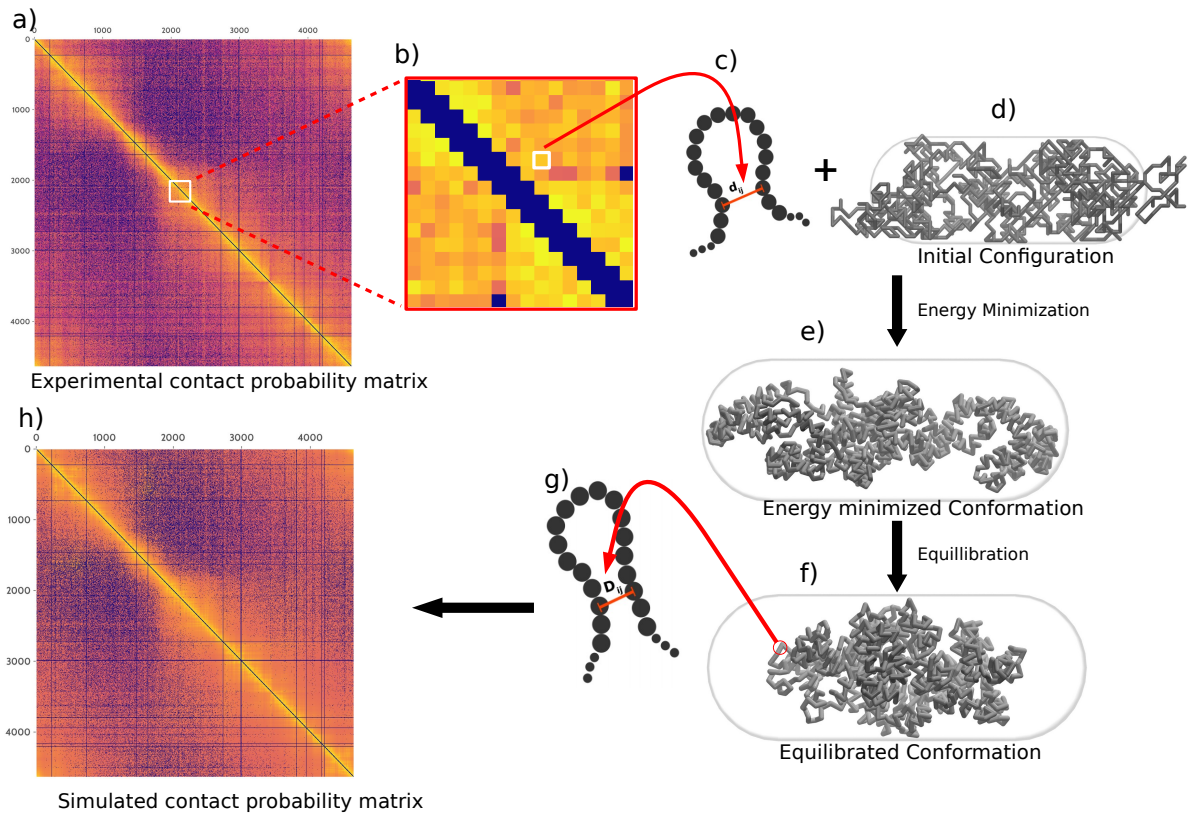

Figure 1: A schematic of the method used to generate an ensemble of final chromosome structures. **a)** Contact probability matrix from Hi-C experiment. **b)** Zoomed section of the matrix. **c)** The experimental contact probabilities are mapped onto bond distances and bond strengths. d_{ij} is the initial bond between particles i and j. **d)** The calculated distances were implemented on a beads-on-a-spring model (having only repulsive Lennard jones potential initially) as harmonic constraints. **(e)** An energy minimized conformation with interactions as described previously. **f)** Chromosome after equilibration of an energy minimized conformation. **g)** D_{ij} is the equilibrium bond distance between i and j. Average bond distances from equilibrium conformations were converted to a contact probability matrix. **h)** Hi-C contact probability matrix from simulation.
The simulated and the experimental matrices are compared and the simulated matrix is used for further analysis.

Materials and Methods

Figure 1 outlines the schematic of the integrative method used in the current work to generate the chromosome structures via combination of bead-in-a-spring model and Hi-C contact matrices. Below we detail different segments of the methods employed in the current work.

Hi-C data processing

We used the data made available by *Lioy et al. (2018)* [20] with the GEO accession number $GSE107301$. SRA (Sequence Read Archive) files were splited into both reads of pair-end sequences using *fastq-dump*. We used hiclib python library (*https://bitbucket.org/mirnylab/hiclib*) provided by Leonid Mirny's group for further processing of the fastqs. Using hiclib, iterative mapping was performed with a minimum sequence length of 20bps and a step length of 5bps. Alignment reads were saved in BAM (Binary Alignment Map) files and processed according to 5kb resolution and filtered using hiclib default fragment level filtering functions, namely: duplicates, large, extreme, dangling ends, etc and they were binned into 928 bins according to 5kbp

resolution. Further bin level filtering was performed to remove low coverage bins, bins with only small area sequenced and diagonal and adjacent to diagonals bins. Then the raw matrix was extracted using h5py python library from HDF5 file generated by hiclib library, we also set the two extreme elements on the off-diagonal of the matrix to zero, since they also are reads from adjacent regions of the chromosome, due to the chromosome being circular. The matrix is then normalized by using sequential component normalization (SCN) [25] in which, first all the column vectors were normalized to one using euclidian norm followed by each row vector and the whole process was iterated until the matrix become symmetric again (3 iterations in our case). The normalized matrix is converted to a contact probability matrix by dividing each row by its maximum value [26] followed by a resymmetrization of the matrix(Figure 1a).

Model details and interaction potentials

We modelled the 4.64×10^6 bp *E. coli* chromosome as a beads-on-spring polymer with each bead representing a 5×10^3 bp (i.e. 5 kbp) nucleotides, which is the resolution of Hi-C interaction maps. The 5 kbp nucleotides are indexed and annotated as per the genetic sequence of wild type *E. coli* MG1655 (GenBank ID: U00096.2) and modelled as non-overlapping van der waals particles, giving rise to a total number of $928 (= 4.64 \times 10^6 / 5 \times 10^3)$ beads. The polymer beads are subjected to a spherocylindrical confinement commensurate with average dimension of *E. coli* at 37°C in LB media (axial length (including end-caps) of 2.482 μm and the diameter of 0.933 μm [1]). As detailed in the SI *Methods*, based on an approximate volume fraction of the chromosome of 0.1[27] (with respect to cell volume) and the spherocylindrical confinement dimensions, the individual bead diameter (σ) was determined to be 0.067 μm. The non-bonded interactions are purely repulsive and are given by the repulsive part of Lenard-Jones potential $V_{nb}(r) = \frac{A}{r^{12}}$ where, A $= 4\epsilon\sigma^{12}$. For the simulations, A=1.0 kJ mol^{-1} σ^{12} has been used.

Adjacent beads of the polymer are connected by strong (300 kJ mol$^{-1}\sigma^2$) harmonic springs with σ as the equilibrium bond length. Hi-C contacts are also modelled as harmonic springs but with distance-dependent force constants and probability-dependent bond lengths(Figure 1b). For Hi-C contact probability matrix P, we define the distance matrix, D as $D_{ij} = \frac{\sigma}{P_{ij}}$, where ij suggests the element in the i-th row and j-th column of the matrices. The restraining potential between a pair of Hi-C contacts at a separation of r_{ij} is given by $V_{HiC}(r_{ij}) = \frac{1}{2}k_{ij}(D_{ij} - r_{ij})^2$. It should be noted that the matrix P is a sparse matrix with large number of the probability matrix elements being close to zero. Therefore a lot of the elements in D would translate to large distance close to ∞. To take this into account, in our simulations, the force constants for the bonds incorporating the Hi-C contacts have been modelled as a gaussian function of the Hi-C distances (as represented in Figure 1b and 1c):

$$k_{ij} = k_0 e^{-\frac{(D_{ij}-\sigma)^2}{w}} \tag{1}$$

Here k_0 is an amplitude term that determines the upper limit to the force constants of the "Hi-C bonds". This function essentially implies weaker values of force constant for larger distances. This function naturally takes $V_{HiC}(D_{ij}) = 0$ for $D_{ij} = \infty$, . In equation 5 in *SI methods*, k_0 and w are parameters that need to be optimised. The metric used to optimize w is a Pearson correlation coefficient between the experimental and the *filtered* simulated contact probability matrices (filtering has been performed as explained in SI *Materials and Methods*).

To implement the effect of spherocylindrical confinement induced by a *E. coli* cell, a restraining potential (Eq. 2) has been used

$$V_{res}(r; R_0) = \frac{1}{2} k_{res} |\vec{r} - \vec{R_0}|^2 \ \mathbf{H}(|\vec{r} - \vec{R_0}|) \tag{2}$$

H is a step function and gets activated only if any chromsome bead attempts to get out of the spherocylindrical confinement. R_0 is the center of the spherocylinder. k_{res} determines extent of elasticity of the cell boundary. For simulations we have used 310 kJ mol$^{-1}\sigma^2$.

Simulation details

All simulations are performed using open source package GROMACS 5.0.7[28]. The source code of the program was modified by us to implement the interaction potential function of the spherocylindrical confinement. All other bonded and non-bonded interaction potentials were introduced by using default GROMACS utilities. At the start of all simulations, a 928 bead polymer is generated bead-wise (using home made python scripts), avoiding any inter-bead overlap. The generated conformations are subsequently energy-minimized by steepest descent algorithm using GROMACS(Figure 1d and 1e). The energy minimized polymer is then subjected to Langevin dynamics in NVT ensemble (Figure 1e and 1f). The temperature of the system was maintained using Langevin thermostat at 303K. The time step for equilibration or production run is 2 $\sqrt{\frac{m\sigma^2}{\epsilon}}$, where m is the mass of 1 bead (see GROMACS manual for units). Each production run has been run for 2×10^6 steps from which only the last 2000 frames have been used for further analysis. For each cell type we have discussed, 200 simulations have been performed with different initial structures to capture the stochastic effect of the ensemble. Using the last 2000 frames of k_{th} simulation, we generate a distance matrix and a probability matrix, whose values have been averaged over the number of frames. We generate the final simulated distance matrix and the probability matrix by averaging over frame averaged matrices (represented by Figure 1f, 1g, and 1h).
We also performed simulations with 12×10^6 steps to check for proper equilibration of the polymer. On comparing the simulated contact probability matrices for simulations with 2×10^6 and 12×10^6 steps (Fig S1 and Fig S2), we concluded that the polymer has equillibrated in 2×10^6. Hence we used the shorter simulations for further analysis.

Results

Simulations reconstruct experimental Hi-C data

The main purpose of the current work is to develop a quantitative model of the *E. coli* chromosme which is rigorously informed with experimental data and is able to elucidate the higher order organization of 4.64 Mbp *E. coli* chromosome. Towards this end, following the protocol outlined in figure 1, we integrate recently reported Hi-C interaction matrix of *E. coli* chromosome [20] within a polymer based beads-on-a-spring model. Each bead represents a 5 kb nucleotide segment which have been indexed and annotated as per the genetic sequence of wild type *E. coli* MG1655. They have been modelled as non-overlapping van der waals particles, giving rise to a total number of 928($= 4.64 \times 10^6/5 \times 10^3$) beads . The adjacent beads are connected by harmonic springs. The raw Hi-C interaction matrices[20] were processed through a python library (see *Methods*). The matrices obtained were normalized with SCN approach[25] and converted to an inter-bead contact probability matrix (see *Methods*). These contact probabilities are modelled as harmonic restraints with the distances and restraining forces scaled as per equations 4 and 5 in SI *Methods*. The chromosome, thereby modelled, is subsequently

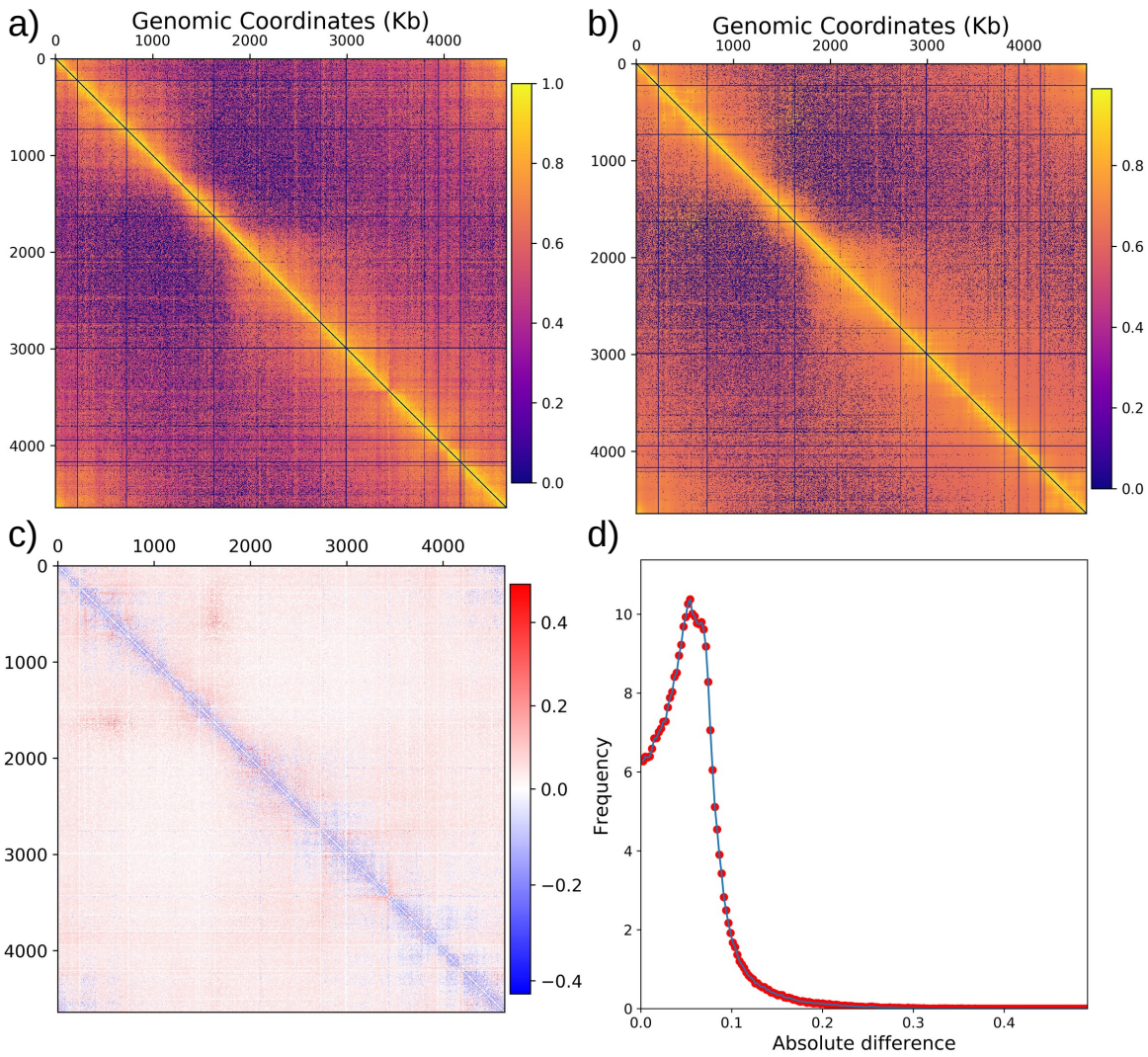

Figure 2: a) heatmap of experimental and **b)** simulated (filtered) Hi-C matrix.**c)** Heatmap of difference between contact probability matrix (simulated - experimental matrix), negative number or blue color indicates higher probability value in experimental matrix and positive number or red color indicates higher probability in simulated matrix. **d)** Histogram of normalized frequency of absolute difference between the two contact probability matrices(experimental and simulated).

subjected to a langevin dynamics simulation at a friction coefficient of 1 $\sqrt{\frac{m\epsilon}{\sigma}}$ and temperature of 303K. The simulation observables are statistically averaged over the last 2000 frames from the 200 trajectories.

Figure 2a and 2b compare the experimental Hi-C contact proabability matrix of *E. coli* chromosome with that obtained from our simulations. A single intense diagonal indicates a smaller distance with a higher contact probability between neighbouring chromosomal regions. Whereas absence of a secondary diagonal in the Hi-C interaction matrix reflects the lack of contacts between the two replication arms of the chromosome. A characteristic feature of prokaryotic chromosome is its circularity. High contact probabilities at the end regions of the Hi-C interaction matrix in Figure 2b (right most upper and left most lower corner) assure circularity of the chromosome in the current model as well.

A Pearson correlation coefficient of 0.89 between experimental and simulated probability

matrix indicates a very good agreement between simulated and in-vivo chromosome conformations. For a direct comparison, we also plotted the heatmap of the difference between the experimental and simulation-derived contact probability matrices (Figure 2c) and the resulting histogram of absolute values from the heatmap (Figure 2d). The difference heatmap shows that for smaller genomic distances (i.e. bins near the diagonal) the contact probability is relatively higher in experimental matrix (blue regions). For rest of the matrix the difference is negligible. The histogram of absolute values from the heatmap in Figure 2c also shows that the disagreement between experiment and simulation is extremely small as the major difference is less than 0.1. Together these analysis suggest a good correspondence between experimental and simulated Hi-C matrices. Thus our model is robust for three-dimensional reconstruction of the *E. coli* bacterial chromosome.

Macrodomain organization underlies the chromosome conformation

Figure 3a is a representative configuration of the *E. coli* chromosome, as obtained from current simulations. Multiple experimental investigations involving *E. coli* chromosome have, in the past, proposed the existence of a set of macrodomains in its genome [8, 9, 7]. Accordingly, in our model, we have color-coded the beads as per the annotation of proposed genetic sequences of chromosome macrodomains. The snapshot in Figure 3a shows that all the annotated macrodomains are well segregated from each other. The separation of macrodomains along the long axis can also be inferred from the distribution of centre of geometry for 4 macrodomains and 2 non structured regions, shown in Figure 3c.

The Left and Right macrodomains have occupied cell's left and right halves which was also found in earlier experiments[29]. Specifically, Ori (red) and Ter (green) macrodomains are aligned at the mid-cell, as visible from the structure in Figure 3a and from their respective position-histograms (Figure 3c). Although position of all the macrodomains are highly variable across replica of trajectories, the position-histograms in Figure 3c suggest that there is effectively an average positional preference for each of the macrodomains inside the cell: Mid-cell for Ori and Ter, left and right half of the cell for Left and Right macrodomains respectively. Interestingly, the representative snapshot also suggests that the chromosome is not completely spread across the entire axial dimension. We also calculated the average linear density profile for the chromosome along both long and and short axes (z- and x- respectively), shown in Figure 3b and d. The chromosome occupies the central region along the long axis without any external restrictions as found in previous experimental and theoretical studies

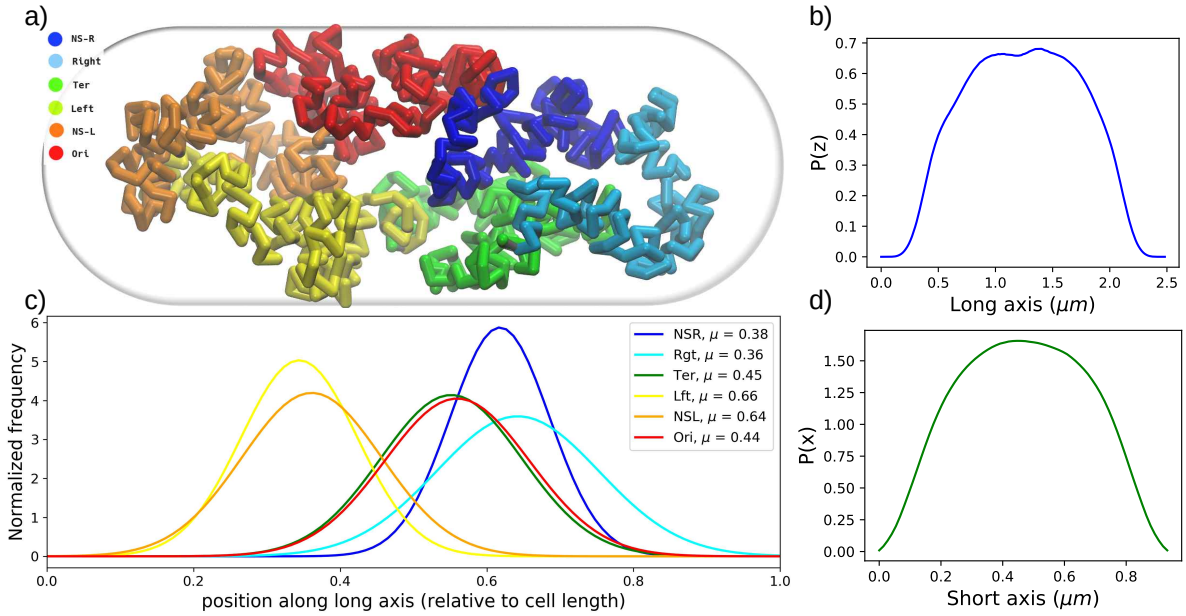

Figure 3: a) Snapshot of the equilibrated chromosome from a trajectory with macrodomains colored and cell dimensions as mentioned in methods. **b)** Linear density of chromosome along z-axis. **c)** distribution of center of geometry of each macrodomain and non-structured regions in 200 trajectories and 2000 frames for wildtype *E. coli* (μ stands for the mean of the distribution). **d)** Linear density of chromosome along the x-axis.

Revisiting genomic architecture of nucleoid in computer simulation

Precedence of FISH data for the interfocal distances between various loci throughout *E. coli* chromosome[7] allowed us to compare distances from our simulated model with experimentally measured distances. A Pearson correlation coefficient of 0.83 (Figure 4a) shows reasonably good agreement between distance calculated from our model and FISH-determined inter-focal distances. Since, the distances are calculated in real units (μm), the slope of the fit gives us the relative size of the simulated chromosome with the in-vivo counterpart. In Figure 4a, a slope of 1.35 indicates that the loci distances are larger in the current model than that found in experimental setup. We speculate that this might have to do with the difference in media and temperature between FISH (25°C in minimal media(MM)) and Hi-C (37°C in LB media) measurements.

For a more comprehensive investigation, we determined the position-distribution of a few loci (genomic locations obtained from *Wiggins et al.*[10]) along the long axis of cell and plotted their mean position (averaged over all trajectories) as a function of their genomic location (relative to that of oriC). As depicted by the linear fit of Figure 4b, we find that the linear organisation of most of the loci along long axis of cell is also retained in the present model of chromosome. However, the Hi-C based model suggests the separation between the loci is smaller than that reported[10] because the fitted line is steeper in the experimental data[10] than our current model. These observations imply that though such elaborate loci arrangement is more evident in slow growing cells, they are also retained in rapidly dividing cells (the present scenario), albeit at a smaller extent.

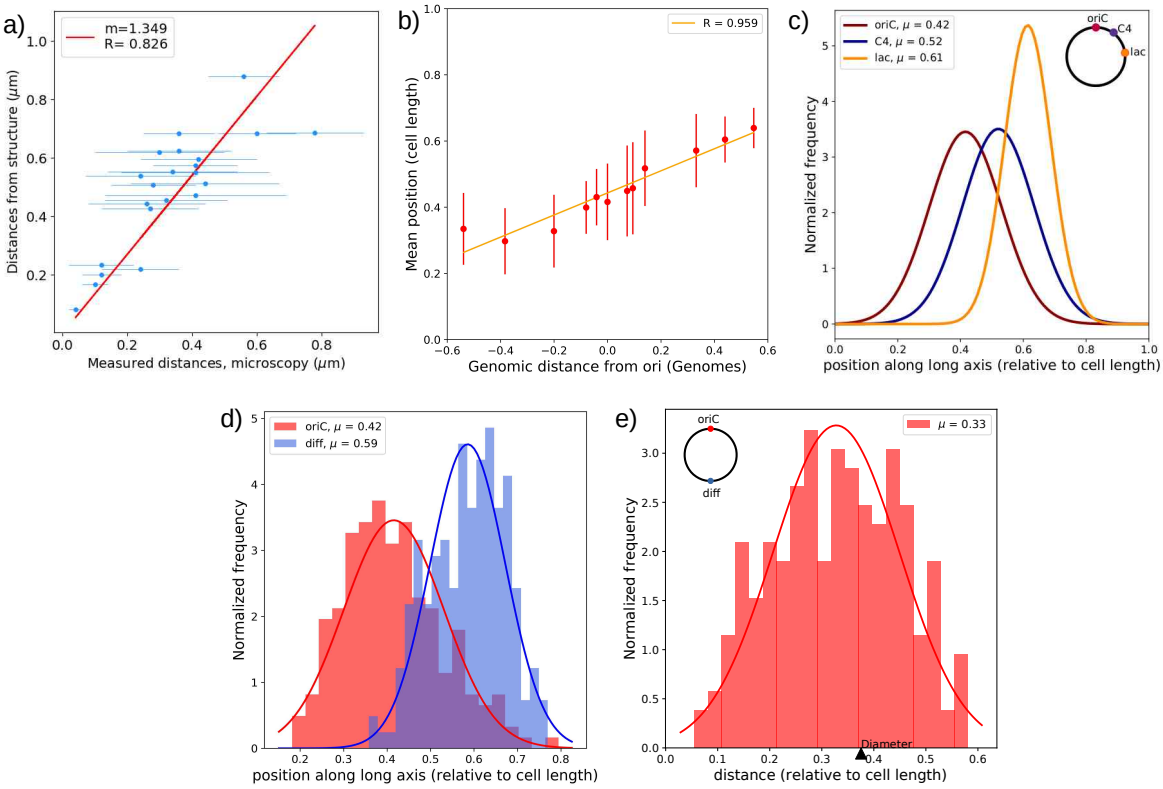

Figure 4: a) distances measured from our simulations (y-axis) for Hi-C data of *E. coli* at 37 $°C$ in LB media vs. distances measured via FISH (x-axis), measured at 25 $°C$ in Minimal media. **b)** mean position along long axis vs. genomic distance from oriC as mentioned by *Wiggins et al. (2010)*[10]. **c)** distribution of position of three loci: oriC, C4 and lac along the long axis(Z-axis). **d)** the distribution of oriC and diff locus with respect to cell length where 0.5 implies the midcell position. **e)** the distribution of distance between oriC and diff locus in 200 simulations and 2000 frames each.

In a fluorescence based assay[10], positions of multiple genetic loci in *E. coli* cells were monitored. From their spatial positioning a 'linearly organised' architecture of the chromosome was proposed. In particular, this investigation hypothesized that these loci are linearly positioned along the long axis of the chromosome. To test this hypothesis in our Hi-C encoded polymer model of chromosome, we plot the distribution of three loci oriC, C4 and lac along long axis of cell(Figure 4c). (These are the same loci position measured in Figure 2A by *Wiggins et al. (2010)*). Our model predicts that these loci are linearly positioned along the long axis of the cell, in accordance with the experiment[10].

Two loci, namely, oriC and diff are known to play a pivotal role in *E. coli* chromosome division and segregation [8, 10, 30]. OriC is the origin of replication from where replication of chromosome starts and diff is another locus present in the Ter macrodomain and opposite to oriC in genomic distance. Diff is the last region on the chromosome to be replicated after which the two chromosomes segregate. In exponentially dividing cells, these two loci have been hypothesized to be present at the mid-cell for majority of the time[29]. To investigate the location of these two loci in our model, we calculated their average positional distribution. As we can see from Figure 4d, the mean position for oriC and diff are 0.42 and 0.59 in cell length along the long axis. The distribution of distance between oriC and diff (Figure 4e) shows that the average distance between the two loci is also approximately equivalent to the diameter of

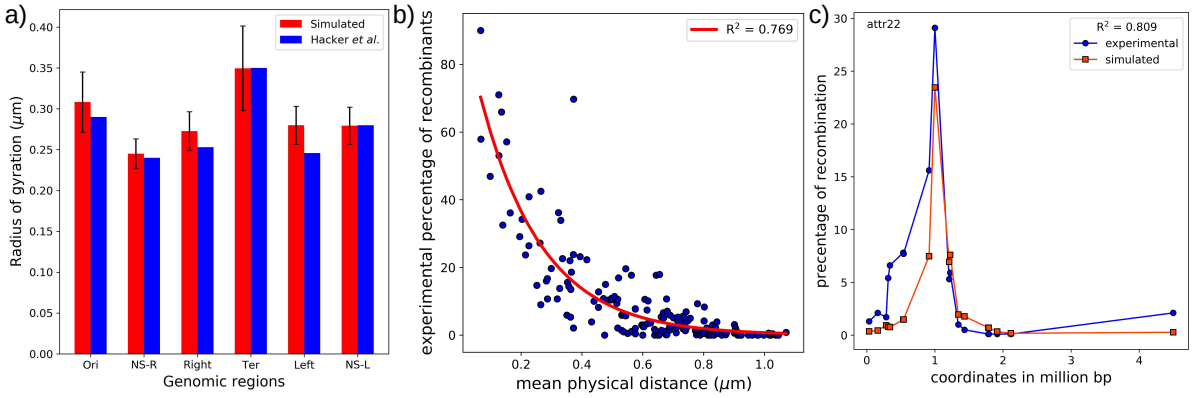

Figure 5: **a**) Radius of gyration from our simulated model as compared to given by *Hacker et al. (2017)*[19] for oriC@midcell plectnomic model. **b)** comparison between the recombination frequencies provided by *Valens et al. 2004* [9] vs. mean physical distance between recombination loci and red line indicates the single exponential fit. **c)** A representative plot of recombination frequencies predicted by mean physical distance, blue: experimental data and red: predicted by simulated data.

the cell (which is around 0.37 units with respect to the cell length). Together, this distribution shows that these two loci are axially present at the mid-cell and radially apart by the cell's periphery, which is an important feature of the rapidly dividing cells in case of *E. coli*.

To determine the approximate size of the macrodomains we calculated the radius of gyration (R_g) for each macro-domain (Figure 5a). Interestingly, we find that R_gs of these domains are matching with R_gs predicted by *Hacker et al.*'s oriC@midcell plectonemic model [19], which is in accordance with our earlier prediction of mid-cell location of oriC loci. We find that Ter has the highest R_g (0.35μm) among all macrodomains and Non-Structured regions (see table S1). This is because in most of the structures, unlike other macrodomains, Ter was present as a linear chain, as evident by its highest end-to-end distance(table S2). This result is also consistent with the model proposed by *Wiggins et al.(2010)*[10] based on fluoroscence-based experiments, that the Ter region acts like a linking part between Right and Left macrodomains which are spanning the right and left cell halves respectively.

We also compared our data with site-specific recombination experiment [9] by fitting an exponential curve between percentage of recombinants and mean physical distance (Figure 5b and c), similar to a previous investigation[19]. The fit suggests a good correlation with R^2 value of 0.769. Figure S3 shows the predicted recombination for other six loci experimented [9] which are also in good correlation with the experimental data.

Local packing of the Chromosome

A common practice for exploring local packing of chromosome is investigation of scaling of contact probability with respect to genomic distance[23]. Earlier, based upon similar analysis of intra-chromosomal contact probability as a function of genomic distance, a power law scaling of contact probability between 500Kb and 7Mb for human chromosome was predicted[23]. The contact probability was found to be scaled as the inverse of genomic distance ($\sim s^{-1}$), which is in agreement with a previously proposed attribute that chromosomes possess, namely, the 'fractal nature'[31]. Fractal polymers are knot free which is important for the segregation of daughter chromosomes during cell replication[23, 32]. For *E. coli*, similar analysis of the intra-genome contact probability as a function of genomic distance, as extracted from experimentally determined Hi-C matrix, produces a scaling factor of ~ -0.77, suggestive of deviation from a

perfect fractal. In our current model (Figure S4a) the contact probability scales as $\sim s^{-0.55}$ in 10Kb to 1Mb range. Therefore the fractal nature might be absent in our model as well. However, on the other hand, scaling of root-mean-squared (RMS) end-to-end distance with genomic distance suggests otherwise (Figure S4b). RMS end-to-end distance scales as $s^{0.36}$ which is close to the expected value for scaling of RMS end-to-end distance with genomic distance for a fractal globule polymer which is $s^{1/3}$. To confirm the fractal nature we explored the presence of knots in our ensemble of structures explicitly using an external python library[33]. Knot analysis revealed that chromosomes in all 2000 frames in 200 simulations are knot free which implies that the chromosome is not an equilibrated globule. Therefore based on current observations, although we are unsure of the presence of perfect fractal nature, we can exclude the possibility of equilibrated polymer nature in the *E. coli* chromosome.

Contact probability scaling with respect to genomic distance provides a one value summary of the organization of the polymer but doesn't show us how the local density of chromosome varies along the chromosome contour. Therefore to determine the variation in local density of the chromosome, we have devised a method based upon moving window averaging of radius of gyration of continuous segments of chromosome. This approach gives rise to a 'radius of gyration map' (R_g map). Using this approach, we calculate the radius of gyration of fine-grained segments for each bead (for example: for n_{th} bead: the R_g will be calculated for 20 beads starting from $n - 10^{th}$ to $n + 9^{th}$ bead) for all the 928 beads in the chromosome. We have calculated the R_g map at four different window sizes: 10, 20, 50, 100 beads, corresponding to 50, 100, 250, and 500 kbp genomic segments respectively, to investigate the features of our model pertaining to the local density of chromosome at different scales.

Figure 6a shows the Rg map for wild type *E. coli* (grown at 37°C in LB) at a window size of 20 beads or 100 kbp region. The black vertical lines, in the Rg map, highlight the most prominent peaks which were calculated with the help of directionality index (DI) at 100 kbp size. Previously it was shown that the DI boundaries, when visualised with the heat map of matrix, appear as the vertices of the triangles along the diagonal of the matrix. Therefore, we also compared our Rg map with the diagonal of the heat map (shown in Figure 6b). The peak positions calculated form the Rg map are also marked in the Figure 6b and qualitatively we can see that these peaks are present at the vertices of triangles in most of the cases. For further validation, we also compared the peak positions with the DI boundaries (Figure 6c) and they are in good agreement with each other (R = 0.99). Therefore, the comparison in Figure 6 suggests that our model is capable of capturing local chromosomal structures such as chromosomal interaction domains) (CID). Therefore, Rg map comes out to be an effective tool in unravelling the local structure from our simulation model.

In a study on *Caulobacter crescentus*, it was found that, generally, one or many highly expressed genes were present at the CID boundaries[21]. In rapidly dividing cells, the rrn operons are expressed at a higher rate than other metabolic genes due to the requirement of ribosomes for protein synthesis[34]. Since, the Hi-C experiment was also performed on rapidly dividing cells, high expression of rrn operons was expected. In another study, rrn operons showed higher transcriptional propensity, measured as RNA/DNA ratio, and overlap with CID boundaries[35]. Therefore, motivated by previous experiments [21, 34, 35] and good agreement between peaks in R_g map and CID boundaries, we compared the peaks in R_g map with rrn operons' genomic locations in *E. coli*. We found that seven of the peaks in R_g map correspond to the genomic location of rrn operons as shown by grey vertical lines in Figure 6b.

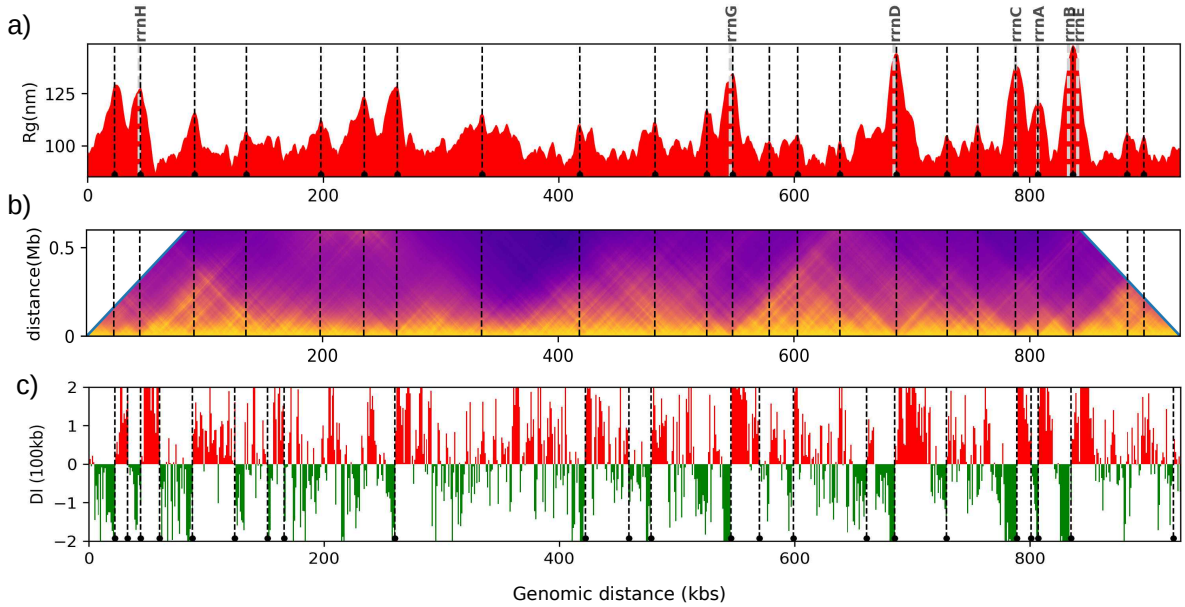

Figure 6: a) Radius of gyration map for wildtype *E. coli* (MG1655 in LB at $37°C$ in exponential growth phase) with respect to genomic coordinates with a moving window size of 20 beads (100kbp), black and grey vertical lines represent the peaks in the map and rrn operon sites respectively. **b)** The main diagonal of the average contact probability matrix rotated 45° clockwise with black vertical lines showing the position of most prominent peaks determined in R_g map. **c)** Directionality index (100kb) for simulated chromosome for wildtype *E. coli* (MG1655 in LB at $37°C$ in exponential growth phase) black vertical lines shows the CID (chromosome interaction domains) boundaries, (calculated as given by *Le et al.(2013)*[21]).

Physically, higher R_g value indicates that the chromosome segment (in case of Figure 6b - 20 beads) is occupying more volume than the adjacent segments of same length. Similarly, lower R_g value implies that the segment is compact and occupies lower volume than its adjacent segments. Therefore our results imply that the local DNA density is lower in the vicinity of highly expressed genes, a phenomenon earlier observed in the eukaryotic cells[36]. These results suggest similarity with the 'transcription centric' approach used in a previous model[19]. These low density regions i.e. with high R_g value in R_g map are thus equivalent to plectoneme free regions(PFR)[19]. Together, our model, via encoding Hi-C data, is able to capture all the spatial informations of the chromosome such as macro-domain structure, plectonemes, CIDs, and transcription details.

Insights on role of NAPs in higher order organization of chromosome

Whole chromosome conformation is a result of complex interactions between regulatory and structural elements such as promoter and repressors, and nucleoid associated proteins (NAPs) respectively. Few of the well known NAPs are HU (heat unstable or histone like), FIS (factor for inversion stimulation), and MatP (macrodomain Ter protein). To investigate the role of these proteins in maintaining chromosome structure, we use our method to develop models of chromosome using Hi-C matrices of mutant cells in which these individual proteins were knocked out[20] and subsequently compared it with appropriate wild type control data.

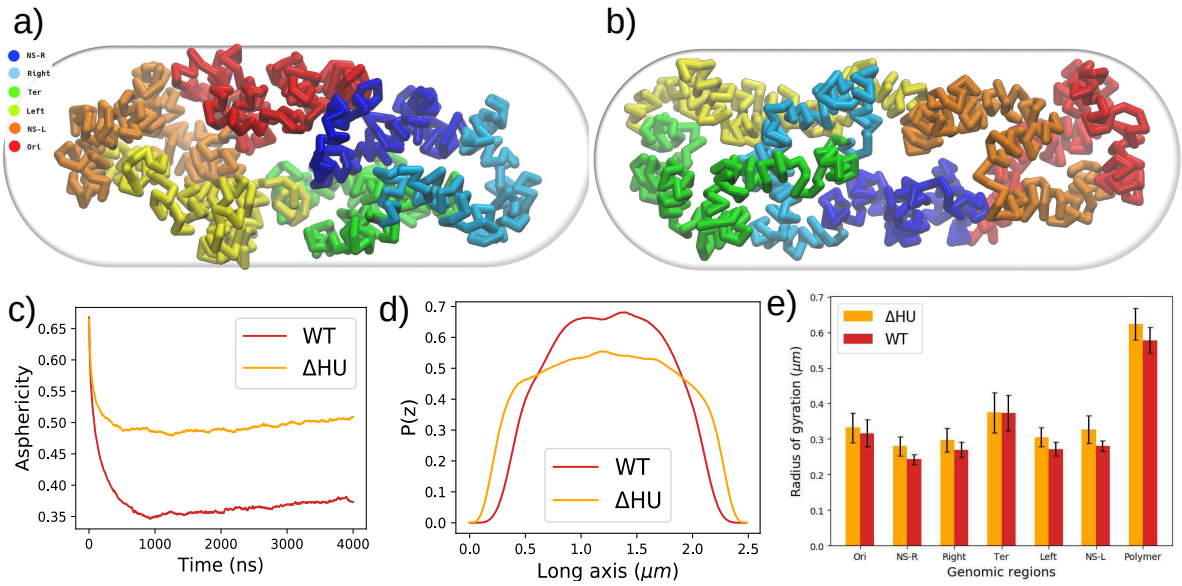

Figure 7: **a)** Snapshot from equilibrated structure for wildtype *E. coli* chromosome simulation. **b)** Snapshot from equilibrated structure for hupAB mutant chromosome simulation. **c)** Asphericity of wildtype and ΔhupAB mutant cells throughout equilibration. **d)** Linear DNA probability density for wildtype and ΔhupAB mutant cells relative to cell length along the long axis (Z- axis). **e)** Radius of gyration (R_g) comparison between wildtype cells and ΔhupAB mutant at same growth conditions.

HU Protein

HU is one of the most abundant and highly conserved nucleoid associated protein (NAP) in bacteria. In rapidly dividing cells its expression is highest[37] where it plays an important role in DNA compaction and negative supercoiling[38]. Therefore to ascertain the effect of HU knockout on whole chromosome structure we have carried out simulations by encoding Hi-C interaction matrices of HupAB mutant cells (*E. coli* MG1655 ΔhupAB at $37°C$ in LB media), further details are given in Table S4. Experimental and simulated (Figure S5a and S5b respectively) Hi-C matrices of the hupAB mutant are in excellent agreement with a correlation coefficient of 0.95. Our simulations predict that whole chromosome structure decondenses in hupAB mutant (Figure 7b), when compared with wildtype (Figure 7a) and gets dispersed across whole cell dimension, consistent with hypothesis of previous experiment[39]. This is reflected in the relative increase in asphericity of the chromosome and axial density near the pole region in hupAB mutant when compared to wild type chromosome (Figure 7c and 7d, Table S3). For more evidence we also calculate radius of gyration for whole chromosome and macrodomains in hupAB mutant, given in Figure 7e. We find that, for each macrodomain, the R_g has increased on an average by $0.030 \pm 0.009\mu m$ and for whole chromosome by $0.082\mu m$. Effect of hupAB knockdown on each individual macrodomain is more apparent in their individual spatial distance versus genetic distance plot (Figure 8). Our simulations predict that Ter macrodomain is least affected by the hupAB knockout. This is due to fact that Ter remains mostly insulated from rest of the chromosome[20, 30], which leaves Ter mostly unchanged in the face of most of the mutations. However, for regions in other macrodomains, especially regions with large genomic distances, their geometric distances increase significantly in hupAB mutant, compared to that in wild-type cell.

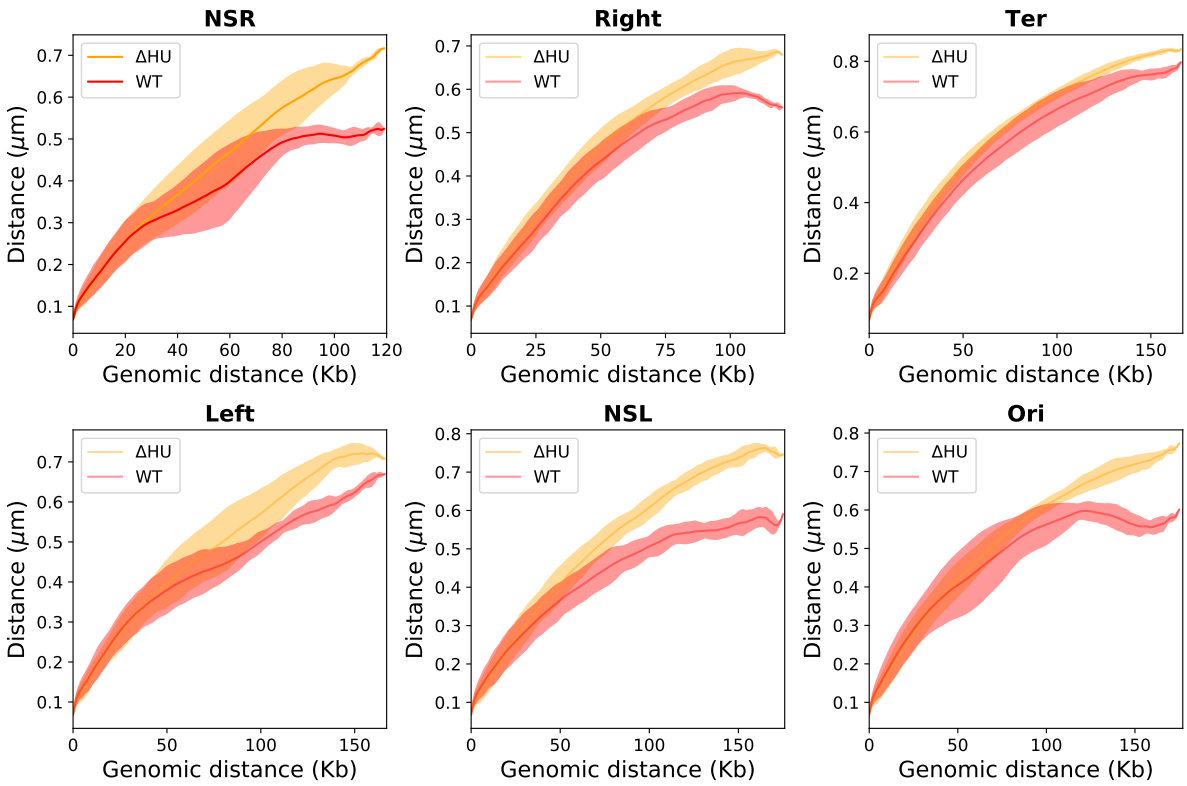

Figure 8: Rms end-to-end distance vs. genomic distance plot for intra-domain distances in hupAB mutant (orange) and wildtype cells at 37°C in LB media (red).

These results from our simulations are consistent with experiments which hypothesized that HU is involved in long range interactions ($> 280kb$) in whole chromosome[20]. As observed in our simulations with the hupAB mutant cell, when compared with wild type *E. coli*, the long range interactions have diminished but local or short range interactions were not affected significantly as evident by the R_g map plot in Figure S5e.

FIS protein

Fis is one of the most abundant NAPs in rapidly growing *E. coli* cells [37]. Its expression is maximum during exponential growth but negligible in stationary phase[37]. We note that the experimental Hi-C matrix of fis mutant cells correspond to a growth condition of minimal media (MM) at 30°C temperature (in contrast to hupAB mutant which corresponds to LB media and temperature of 37°C). Accordingly, for comparison purpose with fis mutant cells, we have also individually modelled the wild type chromosome in MM at 30°C (Figure 9a). In simulations, the chromosome appeared to be compact with respect to wild type cells grown in LB media at 37°C. The average radius of gyration for current chromosome is approximately 0.43μm (Figure S6f) which is quite small when compared to the cells grown in rich media. A smaller R_g value suggests that the whole chromosome's average diameter will be approximately 0.86μm, which is a good estimate for smaller cells with lengths around 1 to 1.5μm as determined in experiments[1, 40]. When compared with FISH data, the the correlation increased from 0.83 in LB at 37°C to 0.87 in MM at 30°C and slope decreased towards 1(Figure S6e). The macrodomains became less structured as their R_g is similar for all of the macrodomains (Figure

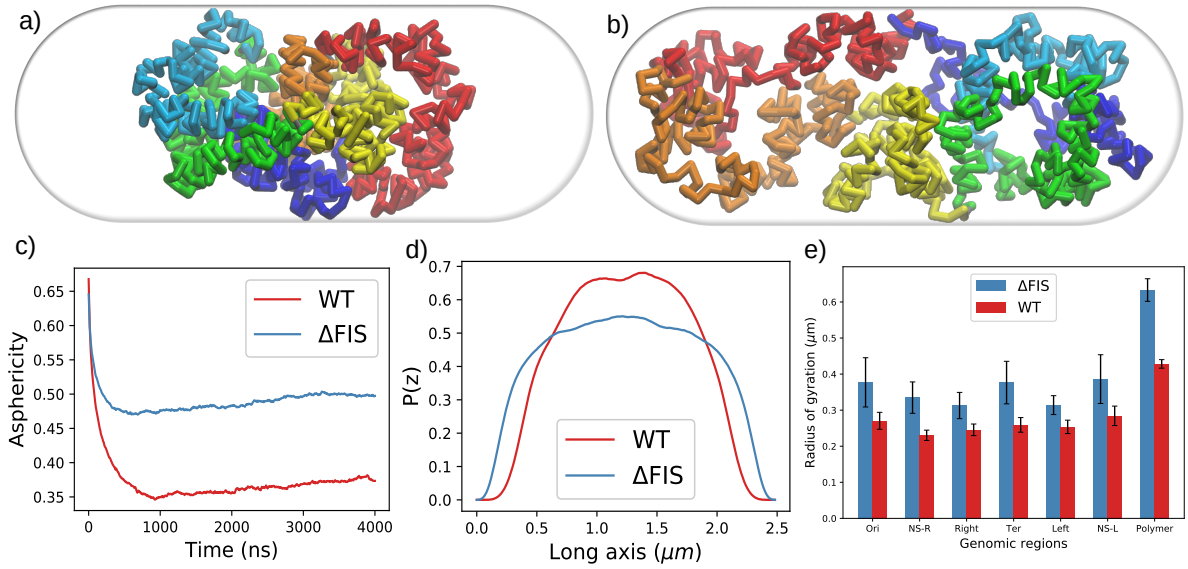

Figure 9: **a)** Snapshot from equilibrated structures for wildtype *E. coli* chromosome simulation. **b)** Snapshot from equilibrated structures for fis mutant chromosome simulation. **c)** Radius of gyration (R_g) comparison between wildtype cells and Δfis mutant at same growth conditions. **d)** Asphericity of wildtype and Δfis mutant cells throughout whole simulation ($4ns$). **e)** Linear DNA probability density for wildtype and Δfis mutant cells relative to cell length along the long axis (Z- axis)

S6f). Thus they have less variability in their packing. Although the chromosome is quite compact for minimal media grown cells but overall features are similar to rich media grown cells (as shown in previous sections) which was also observed in experiments[20].

When we look at the fis mutant results, experimental and simulated matrices (Figure S7a, and S7b respectively) are in mutually good agreement (R = 0.92). Similar to hu, fis is also distributed throughout the chromosome, in both transcriptionally active and inactive regions[41] where it suppresses the expression of non-essential genes[42]. Figure 9b shows one of the representative conformations from simulations for the fis mutant cells, which is visibly more expanded than its corresponding wild-type chromosome at similar conditions (in MM at $30°C$, Figure 9a). The relative change in asphericity and axial chromosome density, in comparison to wild type as shown in Figure 9c and 9d respectively, also predict the relative expansion of the chromosome in fis mutants. R_g values, given in Figure 9e, suggest that each of the individual macrodomains are affected to the similar extent due to lack of fis but the overall effect on whole chromosome is more dramatic than individual macrodomains. This is more prominently visible through their intra-domain rms end-to-end distance plot given in Figure S8.

Our results suggest qualitative similarity in the effect of hupAB and fis knockdown on bacterial chromosome structure. Distribution of centre of geometry for each macrodomains is similar for both models as shown in Figure S9c and S9b. However, the changes in whole chromosome conformation are more dramatic in case of fis mutant than hu, with respect to their wild-type counterparts. The asphericity and axial density of model chromosome for both the mutants are similar as given in table S3. When we compare these effect with their wildtype counterparts (wildtype *E. coli* at $37°C$ in LB and $30°C$ in MM for hupAB and fis mutant, respectively), the asphericity increases from 0.134 to 0.499 in case of fis whereas for hu mutant it only increases from 0.374 to 0.503. Same is true for linear density as well. Since hu helps in sustaining negative supercoils in the chromosome along with gyrase (gyrB)[43] and similar role was found for fis [44], it is probable that the effect of knockout of any of the two proteins

will have similar effect, which we are observing in our computational model as well.

As a part of dual role of fis in gene regulation and chromosome organization, it is known to mediate formation of dense clusters or clumps of DNA at highly transcribed regions such as rrn operons[45]. Through insertion of fluorescent protein binding sites and epifluorescence microscopy, all rrn operons but rrnC were recently shown to colocalize and form a structure similar to eukaryotic nucleolus of the size of around 80 to 130nm[46]. On the other hand, deletion of nearby fis binding sites resulted in disruption of cluster. To quantify the role of fis on potential clustering of rrn operons, we calculated the distribution of distance between all 7 rrn operons as pairs (21 pairs) as shown in Figure S10, and mean values are given in table S5. When compared, pairs with rrnC and without rrnC did not show any sign of colocalization in the wildtype itself. The distribution is independet of genomic distance between the two operons for both wildtype and fis mutant but generally increases with genomic distance. Also, the average distances (table S5) are larger than what was reported in the original experiment[46], ranging from $200 - 750$nm in wildtype and $\sim 300 - 1000$nm in fis mutant. However, we did see a relative increase in the distance between the rrn operons in the fis mutant but it could be due to the overall expansion of the chromosome, rather than any specific effect of fis mutation on the operons. The results are in accordance with the original Hi-C experiment[20] in which the contact frequency didn't indicate any clustering in the rrn operons. In our simulations as well we weren't able to find clustering of rrn operons. Althogether, our results also follow the similar trend as Hi-C experiment[20] and contradict the results of fluorescence study[46]. As suggested, to shed more light on the issue, more experiments are neended[47].

MatP proteins

MatP is a crucial protein for cell division and chromosome segregation, it binds specifically to the Ter macrodomain in the *E. coli* chromosome. MatP has 23 binding sites (called as matS) in whole chromosome (most of which are in Ter)[30]. Currently, two working mechanisms have been proposed for Ter macrodomain insulation by matP protein, based on structural studies[48] and Hi-C experiment[20]. First is, by forming matP tertramers and bridging distant matS sites[48], and second, by excluding mukBEF condensin from Ter macrodomain. Since mukBEF forms long range contacts, Ter interactions are limited to its own self[20]. To determine the effect of matP knockdown on Ter macrodomain along with whole chromosome, we performed similar simulations as in previous sections with Hi-C data from matP knockout cells. Processed Hi-C data details are given in table S4.

Experimental and simulated contact probability matrices (Figure S11a and S11b respectively) are in reasonable mutual agreement with a correlation coefficient of 0.93. Our results are similar to that obtained expereimentally[20]. We see that in the matP mutant the contact probability is similar to the wild type throughout the chromosome. However structurally, the chromosome has become more loosely packed in matP mutant than in wild type cells, shown in Figure 10b and 10a respectively. The spread of the overall chromosome can be seen from the axial density plot (Figure 10c) and the asphericity comparison (Figure 10d). Similar results were obtained from radius of gyration of macrodomains and whole chromosome as shown in Figure 10e. Interestingly, the R_g has increased for only Ori, NSL, and Left macrodomains whereas for Ter, NS-R, and Right it remained unchanged, but average end-to-end distance has increased for all macrodomains except NS-R (table S2). Intradomain distance plot, shown in Figure S12 also predicts similar results i.e. Ter, NS-R, and Right macrodomain's intradomain distance is similar in case of matP mutant and wildtype control.

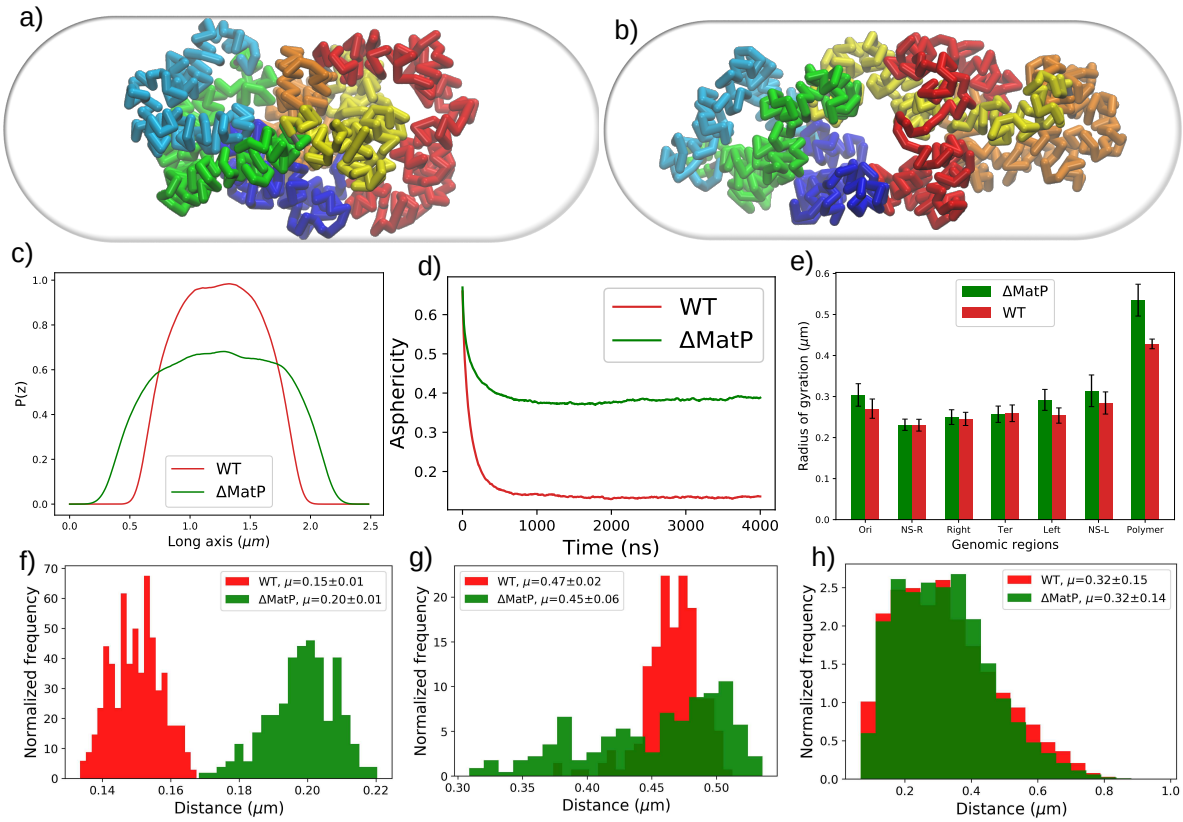

Figure 10: **a)** Snapshot from equilibrated structures for wildtype *E. coli* chromosome simulation. **b)** Snapshot from equilibrated structures for matP mutant chromosome simulation. **c)** Radius of gyration (R_g) comparison between wildtype cells and matP mutant at same growth conditions. **d)** Asphericity of wildtype and matP mutant cells throughout whole simulation. **e)** Linear DNA probability density for wildtype and matP mutant cells relative to cell length along the long axis (Z- axis). **f)** distribution of distance between all 23 matS sites in the *E. coli* genome in wildtype and matP mutant cells. **g)** distance between ter2 and ter4 loci which are 100kb apart in Ter region in *E. coli* chromosome for wildtype and matP mutant cells. **h)** distance between ter2 and ter6 loci which are $\sim$ 350kb apart in Ter region in *E. coli* chromosome for wildtype and matP mutant cells.

Apparently, our results for matP mutant seems to contradict the results of a recombination assay experiment[30]. In the recombination assay experiment for matP mutants, the interaction frequency between Right and Ter macrodomain was increased in contrast with wildtype, indicating Ter disorganization. On the other hand our simulation results show low or negligible disorganization of Right and Ter macrodomain. When we compared the distances between two such loci (ter2 and ter4), which are 100kb apart in Ter region (Figure 10f), we observed an increase in the distance in matP mutant cells. These two loci show that the parts of Ter macrodomains are disorganized while keeping overall properties same. We also calculated the distance between ter2 and ter6 which are $\sim$ 350 kb apart (Figure 10g). We found that in wildtype the average distance is $0.466\mu m$ whereas in matP mutant it decreased to $0.451\mu m$. But in case of wildtype the variance is very low implying greater rigidity in motion between these two loci. In matP mutants, though the average is lower but the variance is more than double from wildtype which indicates the flexibility and disorganization between these two loci. Therefore these results suggest an apparent disorganization of Ter region without significant change in its overall structure.

Similar to the Hi-C experimental results[20], any significant changes were missing in matS

site's contact probability in wildtype and matP mutant cells. As shown in Figure 10h, the distribution of distance between all 23 matS sites showed only slight variation between wildtype and matP mutant. Therefore our results further supporting the argument by *Lioy et al.*[20] that the Ter organization doesn't require matP protein to bring the matS sites in close proximity.

Discussion

General details of the model

For decades the understanding of chromosome organization inside cells and nucleus remained elusive. But with the advent of Chromosome Conformation Capture techniques[23, 49], we are able to determine the interaction frequency of millions of loci throughout the whole genome of any organism. Since Hi-C provides an averaged data over a large ensemble of cells, it is inherently informed with the stochasticity and fluctuations underlying any living cell. More over, even with the help of single cell Hi-C, it has been recently shown that the chromosome conformation at megabase pair scale is similar even in heterogenous population of cells[50]. Therefore we undertook a simulation approach that can integrate the Hi-C data within a physics-based model of chromosome and determine the 3D structure of the chromosome with an in-built stochasticity for the experimental procedure. Starting from the random walk initial configurations our method generates an ensemble of structures of the chromosome whose average properties resemble that of ensemble experimental observables. The chromosome model predicted by the current simulation recovers the major features of the Hi-C interaction matrices reported by recent experimental investigation [20]. Despite starting with a linear polymer chain, our model, being data informed, recapitulates the circularity of *E. coli* chromosome, as confirmed by the high interaction probabilities between the corners of the off diagonal. We also found, in all the wildtype and mutant simulations, the average distance between the end beads is $1.657 \pm 0.122\sigma$, which implies that the end beads are in close proximity. The presented model of chromosome is found to be devoid of knots and lean closely towards a fractal like morphology[23, 31]. Knots, if present during interphase, would result in entangled chromosomes after replication which can hinder chromosome segregation. When we checked for knots in the equilibrated conformation of our models, we found that for all of them, the Alexander polynomial [51] was zero, implying that the structures were knot free. This adds to our belief that our model describes correctly the bacterial chromosome at a coarse grained level.

From equation 4 in *SI Methods*, we can see that only a small part of the Hi-C contact probabilities will actually be large enough to make an impact on the chromosome conformation. To emphasize on the effect that Hi-C makes, we performed a set of control simulations (which are simulations devoid of Hi-C bonds). From Table-S4, it can be seen that only a minor percentage of contacts has been used in each case (these contacts have magnitude of force constant greater than $10^{-6} kJmol^{-1}\sigma^2$). Also we want to emphasize here that along with inverse function for distance calculation from probability matrix, we are also using a vairant of gaussian function to define the force constant values for each Hi-C bond. Here each Hi-C bond is subjected to a unique force constant whose value is determined based on its Hi-C contact probability which makes longer bonds weaker. This feature appears to be a key step in our method which is absent from most of the previous models.

Comparison with existing data

As described in the *Results* section, with the model developed in the current work, we were able to reproduce diverse range of the experimental observations for *E. coli* chromosome. Most prominently, the constructed chromosome model vividly recapitulates the phase-segregation of the chromosome into 4 macrodomains and 2 non-structured regions. In wildtype, our model located oriC and diff preferably at the midcell position, which is important for the chromosome segregation [29, 52]. The positions of Left and Right macrodomains are in accordance with a linear organized architecture of the chromosomes. However we find that the extent of linear order in the current model is less pronounced than originally proposed by fluoroscence-based assay[10]. Most importantly, the close agreement between the distances measured in the current model and prior FISH measurements [7] and site-specific recombination assay[9] also provides a quantitative basis to the macrodomain phase-separation present in the current chromosome model.

When we compared the R_gs obtained with another model, which had incorporated plectonemes [19], we found that our R_gs matched best with the case where oriC was at midcell with a plectonemic polymer model. This not only reveals the multitude of experimental and theoretical data with which our model is consistent with, but also reveals that Hi-C interaction matrices are effectively informed with formation of plectonemes. It might be hard to show plectonemes with our model visually. The average length of a pelctoneme is about 10kbp[53]. At 5kbp resolution a 10kbp plectoneme would only be equivalent 2 polymer beads. However with data at sub kbp resolutions, the length of the plectonemes in the polymer model would be sufficient enough to be detected properly. Signatures of plectonemes being encoded in Hi-C interaction matrices is somewhat expected as negative supercoiling during plectoneme formation results in the regions of the chromosome to come very close, which Hi-C interaction matrices detect as contacts.

Local chromosome structure in wildtype and mutants

Apart from higher order structure such as MDs, we were able to describe the local structure of the *E. coli* chromosome through a physical measurement which we had referred as radius of gyration map. It calculates the average radius of gyration of small segments of the chromosome in a moving window averaging manner and the contour of that map gives us the local density of the chromosome. We found that the peaks in the R_g map were correlating well with the CID boundaries. Higher and lower value of the radius of gyration indicates loosely and densly packed regions in the chromosome, respectively. Therefore our observations from R_g map further, support the argument that the CID boundaries represent the highly expressed genes[21]. These highly experessed genes are plectoneme-free and thus occupy more volume which shows as peaks in the R_g map. For further validation we also compared peak positions with the rrn operons because in rapidly growing cells, rrn operon expression is higher than any other gene in the cell. More over, our results show that out of many peaks, 6 prominent peaks belong to 7 rrn operons (2 rrn operons were in single peak due to very less genomic separation). Therefore our model is capable of showing finer details at the genes/operon scale.

An interesting feature that we noticed in all the mutant and wildtype samples was strong correlation among R_g maps, shown in Figure S13. For smaller regions i.e. at 50-100kb (10, 20 beads) the correlation between R_g map is high for all the mutants and wildtype with a net increase from 50kb to 100kb except for hu37 - matP30 pair. Gradually, it decreases for larger segments starting from window size of 50 beads. However, the correlation between

hu37 and fis30 is almost constant for all window sizes, supporting our claim that the effect of fis and hupAB deletions are similar. The overall comparison suggests that in all mutants and wildtypes the local chromosome packing remains similar which is expected because chromosome organization is largley regulated by transcription. For a cell to survive, expression of essential genes is required, which is going to be same in all cells. Therefore at scale of 50-100kb the overall structure of chromosome is same but at larger scale activity of genes such as hu, fis, matP and growth conditions play a major role.

Role of NAPs in chromosome folding

The current model provides a quantitative basis of the stabilization effects of various NAPs on *E. coli* chromosome. Effect of fis and hupAB mutations were similar in overall structural changes i.e. expansion of whole chromosome and reduction in small range contact. However, when compared with their respective wildtype counterparts, fis deletion showed much more drastic changes than hupAB deletion. Another interesting feature of fis protein is its role in formation of nucleolus-like structure by rrn operons in *E. coli*[46]. But when we computed the average distance between all rrn operons from the wildtype *E. coli* itself, we did not observe any colocalization and the average distance also varied extensively. Although the average distance was increased in case of fis mutant, it was mainly due to overall chromosome expansion. Similarly, the signature of colocalization was absent in original Hi-C interaction matrices, which further supports our results. Another protein, matP is important for chromosome positioning and segregation during cell division. When we simulated chromosome for matP mutant cell, we observed an overall chromosome expansion but individual macrodomain properties remained similar to wildtype. Although the effect on Ter macrodomain seems to be negligible, but the distance between two loci, in Ter region (ter2 and ter4, 100Kb apart), was increased in the matP mutant. Also, loci which were $\sim$ 350kb apart (ter2 and ter6), showed increased variance in distance between them, indicating their disorganization in matP mutant cells. Finally, the matS sites where matP protein binds showed very less variation than wildtype, similar to what was observed in Hi-C experiment. Therefore effect of matP on *E. coli* chromosome appears to negligible but Ter macrodomain showed disorganization along with increased size of Left, and Ori macrodomains and Non-Structured Left.

NAPs' binding sites are a few base pairs long. These sites may be distributed throughout the chromosome, like for hu, fis or mukBEF[5, 20, 54], or can be specific to some macrodomain, which is the case for matP[20, 30]. But the current model is limited by the resolution of Hi-C interaction matrices and hence consider a coarse graining of 5000 bp to 1 bead, where multiple binding sites can be present in one coarse grained bead. This leads to our model being unable to answer questions which are specific to the role of NAPs in gene regulation or the change in the local structure of the chromosome due to the mutations. However, the current work has been largely successful in determining the effect of knock-outs on the overall structure of the chromosome. The model have also been able to mark the important regions of the chromosome, such as CIDs and rRNA operons, which are transcriptionally very active, from our coarse grained model.

Conclusion

In summary, we report a Hi-C data-integrated comprehensive model of *E. coli* chromosome at 5 kb resolution. The model was able to recover and represent the extent of information Hi-C encodes. We noticed that chromosome architecture mainly depends on two factors: *i)* relative

distance among regions of the chromosome, *ii)* relative compactness among the regions. These also decide the transcriptional intensity of a segment of chromosomal DNA. NAPs regulate these two factors for proper growth of the bacteria. We also see that growth conditions affect both the factors which in turn causes difference in shape, size and division time of the bacterial cell. Though we did not investigate the dynamics of the chromosome, but by incorporating mass of each bead one can, in practice, explore the chromosomal dynamics. Currently our model does not include replication arms, but its incorporation would yield a more exact picture of the *E. coli* chromosome. Incorporation of other experimental data would act as refinements on the ensemble averaged Hi-C which has been used as the basis for the modelling. Multiscale simulations can also be attempted in which the coarse-grained interactions can be designed the way we incorporated bonds using Hi-C. In conclusion we can say that our model has been able to predict other independent experimental results from Hi-C only. A lot of improvements can still be performed which can help us reach the goal of achieving a correct, unified and robust chromosome model.

Acknowledgements

This work was supported by computing resources obtained from shared facility of TIFR Centre for Interdisciplinary Sciences, India. JM would like to acknowledge research intramural research grants obtained from TIFR, DAE, India, Ramanujan Fellowship and Core Research grants provided by the Department of Science and Technology (DST) of India (CRG/2019/001219).

References

[1] G Reshes, S Vanounou, I Fishov, and M Feingold. Timing the start of division in e. coli: a single-cell study. *Physical biology*, 5(4):046001, 2008.

[2] Benjamin Volkmer and Matthias Heinemann. Condition-dependent cell volume and concentration of escherichia coli to facilitate data conversion for systems biology modeling. *PloS one*, 6(7), 2011.

[3] Mathew Stracy, Christian Lesterlin, Federico Garza De Leon, Stephan Uphoff, Pawel Zawadzki, and Achillefs N Kapanidis. Live-cell superresolution microscopy reveals the organization of rna polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences*, 112(32):E4390–E4399, 2015.

[4] Esteban Toro and Lucy Shapiro. Bacterial chromosome organization and segregation. *Cold Spring Harbor perspectives in biology*, 2(2):a000349, 2010.

[5] Thøger J Krogh, Jakob Møller-Jensen, and Christoph Kaleta. Impact of chromosomal architecture on the function and evolution of bacterial genomes. *Frontiers in microbiology*, 9:2019, 2018.

[6] Jagannath Mondal, Benjamin P Bratton, Yijie Li, Arun Yethiraj, and James C Weisshaar. Entropy-based mechanism of ribosome-nucleoid segregation in e. coli cells. *Biophysical journal*, 100(11):2605–2613, 2011.

[7] Olivier Espeli, Romain Mercier, and Frédéric Boccard. Dna dynamics vary according to macrodomain topography in the e. coli chromosome. *Molecular microbiology*, 68(6):1418–1427, 2008.

[8] Hironori Niki, Yoshiharu Yamaichi, and Sota Hiraga. Dynamic organization of chromosomal dna in escherichia coli. *Genes & Development*, 14(2):212–223, 2000.

[9] Michèle Valens, Stéphanie Penaud, Michèle Rossignol, François Cornet, and Frédéric Boccard. Macrodomain organization of the escherichia coli chromosome. *The EMBO journal*, 23(21):4330–4341, 2004.

[10] Paul A Wiggins, Keith C Cheveralls, Joshua S Martin, Robert Lintner, and Jané Kondev. Strong intranucleoid interactions organize the escherichia coli chromosome into a nucleoid filament. *Proceedings of the National Academy of Sciences*, 107(11):4991–4995, 2010.

[11] Mark A Umbarger, Esteban Toro, Matthew A Wright, Gregory J Porreca, Davide Bau, Sun-Hae Hong, Michael J Fero, Lihua J Zhu, Marc A Marti-Renom, Harley H McAdams, et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Molecular cell*, 44(2):252–264, 2011.

[12] Miriam Fritsche, Songling Li, Dieter W Heermann, and Paul A Wiggins. A model for escherichia coli chromosome packaging supports transcription factor-induced dna domain formation. *Nucleic acids research*, 40(3):972–980, 2012.

[13] Julien Dorier and Andrzej Stasiak. Modelling of crowded polymers elucidate effects of double-strand breaks in topological domains of bacterial chromosomes. *Nucleic acids research*, 41(14):6808–6815, 2013.

[14] Kang-Jian Hua and Bin-Guang Ma. Evr: reconstruction of bacterial chromosome 3d structure models using error-vector resultant algorithm. *BMC genomics*, 20(1):1–10, 2019.

[15] Charlène Planchenault, Marine C Pons, Caroline Schiavon, Patricia Siguier, Jérôme Rech, Catherine Guynet, Julie Dauverd-Girault, Jean Cury, Eduardo PC Rocha, Ivan Junier, et al. Intracellular positioning systems limit the entropic eviction of secondary replicons toward the nucleoid edges in bacterial cells. *Journal of Molecular Biology*, 432(3):745–761, 2020.

[16] David S Goodsell, Ludovic Autin, and Arthur J Olson. Lattice models of bacterial nucleoids. *The Journal of Physical Chemistry B*, 122(21):5441–5447, 2018.

[17] Jing Huang and Tamar Schlick. Macroscopic modeling and simulations of supercoiled dna with bound proteins. *The Journal of chemical physics*, 117(18):8573–8586, 2002.

[18] Chris A Brackley, Stephen Taylor, Argyris Papantonis, Peter R Cook, and Davide Marenduzzo. Nonspecific bridging-induced attraction drives clustering of dna-binding proteins and genome organization. *Proceedings of the National Academy of Sciences*, 110(38):E3605–E3611, 2013.

[19] William C Hacker, Shuxiang Li, and Adrian H Elcock. Features of genomic organization in a nucleotide-resolution molecular model of the escherichia coli chromosome. *Nucleic acids research*, 45(13):7541–7554, 2017.

[20] Virginia S Lioy, Axel Cournac, Martial Marbouty, Stéphane Duigou, Julien Mozziconacci, Olivier Espéli, Frédéric Boccard, and Romain Koszul. Multiscale structuring of the e. coli chromosome by nucleoid-associated and condensin proteins. *Cell*, 172(4):771–783, 2018.

[21] Tung BK Le, Maxim V Imakaev, Leonid A Mirny, and Michael T Laub. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, 342(6159):731–734, 2013.

[22] Marie Trussart, Eva Yus, Sira Martinez, Davide Bau, Yuhei O Tahara, Thomas Pengo, Michael Widjaja, Simon Kretschmer, Jim Swoger, Steven Djordjevic, et al. Defined chromosome structure in the genome-reduced bacterium mycoplasma pneumoniae. *Nature communications*, 8(1):1–13, 2017.

[23] Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.

[24] Asli Yildirim and Michael Feig. High-resolution 3d models of caulobacter crescentus chromosome reveal genome structural variability and organization. *Nucleic acids research*, 46(8):3937–3952, 2018.

[25] Axel Cournac, Hervé Marie-Nelly, Martial Marbouty, Romain Koszul, and Julien Mozziconacci. Normalization of a chromosomal contact map. *BMC genomics*, 13(1):436, 2012.

[26] Michele Di Pierro, Bin Zhang, Erez Lieberman Aiden, Peter G Wolynes, and José N Onuchic. Transferable model for chromosome architecture. *Proceedings of the National Academy of Sciences*, 113(43):12168–12173, 2016.

[27] Saeed Saberi and Eldon Emberly. Chromosome driven spatial patterning of proteins in bacteria. *PLoS computational biology*, 6(11), 2010.

[28] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilard Pall, Jeremy C. Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19 – 25, 2015.

[29] Xindan Wang, Xun Liu, Christophe Possoz, and David J Sherratt. The two escherichia coli chromosome arms locate to separate cell halves. *Genes & development*, 20(13):1727–1731, 2006.

[30] Romain Mercier, Marie-Agnès Petit, Sophie Schbath, Stephane Robin, Meriem El Karoui, Frédéric Boccard, and Olivier Espéli. The matp/mats site-specific system organizes the terminus region of the e. coli chromosome into a macrodomain. *Cell*, 135(3):475–485, 2008.

[31] A Yu Grosberg, Sergei K Nechaev, and Eugene I Shakhnovich. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de physique*, 49(12):2095–2100, 1988.

[32] Leonid A Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome research*, 19(1):37–51, 2011.

[33] Alexander J Taylor and other SPOCK contributors. `pyknotidknotidentificationtoolkit`. `https://github.com/SPOCKnots/pyknotid`, 2017. Accessed 2020-03.

[34] Julio E Cabrera and Ding J Jin. Active transcription of rrna operons is a driving force for the distribution of rna polymerase in bacteria: effect of extrachromosomal copies of rrnb on the in vivo localization of rna polymerase. *Journal of bacteriology*, 188(11):4007–4014, 2006.

[35] Scott A Scholz, Rucheng Diao, Michael B Wolfe, Elayne M Fivenson, Xiaoxia Nina Lin, and Peter L Freddolino. High-resolution mapping of the escherichia coli chromosome reveals positions of high and low transcription. *Cell systems*, 8(3):212–225, 2019.

[36] Sandra Goetze, Julio Mateos-Langerak, Hinco J Gierman, Wim de Leeuw, Osdilly Giromus, Mireille HG Indemans, Jan Koster, Vladan Ondrej, Rogier Versteeg, and Roel van Driel. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Molecular and cellular biology*, 27(12):4475–4487, 2007.

[37] Talukder Ali Azam, Akira Iwata, Akiko Nishimura, Susumu Ueda, and Akira Ishihama. Growth phase-dependent variation in protein composition of the escherichia coli nucleoid. *Journal of bacteriology*, 181(20):6361–6370, 1999.

[38] Fusheng Guo and Sankar Adhya. Spiral structure of escherichia coli hu$\alpha\beta$ provides foundation for dna supercoiling. *Proceedings of the National Academy of Sciences*, 104(11):4309–4314, 2007.

[39] Mirjana Macvanin and Sankar Adhya. Architectural organization in e. coli nucleoid. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1819(7):830–835, 2012.

[40] HDPP Bremer, Patrick P Dennis, et al. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2(2):1553–69, 1996.

[41] Byung-Kwan Cho, Eric M Knight, Christian L Barrett, and Bernhard Ø Palsson. Genome-wide analysis of fis binding in escherichia coli indicates a causative role for a-/at-tracts. *Genome research*, 18(6):900–910, 2008.

[42] Douglas F Browning, David C Grainger, and Stephen JW Busby. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current opinion in microbiology*, 13(6):773–780, 2010.

[43] Muhammad Malik, Abderazzak Bensaid, Josette Rouviere-Yaniv, and Karl Drlica. Histone-like protein hu and bacterial dna topology: suppression of an hu deficiency by gyrase mutations. *Journal of molecular biology*, 256(1):66–76, 1996.

[44] Christine D Hardy and Nicholas R Cozzarelli. A genetic selection for supercoiling mutants of escherichia coli reveals proteins implicated in chromosome structure. *Molecular microbiology*, 57(6):1636–1652, 2005.

[45] Talukder Ali Azam, Sota Hiraga, and Akira Ishihama. Two types of localization of the dna-binding proteins within the escherichia coli nucleoid. *Genes to Cells*, 5(8):613–626, 2000.

[46] Tamas Gaal, Benjamin P Bratton, Patricia Sanchez-Vazquez, Alexander Sliwicki, Kristine Sliwicki, Andrew Vegel, Rachel Pannu, and Richard L Gourse. Colocalization of

distant chromosomal loci in space in e. coli: a bacterial nucleolus. *Genes & development*, 30(20):2272–2285, 2016.

[47] Subhash C Verma, Zhong Qian, and Sankar L Adhya. Architecture of the escherichia coli nucleoid. *PLoS Genetics*, 15(12), 2019.

[48] Pauline Dupaigne, Nam K Tonthat, Olivier Espéli, Travis Whitfill, Frédéric Boccard, and Maria A Schumacher. Molecular basis for a protein-mediated dna-bridging mechanism that functions in condensation of the e. coli chromosome. *Molecular cell*, 48(4):560–571, 2012.

[49] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *science*, 295(5558):1306–1311, 2002.

[50] Takashi Nagano, Yaniv Lubling, Tim J Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D Laue, Amos Tanay, and Peter Fraser. Single-cell hi-c reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, 2013.

[51] James W Alexander. Topological invariants of knots and links. *Transactions of the American Mathematical Society*, 30(2):275–306, 1928.

[52] Nancy Kleckner, Jay K Fisher, Mathieu Stouf, Martin A White, David Bates, and Guillaume Witz. The bacterial nucleoid: nature, dynamics and sister segregation. *Current opinion in microbiology*, 22:127–137, 2014.

[53] Lisa Postow, Christine D Hardy, Javier Arsuaga, and Nicholas R Cozzarelli. Topological domain structure of the escherichia coli chromosome. *Genes & development*, 18(14):1766–1779, 2004.

[54] Remus T Dame, Fatema-Zahra M Rashid, and David C Grainger. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nature Reviews Genetics*, pages 1–16, 2019.